Typing on a Smartwatch While Mobile: A Comparison of Input Methods

Colton J. Turner, Priority Designs, Columbus, OH, USA, **Barbara S. Chaparro**, Embry-Riddle Aeronautical University, Daytona Beach, FL, USA, and **Jibo He**, Wichita State University, KS, USA

Objective: The user experience of typing on a smartwatch was evaluated with three unique input methods (tap, trace, and handwriting) while standing and while walking.

Background: Despite widespread development within the technology industry, smartwatches have had a relatively slow adoption worldwide compared to smartphones. One limiting factor of smartwatches has been the lack of an efficient means of text entry. The 2017 release of Android Wear addressed this issue by providing support for native text entry (i.e., tap, trace, and handwriting). Determining how user performance and subjective ratings compare across these input methods is essential to understanding their contribution to smartwatch user experience.

Method: Twenty college-age individuals typed phrases using tap, trace, and handwriting input on a smartwatch in three different mobility scenarios (standing, walking a simple course, walking a complex course).

Results: Participants typed faster with trace (30 words per minute; WPM) than with tap (20 WPM) and handwriting (18 WPM), regardless of mobility. Trace also outperformed tap and handwriting across all subjective metrics, regardless of mobility.

Conclusion: Trace input appears to be especially well suited for typing on a smartwatch as it was found to be objectively and subjectively superior to tap and handwriting regardless of user mobility. Objectively, typing speeds with trace are shown to be nearly two times faster than most alternative input methods described in the literature.

Application: Results suggest smartwatch manufacturers should include QWERTY keyboards with trace input as a standard feature in order to provide the best overall typing experience for their users.

Keywords: wearable devices, mobile devices, product design, interface evaluation, usability testing and evaluation

Address correspondence to Barbara S. Chaparro, Department of Human Factors and Behavioral Neurobiology, Embry-Riddle Aeronautical University, Daytona Beach, FL 32114, USA; e-mail: chaparb1@erau.edu

HUMAN FACTORS
Vol. 63, No. 6, September 2021, pp. 974-986
DOI:10.1177/0018720819891291
Article reuse guidelines: sagepub.com/journals-permissions

When the modern smartwatch first hit the consumer market in 2013, it brought with it a completely new and untapped area within personal computing. The promise of ushering in the next wave of convenient technology soon had companies of all sizes and verticals vying for a share of the new market. Despite this widespread interest in smartwatch development, by 2016 smartwatch sales worldwide largely failed to meet expectations, leading some to conclude that the smartwatch industry was dead (Thompson, 2018). According to the International Data Corporation (IDC), smartwatch sales decreased 51.6% from the third quarter of 2015 to the third quarter of 2016 (IDC, 2016). Several factors contributed to the limited sales of smartwatches such as poor battery life, limited apps, and an inability to perform daily activities commonly performed on smartphones (Pulvirent, 2015; Thompson, 2018). In terms of daily activities, smartwatches have historically failed to provide an efficient means of text entry, an extremely common and frequent use of smartphones. According to Jeong et al. (2017) one of the most frequent uses of smartwatches is the checking of notifications, such as messages and emails. However, free form text response to those messages using the smartwatch has been limited and generally required the user to respond using their smartphone instead.

Original smartwatch designs only provided preprogrammed responses and voice input for text entry, yet these methods can be limiting and impractical for real-world use. No smartwatch designs initially allowed for keyboard-based text entry largely due to the initial assumption that traditional keyboard layouts, like the QWERTY, would not work for smartwatches given their small screen sizes. Early thoughts of including a keyboard for smartwatch typing purposes faced much skepticism for three main reasons. First, the size of the smartwatch screen and resulting size of the keyboard was thought to be too small for effective use (Arif et al., 2011; Hong et al., 2015). Second, users' fingers were assumed to be too large in relation to the keyboard to accurately hit the small keys, also known as the "fat finger issue" or "fat finger problem" (Arif et al., 2011; Kim et al., 2006; Oney et al., 2013; Siek et al., 2005). Third, the users' input finger was thought to be too large in relation to the size of screen and could occlude the users' view of the screen (Arif et al., 2011; Funk et al., 2014). The apparent lack of an efficient text entry method for smartwatches sparked a bevy of research into novel text input methods optimized for the small screen size. This included alternative smartwatch designs and attempts to expand the input area to outside of the watch face (Baudisch & Chu, 2009; Funk et al., 2014; Lyons et al., 2012).

Although it was initially assumed that keyboard-based text entry was impossible on a smartwatch, numerous studies have since shown its feasibility. In a review of the existing smartwatch keyboards at the time, Arif and Mazalek (2016) provided a summary table and detailed descriptions of these keyboards. Turner et al. (2018) provided an updated version of the table by Arif and Mazalek (2016) with the latest research findings and added columns for participant mobility (seated, standing, or walking) and subjective measures. See Supplemental Material on the *HF* Web site for a table summarizing the latest literature of text entry on smartwatches and smartwatch sized displays. Given the small screen size of smartwatches, and the resulting small key size, many advocate the need for alternative keyboards that use "zoom" features to enlarge the small keys. While this method has been shown to be feasible, resulting text entry speeds are often slow and accompanied by a steep learning curve. Chaparro et al. (2015) proposed that the creation of novel keyboards and input methods for smartwatches may be unnecessary. The authors demonstrated that an existing smartphone keyboard could be adapted for efficient use on a smartwatch. Because the keyboard was familiar to users, they were able to circumvent the steep learning curves associated with novel, alternative keyboards allowing the users to quickly learn how to use the small keyboard and achieve efficient text entry speeds.

Turner et al. (2017) extended the work of Chaparro et al. (2015) in a comparison of two input methods, tap and trace, finding both input methods as viable and efficient means of text entry on a smartwatch using a small QWERTY keyboard. Additionally, the authors found that regardless of previous experience, users typed faster overall using the trace input method than tap. Furthermore, Turner et al. (2018) investigated how user mobility (standing and walking on a treadmill), input method (tap or trace), and previous experience with trace input (novice or expert) affected typing performance on a smartwatch. Results revealed participants typed faster with trace than with tap, regardless of whether they were standing or walking or whether they had prior experience using trace input. Additionally, participants typed faster overall while standing than while walking on a treadmill.

