ORIGINAL RESEARCH ARTICLE

Perceived User Reachability in Mobile UIs Using Data Analytics and Machine Learning

Lik-Hang LEE[a], Yui-Pan YAU[b] and Pan HUI[c]

[a] The Hong Kong Polytechnic University, Hong Kong SAR; [b]The Hong Kong University of Science and Technology, Hong Kong SAR; [c]The Hong Kong University of Science and Technology (Guangzhou), China

ARTICLE HISTORY
Compiled January 15, 2026

ABSTRACT
One-handed interactions on smartphone interfaces offer a prominent feature of highly mobile inputs. Thus, the design factor of user reachability is essential to realizing the incentive. However, the sole consideration of physical characteristics, such as hand size and thumb length, does not fully reflect the users' perceived choices of hand poses and the corresponding inertia. We first conducted a 6-week questionnaire-based study of UI rating tasks and collected 62,156 responses reflecting user preferences for 3,000 clustered UIs. Our analysis of the responses shows that user perceptions of smartphone UI components are divergent from their physical ability of thumb reaches; e.g., they can reach an icon with a thumb reach, but they prefer alternative hand poses. Accordingly, we propose a machine learning model, i.e., XG-Boost, to predict the user's choices of hand poses, with a reasonable prediction accuracy of 64% that can be regarded as a practical preliminary evaluation tool. With illustrative examples, our model can offer auxiliary information in the assessment of perceived user reachability with one-handed interaction on smartphone interfaces, which paves a path towards a computational understanding of UI designs, and such findings can be further extended to 2D UIs in 3D worlds.

KEYWORDS
Mobile UIs, one-handed interaction, machine learning, reachability, cognitive ergonomics.

1. Introduction

People with smartphones not only view the content on the touchscreen, but also provide inputs with such content that are regarded as clickable items (e.g., icons). One-handed interaction, ideally, brings convenience to people who can leverage their smartphones to connect to the Internet. They accomplish their jobs with one hand when standing, walking, or even riding in a taxi, leaving the other hand available for other activities. One-handed interaction with clickable items on smartphone UIs, according to the WIMP paradigm (i.e., Window, Menu, Icon, Pointer), allows users to preserve a free hand for other tasks in-the-wild (Lee et al., 2019). For instance, the free hand can interact with other people during social interaction (e.g., handshaking) or grasp objects (e.g., an umbrella on rainy days). Therefore, a well-known mobile UI designer,

Lik-Hang Lee is the corresponding author: lik-hang.lee@polyu.edu.hk

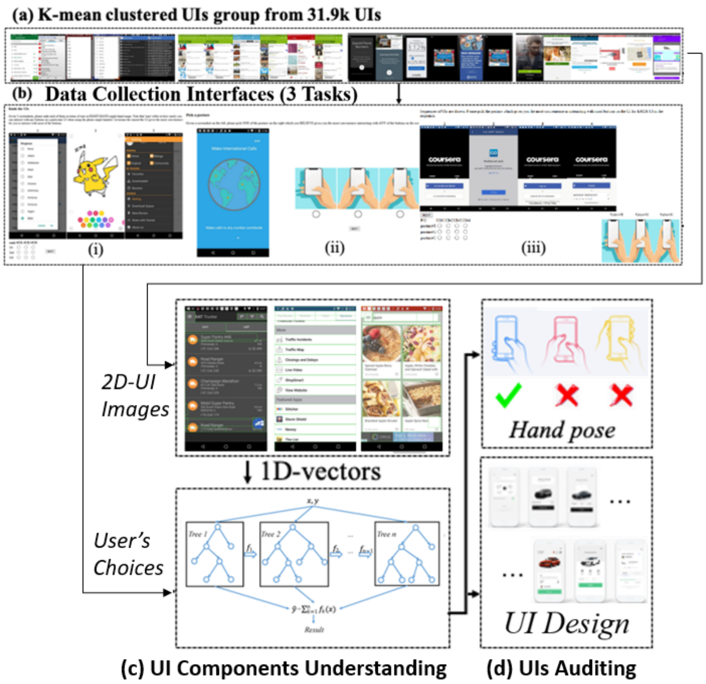


Figure 1. Similar UIs in the K-Means Clustered Groups (a) convert into vectors, by collecting user preferences on UIs via clustered interfaces (b). Under the premise of one-handed interaction, our model (c) can evaluate the likelihood of highly mobile interaction with the target UIs, according to the predicted hand poses (d).

Steven Hoober, has coined the term "*The Thumb Zone*", which describes the most accessible areas (natural, stretch, and unreachable) for one-handed thumb operations on smartphone touchscreens (Hoober & Berkman, 2011). The difficulties of accessing certain icons, menus, and windows, lead to different hand poses, e.g., a stretch hand pose to reach an item on the top edge of a smartphone. Accordingly, UI designers follow the principle and keep important clickable items within the thumb zone (i.e., natural and stretch hand poses). Generally, the principle is to keep frequently used items in easy-to-reach areas, while the infrequently used items are located outside the thumb zone (i.e., unreachable). However, allocating all clickable items within the thumb reach is challenging, under the premise that the users maintain a firm grip on their smartphones (Tsai, Huang, Hsieh, Huang, & Hung, 2016) (Kim, Yu, & Lee, 2012). To alleviate the issue, this paper aims to explore a computational approach encouraging one-handed interaction on smartphones by considering the user's preferred hand poses.

Accordingly, the research community employs linear models, e.g., Fitts's Law (Bi, Li, & Zhai, 2013a), to construct relationships between the user touch performance and the clickable items on UI layouts. Fitts's law is a predictive model of human motion that anticipates duration for swiftly relocating to a designated place. Fitt's Law also describes the touch accuracy of the user's thumb on a small-size touchscreen (Boring et al., 2012). Statistical models also serve similar purposes, in which users' touches on UI components on smartphone layouts are converted into 2D distributions of touch points and, hence, predictive touch areas of user inputs. Meanwhile, existing studies primarily attempt to convert the ergonomic means (e.g., thumb length and grip size) into the ease of touch inputs on smartphones of varying sizes (Larsen, Jacobsen, Boring, Bergström,

& Pohl, 2019; Prange, Buschek, & Alt, 2018). Similarly, a prior study converted the hand grips and finger lengths into a parabola curve that represents the aforementioned "*The Thumb Zone*" (Bergstrom-Lehtovirta & Oulasvirta, 2014).

The above approaches can robustly estimate based on smartphone layouts and users' ergonomic factors. However, they are insufficient to explain the user's mental models, which are complicated and fuzzy. The user perception of smartphone layouts, i.e., the user's mental model[1], is hardly modelled by those linear and statistical models as mentioned earlier (Li, Kumar, Lasecki, & Hilliges, 2020). In brief, the mental model refers to understanding someone's cognitive process that plans how something works in the real world. In the context of this paper, mental models refer to the user perception of smartphone layouts when the planning in one's mind goes before the touch actions. Meanwhile, the recent approaches leveraging data analytics and machine learning demonstrate reasonable predictive powers for users' mental models of smartphones layouts, such as user perceptions on animations, aesthetics, and discoverability of UI components (Deka, Huang, Franzen, Nichols, et al., 2017; Miniukovich & De Angeli, 2015; Swearngin & Li, 2019a; Wu, Jiang, Liu, & Ma, 2020). Nevertheless, the user perception of an array of clickable items in UIs (equivalently, UI component layouts) on whether a particular hand pose is suitable to the current UIs, i.e., cognitive means of UI reachability on smartphone layouts, requires additional explorations with alternative assessment approaches.

This paper begins with an exploratory study revealing the discrepancy between physical factors, by defining an ergonomic metric of Thumb Reachability Scores (TRS) for one-handed interaction, and cognitive means (e.g., user's perception of UI components in UI layouts). According to the 25,884 collected user choices on UIs (Section 4.2), no less than 50% of responses deviate from the ergonomic metric of TRS, indicating that the physical means of UI reachability, on the basis of TRS, are not the absolute answer. Additionally, a separate collection of 30,261 user preferences regarding hand positions reveals that people have a preference for two-handed engagement rather than one-handed interaction when it comes to high TRS user interfaces (Section 4.3). A follow-up study of pose preference across 6,011 sequential UIs (Section 4.4) shows the user's pose inertia (for example, using the same hand poses in different UIs) and their tendency to stick to their two favourite poses while interacting with a series of UIs. The above disagreements between ergonomic factors and cognitive means show the necessity of evaluating UI components in UI layouts (i.e., an array of clickable items in UIs) based on the users' perceived choices of hand poses. By leveraging the 62,156 user responses collected from UI rating tasks, we confirmed that the perceived choices primarily rely on individual UIs, instead of sequential UIs. Accordingly, we extracted 21 features, on the basis of UI components, and built up a machine learning model using XGBoost – XGB (T. Chen & Guestrin, 2016). Our collected user responses, demonstrated by XGB, enable the hand choice assessments, and inform the UI designers about the likelihood of one-handed interaction with mobile UIs.

The primary contribution of the paper is three-fold. First, we study the user preference for three hand poses (one-handed and one-thumb, two-handed and two-thumb, and two-handed and one-thumb) on smartphone UIs. On the basis of 62,156 responses among 3,000+ clusters of similar UIs (31,948 UIs in total), our analytics sheds light on UI reachability from the cognitive perspective of UI component layouts, on top of physical characteristics, albeit only three hand poses (Kirhenstein, 2019) are considered. Second, the discrepancy between physical and perceived choices of hand poses

[1]`https://www.nngroup.com/articles/mental-models/`

reveals aspects related to mental models of hand pose choices. Accordingly, our proposed XGB-based predictive model illustrates the potential of indicating the perceived choices of hand poses: either one-handed or two-handed interactions. The model considers 21 extracted features, and achieves a reasonable prediction accuracy (F1 score = 0.64) for individual UIs. Third, our data analytics and XGB model not only serve as a groundwork for UI auditing of user convenience with one-handed interaction, and allow UI designers to become more knowledgeable about whether the users will maintain one-handed interaction with the newly designed UIs.

The structure of the paper is as follows: Section 2 presents the interaction design of one-handed operations and the existing works on AI-assisted interface auditing. Section 3 explains our data collection pipelines and the details of three UI rating tasks. Next, Section 4 provided in-depth data analytics of the collected data, as the justification for modelling hand poses and pose changes based on the user perception. Accordingly, Section 5 describes feature extraction and establishes machine learning models. Finally, Section 6 discusses the results of model performance, and its implications for the existing knowledge of UI reachability.

2. Related Work

This section discusses the importance of user mobility on smartphone interfaces, and the existing solutions addressing the reachability of icons and menus on smartphone interfaces with one-handed and one-thumb operations.

2.1. Interaction Techniques for One-handed and one-thumb operations

The existing literature primarily considers one-handed and one-thumb operations as the preferred options for interacting with target icons and menus on smartphone interfaces with high levels of user mobility, where another hand is reserved for other tasks in the physical environment, such as holding a handrail or working tools (Lee et al., 2020). Some icons and menus at the hard-to-reach (Tsai et al., 2016) and unreachable (Kim et al., 2012) positions enforce the user to switch into two-handed operations by scarifying the user mobility. To reserve one-handed operations, commercial solutions, e.g., Apple's iPhones, shift the upper half of the user interfaces down to the lower half after double-tapping the home button.

Moreover, various research prototypes enable users to reach targets (e.g., icons and menus (Lai & Zhang, 2015)) on smartphones with the convenience of one-handed and one-thumb operations (Yu, Huang, Hsu, & Hung, 2013), considering multitudinous input modalities, e.g., detecting head movements by a smartphone rear camera (Voelker, Hueber, Corsten, & Remy, 2020), exerting touches (Boring et al., 2012) or press (Corsten, Lahaye, Borchers, & Voelker, 2019; Suzuki, Sakamoto, Sakamoto, & Ono, 2018) on touchscreens, edges areas (Quinn, Lee, Barnhart, & Zhai, 2019), and the back of the smartphones (Corsten, Daehlmann, Voelker, & Borchers, 2017; Le, Bader, Kosch, & Henze, 2016), in either portrait (Hakka, Isomoto, & Shizuki, 2019) or landscape (Borah & Sorathia, 2019) modes. Apart from the interaction techniques for easier access to the hard-to-reach and unreachable icons and menus on smartphone touchscreens, another important aspect is to allocate the icons and menus within the comfortable zones inside UIs that are easily accessible by one-thumb interaction.

Despite the fact that various techniques have been proposed for enhancing single-handed reachability on touchscreens, they ignore the impact of UI designs on reachability. In fact, maximized reachability should not be a concern when the targets on touchscreens are within the proximity of the user's hand. Therefore, this study focuses on the relationship between UI components in layout design and reachability.

2.2. Quantifying the ease of reaches

The majority of the existing works have only paid attention to the physical contexts of the users. Ergonomic factors have been deployed to indicate the user tiredness of particular postures for certain tasks, for instance, the easiness of reaching some locations on touch surfaces (Salazar, Henze, & Wolf, 2016). The difficulties of reaching a target on a touch surface are quantified by anthropometric data, i.e., the users' motion and their maximum reaches (Toney & Thomas, 2007), while biomechanical simulation evaluates the relationship between muscle activation and effective user throughput of one-handed touch operations on smartphones in long-term usage (Bachynskyi, Palmas, Oulasvirta, Steimle, & Weinkauf, 2015). Fitts's Law concerns the pointing performance of certain targets, such as icons and menus, based on the target width and the travel distance (Bi, Li, & Zhai, 2013b), which can estimate the user's pointing errors on smartphone interfaces (Wobbrock, Cutrell, Harada, & MacKenzie, 2008). However, we consider these ergonometric measurements as some relatively indirect indicators of quantifying the reachability of an individual UI component, instead of UI components as a whole. Specifically, each metric usually focuses on one interaction element on the touchscreen at a time. Thus, the metrics neglect the layouts or the relative positions among the targets. Another prior study (Buschek, Auch, & Alt, 2015) quantifies the touch behaviours on individual clickable items inside websites and layouts, and predicts the touchpoint distribution on three types of clickable items, including list entries, images with captions, and links in the text. The work reveals that users did rather reach as far as necessary to hit the target comfortably, and further demonstrates the discrepancy between the designers' expected touch behaviours and the users' mental models, albeit with limited types of clickable items being considered. Due to failing to capture the arrangement of UI components on a touchscreen, ergonometric measurements cannot be used for the holistic evaluation of UIs. In contrast, our work attempts to provide such evaluation on the basis of the user's perceived preference for UI components in a layout design.

Prior studies on the reachability of touchscreen interfaces mainly consider the hand dimensions (Larsen et al., 2019; Prange et al., 2018) and the finger lengths (Yoo, Yoon, & Ji, 2015). The physical context of the user's hands constructs comfortable yet functional zones on smartphone touchscreen in 2D areas (Le, Mayer, Bader, & Henze, 2018) as well as 3D space (Hasan, Kim, Ahlström, & Irani, 2016), implying the easiness of reaching the target contents with various gestures such as pointing and dragging operations (Sarcar et al., 2019). Besides, the kinematics of the gripping hand have established more sophisticated models. A quadratic formula named *Parabolic Motion Trajectory* (Bergstrom-Lehtovirta & Oulasvirta, 2014) estimates the thumb coverage on various touchscreen sizes, and hence decides the maximum extent of reaching various icons and menus on smartphones, where the ways of holding a smartphone can impact the user performance, such as tapping precision and errors (Lehmann & Kipp, 2018). Additionally, other physical factors such as the changes of hand grips (Negulescu & McGrenere, 2015), grip shift (Eardley, Roudaut, Gill, & Thompson, 2017), body gestures (Eardley, Roudaut, Gill, & Thompson, 2018b) and movements (Eardley, Roudaut, Gill, & Thompson, 2018a), as well as the form factors of the touchscreens (Girouard et al., 2015; Pecchioli, Dubois, Irani, & Serrano, 2019) impact the ease of reaching clickable items on mobile UIs.

