Int. J. Human–Computer Studies 163 (2022) 102800

Contents lists available at ScienceDirect

International Journal of Human - Computer Studies

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

Investigating user-defined flipping gestures for dual-display phones

Zhican Yang [a,b], Chun Yu [*,a,b], Xin Chen [a], Jingjia Luo [a], Yuanchun Shi [a,b]

[a] Department of Computer Science and Technology, Tsinghua University, Beijing, China
[b] Key Laboratory of Pervasive Computing, Ministry of Education, China

ARTICLE INFO

Keywords:
Dual-display phone
Phone flipping
Mobile interaction
User-defined gestures
Gesture elicitation

ABSTRACT

Flipping is a potential interaction method for dual-display phones with front and rear screens. However, little is known about users' phone flipping behaviors. To investigate it, we iteratively conduct three user studies in this research. We first elicit 36 flipping gestures from 22 users and present a design space according to the results. We then collect users' flipping data and subjective evaluation of all user-defined gestures through the second user study. We design a flipping detection algorithm based on the data collected and deploy it on an off-the-shelf dual-display phone. Another evaluation study shows that it can detect users' flipping efficiently with an average accuracy of 97.78%. Moreover, users prefer many flip-based applications on dual-display phones to existing non-flipping approaches on regular single-screen phones. In conclusion, our work provides empirical support that flipping is an intuitive and promising input modality for dual-display phones and sheds light on its design implications.

1. Introduction

In recent years, smartphone manufacturers and companies have released many dual-display phones with two screens on both the front and back of the device, such as Vivo NEX Dual Display Edition (Fig. 1a), ZTE Axon M, ZTE Nubia X (Fig. 1b), YotaPhone series (Fig. 1c), Hisense A2, and so on. Some foldable phones like Huawei Mate X series (Fig. 1d) also have screens on both sides in the folded state. By utilizing the two screens, dual-display phones can support precise back-of-device interactions (Cui et al., 2021; Shimon et al., 2015; Wobbrock et al., 2008; Wu and Yang, 2020), present more contents, share content with others, and offer users a sense of uniqueness. One of the typical benefits of dual-display phones is implementing regular shots and selfies with only one single high-quality camera (system). Users can simply flip the phone to choose from taking regular shots or selfies (Fig. 2).

Since dual-display phones offer another screen on the back, it is intuitive for users to access it by flipping the phone. Hence, flipping can be served as a potential input modality for dual-display phones with low cognitive cost. However, though flipping the phone has been mentioned in the previous work as one of the user-defined gestures for mobile interaction (Ruiz et al., 2011) or a gesture delimiter (Ruiz and Li, 2011), none had systematically investigated it in literature, especially for dual-display phones, which leaves the following questions:

- What can the dual-display phone flipping be used for? Do users like applications based on flipping?
- How do users flip a dual-display phone? Is there an agreement among users on used flipping gestures with different applications and use cases? What are the users' subjective evaluations of these gestures?
- How does a phone detect users' flipping motions? What are the patterns of various flipping gestures?

To answer these questions, we iteratively conduct three user studies in this research. The remainder of this paper is organized as follows. First, we give a brief review of related work in Section 2. Next, we discuss potential application scenarios for flipping the dual-display phone in Section 3. After that, we collect user-defined flipping gestures through a user-elicitation study and then conduct another user experience study to collect motion sensor data for all elicited flipping gestures and get users' subjective evaluation of these gestures in Section 4. After labeling and analyzing the data collected, we summarize some findings on users' flipping behaviors in Section 5. Also in Section 5, we design a flip detection algorithm based on the findings and deploy it on an off-the-shelf dual-display phone to implement automatic display switching. We then carry out the third user study to evaluate the algorithm's performance and get subjective feedback of different flipping applications, different display switching methods, and different manual

* Corresponding author at: Department of Computer Science and Technology, Tsinghua University, Beijing, China.
E-mail addresses: yang-zc17@mails.tsinghua.edu.cn (Z. Yang), chunyu@tsinghua.edu.cn (C. Yu), chenxin18@mails.tsinghua.edu.cn (X. Chen), luojj18@mails.tsinghua.edu.cn (J. Luo), shiyc@tsinghua.edu.cn (Y. Shi).

https://doi.org/10.1016/j.ijhcs.2022.102800
Received 13 July 2021; Received in revised form 16 February 2022; Accepted 18 February 2022
Available online 25 February 2022
1071-5819/

intervention methods if the phone fails to detect flip in Section 6. Finally, we discuss other topics related to phone flipping and directions of future work in Section 7 and conclude the paper in Section 8.

In sum, our contributions and main findings are the following:

- We present 36 distinct user-defined flipping gestures for dual-display phones, which can be grouped into 20 categories (Fig. 6).
- The average time duration of flipping is 1144.96 ms. Among all gestures, "Wrist" gestures (flipping gestures that mainly rotate the wrist) are the fastest ($M = 866.96$ ms). On the contrary, the slowest ones are one-handed non-"Wrist" gestures ($M = 1448.43$ ms).
- Users' subjective evaluations of flipping gestures are largely in line with the time costs. Meanwhile, users perceive more differences from gesture types than which hand is involved (dominant hand and non-dominant hand).
- We present a straightforward data-driven flip detection algorithm based on the findings. We also introduce a "turn on both screens during flipping" mechanism to reduce the perceived latency.
- Automatic display switching takes almost half the time of the existing manual switching methods. Therefore, users generally prefer it and like many flip-based interactions and applications.

For convenience, phones mentioned in the remainder of this paper all refer to dual-display phones unless otherwise specified. Also, we will use the term "front screen" to refer to the screen currently facing the user and "rear screen" to refer to the screen that is not.

2. Related work

2.1. Phone flipping and tilting in previous work

Ruiz et al. (2011) investigated motion gesture design for the mobile computing paradigm through a 20-participant guessability study (Wobbrock et al., 2005) that elicits end-user motion gestures to invoke commands on a smartphone device. Flipping the phone was one of the user-defined gestures, which was used to hang up a call. In addition, users designed gestures that quickly rotate the phone around the y-axis (Fig. 3) to the left/right and return back for "pan left/right" tasks. Ruiz and Li (2011) also used the DoubleFlip gesture (flipping the phone around the y-axis and then back) as an always-active input delimiter for mobile motion-based interaction. Their experiments showed that the DoubleFlip gesture is highly resistant to false positive conditions while still achieving a high recognition rate. It indicates that smartphones may be able to detect phone flipping with embedded sensors, though the DoubleFlip gesture consists of two flipping motions in sequence. These use cases inspired us that flipping motion could be used as an input trigger for various applications. We will further push forward the idea to dual-display phones and discuss other use cases in Section 3.

Comparing to flipping, tilting is much more common for regular single-screen phones and has been researched by many researchers. For example, Rahman et al. (2009) explored tilt-based interaction using a mobile device. Results showed that users could comfortably perform up to 16-levels wrist-based input on the pronation/supination axis. Chang et al. (2015) investigated users' touch behavior on large mobile device touchscreens and then utilized tilting gestures to assist one-hand targeting. Furthermore, Eardley et al. (2017) investigated how form factors impact hand usage and movement on mobile phones, including tilt and rotation effects.

2.2. New input modality and interaction for smartphones

Since the smartphone is in users' hands most of the time, many researchers have studied and utilized users' grasps and finger placements as active gestures to trigger an action or passive information to adapt an interface. For instance, Wimmer and Boring (2009) deployed capacitive sensors on the left and right of their HandSense prototype, which correctly classify over 80 percent of all touches, discriminating six different ways of touching the device. Furthermore, Le et al. (2016) investigated finger placement and hand grasp during smartphone interaction. They then presented InfiniTouch (Le et al., 2018c), a system that enables touch input on the whole device surface with capacitive sensors on the front, the back, and three sides. Instead of adding capacitive sensors, Quinn et al. (2019) designed squeeze gestures for the Google Pixel 2 by utilizing strain gauge elements adhered to the inner sidewall of it. Besides, Goel et al. (2012) introduced GripSense, a system that only leverages mobile device touchscreens, built-in inertial sensors, and vibration motor to infer hand postures with 84.3% accuracy.

On the other hand, touchscreens are the most common and successful input method for smartphones. Le et al. (2018a) extended the touch input vocabulary by the palm and presented PalmTouch (Le et al., 2018b), an additional input modality for smartphone interaction that differentiates between touches of fingers and the palm. Xiao et al. (2015) and Mayer et al. (2017) extended the input richness of touchscreens by estimating and utilizing the finger orientation and angle on commodity touchscreens. Yang et al. (2019a) utilized the phone as the input device for indirect touch gesture typing. Liang et al. (2012) combined touch and motion gestures to support 3D manipulation of objects at a distance with a dual-surface concept device.

Since we also explore an input modality new to users (i.e., flipping) in this research, we learn from previous work when designing different parts of our studies, such as how to develop a user-defined gesture set (Ruiz et al., 2011) as well as the algorithm sensitivity adjustment (Quinn et al., 2019).

3. Application scenarios for flipping

We conducted a 16-person brainstorming session (mostly HCI researchers/students and some smartphone practitioners / application developers) to discuss the application scenarios for flipping. We roughly divide the results into three usages as follow and explain some design ideas:

1. Multitasking: dual-display phones have one more screen than regular phones. Taking advantage of this, we can utilize both screens to run

(a) Vivo NEX Dual Display (b) Nubia X (c) YotaPhone 3 (d) Huawei Mate Xs

Fig. 1. Dual-display phones and foldable phones with front and rear screens.

different applications. Typical scenarios would include switching between chatting and watching videos: when a message is coming while watching a video, users can flip the phone to switch to the chatting application automatically. Meanwhile, the video is paused. After replying to the message, flipping the phone again to continue watching the video. Another instance is that flipping to the back of the phone will activate a user-predefined application, e.g., a mirror application or a weather application. Besides application switching, flipping the phone can also be used to change appearance and contents (e.g., change wallpapers of the lock screen, switch between casual mode and business mode, temporarily hide private messages, etc.)

2. Using flipping as a trigger: as mentioned in the introduction, flipping itself can serve as a new input modality for dual-display phones. For example, users can flip the phone to reject an incoming call during a business meeting or delay the alarm for 5 min while sleeping. In both scenarios, users can flip the phone in an eyes-free manner (i.e., no need to look at the phone), and these scenarios are not only for dual-display phones but also for regular phones. As for dual-display phones only, flipping can also be used to switch between taking selfies and regular shots when taking photos. Instead of embedding two cameras (front and rear), dual-display phones can utilize only one single high-quality camera (or camera system) to implement both taking regular shots (Fig. 2a) and selfies (Fig. 2b).
3. Viewing the other side of objects: we all know that we can see the back of an object by flipping it. Hence, it is intuitive and natural that flipping the dual-display phone will display the other side of objects. For example, users can flip the phone to view the back of clothes when shopping online (Fig. 4a and b), or see the meaning/translation of a word when memorizing words (Fig. 4c and d), or view the back of an e-card or an envelope. In addition, based on flipping the phone to change different views, we can also design puzzle video games like Fez (Wikipedia, 2021).

Most use cases mentioned above require the phone to support an automatic display switch when flipping (the rear screen before flipping will become the front screen which should be turned on after flipping, and vice versa). We will discuss how to detect phone flipping in Section 5.

4. User-defined flipping gestures and user preference

In this section, we first carried out a user-elicitation study to collect flipping gestures. We then presented a representative gesture vocabulary in the phone flipping design space based on the results. Finally, we conducted a follow-up user experience study to get users' subjective feedback on these gestures.

4.1. Gesture elicitation

4.1.1. Apparatus

We used Vivo NEX Dual Display Edition (Model V1821A) (VIVO, 2021) as the experimental device in this study and in the subsequent studies. The phone dimensions are 157.2 × 75.3 × 8.1 mm (6.19 × 2.96 × 0.32 in), with a 6.39-inch main display and a 5.49-inch secondary display. It weighs 199 g (7.02 oz). The running OS was Android 10.0.

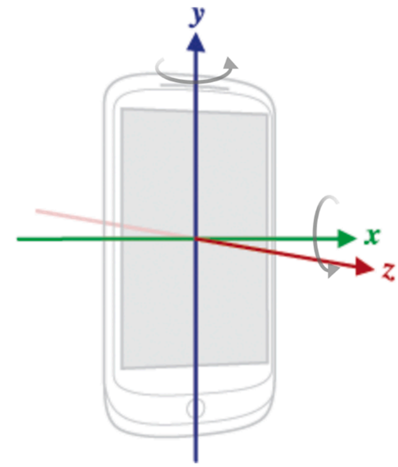


Fig. 3. Definition of the coordinate system. The grey arrows indicate counterclockwise direction (The z-axis is not used when considering flipping).