Although tap and trace have consistently been shown to be efficient means of text entry on a smartwatch, no studies at this time have directly compared the two input methods to other commercially available input methods, such as handwriting. This limitation is especially relevant given the 2017 release of Android 2.0 (renamed Android Wear in 2018). Wear OS now allows smartwatches to have full texting capabilities, allowing users to receive, compose, and send full text messages, just as they would on their smartphone, using either tap, trace, or handwriting input on their smartwatch. Now that tap, trace, and handwriting input are all available for use by the mass consumer market, understanding how they compare is essential to ensuring users have access to the optimal smartwatch typing experience. Additionally, it is unknown how typing performance with these different input methods is affected by mobility, notably naturalistic walking, rather than treadmill walking. Although treadmills can be used to establish a measure of baseline walking and typing performance, they are limited in external validity and may not replicate everyday walking (Dingwell et al., 2001). Given the ubiquitous nature of text entry on a smartphone, understanding how text entry performance on a smartwatch is affected by mobile environments is critical if smartwatches are to ever become as commonplace as smartphones.

HANDWRITING

Perceived issues with onscreen smartwatch keyboards and learning curves associated with novel methods have led some to believe the solution to text input is handwriting (Costagliola et al., 2017; Lin et al., 2018). Costagliola et al. (2017) explains that handwriting is especially well suited for text entry on smartwatches because it is a natural input method, taking advantage of knowledge most users acquire early in life. Traditional paper-and-pen handwriting yields a writing speed of approximately 18 WPM (Connor, 1995; Dutton, 1990). Handwriting text entry speeds for touchscreen devices tend to vary by device, ranging from 16 to 25 WPM for smartphones and tablets (Castellucci & MacKenzie, 2008; Kristensson & Denby, 2009; MacKenzie & Chang, 1999). In a comparison of tap, trace, and handwriting on a smartphone, Castellucci and MacKenzie (2011) found that users typed 10–13 WPM faster with tap and trace than with handwriting. Additionally, the authors reported that users largely disliked handwriting in comparison to other methods, specifically due to the higher incidence of errors and slow typing speed. Observed handwriting speeds on a smartwatch range from 15 to 19 WPM, similar to the speeds observed with alternative keyboards on a smartwatch (Costagliola et al., 2017).

Walking and Typing on a Smartwatch

Although walking and typing is a common user behavior with a smartphone, only three studies to date have evaluated typing performance on a smartwatch in mobile scenarios: Hong et al. (2016), Darbar et al. (2016), and Turner et al. (2018). All three studies found that typing on a smartwatch while standing and while walking is feasible using a variety of input methods, although trace input seems to be particularly well suited. Additionally, all three studies found typing performance significantly worsened while walking than when standing, as it does on a smartphone. Hong et al. (2016) and Darbar et al. (2016) reported typing speeds ranging from 8 to 15 WPM when walking or standing with alternative keyboards and 5 to 13 WPM with a standard QWERTY keyboard using tap input. Turner et al. (2018) reported typing speeds of 35 WPM with trace and 30 WPM with tap using a standard QWERTY keyboard when walking or standing. However, the walking scenarios employed in each study varied. Hong et al. (2016) and Turner et al. (2018) simulated the walking task using a treadmill. Darbar et al. (2016) had participants perform a naturalistic walking scenario, but the details of the walking scenario were not provided.

Mechanics of Tap, Trace, and Handwriting

It is likely the underlying physical mechanics of tap and trace are responsible for the observed performance differences across the literature. Tap input requires users to lift their finger before and after each keystroke, and to manually enter a space. When walking, accurate keystrokes become more difficult as both the keyboard and the user's input finger are in unsynchronized motion due to the constant movement of the body with each step (Kane et al., 2008). In contrast, trace requires users to commit only one initial keystroke, allowing them to drag their finger across all the intended letters in a word, and automatically inserts a space when the finger is lifted from the screen. Similar to tap, the user does not have to tap every key perfectly, but instead only has to be close enough to the intended keys for the autocorrect algorithm to function properly (Zhai & Kristensson, 2012). It is likely the decreased number of keystrokes required with trace is one reason for its superiority over tap across mobile scenarios. With trace, the user's constant contact with the keyboard likely diminishes the difficulty of making accurate keystrokes due to the unsynchronized movement of the keyboard and the user's input finger. Furthermore, trace has been shown to be less physically exerting for muscles in the lower arm than tap which may also account for some of the observed differences between the two input methods (Sonaike et al., 2016). For

Figure 1. Google keyboard with trace input (left), Google keyboard with standard point-and-tap input (middle), and Google Handwriting Input app (right).

handwriting, the mechanics of tap and trace also apply depending on the writing style (cursive vs print).

Purpose

This study evaluates and compares the user experience of typing on a smartwatch with three unique input methods (tap, trace, and handwriting) while standing, walking a simple route, and walking a complex route in a quasi-naturalistic setting.

METHOD

The metrics gathered in this study replicate and expand upon those used in Turner et al. (2017, 2018). Text input method (tap vs. trace vs. handwriting) and mobility (standing vs. walking a simple route vs. walking a complex route) were the independent variables. Typing speed (words per minute; WPM) and accuracy (word error rate; WER), walking behavior, and subjective measures of performance were the dependent variables. Multiple hand dimensions were also gathered to assess if there was a relationship between hand and finger size and typing performance.

Participants

Twenty college-aged participants (12 female, 8 male), ranging from 19 to 25 years of age (M = 21.60, SD = 2.06), participated in this study. All participants were right-hand dominant and typed on the smartwatch using the index finger of their right hand. Participants self-reported their experience level with tap, trace, and handwriting on a smartphone on a 1–7 scale (1 = no experience; 7 = expert). Participants with a self-report rating of 1 or 2 were classified as "Novices," and "Experts" with a 6 or 7 rating (Turner et al., 2018). Using this classification, all participants in the present study were self-reported tap experts (none reported expertise with trace or handwriting). All participants also indicated walking and typing on a smartphone was a common behavior in which they engaged.

