Understanding and Adapting Bezel-to-Bezel Interactions for Circular Smartwatches in Mobile and Encumbered Scenarios

BRADLEY REY*, University of British Columbia, Canada
KENING ZHU†, School of Creative Media, City University of Hong Kong, China
SIMON TANGI PERRAULT, Singapore University of Technology and Design, Singapore
SANDRA BARDOT*
ALI NESHATI*, University of Waterloo, Canada
POURANG IRANI*, University of British Columbia, Canada

Supporting eyes-free interaction, mobility and encumbrance, while providing a broad set of commands on a smartwatch display is a difficult, yet important, task. Bezel-to-bezel (B2B) gestures are valuable for rapid command invocation during eyes-free operation, however we lack knowledge regarding B2B interactions on circular devices during common usage scenarios. We aim to improve our understanding of the dynamics of B2B interactions in these scenarios by conducting two studies and a third analysis: First, we explore the performance of B2B in a seated position; second, we explore the effect of mobility and encumbrance on the B2B interaction; finally, we improve on the B2B accuracies by calculating features and utilizing machine learning. With the limited interaction capabilities on smartwatches and the importance of the scenario of use, we conclude with applications and design guidelines for improved utilization of B2B that enables effective smartwatch control while in common, mobile and eyes-free scenarios.

CCS Concepts: • **Human-centered computing** → **Gestural input**; **Mobile devices**.

Additional Key Words and Phrases: Bezel Gestures; Smartwatch Input; Eyes Free Input; Mobility; Encumbrance

ACM Reference Format:
Bradley Rey, Kening Zhu, Simon T. Perrault, Sandra Bardot, Ali Neshati, and Pourang Irani. 2022. Understanding and Adapting Bezel-to-Bezel Interactions for Circular Smartwatches in Mobile and Encumbered Scenarios. *Proc. ACM Hum.-Comput. Interact.* 6, MHCI, Article 201 (September 2022), 28 pages. https://doi.org/10.1145/3546736

1 INTRODUCTION

With the constant flow of information and connectivity, and their always-available nature, the use of smartwatches is often quick, while mobile, and/or while encumbered [39, 53, 65]. These usage scenarios include walking, during transit or physical activity, while working with one's hands, and/or while communicating with others. In these scenarios, users often must be aware of their core task promoting the need for rapid, and at times, eyes-free interaction. However, while

*Authors noted conducted this work while affiliated with the University of Manitoba.
†Kening Zhu is also affiliated with City University of Hong Kong Shenzhen Research Institute, Shenzhen, P.R. China.

Authors' addresses: Bradley Rey, bradley.rey@ubc.ca, University of British Columbia, Kelowna, British Columbia, Canada; Kening Zhu, keninzhu@cityu.edu.hk, School of Creative Media, City University of Hong Kong, Hong Kong, China; Simon T. Perrault, perrault.simon@gmail.com, Singapore University of Technology and Design, Singapore; Sandra Bardot, ; Ali Neshati, aneshati@uwaterloo.ca, University of Waterloo, Canada; Pourang Irani, pourang.irani@ubc.ca, University of British Columbia, Kelowna, British Columbia, Canada.

2573-0142/2022/9-ART201 $15.00
https://doi.org/10.1145/3546736

smartwatches have become a common and useful device throughout our day-to-day lives, the interactions they offer are often limited and can be affected by these usage scenarios.

Smartwatch bezels, as a means of interaction, provide a unique approach that may be suitable for mobile and/or encumbered use. Bezel interactions include, rotary movement [9], bezel taps [70], bezel initiated sequential tapping [59], bezel pressing [68], bezel gliding [44], and edge based interaction [1, 48]. Furthermore, bezel initiated swipe (BIS) [22, 66] and bezel-to-bezel (B2B) [28, 29] have shown to increase the input vocabulary on smartwatches, allowing for interactions that can be performed quickly, accurately, eyes-free, and without contradiction to current touch gestures [29]. Moreover, BIS and B2B allow for increased interaction with a smartwatch that can provide benefit through leveraging physical traits of the smartwatch, the short and unique gesture set, and do not require additional hardware [29, 66].

While benefits of BIS and B2B interactions are clear, existing work that focuses on these interactions has concentrated towards their use in static and non-encumbered conditions. Furthermore, B2B gestures have been primarily explored on square smartwatches, where edge landmarks are beneficial. For these reasons, this work aims to better understand circular smartwatch B2B interaction while mobile and encumbered. Thus, we focus on three research questions: **(RQ1)** Can B2B interactions be efficiently performed on circular smartwatches for which cornered and straight landmarks are non-existent? **(RQ2)** While mobility and encumbrance creates negative interaction effects on smart-devices, what is the specific impact for B2B interactions? **(RQ3)** Can we algorithmically adapt B2B gestures, resulting in an even more accurate interaction under mobile and encumbered usage scenarios?

To address our questions, we carry out two user studies. Our first study, with participants seated, provides a baseline to further explore the impact of the circular form factor and the effects of later adding mobility and encumbrance. We reveal accuracies of 97.2%, 87.5%, and 74.0% for the 4-, 6-, and 8-Bezel Segments conditions tested respectively. Our second study focuses on mobility and encumbered usage scenarios, where participants perform B2B gestures while walking with or without holding bags in both hands. Results show significantly decreased accuracies across our experimental conditions. However, the underlying characteristics of the gesture itself importantly remain the same. Thus, we utilize the data captured from the two studies to improve the accuracy of B2B interactions using machine learning. Here, we note that a Random Forest model provides the often optimal solution when using a feature set containing the user's touch down and up positions as well as a calculated angle of direction change within their gesture.

Our contributions are twofold: **(C1)** We conduct studies investigating eyes-free B2B interaction on circular smartwatches while seated, mobile, and mobile while encumbered. This provides knowledge of B2B interaction during increasingly common usage scenarios. **(C2)** Through exploration of the underlying components of B2B interactions, we provide a machine learning based approach for adapting and improving B2B interactions through captured and calculated features. Taken together, and through final discussion, our contributions allow us to further promote smartwatch B2B gestures in common usage scenarios for a range of smartwatches.

2 RELATED WORK

2.1 Smartwatch Touch Interaction

Many touch input methods used on smartwatches have been translated from that of smartphones. However, due to the much smaller screen size, adaptations and novel techniques have also been explored. Finger identification [17], distinguishing different touch areas for a single finger [24], multi-touch [32], and sequential tapping [51] all offer feasible interaction and a large set of commands. Specific applications for touch gestures include text-entry [11, 19, 35, 51, 55], command selection [15], list selection [23], and other interactive purposes [52, 61]. Specifically, Fruchard et al. [15] explored a gestural swipe technique that allowed for eyes-free directional interaction of up to 172 commands; achieving an accuracy of 95% after training and with the use of tactile cues around the watch. Notably, the edge (or bezel) of the watch was used by participants to locate the appropriate area of interaction within the screen.

Through these works, we note that touch interactions which are common and natural for users [16], should be considered when designing new interaction techniques. However, some of the interactions proposed have learning periods, longer interaction times, and/or use multiple fingers or taps. This is not ideal under mobility and encumbered scenarios where hands could be occupied. B2B gestures, a form of swipe gesturing, offer a viable alternative as they allow for rapid, simple, and a widely used invocation modality.

2.2 Bezel Interactions

2.2.1 Smartphone and Tablet Bezels. Understanding that screen content competes for visual and situational attention, Bragdon et al. [8] looked at analyzing bezel gestures and taps under various environmental and situational conditions. Bezel gestures offering rapid interaction, and mark-based gestures offering high accuracy, proved to not be significantly affected by the environment in which they were executed in. Furthermore, they found that bezel gestures were suitable for an eyes-free environment, with eye-gaze data showing that users looked at the smartphone 3.5% of the total interaction time while for soft buttons this increased to 98.8%. Jain and Balakrishnan [25] further proposed a text-entry application utilizing bezel gestures and a style of marking menu [31], while others have focused on reducing reachability issues through bezel-based shortcuts [10, 36, 59].

2.2.2 Smartwatch Bezels. Edge based interactions utilize the physical hardware of the smartwatch and focus on bringing interactivity off the touchscreen and onto the side of the device [30, 48]. Ahn et al. [1] further utilized this method in conjunction with the touchscreen for increased interaction capabilities. This gives users the ability to use multiple fingers for tasks involving navigation and selection. Focusing on the touchscreen, smartwatch bezels have been used for data visualization interaction [44], command invocation through bezel initiated swipe (BIS) [22, 66] or bezel to bezel (B2B) [28, 29] gestures, and as a means of filtering during list selection [54]. BIS and B2B have shown to provide many benefits such as their ability to expand the input vocabulary of a smartwatch in a familiar yet non-contradictory manner, as well as work in eyes-free conditions [29, 66]. Han et al. [22] explored using the corners of a square smartwatch to allow for a continuous BIS interaction and Wong et al. [66] explored the range of segments that could be used for BIS on circular smartwatches, noting that after 8 segments performance dropped significantly. As an extension of BIS, B2B has been explored on square form factors [28, 29] to allow invoking of menu items and shortcuts.

Bezel interactions have shown to provide a simple method of increasing the interaction space and ability to interact with a device under varying usage scenarios. While there is promise for the use of bezels on smartwatches in eyes-free conditions due to their accuracy, quick interaction time, and non-contradictory interaction, there are limitations to note. First, most of these studies use square smartwatches which may benefit from landmarks (i.e. the corners/edges of the watch) to be utilized. Second, to our knowledge the use of BIS and B2B in mobile and encumbered scenarios, scenarios which are common when using smartwatches, has yet to be explored. However, we also note the benefit of B2B over BIS in regards to smartwatch interaction. While BIS allows for rapid and eyes-free interaction for up to 8 segments [66], its expressivity is limited to 8 items. Using B2B with 8 segments, the expressivity expands to the squared value of the number of segments, e.g. with 8 segments, B2B allows for 64 possible commands.

2.3 Mobile and Encumbered Smart Device Usage

Mobile, or on-the-go, smart device usage has shown to hinder the performance of touch interactions due to the increase in user's cognitive load [58]. Thus, research has focused on understanding the decrease in interaction performance [7, 37], effects on mobility [7, 58], and effects on awareness of one's surroundings while using smart devices [38, 60]. Furthermore, on-the-go interactions may involve encumbered usage scenarios, such as holding a bag, which can further reduce interaction capabilities [46].

Due to the decrease in interaction performance, research has attempted to provide interaction techniques for smartphones that consider such negative effects [18, 27, 43]. Kane et al. [27] explored adapting the user interface that allowed for larger on-screen contents while walking. Moreover, Goel et al. [18] focused on a purely algorithmic implementation that adapted text-entry for a user in mobile scenarios. Finally, Negulescu et al. [43] noted that motion gestures allowed for less focus required on-screen when interacting compared with tapping. While larger on-screen content mitigates the negative effects of mobility, this is not overly feasible on small-screen devices. Thus, algorithmic and simple interactive changes afford the user interface the ability to be maintained throughout usage scenarios.