Despite all the physical measurements in prior studies, the reachability issues of the smartphone interfaces with the user's cognitive perspective have not been sufficiently addressed. The classical model of users' cognitive loops, namely the *Model Human Processor* (Card, Newell, & Moran, 1983), proposes a useful approximation for understanding and estimating users' actions and reactions to the user interfaces, albeit not the exact modelling of how the brain works. That is, the users have to make decisions on certain interfaces about how to respond, before the motor system (i.e., thumb(s)/ hand grip(s)) carries out the actions and reactions. Our work intends to address the reachability issue of smartphone interfaces by considering the user's hand poses under the cognitive dimension.

2.3. AI-assisted Interface Auditing

The classical models, such as Fitts's and Hick's Laws and their variants (Bi et al., 2013a; Cockburn, Gutwin, & Greenberg, 2007), offer relatively robust estimations of human performance in theoretical or mathematical models. However, these models are less adaptive to realistic environments, and have limited ability to provide analytical insights (Li et al., 2020). In contrast, the data-driven and crowd-powered programmatic methods for interface designs serve as an alternative to scientifically integrating human behaviours into the design process, and subsequently assessing the smartphone interfaces in multitudinous dimensions without costly user tests (Deka, Huang, Franzen, Nichols, et al., 2017). The programmatic methods on the basis of large-scale data analytics (Deka, Huang, Franzen, Hibschman, et al., 2017) and crowd-powered design guidance (Y. Chen, Pandey, Song, Lasecki, & Oney, 2020; Xu, Huang, & Bailey, 2014) enable automatic auditing of smartphone and web interfaces. The existing approaches show diversified goals and metrics, including gestural performance (Deka, Huang, Franzen, Nichols, et al., 2017), time perception (Huhtala, Sarjanoja, Mäntyjärvi, Isomursu, & Häkkilä, 2010), visual search performance (Yuan & Li, 2020), inconsistencies between UI design and actual implementation (C.-F. R. Chen et al., 2017), visual clutter to visual learnability (Oulasvirta et al., 2018), aesthetics (Miniukovich & De Angeli, 2015), visual affordance (Swearngin & Li, 2019a), brand understanding and impressions (Wu, Kim, Li, & Ma, 2019), user engagement (Wu et al., 2020), to name but a few. The above programmatic methods are capable of handling both quantitative and qualitative user-oriented goals and metrics with their specific goals. Nonetheless, relevant studies are rarely found considering both user mobility and user perception of hand poses on smartphones. Thus, our work demonstrates the potential of modelling the user's reaction (i.e., hand pose choices) to smartphone interfaces before the users move their thumbs in the physical world. To the best of our knowledge, this paper is *the first effort* to investigate how to leverage the UI dataset and machine learning algorithms to understand the users' perceived choices of hand poses (i.e., reachability issues) among smartphone interfaces.

3. User Study and Data Collection

We first conducted a 6-week user evaluation and collected user responses to the mobile UIs. This section describes the designs of experimental tasks and user feedback collection on the mobile UIs.

3.1. Design of Our Study

We employ a UI repository named Rico dataset (Deka, Huang, Franzen, Hibschman, et al., 2017) that contains 72k UI layouts, 6,372 user interaction traces (UIT), and corresponding vectors trained from autoencoder embeddings. Rico dataset was selected because of its comprehensiveness in mobile UI collection. Accordingly, our questionnaires include the UIs from the Rico dataset, and our research focuses on user perceptions of the UIs built on such a basis due to its comprehensiveness through including many varying UIs. This allows us to observe the user's choices in achieving mobility with single-handed and right-handed operations (Ko, Hwang, & Liang, 2016) for a longer UI sequence. Considering some individual UIs are highly similar (and thus their UI layout vectors) within the 72k pool of UIs, we apply k-means to cluster all 72,103 UI layout vectors into 3,000 groups based on the pairwise Euclidean distance between UI vectors. Within each group, we extract around 10 examples that are closest to the mean vector within the group, totaling up to 31,948 individual UIs, which become the dataset for our study. The clustering can effectively reduce computation complexity, and save effort in user evaluation by eliminating highly similar UIs and retaining those that are representative of the entire UI dataset. Figure 1(a) shows four k-means clustered groups, each of them carrying five UI examples. We can see the UIs are similar in layouts within each group but distinct across groups. Thus, it implies that k-means provide the subsets of UIs that are representative of the original population. We further exclude the UI sequences if any UI hierarchy has less than three UIs. This allows us to observe the user's choices in achieving mobility with single-handed and right-handed operations (Ko et al., 2016) for a longer UI sequence.

We use content-agnostic similarity heuristics to include UIs in our dataset. As a result, our dataset has a total of 31,948 individual UIs and 6,011 UITs in our study. Accordingly, we design *three UI rating tasks* and host them on Amazon Web Services, to study users' perceived convenience in UI component layouts, on the basis of single- and right-handed smartphone usage. The UI component layouts refer to the 'tappable' UI components in a UI (Swearngin & Li, 2019a), such as buttons, spinner, and listviews. We keep this term of UI component layout(s) in Sections 3 and 4.

The UI rating tasks are held on web interfaces. The reasons for deploying the UI rating tasks on web interfaces are as follows. Based on the classical theory of the *Model Human Processor* (Card et al., 1983), human users first make cognitive efforts to plan the user interaction before the corresponding physical actions are executed on an interface. Therefore, our study focuses on the cognitive aspect of reachability in the user's planning, namely perceived reachability. The perceived reachability should be free of restrictions on the choice of the external medium (i.e., web or mobile interfaces), as the UI component layouts do not vary across devices. Deploying the tasks onto web interfaces distinguishes our study from the classical ergonometric-driven approach towards reachability. We did our utmost to dismiss the physical measurements, as one of the confounding factors, that interfere with our measurements for perceived reachability, focusing on the clickable items in an overall UI layout. On the other hand, we observe an increasing number of mobile UIs studies leveraging web interfaces as experimental platforms (Deka, Huang, Franzen, Hibschman, et al., 2017; Li, Bengio, & Bailly, 2018; Swearngin & Li, 2019b), due to the ease of deployment and the relatively higher compatibility across various devices than the counterparts of having a tangible device per user.

3.2. Participants

Afterwards, we recruited 17 participants (p) from our campuses ($\bar{Age}$: 19.65; (19 – 23, σ=1.25)). Additionally, all the participants have at least five-year experiences of smartphone usage ($\bar{Y}$=6.48, σ=3.43), which leads them to a reasonable level of technology literacy and hence confidence in understanding the widgets, notifiers, as well as commonly available user interaction in the UIs from the Rico dataset (Deka, Huang, Franzen, Hibschman, et al., 2017). They responded to the three tasks within six weeks, with the divisions – Task 1: 10p; Task 2: 8p; and Task 3: 10p. Their smartphone sizes range from 67.1 mm to 80 mm in width ($\bar{W}$=72.4, σ=5.06), 135 mm to 168 mm in height ($\bar{H}$=150.56, σ=10.46). The span length of their right hand, i.e., the distance between the tip of the thumb and the tip of the index finger (Figure 5(right)) when stretched horizontally, ranges from 160 mm to 185 mm ($\bar{s}$=175.77, σ=9.21). The index finger length of their right-hand ranges from 70 mm to 100 mm ($\bar{d}$=82.32, σ=8.29). The grasp position when holding the smartphone, i.e., the distance between the bottom of the smartphone and the thumb metacarpal joint, ranges from 30 mm to 86 mm ($\bar{r}$=50.07, σ=13.30). The thumb orientation, i.e., the angle between the proximal phalanx of the thumb and the metacarpal bone of the thumb while holding the phone, ranges from 111° to 155° ($\bar{o}$=131, σ=12.13). The anthropometric properties of the 17 participants are used for the construction of Thumb Reachability Scores to be introduced in Section 4.1. On the basis of informed consent and truly voluntary, one extra course credit (P), with no discrete grades (e.g., A–F), was distributed as the certificate for their participation.

3.3. Task Design and Collected Data

In all three tasks, participants were asked to judge the perceived reachability of UIs. The tasks aim to collect the user's perceived ease of interacting with the most clickable items on the UI(s) when adopting particular pose(s) among the candidate UIs. The study follows the regulations of the GDPR and the IRB of our universities. All user-generated data was de-identified and password-protected, and the participant information will be deleted after the completion of the project.

We consider the UI component layouts in individual UIs and sequential UIs, as the information being received by users (Kim & Lee, 2018; Picard, 2000) for the users' decision of their hand pose choices, in which the hand pose choices consists of the selection of *hand poses* and the occurrence of *pose changes*. We, therefore, collect participants' responses in terms of pose choices for both individual and sequential UIs. During all tasks, the participants were informed about the reference smartphone model of *Samsung Galaxy Note 4* (resolution: 1440 × 2560 pixels; and physical dimensions: 153.5 mm (l) × 78.6 mm (w) × 8.5 mm (h)). The information was displayed every time after the login pages, with comparable references to other popular smartphone models, including iPhones SE/ 7/ 7plus/ XS, and so on.

3.3.1. Task 1

Task 1 aims to understand the user perception of one-handed reachability; thus, the options of two-handed reachability were excluded. Thus, the task allows users to judge the layouts and compare the difficulties of one-handed reachability among items (e.g., icons, menus, and windows). The results offer us a baseline of the difficulties of employing one-handed interaction on particular UIs, which may offer evidence of further

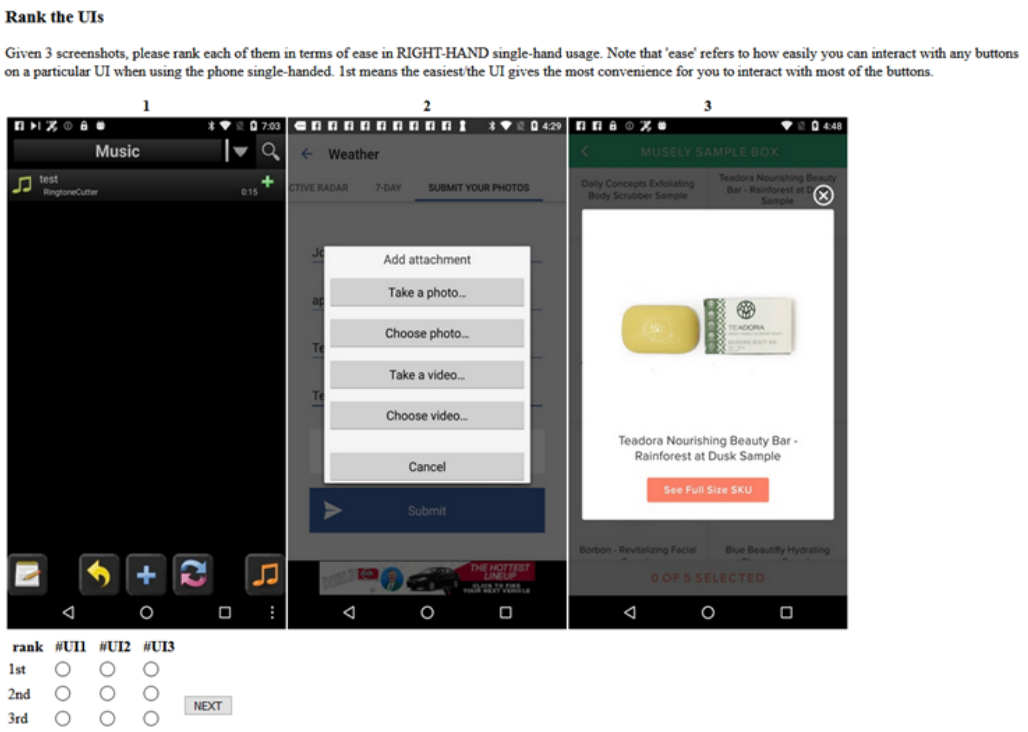


Figure 2. Task 1 Interface: ranking the UIs in terms of the ease of one-handed reachability (only one one-handed hand pose is considered): Three interfaces (top) and Questionnaire (bottom).

considering user perception, on top of the physical ability of thumb reaches. As such, in Task 1 (Figure 2), we asked the participants to rank the UIs in terms of the ease of one-handed interaction with most of the clickable items on three different layouts. In other words, no two-handed hand poses were being considered for the task. The 1st rank UI (leftmost) refers to the one that gives the most convenience to the users, while the 3rd rank (rightmost) refers to the least convenient option. We collected 25,884 responses to the 3-UI sequences. It is also important to mention that the participants did not know the aforementioned ranking, computed by Thumb Reachability Score (TRS) (to be described in Section 4.1), during the task.

3.3.2. Task 2

The design of Task 2 aims to understand the preference of hand poses when they encounter one particular interface. Thus, three options for hand poses, including (1) one-thumb, (2) two-thumb, and (3) two-handed but involving one-thumb, were shown to the users. By collecting the user preference on hand pose, we can verify the difficulties of one-handed reachability and user preference on hand poses. The results also help us to reinforce the need to develop computational models for suggesting hand pose choices based on user perceptions. In Task 2 (Figure 3), we present users with a single UI and ask them to pick one of the three poses that gives them the most convenience in interacting with most of the clickable items on that particular UI. The hand pose of two-handed but involving one-index finger is omitted, as we found, in our pilot study, that users make indistinguishable choices between one-thumb and one-index under

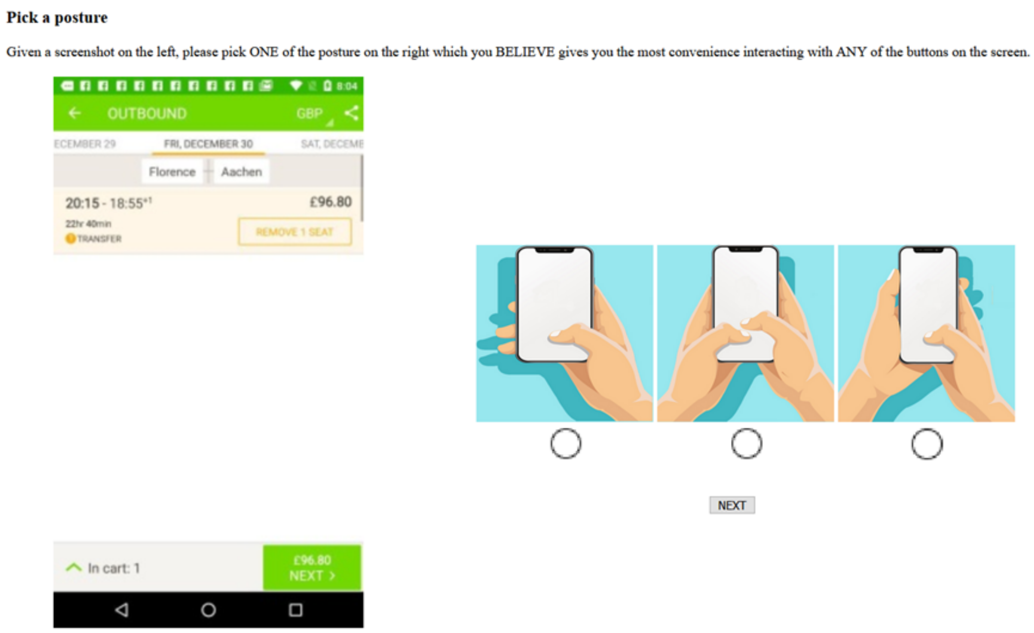


Figure 3. Task 2 Interface: hand pose selection for an individual UI: An interface (Left) and Questionnaires (Right).

two-handed scenarios, as both hand poses are characterized by a supporting hand. We collected 30,261 responses regarding the user preference for hand poses on individual UIs. That is, ∼10 responses link to one UI cluster among the ∼3000+ k-mean cluster.