Materials

A Mobvoi Ticwatch E smartwatch (display size of 1.4 inches) running the Android 2.0 operating system was used in this study (Figure 1). The Ticwatch was equipped with the native Android 2.0 and full QWERTY Google keyboard that supports both tap and trace input (3 mm × 3 mm key size). Additionally, the watch was equipped with the Google Handwriting Input app (~30 mm by ~20 mm writing area). Participants were instructed to print when handwriting.

A subset of phrases were randomly selected from a list of 500 composed by MacKenzie and Soukoreff (2003). Ten practice phrases and 15 experimental phrases were randomly chosen for each condition; there was no overlap between the practice and experimental phrases. Phrases were randomly selected for all participants across all conditions. The phrases contained lowercase

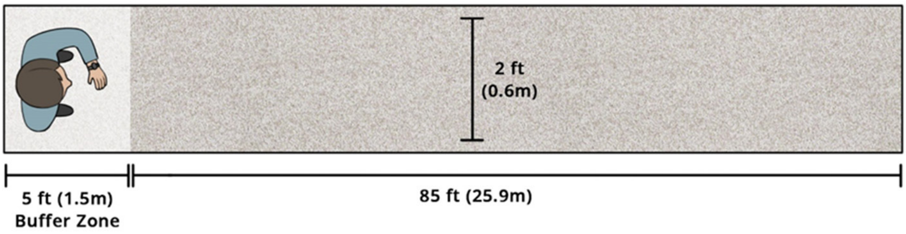

Figure 2. Simple walking route.

letters only (no numbers, symbols, punctuation, or uppercase letters). Phrases ranged from 16 to 42 characters for all conditions.

For the walking conditions, participants walked two predefined routes that were 2 feet (0.6 m) wide, 85 feet long (25.9 m), and of varying complexity. The walking routes were set up inside a quiet, well-lit lab space. Both routes were designed to be long enough so as to prevent participants from reaching the end of the route before they finished typing an intended phrase. Each route was delineated by two blue lines marking the borders of the route. In addition, for the complex route, desks were used on the perimeter of the route to prevent participants from seeing the obstacle arrangements prior to encountering them during the typing task. This was also done to reduce the likelihood of participants memorizing the spatial layout of the obstacles between trials (Turano et al., 1998).

In order to ensure participants were actively engaged in walking before beginning the typing task, they were given a buffer zone of 5 feet (1.5 m) at the beginning of each route. Participants were not allowed to start typing the phrase until they exited the buffer zone. The simple route was a straightaway without any obstacles (Figure 2).

The complex route was a serpentine route with four possible 90° turns and obstacles every 7–12 feet (2.1–3.7 m) that participants were instructed to walk around (Figure 3). The obstacles consisted of three blocks, three foam pads, and three office chairs (nine obstacles in total). The obstacles varied in height and width (Table 1). The arrangement of the obstacles was changed after every five typing phrases. Participants were not allowed to watch as the obstacles were rearranged. The blocks and office chairs were always positioned on the perimeter of the track and occupied 12 inches of space, leaving 12 inches of open space for participants to walk past the obstacles. The foam pads were placed in the center of the route and occupied 8 inches of space, leaving 8 inches of space on either side of the pad for participants to walk past the obstacle.

An Ames Instruments 100 feet (30.5 m) laser tape measure (accurate within ±1/16 of an inch (±1.6 cm)) was used to measure the distance participants walked while typing each phrase.

Procedure

This study was conducted over a 2-day span in order to minimize fatigue. Day 1 consisted of participants completing either one or two conditions. The subsequent condition(s) were completed on Day 2. All conditions were partially counterbalanced such that the two walking conditions did not occur on the same day to minimize fatigue.

On Day 1, participants first completed a brief background survey. They were then introduced to their first condition. For each input method, participants were given a brief typing tutorial by the experimenter and then given 10 practice phrases to type before the experimental trials began. For the experimental trials, 15 phrases were presented one-at-a-time to the participants by the experimenter. Before starting the phrase, participants were instructed to read the phrase aloud to ensure comprehension. Once they felt prepared to begin typing, they verbally indicated so by saying “Start,” and once they had finished typing the phrase, they verbally indicated so by

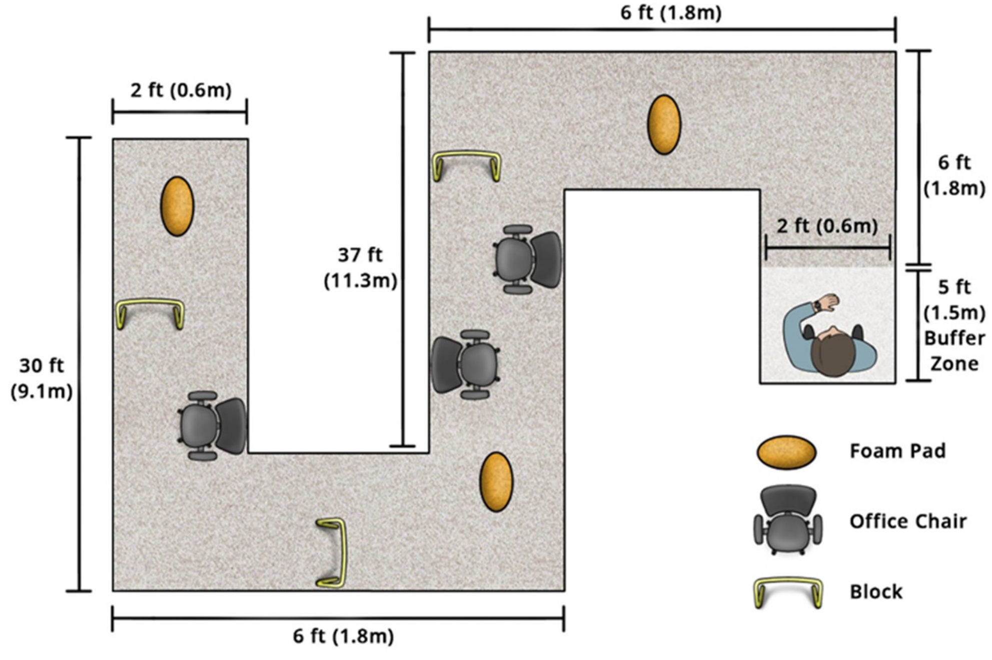

Figure 3. Example obstacle arrangement within the complex walking route.