Smartwatches are often used in spans of under 5 seconds [4, 65], while on-the-go, and to glance at and interact with when accomplishing other tasks [53]. With the many use cases for smartwatches, context is of utmost importance and is known to greatly effect the interaction behaviour [6]. Due to the effects of mobility and encumbrance, Singh et al. [61] focused on creating a design space specifically for smartwatches that overcame the drawbacks of common smartphone touch interactions while mobile and encumbered. They note that *touch efforts* (number of fingers and repetitions used combined with required visual attention) must be reduced to allow for effective interaction while mobile and encumbered. In close relation to this work, Dobbelstein et al. [13] explored tapping, swiping, and wrist flicking across usage scenarios, finding that swiping was more efficient and accurate than the other interaction techniques.

While the aforementioned works focus on adapting interactions due to the decrease in performance while mobile and encumbered, many works that look at smartwatch interaction do not focus on mobile and encumbered conditions [12, 21, 29, 40, 49, 66, 71]. For this reason, this work explores a smartwatch interaction that can be performed efficiently, while eyes-free, and can accommodate mobility and encumbrance. B2B gestures may provide this possibility as they have a very low *touch effort* and show promise in prior research.

3 DEFINITIONS

Bezel Segments. We define the width of the bezel region as an eighth of the diameter of the entire touch screen [44, 66][1]. This allows for the use of typically untouched pixels on the touchscreen to be used in increasing the interaction space [56]. We then divided the bezel of the watch into different numbers of segments. For this we choose 4 (90° per segment), 6 (60° per segment), and 8 (45° per segment) segments, illustrated in Figure 1 (b). As accuracy was greatly impacted when the number of segments increased further than 8 [66], we limited our work to this. Lastly, in exploring orientation of the bezel segments we looked to mimic a compass as our justification for the 4- and 8-Segments conditions. To extend to the 6-Segments condition, we aimed to provide familiarity by having bezel segments encompass the top-most/bottom-most portions of the bezel similar to the 4- and 8-Segments conditions [66].

[1]Similarly found on Samsung's Galaxy Watch Active 2 - https://www.samsung.com/global/galaxy/galaxy-watch-active2/#galaxy-watch-active2

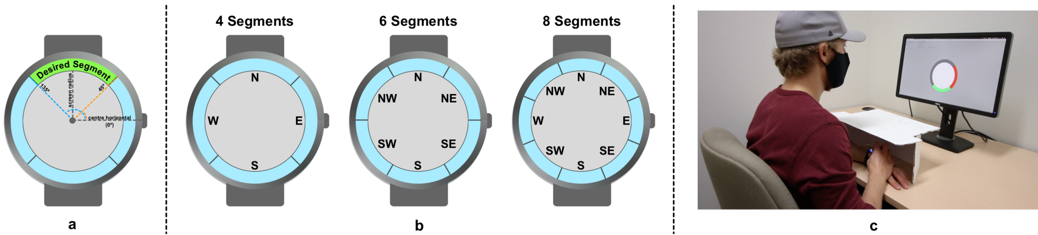

Fig. 1. In *a* we demonstrate how a correct interaction with a single segment occurs (the highlighted green area). First, the touch down (or up) must be within the minimum (orange/45°) and maximum (blue/135°) angles from the centre horizontal (grey/0°). Second, the interaction must occur outside the screen's new radius (radius of the grey circle), thus inside the bezel segment (light blue ring). In *b* we show the bezel segment conditions and their orientations used in our study. The blue portion represents the touch bezel used for the B2B gesture. Note, the bezel regions are not to scale and are enlarged for representation. In *c* we showcase the Study 1 apparatus. Note, the arm is placed parallel to the edge of the desk and the participant's field-of-vision towards the watch is occluded.

B2B Gestures. A B2B gesture is a swipe gesture that starts from a specific bezel segment and ends in the same or a different bezel segment [29]. For a B2B gesture to be correct, users need to perform a gesture with (1) a correct angle from the center, within both segment's angle ranges, and (2) the distance from the centre for both the start and end of the interaction must be greater than the radius of the remaining screen space, thus outside the area of the screen and inside the bezel (Figure 1 (a)). These conditions must be met for both the touch down (initial contact with the bezel, in the start segment), and the touch up (final point of contact, in the end segment).

4 STUDY 1: EYES-FREE B2B INTERACTION IN A STATIC CONTEXT

4.1 Participants

Twelve adults (3 females, 9 males), aged 21 to 39 (M = 26, SD = 4.6) participated. Participants were recruited from our university, all right-handed with the exception of one, and none were color blind; half of our participants (6) were daily smartwatch users.

4.2 Apparatus

The smartwatch used for the study was the Fossil Gen 5 The Carlyle HR[2] with a touch screen of 44 mm in diameter and a resolution of 416 × 416 pixels. This watch has no raised edge around the touch screen, thus has a flat surface from which to interact. During the study, participants sat at a desk where a cardboard covering occluded their vision of the smartwatch. Trials were displayed on a 23", 1080p, external monitor. The study apparatus can be seen in Figure 1.

4.3 Design

We used a within-subject study design with the *Number of Segments* as a main factor (4, 6, and 8); the number of possible B2B gestures was respectively 16, 36, and 64. Each gesture was repeated 3 times within a segment condition, and were randomized within blocks while ensuring no repeated trial was displayed back-to-back. We counterbalanced the segment conditions' order across participants using a Latin Square design. In total, we collected 12 (participants) × 116 (16 + 36 + 64 B2B combinations) × 3 (repetitions) = 4176 total trials (576 trials in the 4-Segments condition + 1296 trials in the 6-Segments condition + 2304 trials in the 8-Segments condition).

[2] https://www.fossil.com/en-us/products/gen-5-smartwatch-the-carlyle-hr-black-silicone/FTW4025.html

4.4 Procedure and Task

After introductions, participants were informed about the ethics and asked to sign a consent form, completed an Ishihara Color Blindness Test[3], and finally a short demographic survey. Before starting the study, the B2B interaction was explained and participants were allowed to practice B2B gestures, while wearing the watch and interacting on a blank screen. Participants were asked to wear the watch on their non-dominant hand's wrist while interacting with their dominant hand's index finger. When interacting with the smartwatch, participants were asked to place their watch wearing arm comfortably parallel to the edge of the table. This ensured that the watchface was oriented in the same manner as the trials being displayed; see Figure 1.

The task required participants to perform a B2B gesture between the bezel segments highlighted on the monitor by touching down in the start bezel segment and then sliding into the end bezel segment (indicated by the colors green and red respectively). For trials where the start and end bezel segments were the same, yellow was used. The participants were instructed to perform the B2B interaction by going through the non-bezel portion of the screen (i.e., starting in the start bezel area, sliding into the screen away from the bezel, and finally sliding into the end bezel area to finish the interaction). Previous works follow a similar approach [28, 29, 66], as keeping interactions around the edge of a round smartwatch has proven to be difficult [3]. Participants were instructed to perform the B2B interaction as fast and accurate as possible. Upon successful completion of a trial, the next trial would be displayed one second later. Otherwise, feedback would be displayed showing the error made. This feedback was given to mimic the real world effect of correctly or incorrectly executing a task. Lastly, participants were given optional breaks after every 12 trials.

4.5 Dependent Variables

We collected the following dependent variables: Accuracy, Time, Offset, and all start, end, and intermediate touch locations. Accuracy is simply whether a correct B2B gesture was performed. Time was further explored through two variables. First, we captured the amount of time taken by the participant from the time the stimuli appeared to the time they released their touch (Trial Time). Second, we isolated the physical touch interaction time from touch-down to touch-up (Movement Time). Offset was registered as the angle from the center horizontal line, as seen in a unit circle, for both the touch-down and -up locations. We use these to compute the offset between the measured angles and the expected angles (center of the respective segments), from which we derived a (signed/relative) offset. Finally, we recorded all touch points for the touch interaction in x,y coordinates. This was done through Android's event handlers, capturing the touch down, all touch move, and finally the touch up locations.

4.6 Statistical Analysis

We analysed our data using R. All statistically significant results are reported, while if not reported there was no significance found. For each dependent variable, we checked the distribution with a Shapiro-Wilks test. If normal, we proceeded in using a parametric analysis. Otherwise, we would transform the data and check for normality again. If after transformation the normality test failed, we would proceed with a non-parametric analysis.

For parametric analysis we used the ANOVA with repeated measures to detect main effects and interactions, and pairwise t-test with Benjamini-Hochberg corrections for post-hoc comparisons. We applied Greenhouse-Geisser sphericity correction when needed. The degrees of freedom may have decimal values after correction, similar to other work [5]. For non-parametric analysis we used Friedman's test to detect main effects and pairwise Wilcoxon tests with Benjamini-Hochberg

[3]https://www.color-blindness.com/ishihara-38-plates-cvd-test/

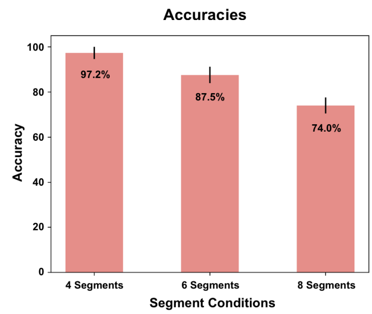

Fig. 2. Accuracy of each segment condition. Bars denote 95% confidence intervals. Statistically significant differences found between all segment conditions.

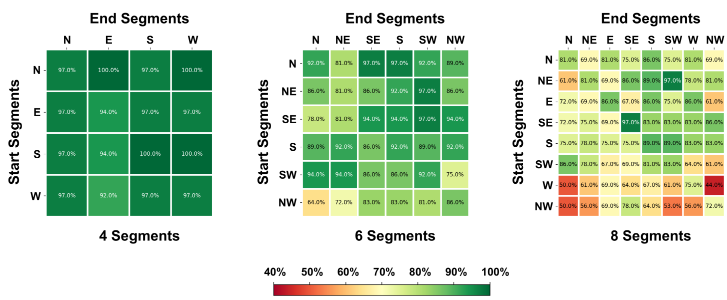

Fig. 3. Accuracy of each B2B combination. The x-axis represents the start bezel and the y-axis represents the end bezel. Thus, each square within the individual segment conditions is a specific B2B gesture combination of start and end segments.

corrections for post-hoc comparisons. We did not use Bonferroni correction, as its conservative approach leads to high rates of false negatives when done with large number of comparisons, which we have for the 6- and 8-Segments conditions [64]. Instead, we relied on Benjamini-Hochberg as it minimizes the problem [42] while still accounting for multiple comparisons.