4.1.2. Study design and procedure

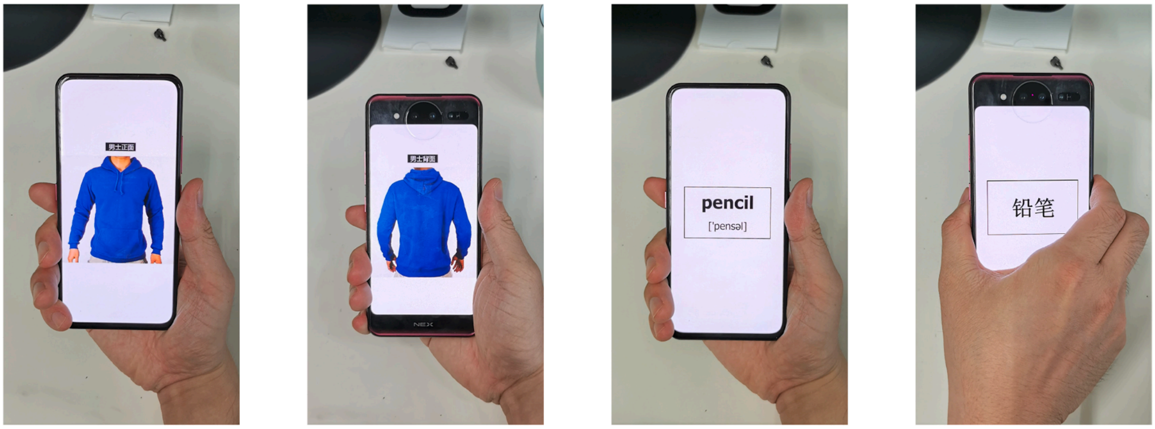
We first gave the participants a brief introduction to the dual-display phone and the idea of flipping. In order to get closer to real usage, participants in this study would design and perform flipping gestures while experiencing various typical application scenarios listed in Table 1. The scenarios were selected from the discussion in Section 3 and could cover all combinations of phone orientations before and after flipping (i.e., portrait, landscape, and landscape rotated 180° around the z-axis, see Fig. 3). The last session, "10. Free", does not include any application scenarios and allows participants to flip the phone at will freely.

We instructed participants to focus on gesture design by assuming the phone would be capable of recognizing any kinds of flipping gestures (Tu et al., 2020). Correspondingly, the experimental software in this study was implemented with a Wizard-of-Oz design: there is no flipping recognition algorithm; both screens of the phone were always on in this study. We also encouraged participants to design as many gestures as possible and explain their designs (think-aloud). Other than these instructions, we would not give any suggestions or ask any questions to participants to avoid bias from experimenters.

A video camera recorded the process of study (Fig. 5). We informed participants that only their hands and the phone were captured, and the recorded videos were for research purposes only.

(a) Taking a regular shot when using the primary display. (b) Taking a selfie when using the second display.

Fig. 2. A Dual-display phone can utilize only one single high-quality camera (system) to support both regular shots and selfies. Flipping the phone switch between taking a regular shot and a selfie.



(a) Online shopping (b) Online shopping (c) Memorize words (d) Memorize words

Fig. 4. Two examples of flipping the phone to view the other side of objects. Fig (a) and Fig (b) show an example that the user flips the phone in hand and keeps the same grip before and after flipping. Fig (c) and Fig (b) show an example that the user flips the phone mainly by rotating the wrist.

Table 1
Application scenarios used in the user-elicitation study.

Session	Phone	Orientation before/after flipping
1–3. Switch between chatting and watching video		Portrait ⇔ Portrait Portrait ⇔ Landscape Portrait ⇔ Landscape (180°)
4 and 5. Swap between two videos / Change videos	In hand	Landscape ⇔ Landscape Landscape ⇔ Landscape (180°)
6. Memorize words		Portrait ⇔ Portrait
7. View the back of commodity clothes		Portrait ⇔ Portrait
8. Reject an incoming call	On table	Portrait ⇔ Portrait
9. Change lock screen's appearance		Portrait ⇔ Portrait
10. Free	Any	Any

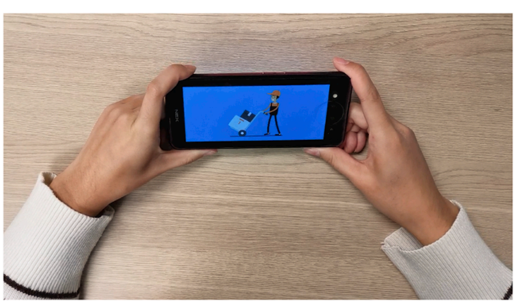


Fig. 5. An example picture of recorded video. The participant was watching a simple video in landscape mode.

4.1.3. Participants

We recruited 22 participants (P1–P22, 11 female, 11 male) from the campus, aged from 19 to 27 years ($M = 21.3$, $SD = 2.2$). All participants were right-handed. Their hand lengths (measured from the middle of inter stylion to the tip of the middle finger) ranged from 160 cm to 204 cm, which comprised samples from the 1tst to 85th percentile of the data reported in an anthropometric survey (Gordon et al., 2014). They have been using smartphones for 5.5 to 13.0 years ($M = 8.03$, $SD = 1.59$). The display of their daily-used smartphones ranged from 4.70-inch to 6.89-inch ($M = 6.16$, $SD = 0.58$). Among them, 17 participants were new to dual-display phones. The study lasted for 5–10 min. Participants were compensated 10 CNY (about 1.5 USD) for their time.

4.1.4. User-defined gestures

According to the videos, each participant designed an average of 16.86 ($SD = 6.26$) distinct flipping gestures, respectively. We divided all of the user-elicited flipping gestures into 20 categories (Fig. 6). There are four main factors for flipping:

1. Hand(s) holding the phone before and after flipping, including left hand (L), right hand (R), and both hands (B).
2. Phone orientations before and after flipping, including portrait (P) and landscape (L). Note that we do not distinguish the two landscape orientations here (0° and 180° around the z-axis) because the next factor (rotation direction) could cover it.
3. Rotation direction and which axis the rotation is around: here, we used the coordinate system (Fig. 3) defined in Android API (Google, 2021b). The rotation direction includes clockwise(CW) and counterclockwise(CCW). The rotation is around the y-axis by default, or around the x-axis if specified. The z-axis is not used when considering flipping.
4. Special cases, including "Wrist" and "OnTable": "Wrist" means the flipping is achieved mainly by rotating the wrist. The phone is held tightly during flipping and the grip postures are relatively steady (e.g., see "10. RR-PL-CW-Wrist" in Fig. 6 for example). "OnTable" means the phone is placed on the table before and after flipping.

We used the above four factors to name each gesture. Taking "16a. LR-PP-CCW" in Fig. 6 as an example, it means the phone is held in the left hand before flipping and in the right hand after flipping. The phone is in portrait orientation before and after flipping. The rotation direction is counterclockwise around the y-axis (Fig. 3).

The gesture set shown in Fig. 6 is a subset of user-performed gestures. Gestures only performed by one participant (e.g., LR-PP-CW) and other low-frequency gestures that are difficult to perform (e.g., RR-PL-CW, designed by two participants) were discarded. In addition, some bimanual symmetric flipping gestures are grouped into one category in Fig. 6 (e.g., "5a. BB-PP-CCW" and "5b. BB-PP-CW"). After filtering, there are 36 gestures grouped into 20 categories in total. Among all gestures, "7.BB-LL-CW" and "18b.RR-PP-CCW-OnTable" have been defined by the most participants.

4.1.5. Agreement rate

Agreement rate (*AR*) (Findlater et al., 2012; Vatavu and Wobbrock, 2015) measures the homogeneity for nominal data (Vuletic et al., 2021). We calculated *AR* for each session listed in Table 1 to evaluate the degree of agreement among flipping gestures elicited from different participants, with the following formula:

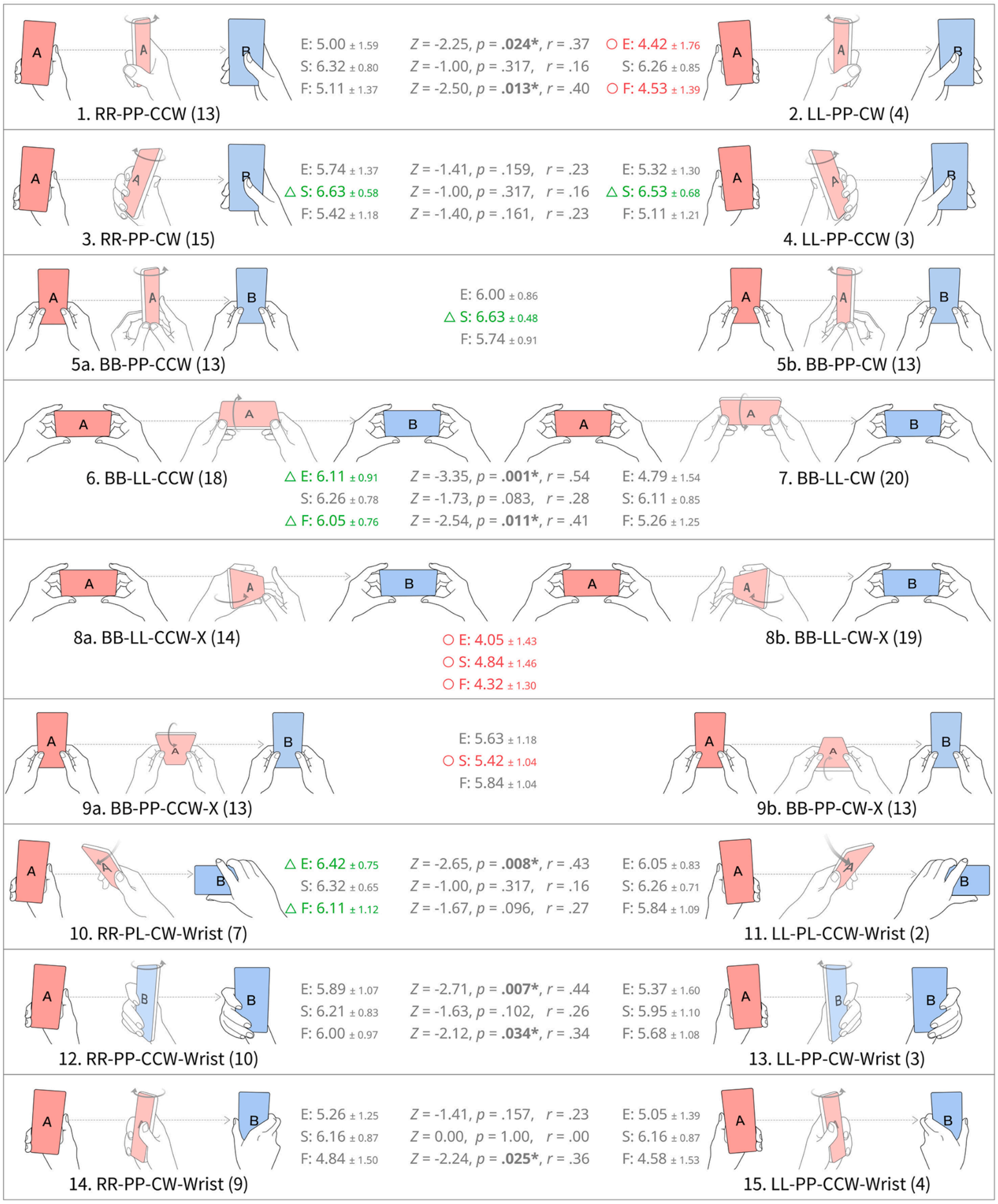


Fig. 6. User-defined flipping gestures. The number in parentheses indicates how many participants have designed this gesture. Users' subjective ratings on Easiness (E), Social Acceptability (S), and Fatigue (F) are also shown in the form of mean ± *SD*, which were collected through a subsequent user study and will be discussed in detail in Section 4.2. The top three and bottom three scores of each criterion are hightlighted in red and green with circle and triangle markers respectively. Results of Wilcoxon signed-rank tests are also shown ($p < 0.05$ are highlighted in bold). Symmetrical gestures colored in grey are omitted and not displayed. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

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

where P is the set of all user-defined gestures for session s, and P_i is a subset i of identical gestures from P[1] According to Vatavu and Wobbrock (2015), the interpretation of AR is: a low agreement when $AR \leq 0.1$, a medium agreement when $0.1 < AR \leq 0.3$, a high agreement when $0.3 < AR \leq 0.5$, and a very high agreement when $AR > 0.5$. The results are shown in Fig. 7.