TABLE 1: Complex Route Obstacles

	Block	Foam Pad	Office Chair
No. of obstacles	3	3	3
Height	10 in	5 in	48 in
Width	12 in	8 in	12 in
Position on route	Perimeter	Middle of route	Perimeter

saying "Stop" (Arif et al., 2011; MacKenzie & Read, 2007). Participants were instructed to type the phrases as quickly and accurately as possible. They were allowed to correct mistakes but were not required to do so. Typing time was recorded by a researcher using a digital stopwatch (Arif et al., 2011; MacKenzie & Read, 2007). Phrases were saved as a text file to an online repository and later scored manually by an experimenter. In total, each participant typed 225 phrases (25 per input method, 75 per mobility condition). Participants were required to participate in the Day 2 condition(s) within 7 days of their Day 1 participation.

For the standing condition, participants were asked to stand in the same spot during the typing of all phrases. The participants were not allowed to lean or stabilize themselves on any objects or furniture. They could sit and rest at the conclusion of each condition if desired.

For the complex route, participants were instructed to walk around the obstacles and to avoid contacting them. Participants were allowed to set their own walking speed but were instructed to walk at a comfortable pace. Participants were not allowed to start typing the phrase until they exited the buffer zone. Once a participant had finished typing a phrase, they were instructed to stop walking and their distance from the start point (after the buffer zone) was recorded using a laser tape measure. Once participants' walking distance had been recorded, they returned to the beginning of the route for the start of a new phrase. Prior to the typing conditions, a baseline walking speed was gathered at the beginning of both walking conditions by having participants walk each route 3 times.

Once participants completed the 15 experimental phrases of a condition, they completed a perceived usability survey and a subjective workload assessment. After finishing the subjective workload assessment, participants were introduced to the second condition and the steps were repeated.

At the end of each mobility condition, participants were asked to rate their perceived performance, preference, and intent to use. After all conditions had been completed, typing hand and finger dimensions were measured. Participants were then debriefed and thanked for their time.

Design

A 3 × 3 repeated measures design was used for this study. Text input method (tap vs. trace vs. handwriting) and mobility (standing vs walking a simple route vs. walking a complex route) were the independent variables. Dependent variables included typing performance, walking behavior, and subjective perceptions of: usability, workload, performance, accuracy, speed, and intent to use.

Typing performance. Typing performance was measured by typing speed (WPM) and typing accuracy (WER). Typing speed was calculated using WPM = $12 \times (T - 1)/S$, where T is the number of transcribed characters, S is the number of seconds, and one word is assumed to be 5 characters (MacKenzie & Tanaka-Ishii, 2010). Typing accuracy, WER, was calculated using the number of word errors per phrase divided by the total number of words per phrase.

Walking behavior. For the walking conditions, walking behavior was measured by using walking speed (WS), lane deviations (line steps), and frequency of obstacle hits. Line steps and frequency of obstacle hits were removed from the analysis due to the extreme rarity of which they occurred.

Subjective measures. The subjective measures were determined by measuring the subjective workload, perceived usability, perceived performance and preference, and intent to use.

Subjective workload. The raw NASA Task Load Index (NASA TLX – R; Hart & Staveland, 1988) was used to measure participants' perceived workload and performance after each condition. Participants provided ratings on a 21-point scale for perceived mental, physical, and temporal effort; performance; overall effort; and frustration. A higher score indicates a more demanding experience or worse perceived performance.

Perceived usability. An adapted System Usability Scale (SUS) was used to measure participants' perceived usability of each input method within each mobility condition. The SUS is an industry-standard 10-item questionnaire with 5 response options (Strongly Disagree to Strongly Agree) that is summarized as a single score between 0 and 100 (Brooke, 2013). Higher scores indicate higher perceived usability. The scale was adapted by replacing "system" with "combination," which referenced the use of the input method during a given mobility condition.

Perceived performance and preference. Perceived accuracy, perceived speed, and overall preference with each input method and mobility condition were measured using a 50-point scale with higher scores reflecting more preferred, or better, in terms of accuracy or speed.

Intent to use. Participants rated the likelihood they would use each input method with each mobility condition using a 0–10 scale, with a 10 being "very likely."

Anthropometric measurements. A sliding digital caliper was used to measure the typing hand of each participant. Hand measurements included the length and width of hand and length, width, and circumference of the index finger in millimeters.

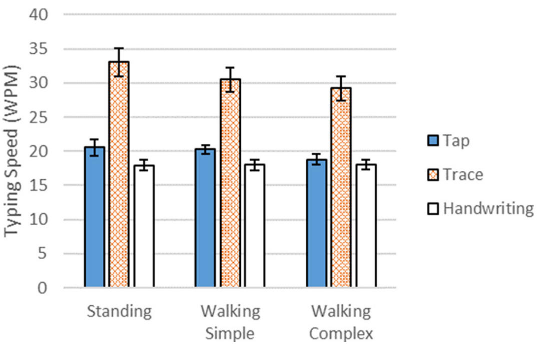

Figure 4. Typing speed. Error bars represent ±1 standard error.

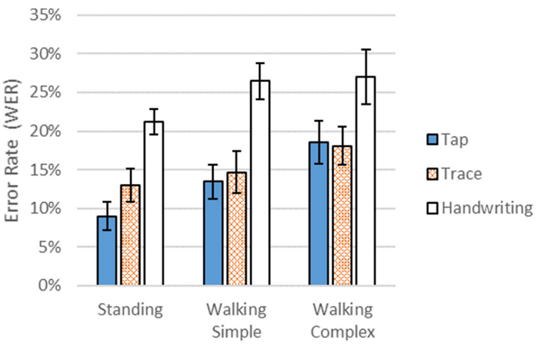

Figure 5. Typing accuracy. Error bars represent ±1 standard error.