5 STUDY 1 RESULTS

5.1 Accuracy

5.1.1 Across Segments Conditions. We found a significant main effect of the Number of Segments (4, 6 or 8) on accuracy ($\chi^2(2) = 24.0$, $p < .0001$). Pairwise comparisons showed significant differences between all segment conditions (all $p < .01$); participants were overall more accurate in the 4-Segments condition ($M = 97.2\%$) [$CI = 95.9\%, 98.6\%$] than in the 6-Segments ($M = 87.5\%$) [$CI = 85.7\%, 89.3\%$] and 8-Segments ($M = 74.0\%$) [$CI = 72.2\%, 75.8\%$] conditions. Figure 2 shows the overall accuracy for each segment condition and Figure 3 shows a complete breakdown of accuracy by segments for each start and end segment combination.

5.1.2 Within Segments Conditions. Further breaking down the B2B gestures by start and end segments revealed significant differences ($\chi^2(7) = 17.09$, $p = .01$) in accuracy in the 8-Segments condition between certain start segments. However, while there was a main effect, due to the number of conditions no significance was found after further analysis. We did not find any other effect of segment locations (start or end) on accuracy for any other condition.

5.2 Time

5.2.1 Total Trial Time. A one-way ANOVA showed a significant main effect of the number of segments on total time ($F_{2,22} = 30.84, p < .0001$). Pairwise comparisons showed significant differences between all conditions (all $p < .01$). The mean total time for 4-Segments was 1756ms [$CI = 1675ms, 1838ms$], for 6-Segments was 2323ms [$CI = 2184ms, 2463ms$], and for 8-Segments was 2711ms [$CI = 2607ms, 2815ms$].

5.2.2 Movement Time. We found a significant main effect of Number of Segments on movement time ($F_{2,22} = 4.67, p = .02$). Participants tended to be significantly slower in the 8-Segments condition with an average movement time of 711ms [$CI = 690ms, 732ms$], as compared to the 603ms [$CI = 557ms, 648ms$] for the 4-Segments conditions ($p < .05$), and 645ms [$CI = 622ms, 688ms$] for 6-Segments ($p < .05$).

5.2.3 Within Segments Conditions. Separating analysis by start and end segments revealed significant differences in the total time taken to perform B2B gestures in both the 4- and 8-Segments conditions. For the 4-Segments, we found a main effect of Segment Location on total time ($\chi^2(3) = 11.1$, $p = .01$). Participants were significantly slower starting from the South segment ($M = 1951$ ms) compared to North ($M = 1646$ ms, $p = .041$) and West ($M = 1665$ ms, $p = .041$) segments. We discovered a similar trend in the 8-Segments condition ($\chi^2(7) = 15.06$, $p = .03$), however we could not find significant differences between individual segment locations.

5.3 Touch Points and Paths

5.3.1 Start and End Touch Points. The touch points for the start and end of the interaction correlate with relative offsets recorded during the interaction, and can be seen with 95% confidence interval ellipses in Figure 4. The relative offset data for these touch points, compared with the centre of the segment, was broken apart by start and end segments across all segment conditions and was normally distributed. Since the segment size differs across the Number of Segments, the analysis of the offsets across this factor does not allow for a fair comparison.

Notably of the start and end touch points, there is no overlap for both start and end segments in the 4-Segments condition. We also note that the West and South clusters are shifted counter-clockwise during the start of the interaction. A significant main effect was found for the start segments in the 4-Segments condition ($F_{3,33} = 6.16, p < .01$). The West segment tended to have an increasingly larger offset ($M = -12.7°$) in the negative direction compared to the North ($M = 0.46°$, $p = .03$) and South ($M = -4.9°$, $p = .05$) segments.

In the 6-Segments condition, we observe a similar counter-clockwise shift for Segment North ($M = -4.0°$), North-East ($M = -9.0°$), South ($M = -4.14°$), South-West ($M = -14.4°$), and North-West ($M = -3.9°$) with little overlap between segments. Interestingly, Segment South-East had a positive mean relative offset ($M = 7.85°$) in the clockwise direction. A very similar shifting phenomenon can be observed with a larger spread of touch points, for the end segments.

Finally, the 8-Segments condition shows a significant amount of overlap, and a slight shifting of cluster centers towards the counter-clockwise direction. A significant main effect was found using a one-way ANOVA for only the start segments ($F_{7,77} = 4.23, p < .001$). We found significant differences between West ($M = -10.48°$), East ($M = -0.77°$, $p = .02$), and South-East ($M = -1.72°$, $p = .03$).

5.3.2 Touch Paths. In order to better understand the entire B2B gesture we visualized the touch paths of participants; see Figure 5 for selected touch paths. We note that, while not always, participants often looked to perform discrete changes in direction/angle. This direction change occurred throughout the centre of the watch. We calculated, with a 95% confidence interval, the *centre* used

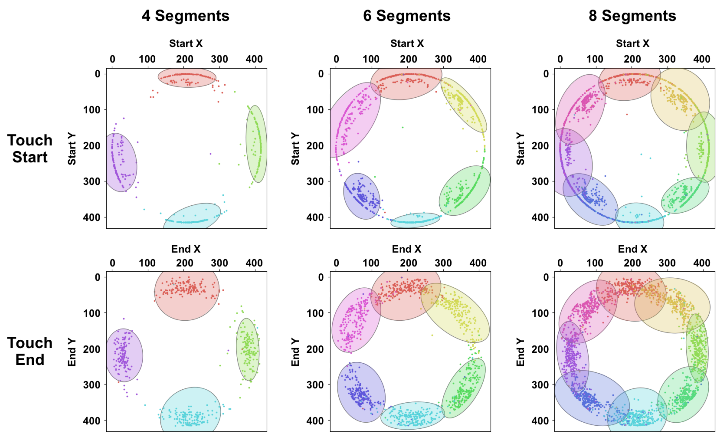

Fig. 4. Start (top row) and end (bottom row) touch points of the B2B gestures in Study 1. Ellipses represent 95% confidence intervals.

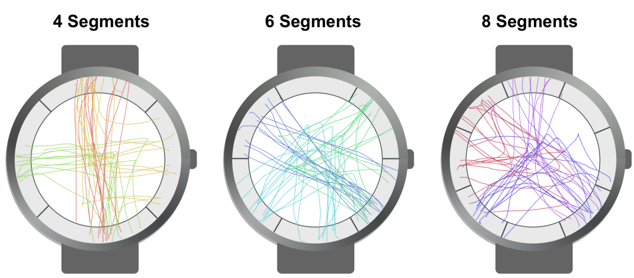

Fig. 5. Selected touch paths of 3 B2B combinations for each of the 3 Segment Conditions. Each line within a single color represents one trial from each participant. As can be seen, across participants, the B2B gestures often have a discrete angle change when moving from the start to end segment. As well, it can be seen that the angle change is often performed within a wide central radius.

for this change in angle across all trials to be 225 pixels wide and 220 pixels high. This suggests a wide area was used to move from start to end segments.

5.4 Summary of Results

5.4.1 B2B gestures are accurate. Accuracy across segment conditions followed an expected pattern, with a decreasing accuracy as we increase the number of segments. Compared with our 4-Segments condition, Kubo et al. found similar results, reporting accuracies of 96.3% in their sighted condition and 92% in the eyes-free condition [29]. This allows us to suggest that the round form factor does not hinder the ability for the a B2B gesture to be performed.

5.4.2 Locating the Start Segment was easier. Participants were seemingly more accurate in performing the start of the B2B interaction than they were in ending the interaction. The 95% confidence interval ellipses for the start segments are much more compressed than those of their end segment counterparts. This implies participants could often locate the start segment, however lose track of location throughout the gesture.

5.4.3 Participants tended to shift counter-clockwise. As observed from the touch point data in Figure 4, we noticed that participants tended to have trouble finding the center of segments, often interacting left of center. This is likely due to bio-mechanical constraints (i.e., due to the interacting hand being on one side of the watch, interaction on that side may be easier and more accessible than across the watchface) and suggest that the segments could potentially be redivided in a more meaningful way, by shifting them slightly in the counter-clockwise direction.

Furthermore, segments on the left and upper-left side of the watch, as well as segments directly next to each other, tended to perform less accurately with absolute offsets higher in these areas. The start and end touch points of the B2B interaction were much less compact on the left and upper-left side in the 6- and 8-Segments condition. This decrease in precision may again have been due to bio-mechanical constraints as previously discussed. With the interaction taking place in a non-natural resting position for the index finger, the accuracy of the interaction then decreases.

5.4.4 Acceptable interaction time. When analyzing time, the total interaction time occurred across all conditions in under 2.75 seconds; for the 4-, 6-, and 8-Segments conditions 1.756s, 2.323s, and 2.711s were respectively calculated. Of importance, is that these times fall well below 5 seconds of which ~50% of smartwatch interactions occur [65].

5.4.5 Shape of gestures use central reference point. To maintain reference, it can be seen that participants tended to attempt to find the centre of the watch face and to change their angle/direction towards the end segment from this reference point. While this strategy promotes the possibility of ending in the end segment, and the change in angle seems to be relatively correct, most participants gesture's changed angle away from the exact centre. This fact caused some interactions to end in the incorrect segment, thus making the B2B gesture incorrect.

6 STUDY 2: EYES-FREE B2B INTERACTION IN MOBILE AND ENCUMBERED SCENARIOS

The second study in this work aims to better understand B2B interactions, remaining eyes-free, while participants are mobile and encumbered. This allows us to evaluate the interaction in an increasingly common usage scenario.

6.1 Participants

For this study, 12 adults (6 females, 6 males) volunteered aged between 18 and 57 (M=30.6, SD=13.9) from and around a local university. All participants were new, thus none had partaken in Study 1. Participants were all right-handed, and 3 were daily smartwatch users. No participants were color blind.

6.2 Apparatus

We used the same smartwatch and external monitor as in Study 1. In addition, participants walked on a treadmill with a cardboard covering placed on their forearm to occlude their vision of the smartwatch. Furthermore, in the encumbered condition, 1.6 kg bags were used [45, 46, 61]. The study apparatus can be seen in Figure 6.

6.3 Design

A 3×2 within-subject study design was utilized, with the Number of Segments conditions (4, 6, and 8) and the Usage Scenarios (walking *WK* and walking while encumbered *WKE*) as the main factors. We keep the 8-Segments condition, despite its weaker performance in Study 1, as our goal is to further improve upon the accuracies in the latter part of this work. Each of the total possible B2B gestures were repeated 3 times within a segment condition, and were randomly displayed to

Fig. 6. Study 2 apparatus. Here the arm is held naturally while still occluded (highlighted by the blue squares). The image on the left demonstrates the walking condition, while the image on the right demonstrates the walking while encumbered condition (with the bags highlighted in red).

the participant while ensuring no repeated trial was displayed back-to-back. We counterbalanced the segment conditions and usage scenarios across participants using a Latin Square design. This ensured that no practice effects were seen across the conditions. Overall, with 12 (participants) × 116 (= 16 + 36 + 64 B2B combinations) × 3 (repetitions) × 2 (usage scenarios) = 8352 total trials (576 in the 4-Segments condition + 1296 in the 6-Segments condition + 2304 in the 8-Segments condition = 4176 trials per usage scenario × 2 = 8352 total trials).