[1] We count by gesture here, not by category. For example, "5a. BB-PP-CCW" and "5b. BB-PP-CW" are counted as different gestures.

According to Fig. 7, the agreement rates for sessions can be grouped into four tiers:

1. Sessions 4 and 5 ($AR \approx 0.5$): participants achieved a near very high agreement when switching between videos. Actually, most participants designed and only designed "6. BB-LL-CCW" and "7. BB-LL-CW" for session 4, "8a. BB-LL-CCW-X" and "8b. BB-LL-CW-X" for session 5, respectively. It shows that users tend to hold the phone in both hands when flipping in landscape orientations because, just as P6 and P10 said, *"it is difficult to flip with one hand."* Correspondingly, the flipping gestures are highly consistent among users, only differ in the direction of rotation.

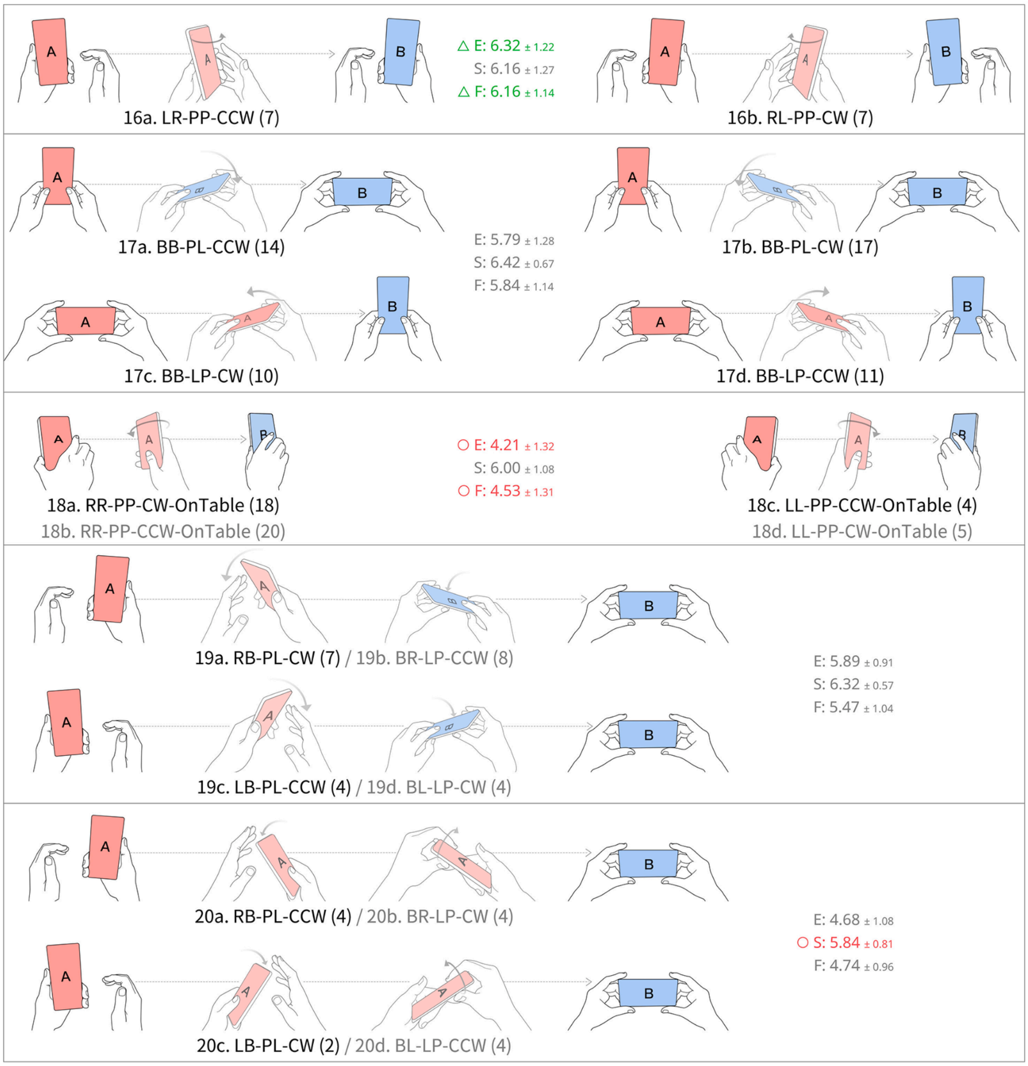


Fig. 6. (*continued*).

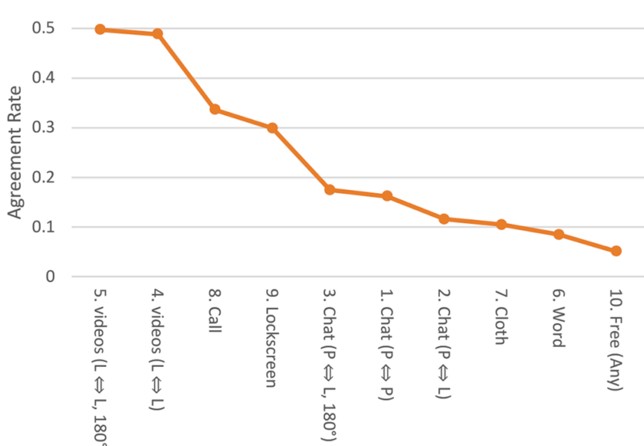


Fig. 7. Agreement rate for each session (L and P are shorts for landscape and portrait).

2. Sessions 8 and 9 ($AR \approx 0.3$): the phone is placed on the table in both sessions. According to Fig. 6, flipping gestures using the right hand ("18a.", "18b.") are performed much more often than gestures using the left hand ("18c.", "18d."), which is in line with the fact that all participants are right-handed. Moreover, few participants used both hands for flipping. As P13 stated, *"using the desktop to support the phone during flipping, flipping with one hand is even more convenient than two hands."* As a result, participants reached a high agreement in sessions 8 and 9.
3. Sessions 1–3 ($0.1 < AR < 0.2$): participants reached a medium agreement when switching between chatting and watching a video. The most often defined gestures for sessions 1, 2, and 3 are "3.RR-PP-CW", "17b. BB-PL-CW", and "17a. BB-PL-CCW", respectively. We observed that participants would actually flip the phone in hand to get a comfortable hand posture after flipping instead of just rotating the wrist. *"I need to type on the screen after flipping"* (P3). *"(I design this flipping gesture) to avoid blocking the screen with my hands"* (P13, P21). As a result, no participant designed "Wrist" gestures in sessions 1–3, which is different from the next group (sessions 6 and 7).
4. Sessions 6, 7, and 10 ($AR \leq 0.1$): participants designed the most various gestures in these sessions. Other than gestures mentioned above, participants also designed "Wrist" gestures: *"Since I only need to view the back briefly, turning my wrist to read word translation and then quickly turning back is more convenient than flipping the phone in my hand"* (P9, P17). Hence, on one hand, the most often defined gestures for sessions 6 and 7 were "12.RR-PP-CCW-Wrist" and "14. RR-PP-CW-Wrist". On the other hand, almost all of the "Wrist" gestures of this study were designed in sessions 6 and 7.

The last session, “10. Free”, has the lowest *AR* as expected. “1.RR-PP-CCW”, “3. RR-PP-CW”, “5a. BB-PP-CCW”, and “7. BB-LL-CW” are most often designed.

4.2. User preference

After the user-elicitation study, we conducted a follow-up user experience study to get users’ subjective evaluation of each gesture shown in Fig. 6.

4.2.1. Study design and procedure

In this study, we asked participants to use each gesture in Fig. 6 to flip the phone and then give subjective ratings.

Fig. 8 shows the experimental setting. Participants sat at the table with the dual-display phone in hand. Similar to the previous elicitation study, the process of this study was also recorded with the camera of another mobile phone.

According to a pilot study, users perceived little difference between some symmetric flipping gestures, especially for bimanual symmetric ones. Therefore, in order to reduce the time of the formal study and the total number of rating items, some symmetric flipping gestures were grouped into the same session (i.e., be performed and rated together). As a result, there were 20 sessions in the study, each consist one gesture category (name from 1. to 20. in Fig. 6).

At the beginning of each session, participants first watched a demonstration video of the flipping gesture and then imitated it. After learning the gesture, participants tapped the START button of the experimental program on the phone to start this session’s recording. After 1 s, the phone would vibrate for 0.5 s to inform the participant to start flipping, and then the participant performed the required gesture to flip the phone. After flipping, participants should hold the phone relatively steadily for at least 0.5 s, and then another vibration would come in 1 s. This process would repeat ten times, which means participants would perform the flipping gestures ten times in each session. This design was derived from the pilot study, allowing participants to focus on flipping without counting the number of flips. The stability testing was achieved by the phone’s inertial measurement units (mainly gyroscopes). Before the formal study, participants would take a test session to familiarize themselves with the study procedure and the dual-display phone.

During the study, the phone logged values and timestamps of all supported motion sensors (Google, 2021a). These data were collected for subsequent statistical analysis of flipping and for designing flip detection algorithm, which will be discussed in the next section, Section 5.

If a session contains two symmetric gestures (e.g., “16a. LR-PP-CCW” and “16b. RL-PP-CW”), both gestures would be performed five times alternately (i.e., 16a, 16b, 16a, 16b...). If a session contains four symmetric gestures (e.g., 20a–20d), participants could perform the gestures in any combination at will, ten times in total.

One of the objectives of this study is to investigate the difference between flipping with the dominant hand and the non-dominant hand. Therefore, sessions containing gesture categories in the same row in Fig. 6 would be arranged successively (mostly symmetrical and similar gestures performed by the left and right hand respectively, e.g., “1. RR-PP-CCW” and “2. LL-PP-CW”). Other than that, the order of sessions was counterbalanced across participants to minimize the carryover learning effect. Participants could take a break between sessions.

At the end of each session, after flipping ten times, participants rated the gesture through a questionnaire which contains three criteria on a 7-point Likert scale (1 - worst, 7 - best) (Likert, 1932), with the descriptions as follows:

- **Easiness**: “Can you perform the gesture easily?”.
- **Social Acceptability**: “Does the gesture look strange from others? Is it easy to attract attention? Can others understand and accept that you are flipping the phone?”
- **Fatigue**: “Do you feel tired after performing the gesture repeatedly?”

These subjective criteria have been commonly used in gesture evaluation studies (Hsieh et al., 2016; Kim and Xiong, 2021; Lu et al., 2020; Serrano et al., 2014; Xu et al., 2020a; Yang et al., 2019b). As a 20-session study, participants were allowed to modify previously given ratings after experiencing and comparing different flipping gestures. We also encouraged participants to make think-aloud comments when rating.

4.2.2. Participants

We recruited 19 participants (P1–P19, 10 female, 9 male) from the campus, aged from 19 to 25 years ($M = 21.0$, $SD = 1.6$). Almost all (17) participants have participated the previous gesture elicitation user study. All participants were right-handed. Their hand lengths ranged from 160 cm to 203 cm, which comprised samples from the 1tst to 85th percentile of the data reported in an anthropometric survey (Gordon et al., 2014). They have been using smartphones for 5.5 to 9.5 years ($M = 7.59$, $SD = 1.12$). The display of their daily-used smartphones ranged from 4.70-inch to 6.83-inch ($M = 6.18$, $SD = 0.57$). Among them, 15 participants were new to dual-display phones. The study lasted for about 60 min. Participants were compensated 100 CNY (about 15 USD) for their time.

4.2.3. Overall results

All participants were able to perform each flipping gesture. The ratings of all 20 gesture categories are also shown in Fig. 6. The average ratings of all gestures were 5.40 ($SD = 0.71$) on **Easiness**, 6.14 ($SD = 0.40$) on **Social Acceptability**, and 5.36 ($SD = 0.60$) on **Fatigue**. Friedman tests found significant effects of gesture category on the ratings of **Easiness** ($\chi^2(19) = 116.89$, $p < 0.001$), **Social Acceptability** ($\chi^2(19) = 73.39$, $p < 0.001$), and **Fatigue** ($\chi^2(19) = 100.36$, $p < 0.001$).

