Effect of running on throughput in pointing tasks: a Fitts' law experiment

Fabian van Noort, MSc

29th August 2014

MASTER THESIS

Applied Cognitive Psychology

Effect of running on throughput in pointing tasks: a Fitts' law experiment

Fabian van Noort, MSc

29th August 2014

MASTER THESIS

Applied Cognitive Psychology

Utrecht University

Faculty of Social Sciences

Heidelberglaan 8

3584 CS Utrecht

Supervisor:

dr. S.F. Donker

First examiner:

dr. S.F. Donker

Second examiner:

dr. S. van der Stigchel

Abstract

Studies have shown that walking has a negative effect on input performance on mobile devices. Runners often use their GPS-enabled smartphones to record their exercise, but the effect of running on touch screen input performance has not yet been quantified. This study aims to fill this gap by performing a Fitts' law experiment on a smartphone. Pointing tasks were performed in conditions where participants were either running or stationary, and the device was either handheld in a running armband or worn on the arm in a running armband. Compared to stationary conditions, running conditions showed higher motion time and lower accuracy. Overall input performance was approximately 47% lower in running conditions, compared to stationary conditions. Furthermore, running speed decreased by 26% when interacting with the device. Therefore, common interactions while running should be foreseen by developers and designers of mobile applications. And although easier targets contribute to higher input performance, most importantly, the GUI should be forgiving of user error in the typical context of the user.

Introduction

Interacting with computers through graphical user-interfaces (GUIs) has become part of daily life for many people. Computers have evolved from room-sized machines to portable and even wearable devices. These mobile devices are no longer mere tools, but companions in our daily lives. Consequently, these devices are no longer operated exclusively in stable environments such as homes and offices but in a wide array of contexts. These contexts cause an equally wide array of possible limitations on the user's ability to interact with a mobile device. These limitations are known as situationally-induced impairments and disabilities (SIIDs) and, as Lin, Goldman, Price, Sears and Jacko stressed, need to be recognized and understood before a suitable GUI can be developed [1, 2]. An ill suited GUI often reflects the assumption of cognitive and motor resources being available in the user, but these resources depend heavily on the context of use [3].

Many mobile smartphone applications support or enrich an existing activity. Knowing about that activity provides insights in the user's context, and thus their ability to interact with the mobile device. For instance, many runners use specialized mobile applications on their smartphones to track their running efforts. Therefore, it is to be expected that sports apps such as Runkeeper[1] and Endomondo[2] will be used while running. As for the user's ability to interact with the device in such a context, unfortunately, the task of controlling the device competes for motor resources and attention with the running task. Several studies have shown that even walking at increased pace has a detrimental effect on the user's ability to provide input to a mobile device, which does not bode well for runners [4, 5]. However, the effect of running on the interaction with, e.g., a smartphone has not yet been quantified specifically. This study aims to

[1] `http://www.runkeeper.com`
[2] `http://www.endomondo.com`

fill this gap by performing a Fitts' law experiment.

To provide insight in the current state of research, the next section discusses some key works on interacting with mobile devices in mobile settings.

Related Work

Interaction with a mobile device requires attention. Interplay between cognitive and motor resources, mediated by attention, is well investigated and has become apparent in several studies. A well-known case is the prediction of falls in elderly, based on the ability of walking and talking simultaneously [6]. The observation that an elderly stops walking when conversation started, proved a useful predictor for falls. In addition, numerous studies have shown that postural sway is impacted by the attentional demands of a cognitive load and is specifically apparent in elderly and balance-impaired persons [7–10]. Pellecchia states that, together with previous studies, the result speaks for a substantial effect of attention on motor performance [11].

Brewster reports finding that walking hampered the user's ability to give input to a mobile device in an experiment using a Palm III handheld computer [12]. Participants entered 5-digit strings into the device using either large or smaller on-screen buttons. A comparison was made between a stationary laboratory setting and an outdoor walking setting. In the latter, less data was entered by participants and subjective workload was higher.