To simulate mobility, participants walked on the treadmill at a speed of 2.5 km/h; this speed has been used to mimic average walking speeds in previous studies [26, 57]. For the encumbrance conditions we used two bags, one held in each hand and each weighing 1.6kgs; previous work explored this specific weight and usage scenario [45, 46, 61] and is noted as a common and worst case encumbrance situation, as both hands are busy.

6.4 Procedure and Task

A similar procedure was used from Study 1. Participants were allowed to practice the B2B gesture on a blank screen in a standing position to gain familiarity with the watchface. After this, the participants were brought to the treadmill to cover operation and safety procedures. The same task and instructions were given to participants as from Study 1.

7 STUDY 2 RESULTS

Similar statistical analysis techniques and methods were used from that of Study 1. Our analysis mainly focuses on the effect of the usage scenarios, both walking and walking while encumbered.

7.1 Accuracy

7.1.1 Usage Scenario. We found a significant main effect of Usage Scenarios on Accuracy ($F_{1,11}$ = 19.24, $p < .001$), as our participants performed better in the Walking Usage Scenario (WK) (M = 74.7%) [CI = 69.4%, 80.1%], compared to the Walking While Encumbered Usage Scenario (WKE) (M = 66.7%) [CI = 60.1%, 73.2%].

7.1.2 Number of Segments. We also found a significant main effect of Number of Segments on Accuracy ($F_{2,22}$ = 34.67, $p < .0001$). Pairwise comparisons showed significant differences between all three conditions, with a higher accuracy for 4-Segments (M = 89.4%) [CI = 86.3%, 92.5%], followed by 6-Segments (M = 79.1%) [CI = 73.1%, 83.1%] and 8-Segments (M = 61.9%) [CI = 56.4%, 67.3%] (all $p < .01$). We did not find any interaction effects (p = .06). The overall accuracy is shown in Figure 7 and the accuracy breakdown for each B2B combination can be seen in Figure 8.

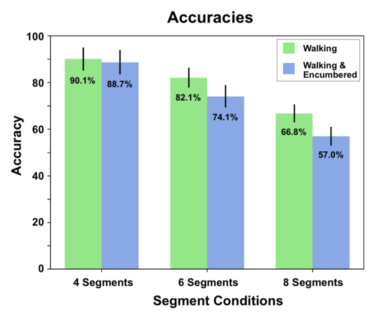

Fig. 7. Overall accuracy in Study 2 across segment conditions. Bars denote 95% confidence intervals.

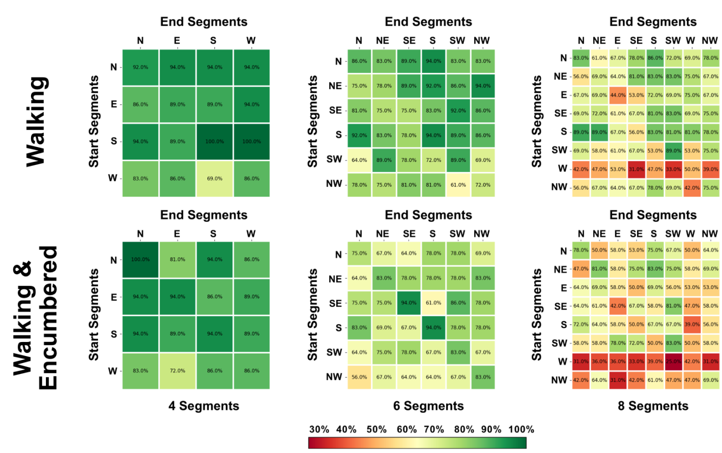

Fig. 8. Accuracy for all combinations of B2B gestures in Study 2.

7.2 Time

7.2.1 Total Trial Time. We did not find any significant effect of Usage Scenarios on Total Trial Time (p = .18). It took our participants 2093 ms on average $[CI = 1902ms, 2284ms]$ to perform a trial.

There was a significant main effect of Number of Segments on Total Trial Time ($F_{2,22} = 29.30$, $p < .0001$). Post-hoc comparisons showed significant differences between all three conditions, with participants performing faster in the 4-Segments condition ($M = 1758ms$) $[CI = 1591ms, 1923ms]$ than in the 6-Segments condition ($M = 1935ms$) $[CI = 1733ms, 2136ms]$ or in the 8-Segments condition ($M = 2350ms$) $[CI = 2120ms, 2580ms]$ (all $p < .01$). We did not find any interaction effects ($p = .10$).

7.2.2 Movement Time. We found no differences between WK and WKE conditions ($p = .97$), for an average performance of 651 ms $[CI = 596ms, 707ms]$ to perform a B2B gesture.

We found a main effect of Number of Segments on Movement Time ($F_{1.34,14.73} = 10.59$, $p < .01$). Overall, participants were significantly slower in the 8-Segments condition ($M = 697ms$) $[CI = 636ms, 758ms]$ compared to the 4-Segments condition ($M = 612ms$, $p < .01$) $[CI = 558ms, 667ms]$ and the 6-Segments condition ($M = 610ms$, $p < .01$) $[CI = 548ms, 672ms]$. We did not find any interaction effects ($p = .32$).

7.3 Touch Points and Paths

7.3.1 Start and End Touch Points. Figure 9 shows the start and end touch points of the B2B gestures across all segment and mobility conditions. It is again clear that the start of the interaction is more accurate than the end, and we note that comparing walking and walking while encumbered shows the encumbrance to have a large effect on the spread, and thus the overlap of the confidence ellipses. Interestingly the charts from the walking data compare relatively similarly with the data from Study 1. This suggests that walking alone has a minimal effect on the estimated location of the start and end segments.

We analyzed the relative offsets comparing our two usage scenarios as factors. This did not have any effect (p = .32) on the offset for the start segments, with an average offset of -3.54° $[CI = -5.90°, -1.18°]$. We did not observe any significant effect of Number of Segments (p = .06) nor any interaction effects (p = .47). We observed a similar trend for the end segments (p = .36), with an average offset of -3.31° $[CI = -5.82°, -0.80°]$, and no effect of Usage Scenario (p = .18), Number of Segments (p = .66) or interaction effects (p = .14).

7.3.2 Touch Paths. Participants again tended make a discrete angle change in the centre of the watch, very similar to those shown in Figure 5. The 95% confidence interval ellipse was slightly larger than in Study 1 at 246px wide and 240px high. No differences in this perceived centre, where the change in angle occurred, was found between the walking and walking while encumbered conditions. Interestingly, the centre of the confidence ellipse was skewed slightly right and low of the actual centre of the watch in both usage scenarios. This further reinforces that interacting across the watch face is more challenging in mobile and encumbered scenarios.

7.4 Summary of Results

7.4.1 Reduced Accuracy. Compared to Study 1, participants were understandably less accurate. Accuracy dropped from an average 86.3% in Study 1 to 79.6% in the WK condition and 73.3% in the WKE condition. The drop in accuracy is consistent across segment conditions, with the largest difference of 17 points calculated between the 8-Segments condition in Study 1 and the 8-Segments while walking and encumbered condition in Study 2 (74.0% vs. 57.0% respectively). We note, participants generally struggled with the 8-Segments condition, often taking breaks and commenting that the task was much too difficult. When comparing between walking and walking while encumbered, we find that accuracy only shows a significant difference in the 8-Segments condition. Accuracies hovered around the 80% mark for the 6-Segments condition and around the 90% mark for the 4-Segments condition. Thus, importantly towards our goal, these two segment conditions can accommodate potential usage across scenarios without hindering performance.

Furthermore, similar to the first study, the left and upper left side were significantly inaccurate in the 6- and 8-Segments conditions. While this was noticeable in Study 1, the results were exaggerated across Study 2. We speculate this may be due to the user's posture when interacting while walking, participants' arms were often aimed towards the ground, and the weight of the bag pulling down during the interaction when encumbered. As seen in Figure 8 the North-West segment in the 6-Segments condition and and West and North-West segments in the 8-Segments condition proved to show the worst accuracy when walking while encumbered. Interestingly, these segments performed at par with the others when they were the end segment of the B2B gesture. Furthermore, we highlight that B2B gestures where the segments are located physically next to each other tend to provide lower accuracy rates.

7.4.2 Similar interaction time across studies. Despite the reduced accuracy, we note that the total (1700 to 2700 ms) and movement (600 to 750 ms) times remained quick and relatively unchanged

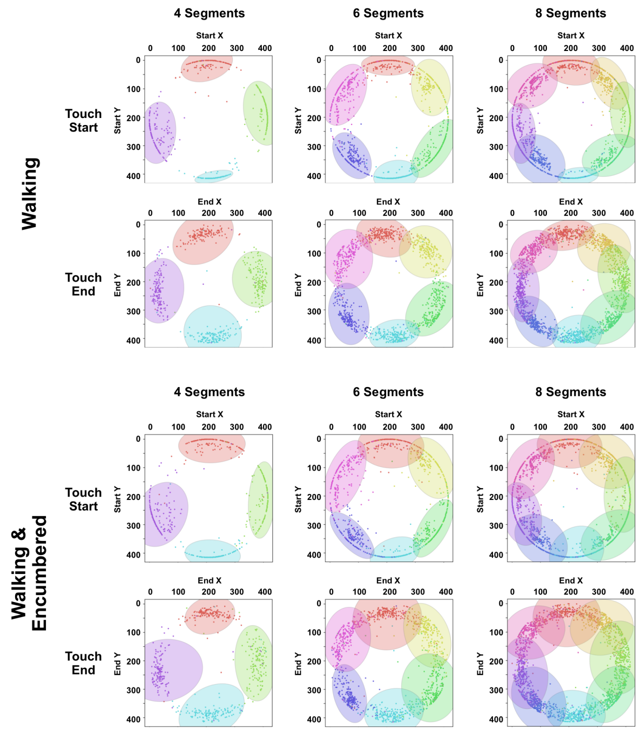

Fig. 9. Start and end touch points of the B2B gestures for all segment conditions and both usage scenarios. Ellipses represent 95% confidence intervals.

from Study 1. The time taken to perform the B2B gesture promotes the use of B2B gestures for rapid command selection and common interaction length on smartwatches, even under different usage scenarios.

7.4.3 Gesture similarities across studies. While Study 1 and Study 2 provided significantly different accuracy results, we note similarities in the interaction characteristics across our studies. These include start and end locations, their offsets to the centre of a segment, as well as touch path shape. Furthermore, the same counter-clockwise shift and lower accuracy on the left side are evident within the Study 2 results. This informs us that underlying gesture characteristics are not strongly affected by walking and walking while being encumbered.

8 OPTIMIZATION OF B2B-SWIPE DETECTION

We aimed to improve the B2B gestures captured over the course of Studies 1 and 2 using adaptive methods. First, we implemented an algorithm to calculate the angle of direction change in a B2B gesture. Second, as this empirical rule-based detection algorithm may miss hidden patterns, we trained a set of supervised-machine-learning classification models. We note, we chose not to utilize common gesture recognizers, such as the lightweight $ family, due to the need for nuanced extraction of similar gesture directions and orientations.