4.2.4. Dominant hand vs. non-dominant hand

To investigate differences in users’ perceptions of flipping with the

Fig. 8. Experimental setting of the user experience study. The participant was watching the demonstration video.

dominant hand and the non-dominant hand, we conducted Wilcoxon signed-rank tests to compare five pairs of symmetric unimanual flipping gestures (1.–4. and 10.–15.) respectively. Results are also shown in Fig. 6. In most cases, there is a significant difference either on **Easiness** or on **Fatigue**. It indicates that users generally agreed that flipping with the non-dominant hand is more difficult and tiring. The only exception is the pair of "3. RR-PP-CW" and "4. LL-PP-CCW". As P14 said, *"Using my thumb to push the edge of the phone (to flip the phone) seems no difference with my left hand and right hand."*

On the other hand, no significant difference was found for all symmetric unimanual gesture pairs in terms of **Social Acceptability**. *"As long as I don't drop the phone, there is no difference which hand I use to flip the phone"* (P6, P18). P3 echoed that *"It's as if people don't care whether you hold the phone with your left hand or right hand."*

To get further insight, we ran a 2×5 within-subjects test on the five pairs of symmetric gestures, with a 2-level factor *Hand* (dominant and non-dominant) and a 5-level factor *Gesture*. We used the aligned rank transform tool (ARTool) (Wobbrock et al., 2011) developed by Kay and Wobbrock (2021) to align and rank data for nonparametric ANOVAs. Table 2 shows the results, where we used the partial eta squared (η_p^2) to report effect sizes (explained in Appendix A). As expected, we observed significant main effects of *Hand* on both **Easiness** and **Fatigue**. However, according to the effect sizes, *Gesture* played a more significant role than *Hand* in all criteria, i.e., participants did feel the difference when flipping with the dominant hand and non-dominant hand, but they perceived more differences from gesture types rather than which hand is involved. *Hand* $\times$ *Gesture* interaction effects were not significant in all three criteria.

4.2.5. Other findings

We sorted all 20 gesture categories by their total scores (Fig. 9). Among all gestures, "10. RR-PL-CW-Wrist", "16.LR-PP-CCW", and "6. BB-LL-CCW" are the Top-3 favorite gestures. In addition, all of the Top-7 gestures (35%) are either gestures performed by both hands or gestures mainly rotating the wrist. Most participants agreed that these gestures are easier to perform, e.g., *"I felt more confident (in successfully flipping) when using both hands"* (P1, P10, P14), *"It ('Wrist' gesture) is easy to perform since I can hold the phone tightly when flipping."* (P11). However, not all bimanual gestures were favored by participants, e.g., "8. BB-LL-CCW-X" received the lowest score. Participants reported that flipping the phone around the *x*-axis is much harder than flipping around the *y*-axis (P3, P6, P8). "9. BB-PP-CCW-X" is a bit better than "8. BB-LL-CCW-X" because, as P1 said, *"when holding the phone in portrait orientation with both hands, I can move the phone slightly down in my hands, pinch the two ends (of the x-axis) between my thumbs and forefingers, and then flip it (around the x-axis)."*

As mentioned in Section 4.2.1, users perceived little difference between most bimanual symmetric gestures. However, the pair of "6.BB-LL-CCW" and "7.BB-LL-CW" is an exception. According to the results of Wilcoxon signed-rank tests shown in Fig. 6, the ratings of "6. BB-LL-CCW" are significantly higher than the ratings of "7. BB-LL-CW" on **Easiness** and **Fatigue**. The difference is that when performing the former, users would use their thumbs to push the bottom edge of the phone from the back of the phone; but when performing the latter, the users' thumbs would instead reach the top edge of the phone and pull it. As also described in Section 4.2.4, users preferred to push the phone from behind rather than stretching their thumbs first to reach the phone and then pull it. Similar conclusions would also be reached when comparing "1. RR-PP-CCW" and "3. RR-PP-CW" (**Easiness**: $Z = -2.507, p < 0.05, r = 0.41$; **Fatigue**: $Z = -1.778, p = 0.075, r = 0.29$, trending), as well as comparing "2. LL-PP-CW" and "4. LL-PP-CCW" (**Easiness**: $Z = -2.602, p < 0.01, r = 0.42$; **Fatigue**: $Z = -2.391, p < 0.05, r = 0.39$)

Table 2
Results of repeated measures ANOVAs using ARTool.

Criterion	Factor	$F_{1,162}/F_{4,162}$	p	η_p^2
Easiness	*Hand*	4.73	**0.031***	0.0284
	Gesture	7.02	**<0.0001***	0.1477
	Hand × *Gesture*	0.06	0.994	0.0014
Social Acceptability	*Hand*	1.41	0.237	0.0086
	Gesture	2.57	**0.040***	0.0597
	Hand × *Gesture*	0.18	0.950	0.0044
Fatigue	*Hand*	5.42	**0.021***	0.0324
	Gesture	9.79	**<0.0001***	0.1947
	Hand × *Gesture*	0.19	0.941	0.0048

5. Flipping data analysis and flip detection algorithm

In this section, we first analyzed motion data collected in the previous user study. We then designed a flip detection algorithm based on the findings.

5.1. Data labeling and processing

As mentioned in Section 4.2.1, we collected raw data from motion sensors when participants performed different flipping gestures. Since participants flipped ten times in each session, we needed to segment the data and label the beginning and end of each flip. We first used a simple motion detection program to get coarse-grained annotations by segmenting data (flipping and not flipping). We then manually checked and adjusted the annotations with the help of data visualization (or relabeled them if needed). Fig. 10 shows an example of labeling. The two sensors shown in Fig. 10, gravity and gyroscope, are the most frequently referred to when labeling.

After labeling, the time durations of each flipping action can be easily calculated. We removed 44 outliers (1.16% of the data) that were more than three times the standard deviation away from the mean of its gesture category. As a result, up to 5 data were removed for each gesture category.

5.2. Time duration

The average time duration of all flipping gesture categories was 1144.96 ms (*SD* = 397.18 ms). Fig. 11 shows the mean time durations of every gesture category. We ran a repeated measures ANOVA (RM-ANOVA) with the single 20-level factor *Gesture Category*. Mauchly's Test of Sphericity indicated that the assumption of sphericity had been violated ($\chi^2(189) = 655.13, p < 0.0001$). Since the Univar Greenhouse-Geisser Epsilon ($\varepsilon = 0.322$) is below 0.75 (Verma, 2015), we reported the results with the Greenhouse-Geisser correction: $F_{6.109,109.963} = 33.656, p < 0.0001, \eta_p^2 = 0.652$. There is a significant main effect of *Gesture Category* on gesture duration. The slowest gesture ("2. LL-PP-CW": $M = 1593.44$ ms, $SD = 292.44$ ms) takes almost twice as long as the fastest one ("10. RR-PL-CW-Wrist": $M = 815.18$ ms, $SD = 217.92$ ms).

According to Fig. 11, all of the 6 "Wrist" gestures were the fastest ($M = 866.96$ ms, $SD = 279.56$ ms). As P11 stated, *"I don't need to change grip posture during flipping."* P1 echoed that *"Not only my wrist but also my arm will rotate; this makes phone flipping more easily and quickly."* Behind the "Wrist" gestures, two-handed gestures constitute the second tier ($M = 1127.98$ ms, $SD = 329.02$ ms), except "8. BB-LL-CCW-X" ($M = 1423.10$ ms, $SD = 283.13$ ms). The one-handed non-"Wrist" gestures were the slowest ($M = 1448.43$ ms, $SD = 397.11$ ms). Among them, "3. RR-PP-CW" and "4. LL-PP-CCW" were significantly faster than "1. RR-PP-CCW" and "2. LL-PP-CW", respectively (1. vs. 3.: $F_{1,18} = 21.03, p < 0.0005, \eta_p^2 = 0.539$; 2. vs. 4.: $F_{1,18} = 14.42, p < 0.005, \eta_p^2 = 0.445$). These results are all consistent with users' subjective evaluations in Section 4.2.5.

However, different from the users' subjective feelings, there is no significant difference between flipping with the dominant and non-dominant hand on time durations ($F_{1,18} = 0.398$, *n.s.*, $\eta_p^2 = 0.022$).

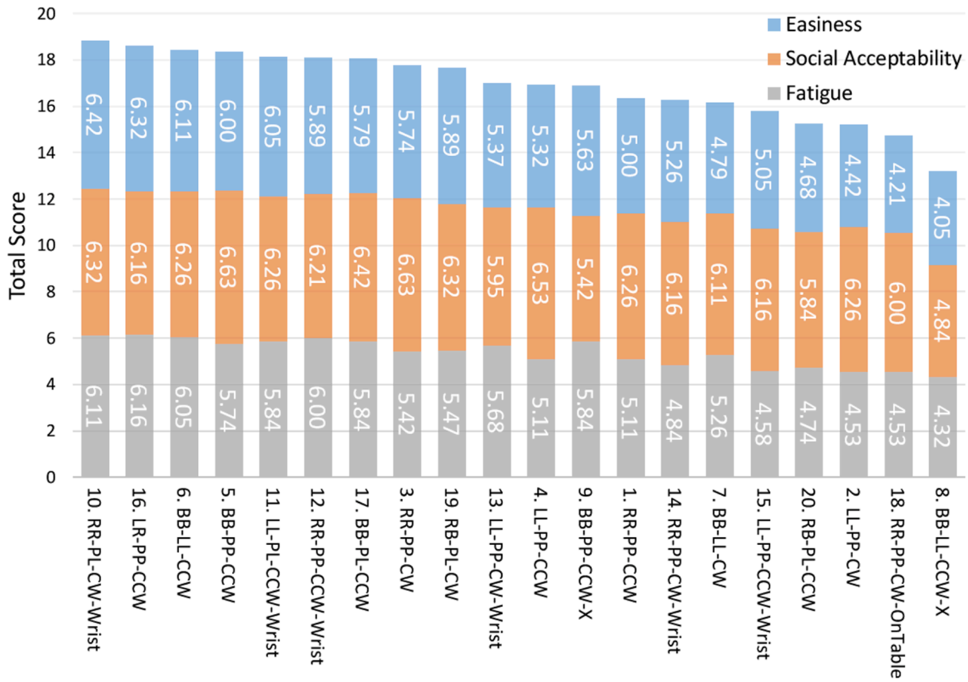


Fig. 9. Subjective ratings of all 20 gesture categories in terms of easiness, social acceptability, and fatigue (rate from 1 to 7, the higher the better), sorted by total score.