By contrast, Kjeldskov and Stage reported finding no difference in input performance between between a stationary and walking laboratory setting referenced to walking in a pedestrian street [13]. Using a Compaq iPAQ PDA, participants completed a set of tasks. The authors report finding that stationary participants, compared to walking participants, experienced less workload while interacting with a mobile device. No difference in workload was reported between the laboratory walking settings and the pedestrian street setting.

Remarkably, the authors stated letting the walking serve as a distracting factor to divide attention, and not being interested in measuring the user's performance on the walking task. However, performance on the walking task might have been neglected by the participants to balance the combined workload of the two tasks. Measurement of walking speed could have provided insight in this potential performance trade-off.

By contrast, measurement of walking speed was not left out by Barnard, Yi, Jacko and Sears [4]. The authors compared walking on a treadmill with walking along a defined path in a controlled environment, under two lighting conditions. Tasks were focussed on retrieving information from a PDA and required some stylus-based input. Although the lower lighting level did cause a higher workload, the authors reported finding only an effect of walking condition and lighting level on the time taken to finish the task. Participants had decreased their walking speed by a third, likely mitigating potential attentional effects on input performance while walking in lower lighting conditions.

Whereas an attentional effect on input performance was not found in the study by Barnard *et al.*, likely due to a compensatory decrease in walking speed, work by Lin *et al.* shows this needs not be the case [1]. The authors report on a study comparing stylus-based target acquisition performance across several experimental conditions. Participants were either seated, walking slowly on a treadmill, walking fast on a treadmill, or walking along an obstacle course. Slow and fast walking was defined as 80% and 120%, respectively, of individual participants' comfortable walking speed. As the authors reported, input performance suffered as a function of walking speed. This was especially true when walking along the obstacle course, even though participants had made use of their freedom to reduce their walking speeds.

Similarly, Schildbach and Rukzio report on an experiment comparing performance on a target acquisition task and a reading task [14]. Different sizes of targets and text were used. Participants were instructed to walk along a predefined path. The authors reported participants needed forty percent more time for target acquisition when walking, compared to being stationary. The reported results also showed that error rates for the walking condition increased by seven percent compared to the stationary condition. While walking, the use of larger targets resulted in a twenty-five percent decrease in error rate, but target selection times remained elevated. When given the liberty, participants reduced their walking speeds by about twenty-five percent, regardless of target size. In another target acquisition experiment, Bergstrom-Lehtovirta, Oulasvirta and Brewster found that target acquisition times remained stable, however, increased walking speed was paired with increased error rates [5].

These experiments used several metrics for the quantification of input performance. Metrics such as error-rates, number of correct inputs, target selection times and total trial times were commonly used as quantifiers for input performance. Surprisingly, a very suitable measure was used very little in the discussed works. The Fitts' law paradigm describes throughput as an index of performance, quantified in bits per second [15]. This paradigm, based on target acquisition tasks also known as pointing tasks, has been widely accepted among HCI researchers and has been verified for a wide range of conditions [16, 17]. Lin *et al.* show that the throughput measure also facilitates comparison of input performances to mobile devices. Using tasks from the Fitts' law paradigm, the present study aims to quantify the effect of running on input performance on a mobile device. For further narrative on the Fitts' law paradigm, and overview on many years of its application, the reader is referred to the work of Soukoreff and MacKenzie [17].

Method

By far, the most common way of providing input to a mobile device is tapping a finger-based touch screen. To quantify the effect of running on input performance on mobile devices, a pointing task experiment as familiar from the Fitts' law paradigm was performed on a mobile device with such a screen.

