Recognizing Unintentional Touch on Interactive Tabletop

XUHAI XU, Tsinghua University and University of Washington
CHUN YU*, YUNTAO WANG, and YUANCHUN SHI, Tsinghua University

A multi-touch interactive tabletop is designed to embody the benefits of a digital computer within the familiar surface of a physical tabletop. However, the nature of current multi-touch tabletops to detect and react to all forms of touch, including unintentional touches, impedes users from acting naturally on them. In our research, we leverage gaze direction, head orientation and screen contact data to identify and filter out unintentional touches, so that users can take full advantage of the physical properties of an interactive tabletop, *e.g.*, resting hands or leaning on the tabletop during the interaction. To achieve this, we first conducted a user study to identify behavioral pattern differences (gaze, head and touch) between completing usual tasks on digital versus physical tabletops. We then compiled our findings into five types of spatiotemporal features, and train a machine learning model to recognize unintentional touches with an F1 score of 91.3%, outperforming the state-of-the-art model by 4.3%. Finally we evaluated our algorithm in a real-time filtering system. A user study shows that our algorithm is stable and the improved tabletop effectively screens out unintentional touches, and provide more relaxing and natural user experience. By linking their gaze and head behavior to their touch behavior, our work sheds light on the possibility of future tabletop technology to improve the understanding of users' input intention.

CCS Concepts: • **Human-centered computing Ubiquitous and mobile computing**; *Empirical studies in HCI*.

Additional Key Words and Phrases: Unintentional input, Interactive tabletop, Behavior pattern, Gaze and head

ACM Reference Format:
Xuhai Xu, Chun Yu, Yuntao Wang, and Yuanchun Shi. 2020. Recognizing Unintentional Touch on Interactive Tabletop. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 1, Article 33 (March 2020), 24 pages. https://doi.org/10.1145/3381011

1 INTRODUCTION

Nowadays, interactive tabletops are widely used in exhibition [49], education [23, 27], military [4] and emergency control [6, 26], *etc.* Such a tabletop is expected to combine the advantages of a physical table with a digital computer, by allowing users to directly touch the table's surface to issue input commands. A touchable and unmovable tabletop is distinguished from mobile touchable devices by providing benefits such as body support [10], large visual information display [4, 26, 49] and simultaneous multiperson collaboration [3, 4].

An interactive tabletop allows for tangible interaction, but not all contacts on the tabletop surface are intended to trigger a digital response. For example, when writing or drawing on the surface, a user may support her body, or rest her palms, wrists, and forearms on the surface to reduce fatigue [2]. In such cases, an ideal tabletop system should have the intelligence to filter unintentional touches. We define *unintentional touches* as those touches that

[*] Corresponding Author

Authors' addresses: Xuhai Xu, xuhaixu@uw.edu, Tsinghua University, Search Results 30 Shuangqing Rd, Beijing, Beijing, 100084, University of Washington, 1410 NE Campus Parkway, Seattle, WA, 98195; Chun Yu, chunyu@mail.tsinghua.edu.cn; Yuntao Wang, yuntaowang@mail.tsinghua.edu.cn; Yuanchun Shi, shiyc@mail.tsinghua.edu.cn, Tsinghua University.

2474-9567/2020/3-ART33 $15.00
https://doi.org/10.1145/3381011

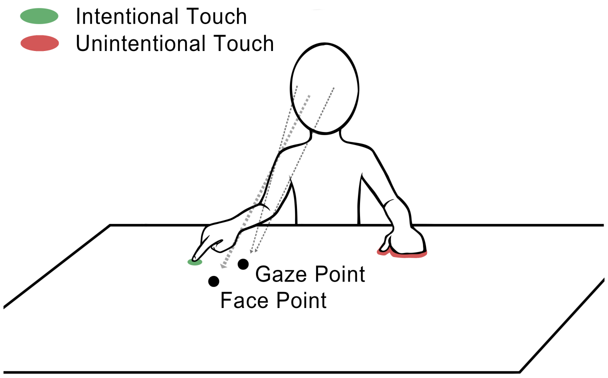


Fig. 1. Intentional and unintentional touches on a tabletop. *Face point* is the intersection of the tabletop surface and the "face ray", which is emitted from the center of the face, in the forward direction. *Gaze point* is the midpoint of two interactions of the surface and "gaze rays", which are emitted from the center of two pupils respectively, in the gaze direction.

do not contribute to any interaction goal. Previous works also use *accidental/unwanted touches* [32, 36]. In our paper, we use them interchangeably with the same definition as *unintentional touches*.

However, currently, the issue of avoiding unintentional touch on tabletops has not been systematically studied in academia; a tabletop system usually recognizes any contact of a human body with the tabletop surface as a touch event. As a result, users tend to behave carefully and prudently on tabletops to avoid triggering unwanted touch events. For example, Annett et al. [1] found that users floated their hand over the screen when handling a stylus and drawing on a tablet to avoid accidental palm touches. In other words, it is hard for users to exhibit many of their natural behaviors on interactive tabletops, *e.g.*, resting hands/arms on the surface, stretching arms/elbows to support the body, gesticulating tentatively during contemplation, *etc.* Users may get even more frustrated when committing unintentional touches disrupts their workflow, despite their effort to avoid them.

In literature, the issue of unintentional touches has been extensively studied [9, 13, 37, 65], but mostly on mobile devices such as phones and tablets [2, 32, 36]. For instance, some research has looked into palm rejection on mobile devices during tasks such as note-taking and drawing [2, 14, 52]. However, the interaction paradigm on a large-scale tabletop is different from that on a mobile device because of its physical size and lack of mobility. For example, users can hold and move the mobile phone or tablet when needed, but when interacting with a tabletop, users must move themselves. In our study involving two usual tasks on multi-touch tabletops, we found that on average, 17.1%/45.3% of touches are unintentional on an interactive/physical tabletop. The high frequency of unintentional touches on tabletops motivates our research.

In this research, we propose a novel approach that leverages the intrinsic relationships between the gaze and face behavior and contact behavior to recognize unintentional touches on interactive tabletops (Figure 1). This is inspired by the previous observation that face and gaze direction usually indicates where people's interest and attention lie [59, 74]. Our work sheds light on the possibility of future tabletop technology to better understand users and their input intention by incorporating the information conveyed by their gaze and head behavior.

This research has three phases, each contributing to answer one of the three research questions.

(1) **RQ1**) *What are the differences between users' gaze, head, and touch behavior patterns on a digital interactive tabletop and those on a physical tabletop?* We conduct a user study to empirically investigate and compare users' gaze, head, and touch behavior on a digital interactive tabletop versus on a physical tabletop, for

completing two usual tasks (map navigation and photo categorization). We identify significant differences in terms of the number of unintentional touches, as well as coordination patterns of gaze and head between the two modalities.

(2) **RQ2**) *How to filter out unintentional touches on a digital tabletop to allow relaxing postures on the surface?* We propose a set of features according to the spatiotemporal characteristics of gaze, head and touch data. Based on these features, we train and compare several machine learning models to recognize unintentional touches. The performance of the best Gradient Boosting model (0.914, 0.913 and 0.913 on average precision, recall, and F1 score, respectively) significantly outperforms baseline models.

(3) **RQ3**) *What is the stability of the algorithm and the usability of the new interactive tabletop?* We implement a system that can recognize and filter unintentional touches on an interactive tabletop in real-time. A user study validates the stability of our algorithm (F1 score equals 90.6%), and shows that our new system can effectively screen out unintentional touches, and provide more natural and relaxing user experience.

2 BACKGROUND

We review related work on addressing unintentional touches on mobile devices, how the face/gaze direction indicates a human's intention, and how to use gaze to facilitate touch during interactions. In addition, we also review another body of related work that leverages unintentional touches as an alternative type of input method.

2.1 Recognizing Unintentional Touch on Interactive Devices

Several works have been conducted to recognize unintentional input on small-scale touchscreen devices. Matero and Colley [36] tried to classify accidental touch events on mobile phones in daily usage. They investigated numerous typical operations on mobile devices (*e.g.*, sweeping on the screen, device handling and phone calls) and proposed a rule-based classification algorithm according to the contact area and touch duration. Their best algorithm can eliminate 79.6% of unintentional touches with a 0.8% false-positive rate. Lu and Li [32] analyzed accidental touches during the standby mode of mobile phones. They compared intentional touch gestures and unintentional touches on a turned-off screen, *e.g.*, putting the device in a pocket. They used a number of touch spatiotemporal features such as duration, touch area trajectory, pressure level on the screen, and acceleration deviation from the motion sensor to classify whether a touch on a turned-off screen was accidental. Their final decision tree model achieved 98.2% on precision and 97.6% on recall.

Another big family of touch-based devices are tablets. Annett et al. [1] compared the unintended touches that occurred when using a stylus on digital versus media. They found that users maintained some unnatural and tiring postures on the digital tablets to avoid unintentional touches caused by their palms, *e.g.*, floating their hands over the screen. They further investigated different algorithms to remove accidental touches on stylus-based tablets [2]. Their best algorithm used the distance between the screen and the stylus as the threshold and achieved approximately 86% accuracy for the classification. Julia Schwarz et al. [52] focused on unwanted interaction triggered by palms on tablets. They used similar spatiotemporal features in [32] to distinguish palm touches from stylus input. Their decision forest model achieved a precision of 97.9% and 0.016 errors/stroke.

To our knowledge, previous works on recognizing unintentional touch were performed on mobile phones and tablets. However, tabletops can accommodate distinct interaction activities. Beyond that, they are more prone to unintentional touches due to their increased surface area and a wider range of possible inputs. For instance, users can easily lean on the tabletop or put an arm on the surface to support themselves. These actions would be far less common when interacting with smartphones or tablets. Although the interactions on tablets have some overlap with those of tabletops, previous works focus on the distinction between the stylus and accidental finger/palm touches. This relieves the technical difficulty, because the stylus has unique properties such as small static touch area, as pointed out in [52]. The difference on the interface will also lead to the difference in features such as touch trajectories [40]. In this research, we address the gap by investigating unintentional touches on digital tabletops.

2.2 Gaze and Face for Interaction and Intention

People primarily direct their gaze towards regions of interest [17, 20, 25, 56, 59, 68, 69, 71, 72, 74]. Several works evaluated the coordination patterns of mouse cursor and gaze during daily interaction on PCs. Huang et al. [19] investigated the behavioral patterns of gaze and mouse cursor during web searches. They found that while users' gaze-cursor alignment varied a lot, users lagged their cursor behind their gaze by at least 250 ms and on average 700 ms. Liebling and Dumais [30] looked into users' daily work on PC in the wild. In contrast to the previous findings, they suggested that the gaze leads mouse only in two-thirds of the time. However, the results in [30] still supported the consistency between the cursor and gaze during intentional interactions.