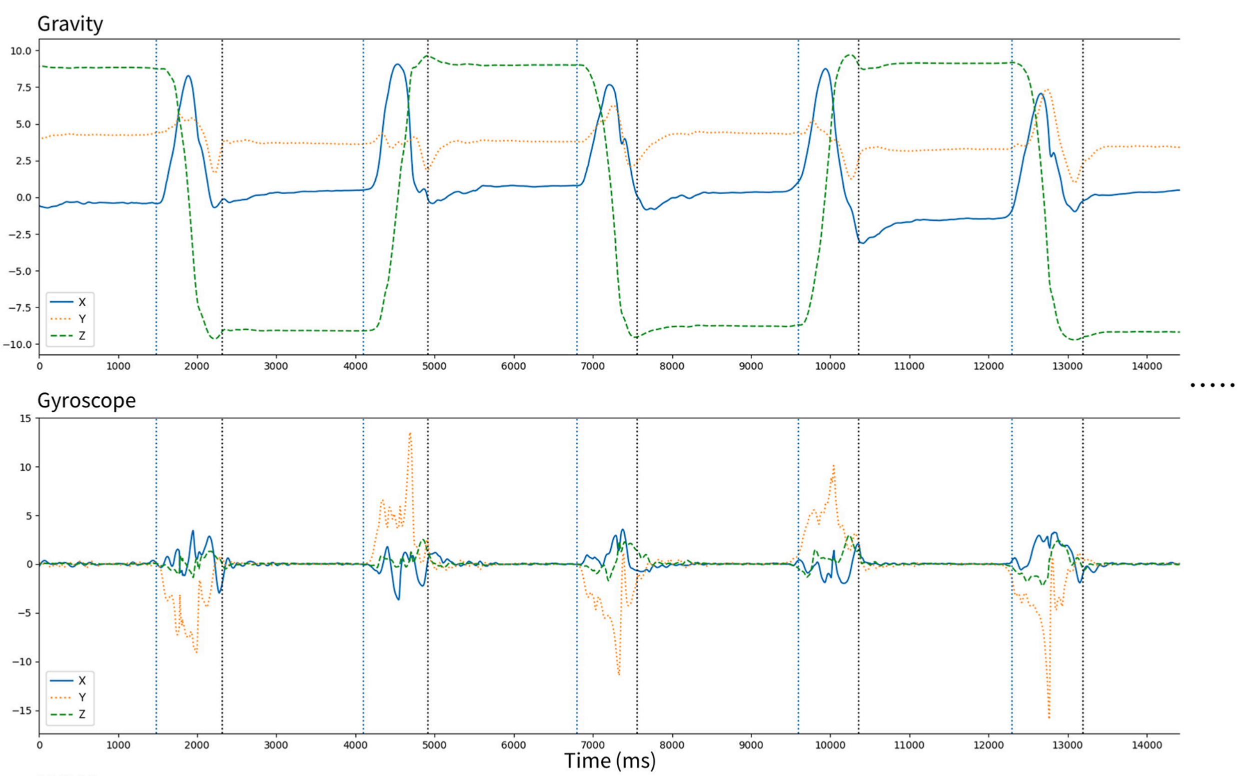


Fig. 10. Data visualization and labeling. The example picture here shows the motion sensor data from one participant's "16a. LR-PP-CCW" and "16b.RL-PP-CW" session. The vertical dotted lines indicate the labeled beginning/end of a flipping action. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 12 shows the distribution. According to Figs. 11 and 12, participants spend about the same amount of time performing symmetric unimanual gestures with the dominant and non-dominant hand.

5.3. *Gyroscope data*

When considering flipping, the gyroscope is naturally the most important sensor. Based on which axis to flip around, we classified the

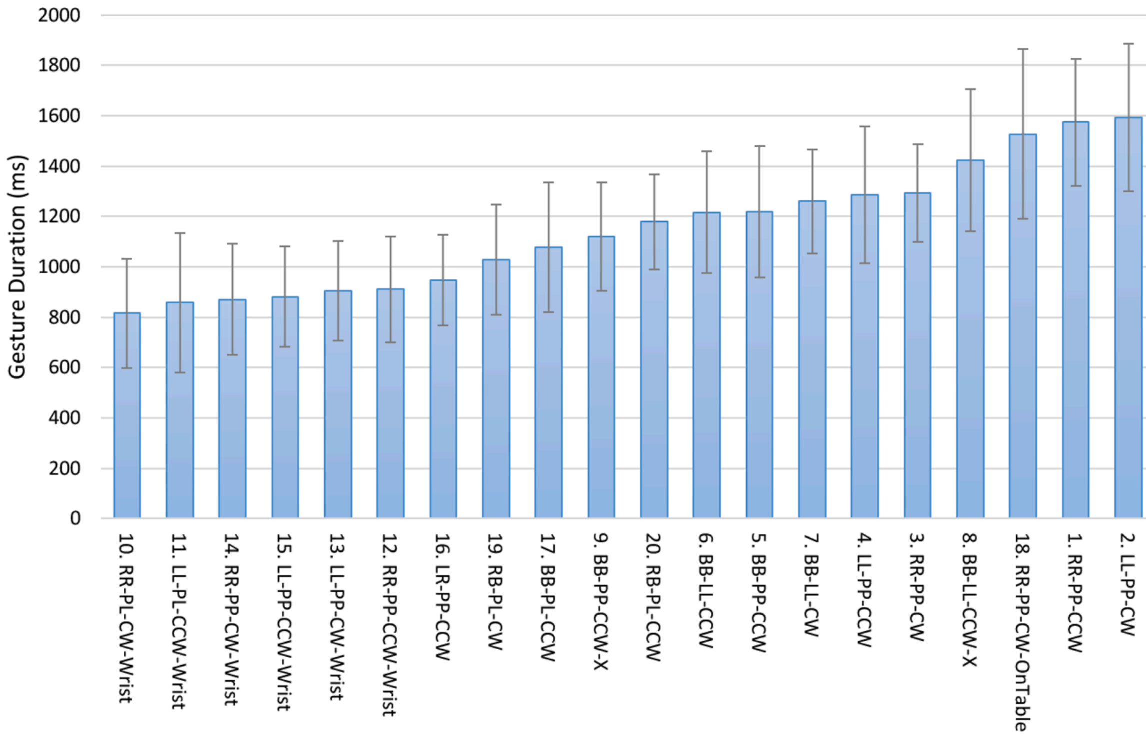


Fig. 11. The mean time durations of all 20 flipping gesture categories. Error bars indicate one standard deviation applying Bessel's correction.

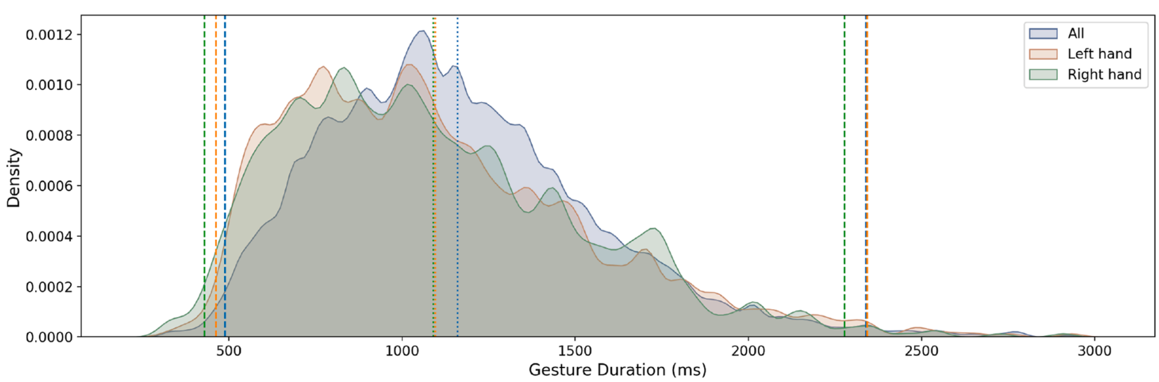


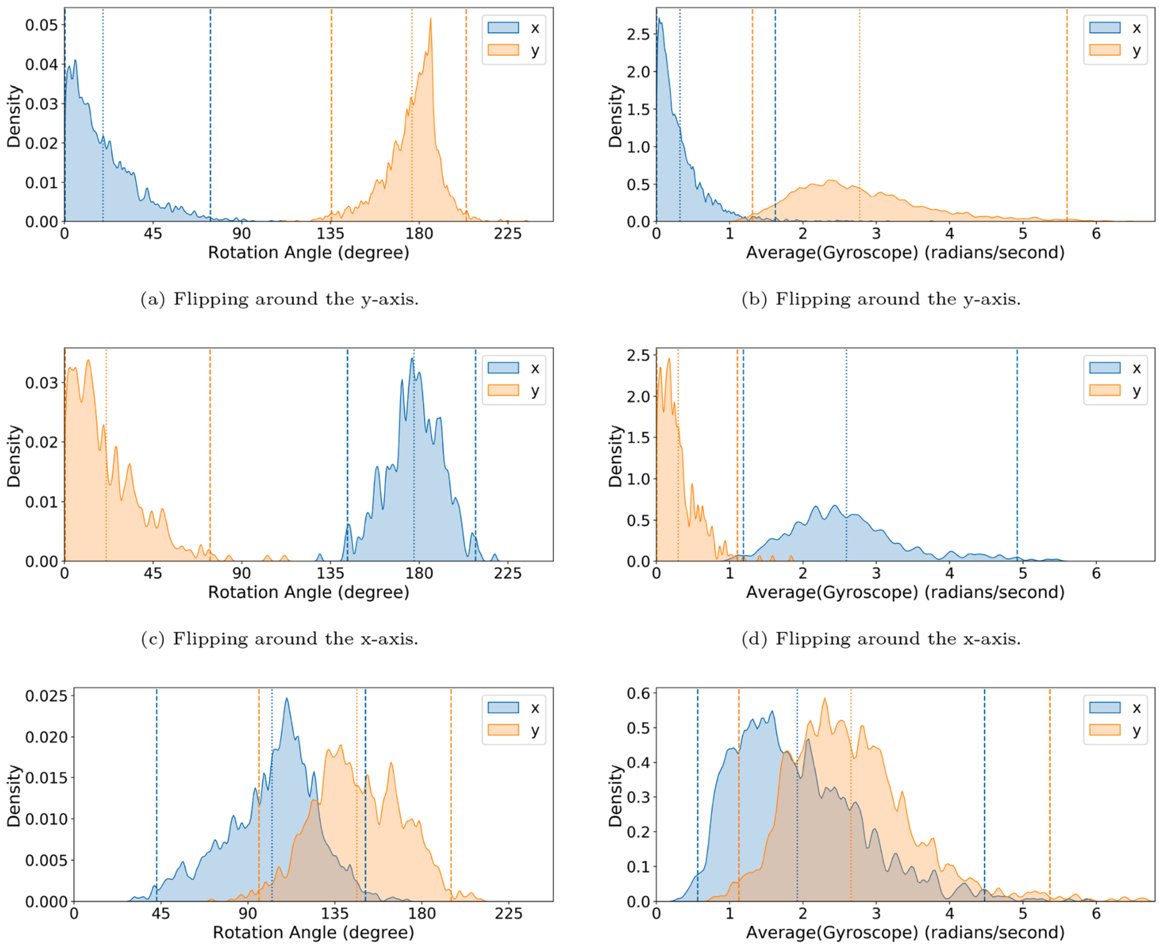
Fig. 12. The distribution of flipping gestures' duration. The vertical dashed lines indicate the 1tst and 99th percentile of the duration of that type and the vertical dotted lines indicate the averages.

flipping gestures into three types: (1) gestures that are mainly flipping around the *y*-axis, (2) gestures that are mainly flipping around the *x*-axis, and (3) gestures flipping around both the *x*-axis and *y*-axis simultaneously, namely all of the "PL" and "LP" gestures which switch between portrait and landscape orientations. Fig. 13 shows the distributions of the three flipping gesture types. From it we can see the obvious difference between the three gesture types, which will be used for flip detection.

5.4. Flip detection

Based on the findings, we designed a straightforward data-driven flip detection algorithm as follows. The algorithm buffers the last 2500 ms of motion sensor data, which can cover over the 99th percentile of the flipping duration according to Fig. 12. It then detects flipping based on the buffered data at 100 Hz. If the current gyroscope readings do not exceed a pre-defined small threshold on both the *x*-axis and y-axis, the algorithm simply determines that the phone is not flipping for energy saving. The threshold is set to 1.0 radians/s at default, which is less than the 1tst percentile of the flipping gyroscope data according to Fig. 13. If passing the threshold, the algorithm will further try to find if there is a time interval ending with the current time that meet the following conditions: the average gyroscope readings of *x*-axis and *y*-axis are between the 1tst and 99th percentile of the data of any one of the three flipping gesture types (Fig. 13), as well as the accumulated rotation angles. If such interval(s) exists, it will report all of the possible time intervals along with the confidences based on statistical density (the more it appears in the statistics, the higher the confidence is). If the confidence exceeds a threshold, then a flip is detected and reported. The minimum time interval length is set to 400 ms, i.e., each flipping will last for at least 400 ms (below the 1tst percentile of the data in Fig. 12).

The above algorithm briefly describes how to detect a complete flipping. However, according to a pilot study, if the phone switches the display only until it detects a complete flipping, almost all users perceived the delay and felt unsatisfied with the sudden brightening of the screen. Therefore, we improved the algorithm to support the detection of partial flipping. The main idea is the same as mentioned above; the only difference is the statistical data used, i.e., data from parts



(a) Flipping around the y-axis. (b) Flipping around the y-axis.

(c) Flipping around the x-axis. (d) Flipping around the x-axis.

(e) “PL” and “LP” gestures. (f) “PL” and “LP” gestures.