3.3.3. Task 3

The goal of Task 3 is to investigate whether any relationship exists between the user's hand poses and the sequential UIs. We attempt to check whether the users view each UI independently or not. Therefore, the task displayed a series of UIs, and the users gave responses as the selected hand poses. The results will impact the design of the computational model in Section 5. In Task 3 (Figure 4), we assess the participants' responses to the sequences of UI interaction trace. The participants were asked to pick hand pose(s) among those three (the identical hand poses being described in Task 2, on the bottom right of the screen). The participants chose a hand pose that gave them the most convenience in interacting with most of the clickable items on each of the UI layouts in the sequence, i.e., considering multiple UIs in the sequence. We collected 6,011 responses of sequential UIs by employing 6,011 UIT.

4. Observations among the user choices of hand poses

This section explains the observation of the collected responses. We generalize single-handed user behaviours with clickable items in the large set of UI layouts. Accordingly, the collected dataset of perceived reachability gives justifications for building our predictive model (Section 5).

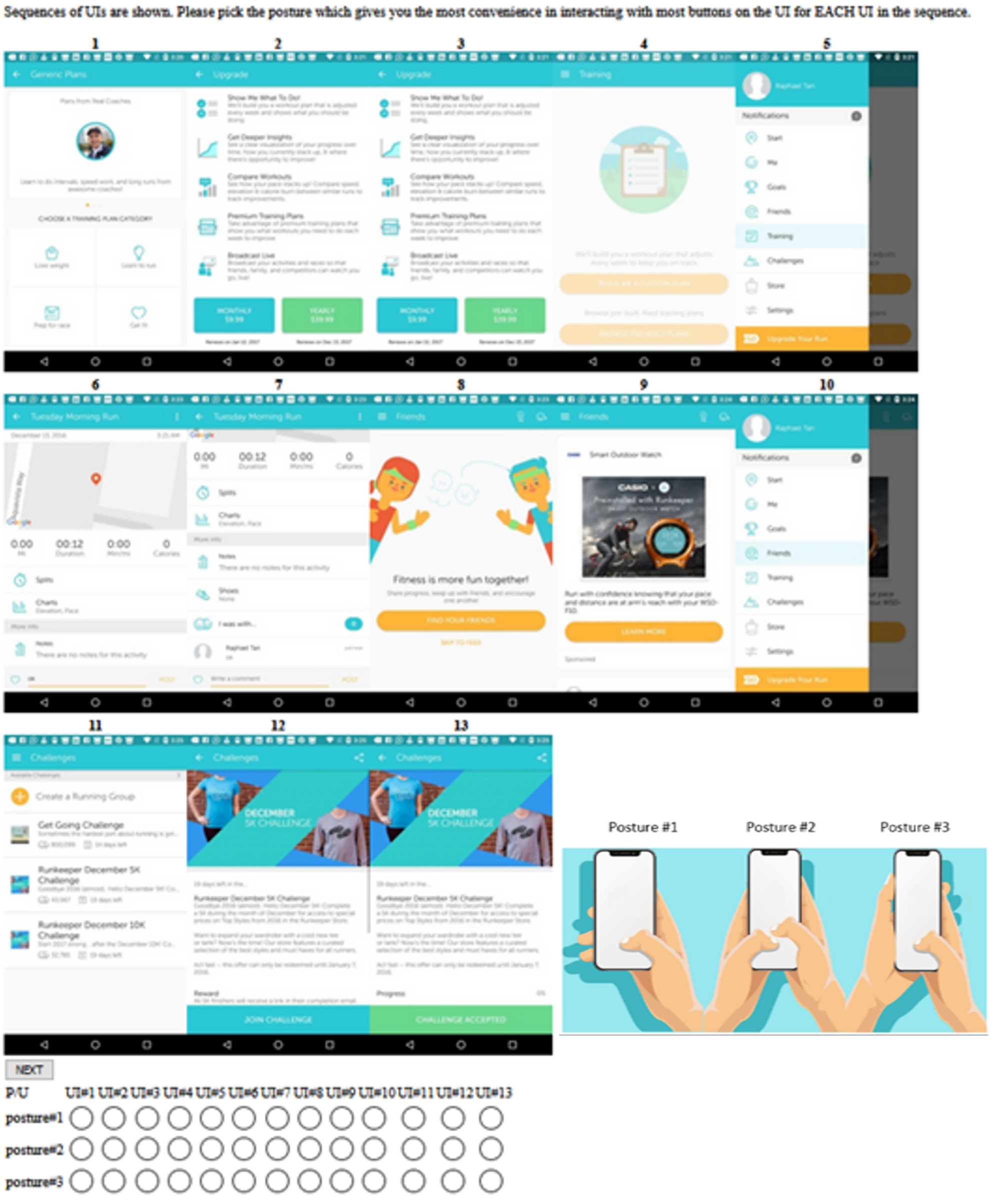


Figure 4. Task 3 Interface: selecting hand poses for multiple UIs in sequential order (e.g., 14-length sequential UIs in this figure), extracted by the UIT generated by Rico (Deka, Huang, Franzen, Hibschman, et al., 2017): Sequential Interfaces (top) and Questionnaire (bottom)

4.1. Preamble: Thumb Reachability Scores

Since our participants' responses are regarded as the perceived ease of reachability of particular UIs, we first attempt to grab the big picture of the user's perceived choices. Then, we investigate any divergence between such perceived choices and their physical reachability, named thumb reachability score (TRS) originated from the study in (Bergstrom-Lehtovirta & Oulasvirta, 2014).

Our baseline of TRS is explained as follows, which is compared to the perceived choices in the next section. As shown in Figure 5, the key idea of TRS is to estimate

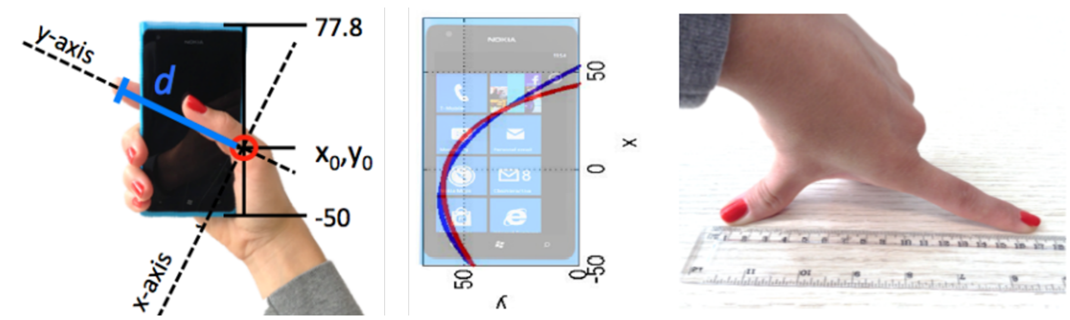


Figure 5. (Left) Measurements of physical thumb reaches, (Middle) Parabola for Thumb Functional Area, sourced from (Bergstrom-Lehtovirta & Oulasvirta, 2014), and (Right) the measurement of a user's hand between the tip of the thumb and the tip of the index finger.

the thumb coverage by computing a parabola, i.e., the estimated functional area of the thumb (Bergstrom-Lehtovirta & Oulasvirta, 2014). That is, the hand parameters of users serve as the hard fact to depict the functional area of the thumb physically on a mobile UI. The area below the parabola is regarded as the reachable area with one-thumb interactions, while the remaining areas above the parabola are classified as unreachable. Accordingly, the UI components in a mobile UI are represented by the physical-context value named TRS. We acknowledge that the binary classification of areas as 'reachable' and 'unreachable' could be overly simplistic. It might not capture the real nuances of the user experience, which needs gradations or ranges to represent different levels of ease or difficulty in reachability.By substituting the results from (Bergstrom-Lehtovirta & Oulasvirta, 2014) and obtaining a_y=-0.013,h=-15.996,a_k=-0.257,b_k=64.655, and the averaged results from our participants' statistics of s=175.77 mm,d=82.32 mm, we derive Eq. 1, with x and y being the coordinates of the phone along with the height and width dimension respectively. 's' indicates the span of the user's hand, and 'd' refers to the length of an index finger. a_k and b_k are the estimated coefficients, and a_y as the curvature. Next, we apply Eq. 1 into our dataset to assess the UI component layouts, with the average grasp position of 50.07 mm from the bottom of the smartphone to the thumb metacarpal joint, and the average thumb orientation of 110° between the proximal phalanx of the thumb and the metacarpal bone of the thumb. Finally, TRS measures the proportion of UI-component pixels of an interface laying underneath the parabola curve against the total number of pixels from all clickable items on the same interface. TRS is perceived as the percentage of thumb reachability (Eq. 2), where b_i stands for the total number of pixels for the i^{th} UI component on an interface, while the binary-valued $\hat{r_{i,j}}$ stands for whether the j^{th} pixel of the i^{th} UI component lies beneath the parabola curve of the interface (value 1) or not (value 0). N refers to the total number of clickable items on a UI, and T refers to the total number of pixels in a UI component. Thus, the normalized value of TRS [0,1] reflects the user's accessibility in a UI with one-handed operations, where 1 represents the best thumb reachability for a UI component layout on an interface as all clickable items are under the parabola curve, or vice versa.

$$f_y(x) = a_y(x+h)^2 + s - d(1+a_k) - b_k = -0.013(x-15.996)^2 + 49.951 \qquad (1)$$

$$TRS = \frac{\sum_{i=1}^{N} \sum_{j=1}^{T} \hat{r_{i,j}}}{\sum_{i=1}^{N} b_i} \tag{2}$$

4.2. Analysis of User Responses in Task 1: UI Ranking with a Fixed Pose

From the collected 25,884 users' ranking of UIs in Task 1 (Section 3.3.1), we study the ease of single-handed (right-handed) interaction considering all clickable items in an individual UI.

4.2.1. Confusion Types

Under the consideration of one-handed interaction on mobile UIs, we study the ranked UIs by comparing the collected user preferences (i.e., perceived choices of hand poses) and the computed TRS (i.e., theoretical physical reaches), defined as Confusion Types (CT). The parabola curve (Bergstrom-Lehtovirta & Oulasvirta, 2014) represents the TRS-supported rank. Meanwhile, CT arises once the mismatch between users' cognitively preferred rankings and the TRS-supported rank exists. To make Confusion Types (CT) easier to understand, we first consider the three UIs that are ranked in ascending rankings based on TRS, leading to a CT choice of '123'. Therefore, the ideal case of the user's perceived choices should follow the same order, and hence *no confusion* in the user's ranking has been found. This means that the user ranks the UIs completely aligned with the physical context of thumb reachability, i.e., ascendingly ordered TRS of the UI sequence in {1,2,3}.

Nevertheless, the most occurred CT is '213' (33.9%), i.e., the participants reversed the ranking between the 1st and 2nd UIs inside the ascendingly TRS-ranked UI sequences. The 2nd (CT '123', 20.9%) and 3rd (CT '132', 16.1%) ranked CTs follow. The orders of '213' and '132' represent adjacently reversed ranking, indicating the physical contexts of TRS show sight discrepancy from the user's perception of UI components in UI layouts. CTs '321' and '231' (14.4%) reveal more significant deviations from the physical context TRS, with the ranking shift by two positions. Figure 6(left) summarizes the counts of position shifts among aggregated CTs.

4.2.2. Quantification of Ranking Discrepancy

Furthermore, we quantify such reversed rankings, defined by Sum of Score Difference (SSD) in each CT, as follows. We analyze all ascendingly ranked sequence UIs being employed in the interfaces of Task 1. The ascending score sequences are manipulated as in Eq. 3, where S(u_i) and S(u_{i+1}) represent the TRS for the UIs ranked at i and i+1, respectively. Since each score difference must be positive, the re-arranged form of Eq. 3 quantifies the sum of differences in TRS between UIs in every three-sequenced UIs.

$$SSD = \sum_{i=1}^{2} S(u_{i+1}) - S(u_i) \quad = S(u_3) - S(u_1) \tag{3}$$

Figure 6(right) visualizes the distribution of SSD in each shift type of aggregated CT groups. We find a remarkable trend: 1) the CTs with higher counts tend to have

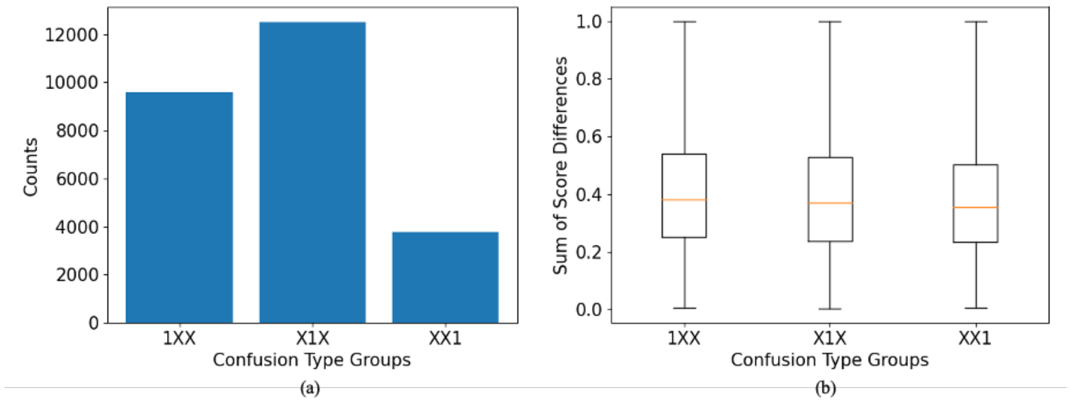


Figure 6. Statistics of Position Shifts of TRS-preferred ranking (1XX = CTs '123' and '132'; X1X = CTs '213' and '312'; XX1 = CTs '231' and '321') – (left) The Counts; (right) SSD of all CT groups.

a larger SSD in their UI sequences; 2) the range of the SSD has been decreasing with the popularity of CT, highlighting the fact that the more subtle the TRS differences between UIs within the sequences, the more likely confusion might occur. Additionally, we take a closer examination into the SSD between reversely ranked UIs, i.e., the reference CT '123' is excluded. The SSD in CT '213' is the lowest among other CTs, illustrating why CT '213' has a more significant number of counts among all CTs. The increasing trend in the SSD in Figure 6(right) complies with the decreasing trend of CT counts in Figure 6(left) and provides quantitative explanations for the trend of occurrence in Figure 6(left).