The consistency between gaze and intention inspired a number of new interaction techniques. Jacob [20] proposed several novel interaction methods using eye gaze as the independent channel, *e.g.*, object selection on a PC screen. Since then, many gaze-enhanced input approaches have been proposed. Researchers have tried to facilitate pointing/touch interaction with the gaze. For example, Zhai et al. [76] proposed MAGIC pointing, which used gaze for object suggestions on screens and hand-control for confirmation. Stellmach et al. [58] designed a gaze-supported selection method for distant display. They used gaze for selection and hand gestures on a handheld device for manipulation. Sidenmark et al. [54] leveraged the coordination of gaze and head to acquire gaze targets in virtual reality. Turner et al. [62] extended the idea with more gaze-touch combination patterns, such as performing RST (rotate, scale, translate) by touch on the trajectory suggested by gaze. Voelker et al. [64] developed an indirect input system, where a user touched on a horizontal surface and looked at a vertical screen. Both Turner et al. [63] and Mauderer et al. [39] suggested selecting out-of-reach objects through combination of far gaze and close touch on a tabletop. Pfeuffer et al. [43, 45] proposed several scenarios on a tablet screen, using gaze for object selection or button selection and using touch for RST. Sidenmark et al. [55] utilized eye-hand coordination patterns on interacted objects for eye-tracking calibration in the virtual reality setting. All these works employed gaze as a viable input channel, showing that gaze could play a significant role in enhancing interaction across mobile devices such as tablets and phones or non-mobile devices such as tabletops.

However, these works mainly employed gaze as an active channel during interactions. Fewer works discussed using gaze as a cue for human intention. Schwarz and her colleagues [51] used the gaze and body direction to determine whether the user was engaged in a Kinect game. Mariakakis et al. [34] detected users' gaze to determine whether they were focusing on their phones. Pfeuffer et al. [44] proposed a mechanism to switch between direct and indirect input mode based on the alignment of the input area and the user's visual attention. Compared to gaze direction, a relatively lower-cost alternative is to use facial orientation, which also proves to be a powerful proxy for attention. Hollands et al. [18] suggested that in approximately 70 percent of scenarios, the gaze direction and face direction are the same. Liao [29] used facial direction as one of the clues to determine a user's most interested picture in a photo collection. Maglio et al. [33] built a system that used face direction to determine where the user intends to interact with the environment. All these works indicate that eye gaze and head pose can be used to predict regions of interest. To the best of our knowledge, this is the first work evaluating the performance using gaze and face features for recognizing the unintentional inputs on touchscreens.

2.3 Leveraging Unintentional Touches as Intentional Input Methods

Our goal is to remove unwanted effects of unintentional touches on tabletops. However, some literature interprets the nature of these touches differently: instead of "ignoring" unintentional interactions on tabletops, they tried to "use" those interactions. For example, Koura et al. [24] proposed to use the forearm, which was usually considered problematic (*e.g.*, incorrect recognition and occlusions), as a new interaction technique for menu manipulation and data storage. Le et al. [28] proposed to interpret leaning into a new class of gestures to enhance interaction. Matulic et al. [38] extended hand interactions from fingertips to the whole hand in hand-shape based interaction. Zhang et al. [77] proposed to leverage various hand postures such as using the palm to augment pen and touch interactions. These works provide another perspective to deal with unintentional touches. However, any interaction technique requires additional attention from users. These works do not solve users' concerns about accidentally creating unwanted interaction. Recently, Serim and Jacucci had an interesting discussion on the distinct definition of explicit vs. implicit interaction, and identified new considerations for design and evaluation of implicit interaction [53]. Similarly, our work tries to obviate the concerns about accidental actions and enable users to operate on tabletops freely.

In the rest of the paper, we try to answer the three research questions one by one. We begin by answering **RQ1** via Study 1.

3 STUDY 1: INTERACTION BEHAVIOR ON TABLETOPS

We conducted Study 1 to obtain empirical knowledge on how frequently unintentional touches occurred while completing common daily tasks on tabletops. We also wanted to investigate the differences among touch behaviors on the two types of tabletops: digital tabletops equipped with multi-touch interactive touch screens versus physical tabletops with unresponsive surfaces. Measuring the difference helped us understand how users changed their interaction behaviors when using touch-sensitive technology. In addition, the data collected from in this study was leveraged to train classification models to recognize and filter unintentional touches.

3.1 Participants

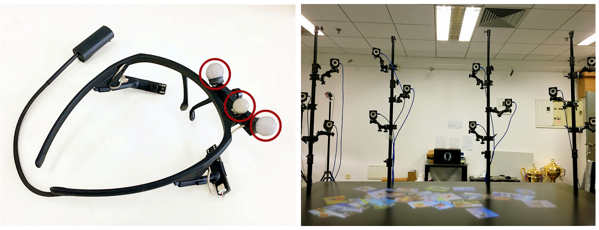
We recruited 12 participants (8 males, 4 females, Age = 23.2 ± 1.47) from a local university through email. All participants had at least four years of experience with touchscreen devices such as smartphones and tablets, and used mobile phones on a daily basis. None of them had used an interactive large-scale tabletop before.

3.2 Apparatus

Figure 2 illustrates the apparatus used in this study. We used a customized interactive tabletop as the experimental platform. The size of the system was 250 cm × 160 cm × 80 cm, with a 220cm × 135cm screen at the center of the table surface, powered by a built-in computer with Windows 10 OS. Its capacitive screen was realized via an iPCT transparent touch foil [75] that could sense up to 20 touch events simultaneously at 100Hz, recording the timestamp, location and state (touch down/up/move) of each touch point. Here we defined a *touch event* as a detected touch point from being put down to lifted up. Note that the system did not provide contact area information. A wide touch area registered as multiple simultaneous touch events. Therefore, if a palm was put on the surface unintentionally, there would be a group of unintentional touch events. Moreover, it did not have any pre-filtering algorithm and registered all touch events, which served as a good platform for answering our research questions.

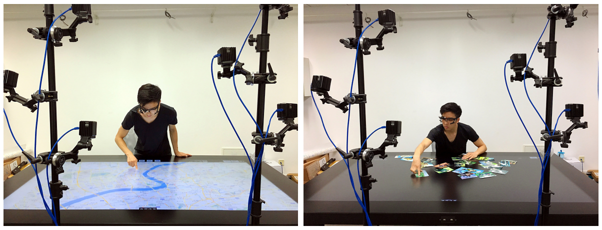
Additionally, we used a head-mounted gaze tracker, Binocular 120Hz Pupillab Eye-tracker [21] to track users' gaze direction relative to the gaze tracker. The Pupillab was calibrated with its marker system for each participant.

We employed an Optitrack system with 12 cameras [46] to capture the 3D location and angle of the gaze tracker at 120 Hz. The direction of the gaze trackers indicated the head pose and face direction. It could then be used to calculate the face/gaze fixation points. The fixation position on the tabletop where a user was facing/looking at was computed in real-time in Unity 5.5.4. Figure 1 shows the definition and the calculation of the *face point* and *gaze point*. We tested the eye-tracking on two authors with 9 markers (3 × 3, aspect ratio, for all regions of the screen) on the table, with an average gaze angle of 25 ° and an average vision distance of 50 cm. The



(a) Pupillab with Markers (b) Optitrack cameras Setup

Fig. 2. System Apparatus



(a) Digital, Stand, Map Task (b) Physical, Sit, Photo Task

Fig. 3. Experiment Setting Examples

tracking accuracy is 2.39 ± 1.48 ° and the precision is 1.97 ± 1.05 °. The tabletop, Pupillab, and Optitrack were three separate systems. All data collected by different devices were synchronized using timestamps in Unity.

3.3 Design

We used a two-factor within-subject design. The two independent variables are the type of the tabletop (digital vs. physical) and the user posture (sitting vs. standing, initially in the middle of the bottom edge of the tabletop). The presentation order of the conditions is counterbalanced using a Latin Square design.

3.3.1 Tasks. We experimented with two typical tasks on tabletops: map navigation [4, 5] and photo categorization [31, 60]. These two tasks were commonly used to investigate interactions in previous research. They involve the most common interactions on the surface, such as tapping, dragging, zooming, rotating, *etc.*

In the map navigation task, a mid-sized city map is shown on the tabletop, initially at the same size as the screen and centered. None of the participants were familiar with the city. Two subtasks needed to be finished under each condition. In the first subtask, participants were asked to find three landmarks on the map, after which they were instructed to draw a route connecting them. In the second subtask, participants needed to find two landmarks and design a subway route from one to the other. There were several candidate routes, and the subtask could be completely by sketching out any one of them. To avoid memory effect, the details of the subtasks (*e.g.*, the landmarks) between the digital and physical tabletops were different. But the tasks are kept the same among all participants to maintain the difficulty level.

In the photo categorization task, 40 fixed-size photos (standard A5 size, 148 x 210 mm) were displayed on the surface. These photos belonged to four different groups such as forest, snowfield, beach, and house. 10 samples from each category were randomly scattered on the table in a reachable region. Participants were required to sort them into four piles by category. We prepared 16 different categories to avoid the learning effect.

3.3.2 Setup of The Digital Tabletop. For the interactive tabletop, the operations in the map navigation task included 1) moving and zooming the map in the dragging mode, with at least one touch point for moving and at least two touch points for zooming in/out; 2) drawing lines/markers on the map in the drawing mode, where a marker would appear with a tap and a line would be drawn following a finger's movement; 3) erasing lines in the clearing mode, where any line crossed by a finger trajectory would be removed. Participants could press a button to switch between the modes.

The operations in the photo categorization task were simple. The photos could be moved by at least one finger dragging and rotated by at least two fingers rotating. The photos were not zoomable.

3.3.3 Setup of The Physical Tabletop. We set up the physical interface on the same interactive tabletop (Figure 2b), on which we disabled all touch responses and displayed a pure black wallpaper to serve as a plain physical table. All electronic elements on the digital tabletop were substituted with paper materials. We verified the robustness of the capacitive touchscreen of our interactive tabletop to sense touch through papers. Therefore, all touch behaviors on the physical tabletop could also be recorded.

For the map navigation task, we placed a paper map with identical size (before zooming) and color on top of the surface. Participants were asked to finish the same two sub-tasks as described. They needed to move fingers along the route for the route drawing task. For the photo categorization task, we placed 40 printed photos of the same size as the digital photos on the tabletop. To retain consistency with the digital tabletop, participants were free to move the paper map and photos on the desktop but not allowed to pick them up from the surface.