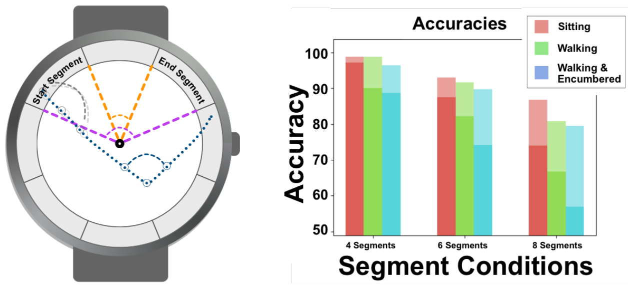

Fig. 10. (Left:) the blue dotted line represents a user's touch path, with significantly less touch points shown for demonstrative purposes. The first two iterations of the angle of change detection algorithm are highlighted from the start segment. The algorithm iterates through the touch path, running until no touch points remain, taking three touch points per iteration to calculate the angle between them. After iterating through all touch points, the smallest angle calculated is found to be the angle of change; here that angle is highlighted in blue. For demonstrating effectiveness, we can utilize the minimum (orange) and maximum (purple) angles to test against. In this example the gesture would be seen as correct, as the blue angle is within the orange and purple angles. (Right): the graph denotes the increase in accuracy (faded portion of the bar charts) for all conditions in Study 1 and 2 when angle optimization is utilized and tested against the minimum and maximum angles.

8.1 Angle Change Detection

We created an angle change detection algorithm that analyzes the points recorded from a B2B gesture and calculates the angle at which participants changed direction within the gesture. This takes into consideration our assumption that participants understood the angle between the segments, but potentially lost reference to the start and end segments during the eyes-free aspect of the interaction. Pseudo-code for the algorithm can be found in Appendix A while Figure 10 provides a rudimentary example of the algorithm at work.

To explore how often participants utilized a correct angle of change, we compared the angle of change calculated to the minimum and maximum allowed angles between the given start and end segments for a specific trial. If the value of the angle of change fell within the minimum and maximum allowed angles, the trial was considered a success. As seen in Figure 10, performance gain is obtained across the study scenarios, with an up to 19.5% increase in accuracy in the walking while encumbered 8-Segments condition. We note, that in-the-wild the angle of change could be calculated in real-time for each gesture; however, it can not be used as the sole model for accuracy as the desired start and end segments would be unknown. As this is the case, yet due to the large accuracy increase found, we aim to utilize this calculation within the machine learning models below.

8.2 Classifying B2B Swipes with Supervised Machine Learning

To utilize machine learning techniques, we first derived a set of 7 features for a B2B swipe gesture, including the gesture's start and end position (i.e. two pairs of x and y coordinates), start and end touch angles calculated from the centre horizontal, minimum angle offset values at the start and end position, and the angle of direction change during the gesture. Note, these features can be grouped into four categories: 1) positional touch information 2) angle information 3) offset information and 4) the maximum angle of direction change during the gesture. The combinations of these feature categories resulted in 15 possible feature sets as shown in Table 1 within the appendix.

In our user studies, each participant performed each B2B swipe for three repetitions, resulting in 12 participants × 3 repetitions = 36 samples for each bezel combination in each condition. As this sample size may not be sufficient for training the classification models, we adopted a data-augmentation process for the collected data. Pearson Correlation analysis on the selected features shows a weak correlation among the start and end features. As there is no obvious causal relation between start and end features, we assume that for each B2B gesture performed by participants, given the feature group of one particular start position, it is then possible to result in any end position and angle of direction change. Thus, for each Study 1 participant, we augmented their data with the combinations of 3 start features × 3 end features × 3 angle features = 27 data entries for each type of B2B swipe. This led to 12 participants × 27 augmented entries = 324 samples for each bezel condition in each condition.

For each combination of segment and encumbrance condition, we adopted the strategy of leave-one-out cross validation. That is, we trained, validated, and tested each classification model for 12 rounds, each round with one user's data as the testing set and the rest for training and validation, and calculated the average performance. In each round, the training-validation set was further randomly divided, 70% for training and 30% for validation. The classification models were trained and validated using the training-validation set, and then evaluated with the testing set.

For all models tested below, training and testing were performed on a MacBook Pro with a 2.3GHz quad-core 10th-generation Intel Core i7 processor, 16GB RAM, and 1TB SSD hard disk. We considered four commonly used models throughout, including Support-Vector Machine (SVM) [47], K-Nearest Neighbors (KNN) [14], Logistic Regression (LR) [67], and Random Forest (RF) [63]. As we were working with a small amount of feature sets, we compared the performance of each model with all 15 feature sets across each test.

8.2.1 Classification Experiment on Study 1 Data. As shown in Figure 11 (left), all the selected classification models improved the accuracy of the testing data sets for each layout. On average, the four models improved the accuracy on the testing sets by 1.8% (from 97.2% to 99.0%), 6.8% (from 87.5% to 94.34%), and 9.6% (from 74.0% to 83.6%) for the 4-Segments, the 6-Segments, and the 8-Segments conditions respectively. In addition, the inference time captured was within the range of real-time response (i.e., 100ms) [41].

8.2.2 Classification Experiment on Study 2 Data. With the classification models trained with the Study 1 data, we first experimented the generalizability of these models on the data recorded in Study 2. Figure 11 (right) summarizes the classification performance of the trained models on the Study 2 data.

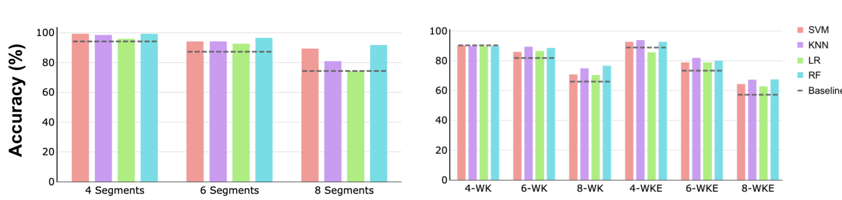

Fig. 11. (Left): Performance of B2B classification on Study 1 data. The baseline accuracy is denoted by the dark dashed line. Numbered results can be found in Table 2 of the appendix. (Right): Classification performance of the Study 1-data-trained models on all the Study 2 data. Numbered results can be found in Table 3 of the appendix. In both left and right, the baseline accuracy is denoted by the dark dashed line.

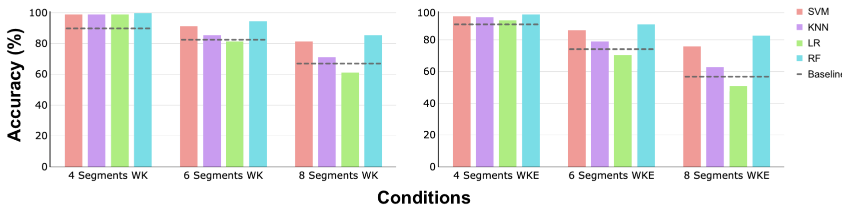

Fig. 12. Performance of B2B classification trained by the Study 2 data for the walking condition (left) and walking while encumbered condition (right). In both, the baseline is denoted by the dark dashed line. Numbered results can be found in Table 4 and Table 5 of the appendix.

Overall, the classification models improved the accuracy from their respective baselines. Among the three segment conditions and the two usage scenarios, the pre-trained SVM model achieved the highest accuracy of 90.28% in the 4-Segments condition while walking without encumbrance. KNN achieved the highest accuracy of 89.35% and 83.02% for the 6-Segments condition in walking without and with encumbrance respectively, and 93.75% for the 4-segments condition in walking with encumbrance. The pre-trained RF model achieved the highest accuracy for the rest of the segment conditions and usage scenarios (8-Segments while walking without encumbrance: 76.21%; 8-Segments while walking with encumbrance: 67.71%).

While the models trained with the Study 1 data showed a considerable improvement on the accuracy, we were also interested in how the models would perform when specifically trained with the data from Study 2. For the data recorded in Study 2, we adopted the same feature-sets and processes as that of the Study 1 trained model to train two new sets of classification models, one for each of the usage scenarios. Figure 12 shows the classification accuracy of the two sets of models trained with the Study 2 data.

8.3 Summary of Results

8.3.1 Features, Model, and Generalizability. The feature set which most often provided the greatest accuracy increase contained the starting and ending position of the gesture as well as the angle of direction change; beneficially this is a relatively small feature set. Furthermore, the majority of feature sets which provided high accuracy increase included the angle of direction change, highlighting the ability for a mid-gesture characteristic to benefit the overall accuracy.

While all models across our testing, except LR, provided respectable results, the RF model often provided the largest increase in accuracy and was consistently a top performing model. Importantly, the RF model clearly outperformed the models aiming to improve on the walking and walking while encumbered data.

An independent-sample T-test showed no significant difference between the classification accuracy of the models trained by the Study 1 data and the Study 2 data respectively, showcasing successful model use both within and across study data. As Study 1 and Study 2 involved two different groups of participants, this may indicate the generalizability of the classification models trained with the Study 1 data. Furthermore, this may imply that the data collected when performing B2B gestures while walking without and with encumbrance yielded similar behavioral patterns as the B2B gesture data during seated interaction.

8.3.2 Smartwatch Deployment. While we did not run the prediction models in real time, we note that research has begun to take advantage of greatly improved hardware to successfully train and run models on smartwatches [33, 34, 62]. For now, we deployed the trained RF model to an emulator with the same configuration as the watch used for the data collection. The inference time for one data entry, using our optimal feature set, took just over half a second. This time elementally indicates the real-time speed of the trained prediction model. As an important future work, we will deploy the prediction model to a range of smartwatches, and conduct real-life usability validation.

Furthermore, as performance varies among different users, personal models could outperform the general study models, as seen in prior works [66]. Thus, future work will also investigate the effectiveness of adaptive and continual learning models that can dynamically improve the model for each user over time, both gradually and automatically. This would require a much larger study, in both time from participants and data captured, and was thus not explored within this work. However, through the positive results obtained, we recognize this as an important and plausible step to consider for future real-world, in-situ, deployment.

9 DISCUSSION

As each study section above includes a discussion, here we consider overall design guidelines for B2B usage and their comparison to previous work, broader implications of eyes-free and B2B gestures, potential applications, and limitations and future directions for our work.

9.1 B2B Guidelines and Comparison to Prior Work

Our assumption was that due to a decreased amount of physical landmarks (i.e., the physical discrete edges of a square smartwatch), a circular form factor may not allow for as accurate of B2B gestures. When comparing with the work done by Kubo et al. [29], our 4-Segments eyes-free condition actually achieved greater accuracy than their sited (96.3%) and their eyes-free (92%) conditions. Thus, B2B gestures work equally as well for 4-Segments on circular form factors, and can therefore be used across the range of smartwatches available on the market today. While accuracy is not affected, we do note that movement time of the gesture almost doubled (from roughly 300 to 600 ms). This suggests that slightly more cognitive effort was needed on the circular form factor, yet B2B remains to be a rapid and accurate gesture.