Participants

Twenty-four participants (11 females; 13 males) with a mean age of 21.2 years (SD = 1.98) were recruited among Dutch students from the Utrecht University, mainly from the department of Computer Science. All participants had normal or corrected-to-normal vision. Three participants were left-handed. Participants received information about the experiment location via e-mail, including reminders the day before participation. Participants were offered a cookie for their participation, or a credit point. Participants were given a brief description of the experiment on arrival. The researcher indicated that all running of the participant would be under direct supervision of the researcher. It was also made clear that the researcher would always be directly available to repeat any instructions given during the experiment. All participants signed an informed consent form.

Tasks

Pointing tasks, known as Fitts' tasks, were created in an Android smartphone application [17]. Figure 1a shows the home screen of the application with three buttons leading to the pointing tasks. During all pointing tasks, circular targets were indicated on screen on-by-one for the participant to tap on. The next target was indicated directly after the screen was touched by the participant, regardless of the participant having hit or missed the target. The touch point was recorded as the approximate centroid of contact area between the finger and the screen, at the moment the finger landed on the screen. At the end of each task, a screen was shown stating that the end of that task was reached and that the researcher would attend to the participant. This screen also included a button allowing return to the home screen of the application.

Training task

A training block of Fitts' tasks was used to let participants familiarize with interacting with the device. The training task consisted of 24 trials of a two-dimensional Fitts' law pointing task (see Figure 1b). The circular array of six targets consisted of grey circles, the target that should be tapped was filled black. Only one configuration of targets, a combination of a target width W and movement distance D, was presented. This choice was made as the intention of the training was to familiarize the participant with using the device in a specific condition, not necessarily to mitigate learning effects across all indices of difficulty. W was 40 pixels and D was 500 pixels for all trials. Performance of the participant was not logged during the training task.

Finger input calibration task

Based on recommendations made by Bi, Li and Zhai, a finger input calibration task was created to determine a baseline of touch point distribution [18]. Single targets were displayed as filled 40 pixel wide black circles in random positions near the center of the screen. This target width was chosen subjectively as to yield a target that is small, yet easily legible in the most challenging condition. The task consisted of 20 trials (see Figure 1c). No feedback on deviation from target, such as 'hit' or 'miss', was given because only the distribution of touch points is of interest for the present study.

Fitts' law pointing task

The experimental task consisted of two-dimensional Fitts' law pointing tasks (see Figure 1d). The targets were shown in 18 configurations, as two versions of nine different nominal indices of difficulty. The indices of difficulty and target configurations were chosen as a large representative range of what may be encountered in practice, following recommendations by Soukoreff *et al.* [17]. The dimensions of the screen, combined with the target sizes that were expected to be legible, provided an upper limit to index of difficulty of 5.85 bits. Values of index of difficulty, ID, were calculated using the Shannon formulation and were rounded to two decimals[3] (see Table 1). The order of the 18 target configurations in the Fitts' task was randomized separately for each participant and condition. For each configuration of targets, 12 trials were executed. The arrays of targets were displayed in the same manner as in the training task.

Apparatus

Participants performed the tasks of the experiment on a smartphone. The device was a Sony LT26i Xperia S smartphone, running on Android OS version 4.1.2. The capacitive LCD touch screen had a diagonal size of 4.3 inches and a resolution of 1280 by 720 pixels[4]. Its brightness was set to maximum at all times. As is common for runners, the device was kept in a generic running armband which was worn on the non-dominant upper arm during the experiment (see Figure 2).

The location of the participant was tracked using GPS with the Runkeeper application for iOS on an Apple iPhone 5C. This device was kept in an identical, separate generic running armband. The armband with the GPS tracker was worn on the dominant upper arm.

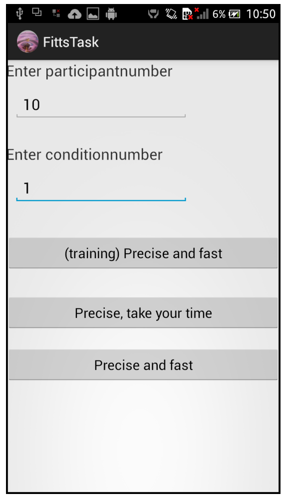
[3] The 12 unrounded unique values were used during analysis.
[4] 1 cm on screen equals 135 pixels.