3.4 Procedures

Participants first signed the consent form. Before the experiment, we notified the participants to act naturally and not to worry about the task performance. The only requirement was to finish each task in 10 minutes. Participants first put on the Pupillab device and went through the calibration procedure. They were given 3 minutes to familiarize themselves with the interactions on the tabletop. After that, they performed the aforementioned tasks under different conditions. On average a map navigation task took about 5 minutes and a photo categorization task took roughly 3 minutes. The full experiment with four conditions, two digital and two physical, lasted approximately 30 minutes. After the experiment, we briefly interviewed each participant about their subjective feelings during the tasks, especially on the perceived difference between the digital and the physical tabletops. Finally, participants were thanked and dismissed. Each participant was offered $10 as compensation.

3.5 Data Collection and Annotation

During the experiment, participants' direction of face and gaze (as described in Section 3.2), *face/gaze point*s (Figure 1) and touch points were recorded for each frame. On average, about 100,000 data points were collected per participant. Furthermore, we recorded the table screen and videotaped their behavior throughout the experiment. After data collection, we smoothed and resampled the gaze trajectory using Stampe's two-stage filter [57] and two sample weighted average [30].

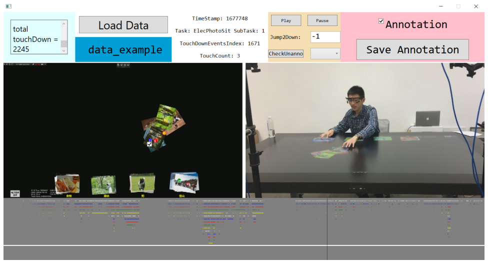
We collected a total of 1,228,250 frames and 27,384 touch events from twelve participants. Each touch event had 97.1 frames of data (SD = 211.9) on average. Two authors independently annotated every touch event (intentional or not) throughout the data using a simple self-developed tool, which visualized and replayed the participants' behavior (see Figure 4). Cohen's Kappa inter-rater reliability equaled 0.73. Conflicts were solved by a collective review of the two authors.

3.6 Results

Participants spent similar amounts of time on digital and physical tables to finish the tasks ($t_{11} = -0.19, p = 0.85$). Overall, we discovered a large proportion of unintentional touches: 17.1 (SD = 13.0) / 45.3 (SD = 10.5) percent of touches were unintentional on an interactive/physical tabletop. We observed interesting differences between the two types of tabletops, in terms of behavior patterns (Section 3.6.1), number of unintentional touches (Section 3.6.2), and mental models (Section 3.6.3). We summarize these differences one by one.

3.6.1 *Different Behavior Patterns on Physical and Digital Tabletops.*

We found remarkable differences in touch behavior between the digital and the physical tabletops. Figure 5 depicts the typical behavior patterns on these two forms of tabletop. The specific difference is summarized in Table 1. Although participants were instructed to behave as natural as possible, they tended to be quite cautious on the digital tabletop, which is similar to the findings on tablets [1]. 8 out of 12 participants mentioned "careful" or "safe" during the interview. *"It always came to my mind that I need to be careful with the screen, otherwise I will trigger unwanted touch events"* (P2). *"I soon got to float my hands above the surface. That's safe"* (P10). Therefore, only a small number of unintentional


total touchDown = 2245

TimeStamp: 1677748

Task: ElecPhotoSit SubTask: 1

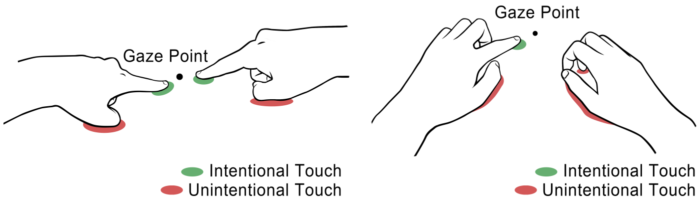
TouchDownEventsIndex: 1671

TouchCount: 3

Jump2Down: -1

☑ Annotation

Fig. 4. GUI of Annotation Tool: The video of the tabletop screen (left part) and the participant (right part) are presented after calibration. Annotators can navigate back and forth to any time frame. All touch events and *face/gaze point* are visualized in the left part at frame level.



(a) Patterns on Digital Table

(b) Patterns on Physical Table

Fig. 5. Typical Behavioral Patterns on the Two Tabletops

Table 1. Summary of the behavior difference between the digital and physical tabletop.

	Digital Tabletop	Physical Tabletop
Intentional Touch	• Users are cautious, with arms floating over the surface. • Mostly one or two fingers are used for interaction. • Other parts of hand/arm (except the finger involved in interactions) stay far away from the screen. • The unused hand rest on the edge of the tabletop. • **Usually gaze leads or stays in line with the touch**.	• Users are casual, with arms often resting on the surface. • Mostly two or three fingers are used for interaction. • Other parts of hand/arm (except the finger involved in interactions) lay on the tabletop surface. • The unused hand rest on the surface casually. • **Gaze mostly stays in line with touch**.
Unintentional Touch	• Touches are less frequent than those on the physical tabletop due to the carefulness. • Edge accidental touches are inevitable and usually triggered by the palm/elbow resting on the edge. • Usually touches are relatively far from the area of the participants' attention. • Many touches are static and ephemeral. • **Gaze does not stay in line with touches**.	• Touches are very common and appear with intentional touches simultaneously. • Many touches are triggered by the hand/arm resting on the surface (no matter the hand is being used or not). • Usually touches gather together if triggered by the hand that is being used, and are far from the attentive area if they are triggered by the hand that is not being used. • Touches are more dynamic and long-lasting than those on the digital tabletop. • **Gaze does not stay in line with touches**.

touch events were recorded on the digital surface. In most cases, participants used only one or two fingers to operate, keeping the rest of their hands and arms far from the surface. In contrast, on the physical table, users were more relaxed. They naturally rested their arms on the surface. *"I did the same as what I will do with my wooden desk"* (P2). When users pointed at the target (usually with two or three fingers), other fingers or even the palm were often put on the tabletop casually (see Figure 3b). These touches were unwanted but very common on the physical table.

3.6.2 More Unintentional Touches When Operating with The Physical Tabletop. We first ran two-way ANOVAs (table type × posture) according to the experiment design on the number of intentional touches and unintentional touches separately. Neither table type nor posture had a significant effect on the number of intentional

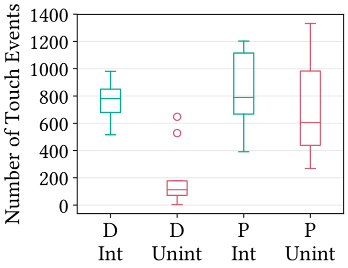


Fig. 6. Boxplot of touches on two tabletops. Data of sitting and standing postures are merged. "D/P" refers to digital/physical tabletop settings. "Int/Unint" refers to intentional/unintentional touches. The same below.

touches ($F_{table}(1,11) = 0.826, p = 0.38, F_{posture}(1,11) = 0.13, p = 0.73$). The effect of the two main factors' interaction was also not significant ($F_{table\times posture}(1,11) = 2.67, p = 0.13$). As for the number of unintentional touches, the results indicated significance on both main factors and their interaction ($F_{table}(1,11) = 45.16, p < 0.01, F_{posture}(1,11) = 14.13, p < 0.01, F_{table\times posture}(1,11) = 29.54, p < 0.01$). More unintentional touches were observed on the physical tabletop and in the sitting posture.

The posture factor was designed to incorporate a variety of different behavior patterns on the tabletop. We hence combine the data of the standing and sitting postures. In Figure 6, the boxplot shows the 12 participants' average number of the intentional/unintentional touches on the digital/physical table respectively. We ran two paired t-tests to compare the number of intentional and unintentional touches between two tabletops. The results do not show significance of tabletop types on the number of intentional touches ($t_{11} = -0.73, p = 0.48$), but revealed significance on unintentional touches ($t_{11} = -5.86, p < 0.01$). Similar numbers of intentional touches were triggered on two tabletops, but more unintentional touch events appeared on the physical tabletop than the digital tabletop. In addition, we found an interesting case where two participants exhibited noteceably higher numbers of unintentional touches on the digital tabletop. From the video, the two participants made attempts to "imitate" operations on the physical table by placing their palm on the surface, according to their understanding of "being natural". This also reflects the difference in behavior patterns between the digital and physical settings.

3.6.3 Different Mental Models Towards Two Types of Interface. The interview results are also interesting: although participants behaved quite differently on two surfaces, they did not mind the differences: it was natural to be careful on the digital screen and casual on the physical surface. Almost all users (11 out of 12) expressed that they would feel uncomfortable if they were asked to interact with the touch screen in the same way as the plain table. *"I never touch screens...in the way on ordinary tables, otherwise, it will annoy me with a lot of unwanted triggers!"* (P1). *"It does sometimes make me tired, but compared to unwanted triggers, I'd rather be careful"* (P11). One of the two outliers said *"Although I tried to operate naturally on the digital screen, I cannot resist the temptation to raise my hands. It is an awkward experience..."* (P7). We speculate this phenomenon is caused by different mental models towards digital and physical tables. Users are accustomed to use digital screens more cautiously. In order to avoid triggering unwanted touch events, they are willing to pay more attention to their actions.

4 TOUCH INTENTION IDENTIFICATION

Our goal was to build an intelligent tabletop that allows users to operate on an interactive tabletop in a similar manner as on a physical table. To answer RQ2, we extracted features from user behavior data and trained a binary classifier. We blended the data of two tabletops, only focusing on whether the touches were intentional or not.

The analysis of the behavior patterns (Table 1) provided insights for feature extraction, which shed light on the features that were useful for classification. Specifically, the difference between intentional and unintentional touches is summarized as follows:

- The majority of the gaze is in line with intentional touches, but not with unintentional ones (*Gaze/Face Distance*)
- A large number of unintentional touches are more static, ephemeral and closer to the edge compared to intentional ones (*Side Distance*)
- Intentional touches usually involve one to three fingers while unintentional ones often appear in groups (*Clustering*)
- Historical behavior affects current touches. *E.g.*, sometimes gaze leads intentional touches shortly, the unintentional touches clustering area often spawn more unintentional touches subsequently (*History*)

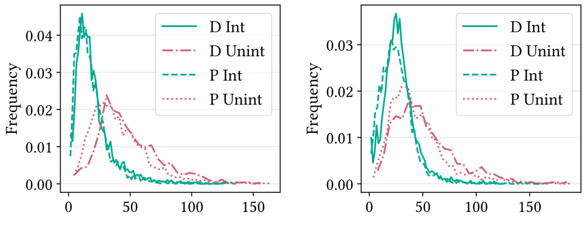
We randomly split our dataset into an optimization set (30%) for feature analysis, and a traintest set (70%) for model training and testing. In Section 4.1, we introduce the features extracted from the gaze, face and touch data (as summarized in Table 2). We present an overview of these features through descriptive statistics, using data from the optimization set. Note that we purposefully only extract features that are compatible with real-time system implementations. Other features, *e.g.*, the lifetime of a touch event, are not included in our analysis.