Fig. 13. The distributions of gyroscope data for three flipping gesture types: (1) gestures that are mainly flipping around the *y*-axis, (2) gestures that are mainly flipping around the *x*-axis, and (3) “PL” and “LP” gestures, i.e., switching between portrait orientation and landscape orientation. The vertical dashed lines indicate the 1tst and 99th percentile of the data, and the vertical dotted lines indicate the averages.

of the flip (duration, gyroscope, etc.) were used instead of the entire flipping. In addition, since users reported dissatisfaction with the display switching delay and sudden brightening, the more sensitive and aggressive the algorithm detects the flipping, the better. However, increasing the algorithm’s sensitivity would lead to more false positive errors, which also negatively affecting user experience. Thus, we introduced a trick to prevent it: temporarily turn on both screens during flipping. We suspected that users could only see one screen in most cases, even more so when flipping. Therefore, it would not hurt user experience, and users would not perceive the “delay of display switching” since the rear screen was already on as soon as the user saw it. In our implementation, both screens will be on when flipping reaches a certain angle (more aggressive compared to flip detection), and the screens reject all touch inputs during flipping to prevent unintended touch.

To test whether users like this design and get suitable parameters for flip detection, we designed flip detection sensitivity as a 7-level user-configurable setting (Table 3). The higher the sensitivity, the more aggressive the detection algorithm. For example, the phone will switch displays when it detects a partial flipping and the rotation angle reaches a total of 115 degrees around the *x*-axis and *y*-axis at sensitivity 1, 90 degrees instead at sensitivity 4 or higher. The phone will turn on both screens temporarily when the flipping reaches a specific angle at sensitivity 5 or higher. Since the phone’s front screen does not always face the user directly before flipping, we also tested when to turn on both screens by adjusting the response angle in sensitivities 5–7. The testing

Table 3
Parameters of the flip detection algorithm.

Sensitivity	Switch displays	Turn on both screens
1	115°	never
2	105°	never
3	95°	never
4	90°	never
5	90°	75°
6	90°	65°
7	90°	55°

will be conducted in the next user study (Section 6). Fig. 14 shows the user interface of sensitivity adjusting (Quinn et al., 2019), in which users can experience flipping with the newly set sensitivity at will. The default sensitivity is 4. The parameters in Table 3 are not exposed to users.

Besides the gyroscope, we also utilized linear acceleration sensors to reject some false positive errors, i.e., the motion that is not a phone flipping but looks similar to it from gyroscope readings. For example, users sit by the bed and then lie down while holding the phone aloft. Another example is that users turn backward while holding the phone in front of their chests. In these situations, the screen facing the user has not changed, but gyroscopes will sense rotations from one or more axes. In other words, the relative position and orientation relationship of the phone and the user’s face have not changed, but the phone moves significantly in the world coordinate system (compared to flipping in

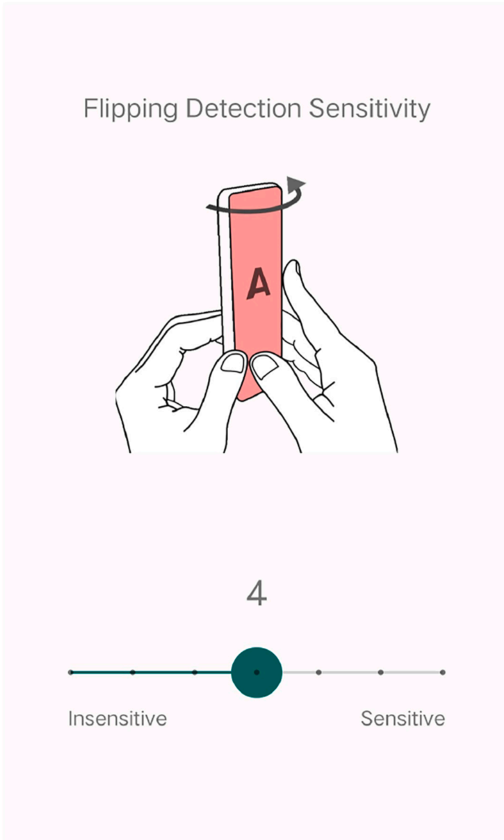


Fig. 14. User interface for adjusting detection sensitivity.

hand). Using linear acceleration sensors to detect phone movement can simply handle it.

6. User evaluation of flip detection and flipping applications

Based on the collected data and findings in Section 5, we designed and implemented the flip detection algorithm and deployed the prototype on the dual-display phone. In this section, we conducted a user study to test the detection algorithm and get users' subjective evaluation of different flipping applications, different display switching methods, and different manual intervention methods when the display does not switch.

6.1. Study design and procedure

We first gave the participants a brief introduction to the dual-display phone and the idea of flipping. They then got familiar with the phone. The experimental setting was similar to the previous studies (Fig. 8).

There are three parts to this study. Participants will experience various applications or methods in each part and rate them on different criteria.

The first part is experiencing and evaluating six application prototypes based on flipping. The six applications are: 1) switching between watching the video and chatting, 2) switching between regular shot and selfie, 3) rejecting a phone call, 4) delaying an alarm, 5) viewing the back of clothes, and 6) memorizing words. Participants were instructed to use every application ten times, which also meant flipping the phone ten times. Take rejecting the call as an example: a virtual phone call would be made to the phone; turning it over would reject it. After a random 2–4 s, another phone call would come. After experiencing, participants rated the application through a questionnaire which contains the following four descriptions on a 7-point Likert scale (1-strongly disagree, 7-strongly agree): 1) "I can successfully flip the phone", 2) "The phone can successfully detect flips", 3) "I like this application", and 4) "I prefer the flipping approach on the dual-display phone to the current non-flipping implementation on regular single-screen phones". Also, we evaluated the performance of the flip detection algorithm in this part.

In addition to the flipping applications, we also introduced the flip detection sensitivity to participants. They could adjust and experience different sensitivities (Fig. 14) before and during the first part of the study. In the end, we collected the sensitivity that participants changed to and their subjective feelings through the questionnaire. Note that the specific parameters in Table 3 were not exposed to participants.

The second part of this study compares flipping to other existing manual display switching methods on the dual-display phone. They are: 1) pressing the power button on the right, 2) pressing power buttons on both sides together, 3) sliding horizontally with three fingers, and 4) tapping the "Display on the other screen" UI button in the slide-up control center. There are power buttons on both sides of the phone in symmetrical positions, so pressing them together is feasible and can be used for display switching. Similarly, there will always be a power button on the right, no matter which screen faces users. Pressing it will turn on the front screen (and turn off the rear screen), which can be used to switch displays after flipping the phone. We instructed participants to flip the phone and use each display switching method ten times. They then rated these four manual methods along with automatic display switching when flipping on a 7-point Likert scale (the higher, the better): **Simplicity, Efficiency**, (the possibility of) **Trigger by mistake**, and **Preference**.

The third and final part of the study evaluates five manual intervention methods when the algorithm fails to detect the flipping and the display does not switch. In such cases, the front screen after flipping is off, and users need to switch displays manually. The five intervention methods are: 1) single-finger double tap, 2) single-finger circle draw, 3) pressing the power button on the right, 4) pressing power buttons on both sides together, and 5) multi-fingers double tap. Note that methods 1, 2, and 5 require the touchscreen to be able to always read touch input (or turn on for a while without displaying anything if the gyroscope readings pass a threshold). The flip detection algorithm was turned off, and participants would flip the phone and then manually intervene as if the phone failed to detect the flipping. We asked participants to use each intervention method ten times in both one-handed and two-handed use, respectively. Method 5, multi-fingers double tap, is for two-handed use only. Participants then rated the intervention methods with the same criteria in the second part: **Simplicity, Efficiency, Trigger by mistake**, and **Preference**.

In this study, we did not give participants any instructions or guidance about flipping gestures. They could use any flipping gestures they like to turn over the phone freely. In each part of the study, the order of applications/methods was counterbalanced across participants, and participants could modify previously given ratings after experiencing and comparing different applications/methods. We also encouraged participants to make think-aloud comments when rating.

6.2. Participants

We recruited 12 participants (P1–P12, 5 female, 7 male) from the campus, aged from 19 to 27 years (M = 22.4, SD = 2.6). None had participated in the previous data-collection study. All participants were right-handed. Their hand lengths ranged from 161 cm to 204 cm, which comprised samples from the 2nd to 85th percentile of the data reported in an anthropometric survey (Gordon et al., 2014). They have been using smartphones for 4.5 to 11.0 years (M = 8.67, SD = 1.88). The display of their daily-used smartphones ranged from 5.65-inch to 6.89-inch (M = 6.12, SD = 0.37). Among them, 10 participants were new to dual-display phones. The study lasted for about 40 min. Participants were compensated 60 CNY (about 9 USD) for their time.

6.3. User evaluation of flipping applications

Fig. 15 shows the subjective ratings of flipping applications. Based on the high rating of "I can successfully flip the phone" ($M = 6.78$, $SD = 0.11$), users have no difficulty using all of the applications by flipping. Moreover, users generally liked all of the offered flipping applications ($M = 5.57$, $SD = 0.41$). Among the six applications, users liked 2) switching between selfies and regular shots ($M = 6.08$, $SD = 0.76$) and 4) delaying an alarm the most ($M = 5.92$, $SD = 1.61$), followed by 1) switching between watching the video and chatting ($M = 5.58$, $SD = 1.66$) and 6) memorizing words ($M = 5.67$, $SD = 1.11$). Users consistently preferred 4) delaying an alarm by flipping to non-flipping implementation on regular phones ($M = 6.17$, $SD = 1.34$). As P3 said, *"It is fantastic to be capable of delay the annoying alarm in an eyes-free manner."* In contrast, users did not prefer to reject a phone call ($M = 4.08$, $SD = 1.26$) or view the back of clothes ($M = 4.50$, $SD = 1.76$) by flipping: *"In most cases, I will check who is calling and decide whether to answer it. Hence, flipping the phone in an eyes-free manner to reject the call is not that attractive to me."* (P12) *"It is too cumbersome to turn over the phone to view photos from different angles comparing to sliding."* (P11) However, the cost of flipping may turn into a benefit sometimes. Memorizing words is a typical example ($M = 5.25$, $SD = 1.23$): *"I like this design (flip to read the meaning/translation of the word) because it makes me carefully recall the meaning of the word before flipping. Only when I can not remember the word at all, I will flip the phone to check it."* (P6)

6.4. Performance of flip detection

Every participant flipped the phone $6 \times 10 = 60$ times in the first part of the study. The average accuracy of the flip detection algorithm was 97.78% ($SD = 2.08\%$), with 0–4 undetected flips for each participant. Participants also gave a high score on "The phone can successfully detect flips" (Fig. 15, $M = 6.61$, $SD = 0.22$).

As for flip detection sensitivity, four, six, and two participants chose sensitivity 4, 5, and 6, respectively. It indicates that users want the rear screen to be turned on before seeing it during flipping. The "turn on both screens" mechanism helped achieve this and was accepted by most participants. In addition, participants rated an average of 6.50 (1 = strongly disagree, 7 = strongly agree) on "I'm satisfied with the speed of display switching with the chosen sensitivity." Besides, though some participants reported that they did not feel the difference between adjacent sensitivities (P4, P12), all of them could perceive the difference between sensitivities 1 and 7.

6.5. User evaluation of display switching methods

Fig. 16 shows the subjective ratings of display switching methods. As expected, simply flipping the phone to switch displays received the highest scores on both **Simplicity** ($M = 6.58$, $SD = 0.64$) and **Efficiency** ($M = 6.50$, $SD = 0.87$) since existing manual display switching methods all require more steps on top of flipping. In particular, tapping the UI button in the control center requires three steps: first sliding up the control center, then tapping the "Display on the other screen" button, and finally flipping the phone. Hence, it received the lowest **Efficiency** score ($M = 3.50$, $SD = 1.32$) but the highest **Trigger by mistake** score ($M = 5.92$, $SD = 1.11$; the higher the score is, the better). As for sliding horizontally with three fingers, P1 and P4 complained that *"it can only be performed by two hands, which may cause inconvenience in some cases."*

Users' subjective evaluations of **Efficiency** were consistent with the actual time costs of display switching methods, as also shown in Fig. 16. Simply flipping the phone to switch displays spent significantly less time than manual methods. As a result, users consistently preferred it ($M = 6.08$, $SD = 0.76$) as they were satisfied with the accuracy and latency of the flip detection algorithm.