D (pixels)	*W (pixels)*	*ID (bits)*
200	40.00	2.58
200	70.00	1.95
260	26.00	3.46
300	16.00	4.30
300	20.00	4.00
300	40.00	3.09
400	80.00	2.58
400	140.00	1.95
428	10.00	5.45
496	16.00	5.00
500	50.00	3.46
560	30.00	4.30
600	10.00	5.93
600	40.00	4.00
600	80.00	3.09
620	20.00	5.00
682	12.00	5.85
682	16.00	5.45

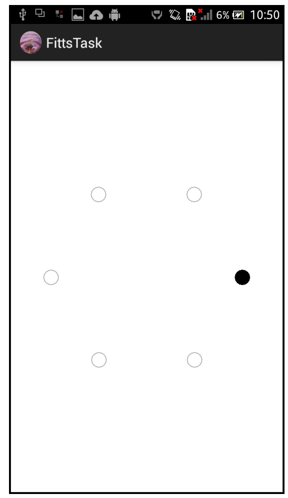
Table 1: The implemented configurations of targets for the Fitts' law pointing task.

Design

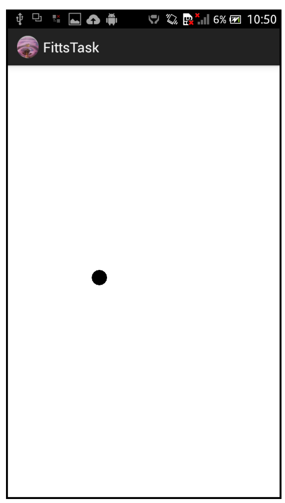
Runners that use their smartphones during their exercise commonly carry the device in a specialized armband. The armband allows for touch screen input through transparent plastic. Comparing this usage to otherwise common mobile phone use, two factors are clearly abnormal: the motion of the participant and the placement of the device. Taking this into consideration, four conditions were examined in a 2x2 within-subjects design. That is, participants were either stationary with the device held in hand (C1), stationary with the device worn on the non-dominant upper arm (C2), running with the device held in hand (C3), or running with the device worn on the non-dominant upper arm (C4). The device was carried in the running armband in all conditions. With 24 possible unique orders of four conditions, and an equal amount of participants, the orders of the four conditions were counterbalanced. For each condition,



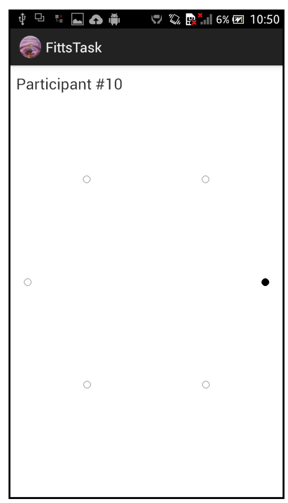
(a) Home screen showing the buttons that lead to the tasks.



(b) The training block of Fitts' tasks.



(c) The finger input calibration task.



(d) The experimental Fitts' task.

Figure 1: Screenshots of the experimental smartphone application. During the experiment, the texts were in Dutch.

the training block was performed first, the finger input calibration task second, and the experimental task third. Movement directions for the trials within the

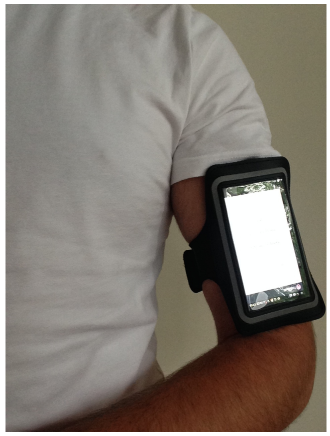


Figure 2: The smartphone worn on the left upper-arm in a running armband.

training task and experimental task were balanced as described by Soukoreff *et al.* [17].