4.1 Feature Definition

4.1.1 Gaze/Face Distance. Within each frame, the distance between each touch point and *face/gaze point* (defined in Figure 1) is named as *face/gaze distance*. The distribution of distances of four categories (intentional/unintentional × digital/physical) is shown in Figure 7. In comparison with unintentional touches, intentional touches have smaller *gaze/face distance* in both digital and physical conditions. Compared with Figure 7b, Figure 7a has a more pronounced difference between intentional and unintentional touches. This indicates that the *gaze point* stays more in line with touches than the *face point*.

Figure 8a shows a heatmap of touch locations. Intentional touches are distributed in the center of the surface while the majority of unintentional touches are scattered near the bottom edge. *Side distance* is defined as the perpendicular distance of a touch point to the nearest screen edge. The near-margin property of unintentional touches on both tabletops is salient (Figure 8b).

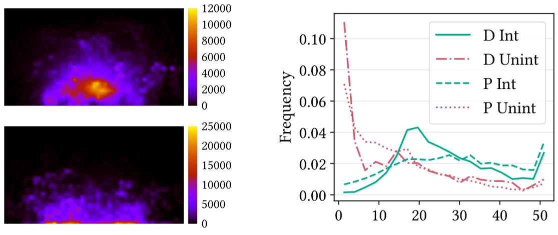
4.1.2 Clustering. We observed obvious spatial clusters of touch points during the study, especially on the physical table. This may be explained by users' casualness: users would usually rest their hand or even their arm on the table, which triggered a wide contacting region. Owing to the hardware properties of the customized tabletop, the system recognized a large contacting region as a group of separate touch points rather than a continuous area. A dynamic distance matrix was created. Each row and column represents a touch point at the frame. *Touch*



(a) Gaze Distance/cm

(b) Face Distance/cm

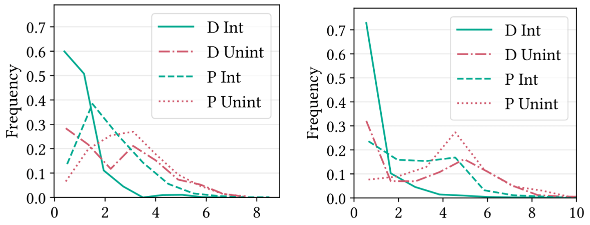
Fig. 7. Gaze/Face Distance



(a) Heatmaps of touches

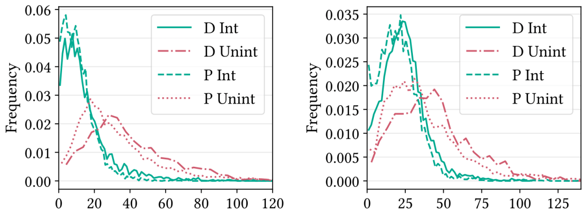
(b) Side Distance/cm

Fig. 8. Surface Distribution of Touches



(a) Number of Touches in 10cm (b) Number of Touches in 20cm

Fig. 9. Number of Touches within Distance Threshold



(a) Gaze Distance/cm (b) Face Distance/cm

Fig. 10. Historical Gaze/Face Distance

distance, the mean of the element of the matrix, is the distance between two touch points. *Number of adjacent touches* of each touch point is defined as the count of other touches whose distance to this touch point is smaller than some particular threshold. Figure 9 suggests apparent clustering of unintentional touches on both tabletops: most intentional touches have zero or one *adjacent touch point*, while the majority of unintentional touches have at least three *adjacent touch points*. Several distance thresholds were tested (Figure 9 only shows two examples). Compared to a 10 cm threshold (see Figure 9a) and other values, the threshold of 20 cm appeared to have the strongest splitting power between intentional and unintentional touches (see Figure 9b). Hence, the clustered threshold is set to 20 cm, which is approximately the width of a stretched palm.

4.1.3 History. As found by previous literature, the gaze leads hands in many cases [19, 30]. This indicates that the position of current intentional touches may be close to previous *gaze points*. Moreover, we observe the phenomenon of clustering not only on the spatial dimension but also on the temporal dimension: if a location is clustered with a group of touches, a new touch will be more likely to appear within or near the clustering area shortly afterward, especially when users are unaware of the touches. Hence a historical time window may carry useful information. *700 ms* is selected as the width of the window according to [19], counted backward from the down time of a touch event. *Historical gaze/face distance* represents the distance between the current touch point and previous *gaze/face points* in the window. *Historical touch distance* and *historical number of adjacent touches* can be similarly obtained based on the distance between the current touch point and previous ones. Figure 10 shows the *historical gaze/face distance*. It is similar to Figure 7 but with a higher degree of separation.

We categorized all features into four different groups based on their characteristics (see Table 2). Some features appear more than once because they belong to different types simultaneously. Then we trained machine learning models based on these features.

Table 2. Feature Groups for Classification

Gaze/Face	Side	Clustering	History
Gaze/Face Distance Historical Gaze/Face Distance	Side Distance	Touch Distance Historical Touch Distance Number of Adjacent Touches (20cm) Historical Number of Adjacent Touches (10cm)	Historical Gaze Distance Historical Face Distance Historical Touch Distance Historical Number of Adjacent Touches (10cm)

4.2 Model Comparison and Feature Comparison

To obtain the best binary classifier for unintentional touch identification, we compared different machine learning models using our features in Section 4.2.1. After obtaining the best model, we compared it with other baseline models in Section 4.2.2. In addition, we also conducted a feature ablation study in Section 4.2.3 to investigate the relative importance among different feature types.

4.2.1 Model Selection. A good classifier is capable of identifying intentional touches and therefore filtering unintentional ones. Yet more significantly, the effect of each type of feature in the model is also worthy of exploration. This indicates the importance of features during classification, which can provide insights into human behavior patterns.

Six machine learning models were compared on the traintest set, including LR (Logistic Regression), NB (Naïve Bayes), KNN (K-Nearest Neighbor), RF (Random Forest), GB (Gradient Boosting) [11], and MLP (Multi-Layer Perceptron). The first three models are typical approaches to classification problems. Both Random Forest and Gradient Boosting employ decision trees as weak classifiers. They are suitable for feature selection [7]. MLP also proves to be a suitable model for complex problems where the relative importance of features is unknown [12].

Every person's behavioral pattern has consistency, thus merging all users' data for training will lead to information leaks, which makes the classification naive and impractical. Therefore, we trained models using the leave-one-user-out method, where the models were trained on the data of 11 participants and tested on the remaining one. This can better measure the model's generalizability. The tuning of the hyperparameters of each model was conducted within the leave-one-user-out set using an inner five-fold cross-validation loop. After the best parameters were chosen, the ignored user's data was used for testing. Final metrics (precision, recall and F1 score) were the average of 12 models (same as the number of participants). Table 3 summarizes the testing result of all models. We chose the GB model for our later analysis since this model significantly outperformed than other models.

4.2.2 Baseline Model Comparison. We further compared our best model with two baseline models:

(1) A naive threshold-based model using *gaze distance*. The threshold was determined from the optimization set as 23.9cm which maximizes the split of the two types of touches.

(2) A modified version of the state-of-the-art decision-tree model from [52]. Although this model is intended for distinguishing palm touches from stylus input, many features are transferable to our system. We implemented the instant version for a fair comparison. The features include touch distance, touch points travelling speed and acceleration. As our tabletop does not provide touch area information, we omitted this feature type.

Table 3. Results of Model Comparison. A paired t-test on the cross-validation F1 scores between the GB and the RF shows significance $p < 0.001$. The hyperparameters of the GB model among cross-validation is consistent. The best GB model has 100 as the number estimator and 3 as the maximum depth of each tree.

Models	Prec	Rec	F1
NB	0.832 ± 0.016	0.836 ± 0.020	0.833 ± 0.008
KNN	0.842 ± 0.029	0.844 ± 0.026	0.843 ± 0.027
MLP	0.850 ± 0.018	0.848 ± 0.033	0.848 ± 0.025
LR	0.874 ± 0.020	0.869 ± 0.012	0.872 ± 0.009
RF	0.893 ± 0.020	0.892 ± 0.011	0.893 ± 0.009
GB	**0.914** ± 0.020	**0.913** ± 0.026	**0.913** ± 0.014

Table 4. Results of Baseline Comparison. A paired t-test on the cross-validation F1 scores between the full GB and the Palm rejection model shows significance $p < 0.01$.

Models	Prec	Rec	F1
Gaze-threshold-based	0.794 ± 0.026	0.791 ± 0.047	0.793 ± 0.033
Palm rejection [52]	0.873 ± 0.057	0.869 ± 0.021	0.870 ± 0.032
Full GB	**0.914** ± 0.020	**0.913** ± 0.026	**0.913** ± 0.014

Table 4 summarizes the results on the test set. Our model significantly outperforms baseline 1 by 12.0% and baseline 2 by 4.3% on the average F1 score. The big gap between the result of the reimplemented model from [52] and that of the original paper can be caused by a few reasons. The interaction paradigms on tabletops are different from tablets. The features of a stylus's input in [52] can be very different from a finger's input. [52] considers palm as the major source of unintentional touches, while in our case other parts of the hand and the arm were also a source. For example, Figure 5a shows that the thumb can often trigger unwanted touches.

4.2.3 Feature Ablation Comparison. We enhanced our findings through a feature ablation study to investigate the effect of gaze and face features, *i.e.*, removing certain feature types and observing the performance decrease of the model.

The removal of the gaze features, including any features that involved gaze information, both in *Gaze* and *History* feature types, led to a significant drop in the model performance. This emphasized the importance of gaze-based features, which is a reflection of previous research concluding that people tend to look at the place where they have interest [17, 74]. Our results further reveal its positive effect in classification and show that *Gaze* is a strong clue that conveys human intention. These findings are consistent with previous studies on human gaze-attention correlation [17, 20, 25, 59]. The intuition of "where we look is where we have interests, therefore where we touch" is validated again by our study in the context of tabletop interaction.

Surprisingly, compared to *Gaze*, the removal of *Face* did not cause much decrease in the performance. Compared to the model with gaze features ablated, the model with all gaze and face features removed just decreased by 0.009 on the F1 score. These results suggest that *Face* may not be an appropriate alternative for *Gaze*. This indicates that there will be a significant drop in classification performance by substituting gaze tracking with a lower-cost head tracking. This may be explained by the lack of head-gaze consistency during the interaction on the large-scale surface.

To further understand the consistency, we investigated the included angle between the *face ray* and *gaze ray* during the experiment. Figure 11 shows the heatmaps of the average included angle between the two rays throughout the Study 1. The angle is calculated frame by frame and takes the *face point* as the position of the

Table 5. Results of Feature Ablation. A paired t-test on the cross-validation F1 scores between the full model and the model with Face feature ablated shows significance $p < 0.05$.