Results

All dependent measures of typing performance and subjective ratings were analyzed using a 3 × 3 repeated-measures ANOVA. All walking metrics were analyzed using a 2 × 3 repeated-measures ANOVA. One-way ANOVAs were conducted in order to determine if any practice or learning effects occurred with any of the input methods over time. Partial eta squared (η_p^2) was used to estimate effect size for all ANOVA tests. Analyses of simple main effects were conducted to follow up on all significant interactions. Bonferroni correction was used to control for family-wise type I error across multiple comparisons.

Typing Speed

A significant main effect of input method was found for typing speed (WPM) along with a significant interaction of input method and mobility: $F(2, 38) = 63.86$, $p < .001$, $\eta_p^2 = .77$; $F(4, 76) = 3.09$, $p = .02$, $\eta_p^2 = .14$, respectively. Follow-up analysis revealed participants typed faster with trace than with both tap and handwriting in each of the mobility conditions. Additionally, participants typed faster with trace when standing ($M = 33.06$, $SD = 9.15$) than when walking the complex route ($M = 29.23$, $SD = 8.08$), $p < .05$ (see Figure 4).

Typing Accuracy

Significant main effects of input method and mobility were found for typing accuracy (word error rate). Participants typed more accurately with tap ($M = 14\%$, $SD = 10\%$) and trace ($M = 15\%$, $SD = 11\%$) than with handwriting ($M = 25\%$, $SD = 11\%$); $F(2, 38) = 24.03$, $p < .001$, $\eta_p^2 = .56$. Participants also typed more accurately when standing ($M = 14\%$, $SD = 8\%$) than when walking the complex route ($M = 21\%$, $SD = 13\%$); $F(2, 38) = 8.98$, $p = .001$, $\eta_p^2 = .32$ (see Figure 5).

Typing Speed and Accuracy Over Time

Multiple one-way ANOVAs were conducted in order to determine if any practice or learning effects occurred with any of the input methods over time with respect to typing speed and with respect to accuracy. No significant main effects were found for typing speed or typing accuracy over time, $p > .05$.

Subjective Workload

Using the NASA TLX – R, significant main effects of input method and mobility were found for subjective workload. Participants rated their mental demand, physical demand, effort, and frustration lower when using trace than when using tap and handwriting: $F(2, 38) = 5.85$, $p = .006$, $\eta_p^2 = .24$; $F(2, 38) = 15.16$, $p < .001$, $\eta_p^2 = .44$; $F(2, 38) = 10.81$, $p < .001$, $\eta_p^2 = .36$; $F(2, 38) = 3.25$, $p = .05$, $\eta_p^2 = .15$, respectively. Participants also rated their performance better when using trace than when using handwriting: $F(2, 38) = 4.10$, $p = .02$, $\eta_p^2 = .18$ (see Figure 6).

Participants rated their mental demand and effort lower when standing than when walking the simple and complex routes: $F(2, 38) = 9.73$,

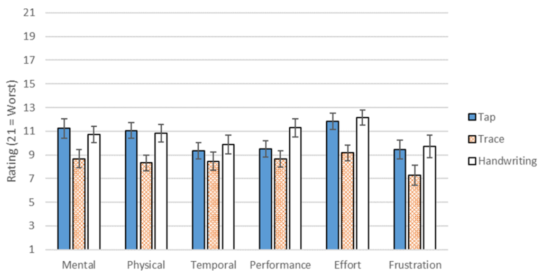

Figure 6. Perceived workload dimensions by input method. Error bars represent ±1 standard error.

$p < .001$, $\eta_p^2 = .34$; $F(2, 38) = 4.60$, $p = .02$, $\eta_p^2 = .20$, respectively.

Perceived Usability

Using the SUS, a significant main effect of input method was found for perceived usability. Participants reported higher perceived usability when using trace than when using tap and handwriting: $F(2, 38) = 3.80$, $p = .03$, $\eta_p^2 = .17$ (see Figure 7).

Perceived Accuracy, Speed, and Preference

Using a 50-point scale, significant main effects of input method were found for perceived accuracy, speed, and preference. Participants reported higher perceived accuracy ratings when using trace than when using handwriting: $F(2, 38) = 3.34$, $p = .05$, $\eta_p^2 = .15$. Participants also reported higher perceived accuracy ratings when standing than when walking the complex route: $F(2, 38) = 4.27$, $p = .02$, $\eta_p^2 = .18$. Participants reported higher perceived speed ratings when using trace than when using tap and handwriting: $F(2, 38) = 14.39$, $p < .001$, $\eta_p^2 = .43$. Participants also reported higher preference ratings when using trace than when using tap and handwriting: $F(2, 38) = 5.59$, $p = .01$, $\eta_p^2 = .23$.

Intent to Use

Using an 11-point scale, a significant main effect of input method was found for intent to use. Participants reported higher intent to use ratings for trace than for tap and handwriting: $F(2, 38) = 7.24$, $p = .002$, $\eta_p^2 = .28$ (see Figure 8).

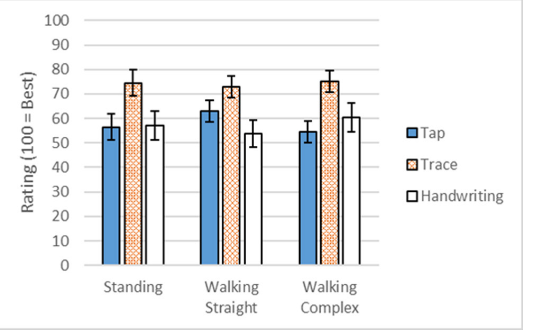

Figure 7. Perceived usability. Error bars represent ±1 standard error.

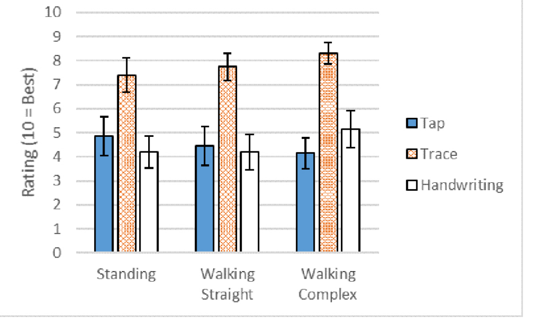

Figure 8. Intent to use. Error bars represent ±1 standard error.