4.2.3. The Key Implication

To conclude, a significant number of user responses reflects the ranking reverse of one position, e.g., CTs, '213'. However, only 21% of user responses follow the TRS-ranked '123' under the physical-context of TRS. The general trend shows that CTs with higher counts have higher SSDs than their counterparts with lower SSDs. Our findings highlight the discrepancy between human perception of UI component layouts and the physically formulated TRS (i.e., sole reliance on the physical metric of thumb length). This implies that the TRS derived from (Bergstrom-Lehtovirta & Oulasvirta, 2014) can only partially serve as an indicator of single-handed thumb reachability on UIs, as it misses the user's perception about the UIs before executing thumb reaches. As such, in terms of UI design, both means of user perception and physical metrics should be considered holistically.

4.3. Analysis of User Responses in Task 2: User Preferred Poses in a UI

According to the collected 30,261 user preferences on the three hand poses in Task 2 (Section 3.3.2), we examine the pose choices in single-handed and two-handed interaction with the most clickable items on the UI in the dataset. Figures 7(left) and (right) depict the popularity of three hand poses and the relationship between UIs' TRS and the popularity of the hand poses. Decreasing preference across Poses 1 to 3 occurs (Figure 7(left))– 65% of participants prefer Pose 1 (single-handed), which is significantly more than that of Pose 2 (20%, two-handed and two-thumb) and Pose 3

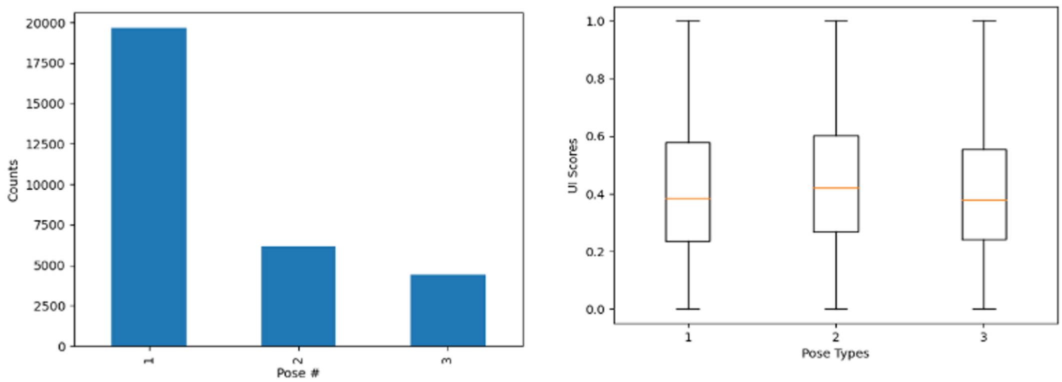


Figure 7. The counts of hand poses (1 – 3) were chosen by the participants (left) and the boxplot for TRS scores of three hand poses (right).

(15%, two-handed and one-thumb). Pose 2 is more popular than Pose 3, as the participants attempt to maximize the utility of two-handed interactions, considering the left hand in Pose 3 acts only as a supporting hand, instead of an actor for interacting with the UI. Accordingly, we investigate the thumb reachability score (TRS) distributions for each pose. Under the definition of TRS values normalized within 0 and 1, most UI-component pixels in a UI with Pose 1 are under the coverage of the physical representation of the parabola curve (Bergstrom-Lehtovirta & Oulasvirta, 2014) (Figure 5), while UIs with Poses 2 and 3 mean the UI-component pixels are hard to reach and hence trigger two-handed interaction. Intuitively, the UIs with Pose 1 should reflect higher TRS than Poses 2 and 3.

However, this intuition does not hold. Pose 1 ($T\bar{R}S_1$= 0.41, $\sigma_1 = 0.241$) is lower than Poses 2 ($T\bar{R}S_2$= 0.44, $\sigma_2 = 0.227$), and show a similar score to Pose 3 ($T\bar{R}S_3$= 0.41, $\sigma_3 = 0.222$). Among the TRS distribution for each pose (Figure 7(right)), Pose 2 unexpectedly owns the highest score distributions, instead of Pose 1. The phenomenon indicates that the computed TRS from the predictive formula does not completely dominate the choice of hand poses, i.e., not capturing the user's (cognitive) perception of thumb reachability. Importantly, the UIs, matched with the two-handed interaction of Pose 2, result in unaligned high TRS, which serve as additional evidence for the existence of discrepancy between perceived choices to UI component layouts and physical reachability.

Although the TRS score aims to quantify how thumb-friendly a UI should be, TRS does not sufficiently address two aspects. First, TRS employs a binary classification of UI-component pixels into either above or under the parabola, but fails to consider the relative positions of clickable items within an individual UI; Second, TRS neglects the effect beyond UI component layouts, e.g., the user's habits and their preference for hand poses. Therefore, the most popular pose (Pose 1), despite its lower TRS (Figure 7(right)), implies other considerable factors in the hand poses, perhaps the user's habits to the overall UI component layouts in a UI (More details in Section 4.4). In other words, the participants select Pose 1 even though some clickable items are less convenient to reach with one-handed interaction.

4.4. Analysis of User Responses in Task 3: Pose Preferences on Sequential UIs and User Habits

While Sections 4.2 and 4.3 focus on the analysis of pose preferences on individual UIs, this section attempts to strengthen our understanding of such divergences in hand poses (perceived vs. physical) from the perspective of user habits. To this end, we attempt to examine the dynamics by which the user interacts with a sequence of UIs, i.e., the impacts of UI component layouts in sequential UIs on the user's pose preferences. We also investigate any relationship between the preferred hand pose on individual UIs and the user habits caused by sequential UIs. As the Rico Dataset only offers sequential UIs and corresponding user interaction traces in the same application, our findings below cannot reflect the user interactions when a user switches from one application to another.

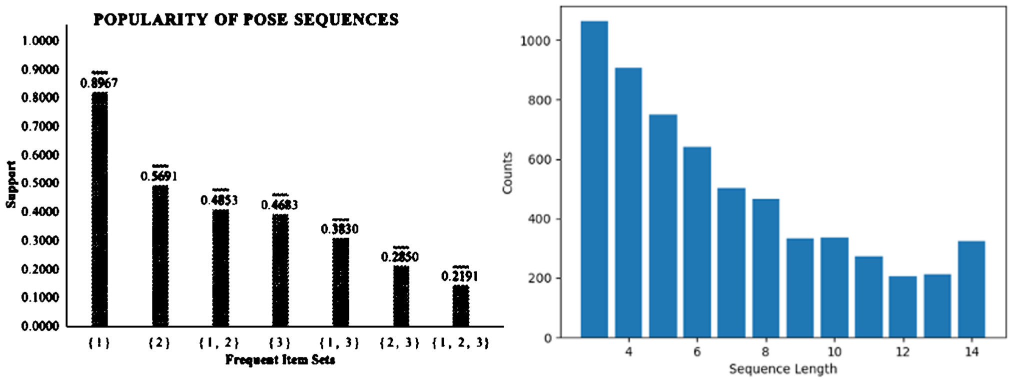


Figure 8. (left) Popularity of Pose Sequences that shows the (co-)occurrence of hand poses in sequential UIs of varying lengths; for instance, 1,2 indicates that users with only Poses 1 and 2 can accomplish the interaction with one particular set of sequential UIs. Among the pose sequences, 1, 2, and 1,2 are the most popular options; and, (right) Distribution of Sequential UI length indicates the counts of sequential UIs in varying lengths, ranging from 3 to 14. The statistics show that the participants interacted mostly with the sequential UIs of 3 – 9 lengths.

4.4.1. Does the inertia of hand poses exist?

By leveraging the 6,011 sequences collected in Task 3, we start the analysis by mining the most Frequent Itemset (FIS) in pose sequences with apriori algorithm (Huang, Chen, Wang, & Chen, 2000). FIS indicates a combined option of hand pose(s) to accomplish a series of user interactions on a sequence of UIs. For instance, {1} indicates that hand pose 1 can accomplish the interaction throughout a sequence of UIs, and {1,2} indicates hand poses 1 and 2 are sufficient to interact with a sequence of UIs (Figure 8(left)). Quantitatively, '*support*' refers to the percentage of the time, in the entire sequence dataset, that an FIS appears. We identify the most popular pose sequences during the user's interaction with a UI sequence. Figure 8(left) shows only those FIS with *support* greater than 20%. And Figure 8(right) illustrates the distribution of UI sequence lengths. It is important to note that the FIS has a non-repetitive property, which does not allow duplicating itemsets and relationships among items across itemsets. Thus, each element will only appear once in any FIS. For example, FIS {1} refers to the existence of Pose 1 in any *sub*-sequences of at least one or more consecutive UI(s). Figure 8(left) has shown that the most frequent FIS is {1}, indicating the prevalence of single-handed smartphone usage. With the *support* of nearly

90%, the FIS {1} is the most prevalent across all UI sequences of varying lengths. The *support* values of {1}, {2}, {3} are in descending popularity, which aligns with the results in Section 4.3 (Figure 7(left)), with the remarkable evidence of that {1,2} ranks over {3}. The users prefer to fully utilize their thumb(s) from the operating hand(s) to achieve the greatest ease with the interaction with most clickable items inside UIs.

Figure 8(left) also shows that the FIS containing more elements tends to have significantly lower *support*, with the following two key implications. First, the participants do not change their pose frequently across UI sequences of varying lengths. Either consecutive UIs tend to be very similar, perhaps due to the industry practices, or the inertia of in-use hand pose exists, i.e., the user is reluctant to change poses, albeit the user's thumb can only barely reach the target without the maximal comfort of physical reaches. Second, most participants in the studies tend to maintain one or two most preferred poses, with a low preference for employing all three poses in a sequence.

4.4.2. How does the length of sequential UIs impact the hand poses cognitively?

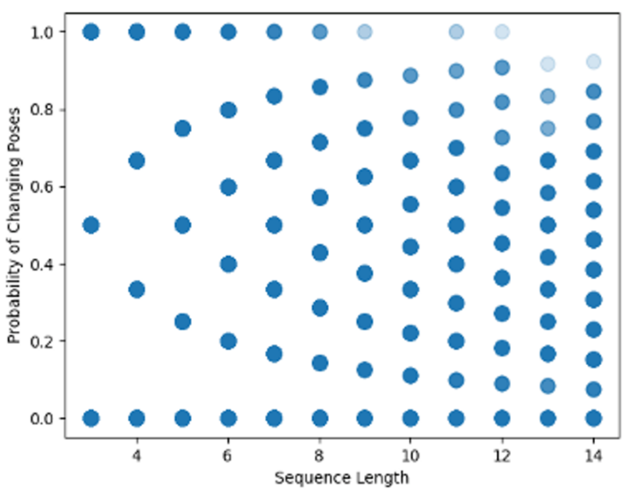


Figure 9. Pose-Change Probability Distribution represents the relationship between the length of sequential UIs and the probability of triggering the switch(es) of hand poses, in which the faded bubble means a lesser chance of switching hand poses, while a bubble of solid colour represents a higher chance of switching hand poses. For example, at the length of 4, the four bubbles describe the number of pose changes among the transitions between the four consecutive UIs, representing probabilities for switching 0 (no changes) – 3 poses (0/3, 1/3, 2/3, and 3/3 from the bottom to the top along the y-axis) in a 4-length UI sequence. Among all the sequential UIs of varying lengths, it is obvious that pose changes are unavoidable, especially in lengthy sequential UIs. However, the solid colours of the bubbles at the bottom shed light on having no pose changes, in case the UIs consider the users' abilities.

Next, we further examine the effects of the length of UI sequences. Figure 9 presents the probability of changing poses at UI sequence length [3,14]. Each probability, represented by a bubble, is computed by Eq. 4, measuring the frequency of the user will alter poses across a set of UIs, where u_i stands for a UI at position i, and y_i represents whether a user alters one's hand pose between two subsequent UIs u_i and u_{i+1}. $Q(u_i)$ represents a user's pose at u_i.

$$P(pose_{k\to j}) = \frac{\sum_{i=1}^{N-1} y_i}{N-1}, y_i = \begin{cases} 1 & Q(u_{i+1}) \neq Q(u_i) \\ 0 & else \end{cases} \tag{4}$$

Within the normalized score [0,1], 0 refers to maintaining the same pose from the beginning to the end in an interaction trace with a particular UI sequence, while 1 indicates that the user's pose will change upon every new UI in the trace, indicating the user inconvenience caused by altering hand poses across UI sequences. A more transparent bubble indicates a lower *count* of the probability the bubble is representing, and vice versa. For example, at the length of 3, three bubbles describe the number of pose changes among the transitions between three consecutive UIs, representing probabilities for switching 0 – 2 poses (0/2, 1/2, 2/2, from the bottom to the top along the y-axis) in a 3-length sequence. The three bubbles have comparable transparency, indicating these three probabilities are approximately equally likely to occur. It is important to note that the UI sequences with length [¿8] show bubbles of faded colours over 0.9 probability. Thus, a lengthy UI sequence is less likely to trigger pose changes for every UI interstice (i.e., from one UI to another). In other words, users with real-life applications have less tendency to change their hand poses for each UI. This can be explained by long sequential UIs that usually stay on a similar interface, e.g., an additional button or menu on top of the existing interfaces. On the contrary, the UI sequences with length [≤8] demonstrate more solid colours, i.e., shorter UI sequences are more likely to trigger high numbers of pose changes. It is worth mentioning that we inspected some samples of UI sequences in our datasets and spotted similarity among UIs in a sequence (to be discussed in Section 4.4.3).

4.4.3. What causes lesser switching of hand poses in longer UI sequences?

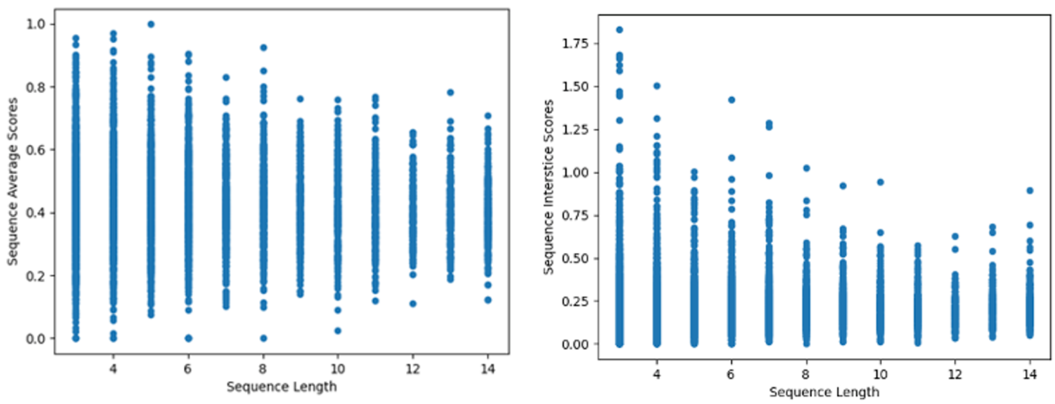


Figure 10. Based on the normalized score TRS, we compute (left) Inter-sequence Scores (SAS) to examine the averages of TRS scores in sequential UIs of varying lengths. SAS further reveals the distribution of TRS, in which lengthy sequences tend to have a centered distribution between 0.2 and 0.8. On the other hand, (right) Interstice (Intra-sequence) Scores (SIS) depict the dynamics (changes of values) among UIs in sequences of varying lengths, which indicates lengthy sequences own similar designs and hence less varied physical reachability.