6.6. User evaluation of manual intervention methods

Fig. 17 shows the subjective ratings of manual intervention methods. Single-finger double tap was the favorite choice for both one-handed and two-handed use, which received the highest **Simplicity, Efficiency**, and **Preference** ratings. In contrast, pressing power buttons together was the least favorite one. P6 explained the reason: *"When holding the phone in both hands, it is uncomfortable to press the power buttons simultaneously no matter using the thumbs or index fingers of both hands. When holding it in one hand, pressing buttons with the thumb and index finger does feel good, just like gripping. However, to do so needs to change the grip posture after flipping the phone, which is inconvenient."* On the other hand, P4 shared concern for pressing the button on the right: *"it is not intuitive to tell which power button is 'on the right' when using the phone in landscape orientation, especially after flipping it."* Moreover, *"To avoid accidentally triggering, I need to be careful not to press the power button during flipping"* (P5, P8). Therefore, users generally liked double tap and draw gestures on the front screen, especially for two-handed use in which users can hold the phone in one hand and do touch input with the other hand.

7. Discussion, limitation, and future work

We now discuss some topics related to phone flipping but not covered in our studies. Meanwhile, we summarize some limitations of this paper, which we also see as opportunities for future work.

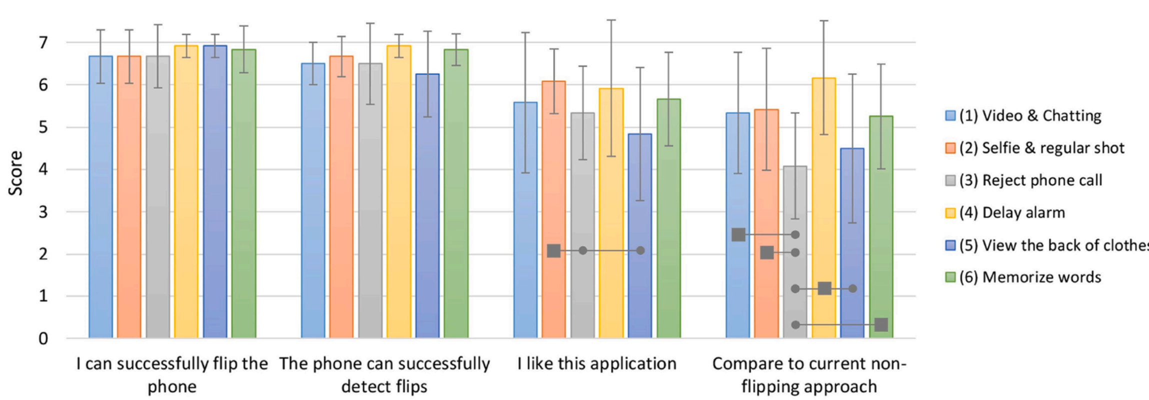


Fig. 15. Subjective ratings of flipping applications collected on a 7-point Likert scale (1 = strongly disagree, 7 = strongly agree). The horizontal lines represent results of Wilcoxon signed-rank tests: for a given line, the rating of the square application is significantly higher than that of the dot one ($p < 0.05$). Error bars indicate one standard deviation.

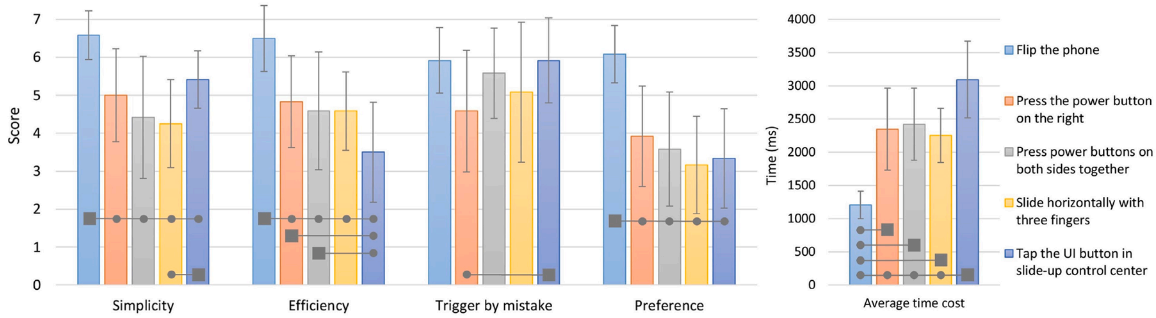


Fig. 16. Subjective ratings of display switching methods collected on a 7-point Likert scale (the higher, the better) and average time costs. The horizontal lines represent Wilcoxon signed-rank tests / RM-ANOVA results: for a given line, the rating / time cost of the square method is significantly higher than that of the dot one ($p < 0.05$). Error bars indicate one standard deviation.

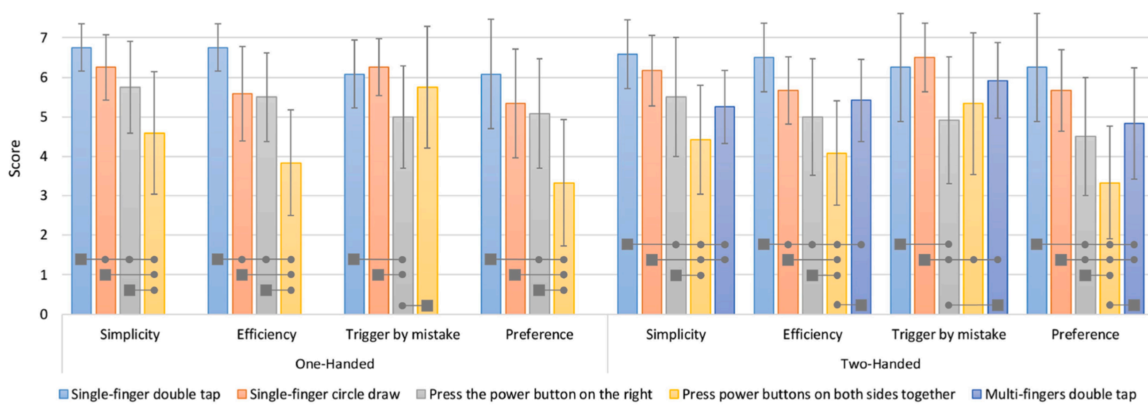


Fig. 17. Subjective ratings of manual intervention methods collected on a 7-point Likert scale (the higher the better). The horizontal lines represent results of Wilcoxon signed-rank tests: for a given line, the rating of the square method is significantly higher than that of the dot one ($p < 0.05$). Error bars indicate one standard deviation.

7.1. Turning on which screen when waking on the phone

We have discussed how to detect flipping to support display switching when flipping. However, in addition to display switching, there still remains a problem of which screen should be turned on when waking on the dual-display phone. In other words, which screen is facing the user. Possible solutions and clues may include:

- Touch input. Users can explicitly tap the desired screen to wake it, similar to the existing Tap to Wake feature (Apple, 2021). Similarly, users can press the fingerprint identity sensor to unlock the desired screen (if both sides of the phone support fingerprint recognition).
- Lifting direction (linear acceleration of z-axis). It is especially useful when users wake on the phone with the Raise/Lift to Wake feature (Apple, 2021).
- Power button on the right. As mentioned in Section 6, pressing the power button on the right will wake the front screen. It requires the phone to have two symmetric buttons on both the left and right sides.

7.2. Unintended touch

Some participants in our user studies were concerned about the unintended touch problem during flipping. In our prototype implementation, we simply reject all touch inputs during flipping to prevent unintended touch. The condition is whether the gyroscope readings in the x-axis and y-axis (Fig. 3) pass a threshold, which is straightforward and preliminary. The patterns of unintended touch during flipping and the corresponding error recovery method are both worth researching and require thorough investigations, which can refer to previous works on unintended touch for direct pen interaction (Annett et al., 2014), back-of-device touch panels (Le et al., 2019), and interactive tabletops (Xu et al., 2020b).

7.3. Impact of hand size and mobile form factors

In this research, we used a specific commercial off-the-shelf phone as the only experimental device (described in Section 4.1.1) in user studies. However, mobile form factors, especially the phone size, may impact hand movements of flipping (Eardley et al., 2017). For example, if the phone size gets bigger, some users may not be able to perform some unimanual flipping gestures or spend more time and effort flipping. Correspondingly, users' hand sizes may also influence their flipping behaviors. In fact, we already found a weak negative correlation between hand length and average flipping time duration in our study ($R^2 = 0.10$).

7.4. Limitations of collected flipping data

On one hand, we collected users' flipping data in a lab setting, in which users were relatively concentrated. The patterns of flipping behavior may be different if the user is distracted. For example, the user flips the phone while conversing with others, i.e., with the dual-task interference (Pashler, 1994). Also, we only collected flipping data while the user is sitting. It is worth exploring whether there is a difference in how users flip the phone under different conditions, such as walking, running, and lying down. Moreover, the sensor reading would be much noisier when the user is moving or in a vibration environment (e.g., on a bus). Although they mainly affect accelerometers and flip detection mostly depends on the gyroscope, it may still be challenging to accurately and efficiently detect users' flipping behaviors in these circumstances.

On the other hand, since dual-display smartphones are a relatively new product type and not very popular yet, most participants of the user studies did not use a dual-display phone in their daily lives. Therefore, data collected in this research mainly reflected the flipping behaviors of users who are new to dual-display phones. Possible learning effects and their implications have not been adequately studied. Finally, as a preliminary study of a new interaction method, the number and diversity of the participants in our user studies was limited. Hence, some results in this research may not be directly generalized to different cultural contexts (e.g., the social acceptance subjective rating of gestures).

8. Conclusion

In this research, we proposed flipping as a new input modality for dual-display phones. We first discussed potential application scenarios for flipping the dual-display phone through brainstorming, with the following three main usages: multitasking, using flipping as a trigger, and viewing the other side of objects. We then elicited 36 user-defined flipping gestures from 22 participants (Fig. 6) and showed the agreement rates of the flipping gestures in different application scenarios (Fig. 7).

Next, we conducted a user experience study with 19 participants to collect sensor data for all flipping gestures and get users' subjective evaluations regarding easiness, social acceptability, and fatigue (Fig. 9). Results showed that the average time duration of flipping is 1144.96 ms. The slowest gesture ($M = 1593.44$ ms) takes almost twice as long as the fastest one ($M = 815.18$ ms) (Fig. 11). "Wrist" gestures (flipping gestures that are mainly achieved by rotating the wrist) are the fastest ($M = 866.96$ ms) among all gestures, followed by two-handed gestures ($M = 1127.98$ ms). Meanwhile, users also generally preferred "Wrist" gestures and two-handed gestures (Fig. 9). Besides, users perceived more differences from different gesture types than which hand is used to flip (Table 2). In fact, there is no significant difference between flipping with the dominant and non-dominant hand on time costs (Fig. 12).

Based on the findings and data collected, we designed the flip detection algorithm to automatically switch displays when flipping and then implemented it on a commercial off-the-shelf dual-display phone. We introduced the "turn on both screens temporarily during flipping" mechanism to reduce the perceived latency of display switching without modifying the detection sensitivity of the algorithm. A user evaluation study showed that the algorithm achieved 97.78% detection accuracy, and users were satisfied with display switching speed. Compared to existing manual display switching methods, users consistently preferred automatic display switching when flipping since it is much more efficient (Fig. 16). We also collected users' subjective evaluations of flip-based applications (Fig. 15) and manual intervention methods (when the display does not switch) (Fig. 17).

Overall, our work demonstrates that flipping is a practical and promising interaction modality for dual-display phones and provides design implications for it.

CRediT authorship contribution statement

Zhican Yang: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing. **Chun Yu:** Conceptualization, Supervision, Project administration. **Xin Chen:** Formal analysis, Investigation, Data curation. **Jingjia Luo:** Formal analysis, Investigation, Data curation. **Yuanchun Shi:** Project administration, Funding acquisition.

Declaration of Competing Interest

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

Acknowledgments

This work is supported by the Natural Science Foundation of China under Grant no. 62132010, and National Key R&D Program of China no. 2020AAA0105200, and also by Beijing Key Lab of Networked Multimedia, the Institute for Guo Qiang, Tsinghua University, Institute for Artificial Intelligence, Tsinghua University (THUAI), and Beijing Academy of Artificial Intelligence (BAAI).