Walking Speed

For the simple route, participants walked faster for their baseline walking speed (2.5 mph using no input method) than they did while using any of the input methods: $F(3, 57) = 21.14$, $p < .001$, $\eta_p^2 = .53$. Likewise, for the complex route, participants walked faster for their baseline walking speed (2.1 mph using no input method) than they did while using any of the input methods: $F(3, 57) = 75.97$, $p < .001$, $\eta_p^2 = .80$.

Participants walked faster when using trace than tap: $F(2, 38) = 3.37$, $p < .05$, $\eta_p^2 = .15$. Participants also walked faster when walking the simple route than when walking the complex route: $F(1, 19) = 52.98$, $p < .001$, $\eta_p^2 = .74$. A significant interaction of input method and mobility also was found for walking speed: $F(2, 38) = 4.90$, $p = .01$, $\eta_p^2 = .15$. Follow-up

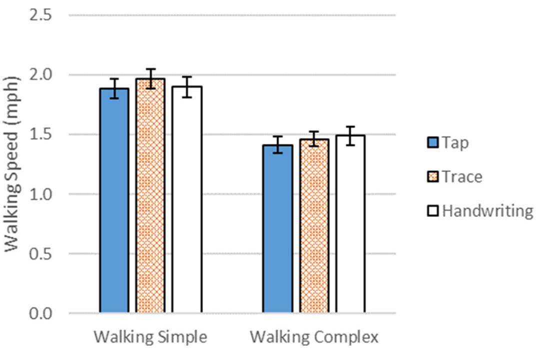

Figure 9. Walking speed. Error bars represent ±1 standard error.

analysis revealed participants walked faster when using trace in the simple route but faster when using handwriting in the complex route than tap, $p < .05$ (Figure 9).

Hand Measurements

To determine whether there was any evidence of the "fat finger" issue, a series of correlations were conducted between WPM, WER, hand width, index finger width, and index finger length for all conditions. The range of participants' hand widths was representative of the 1st–75th percentile of adult men and women (White, 1980). No significant correlations were found, $p > .05$ (r values ranged from –.35 to +.36).

DISCUSSION

This is the first known study to examine the user experience of typing on a smartwatch using three different text input methods in three different mobile environments. The results show tap, trace, and handwriting are all viable means of text entry on a smartwatch, yet they are not equally efficient, nor effective. Participants were able to achieve the fastest typing speeds with trace across all three mobility conditions (29–33 WPM), while participants were only able to achieve 18–20 WPM for both tap and handwriting across all three mobility conditions. The observed typing speeds, and observed superiority, of trace are consistent with previous findings and continue to be among the fastest in the reported literature regardless of participant mobility (sitting, standing, walking on a treadmill, quasi-naturalistic walking).

Across all mobility conditions, participants did not differ in their typing accuracy between tap and trace (14%–15% WER), yet participants were significantly less accurate with handwriting (25% WER). The observed WERs for tap and trace are consistent with other observed error rates on smartwatches, but the observed WERs for handwriting are somewhat higher than what is reported in the literature. It is likely that the different error rates for handwriting are due to the handwriting keyboard used in Costagliola et al. (2017), as it only offered a static writing area. Google Handwriting implements an automatic horizontal scroll, or dynamic flow. As participants are writing characters, the screen automatically scrolls to the left continuously creating a blank writing area for new characters to be written. Subjectively, some participants commented the flow of the writing surface was not fast enough to allow them to write at a natural pace, ultimately limiting them from writing faster. Additionally, participants commented on the inefficiency of the auto-correct for handwriting, while they praised the auto-correct abilities of trace and tap. These are some of the same reasons why participants preferred tap and trace input over handwriting on a smartphone (Castellucci & MacKenzie, 2011).

Based on the results of Turner et al. (2018) it was expected participants would have higher WERs with tap and handwriting than with trace when walking. Mechanically, tap and handwriting input (participants were instructed to use print handwriting) require the user to lift their finger before and after each keystroke. When walking, accurately inputting text is even more difficult due to the constant motion of the body with each step. In contrast, trace requires the user to use one continuous motion to type, so the finger is always in contact with the screen. According to Schildbach and Rukzio (2010) input methods that require very accurate finger movements, such as tap and handwriting, cannot be used effectively while walking. Future research should examine the biomechanics of each input method to investigate further.

Given that all participants were naïve to typing on a smartwatch, typing speed and accuracy over time were analyzed in order to determine if any learning effects occurred. No learning effects were found for any of the input methods. This result is especially interesting for trace as all of the participants indicated little to no experience with it prior to the study. Despite this, they were quickly able to achieve an extremely high rate of typing speed and high accuracy with minimal practice across all three mobility conditions. The ability to quickly learn and be efficient with a new input method meets the key requirements for an input method to be accepted by the mass consumer market according to Zhai and Kristensson (2012).

In addition to its superior objective performance, trace was also subjectively superior to tap and handwriting, regardless of mobility. Participants indicated they would prefer, and be more likely, to use trace over tap and handwriting when typing on a smartwatch.

Better typing performance while standing versus walking is a consistent trend in the literature (Clawson et al., 2014; Conradi et al., 2015; Darbar et al., 2016; Hong et al., 2016; Mizobuchi et al., 2005; Turner et al., 2018). Yet, interestingly, typing speed differed very little for all input methods across the three mobility conditions. Participants typed more accurately when standing (14% WER) than when walking the complex route (21% WER), yet there was no difference when walking the simple route (17% WER). This is somewhat surprising. It was expected that typing accuracy would decrease from standing to walking the simple route, as is consistent with the literature (Bergstrom-Lehtovirta et al., 2011; Darbar et al., 2016; Hong et al., 2016). Based on previous research, the authors assumed any form of walking would result in worse typing performance on a smartwatch than when standing. The data from this study suggest walking a straight route, free of obstacles is not actually more difficult than standing and typing, especially for the population in this study (college students) who are used to texting on their phone while walking. Additionally, there was no evidence that participants were sacrificing their walking performance to bolster their typing performance.