In order to verify that UIs in longer sequences adopt similar designs and thus lead to lower switching occurrences, we inspect the TRS of UIs in each sequence and visualize the results in Figure 10. TRS can implicitly define the layouts or interaction tasks in a UI (Bergstrom-Lehtovirta & Oulasvirta, 2014). Thus, multiple TRS for sequential UIs can be considered as a reference for the changes in physical requirements of user interaction. Intuitively, the high variance of TRS in a sequence, in general, reflects sudden changes in layouts or interaction tasks within the sequential UIs and, hence, triggers pose changes. Therefore, we propose Sequence Average Score (SAS) and Sequence Interstice Scores (SIS) to evaluate the averaged values of TRS in UI sequences of varying lengths and the changes of TRS values in UI sequences, respectively.

We define the SAS in Eq. 5 as a measure of the average TRS for each UI in a sequence. S(u_i) refers to the TRS of u_i, which is the UI at the i^{th} position in a sequence. N is the length of the entire sequence. The normalized measure of SAS allows the comparison of the UI sequences of various lengths. Figure 10(left) indicates that the range of SAS is reversely proportional to the sequence length [3,14]. That is, shorter UI sequences tend to have highly polarized SAS values (e.g., 3-length: $S\bar{A}S$= 0.446, σ= 0.169). Reversely, longer UI sequences tend to have narrower SAS ranges than the shorter counterparts, and concentrate at relatively lower SAS (e.g., 14-length: $S\bar{A}S$= 0.394, σ = 0.093). Figure 10(left), however, only reflects the SAS dispersion in UI sequences.

To reinforce the above findings, another metric – SIS (Eq. 5), measures the average value of the absolute TRS difference of all UIs within a sequence of UI transitions (i.e., *Interstice*), where notations, as defined previously, are applicable here. Figure 10(right) shows the distribution of SIS at UI sequence length [3,14]. We observe a decreasing trend of SIS (i.e., the average differences in TRS between UIs in a sequence) as the length of UI sequence increases (e.g., 3-length: $S\bar{I}S$= 0.261, σ = 0.240; 14-length: $S\bar{I}S$= 0.217, σ = 0.1). Considering both the SAS and hence SIS, UIs in longer sequences tend to have a similar design, due to their similar TRS values, and therefore trigger a lesser number of pose switching. The implication is counter-intuitive to the (rejected) speculations that the number of pose switching is directly proportional to the UI sequence length.

$$SAS = \frac{\sum_{i=1}^{N} S(u_i)}{N}, \quad \text{and} \quad SIS = \frac{\sum_{i=1}^{N-1} |S(u_{i+1}) - S(u_i)|}{N-1} \tag{5}$$

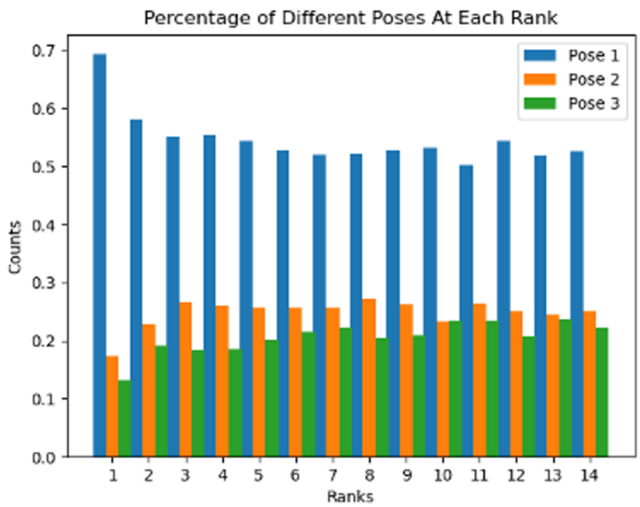


Figure 11. Distribution of in-sequence UI Positions: Ranks (x-axis) indicate the occurrence of hand poses at a particular UI position in sequential UIs, while Counts (y-axis) refer to the breakdown of hand pose choices (in percentages). We observe that no significant difference exists between hand pose choices and their occurrences in sequential UIs of varying lengths, serving as evidence of omitting the potential effects of sequential UIs on the choices of hand poses.

4.4.4. Are there any impacts of UI positions in sequential UIs to hand poses?

Finally, we inspect the effect of UI positions in a sequence, as follows. We hypothesize that the UI positions in a sequence do not influence the user's pose preference. We observe constant patterns across the orders in Figure 11 – Pose 1 (23,128 counts) is overwhelmingly larger than that of Pose 2 (9,871 counts) and Pose 3 (7,766 counts). Such consistency in patterns implies the UI positions (i.e., the order) within a sequence show no significant effects of UI positions on the hand pose preference. Remarkably, one-handed interaction (Pose 1, 4,166 counts) appears more frequently, as the 1st-position UI in the sequences, than the alternatives of two-handed interactions (Pose 2, 1,047 counts; and Pose 3, 798 counts). One of the key reasons leading to the highest count of Pose 1 can be accounted for by the influence of user habit, i.e., the pose of initialization or 'warm-up' in their daily app's usage (Ko et al., 2016).

4.5. A Summary of Key Findings

By summarizing the key findings as discussed above, we justify the design rationale for the proposed predictive models of hand poses. In Sections 4.2 and 4.3, we clearly observe the discrepancy between physical context TRS and the perceived hand poses in individual UIs. As proposed by *Nielsen Norman Group*, a mental model describes "*What users believe they know about a UI strongly impacts how they use it. Mismatched mental models are common.*"

In the context of this paper, the mental model highlights the gap between the full information inherited in the UIs and the perceived information by the users. In other words, the physical layouts, reflected by TRS, represent full information, while users leverage their own perceived information to analyze the UIs. The first implication of our findings for existing knowledge is that, by examining the crowdsourced data, we detail the relationship between TRS and the user's cognitive choices on individual and sequential UIs for one-handed interactions. According to our results, the perceived hand poses deviated from the TRS, representing a gap between the users' mental model and the external medium embedded in the physical contexts of UIs. As such, we connect all key findings in Section 4.4, offering insights into the gap and its validity in sequential UIs.

The second implication is that while thumb length and grip sizes support TRS's justifications, the existing theory oriented to physical ability ignores the user's habit or pose preferences as well as the inertia of in-use hand poses. A phenomenon of hand-pose inertia exists (Section 4.4.1), in which users attempt to complete a sequence of UIs with a limited set of hand poses. For instance, Pose 1, Pose 2, Poses 1 and 2, and Pose 3 are the most popular combinations (in descending order), as depicted by the Frequent Itemset (FIS). Also, as shown in Section 4.4.2, the occurrences of hand pose changes decrease in longer sequential UIs. Next, we further examine the TRS among sequential UIS of varying lengths (Section 4.4.3). The external medium, i.e., physical context TRS, demonstrates decreasing values of averaged TRS (SAS) and narrowing ranges of TRS (SIS) when sequential UIs become more lengthy. The above findings have confirmed that, under the scenarios of sequential UIs, the gap between a physical layout and a mental model remains valid.

More importantly, the next implication reflects our key question: ' *Do we need alternative predictive modelling between individual and sequential UIs?*'. As revealed in Section 4.4.4, we do not observe a significantly different breakdown of hand pose choices in varying sequential UIs. That is, the occurrence of hand poses is independent of the UI order (UI position in a sequence of UIs). However, we note that alternative predictive models for sequential UIs will become necessary if the breakdowns are different in certain UI position(s) (i.e., the 'Rank', x-axis in Figure 11). Thus, we can apply the same predictive model to both individual and sequential UIs.

5. Predicting Perceived Pose Reachability – the Predictive Model

The findings in Section 4 depict that the users choose the hand poses and perceived pose reachability differently from the physical context TRS. The user habits, and the inertia among hand poses, in particular, lead to such differences. The detailed investigation of the inertia in sequential UIs of varying lengths demonstrates that the users do not change pose frequently in a lengthy UI sequence. Albeit the inertia exists, our findings also help us to understand that the order of UIs in a sequence does not significantly influence user choices. As such, a predictive model at the level of individual UIs is sufficient to account for the user choices of hand poses and, hence, the perceived pose reachability. The predictive models in this section are solely built on the perceived choices of hand poses, instead of the physical context TRS. It is important to note that the previous section only serves as a piece of evidence for the necessity of the proposed predictive model. Accordingly, this section outlines the designs of our predictive model, in which component-based feature engineering and XGBoost(T. Chen & Guestrin, 2016) Classifier leverages the collected user responses on the basis of the Rico datasets. Generally, XGBoost is often used because of its exceptional performance on extensive datasets and complicated tasks. XGBoost was chosen because of its nature as an ensemble learning methodology, whereby it amalgamates the forecasts of several feeble learners (decision trees) to generate a potent, resilient, and precise predictive model. The use of this collective approach often surpasses the performance of individual models, resulting in a decrease in overfitting and an enhancement in generalisation. Thus, we develop classification models for predicting a user's pose preference over UIs. We further consider the UI Components (Table 1) as the key features that trigger user behaviours. Thus, our model, supported by XGBoost's decision tree, aims to check the existence of certain UI components that will lead to the user's choice of hand poses.

5.1. Data Description

We employed the data collected in the previous sections, i.e., the 30,242 UIs subsampled from the Rico dataset (Deka, Huang, Franzen, Hibschman, et al., 2017) and our collected participant feedback. As described in Section 3.3.3, we showed the UIs to 17 participants who rated the UIs, and we classified the hand poses into two classes: one-handed (Pose 1) and two-handed (Poses 2 and 3). The distribution of one-handed and two-handed is 65% (19,640 UIs) and 35% (10,602 UIs), respectively. We utilized the dataset for training a binary classifier to predict the preferred hand pose. We first transform the UIs into component-based embedding and then deploy the tree-based classification method. Table 1 depicts the UI components, and the details are available in the next paragraph.

Table 1. Association map of UI Components and Class name, with UI Components as shown in Figure 2 (Appendix): the table lists the top-20 most common UI Components in the Android Mobile UIs, while the 'Other' at the bottom include all the rest of UI Components with lesser occurrences exhaustively in the Android Mobile UIs.

UI Component	Associated Class Name
Button	Button, AppCompatButton, ToggleButton
Card View	CardView, SquareCardView, RippleCardView
Text View	TextView, AppCompatTextView, FontTextView
Tab View	TabLayout#TabView, TabPageIndicator#TabView
Web View	WebView, AdWebView, HtmlBannerWebView
Grid View	GridView, HeaderGridView, DynamicGridView
Toolbar	Toolbar, ToolbarButton
Spinner	AppCompatSpinner, Spinner, CustomSpinner
Slider	Slider, SliderButton, SliderAdView
Menu Item View	ActionMenuItemView, NavigationMenuItemView
List View	ListView, ExpandableListView, ObservableListView
Image View	ImageView, AppCompatImageView, NetworkImageView
Image Button	ImageButton, AppCompatImageButton
View Group	ViewGroup, ViewGroupWithContextInfo
Checkbox	AppCompatCheckBox, CheckBox, QCCheckBox
Radio Button	RadioButton
Edit Text	EditText, AppCompatEditText, TextInputEditText
Linear Layout	LinearLayout, SquareLinearLyout, ForgroundLinearLayout
Relative Layout	RelativeLayout, PercentRelativeLayout, SquareRelativeLayout
Frame Layout	FrameLayout, SquarbleFrameLaout
Other	e.g., SwitchCompat, and other components not listed above.

5.2. Feature Extraction

For the feature extraction part, we first identify a set of UI component types. The criteria of component type selection are based on the occurrence of the component in the Rico dataset (Deka, Huang, Franzen, Hibschman, et al., 2017) that consists of more than 72k UI screens and metadata. The main idea of feature extraction is associating the occurrence of UI components with the hand pose choices through a series of examinations in the decision trees of the XGBoost model (Figure 1, Appendix). Thus, the metadata describes UI components in a layout, and further includes the box coordinate and component class name. As the original metadata has a huge number of classes, we clustered the class names into a smaller number of classes. As a result, a total of 21 UI component types are defined, including 1) BUTTON, 2) CARD VIEW, 3) MENU ITEM VIEW, 4) TAB VIEW, 5) WEB VIEW, 6) GRID VIEW, 7) TOOLBAR, 8) SPINNER, 9) SLIDER, 10) LIST VIEW, 11) TEXT VIEW, 12) IMAGE VIEW, 13) IMAGE BUTTON, 14) VIEW GROUP, 15) CHECKBOX, 16) RADIO BUTTON, 17) EDITTEXT, 18) LINEAR LAYOUT, 19) RELATIVE LAYOUT, 20) FRAME LAYOUT and 21) OTHER. Table 1 matches the UI components with the associated class names. Figure 2 in the Appendix also shows examples of the 21 types of UI components. When certain associate class names exist (available in the Rico dataset), the XGBoost model can record the occurrence of UI components in a mobile UI. Then, each UI and metadata were transformed into a one-dimensional vector with a size of 21 that represents the occurrence of the aforementioned component types. Each scalar value represents the occurrence count of the component in the UI. As such,

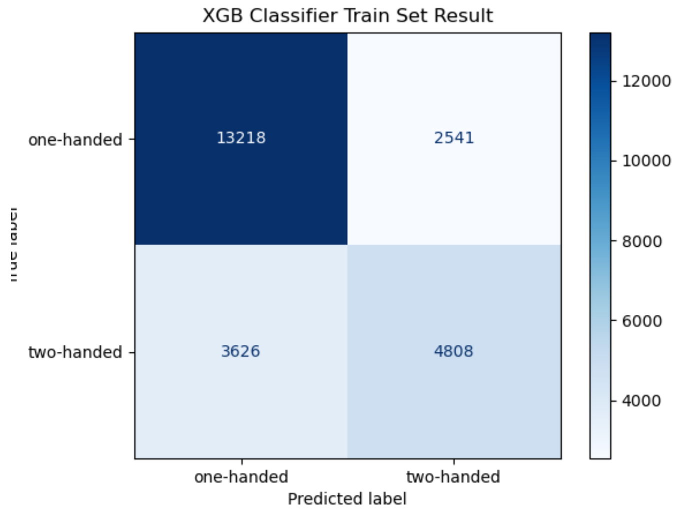


Figure 12. The confusion matrix during the training phase under four metrics: During the *training* phase: the values of precision, recall and f1-score for one-handed interaction are 0.78, 0.84, and 0.81, while the values of precision, recall and f1-score for two-handed interaction are 0.65, 0.57, 0.61. These result in an overall accuracy of 0.75 and an F1-score of 0.71.

we regard the UI components as the features. The occurrence count of the component in the UI, through the model supported by decision trees, will lead to the classification of the hand poses.

5.3. Classification Model

We utilize the XGBoost classifier (T. Chen & Guestrin, 2016) to address the classification problem of hand pose mapping to UIs. XGBoost (XGB) is a decision tree-based ensemble machine learning algorithm. The XGB classifier owns two classes: one-handed interaction (hand pose 1) and two-handed interaction (Aggregation of hand poses 2 and 3), where the hand poses 1 – 3 originated from the user evaluation in Section 3. We first transformed the UIs and metadata into numerical embeddings according to the aforementioned feature engineering method. Then, we separated the dataset into 80% for training and 20% for testing. Since the dataset is imbalanced, its distribution is around 65% one-handed (Pose 1) and 35% (Pose 2). Therefore, we chose to use the random oversampling method (Chawla, Bowyer, Hall, & Kegelmeyer, 2002) to avoid biased situations. After cleaning and preparing the dataset, we separated the training dataset into ten portions and conducted k-fold cross validation to validate the generality. We observed that the model performance between different folds performed similarly during the procedures (Figures 12 and 13). This implies that the model does not over-fitting to any specific portion of our dataset. Thus, we proceed to the model evaluation with the testing dataset.