As expected, mobile and mobile while encumbered conditions decreased the accuracy of the B2B gestures. Yet, while this is true, 4- and 6-Segments still remain feasible for use, and the interactions show little difference between the two usage scenarios tested within Study 2. As smartwatches have the potential for use across usage scenarios, these results show promise for a smartwatch interaction technique that allows for accurate and efficient interaction.

Finally, we aimed to explore the use of machine learning methods and calculated features to increase the accuracy, and thus the potential, of B2B gestures on smartwatches. We suggest to improve upon basic B2B gesturing through the calculated use of the B2B gesture's change in angle naturally performed by the user and a machine learning model. This can provide at times more than a 15% increase in accuracy. This was achievable due to the common characteristics seen from the B2B gestures across our studies.

9.2 Eyes-Free Interaction

Studying eyes-free interaction within this work, while not performed regularly throughout one's day, allows for two factors to be taken into consideration. First, smartwatches are often used in diversified short interactive bursts while on-the-go and during day-to-day activities [20, 53, 65]. Thus, a fundamental motivation for the ability to perform eyes-free interaction is to allow for a user to be visually uninhibited during the interaction, so as to remain focused on the scenario of use such as walking or talking with others [50]. Further motivation has been explored by Yi et al. [69], citing environmental, social, device features, and personal aspects as reasons for the need of eyes-free interaction. These situated usage scenarios and motivations must be considered when designing interactions, as is a motivation of this work. Second, through our studies and prior works exploring eyes-free interaction [29, 52, 66], we highlight the ability for B2B gestures to be used in eyes-free scenarios if needed, while understanding this may not always be the case. Furthermore, eyes-free conditions allow us to encapsulate a worst-case usage environment, thus reporting a lower bound on our results.

9.3 B2B Versus BIS

An advantageous aspect of B2B, beyond that of BIS, is the increased number of possible interactions within the same screen real estate. While the results of our work are promising for 4 segments and plausible for 6 segments, which are only slightly less accurate, and matched through machine learning, to the BIS work done by Wong et al. [66], this allows for 16 and 36 possible interactions compared with only 4 and 6 in BIS respectively. However, this raises the questions of memorability and need. In terms of memorability, Appert and Zhai [2] explored stroke gestures, arbitrary to the command being performed, and found that with 14 commands users could achieve ~80% recall accuracy after only 12 sessions. We believe that through the proper mapping of commands, users could achieve similar recall when utilizing B2B gestures. This result bodes well for the 4-Segments condition in full, however the 6-Segments condition offers a much increased set of possibilities. Here, we understand that the potentiality for users to need all 36 B2B commands is limited. However, while users may not utilize every command, an increased set of possibilities allows designers to create a range of options within their applications from which users can utilize a select few that regularly pertain to them. Ultimately, this expanded B2B gesture set allows designers to provide greater option and generalizability for all users.

9.4 Applications

9.4.1 Smartwatch Menu. B2B allows for applications that can benefit and provide users with extended options or easier access to many controls that may be otherwise hard to utilize. In this regard, we suggest using marking menu shortcuts that can improve smartwatch applications where reduced interaction is currently an issue or many controls are available. For everyday smartwatch users, applications that could benefit include messaging, IoT control, or music players. Messaging could benefit from the quick and rapid responses B2B makes available, see Figure 13 (a). Ample common phrases or responses could be made available to the user through B2B gestures, without the need for additional time spent looking at the smartwatch after reading the message.

9.4.2 Augmented and Virtual Reality. To extend on this, virtual reality (VR) and/or augmented reality (AR) applications with at times complex menuing systems could utilize a smartwatch as an external input device; see Figure 13 (b). VR applications would often be performed while encumbered, through use of the controllers, and while eyes-free, due to the headset. Throughout, additional and rapid menuing enabled by B2B interaction could be utilized across applications (e.g., gaming and paint/drawing applications) and even for native UI navigation. On the other hand, AR applications have future potential in being used throughout our daily lives where mid-air and controller interaction may not be suitable. Again, due to the nature of viewing what is being displayed and even your surroundings, the B2B interaction with the smartwatch can be seen as potentially eyes-free and allow for a range of interactive capabilities. We note that using the smartwatch as an input modality for these head worn displays provides an important distinction. The B2B commands no longer need to be memorized by the user, as the controls could be displayed within the virtual content while the interaction remains eyes-free. This allows for the use of the recommended 6-Segments and allows designers of VR/AR systems to provide ample controls to users as needed.

Fig. 13. In *a* (left) we demonstrate quick replies utilizing B2B and marking menu responses. In *b* (right) we demonstrate how B2B on a smartwatch could be used in AR. Note, the interaction could be eyes-free of the actual smartwatch while controls are displayed within the virtual content.

9.5 Limitations and Future Work

Throughout our work we highlight limitations. The second study's design utilized a treadmill to simulate the walking conditions. While this provides us a means of studying the effect of mobility, as is often done in HCI research [26, 44, 61], testing B2B interactions outdoors or in-the-wild may produce slightly different results. External factors and the need to better understand one's surroundings may influence the B2B gesture. Second, our sample size limited two factors. First, all participants except one were right handed. Thus, our results are focused only on participants wearing the watch on their left wrist and interacting with their right hand. We presume the results we have found could be mirrored along the centre y-axis, but this remains to be studied. Second, as our dataset was relatively small, data-augmentation techniques were needed to create large enough training and testing data sets for the machine learning models. While this technique is utilized throughout data analysis domains, true raw data can be seen as better.

Future work in this space will mainly focus in three areas. We believe that exploring the B2B gesture past just a start and end segment could allow for the possibility of a swipe text-entry or password pattern entry. Second, as we note that B2B provides extended input options over BIS we also understand that memorability of options may become an issue. As such, we aim to explore this aspect in future work. Lastly, with the knowledge from this work we aim to further explore the effect that arm and hand position has on the smartwatch interaction. These areas of research may further expand the feasibility of utilizing the bezel and machine learning models to their full potential on a smartwatch.

10 CONCLUSION

Due to the always-available nature of smartwatches, input and output can occur at any time across many usage scenarios. Thus, a simple means to perform commands that allow for precise manipulation of on-screen contents is required. Most importantly, this interaction must not interfere with the core task being accomplished. Through this work, we have shown that B2B gestures can do just this; allowing for simple, quick, accurate, and eyes-free interaction across mobility and encumbrance scenarios. This can be utilized on round smartwatches, granting the user the ability to rapidly perform accurate interactions to invoke commands. Through the results of this work, we suggest that designers focus on utilizing either 4 or 6 bezel segments. Furthermore, we also highlight that using a mid-gesture feature, the angle of change made during the B2B gesture, and machine learning models can further improve accuracy across our usage scenarios. Through these findings, end users can benefit from an increased smartwatch interaction space which allows for use across a range of typical usage scenarios.

ACKNOWLEDGMENTS

We would like to first thank all of our study participants. We also acknowledge funding support from the NSERC CREATE grant on Visual and Automated Disease Analytics as well as the National Natural Science Foundation of China (Project No. 61907037 62172346), the Guangdong Basic and Applied Basic Research Foundation (Project No. 2021A1515011893), and ACIM, School of Creative Media, City University of Hong Kong. Finally, we would like to acknowledge Pui Chung Wong for sharing his code base from which we were able to adapt for our study.

REFERENCES

[1] Sunggeun Ahn, Jaeyeon Lee, Keunwoo Park, and Geehyuk Lee. 2018. Evaluation of edge-based interaction on a square smartwatch. *International Journal of Human-Computer Studies* 109 (2018), 68–78. https://doi.org/10.1016/j.ijhcs.2017.08.004

[2] Caroline Appert and Shumin Zhai. 2009. Using strokes as command shortcuts: cognitive benefits and toolkit support. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2289–2298. https://doi.org/10.1145/1518701.1519052

[3] Daniel Ashbrook, Kent Lyons, and Thad Starner. 2008. An investigation into round touchscreen wristwatch interaction. In *Proceedings of the 10th international conference on Human computer interaction with mobile devices and services*. 311–314. https://doi.org/10.1145/1409240.1409276

[4] Daniel Lee Ashbrook. 2010. *Enabling mobile microinteractions*. Ph.D. Dissertation. Georgia Institute of Technology.

[5] Marc Baloup, Thomas Pietrzak, and Géry Casiez. 2019. RayCursor: A 3D Pointing Facilitation Technique Based on Raycasting. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300331

[6] Leon Barnard, Ji Soo Yi, Julie A Jacko, and Andrew Sears. 2007. Capturing the effects of context on human performance in mobile computing systems. *Personal and Ubiquitous Computing* 11, 2 (2007), 81–96. https://doi.org/10.1007/s00779-006-0063-x

[7] Joanna Bergstrom-Lehtovirta, Antti Oulasvirta, and Stephen Brewster. 2011. The effects of walking speed on target acquisition on a touchscreen interface. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*. 143–146. https://doi.org/10.1145/2037373.2037396

[8] Andrew Bragdon, Eugene Nelson, Yang Li, and Ken Hinckley. 2011. Experimental analysis of touch-screen gesture designs in mobile environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 403–412. https://doi.org/10.1145/1978942.1979000

[9] Emeline Brulé, Gilles Bailly, Marcos Serrano, Marc Teyssier, and Samuel Huron. 2017. Investigating the design space of smartwatches combining physical rotary inputs. In *Proceedings of the 29th Conference on l'Interaction Homme-Machine*. 13–20. https://doi.org/10.1145/3132129.3132139

[10] Chen Chen, Simon T Perrault, Shengdong Zhao, and Wei Tsang Ooi. 2014. BezelCopy: an efficient cross-application copy-paste technique for touchscreen smartphones. In *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces*. 185–192. https://doi.org/10.1145/2598153.2598162

[11] Xiang'Anthony' Chen, Tovi Grossman, and George Fitzmaurice. 2014. Swipeboard: a text entry technique for ultra-small interfaces that supports novice to expert transitions. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*. 615–620. https://doi.org/10.1145/2642918.2647354

[12] Rajkumar Darbar, Prasanta Kr Sen, and Debasis Samanta. 2016. PressTact: Side pressure-based input for smartwatch interaction. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. 2431–2438. https://doi.org/10.1145/2851581.2892436

[13] David Dobbelstein, Gabriel Haas, and Enrico Rukzio. 2017. The effects of mobility, encumbrance, and (non-) dominant hand on interaction with smartwatches. In *Proceedings of the 2017 ACM International Symposium on Wearable Computers*. 90–93. https://doi.org/10.1145/3123021.3123033

[14] S. A. Dudani. 1976. The Distance-Weighted k-Nearest-Neighbor Rule. *IEEE Transactions on Systems, Man, and Cybernetics* SMC-6, 4 (1976), 325–327. https://doi.org/10.1109/TSMC.1976.5408784