We believe the observed superiority of trace over other keyboards and input methods listed in the literature is attributable to several factors (Turner et al., 2018). First, participants are inherently familiar with the QWERTY keyboard layout as they generally use it with their primary input method. This results in a shorter learning curve than alternative keyboard layouts or alternative input methods allowing participants to immediately be effective and efficient. Second, the small screen size of the smartwatch is actually a benefit to trace input as it requires less distance for the participant's tracing finger to travel while typing. This, coupled with the fact that the trace keyboard used in this study employed an effective autocorrect algorithm, resulted in more efficient typing.

Conclusion

These results demonstrate that all three input methods are viable means of inputting text in a quasi-naturalistic mobility scenario. Participants completely naïve to typing on a smartwatch were able to achieve typing speeds consistent, or superior, to speeds achieved with other methods listed in the literature with little practice. Trace input appears to be especially well suited for typing on a smartwatch regardless of mobility as participants were able to type 29–33 WPM depending on the mobility condition with minimal error (15% WER), while only achieving 18–20 WPM with tap and handwriting. In addition to performance, trace also outperformed tap and handwriting across all subjective metrics, indicating a more positive user experience overall. Participants rated trace as easier and less demanding to use, preferred it over tap and handwriting, and suggested they would use trace over tap and handwriting on a smartwatch in the future.

The present study provides evidence that consumers can use a smartwatch as a text input device efficiently and effectively with a standard QWERTY keyboard with trace input. Future research should be conducted to further generalize these results with non-college-age participants with a wider range of text stimuli (including symbols, digits, emojis, and punctuation) and naturalistic environments.

Limitations

The current study examines performance while walking in a natural setting, albeit an obstacle course indoors. While this can be considered "quasi-naturalistic," it is a necessary and logical next step in the question of efficacy of text input methods. We understand that the most generalizable test environment may be an outdoor naturalistic setting; however, the lack of experimental control and potential confounds of variable sunlight, temperature, and so on may complicate the interpretation of the results. Future studies should seek to understand the effect walking in the "wild" has on smartwatch text entry.

KEY POINTS

- Trace input provides the most positive user experience for typing on a smartwatch while both standing and walking.
- Participants reported poor user experience when handwriting using the dynamic writing area of the Google Handwriting app.
- Collecting objective and subjective metrics is essential for assessing the complete user experience of text input.
- Evaluating typing performance in a naturalistic environment is important for generalizing user experience beyond the laboratory.

ORCID iD

Barbara S. Chaparro https://orcid.org/0000-0002-9140-059X

SUPPLEMENTAL MATERIAL

The online supplemental material is available with the manuscript on the *HF* website.

REFERENCES

Arif, A. S., Iltisberger, B., & Stuerzlinger, W. (2011, November). Extending mobile user ambient awareness for nomadic text entry. In *Proceedings of the 23rd Australian Computer-Human Interaction Conference* (pp. 21–30). ACM.

Arif, A. S., & Mazalek, A. (2016). A survey of text entry techniques for smartwatches. In Masaaki. Kurosu (Ed.), *Human-Computer Interaction. Interaction Platforms and Techniques: 18th International Conference, HCI International, Proceedings, Part II-Volume 9732* (pp. 255–267). Springer-Verlag, New York, Inc.

Baudisch, P., & Chu, G. (2009). Back-of-device interaction allows creating very small touch devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1923–1932). ACM.

Bergstrom-Lehtovirta, J., Oulasvirta, A., & Brewster, S. (2011). The effects of walking speed on target acquisition on a touchscreen interface. *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, *2011*, 143–146.

Brooke, J. (2013). SUS: A retrospective. *Journal of Usability Studies*, *8*, 29–40.

Castellucci, S. J., & MacKenzie, I. S. (2008). Graffiti vs. unistrokes: an empirical comparison. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 305–308). ACM.

Castellucci, S. J., & MacKenzie, I. S. (2011). Gathering text entry metrics on Android devices. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems* (pp. 1507–1512). ACM.

Chaparro, B. S., He, J., Turner, C., & Turner, K. (2015). Is touch-based text input practical for a smartwatch? In *HCI International 2015-posters' extended abstracts* (pp. 3–8). Springer.

Clawson, J., Starner, T., Kohlsdorf, D., Quigley, D. P., & Gilliland, S. (2014). Texting while walking: an evaluation of mini-QWERTY text input while on-the-go. In *Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices & Services* (pp. 339–348). ACM.

Connor, M. (1995). Handwriting performance and GCSE concessions. *Handwriting Review*, *9*, 7–21.

Conradi, J., Busch, O., & Alexander, T. (2015). Optimal touch button size for the use of mobile devices while walking. *Procedia Manufacturing*, *3*, 387–394. https://doi.org/10.1016/j.promfg.2015.07.182

Costagliola, G., De Rosa, M., & Fuccella, V. (2017). Handwriting on smartwatches: An empirical investigation. *IEEE Transactions on Human-Machine Systems*, *47*, 1100–1109. https://doi.org/10.1109/THMS.2017.2754938

Darbar, R., Dash, P., & Samanta, D. (2016). ETAO Keyboard: Text input technique on smartwatches. *Procedia Computer Science*, *84*, 137–141. https://doi.org/10.1016/j.procs.2016.04.078

Dingwell, J. B., Cusumano, J. P., Cavanagh, P. R., & Sternad, D. (2001). Local dynamic stability versus kinematic variability of continuous overground and treadmill walking. *Journal of Biomechanical Engineering*, *123*, 27–32. https://doi.org/10.1115/1.1336798

Dutton, K. P. (1990). Writing under examination conditions: Establishing a baseline. *Professional Development Initiatives*, *1989–90*, 189–210.

Funk, M., Sahami, A., Henze, N., & Schmidt, A. (2014). Using a touch-sensitive wristband for text entry on smart watches. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems* (pp. 2305–2310). ACM.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology*, *52*, 139–183.