Feature Ablated	Prec	Rec	F1
− (Full)	**0.914** ± 0.020	**0.913** ± 0.026	**0.913** ± 0.014
Face	0.900 ± 0.021	0.892 ± 0.020	0.895 ± 0.018
Gaze	0.842 ± 0.029	0.837 ± 0.031	0.838 ± 0.022
Gaze&Face	0.831 ± 0.032	0.826 ± 0.022	0.829 ± 0.026

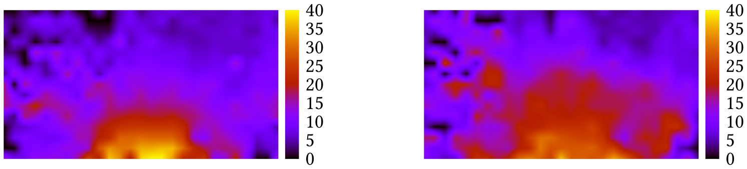


Fig. 11. Face-gaze average included angle (in degree unit) heatmaps on the tabletop. The *gaze/face points* outside the screen are discarded. Left) sitting posture. Right) standing posture

value. Two heatmaps of the sitting and standing postures share similar patterns. Users' gaze direction has a greater deviation from their face direction in the main working area. The figure of standing posture has a wider red region than that of sitting because of the larger movement space for the upper limbs during the standing posture. The heatmaps indicate that when people are focusing on tasks within their reaching range, they may have more gaze movement rather than head movement. The gaze direction greatly deviates from the normal direction of the face – users do not always stare strictly forward, but often move their eyes around, especially when they are operating on the surface close to them.

4.3 Misclassification Analysis

We thoroughly investigated the misclassified samples during training and testing. This can equip us with more insights about the similarities between intentional and unintentional touches that fool our algorithm.

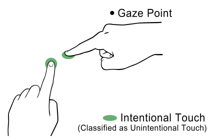
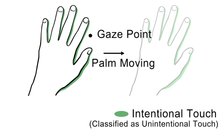
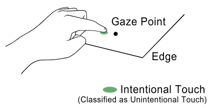
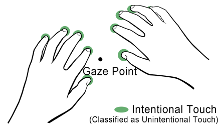
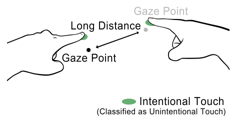
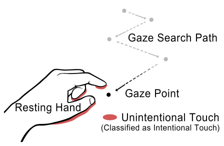
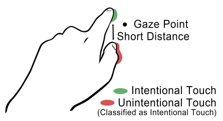
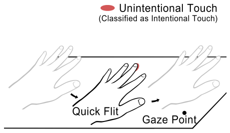
Close inspection shows that some unintentional touches have almost the same values on most features as some intentional ones or vice versa. Table 6 summarizes the relatively frequent patterns of wrongly classified samples in the GB full model. These samples indicate the shortage of the model. The reasons may lie in the fact that some features that hide deep beneath the data are not extracted effectively, or even some patterns that can not be captured by the current feature set.

The comparison between the GB model with human annotation can reveal some information that is not observable from the data. The videotape allowed the authors to obtain the context information of operations with two additional dimensions, both historical and future, either task-specific or person-specific. These two dimensions can further provide insights into the misclassification issue. For instance, *"The participant just found the start place on the map. So in the next step, he needed to find the destination, so he was searching at that time."* (Author 1, historical and task-specific). *"Several seconds after this touch, he switched his attention to another place far from the current position. Then he got excited since he found the target, which was quite different from his current state."* (Author 2, future and user-specific). These messages cannot be reflected in the data from the tabletop or eye tracker because both future and task-specific information are missing. This indicates the limitation of just leveraging gaze and touch information. The task information and emotional state (*e.g.*, getting excited) can be two promising candidate features that worth further exploration to achieve better classification performance.

5 STUDY 2: USABILITY STUDY

To answer RQ3 regarding the viability of the algorithm, we implemented a real-time filtering system by integrating the GB model with the tabletop. This tabletop provided a new experience to users. How well does the algorithm perform in real-time? Will users accept it? Will they feel the benefit? In Study 2 we answer these questions by testing the system's usability and effectiveness.

Table 6. Misclassification Pattern Visualization and Summary

Patterns	Details	Patterns	Details
 Gaze Point Intentional Touch (Classified as Unintentional Touch)	(a) Proficient Operation Without Gaze: Proficient operations without the need of gaze. The misclassification may be caused by large gaze distances since proficient operations do not involve gaze extensively.	 Gaze Point Palm Moving Intentional Touch (Classified as Unintentional Touch)	(b) Whole Palm Continuous Interaction: Using the whole palm for moving, rotating and zooming in/out. Close touch distances and large numbers of touch points are common features of unintentional touches.
 Gaze Point Edge Intentional Touch (Classified as Unintentional Touch)	(c) Touch Close to The Bottom Edge: Touches near the bottom edge of the tabletop. This may be biased by the large number of unintentional touches that appear near the edge of the tabletop.	 Gaze Point Intentional Touch (Classified as Unintentional Touch)	(d) Two-hand Close Cooperation: Close hands cooperation with many simultaneous touches. Touches may be misclassified because of small touch distances.
 Gaze Point Long Distance Gaze Point Intentional Touch (Classified as Unintentional Touch)	(e) Two Touches with Gaze Shifting: Two touch points are far from each other, with gaze switching focus between them. The possible reason of misclassification is large gaze distance when the gaze moves away from one touch point to another.	 Gaze Search Path Gaze Point Resting Hand Unintentional Touch (Classified as Intentional Touch)	(f) Resting Hand with Gaze Passing by: Fingers resting on the surface with unconsciously close gaze point when the user is searching target or thinking. The small gaze distance can lead to misclassification.
 Gaze Point Short Distance Intentional Touch Unintentional Touch (Classified as Intentional Touch)	(g) Touch Close to Intentional Area: Touches caused by a part of the palm or redundant fingers of the hand being used, close to the intentional area. The gaze may be close to unintentional touches, which is a common feature for intentional ones.	 Unintentional Touch (Classified as Intentional Touch) Quick Flit Gaze Point	(h) Hand Flying over the Surface: Unintentionally flying over fingers or transient resting palms on the surface. The combination of moderate touch duration and gaze distance may confuse the classifier.

5.1 System Implementation

We obtained our final GB model by training on the whole dataset collected in Study 1 with the same hyperparameters. When a touch event is registered, it will pass through the classifier to determine whether it is intentional. If the event is classified as intentional, then it will be processed as a normal operation. Otherwise, it will be removed immediately and won't change anything on the surface. The average computational time of our algorithm is 83.3 ms.

5.2 Participants and Apparatus

We invited another 12 participants from the local university through email (7 males, 5 females, Age = 22.7 ± 2.1). They were all familiar with mobile phones/tablets. None of them attended Study 1. The equipment remained unchanged. The only difference was the new classifier integrated with the tabletop.

5.3 Design

We used a one-factor within-subject design. The independent factor was the classifier being enabled/disabled. The study consisted of two sessions: one session with our classifier (namely *with* session) and one without the classifier (namely *without* session). The order of the two sessions was counterbalanced.

In Study 2, participants were only asked to finish tasks on the digital surface in the sitting posture. As found in Study 1, this posture condition provoked users to trigger more accidental touches. Similar to Study 1, the tasks consisted of map navigation and photo categorization. In each session of Study 2, participants needed to finish a long map navigation task and a long photo categorization task with a balanced order. The operations in two tasks were the same as Study 1. A 5-minute warm-up stage was scheduled before each session for users to get familiar with the system. Both tasks were designed to take approximately 15 minutes and a 2-minute break was inserted between the two tasks. Participants needed to perform each task in the sitting posture for a long time, henceforth their arms would easily be affected by fatigue, which could make them more "willing" to rest their arms on the surface. These modifications could provide more "opportunities" for users to evaluate the performance of this new system and have a better sense to compare the systems in the two sessions.

We manipulated users' expectations of the tabletop in Study 2. In both sessions, the experiment was introduced to the participants as an inspecting study. Participants were told that the tabletop was expected to be intelligent so that it could recognize and filter any unintentional touches. However, the system was not yet perfect. They were asked to perform either carefully or casually as long as they were comfortable with the system, and report any touches they deemed "wrong" during the experiment, which can reflect users' subjective evaluation on system's performance in practice. The word "wrong" is used to describe those touches that ideally should have been removed (false positive) or should have appeared (false negative), but are noticed by users. Specifically, in the *with* session, the "wrong" touches represent those misclassified touches observed by participants (either intentional touches are classified as unintentional ones or vice versa). In the *without* session, the "wrong" touches mean the accidental touches that trigger responses in the system and are noticed by users. Whenever participants found anything wrong, they needed to speak up to the instructor immediately.

At the end of each session, users rated the system with a questionnaire. The questions were all rated on a 7-point Likert Scale (1: Not at all - 7: A lot), consisting of the following parts.

- One question for the subjective feeling of intuitiveness of the system during the operation
- Two questions for the efficiency and learnability of the system from Perceived Usefulness and Ease of Use Questionnaire [8]
- Five questions for task load during the experiment from the NASA Task Load Index (NASA-TLX) [16]
- Two questions for the system's capability of intention recognition and error prevention from Nielsen's Heuristic Evaluation [42] (only in the *with* session)

5.4 Procedure

In the study, participants first went through the adjustment and the calibration of the Pupillab, and finished two tasks in each session, with a 3-minute break in between. After completing each session, participants were asked to answer the questionnaire to evaluate the system. Each participant was offered $15 as compensation.

5.5 Result

We compared the numeric results between the best model and baseline models to validate the advantages of our method (Section 5.5.1). We observed an interesting phenomenon that users did not notice a number of misclassified touches (Section 5.5.2). We then compared the two sessions in terms of the number of "wrong" touches (Section 5.5.3) and users feedback (Section 5.5.4). The results revealed the advantage of our algorithm.

5.5.1 Validation of The New Algorithm. We first evaluated our best model with Study 2 data. Like Study 1, two authors manually annotated the data with the tool and solved conflicts with a collective review. Then, we applied our model and two baselines to the Study 2 data. Table 7 summarized the results. All models had similar performance as Table 4 and our model consistently outperformed the two baselines. In the real-time system, the model achieved an F-1 score of 0.906. Compared to the model modified from [52], our model still had an advantage of 4.0% on the F1 score (with statistical significance). This indicates that our algorithm is stable for new users in the real-time system.

5.5.2 Good Numeric Results According to User-report Data. Alternatively, user subject reports on the false positive/negative touch event also provided another perspective of the system performance. In the *with* session,

Table 7. Testing Results on Study 2 Data. A paired t-test on the cross-validation F1 scores between the full model and the palm rejection model shows significance $p < 0.01$.