Procedure

The experiment took place in a natural setting, a cleared hockey field, based on recommendations for comparison studies given by Barnard *et al.* [4]. This setting allowed participants to naturally regulate their running speed, creating higher ecological validity compared to the usage of treadmills. With field dimensions of approximately 91.4 meters by 55 meters, plenty of safe and unobstructed running space was available.

Participants were explicitly allowed to adjust the fit of the GPS tracker armband at their own discretion, to avoid interference of the armband in the execution of the pointing tasks. Participants were then instructed to run for about one minute along the inside edge of the field, as to establish a comfortable baseline running speed. Participants were instructed to choose a speed at which they expected to be able to perform constantly for the duration of the experiment. In all cases, the researcher ran along with the running participant for observational and safety purposes. To avoid influencing the running speed of the participant, the researcher never ran in direct sight of the participant. After this measurement of baseline speed, the participant was introduced to the device in the armband on which the tasks were to be executed, and to the tasks themselves. It was also made clear that brief breaks were allowed during the tasks. For the first handheld condition and for the first condition with the armband worn on the non-dominant upper arm, the participant was allowed to find preference for a specific way of holding and controlling the device during the training task. These preferences were then noted by the researcher to ensure consistency across the running and stationary conditions. In the conditions where the pointing tasks were executed with the device worn on the arm, the researcher assisted the participants in putting on the armband on the non-dominant upper arm. After all tasks were completed in all conditions, participants were thanked for their participation.

Dependent variables

Motion time on the experimental tasks was measured by the device as the time between the current and previous touch point in milliseconds. For the finger input calibration task and the experimental task, bivariate touch point deviation in pixels (SD_{xy}) was recorded by the device. This measure has shown to provide a better model fit, compared to usage of a univariate deviation (SD_x) [19]. Positive values on these axes represent overshoot in the direction of motion. Throughput, in bits per second, was measured by combining the index of difficulty of the pointing tasks and motion time. Running speed, in meters per second, was measured using GPS and was normalized against individual baseline.

Statistical analysis

Means of motion time, bivariate touch point distribution, and throughput were first of all calculated within participants, secondly across participants. The mean motion time, bivariate touch point deviation, and throughputs of participants were subjected to two-way repeated measures analysis of variance. Normalized mean running speed of participants was subjected to one-way repeated measures analysis of variance. All statistical analysis were performed using IBM SPSS Statistics for Macintosh, Version 22.0.

Results

Each session took approximately 40 minutes. The mean total travelled distance by the participants was 1.26 kilometers ($SD = 0.28$).

Adjustment of Data

Trials were marked as outliers, and removed from analysis, if distance from the touch point to the center of the target (on x or y axis) deviated more than three standard deviations from the mean for that specific participant, condition, and target configuration. The bivariate touch point deviation (SD_{xy}) was used as standard deviation. From the finger input calibration task, 11 out of 2016 trials were marked as outliers. As Fitts' law is intended for rapid aimed movements, trials with motion time exceeding 2000 milliseconds were also marked as outliers. Furthermore, trials with motion time deviating more than three standard deviations from the mean for that specific participant, condition, and target configuration were also marked as outliers. From the experimental task, 408 out of 20736 trials were marked as outliers.

By experiment design, the SD_{xy} in the finger calibration task would always be of lower value than in the experimental task. The SD_{xy} from the finger input calibration task is supposed to indicate maximum pointing precision. However, this assumption was violated in 39 (out of 24x4) cases. The measure was redefined as the lowest SD_{xy} occurring in either the finger input calibration task or the experimental task, for that participant and condition. If the SD_{xy} value for maximum pointing precision was taken from a set of experimental task trials, instead of the finger input calibration task trials, those trials were removed from further analysis.