[15] Bruno Fruchard, Eric Lecolinet, and Olivier Chapuis. 2020. Side-Crossing Menus: Enabling Large Sets of Gestures for Small Surfaces. *Proceedings of the ACM on Human-Computer Interaction* 4, ISS (2020), 1–19. https://doi.org/10.1145/3427317

[16] Bogdan-Florin Gheran, Jean Vanderdonckt, and Radu-Daniel Vatavu. 2018. Gestures for smart rings: empirical results, insights, and design implications. In *Proceedings of the 2018 Designing Interactive Systems Conference*. 623–635. https://doi.org/10.1145/3196709.3196741

[17] Hyunjae Gil, DoYoung Lee, Seunggyu Im, and Ian Oakley. 2017. TriTap: identifying finger touches on smartwatches. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 3879–3890. https://doi.org/10.1145/3025453.3025561

[18] Mayank Goel, Leah Findlater, and Jacob Wobbrock. 2012. WalkType: using accelerometer data to accomodate situational impairments in mobile touch screen text entry. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2687–2696. https://doi.org/10.1145/2207676.2208662

[19] Mitchell Gordon, Tom Ouyang, and Shumin Zhai. 2016. WatchWriter: Tap and gesture typing on a smartwatch miniature keyboard with statistical decoding. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 3817–3821. https://doi.org/10.1145/2858036.2858242

[20] Rúben Gouveia, Evangelos Karapanos, and Marc Hassenzahl. 2018. Activity tracking in vivo. *Conference on Human Factors in Computing Systems - Proceedings* 2018-April. https://doi.org/10.1145/3173574.3173936

[21] Anhong Guo and Tim Paek. 2016. Exploring tilt for no-touch, wrist-only interactions on smartwatches. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*. 17–28. https://doi.org/10.1145/2935334.2935345

[22] Teng Han, Jiannan Li, Khalad Hasan, Keisuke Nakamura, Randy Gomez, Ravin Balakrishnan, and Pourang Irani. 2018. PageFlip: Leveraging Page-Flipping Gestures for Efficient Command and Value Selection on Smartwatches. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–12. https://doi.org/10.1145/3173574.3174103

[23] Johannes Hartmann, Maximilian Schirmer, and Florian Echtler. 2019. BinarySwipes: Fast List Search on Small Touchscreens. In *Proceedings of Mensch Und Computer 2019* (Hamburg, Germany) *(MuC'19)*. Association for Computing Machinery, New York, NY, USA, 333–340. https://doi.org/10.1145/3340764.3340774

[24] Da-Yuan Huang, Ming-Chang Tsai, Ying-Chao Tung, Min-Lun Tsai, Yen-Ting Yeh, Liwei Chan, Yi-Ping Hung, and Mike Y Chen. 2014. TouchSense: expanding touchscreen input vocabulary using different areas of users' finger pads. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 189–192. https://doi.org/10.1145/2556288.2557258

[25] Mohit Jain and Ravin Balakrishnan. 2012. User learning and performance with bezel menus. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2221–2230. https://doi.org/10.1145/2207676.2208376

[26] Seungwoo Je, Minkyeong Lee, Yoonji Kim, Liwei Chan, Xing-Dong Yang, and Andrea Bianchi. 2018. Pokering: Notifications by poking around the finger. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–10. https://doi.org/10.1145/3173574.3174116

[27] Shaun K Kane, Jacob O Wobbrock, and Ian E Smith. 2008. Getting off the treadmill: evaluating walking user interfaces for mobile devices in public spaces. In *Proceedings of the 10th international conference on Human computer interaction with mobile devices and services*. 109–118. https://doi.org/10.1145/1409240.1409253

[28] Yuki Kubo, Buntarou Shizuki, and Shin Takahashi. 2016. Watch Commander: A Gesture-based Invocation System for Rectangular Smartwatches using B2B-Swipe. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. 37–39. https://doi.org/10.1145/2984751.2985697

[29] Yuki Kubo, Buntarou Shizuki, and Jiro Tanaka. 2016. B2B-Swipe: Swipe gesture for rectangular smartwatches from a bezel to a bezel. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 3852–3856. https://doi.org/10.1145/2858036.2858216

[30] Kapil Kumar, Arindam Mondal, and Gaurav Gupta. 2019. Watch360: A Device to Enable and Detect Tilt, Translation and Rotation of a Watch Bezel. In *2019 16th IEEE Annual Consumer Communications & Networking Conference (CCNC)*. IEEE, 1–6. https://doi.org/10.1109/CCNC.2019.8651875

[31] Gordon Kurtenbach and William Buxton. 1994. User learning and performance with marking menus. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 258–264. https://doi.org/10.1145/191666.191759

[32] Benjamin Lafreniere, Carl Gutwin, Andy Cockburn, and Tovi Grossman. 2016. Faster command selection on touchscreen watches. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 4663–4674. https://doi.org/10.1145/2858036.2858166

[33] Gierad Laput, Robert Xiao, and Chris Harrison. 2016. ViBand: High-Fidelity Bio-Acoustic Sensing Using Commodity Smartwatch Accelerometers. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (Tokyo, Japan) *(UIST '16)*. Association for Computing Machinery, New York, NY, USA, 321–333. https://doi.org/10.1145/2984511.2984582

[34] Kyung-Taek Lee, Hyoseok Yoon, and Youn-Sung Lee. 2018. Implementation of smartwatch user interface using machine learning based motion recognition. In *2018 International Conference on Information Networking (ICOIN)*. IEEE, 807–809. https://doi.org/10.1109/ICOIN.2018.8343229

[35] Luis A Leiva, Alireza Sahami, Alejandro Catala, Niels Henze, and Albrecht Schmidt. 2015. Text entry on tiny qwerty soft keyboards. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 669–678. https://doi.org/10.1145/2702123.2702388

[36] Wing Ho Andy Li, Hongbo Fu, and Kening Zhu. 2016. BezelCursor: bezel-initiated cursor for one-handed target acquisition on mobile touch screens. *International Journal of Mobile Human Computer Interaction (IJMHCI)* 8, 1 (2016), 1–22. https://doi.org/10.4018/IJMHCI.2016010101

[37] Min Lin, Kathleen J Price, Rich Goldman, Andrew Sears, and J Jacko. 2005. Tapping on the move-Fitts' law under mobile conditions. In *Proc. IRMA*, Vol. 5. 132–135.

[38] Ming-I Brandon Lin and Yu-Ping Huang. 2017. The impact of walking while using a smartphone on pedestrians' awareness of roadside events. *Accident Analysis & Prevention* 101 (2017), 87–96. https://doi.org/10.1016/j.aap.2017.02.005

[39] Xing Liu, Tianyu Chen, Feng Qian, Zhixiu Guo, Felix Xiaozhu Lin, Xiaofeng Wang, and Kai Chen. 2017. Characterizing smartwatch usage in the wild. In *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*. 385–398. https://doi.org/10.1145/3081333.3081351

[40] Jess McIntosh, Paul Strohmeier, Jarrod Knibbe, Sebastian Boring, and Kasper Hornbæk. 2019. Magnetips: Combining Fingertip Tracking and Haptic Feedback for Around-Device Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12. https://doi.org/10.1145/3290605.3300638

[41] Robert B Miller. 1968. Response time in man-computer conversational transactions. In *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*. 267–277. https://doi.org/10.1145/1476589.1476628

[42] Shinichi Nakagawa. 2004. A farewell to Bonferroni: the problems of low statistical power and publication bias. *Behavioral Ecology* 15, 6 (11 2004), 1044–1045. https://doi.org/10.1093/beheco/arh107

[43] Matei Negulescu, Jaime Ruiz, Yang Li, and Edward Lank. 2012. Tap, swipe, or move: attentional demands for distracted smartphone input. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*. 173–180.

[44] Ali Neshati, Bradley Rey, Ahmed Shariff Mohommed Faleel, Sandra Bardot, Celine Latulipe, and Pourang Irani. 2021. BezelGlide: Interacting with Graphs on Smartwatches with Minimal Screen Occlusion. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 501, 13 pages. https://doi.org/10.1145/3411764.3445201

[45] Alexander Ng, Stephen A Brewster, and John H Williamson. 2014. Investigating the effects of encumbrance on one-and two-handed interactions with mobile devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1981–1990. https://doi.org/10.1145/2556288.2557312

[46] Alexander Ng, John Williamson, and Stephen Brewster. 2015. The effects of encumbrance and mobility on touch-based gesture interactions for mobile phones. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*. 536–546. https://doi.org/10.1145/2785830.2785853

[47] William S Noble. 2006. What is a support vector machine? *Nature Biotechnology* 24, 12 (2006), 1565–1567. https://doi.org/10.1038/nbt1206-1565

[48] Ian Oakley and Doyoung Lee. 2014. Interaction on the edge: offset sensing for small devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 169–178. https://doi.org/10.1145/2556288.2557138

[49] Ian Oakley, DoYoung Lee, MD Rasel Islam, and Augusto Esteves. 2015. Beats: Tapping gestures for smart watches. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 1237–1246. https://doi.org/10.1145/2702123.2702226

[50] Ian Oakley and Jun-Seok Park. 2007. Designing eyes-free interaction. In *International Workshop on Haptic and Audio Interaction Design*. Springer, 121–132. https://doi.org/10.1007/978-3-540-76702-2_13

[51] Stephen Oney, Chris Harrison, Amy Ogan, and Jason Wiese. 2013. ZoomBoard: a diminutive qwerty soft keyboard using iterative zooming for ultra-small devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2799–2802. https://doi.org/10.1145/2470654.2481387

[52] Simon T Perrault, Eric Lecolinet, James Eagan, and Yves Guiard. 2013. Watchit: simple gestures and eyes-free interaction for wristwatches and bracelets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1451–1460. https://doi.org/10.1145/2470654.2466192

[53] Stefania Pizza, Barry Brown, Donald McMillan, and Airi Lampinen. 2016. Smartwatch in vivo. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 5456–5469. https://doi.org/10.1145/2858036.2858522

[54] Katrin Plaumann, Michael Müller, and Enrico Rukzio. 2016. CircularSelection: Optimizing List Selection for Smartwatches. In *Proceedings of the 2016 ACM International Symposium on Wearable Computers* (Heidelberg, Germany) *(ISWC '16)*. Association for Computing Machinery, New York, NY, USA, 128–135. https://doi.org/10.1145/2971763.2971766

[55] Gulnar Rakhmetulla and Ahmed Sabbir Arif. 2020. SwipeRing: Gesture typing on smartwatches using a segmented QWERTY around the bezel. In *Graphics Interface 2021*. https://openreview.net/forum?id=eVyaNs3sm8a

[56] Volker Roth and Thea Turner. 2009. Bezel swipe: conflict-free scrolling and multiple selection on mobile touch screen devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1523–1526. https://doi.org/10.1145/1518701.1518933