Models	Prec	Rec	F1
Gaze-threshold-based	0.790 ± 0.016	0.788 ± 0.027	0.788 ± 0.015
Palm rejection [52]	0.867 ± 0.014	0.863 ± 0.034	0.866 ± 0.021
Full GB	**0.909 ± 0.022**	**0.904 ± 0.029**	**0.906 ± 0.025**

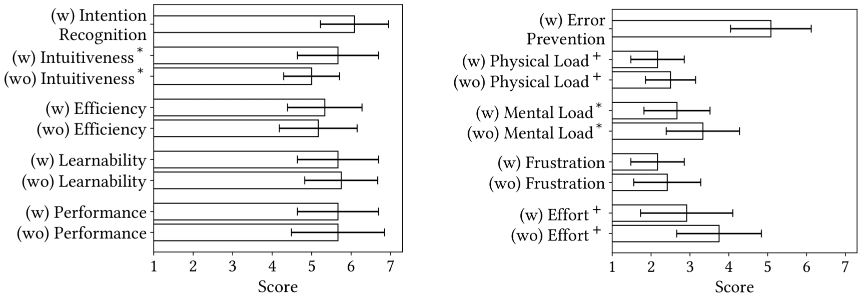


Fig. 12. Results of The Questionnaire. Error bar indicates the standard error. “w” represents the *with* session and “wo” represents the *without* session. $^+$ means marginally significance ($p < 0.1$) and * means significance ($p < 0.05$).

the average subjective F1 score among 12 participants achieved a surprisingly high value (0.956, 0.954, 0.955 for precision, recall and F1 score, respectively). Inspection of the data reveals that some misclassified touches were not perceived by participants. For instance, when a user operates with multiple fingers intentionally (moving or rotating), one finger's touch events may be filtered (false negative) but the operation can still continue normally. In another example, some unintentional touches are indeed classified as intentional (false positive), but the duration of the touches is too short to affect the normal operation. These gesture-related or task-related misclassified touches do not change the procedure of the on-going operation, and thus are difficult to be noticed by users. In other words, users only detect misclassification when superficial operations do not go as expected. Therefore, they will ignore lots of misclassification cases in the system.

5.5.3 Less “Wrong” Touches on The New Tabletop. We compared the percentage of the “wrong” touches in two sessions, *i.e.*, the perceived classification error rate in the *with* session and the perceived accidental touch rate in the *without* session. A Wilcoxon signed-rank test shows that the ratio of “wrong” touches in the *without* session is significantly higher than that of the *with* session (8.0% vs. 4.7%, $V = 3, p = 0.002$). This indicates that our new system can reduce the number of noticeable unintentional touches considerably.

5.5.4 Positive Feedbacks of The New Tabletop. We ran Wilcoxon signed-rank tests on the questionnaire results (see Figure 12). The intuitiveness ratings in the *with* session were significantly higher than the scores in the *without* session ($V = 37.0, p < 0.05$). And the mental load scores in the *with* sessions were significantly lower than those in the *without* session ($V = 9.0, p < 0.05$). In addition, the comparison of physical load and effort in two sessions showed marginal significance. Participants had slightly lower physical load and effort in the *with* session (for physical load, $V = 9.0, p = 0.09 < 0.1$, for effort, $V = 15.5, p = 0.06 < 0.1$). However, the ratings of the system's efficiency and learnability, as well as the performance and frustration in the two sessions did not show any significance. Participants gave positive feedback on the two metrics of intention recognition and error prevention for the *with* session. Most participants agreed that the system could correctly classify their intention (with the average score as 6.1 ± 0.9). They also rated 5.1 ± 1.1 scores on average for evaluating the system's ability for error prevention. These results show that the purpose of the new system was recognized by the participants. Our new system organically integrates the advantages of digital and physical tabletops.

6 DISCUSSION

6.1 Generalizability of the Model for Other Scenarios

In this section, we discuss how the feature sets and classification models can be generalized. We first summarize the similarities and differences between the features of digital tabletops and mobile devices. We then discuss the relationships between our model based on the capacitive technique and other digital table technologies, and show the potential of generalizing our model to other technologies. We also discuss how our single-user scenario can provide insights for multi-user scenarios.

6.1.1 Similarities and Differences of Features on Mobile Devices. We investigated five spatio-temporal feature types for unintentional touches classification on multi-touch tabletops (see Table 2). Our results support the importance of the *Gaze* for intention classification on a large-scale touchscreen. Comparatively, *Face* does not have such a great performance.

Other features have been discussed in previous literature on mobile phones and tablets. Our findings suggest that those features share similarities between tabletops and mobile devices. For example, unintentional touches are more prone to cluster together (*Clustering* [2, 52]). They appear more often near the edge of the device, in spite of the size of the touch screen (*Side* [32]). The distribution of the unintentional touch duration has a long tail [32, 36]. These similarities suggest the potential to transfer our method to mobile devices.

6.1.2 Features with Other Digital Table Technology. Our tabletop used capacitive sensing technology to register touch events. However, the features used in our model are actually independent of the sensing technique. Features in Table 2 only require positional and temporal information of touch events. *E.g.*, *Gaze distance* can be calculated once the positions of both touch points and gaze points are obtained, *History* can be calculated once the timestamps of touch points are recorded. As long as a system's technology can provide accurate information, it will be compatible with our model. Therefore, our system has the potential to be generalized to tabletops with other technologies such as FTIR [15], diffused illumination [50], etc. However, we point out that some technical issues related to the sensing accuracy of the positional and temporal information are beyond the scope of our paper, *e.g.*, the noise effect of environment lightness on FTIR/DI desktops.

6.1.3 Multi-user Scenarios. In this paper, we only investigated single-user scenarios. A multi-user scenario is another major use case of large-scale interactive surfaces. If different users' interaction areas have minute overlap, each user's interactions can be treated as a single-user case and our model can be generalized easily. The major differences between multi-user and single-user scenarios lie in two aspects [35]: 1) users can have close collaboration, where their touch points are mixed together. 2) users can have human-human interaction, where their attention is drawn away from the surface. These greatly increase the complexity of the system. For the first aspect, there are some works trying to distinguish users when multiple users are interacting with a tabletop (*e.g.*, [10]). This may be one of the promising methods to leverage our model, *i.e.*, dividing a multi-user scenario into several single-user cases, so that our model can be applied. As for the second aspect, the intuition of "where we look is where we have interests" holds in social interaction as well [22]. Therefore, gaze can still be an important feature showing users' intention for interaction. We expect more exploration in the future work.

6.2 Awareness of Unintentional Touch

There are actually two types of unintentional touches during the interaction on the tabletop. Some touches are *unconscious*. Users do not realize that they triggered an event. If the system does not respond to such contact, users will not notice it. For example, users will only notice the unintentional touches triggered by the thenar (the bases of the thumb) when they see accidental reactions happening around it.

On the other hand, some unintentional touches are *inevitable*. Users "have to" do some specific operations during the interaction to achieve their goal, even having the expectation of the consequence that it will lead to accidental touches. For example, users "have to" put the palms on the surface to support their body while reaching out to distant targets, or they "have to" rest their arms because of fatigue.

The mental model essentially works only on the *inevitable* touches but not those *unconscious* touches, since the model works only when users are conscious of their touches, either explicitly or implicitly, as reflected in our Study 2. However, designers and developers should be careful about *unconscious* touches. If they are not filtered, it will have a more significant effect on user experience than the unfiltered *inevitable* touches, since users do not expect any responses from these *unconscious* touches.

6.3 Limitations and Future Work

First, in this research, we explore user touch behavior in only two tasks: map navigation and photo categorization. Although these tasks are common on multitouch tabletops, there is other unintentional touch behavior we did not investigate which may appear in other interactive scenarios such as drawing sketches. Task-specific patterns may also affect the filtering, which will make the classification more complicated. Studies on more tasks will be involved in future work. Our technique leverages the coordination between touch and gaze. There are some cases where touch input is less guided by visual attention, such as blind typing. The technique might also conflict with other interaction paradigms such as remote object manipulation or peripheral-vision-based interaction [47], as summarized in Section 2.3. In these tasks, the benefit of our method is diminished. In addition, although with a limited number of users we found a consistent result (see the 2nd and 4th box in Fig 6), 12 people is a relatively small number for a two-factor within-subject design. The manually labeling of intentional and unintentional touches can introduce potential bias. In the future, the number of participants needs to be increased to incorporate more diverse behavior patterns across different users. Better palm-rejection models [52] and more sophisticated deep learning (*e.g.*, LSTM) methods are worth exploring.

Second, due to the hardware properties of the tabletop, the screen does not have the ability to measure the contact area and pressure level of any touch events [10, 32, 52, 70]. Although the touch area is partially reflected by the *Clustering* features, the lack of these features limits the performance of our system. For example, the accuracy of our system is not as good as previous work on tablets which leveraged contact size features (91.3% vs. 97.6% [52]). Regardless of the different interaction paradigms on tablets [52] and tabletops, it reflects how the contact size feature can improve recognition. Meanwhile, touch pressure can also reflect interaction intention to some extent [61]. These pieces of literature indicate the potential of enhancing our model by including these features. There are some commercial multi-touch tabletops that can provide contact images (*e.g.*, Microsoft PixelSense [38]) or pressure distribution (in the foreseeable future) with fairly high resolution, which needs further exploration. However, nowadays there are a great number of tabletops that cannot provide area or pressure information, our feature sets can be applied for those tables to support intention recognition.

Third, to obtain accurate head orientation and gaze direction, a head-mounted eye tracker is still necessary, which hinders the direct implementation of our model into an ordinary digital surface. This reduces the system's viability and scalability. Some previous researches have worked on head/gaze direction estimation directly from a camera [41, 48]. It is enticing to obviate the necessity of additional devices. Our work provides an example of improving user understanding by leveraging other aspects of user behavior [66, 67, 73]. More behaviors beyond gaze and face are worth exploration in the future.

7 CONCLUSION

In this paper, we investigated the face-gaze-touch coordination patterns of touches on large-scale multi-touch tabletops. In the first user study, we collected empirical behavioral data from 12 participants' performing usual daily tasks on both digital and physical tabletops. Inspection of the data indicates a significant difference between the two types of tabletop: users tend to be more careful and prudent when interacting with the digital tabletop than the physical table. We then extracted five types of spatio-temporal features to classify real-time unintentional touches. A Gradient Boosting model was trained on the data and achieved 0.914, 0.913 and 0.913 on precision, recall and F1 score respectively. It significantly outperformed the start-of-the-art model by 4.3% on the F1 score. The model reveals the importance of gaze features for unintentional touch recognition. Our second user study evaluated our model in a real-time system. The results validated the advantage of our system over baselines. Participants' positive subjective feedback indicates the effectiveness of our algorithm. provide a user friendly experience by reducing the cognitive barrier preventing users from interacting naturally with the table. This work sheds light on the possibility of future tabletop technology to improve the understanding of users' input intention by linking their gaze and head behavior to their touch behavior.