Hong, J., Heo, S., Isokoski, P., & Lee, G. (2015). SplitBoard: A simple split soft keyboard for wristwatch-sized touch screens. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 1233–1236). ACM.

Hong, J., Heo, S., Isokoski, P., & Lee, G. (2016). Comparison of three QWERTY Keyboards for a Smartwatch. *Interacting with Computers*, *28*, 811–825. https://doi.org/10.1093/iwc/iww003

IDC. (2016). Smartwatch market declines 51.6% in the third quarter as platforms and vendors realign, IDC finds. https://www.idc.com/getdoc.jsp?containerId=prUS41875116

Jeong, H., Kim, H., Kim, R., Lee, U., & Jeong, Y. (2017). Smartwatch wearing behavior analysis: A longitudinal study. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *1*, 60.

Kane, S. K., Wobbrock, J. O., & Smith, I. E. (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* (pp. 109–118). ACM.

Kim, S., Sohn, M., Pak, J., & Lee, W. (2006). One-key keyboard: A very small QWERTY keyboard supporting text entry for wearable computing. In *Proceedings of the 18th Australia Conference on*

Computer-Human Interaction: Design: Activities, Artefacts and Environments (pp. 305–308). ACM.

Kristensson, P. O., & Denby, L. C. (2009). Text entry performance of state of the art unconstrained handwriting recognition: A longitudinal user study. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 567–570). ACM.

Lin, X., Chen, Y., Chang, X. -W., Liu, X., & Wang, X. (2018). Show: Smart handwriting on watches. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *1*, 151.

Lyons, K., Nguyen, D., Ashbrook, D., & White, S. (2012). Facet: A multi-segment wrist worn system. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology* (pp. 123–130). ACM.

MacKenzie, I. S., & Chang, L. (1999). A performance comparison of two handwriting recognizers. *Interacting With Computers*, *11*, 283–297. https://doi.org/10.1016/S0953-5438(98)00030-7

MacKenzie, I. S., & Soukoreff, R. W. (2003). Phrase sets for evaluating text entry techniques. In *CHI'03 Extended Abstracts on Human Factors in Computing Systems* (pp. 754–755). ACM.

MacKenzie, I. S., & Read, J. C. (2007). Using paper mockups for evaluating soft keyboard layouts. In *Proceedings of the 2007 Conference of the Center for Advanced Studies on Collaborative Research* (pp. 98–108). IBM Corp.

MacKenzie, I. S., & Tanaka-Ishii, K. (2010). *Text entry systems: Mobility, accessibility*. Elsevier.

Mizobuchi, S., Chignell, M., & Newton, D. (2005). Mobile text entry: Relationship between walking speed and text input task difficulty. In *Proceedings of the 7th International Conference on Human Computer Interaction With Mobile Devices and Services* (pp. 122–128). ACM.

Oney, S., Harrison, C., Ogan, A., & Wiese, J. (2013). ZoomBoard: A diminutive qwerty soft keyboard using iterative zooming for ultra-small devices. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2799–2802.

Pulvirent, S. (2015). Does Anyone Really Want a Smartwatch?. https://www.bloomberg.com/news/articles/2015-01-08/does-anyone-really-want-a-smart-watch

Schildbach, B., & Rukzio, E. (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* (pp. 93–102). ACM.

Siek, K. A., Rogers, Y., & Connelly, K. H. (2005). Fat finger worries: How older and younger users physically interact with PDAs. In M. F. Costabile & F. Paternò (Eds.), *Human Computer Interaction - INTERACT 2005. Lecture Notes in Computer Science* (Vol. 3585, pp. 267–280). Springer.

Sonaike, I. A., Bewaji, T. A., Ritchey, P., & Peres, S. C. (2016). The ergonomic impact of Swype. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *60*, 1374–1378. https://doi.org/10.1177/1541931213601317

Thompson, J. (2018). A Concise History of the Smartwatch?. https://www.bloomberg.com/news/articles/2018-01-08/a-concise-history-of-the-smartwatch

Turano, K. A., Geruschat, D. R., & Stahl, J. W. (1998). Mental effort required for walking: Effects of retinitis pigmentosa. *Optometry and Vision Science*, *75*, 879–886. https://doi.org/10.1097/00006324-199812000-00010

Turner, C. J., Chaparro, B. S., & He, J. (2017). Text input on a smartwatch QWERTY keyboard: Tap vs trace. *International Journal of Human–Computer Interaction*, *33*, 143–150. https://doi.org/10.1080/10447318.2016.1223265

Turner, C. J., Chaparro, B. S., & He, J. (2018). Texting while walking: Is it possible with a Smartwatch? *Journal of Usability Studies*, *13*.

White, R. M. (1980). *Comparative anthropometry of the hand.* (No NATICK/CEMEL-229). Army Natick Research and Development Labs Ma Clothing Equipment and Materials Engineering Lab.

Zhai, S., & Kristensson, P. O. (2012). The word-gesture keyboard: Reimagining keyboard interaction. *Communications of the ACM*, *55*, 91–101.

Colton J. Turner has a PhD in Human Factors Psychology from Wichita State University. His research interests are in the fields of mobile device text input, smartwatch usability, user experience, and applied human–computer interaction. As a Human Factors Specialist at Priority Designs, a product development consultancy in Columbus, OH, Colton works on a variety of applied research projects across an assortment of domains, namely medical devices and consumer electronics.

Barbara S. Chaparro has a PhD in Experimental Psychology from Texas Tech University. She is a Professor in the Human Factors and Behavioral Neurobiology Department and head of the Research in User eXperience (RUX) Lab at Embry-Riddle Aeronautical University. Her research interests include the study of factors that influence the user experience (UX) of products and systems, the investigation of usability assessment methods, and the efficacy of mobile computing devices.

Jibo He has a PhD in Engineering Psychology from the University of Illinois and bachelor’s degrees from Peking University. He is the director of the Human Automation Interaction Lab at Wichita State University. His research interests include human factors, driver distraction, eye movement, human–machine interaction, usability, and mobile devices.

Date received: June 10, 2019
Date accepted: October 1, 2019