[57] Thijs Roumen, Simon T Perrault, and Shengdong Zhao. 2015. Notiring: A comparative study of notification channels for wearable interactive rings. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 2497–2500. https://doi.org/10.1145/2702123.2702350

[58] Bastian Schildbach and Enrico Rukzio. 2010. Investigating selection and reading performance on a mobile phone while walking. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*. 93–102. https://doi.org/10.1145/1851600.1851619

[59] Marcos Serrano, Eric Lecolinet, and Yves Guiard. 2013. Bezel-Tap gestures: quick activation of commands from sleep mode on tablets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 3027–3036. https://doi.org/10.1145/2470654.2481421

[60] Daniel J Simons and Christopher F Chabris. 1999. Gorillas in our midst: Sustained inattentional blindness for dynamic events. *perception* 28, 9 (1999), 1059–1074. https://doi.org/10.1068/p281059

[61] Gaganpreet Singh, William Delamare, and Pourang Irani. 2018. D-SWIME: A Design Space for Smartwatch Interaction Techniques Supporting Mobility and Encumbrance. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13. https://doi.org/10.1145/3173574.3174208

[62] Andrew L Skinner, Christopher J Stone, Hazel Doughty, and Marcus R Munafò. 2019. StopWatch: The preliminary evaluation of a smartwatch-based system for passive detection of cigarette smoking. *Nicotine and Tobacco Research* 21, 2 (2019), 257–261. https://doi.org/10.1093/ntr/nty008

[63] Vladimir Svetnik, Andy Liaw, Christopher Tong, J Christopher Culberson, Robert P Sheridan, and Bradley P Feuston. 2003. Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling. *Journal of Chemical Information and Computer Sciences* 43, 6 (nov 2003), 1947–1958. https://doi.org/10.1021/ci034160g

[64] David Thissen, Lynne Steinberg, and Daniel Kuang. 2002. Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons. *Journal of educational and behavioral statistics* 27, 1 (2002), 77–83. https://doi.org/10.3102/10769986027001077

[65] Aku Visuri, Zhanna Sarsenbayeva, Niels van Berkel, Jorge Goncalves, Reza Rawassizadeh, Vassilis Kostakos, and Denzil Ferreira. 2017. Quantifying sources and types of smartwatch usage sessions. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 3569–3581. https://doi.org/10.1145/3025453.3025817

[66] Pui Chung Wong, Kening Zhu, Xing-Dong Yang, and Hongbo Fu. 2020. Exploring Eyes-free Bezel-initiated Swipe on Round Smartwatches. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–11. https://doi.org/10.1145/3313831.3376393

[67] Raymond E Wright. 1995. Logistic regression. In *Reading and understanding multivariate statistics*. American Psychological Association, Washington, DC, US, 217–244.

[68] Hui-Shyong Yeo, Juyoung Lee, Andrea Bianchi, and Aaron Quigley. 2016. WatchMI: pressure touch, twist and pan gesture input on unmodified smartwatches. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*. 394–399. https://doi.org/10.1145/2935334.2935375

[69] Bo Yi, Xiang Cao, Morten Fjeld, and Shengdong Zhao. 2012. Exploring user motivations for eyes-free interaction on mobile devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2789–2792. https://doi.org/10.1145/2207676.2208678

[70] Cheng Zhang, Junrui Yang, Caleb Southern, Thad E Starner, and Gregory D Abowd. 2016. WatchOut: extending interactions on a smartwatch with inertial sensing. In *Proceedings of the 2016 ACM International Symposium on Wearable Computers*. 136–143. https://doi.org/10.1145/2971763.2971775

[71] Junhan Zhou, Yang Zhang, Gierad Laput, and Chris Harrison. 2016. AuraSense: enabling expressive around-smartwatch interactions with electric field sensing. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. 81–86. https://doi.org/10.1145/2984511.2984568

A ANGLE CHANGE DETECTION ALGORITHM

Algorithm 1: Angle calculation algorithm

input : An array of all touch points (containing x and y locations) in a B2B gesture
output: The angle of change within the B2B gesture. This value can be used within the machine learning model's feature sets or to compare to the minimum/maximum allowable angle of change for a B2B gesture.

```
/* Set variables for use in algorithm. A granularity value of 5 was
   chosen as optimal after testing a range of values.                */
granularityValue ⟵ 5;
currentAngleChange ⟵ 361;
/* Iterate through touch point array and calculate angle between points
   and test against current angle change.                             */
for i ← 0 to len(touchPointArray) do
    try :
    firstPoint ⟵ touchPointArray[i];
    midPoint ⟵ touchPointArray[i + granularityValue];
    lastPoint ⟵ touchPointArray[i + granularityValue * 2];
    /* Calculate the angle between the three points. calculateAngle
       trivially calculates the angle using the tangent function.
       calculateAngle also checks if the calculated value needs to be
       inverted before returning it as tempAngle.                     */
    tempAngle ⟵ calculateAngle(firstPoint, midPoint, lastPoint);
    /* If the newly calculated angle is less than the currentAngleChange
       then we can set it as our currentAngleChange                   */
    if tempAngle < currentAngleChange and touch points used for angle calculation are
      not within the bezel area then
        currentAngleChange ⟵ tempAngle
    except : pass
/* After iterating through the touch path, return the calculated angle
   of change.                                                         */
return currentAngleChange
```

B FEATURE SETS

Here we provide a table containing all feature sets tested within our machine learning models.

Feature Set	Features
1	starting position, ending position, starting angle, ending angle, minimum starting offset, minimum ending offset, angle of direction change
2	starting position, ending position
3	starting position, ending position, minimum starting offset, minimum ending offset
4	starting position, ending position, starting angle, ending angle, minimum starting offset, minimum ending offset
5	starting position, ending position, minimum starting offset, minimum ending offset, angle of direction change
6	starting position, ending position, starting angle, ending angle
7	starting position, ending position, starting angle, ending angle, angle of direction change
8	starting position, ending position, angle of direction change
9	minimum starting offset, minimum ending offset
10	starting angle, ending angle, minimum starting offset, minimum ending offset, angle of direction change
11	starting angle, ending angle, minimum starting offset, minimum ending offset
12	minimum starting offset, minimum ending offset, angle of direction change
13	starting angle, ending angle, angle of direction change
14	starting angle, ending angle
15	angle of direction change

Table 1. Feature sets for B2B-swipe classification

C EXPANDED MACHINE LEARNING RESULTS

Here we provide tables with expanded results from the machine learning done in Section 8.2. The green in each table highlights the optimal results.

		SVM	KNN	LR	RF
4-segment Layout	Optimal Feature Set	#1	#8	#1	#8
	Training Accuracy	100.00%	98.96%	99.14%	100.00%
	Validation Accuracy	97.38%	98.83%	97.75%	98.63%
	Testing Accuracy	99.88%	98.61%	97.80%	99.83%
	Baseline Overall Accuracy	97.22%			
6-segment Layout	Optimal Feature Set	#1	#8	#1	#10
	Training Accuracy	99.81%	95.69%	95.31%	100.00%
	Validation Accuracy	89.87%	93.02%	89.84%	91.57%
	Testing Accuracy	94.42%	94.31%	92.21%	96.42%
	Baseline Overall Accuracy	87.50%			
8-segment Layout	Optimal Feature Set	#1	#8	#1	#5
	Training Accuracy	98.84%	89.02%	81.09%	100.00%
	Validation Accuracy	79.78%	82.30%	71.48%	82.86%
	Testing Accuracy	89.12%	80.51%	73.60%	91.25%
	Baseline Overall Accuracy	74.00%			

Table 2. Performance of B2B-swipe classification on Study 1 data. The baseline accuracy is calculated with the same way in Study 1's result analysis.

		Baseline	SVM	KNN	LR	RF
Walking without Encumbrance	4-segment Layout	90.10%	90.28%	89.58%	89.58%	89.58%
	6-segment Layout	82.10%	85.03%	89.35%	86.03%	88.66%
	8-segment Layout	66.80%	70.88%	75.00%	68.97%	76.21%
Walking with Encumbrance	4-segment Layout	88.71%	92.36%	93.75%	86.11%	92.36%
	6-segment Layout	74.10%	78.78%	83.02%	78.54%	80.02%
	8-segment Layout	57.00%	64.41%	67.32%	62.33%	67.71%

Table 3. Classification performance of the Study1-data-trained models on all the Study 2 data. The baseline accuracy is calculated with the same way in Study 2's result analysis.

		SVM	KNN	LR	RF
4-segment Layout	Optimal Feature Set	#7	#13	#7	#1
	Training Accuracy	100.00%	96.68%	97.24%	100.00%
	Validation Accuracy	95.33%	95.34%	95.28%	96.75%
	Testing Accuracy	98.38%	98.32%	98.38%	99.31%
	Baseline Overall Accuracy	90.10%			
6-segment Layout	Optimal Feature Set	#1	#5	#7	#8
	Training Accuracy	99.84%	92.87%	89.85%	100.00%
	Validation Accuracy	87.35%	89.94%	86.41%	89.89%
	Testing Accuracy	91.64%	84.59%	81.02%	94.03%
	Baseline Overall Accuracy	82.10%			
8-segment Layout	Optimal Feature Set	#1	#8	#1	#13
	Training Accuracy	98.05%	86.78%	78.48%	100.00%
	Validation Accuracy	69.09%	73.55%	64.88%	72.27%
	Testing Accuracy	81.45%	71.01%	61.37%	85.20%
	Baseline Overall Accuracy	66.80%			

Table 4. Performance of B2B-swipe classification trained by the Study 2 data of walking without and with encumbrance. The baseline accuracy is calculated in the same manner as Study 2's result analysis.

		SVM	KNN	LR	RF
4-segment Layout	Optimal Feature Set	#2	#3	#7	#6
	Training Accuracy	97.62%	95.89%	96.05%	100.00%
	Validation Accuracy	92.87%	95.02%	90.16%	94.08%
	Testing Accuracy	96.35%	95.25%	92.25%	97.86%
	Baseline Overall Accuracy	88.71%			
6-segment Layout	Optimal Feature Set	#1	#1	#1	#8
	Training Accuracy	98.66%	89.65%	84.29%	100.00%
	Validation Accuracy	75.93%	80.51%	75.30%	81.56%
	Testing Accuracy	86.47%	78.11%	71.15%	89.25%
	Baseline Overall Accuracy	74.10%			
8-segment Layout	Optimal Feature Set	#1	#7	#1	8
	Training Accuracy	94.67%	80.23%	67.03%	100.00%
	Validation Accuracy	62.36%	66.08%	58.57%	68.94%
	Testing Accuracy	76.06%	63.41%	52.29%	82.96%
	Baseline Overall Accuracy	57.00%			

Table 5. Performance of B2B-swipe classification trained by the Study 2 data for walking with encumbrance. The baseline accuracy is calculated in the same manner as Study 2's result analysis.

Received February 2022; revised May 2022; accepted June 2022