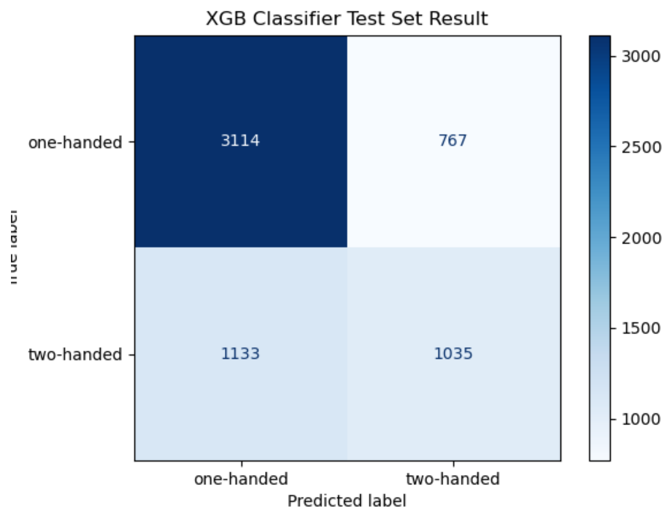


Figure 13. The confusion matrix during the testing phase under four metrics: During the *testing* phase: the values of precision, recall and f1-score for one-handed interaction are 0.73, 0.80, and 0.77, while the values of precision, recall and f1-score for two-handed interaction are 0.57, 0.48, 0.52. These result in an overall accuracy of 0.69 and an F1-score of 0.64.

$$\begin{aligned} Accuracy &= \frac{TP+TN}{TP+TN+FP+FN} \\ Precision &= \frac{TP}{TP+FP} \\ Recall &= \frac{TP}{TP+FN} \\ F1-score = \frac{2*Precision*Recall}{Precision+Recall} &= \frac{2*TP}{2*TP+FP+FN} \end{aligned} \tag{6}$$

Evaluation metrics of the UI Hand Pose Prediction Problem

5.4. Prediction Accuracy

We regard the prediction of hand poses in UIs as a classification problem. We consider four metrics: Accuracy, Precision, Recall, and F1-score, as shown in Eq. 6, to evaluate the performance of the proposed model. The XGboost classifier with training dataset reaches 75% accuracy and 71% f1-score, while the classifier with testing dataset achieves 69% and 64% f1-score. Due to the skewed distribution of the collected hand poses, the current model works better with predicting the one-handed interaction (80%) than the two-handed interaction (48%). The full details of the four metrics for training and testing phases are available in Figures 12 and 13, respectively. Furthermore, we take a closer examination of the decision trees in the XGB models. But its decision process is highly complicated (Figure 1 in Appendix), and hence, we employ an alternative – logistic regression, to assist the explanations.

Figure 14 visualizes the coefficients of UI components to the prediction of hand poses. The positive values of coefficients demonstrate the positive correlation to the preference for two-handed interaction, while the zero and negative values of coefficients imply the preference for one-handed interaction. In general, 17 out of 20 types of UI components can induce two-handed interaction. We find that two classes of UI components, namely Slider and Edit text, demonstrate significantly larger coefficients than the rest of the UI components. Whenever the Slider or the Edit text occurs in a UI, the model will generate a high likelihood of two-handed interaction. Slider requires continuous touches (i.e., rubbing) on the touchscreen, and one-handed interaction makes difficult thumb movements, while Edit text calls the soft keyboard and hence triggers two-handed interaction. Spinner, Gridview, Webview and Menuview, usually trigger clickable items widely distributed in a UI, which turn two-handed interactions into a more accessible choice. For instance, Spinner will trigger exploded views of clickable items vertically. Apart from these two components, we find that other components, such as Spinner, Grid view, and Menu item, also significantly influence the classification of perceived user reachability into two-handed interactions.

Additionally, once Toolbar and Viewgroup are spotted in a UI, one-handed interactions are likely to appear in predictive results. Intuitively, under the current UI/UX practice, Toolbars are usually located at the upper screen position, while Viewgroup is a bulky item spreading across a significant area in the touchscreen. And the physical-context measurements would define such interactive UI components as unreachable and hence mistakenly detriment to the reachability score (e.g., TRS). Nevertheless, such UI components do not turn the perceived user reachability into two-handed interaction, for the following key reasons. These UI components enable users to perform scrollable interaction on touchscreens (i.e., swipe actions) on the collection of information. As such, a user with a thumb manages to explore the exploded items in the collection of information. Conversely, the scrolling process will be interrupted if another thumb is also touching the screen, not to mention the occluded views on such large-size UI components. On the other hand, Toolbars, usually located at the upper-right or -left corners of a UI (e.g., a navigation bar), are usually regarded as some auxiliary operations or optional functions. Therefore, users would consider the minimal necessity of reaching such top positions, and give perceived hand poses of one-handed interaction that are sufficient to interact with the rest of the UI components.

5.5. An illustration of the XGB-based Auditing of UI Reachability

Mobile applications are no doubt deployed in various scenarios, for instance, holding a handrail inside a train or carrying a shopping bag in urban areas. Considering such in-the-wild scenarios, users prefer to interact with smartphone UIs with one hand. The UI designer should consider the aforementioned use case as the mobility constraint in the UI design, i.e., Design for Mobility. The proposed XGB model can reveal that the user will interact with a UI via either one-handed or two-handed interaction. Our predictive model would be beneficial to UI designers by revealing UIs that suffer from mobility due to two-handed interactions. The predictive results can serve as a design clue to re-arrange the UI components in a layout. The UI designer, with such design clues, can improve user mobility accordingly.

Figures 15 and 16 illustrate prediction results of perceived reachability in individual UIs. It is important to note that the green boxes pinpoint the UI components in the example UIs. In the first row, *True Label* indicated the user preference being collected in the 6-week user evaluation. Remarkably, the *True Labels* in Figures 15 and 16 are 1 and 2, representing one-handed and two-handed interactions chosen by

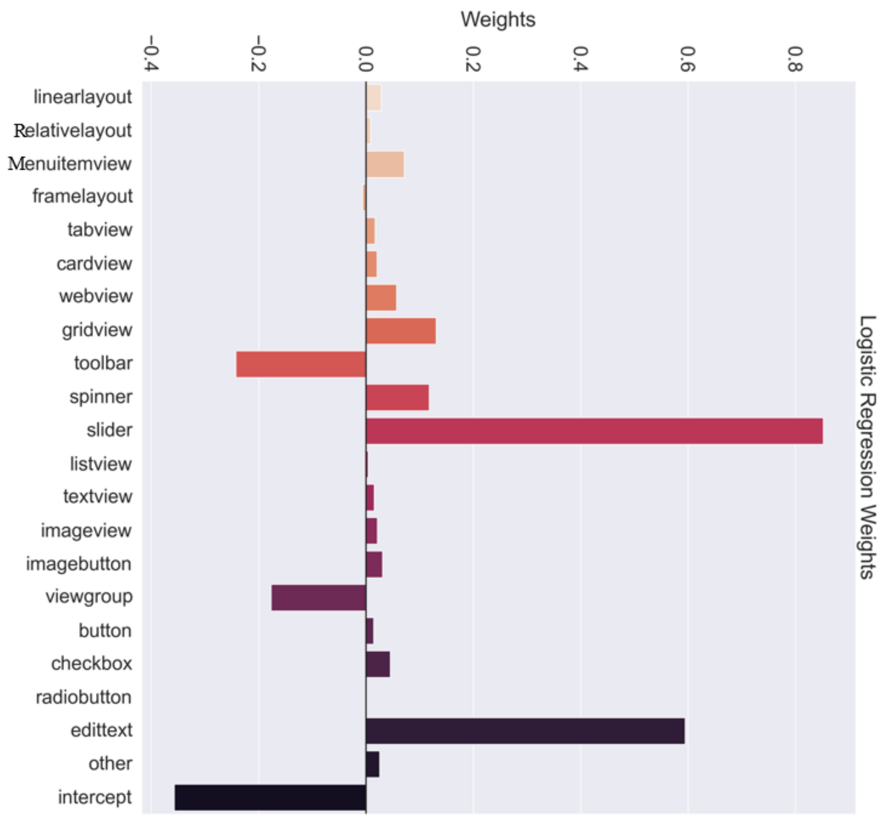
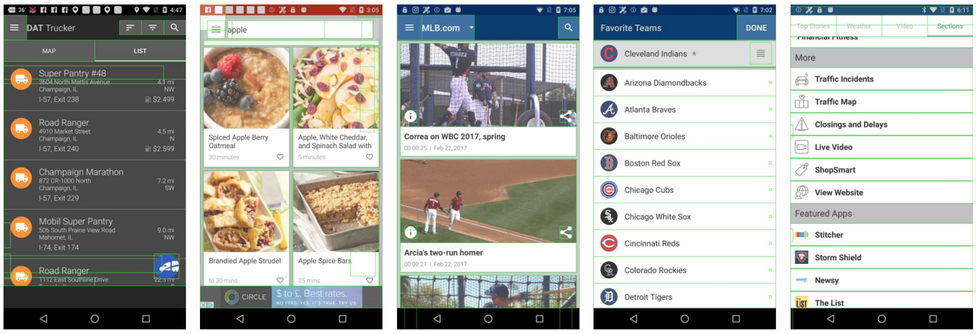


Figure 14. The Visualization of the Weighting of UI components by Logistic Regression. Starting from the top to bottom, they are Linear Layout, Relative Layout, Menu Item View, Frame Layout, Tab View, Card View, Web View, Grid View, Toolbar, Spinner, Slider, List View, Text View, Image View, Image Button, View Group, Button Checkbox, Radio Button, Edit Text and Other(s).

the participants. The second and third rows depict the probability of two classes for either one-handed or two-handed interaction. The bottom row illustrates the predicted results, in which the numbers [1 or 2] in the colours of green or red refer to either correct or incorrect predictions – 'does the prediction match with the *True Label*?'.

Generally speaking, most illustrated UI examples contains Toolbar and Viewgroup, but the predictive model generates different classification. UIs with *True Label of 1* (Figures 15), at first glance, present multiple items physically beyond the thumb reach (on the basis of *Parabolic Motion Trajectory* (Bergstrom-Lehtovirta & Oulasvirta, 2014)). However, users precept the remaining UI components in such UIs as scrollable items and hence maintain one-handed yet one-thumb interactions. In contrast, UIs with a *True Label of 2* contain interactive components that require precise taps to select such components beyond the thumb's reach. As shown in Figure 16, the rightmost three UIs have icons at upper-left positions, while the leftmost three UIs contain numerous smaller items, triggering two-handed operations for the tap-and-select. In the leftmost UI, miniature UI components are at the upper-right corner to navigate to the next article (i.e., the arrow), while the next UI (the second) contains arrays of UI components on the right-hand side.



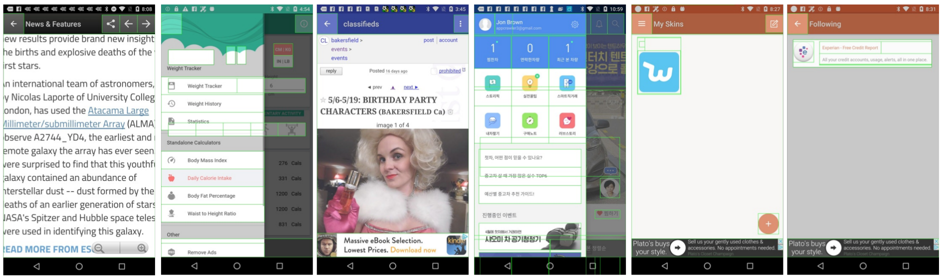
True Label	1	1	1	1	1
Prob.(1) [1-hand]	.55	.42	.63	.91	.56
Prob.(2) [2-hand]	.45	.58	.37	.09	.44
Cognitive Choices	1	2	1	1	1

Figure 15. Five UI examples (with *True Label* = 1, one-handed interaction) were audited by the proposed XGB model.

6. Discussion

By examining the perceived thumb reachability data, we detail the relationship between TRS and the user's perceived choices on individual UIs for one-handed interactions. TRS, albeit its rationales behind thumb lengths and grip sizes, achieves binary classifications of *Parabolic Motion Trajectory* (Bergstrom-Lehtovirta & Oulasvirta, 2014), overlooks the user's habit or pose preferences on individual UIs and neglects the inertia being triggered by sequential UIs. We do not intend to disprove the prior models supported by such physical-context approaches. Instead, we introduce an alternative assessment of UI reachability motivated by the above cognitive means. The predictive model of hand poses attempts to evaluate UI reachability before the physical actors of hands and thumbs work on the UIs. Supported by our analytics in Section 4, the models highlight the importance of the user perception as well as the perceived ease of interaction with most UI components in individual UIs, i.e., considering UI components in a layout as user affordance (Gaver, 1991; Petersen, Rasmussen, & Trettvik, 2020). The models also broaden our knowledge of UI reachability at the cognitive phase (i.e., reveal aspects possibly relevant to mental models), and works complementary with the existing physical-context assessments (Girouard et al., 2015; Le et al., 2018; Pecchioli et al., 2019; Prange et al., 2018). Combining cognitive and physical evaluation could offer us a more comprehensive understanding of user-centered UI designs for one-handed interactions.

The benefit of machine learning-based modelling of human perceived choices on UI reachability is multi-fold. First, through learning from the crowdsourced UIs (Deka, Huang, Franzen, Hibschman, et al., 2017) and our user response data, the models naturally capture the priors that humans have made perceived choices on UIs on the basis of UI components in layouts. In terms of user interaction with UIs, the captured priors allow us to distinguish the convenient designs that allow one-handed interactions on smartphones. The extracted features of UI components on the UIs could inform



True Label	2	2	2	2	2	2
Prob.(1) [1-hand]	.42	.20	.48	.14	.43	.52
Prob.(2) [2-hand]	.58	.80	.52	.86	.57	.48
Cognitive Choices	2	2	2	2	2	1

Figure 16. Six UI examples (with *True Label* = 2, two-handed interaction) were audited by the proposed XGB model.

the interaction designers to understand how the users think about their designs of new applications, i.e., the cognitive means of a UI. It is important to note that XGB models, with a strong capability of extracting the features of UI components, have been extensively employed in the banking and finance industrial sectors. Our works consider the user perceptions of UIs as an entire interaction task and the corresponding UI components, in which the decision trees in XGB can parse the likelihood of the hand pose happening in UIs. Second, when the UI becomes more complicated, e.g., numerous UI components at irregular positions, the abstract user perception among multiple clickable items (or buttons) in a UI cannot be simply represented by a single metric or multi-criteria heuristics. Third, the prior works do not support modelling abstract perceived choices for user interactions with mobile UIs. In contrast, the promising capacity of data analytics and machine learning enables us to easily leverage intrinsic clues between the user's choices and the spatial relations among UI components in a layout.