Appendix A. Effect size

In this research, we reported effect sizes along with p values to illustrate both the substantive significance (effect size) and statistical significance (p value) in quantitative studies (Sullivan and Feinn, 2012). Specifically, the partial eta squared (η_p^2) and the correlation coefficient (r) (Rosenthal et al., 1994) were reported for RM-ANOVA and the Wilcoxon signed-rank test respectively (Tomczak and Tomczak, 2014). The formulas are presented below:

$$\eta_p^2 = \frac{SS_{\text{effect}}}{SS_{\text{effect}} + SS_{\text{error}}} \tag{A.1}$$

where SS_{effect} is the sum of squares for the effect and SS_{error} is the sum squared errors.

$$r = \frac{Z}{\sqrt{n}} \tag{A.2}$$

where n is the total number of observations on which Z is based.

The thresholds for interpreting effect size (Cohen, 2013) are summarized as follows (Table A.1):

Table A.1
Thresholds for interpreting effect size.

Test	Relevant effect size	Effect size threshold			
		Small	Medium	Large	Very large
Repeat measures ANOVA	η_p^2	0.0099	0.0588	0.1379	/
Wilcoxon signed-rank test	r	0.10	0.30	0.50	0.70 (Rosenthal, 1996)

References

Annett, M., Gupta, A., Bischof, W.F., 2014. Exploring and understanding unintended touch during direct pen interaction. ACM Trans. Comput.-Hum. Interact. 21 (5) https://doi.org/10.1145/2674915.

Apple, 2021. Use Raise to Wake on your iPhone - Apple Support. https://support.apple.com/en-gb/HT208081.

Chang, Y., L'Yi, S., Koh, K., Seo, J., 2015. Understanding users' touch behavior on large mobile touch-screens and assisted targeting by tilting gesture. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, pp. 1499–1508. https://doi.org/10.1145/2702123.2702425.

Cohen, J., 2013. Statistical Power Analysis for the Behavioral Sciences. Academic Press.

Cui, W., Zhu, S., Li, Z., Xu, Z., Yang, X.-D., Ramakrishnan, I.V., Bi, X., 2021. BackSwipe: Back-of-Device Word-Gesture Interaction on Smartphones. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3411764.3445081.

Eardley, R., Roudaut, A., Gill, S., Thompson, S.J., 2017. Understanding Grip Shifts: How Form Factors Impact Hand Movements on Mobile Phones. Association for Computing Machinery, New York, NY, USA, pp. 4680–4691. https://doi.org/10.1145/3025453.3025835.

Findlater, L., Lee, B., Wobbrock, J., 2012. Beyond QWERTY: augmenting touch screen keyboards with multi-touch gestures for non-alphanumeric input. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, pp. 2679–2682. https://doi.org/10.1145/2207676.2208660.

Goel, M., Wobbrock, J., Patel, S., 2012. GripSense: using built-in sensors to detect hand posture and pressure on commodity mobile phones. Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology. Association for Computing Machinery, New York, NY, USA, pp. 545–554. https://doi.org/10.1145/2380116.2380184.

Google, 2021a. Motion sensors | Android Developers. https://developer.android.com/guide/topics/sensors/sensors_motion.

Google, 2021b. Sensor Event | Android Developers. https://developer.android.com/reference/android/hardware/SensorEvent.

Gordon, C.C., Blackwell, C.L., Bradtmiller, B., Parham, J.L., Barrientos, P., Paquette, S.P., Corner, B.D., Carson, J.M., Venezia, J.C., Rockwell, B.M., et al., 2014. 2012 Anthropometric Survey of US Army Personnel: Methods and Summary Statistics. Technical Report. Army Natick Soldier Research Development and Engineering Center MA.

Hsieh, Y.-T., Jylhä, A., Orso, V., Gamberini, L., Jacucci, G., 2016. Designing a willing-to-use-in-public hand gestural interaction technique for smart glasses. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, pp. 4203–4215. https://doi.org/10.1145/2858036.2858436.

Kay, M., Wobbrock, J. O., 2021. ARTool: Aligned Rank Transform for Nonparametric Factorial ANOVAs. R package version 0.11.0. https://github.com/mjskay/ARTool.

Kim, W., Xiong, S., 2021. User-defined walking-in-place gestures for VR locomotion. Int. J. Hum.-Comput. Stud. 152, 102648. https://doi.org/10.1016/j.ijhcs.2021.102648. https://www.sciencedirect.com/science/article/pii/S1071581921000665.

Le, H.V., Kosch, T., Bader, P., Mayer, S., Henze, N., 2018. PalmTouch: Using the Palm as an Additional Input Modality on Commodity Smartphones. Association for Computing Machinery, New York, NY, USA, pp. 1–13. https://doi.org/10.1145/3173574.3173934.

Le, H.V., Kosch, T., Mayer, S., Henze, N., 2018. Demonstrating palm touch: the palm as an additional input modality on commodity smartphones. Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct. Association for Computing Machinery, New York, NY, USA, pp. 353–358. https://doi.org/10.1145/3236112.3236163.

Le, H.V., Mayer, S., Henze, N., 2018. InfiniTouch: finger-aware interaction on fully touch sensitive smartphones. Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology. Association for Computing Machinery, New York, NY, USA, pp. 779–792. https://doi.org/10.1145/3242587.3242605.

Le, H.V., Mayer, S., Steuerlein, B., Henze, N., 2019. Investigating unintended inputs for one-handed touch interaction beyond the touchscreen. Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3338286.3340145.

Le, H.V., Mayer, S., Wolf, K., Henze, N., 2016. Finger placement and hand grasp during smartphone interaction. Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, pp. 2576–2584. https://doi.org/10.1145/2851581.2892462.

Liang, H.-N., Williams, C., Semegen, M., Stuerzlinger, W., Irani, P., 2012. User-defined surface + motion gestures for 3Dmanipulation of objects at a distance through a mobile device. Proceedings of the 10th Asia Pacific Conference on Computer Human Interaction. Association for Computing Machinery, New York, NY, USA, pp. 299–308. https://doi.org/10.1145/2350046.2350098.

Likert, R., 1932. A technique for the measurement of attitudes. Arch. Psychol. 140, 5–55.

Lu, Y., Huang, B., Yu, C., Liu, G., Shi, Y., 2020. Designing and evaluating hand-to-hand gestures with dual commodity wrist-worn devices. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 4 (1) https://doi.org/10.1145/3380984.

Mayer, S., Le, H.V., Henze, N., 2017. Estimating the finger orientation on capacitive touchscreens using convolutional neural networks. Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces. Association for Computing Machinery, New York, NY, USA, pp. 220–229. https://doi.org/10.1145/3132272.3134130.

Pashler, H., 1994. Dual-task interference in simple tasks: data and theory. Psychol. Bull. 116 (2), 220.

Quinn, P., Lee, S.C., Barnhart, M., Zhai, S., 2019. Active Edge: Designing Squeeze Gestures for the Google Pixel 2. Association for Computing Machinery, New York, NY, USA, pp. 1–13. https://doi.org/10.1145/3290605.3300504.

Rahman, M., Gustafson, S., Irani, P., Subramanian, S., 2009. Tilt techniques: investigating the dexterity of wrist-based input. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, pp. 1943–1952. https://doi.org/10.1145/1518701.1518997.

Rosenthal, J.A., 1996. Qualitative descriptors of strength of association and effect size. J. Soc. Serv. Res. 21 (4), 37–59.

Rosenthal, R., Cooper, H., Hedges, L., 1994. Parametric measures of effect size. Handb. Res. Synth. 621, 231–244.

Ruiz, J., Li, Y., 2011. DoubleFlip: A Motion Gesture Delimiter for Mobile Interaction. Association for Computing Machinery, New York, NY, USA, pp. 2717–2720. https://doi.org/10.1145/1978942.1979341.

Ruiz, J., Li, Y., Lank, E., 2011. User-Defined Motion Gestures for Mobile Interaction. Association for Computing Machinery, New York, NY, USA, pp. 197–206. https://doi.org/10.1145/1978942.1978971.

Serrano, M., Ens, B.M., Irani, P.P., 2014. Exploring the use of hand-to-face input for interacting with head-worn displays. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, pp. 3181–3190. https://doi.org/10.1145/2556288.2556984.

Shimon, S.S.A., Morrison-Smith, S., John, N., Fahimi, G., Ruiz, J., 2015. Exploring user-defined back-of-device gestures for mobile devices. Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services. Association for Computing Machinery, New York, NY, USA, pp. 227–232. https://doi.org/10.1145/2785830.2785890.

Sullivan, G.M., Feinn, R., 2012. Using effect size–or why the *P* value is not enough. J. Grad. Med. Educ. 4 (3), 279–282.

Tomczak, M., Tomczak, E., 2014. The need to report effect size estimates revisited. An overview of some recommended measures of effect size. Trends Sport Sci. 21 (1), 19–25.

Tu, H., Huang, Q., Zhao, Y., Gao, B., 2020. Effects of holding postures on user-defined touch gestures for tablet interaction. Int. J. Hum.-Comput. Stud. 141, 102451. https://doi.org/10.1016/j.ijhcs.2020.102451. https://www.sciencedirect.com/science/article/pii/S1071581920300537.

Vatavu, R.-D., Wobbrock, J.O., et al., 2015. Formalizing agreement analysis for elicitation studies: new measures, significance test, and toolkit. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, pp. 1325–1334. https://doi.org/10.1145/2702123.2702223.

Verma, J.P., 2015. Repeated Measures Design for Empirical Researchers. John Wiley & Sons.

vivo, 2021. vivo NEX Dual Display|vivo Global. https://www.vivo.com/en/products/nexdualdisplay.

Vuletic, T., Duffy, A., McTeague, C., Hay, L., Brisco, R., Campbell, G., Grealy, M., 2021. A novel user-based gesture vocabulary for conceptual design. Int. J. Hum.-Comput. Stud. 150, 102609. https://doi.org/10.1016/j.ijhcs.2021.102609. https://www.sciencedirect.com/science/article/pii/S1071581921000276.

Wikipedia, 2021. Fez (video game) - Wikipedia. https://en.wikipedia.org/wiki/Fez_(video_game).

Wimmer, R., Boring, S., 2009. HandSense: discriminating different ways of grasping and holding a tangible user interface. Proceedings of the 3rd International Conference on Tangible and Embedded Interaction. Association for Computing Machinery, New York, NY, USA, pp. 359–362. https://doi.org/10.1145/1517664.1517736.

Wobbrock, J.O., Aung, H.H., Rothrock, B., Myers, B.A., 2005. Maximizing the Guessability of Symbolic Input. Association for Computing Machinery, New York, NY, USA, pp. 1869–1872. https://doi.org/10.1145/1056808.1057043.

Wobbrock, J.O., Findlater, L., Gergle, D., Higgins, J.J., 2011. The aligned rank transform for nonparametric factorial analyses using only ANOVA procedures. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, pp. 143–146. https://doi.org/10.1145/1978942.1978963.

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. Int. J. Hum.-Comput. Stud. 66 (12), 857–875. https://doi.org/10.1016/j.ijhcs.2008.03.004. https://www.sciencedirect.com/science/article/pii/S107158190800027X.

Wu, H., Yang, L., 2020. User-defined gestures for dual-screen mobile interaction. Int. J. Hum.-Comput.Interact. 36 (10), 978–992. https://doi.org/10.1080/10447318.2019.1706331.

Xiao, R., Schwarz, J., Harrison, C., 2015. Estimating 3D finger angle on commodity touchscreens. Proceedings of the 2015 International Conference on Interactive Tabletops & Surfaces. Association for Computing Machinery, New York, NY, USA, pp. 47–50. https://doi.org/10.1145/2817721.2817737.

Xu, X., Shi, H., Yi, X., Liu, W., Yan, Y., Shi, Y., Mariakakis, A., Mankoff, J., Dey, A.K., 2020. EarBuddy: enabling on-face interaction via wireless earbuds. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, pp. 1–14. https://doi.org/10.1145/3313831.3376836.

Xu, X., Yu, C., Wang, Y., Shi, Y., 2020. Recognizing unintentional touch on interactive tabletop. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 4 (1) https://doi.org/10.1145/3381011.
Yang, Z., Yu, C., Yi, X., Shi, Y., 2019. Investigating gesture typing for indirect touch. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 3 (3) https://doi.org/10.1145/3351275.
Yang, Z., Yu, C., Zheng, F., Shi, Y., 2019. ProxiTalk: activate speech input by bringing smartphone to the mouth. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 3 (3), 1–25. https://doi.org/10.1145/3351276.