ACKNOWLEDGMENTS

This work is supported by the National Key Research and Development Plan under Grant No. 2016YFB1001402, the Natural Science Foundation of China under Grant No. 61521002, No. 61672314, and also by Beijing Key Lab of Networked Multimedia. We thank Tianheng Wang for helping code development and Gero Bergk for the proof-reading.

REFERENCES

[1] Michelle Annett, Fraser Anderson, Walter F. Bischof, and Anoop Gupta. 2014. The Pen is Mightier: Understanding Stylus Behaviour While Inking on Tablets. In *Proceedings of Graphics Interface 2014 (GI '14)*. Canadian Information Processing Society, Toronto, Ont., Canada, Canada, 193–200. http://dl.acm.org/citation.cfm?id=2619648.2619680

[2] Michelle Annett, Anoop Gupta, and Walter F. Bischof. 2014. Exploring and Understanding Unintended Touch During Direct Pen Interaction. *ACM Trans. Comput.-Hum. Interact.* 21, 5, Article 28 (Nov. 2014), 39 pages. https://doi.org/10.1145/2674915

[3] Florian Block, James Hammerman, Michael Horn, Amy Spiegel, Jonathan Christiansen, Brenda Phillips, Judy Diamond, E. Margaret Evans, and Chia Shen. 2015. Fluid Grouping: Quantifying Group Engagement Around Interactive Tabletop Exhibits in the Wild. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 867–876. https://doi.org/10.1145/2702123.2702231

[4] Christophe Bortolaso, Matthew Oskamp, Greg Phillips, Carl Gutwin, and T.C. Nicholas Graham. 2014. The Effect of View Techniques on Collaboration and Awareness in Tabletop Map-Based Tasks. In *Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces (ITS '14)*. ACM, New York, NY, USA, 79–88. https://doi.org/10.1145/2669485.2669504

[5] Mark Brown, Winyu Chinthammit, and Paddy Nixon. 2014. A Comparison of User Preferences for Tangible Objects vs Touch Buttons with a Map-based Tabletop Application. In *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: The Future of Design (OzCHI '14)*. ACM, New York, NY, USA, 212–215. https://doi.org/10.1145/2686612.2686645

[6] Y.-L. Betty Chang, Stacey D. Scott, and Mark Hancock. 2014. Supporting Situation Awareness in Collaborative Tabletop Systems with Automation. In *Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces (ITS '14)*. ACM, New York, NY, USA, 185–194. https://doi.org/10.1145/2669485.2669496

[7] Manoranjan Dash and Huan Liu. 1997. Feature selection for classification. *Intelligent data analysis* 1, 1-4 (1997), 131–156.

[8] Fred D Davis. 1989. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly* (1989), 319–340.

[9] Tulio de Souza Alcantara, Jennifer Ferreira, and Frank Maurer. 2013. Interactive Prototyping of Tabletop and Surface Applications. In *Proceedings of the 5th ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS '13)*. ACM, New York, NY, USA, 229–238. https://doi.org/10.1145/2494603.2480313

[10] Abigail C. Evans, Katie Davis, James Fogarty, and Jacob O. Wobbrock. 2017. Group Touch: Distinguishing Tabletop Users in Group Settings via Statistical Modeling of Touch Pairs. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 35–47. https://doi.org/10.1145/3025453.3025793

[11] Jerome H Friedman. 2001. Greedy function approximation: a gradient boosting machine. *Annals of statistics* (2001), 1189–1232.

[12] Matt W Gardner and SR Dorling. 1998. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric environment* 32, 14 (1998), 2627–2636.

[13] Jens Gerken, Hans-Christian Jetter, Toni Schmidt, and Harald Reiterer. 2010. Can "Touch" Get Annoying?. In *ACM International Conference on Interactive Tabletops and Surfaces (ITS '10)*. ACM, New York, NY, USA, 257–258. https://doi.org/10.1145/1936652.1936704

[14] Jason Tyler Griffin. 2013. Touch screen palm input rejection. US Patent App. 13/469,354.

[15] Jefferson Y Han. 2005. Low-cost multi-touch sensing through frustrated total internal reflection. In *Proceedings of the 18th annual ACM symposium on User interface software and technology*. ACM, 115–118.

[16] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology* 52 (1988), 139–183.

[17] John M Henderson. 2003. Human gaze control during real-world scene perception. *Trends in cognitive sciences* 7, 11 (2003), 498–504.

[18] Mark A Hollands, Aftab E Patla, and Joan N Vickers. 2002. "Look where you're going!": gaze behaviour associated with maintaining and changing the direction of locomotion. *Experimental brain research* 143, 2 (2002), 221–230.

[19] Jeff Huang, Ryen White, and Georg Buscher. 2012. User See, User Point: Gaze and Cursor Alignment in Web Search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 1341–1350. https://doi.org/10.1145/2207676.2208591

[20] Robert J. K. Jacob. 1991. The Use of Eye Movements in Human-computer Interaction Techniques: What You Look at is What You Get. *ACM Trans. Inf. Syst.* 9, 2 (April 1991), 152–169. https://doi.org/10.1145/123078.128728

[21] Moritz Kassner, William Patera, and Andreas Bulling. 2014. Pupil: An Open Source Platform for Pervasive Eye Tracking and Mobile Gaze-based Interaction. In *Adjunct Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '14 Adjunct)*. ACM, New York, NY, USA, 1151–1160. https://doi.org/10.1145/2638728.2641695

[22] Adam Kendon. 1967. Some functions of gaze-direction in social interaction. *Acta psychologica* 26 (1967), 22–63.

[23] Ahmed Kharrufa, Madeline Balaam, Phil Heslop, David Leat, Paul Dolan, and Patrick Olivier. 2013. Tables in the Wild: Lessons Learned from a Large-scale Multi-tabletop Deployment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 1021–1030. https://doi.org/10.1145/2470654.2466130

[24] Seiya Koura, Shunsuke Suo, Asako Kimura, Fumihisa Shibata, and Hideyuki Tamura. 2012. Amazing Forearm As an Innovative Interaction Device and Data Storage on Tabletop Display. In *Proceedings of the 2012 ACM International Conference on Interactive Tabletops and Surfaces (ITS '12)*. ACM, New York, NY, USA, 383–386. https://doi.org/10.1145/2396636.2396706

[25] Yoshinori Kuno, Tomoyuki Ishiyama, Satoru Nakanishi, and Yoshiaki Shirai. 1999. Combining Observations of Intentional and Unintentional Behaviors for Human-computer Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '99)*. ACM, New York, NY, USA, 238–245. https://doi.org/10.1145/302979.303051

[26] Andreas Kunz, Ali Alavi, Jonas Landgren, Asim Evren Yantaç, PawełWoźniak, Zoltán Sárosi, and Morten Fjeld. 2013. Tangible Tabletops for Emergency Response: An Exploratory Study. In *Proceedings of the International Conference on Multimedia, Interaction, Design and Innovation (MIDI '13)*. ACM, New York, NY, USA, Article 10, 8 pages. https://doi.org/10.1145/2500342.2500352

[27] Ricardo Langner, John Brosz, Raimund Dachselt, and Sheelagh Carpendale. 2010. PhysicsBox: Playful Educational Tabletop Games. In *ACM International Conference on Interactive Tabletops and Surfaces (ITS '10)*. ACM, New York, NY, USA, 273–274. https://doi.org/10.1145/1936652.1936712

[28] Khanh-Duy Le, Mahsa Paknezhad, Paweł W. Woźniak, Maryam Azh, Gabrielė Kasparavičiūtė, Morten Fjeld, Shengdong Zhao, and Michael S. Brown. 2016. Towards Leaning Aware Interaction with Multitouch Tabletops. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction (NordiCHI '16)*. ACM, New York, NY, USA, Article 4, 4 pages. https://doi.org/10.1145/2971485.2971553

[29] Wen-Hung Liao. 2009. A Framework for Attention-based Personal Photo Manager. In *Proceedings of the 2009 IEEE International Conference on Systems, Man and Cybernetics (SMC'09)*. IEEE Press, Piscataway, NJ, USA, 2128–2132. http://dl.acm.org/citation.cfm?id=1732003.1732067

[30] Daniel J. Liebling and Susan T. Dumais. 2014. Gaze and Mouse Coordination in Everyday Work. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication (UbiComp '14 Adjunct)*. ACM, New York, NY, USA, 1141–1150. https://doi.org/10.1145/2638728.2641692

[31] Roman Lissermann, Jochen Huber, Martin Schmitz, Jürgen Steimle, and Max Mühlhäuser. 2014. Permulin: Mixed-focus Collaboration on Multi-view Tabletops. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 3191–3200. https://doi.org/10.1145/2556288.2557405

[32] Hao Lu and Yang Li. 2015. Gesture On: Enabling Always-On Touch Gestures for Fast Mobile Access from the Device Standby Mode. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 3355–3364. https://doi.org/10.1145/2702123.2702610

[33] Paul P. Maglio, Teenie Matlock, Christopher S. Campbell, Shumin Zhai, and Barton A. Smith. 2000. Gaze and Speech in Attentive User Interfaces. In *Proceedings of the Third International Conference on Advances in Multimodal Interfaces (ICMI '00)*. Springer-Verlag, London, UK, UK, 1–7. http://dl.acm.org/citation.cfm?id=645524.656806

[34] Alexander Mariakakis, Mayank Goel, Md Tanvir Islam Aumi, Shwetak N. Patel, and Jacob O. Wobbrock. 2015. SwitchBack: Using Focus and Saccade Tracking to Guide Users' Attention for Mobile Task Resumption. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 2953–2962. https://doi.org/10.1145/2702123.2702539

[35] Paul Marshall, Richard Morris, Yvonne Rogers, Stefan Kreitmayer, and Matt Davies. 2011. Rethinking 'Multi-user': An In-the-wild Study of How Groups Approach a Walk-up-and-use Tabletop Interface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 3033–3042. https://doi.org/10.1145/1978942.1979392

[36] Juha Matero and Ashley Colley. 2012. Identifying Unintentional Touches on Handheld Touch Screen Devices. In *Proceedings of the Designing Interactive Systems Conference (DIS '12)*. ACM, New York, NY, USA, 506–509. https://doi.org/10.1145/2317956.2318031

[37] Fabrice Matulic and Moira Norrie. 2012. Empirical Evaluation of Uni- and Bimodal Pen and Touch Interaction Properties on Digital Tabletops. In *Proceedings of the 2012 ACM International Conference on Interactive Tabletops and Surfaces (ITS '12)*. ACM, New York, NY, USA, 143–152. https://doi.org/10.1145/2396636.2396659

[38] Fabrice Matulic, Daniel Vogel, and Raimund Dachselt. 2017. Hand Contact Shape Recognition for Posture-Based Tabletop Widgets and Interaction. In *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces (ISS '17)*. ACM, New York, NY, USA, 3–11. https://doi.org/10.1145/3132272.3134126

[39] Michael Mauderer and Florian Daiber. 2013. Combining Touch and Gaze for Distant Selection in a Tabletop Setting. (2013).