It is well-recognized that the static equations and mathematical programmings (Bi et al., 2013a; Wobbrock et al., 2008) provide the outstanding property of interpretability (Reinecke & Gajos, 2014). Machine learning models, in general, lack the above property in spite of their superior ability of feature engineering (Yuan & Li, 2020), e.g., user perceptions of UIs. In this paper, we analyzed the user perception of the UI layout on an individual level. We build our models according to a grounded theory leveraging such generalized user perceptions (e.g., SSD for individual UIs and FIS for sequential UIs). Thus, rather than building deep learning models in an opportunistic manner, our machine learning approach, on the basis of UIs big data analytics, can be viewed as a justified tool for predicting hand pose choices to discover UI design paradigms of one-handed interaction, and to advance UI reachability research. It is expected that, in real-life scenarios, UI Designers, with our findings and proposed models, can understand the user perceptions of UI components inside smartphone layouts. Accordingly, they can build UIs that encourage one-handed interaction at the cognitive phase. For instance, UI designers can do several trials regarding the arrangement of UI components on a smartphone. More specifically, our model, without recruiting real users, can assess the UIs to see if users find one-handed or two-handed interaction convenient, as the model can give early indications about the user's pose choices. In other words, UI designers, during the design phases of smartphone UIs, can perform a preliminary evaluation with the benefits of saving time and reducing financial burdens caused by recruiting human participants and, hence, user studies. Next, the proposed tool of our work can be further combined with other existing tools, e.g., TRS (Bergstrom-Lehtovirta & Oulasvirta, 2014). UI designers can leverage ergonomic tools or other traditional evaluation methods simultaneously, which can better guarantee the alignment of one-handed interactions on smartphone layouts in both cognitive and action phases. A practical and illustrative example is the implementation of single-handed interactions across a range of frequently used interfaces for users in the workforce engaged in daily commuting. In the public transportation system, users need one hand to carry items or hold a handrail. Thus, the UI design with our proposed model owns the competitive advantage by allowing better user experiences.

Several limitations exist in this work. First of all, we notice that the current dataset only reflects three types of hand poses, while smartphone users may have more diversified hand poses, e.g., two-handed interaction, involving multiple fingers on one hand and another hand as the support. More importantly, our dataset is collected by judging the ranking of UIs purely on sight and interpretation, leading to decoupling the ranking of UIs from a physical context. The combined models of the physical context of TRS and user perception will become an interesting direction.

Despite our analysis successfully establishing the cues between UI components in layouts and users' perceived reachability, our model results have limited explainability of the predictive results. It is important to communicate the design issue at the level of UI-component pixels, and pinpoint the design suggestions to the practitioners. We also acknowledge that such explainability is essential to building trust between the predictive model and the designers. In the first stage, the explainability can be done by offering multiple new samples that cover various classes of inputs and visualising how the variations of input parameters lead to a different classification.

In addition, our study only focuses on the quantitative results from the data analytics and the proposed predictive model, while the qualitative feedback (trust, usability, mental loading, etc.) from the designers has not yet been investigated. But still, the machine learning model of XGB offers reasonable accuracy of the user pose choices for UIs, serving as a valuable reference of UI reachability towards the cognitive dimension. Additionally, the predictive poses for individual UIs could produce unambiguous yet inexpensive design clues to the practitioners. Opportunities for balancing conflicting choices through extended optimization strategies should be considered.

On the other hand, it is also preferable to scale up the data collection process to cover more participants, further enhancing the generality of our study, at the expense of higher demand for resources. Our current participants are primarily young people, and we have to broaden the participants' age range. Also, our existing dataset encounters a skewed distribution, leading to reduced predictive power in two-handed interaction and thus deteriorating overall prediction accuracy (i.e., F1 score).

Moreover, the original dataset (Deka, Huang, Franzen, Hibschman, et al., 2017) only contains Android UIs. However, no less than 25% of users using iOS smartphones exist in the consumer market. If a dataset can capture UIs from both Android and iOS, the coverage of our study regarding perceived user reachability would be more promising. In other words, a comprehensive study should not neglect this significant population of smartphone users who deal with the iOS interfaces daily, which leads to more inclusive and applicable findings to a broader market. For instance, the interface designs of iOS may induce users to judge their potential actions. Finally, the current assessment primarily relies on the UI components' statistics. Suppose multi-channels of UI features and characteristics are included in our assessment, such as the coordination of UI components and the types of applications and user contexts. In that case, the model performance will result in improved performance as well as more detailed design cues.

7. Conclusion and Future Work

While the physical context of thumb reachability has been extensively studied for one-handed interaction, our work adds to the body of literature by modelling user's perceived choices on mobile UIs. We have made the first attempt at the acquisition of users' perceived reachability. Our data analytics of users' choices on UIs preference confirms that human perception of UI components in layouts should be considered separately from the physical characteristics (in terms of TRS) (Bergstrom-Lehtovirta & Oulasvirta, 2014). We further offer insights into the perceived pose choices on individual UIs and their potential reasons, led by the inertia among sequential UIs. Based on the observations from the analytical analysis, our modelling work provides design cues of perceived choices of hand poses, in terms of understanding how users respond to UIs on smartphones. The reasonable degree of model predictability establishes a significant connection between UI component layouts and perceived reachability, and thus, we have made a novel contribution by uncovering such a relation. The resultant machine learning model can serve as an auditing tool for designers to conveniently evaluate individual UIs and UI sequences without running expensive user testing. Without the user testing that is usually time-consuming, the proposed model can alert the designer regarding certain interfaces triggering inconvenient interactions. The shortened implementation time brings a more agile application development cycle. Thus, our contribution leads to a more cost-effective and responsive design workflow and simultaneously promotes mobile UI designs.

There are several directions for future work. For example, although XGB serves as a use case for the collected user responses, there is still room for improvement in model performance, e.g., feature extraction. We can expand the capability of the models and consider any intrinsic relationships between UI components, for instance, by modelling a graph neural network. In addition, we currently only focus on the prediction of hand poses. We can extend the models to give designers suggestions on reducing pose changes by visualizing the potential issues and considering the weightings of UI components with finer granularity, e.g., at the pixel level.

References

Bachynskyi, M., Palmas, G., Oulasvirta, A., Steimle, J., & Weinkauf, T. (2015). Performance and ergonomics of touch surfaces: A comparative study using biomechanical simulation. In *Proceedings of the 33rd annual acm conference on human factors in computing systems* (p. 1817–1826). New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi-org.lib.ezproxy.ust.hk/10.1145/2702123.2702607`

Bergstrom-Lehtovirta, J., & Oulasvirta, A. (2014). Modeling the functional area of the thumb on mobile touchscreen surfaces. In *Proceedings of the sigchi conference on human factors in computing systems* (p. 1991–2000).

New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi-org.lib.ezproxy.ust.hk/10.1145/2556288.2557354

Bi, X., Li, Y., & Zhai, S. (2013a). Ffitts law: Modeling finger touch with fitts' law. In *Proceedings of the sigchi conference on human factors in computing systems* (p. 1363–1372). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2470654.2466180

Bi, X., Li, Y., & Zhai, S. (2013b). Ffitts law: Modeling finger touch with fitts' law. In *Proceedings of the sigchi conference on human factors in computing systems* (p. 1363–1372). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi-org.lib.ezproxy.ust.hk/10.1145/2470654.2466180

Borah, P. P., & Sorathia, K. (2019). Natural and intuitive deformation gestures for one-handed landscape mode interaction. In *Proceedings of the thirteenth international conference on tangible, embedded, and embodied interaction* (p. 229–236). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi-org.lib.ezproxy.ust.hk/10.1145/3294109.3300996

Boring, S., Ledo, D., Chen, X. Marquardt, N., Tang, A., & Greenberg, S. (2012). The fat thumb: Using the thumb's contact size for single-handed mobile interaction. In *Proceedings of the 14th international conference on human-computer interaction with mobile devices and services* (p. 39–48). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi-org.lib.ezproxy.ust.hk/10.1145/2371574.2371582

Buschek, D., Auch, A., & Alt, F. (2015). A toolkit for analysis and prediction of touch targeting behaviour on mobile websites. In *Proceedings of the 7th acm sigchi symposium on engineering interactive computing systems* (p. 54–63). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2774225.2774851

Card, S. K., Newell, A., & Moran, T. P. (1983). *The psychology of human-computer interaction.* USA: L. Erlbaum Associates Inc.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, *16*, 321–357.

Chen, C.-F. R., Pistoia, M., Shi, C., Girolami, P., Ligman, J. W., & Wang, Y. (2017). Ui x-ray: Interactive mobile ui testing based on computer vision. In *Proceedings of the 22nd international conference on intelligent user interfaces* (p. 245–255). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3025171.3025190

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (p. 785–794). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi-org.lib.ezproxy.ust.hk/10.1145/2939672.2939785

Chen, Y., Pandey, M., Song, J. Y., Lasecki, W. S., & Oney, S. (2020). Improving crowd-supported gui testing with structural guidance. In *Proceedings of the 2020 chi conference on human factors in computing systems* (p. 1–13). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi-org.lib.ezproxy.ust.hk/10.1145/3313831.3376835

Cockburn, A., Gutwin, C., & Greenberg, S. (2007). A predictive model of menu performance. In *Proceedings of the sigchi conference on human factors in computing systems* (p. 627–636). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/1240624.1240723

Corsten, C., Daehlmann, B., Voelker, S., & Borchers, J. (2017). Backxpress: Using back-of-device finger pressure to augment touchscreen input on smartphones. In *Proceedings of the 2017 chi conference on human factors in computing systems* (p. 4654–4666). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi-org.lib.ezproxy.ust.hk/10.1145/3025453.3025565

Corsten, C., Lahaye, M., Borchers, J., & Voelker, S. (2019). Forceray: Extending thumb reach via force input stabilizes device grip for mobile touch input. In *Proceedings of the 2019 chi conference on human factors in computing systems*

(p. 1–12). New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi-org.lib.ezproxy.ust.hk/10.1145/3290605.3300442`

Deka, B., Huang, Z., Franzen, C., Hibschman, J., Afergan, D., Li, Y., ... Kumar, R. (2017). Rico: A mobile app dataset for building data-driven design applications. In *Proceedings of the 30th annual acm symposium on user interface software and technology* (p. 845–854). New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi-org.lib.ezproxy.ust.hk/10.1145/3126594.3126651`

Deka, B., Huang, Z., Franzen, C., Nichols, J., Li, Y., & Kumar, R. (2017). Zipt: Zero-integration performance testing of mobile app designs. In *Proceedings of the 30th annual acm symposium on user interface software and technology* (p. 727–736). New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi-org.lib.ezproxy.ust.hk/10.1145/3126594.3126647`

Eardley, R., Roudaut, A., Gill, S., & Thompson, S. J. (2017). Understanding grip shifts: How form factors impact hand movements on mobile phones. In *Proceedings of the 2017 chi conference on human factors in computing systems* (p. 4680–4691). New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi-org.lib.ezproxy.ust.hk/10.1145/3025453.3025835`

Eardley, R., Roudaut, A., Gill, S., & Thompson, S. J. (2018a). Designing for multiple hand grips and body postures within the ux of a moving smartphone. In *Proceedings of the 2018 designing interactive systems conference* (p. 611–621). New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi-org.lib.ezproxy.ust.hk/10.1145/3196709.3196711`

Eardley, R., Roudaut, A., Gill, S., & Thompson, S. J. (2018b). Investigating how smartphone movement is affected by body posture. In *Proceedings of the 2018 chi conference on human factors in computing systems* (p. 1–8). New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi-org.lib.ezproxy.ust.hk/10.1145/3173574.3173776`

Gaver, W. W. (1991). Technology affordances. In *Proceedings of the sigchi conference on human factors in computing systems* (p. 79–84). New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi.org/10.1145/108844.108856`

Girouard, A., Lo, J., Riyadh, M., Daliri, F., Eady, A. K., & Pasquero, J. (2015). One-handed bend interactions with deformable smartphones. In *Proceedings of the 33rd annual acm conference on human factors in computing systems* (p. 1509–1518). New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi-org.lib.ezproxy.ust.hk/10.1145/2702123.2702513`

Hakka, K., Isomoto, T., & Shizuki, B. (2019). One-handed interaction technique for single-touch gesture input on large smartphones. In *Symposium on spatial user interaction.* New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi-org.lib.ezproxy.ust.hk/10.1145/3357251.3358750`

Hasan, K., Kim, J., Ahlström, D., & Irani, P. (2016). Thumbs-up: 3d spatial thumb-reachable space for one-handed thumb interaction on smartphones. In *Proceedings of the 2016 symposium on spatial user interaction* (p. 103–106). New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi-org.lib.ezproxy.ust.hk/10.1145/2983310.2985755`

Hoober, S., & Berkman, E. (2011). Designing mobile interfaces - patterns for interaction design..

Huang, L., Chen, H., Wang, X., & Chen, G. (2000). A fast algorithm for mining association rules. *Journal of Computer Science and Technology*, *15*(6), 619–624. Retrieved from `https://doi.org/10.1007/BF02948845`

Huhtala, J., Sarjanoja, A.-H., Mäntyjärvi, J., Isomursu, M., & Häkkilä, J. (2010). Animated ui transitions and perception of time: A user study on animated effects on a mobile screen. In *Proceedings of the sigchi conference on human factors in computing systems* (p. 1339–1342). New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi.org/10.1145/1753326.1753527`

Kim, S., & Lee, S. (2018). Three levels of perceiving affordance: Possibility, specificity, and signification levels. In *Proceedings of the 30th australian conference on computer-human interaction* (p. 261–265). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3292147.3292219

Kim, S., Yu, J., & Lee, G. (2012). Interaction techniques for unreachable objects on the touchscreen. In *Proceedings of the 24th australian computer-human interaction conference* (p. 295–298). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi-org.lib.ezproxy.ust.hk/10.1145/2414536.2414585

Kirhenstein, A. (2019, Oct). *Mobile reachability. rules of thumb.* Medium. Retrieved from https://medium.com/@Draward/mobile-reachability-rules-of-thumb-ce37dd0cd3ad

Ko, P.-H., Hwang, Y.-H., & Liang, H.-W. (2016). Influence of smartphone use styles on typing performance and biomechanical exposure. *Ergonomics*, *59*(6), 821–828.

Lai, J., & Zhang, D. (2015). Extendedthumb: A target acquisition approach for one-handed interaction with touch-screen mobile phones. *IEEE Transactions on Human-Machine Systems*, *45*(3), 362-370.

Larsen, J. N., Jacobsen, T. H., Boring, S., Bergström, J., & Pohl, H. (2019). The influence of hand size on touch accuracy. In *Proceedings of the 21st international conference on human-computer interaction with mobile devices and services.* New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi-org.lib.ezproxy.ust.hk/10.1145/3338286.3340115

Le, H. V., Bader, P., Kosch, T., & Henze, N. (2016). Investigating screen shifting techniques to improve one-handed smartphone usage. In *Proceedings of the 9th nordic conference on human-computer interaction.* New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi-org.lib.ezproxy.ust.hk/10.1145/2971485.2971562

Le, H. V., Mayer, S., Bader, P., & Henze, N. (2018). Fingers' range and comfortable area for one-handed smartphone interaction beyond the touchscreen. In *Proceedings of the 2018 chi conference on human factors in computing systems* (p. 1–12). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi-org.lib.ezproxy.ust.hk/10.1145/3173574.3173605

Lee, L. H., Lam, K. Y., Li, T., Braud, T., Su, X., & Hui, P. (2019, September). Quadmetric optimized thumb-to-finger interaction for force assisted one-handed text entry on mobile headsets. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, *3*(3). Retrieved from https://doi.org/10.1145/3351252

Lee, L. H., Zhu, Y., Yau, Y. P., Braud, T., Su, X., & Hui, P. (2020, mar). One-thumb text acquisition on force-assisted miniature interfaces for mobile headsets. In *2020 ieee international conference on pervasive computing and communications (percom)* (p. 1-10). Los Alamitos, CA, USA: IEEE Computer Society. Retrieved from https://doi.ieeecomputersociety.org/10.1109/PerCom45495.2020.9127378

Lehmann, F., & Kipp, M. (2018). How to hold your phone when tapping: A comparative study of performance, precision, and errors. In *Proceedings of the 2018 acm international conference on interactive surfaces and spaces* (p. 115–127). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi-org.lib.ezproxy.ust.hk/10.1145/3279778.3279791

Li, Y., Bengio, S., & Bailly, G. (2018). Predicting human performance in vertical menu selection using deep learning. In *Proceedings of the 2018 chi conference on human factors in computing systems* (p. 1–7). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3173574.3173603

Li, Y., Kumar, R., Lasecki, W. S., & Hilliges, O. (2020). Artificial intelligence for hci: A modern approach. In *Extended abstracts of the 2020 chi conference on human factors in computing systems* (p. 1–8). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi-org.lib.ezproxy.ust.hk/10.1145/3334480.3375147

Miniukovich, A., & De Angeli, A. (2015). Computation of interface aesthetics. In *Proceedings of the 33rd annual acm conference on human factors in computing systems*

(p. 1163–1172). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2702123.2702575

Negulescu, M., & McGrenere, J. (2015). Grip change as an information side channel for mobile touch interaction. In *Proceedings of the 33rd annual acm conference on human factors in computing systems* (p. 1519–1522). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi-org.lib.ezproxy.ust.hk/10.1145/2702123.2702185

Oulasvirta, A., De Pascale, S., Koch, J., Langerak, T., Jokinen, J., Todi, K., ... Weinkauf, T. (2018). Aalto interface metrics (aim): A service and codebase for computational gui evaluation. In *The 31st annual acm symposium on user interface software and technology adjunct proceedings* (p. 16–19). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3266037.3266087

Pecchioli, M., Dubois, E., Irani, P., & Serrano, M. (2019). Investigating screen reachability on an articulated dual-display smartphone. In D. Lamas, F. Loizides, L. Nacke, H. Petrie, M. Winckler, & P. Zaphiris (Eds.), *Human-computer interaction – interact 2019* (pp. 476–485). Cham: Springer International Publishing.

Petersen, M. G., Rasmussen, M. K., & Trettvik, J. (2020). Affordances of shape-changing interfaces: An information perspective on transformability and movement. In *Proceedings of the 2020 acm designing interactive systems conference* (p. 1959–1971). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3357236.3395521

Picard, R. W. (2000, March). Perceptual user interfaces: Affective perception. *Commun. ACM*, *43*(3), 50–51. Retrieved from https://doi.org/10.1145/330534.330539

Prange, S., Buschek, D., & Alt, F. (2018). An exploratory study on correlations of hand size and mobile touch interactions. In *Proceedings of the 17th international conference on mobile and ubiquitous multimedia* (p. 279–283). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi-org.lib.ezproxy.ust.hk/10.1145/3282894.3282924

Quinn, P., Lee, S. C., Barnhart, M., & Zhai, S. (2019). Active edge: Designing squeeze gestures for the google pixel 2. In *Proceedings of the 2019 chi conference on human factors in computing systems* (p. 1–13). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi-org.lib.ezproxy.ust.hk/10.1145/3290605.3300504

Reinecke, K., & Gajos, K. Z. (2014). Quantifying visual preferences around the world. In *Proceedings of the sigchi conference on human factors in computing systems* (p. 11–20). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2556288.2557052

Salazar, C. A. F., Henze, N., & Wolf, K. (2016). Ergonomics of thumb-based pointing while holding tablets. In *Proceedings of the 18th international conference on human-computer interaction with mobile devices and services adjunct* (p. 730–737). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi-org.lib.ezproxy.ust.hk/10.1145/2957265.2961864

Sarcar, S., Silpasuwanchai, C., Delamare, W., Ono, A., Oulasvirta, A., & Ren, X. (2019). Exploring performance of thumb input for pointing and dragging tasks on mobile device. In *Proceedings of asian chi symposium 2019: Emerging hci research collection* (p. 38–45). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi-org.lib.ezproxy.ust.hk/10.1145/3309700.3338434

Suzuki, K., Sakamoto, R., Sakamoto, D., & Ono, T. (2018). Pressure-sensitive zooming-out interfaces for one-handed mobile interaction. In *Proceedings of the 20th international conference on human-computer interaction with mobile devices and services.* New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi-org.lib.ezproxy.ust.hk/10.1145/3229434.3229446

Swearngin, A., & Li, Y. (2019a). Modeling mobile interface tappability using crowdsourcing and deep learning. In *Proceedings of the 2019 chi conference on human factors in computing systems* (p. 1–11). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi-org.lib.ezproxy.ust.hk/10.1145/3290605.3300305

Swearngin, A., & Li, Y. (2019b). Modeling mobile interface tappability using crowdsourcing and deep learning. In *Proceedings of the 2019 chi conference on human factors in computing systems* (p. 1–11). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3290605.3300305

Toney, A. P., & Thomas, B. H. (2007). Modeling reach for use in user interface design. In *Proceedings of the eight australasian conference on user interface - volume 64* (p. 27–30). AUS: Australian Computer Society, Inc.

Tsai, H.-R., Huang, D.-Y., Hsieh, C.-H., Huang, L.-T., & Hung, Y.-P. (2016). Movingscreen: Selecting hard-to-reach targets with automatic comfort zone calibration on mobile devices. In *Proceedings of the 18th international conference on human-computer interaction with mobile devices and services adjunct* (p. 651–658). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi-org.lib.ezproxy.ust.hk/10.1145/2957265.2961835

Voelker, S., Hueber, S., Corsten, C., & Remy, C. (2020). Headreach: Using head tracking to increase reachability on mobile touch devices. In *Proceedings of the 2020 chi conference on human factors in computing systems* (p. 1–12). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi-org.lib.ezproxy.ust.hk/10.1145/3313831.3376868

Wobbrock, J. O., Cutrell, E., Harada, S., & MacKenzie, I. S. (2008). An error model for pointing based on fitts' law. In *Proceedings of the sigchi conference on human factors in computing systems* (p. 1613–1622). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi-org.lib.ezproxy.ust.hk/10.1145/1357054.1357306

Wu, Z., Jiang, Y., Liu, Y., & Ma, X. (2020). Predicting and diagnosing user engagement with mobile ui animation via a data-driven approach. In *Proceedings of the 2020 chi conference on human factors in computing systems* (p. 1–13). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3313831.3376324

Wu, Z., Kim, T., Li, Q., & Ma, X. (2019). Understanding and modeling user-perceived brand personality from mobile application uis. In *Proceedings of the 2019 chi conference on human factors in computing systems* (p. 1–12). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3290605.3300443

Xu, A., Huang, S.-W., & Bailey, B. (2014). Voyant: Generating structured feedback on visual designs using a crowd of non-experts. In *Proceedings of the 17th acm conference on computer supported cooperative work social computing* (p. 1433–1444). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2531602.2531604

Yoo, H., Yoon, J., & Ji, H. (2015). Index finger zone: Study on touchable area expandability using thumb and index finger. In *Proceedings of the 17th international conference on human-computer interaction with mobile devices and services adjunct* (p. 803–810). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi-org.lib.ezproxy.ust.hk/10.1145/2786567.2793704

Yu, N.-H., Huang, D.-Y., Hsu, J.-J., & Hung, Y.-P. (2013). Rapid selection of hard-to-access targets by thumb on mobile touch-screens. In *Proceedings of the 15th international conference on human-computer interaction with mobile devices and services* (p. 400–403). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi-org.lib.ezproxy.ust.hk/10.1145/2493190.2493202

Yuan, A., & Li, Y. (2020). Modeling human visual search performance on realistic webpages using analytical and deep learning methods. In *Proceedings of the 2020 chi conference on human factors in computing systems* (p. 1–12). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi-org.lib.ezproxy.ust.hk/10.1145/3313831.3376870

8. Appendices

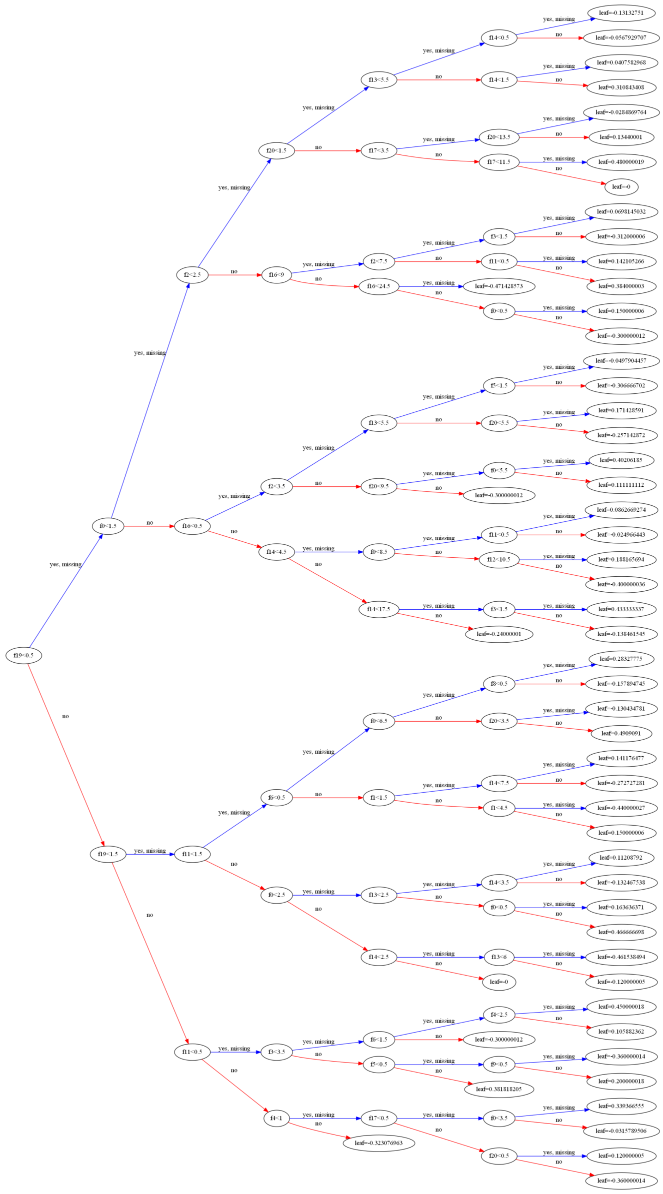


Figure 1. The decision trees in the proposed XGB model: The numbers ranged from F1 – F21 follow the UI components in Table 1.

Appendix A. Author Biographies

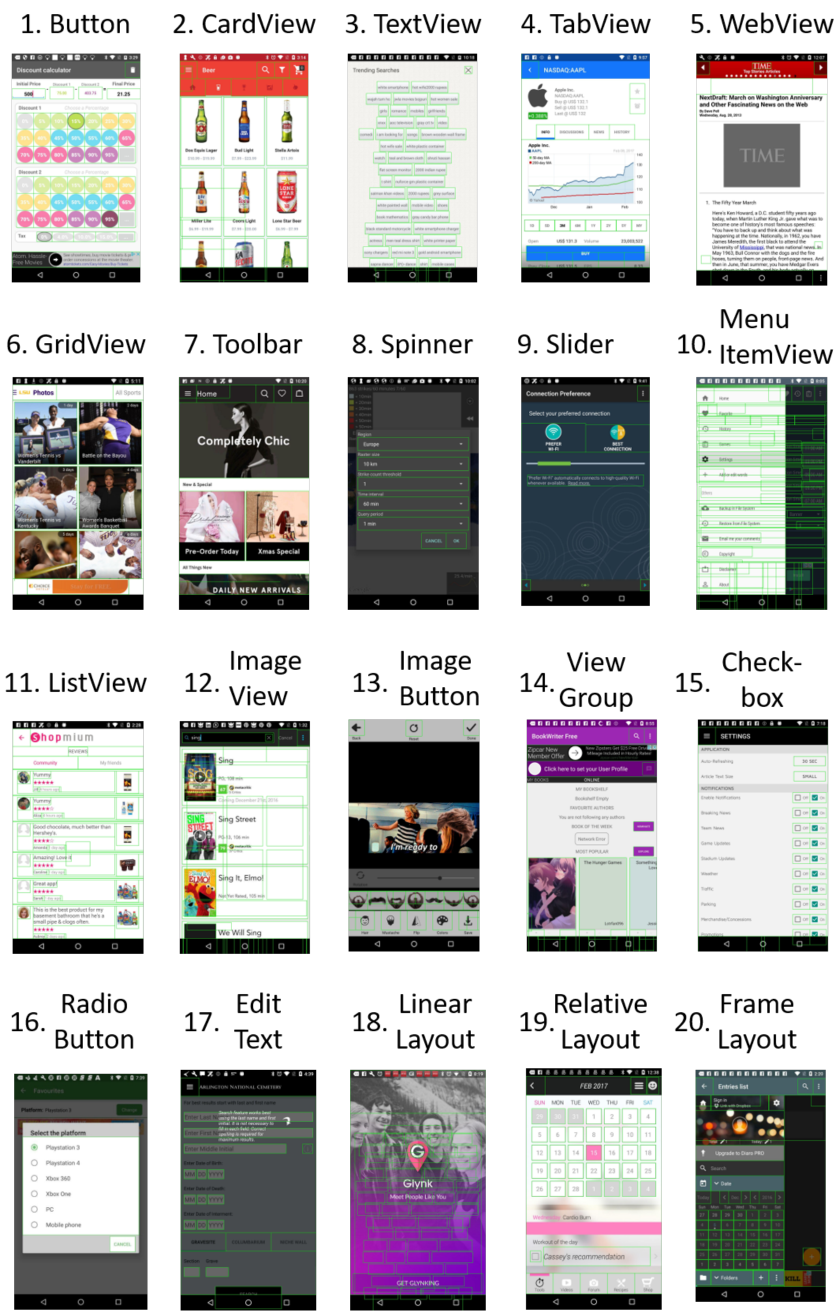


Figure 2. Illustrative examples for the 20 types of UI components (1 – 20, referring F1-F20 in Figure 1), while the type 21 (other) is not displayed as it is a collection of various UIs not included in the types 1 – 20.

Figure A1. Lik-Hang Lee is an assistant professor at the Hong Kong Polytechnic University. Prior to his current position, he was an assistant professor (tenure-track) with the Korea Advanced Institute of Science and Technology (KAIST), South Korea. His primary research interests are Augmented Reality (AR), Virtual Reality (VR), and Education Technologies.

Figure A2. Yui-Pan Yau completed his M.Phil. degree at the University of Science and Technology. His research interests are in designing thumb-driven user interfaces and human-drone interaction. His works appear at several well-recognized venues, including IEEE PERCOM, ACM IMWUT, IEEE TiTS, ACM WebConf, IEEE/RSJ IROS, and Proceedings of the ACM on HCI.

Figure A3. Pan Hui is a Chair Professor of Computational Media and Arts thrust at the Hong Kong University of Science and Technology (Guangzhou), a Chair Professor of the Academy of Interdisciplinary Studies, and a Director of the HKUST-DT Systems and Media Lab, Hong Kong University of Science and Technology.