[40] Pranav Mistry, Pattie Maes, and Liyan Chang. 2009. WUW - Wear Ur World: A Wearable Gestural Interface. In *CHI '09 Extended Abstracts on Human Factors in Computing Systems (CHI EA '09)*. ACM, New York, NY, USA, 4111–4116. https://doi.org/10.1145/1520340.1520626

[41] Erik Murphy-Chutorian and Mohan Manubhai Trivedi. 2009. Head pose estimation in computer vision: A survey. *IEEE transactions on pattern analysis and machine intelligence* 31, 4 (2009), 607–626.

[42] Jakob Nielsen. 1994. Usability inspection methods. In *Conference companion on Human factors in computing systems*. ACM, 413–414.

[43] Ken Pfeuffer, Jason Alexander, Ming Ki Chong, and Hans Gellersen. 2014. Gaze-touch: Combining Gaze with Multi-touch for Interaction on the Same Surface. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology (UIST '14)*. ACM, New York, NY, USA, 509–518. https://doi.org/10.1145/2642918.2647397

[44] Ken Pfeuffer, Jason Alexander, Ming Ki Chong, Yanxia Zhang, and Hans Gellersen. 2015. Gaze-Shifting: Direct-Indirect Input with Pen and Touch Modulated by Gaze. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software &#38; Technology (UIST '15)*. ACM, New York, NY, USA, 373–383. https://doi.org/10.1145/2807442.2807460

[45] Ken Pfeuffer and Hans Gellersen. 2016. Gaze and Touch Interaction on Tablets. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16)*. ACM, New York, NY, USA, 301–311. https://doi.org/10.1145/2984511.2984514

[46] Natural Point. 2011. Optitrack. *Natural Point, Inc.,[Online]. Available: http://www. naturalpoint. com/optitrack/.[Accessed 22 2 2014]* (2011).

[47] Argenis Ramirez Gomez and Hans Gellersen. 2019. SuperVision: Playing with Gaze Aversion and Peripheral Vision. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 473, 12 pages. https://doi.org/10.1145/3290605.3300703

[48] Neil Robertson and Ian Reid. 2006. Estimating gaze direction from low-resolution faces in video. *Computer Vision–ECCV 2006* (2006), 402–415.

[49] Hasibullah Sahibzada, Eva Hornecker, Florian Echtler, and Patrick Tobias Fischer. 2017. Designing Interactive Advertisements for Public Displays. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 1518–1529. https://doi.org/10.1145/3025453.3025531

[50] Johannes Schöning, Peter Brandl, Florian Daiber, Florian Echtler, Otmar Hilliges, Jonathan Hook, Markus Löchtefeld, Nima Motamedi, Laurence Muller, Patrick Olivier, et al. 2008. Multi-touch surfaces: A technical guide. *IEEE Tabletops and Interactive Surfaces* 2, 11 (2008).

[51] Julia Schwarz, Charles Claudius Marais, Tommer Leyvand, Scott E. Hudson, and Jennifer Mankoff. 2014. Combining Body Pose, Gaze, and Gesture to Determine Intention to Interact in Vision-based Interfaces. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 3443–3452. https://doi.org/10.1145/2556288.2556989

[52] Julia Schwarz, Robert Xiao, Jennifer Mankoff, Scott E. Hudson, and Chris Harrison. 2014. Probabilistic Palm Rejection Using Spatiotemporal Touch Features and Iterative Classification. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 2009–2012. https://doi.org/10.1145/2556288.2557056

[53] Barış Serim and Giulio Jacucci. 2019. Explicating &#34;Implicit Interaction&#34;: An Examination of the Concept and Challenges for Research. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 417, 16 pages. https://doi.org/10.1145/3290605.3300647

[54] Ludwig Sidenmark and Hans Gellersen. 2019. Eye&#38;Head: Synergetic Eye and Head Movement for Gaze Pointing and Selection. In *Proceedings of the 32Nd Annual ACM Symposium on User Interface Software and Technology (UIST '19)*. ACM, New York, NY, USA,

1161–1174. https://doi.org/10.1145/3332165.3347921

[55] Ludwig Sidenmark and Anders Lundström. 2019. Gaze Behaviour on Interacted Objects During Hand Interaction in Virtual Reality for Eye Tracking Calibration. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications (ETRA '19)*. ACM, New York, NY, USA, Article 6, 9 pages. https://doi.org/10.1145/3314111.3319815

[56] Misha Sra, Xuhai Xu, Aske Mottelson, and Pattie Maes. 2018. VMotion: Designing a Seamless Walking Experience in VR. In *Proceedings of the 2018 Designing Interactive Systems Conference (DIS '18)*. ACM, New York, NY, USA, 59–70. https://doi.org/10.1145/3196709.3196792

[57] Dave M Stampe. 1993. Heuristic filtering and reliable calibration methods for video-based pupil-tracking systems. *Behavior Research Methods, Instruments, & Computers* 25, 2 (1993), 137–142.

[58] Sophie Stellmach and Raimund Dachselt. 2013. Still Looking: Investigating Seamless Gaze-supported Selection, Positioning, and Manipulation of Distant Targets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 285–294. https://doi.org/10.1145/2470654.2470695

[59] Rainer Stiefelhagen, Michael Finke, Jie Yang, and Alex Waibel. 1999. From gaze to focus of attention. In *Visual Information and Information Systems*. Springer, 765–772.

[60] Lucia Terrenghi, David Kirk, Abigail Sellen, and Shahram Izadi. 2007. Affordances for Manipulation of Physical Versus Digital Media on Interactive Surfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. ACM, New York, NY, USA, 1157–1166. https://doi.org/10.1145/1240624.1240799

[61] Reed L Townsend, Alexander J Kolmykov-Zotov, Steven P Dodge, and Bryan D Scott. 2011. Unintentional touch rejection. US Patent 8,018,440.

[62] Jayson Turner, Jason Alexander, Andreas Bulling, and Hans Gellersen. 2015. Gaze+RST: Integrating Gaze and Multitouch for Remote Rotate-Scale-Translate Tasks. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 4179–4188. https://doi.org/10.1145/2702123.2702355

[63] Jayson Turner, Andreas Bulling, and Hans Gellersen. 2011. Combining Gaze with Manual Interaction to Extend Physical Reach. In *Proceedings of the 1st International Workshop on Pervasive Eye Tracking &#38; Mobile Eye-based Interaction (PETMEI '11)*. ACM, New York, NY, USA, 33–36. https://doi.org/10.1145/2029956.2029966

[64] Simon Voelker, Andrii Matviienko, Johannes Schöning, and Jan Borchers. 2015. Combining Direct and Indirect Touch Input for Interactive Workspaces Using Gaze Input. In *Proceedings of the 3rd ACM Symposium on Spatial User Interaction (SUI '15)*. ACM, New York, NY, USA, 79–88. https://doi.org/10.1145/2788940.2788949

[65] Daniel Vogel and Ravin Balakrishnan. 2010. Occlusion-aware Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, USA, 263–272. https://doi.org/10.1145/1753326.1753365

[66] Xuhai Xu, Ahmed Hassan Awadallah, Susan T. Dumais, Farheen Omar, Bogdan Popp, Robert Routhwaite, and Farnaz Jahanbakhsh. 2020. Understanding User Behavior For Document Recommendation. In *The World Wide Web Conference (WWW '20)*. Association for Computing Machinery, New York, NY, USA, 7. https://doi.org/10.1145/3366423.3380071

[67] Xuhai Xu, Prerna Chikersal, Afsaneh Doryab, Daniella K. Villalba, Janine M. Dutcher, Michael J. Tumminia, Tim Althoff, Sheldon Cohen, Kasey G. Creswell, J. David Creswell, and et al. 2019. Leveraging Routine Behavior and Contextually-Filtered Features for Depression Detection among College Students. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 3, Article Article 116 (Sept. 2019), 33 pages. https://doi.org/10.1145/3351274

[68] Xuhai Xu, Alexandru Dancu, Pattie Maes, and Suranga Nanayakkara. 2018. Hand Range Interface: Information Always at Hand with a Body-centric Mid-air Input Surface. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '18)*. ACM, New York, NY, USA, Article 5, 12 pages. https://doi.org/10.1145/3229434.3229449

[69] Xuhai Xu, Haitian Shi, Xin Yi, Wenjia Liu, Yukang Yan, Yuanchun Shi, Alex Mariakakis, Jennifer Mankoff, and Anind K. Dey. 2020. EarBuddy: Enabling On-Face Interaction via Wireless Earbuds. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 14. https://doi.org/10.1145/3313831.3376836

[70] Xuhai Xu, Chun Yu, Anind K. Dey, and Jennifer Mankoff. 2019. Clench Interface: Novel Biting Input Techniques. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 275, 12 pages. https://doi.org/10.1145/3290605.3300505

[71] Yukang Yan, Yingtian Shi, Chun Yu, and Yuanchun Shi. 2020. HeadCross: Exploring Head-Based Crossing Selection on Head-Mounted Displays. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 1 (March 2020), 22. https://doi.org/10.1145/3380983

[72] Yukang Yan, Chun Yu, Xiaojuan Ma, Xin Yi, Ke Sun, and Yuanchun Shi. 2018. VirtualGrasp: Leveraging Experience of Interacting with Physical Objects to Facilitate Digital Object Retrieval. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 78, 13 pages. https://doi.org/10.1145/3173574.3173652

[73] Yukang Yan, Chun Yu, Wengrui Zheng, Ruining Tang, Xuhai Xu, and Yuanchun Shi. 2020. FrownOnError: Interrupting Responses from Smart Speakers by Facial Expressions. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 14. https://doi.org/10.1145/3313831.3376810

[74] Alfred L Yarbus. 1967. *Eye movements during perception of complex objects*. Springer.

[75] Yaning Luo Zejiang Liu. 2012. Nanometer touch film production method.

[76] Shumin Zhai, Carlos Morimoto, and Steven Ihde. 1999. Manual and Gaze Input Cascaded (MAGIC) Pointing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '99)*. ACM, New York, NY, USA, 246–253. https://doi.org/10.1145/302979.303053

[77] Yang Zhang, Michel Pahud, Christian Holz, Haijun Xia, Gierad Laput, Michael McGuffin, Xiao Tu, Andrew Mittereder, Fei Su, William Buxton, and Ken Hinckley. 2019. Sensing Posture-Aware Pen+Touch Interaction on Tablets. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 55, 14 pages. https://doi.org/10.1145/3290605.3300285