Fitts' law

Table 2 shows the mean motion time for each target configuration and each condition in the experimental task. To analyse potential difference in motion time in the experimental tasks between the four conditions, a two-way analysis of variance with repeated measures was used (see Figure 3). Mauchly's test of sphericity indicated no violation. Motion time was significantly affected by participant motion in the experiment, $F(1, 23) = 17.97$, $p < .001$. Motion time in the running conditions ($M = 616.33$, 95% CI $[600.49, 632.17]$) was significantly higher compared to the stationary conditions ($M = 551.41$, 95% CI $[535.57, 567.25]$). Motion time was also significantly affected by device placement, $F(1, 23) = 20.97$, $p < .001$. Motion time in the conditions with the device worn on the arm ($M = 628.11$, 95% CI $[608.13, 648.09]$) was significantly higher compared to the conditions with the device held in hand ($M = 539.63$, 95% CI $[519.65, 559.61]$). The interaction between participant motion and device placement was not significant, $F(1, 23) = 3.85$, $p = .062$. Linear regressions for the four conditions provided the following Fitts' law models.

$$MT_{C1} = 312.184 + 53.319 \cdot ID$$

$$MT_{C2} = 285.135 + 78.421 \cdot ID$$

$$MT_{C3} = 314.261 + 63.456 \cdot ID$$

$$MT_{C4} = 318.130 + 90.043 \cdot ID$$

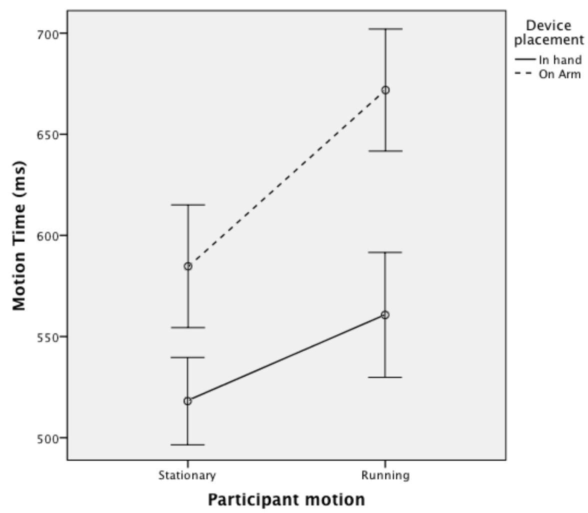


Figure 3: Mean motion time in the experimental tasks, for each condition. Error bars represent 95% confidence interval.

Motion time increased approximately linearly with the index of difficulty of the pointing task in all four conditions ($R^2_{C1} = .872$; $R^2_{C2} = .851$; $R^2_{C3} = .862$; $R^2_{C4} = .902$), indicating that the data gathered in this experiment may be evaluated using Fitts' law as intended.

As the main purpose of this study is comparison of input performance between several conditions, rather than prediction of motion time per se, the constructed Fitts' law models are of little further interest. As described, throughput is the suitable measure in this case. For accurate quantification of throughput, however, measures of just motion time will not suffice. A speed-accuracy trade-off may allow motion time to decrease at the cost of accuracy.

Dimensions (pixels)		*Mean Motion Time (ms)*			
D	*W*	*C1*	*C2*	*C3*	*C4*
200	40	464.60	474.82	482.38	557.80
200	70	457.05	476.11	456.01	540.58
260	26	479.04	504.37	486.56	605.02
300	16	499.02	534.31	530.02	646.01
300	20	474.04	519.62	532.92	624.73
300	40	459.02	489.46	487.76	547.00
400	80	443.16	493.43	478.51	542.42
400	140	441.21	476.86	463.17	517.62
428	10	573.66	663.48	653.36	788.43
496	16	515.08	578.00	601.28	774.49
500	50	468.39	543.96	531.53	605.46
560	30	531.40	597.39	540.71	680.53
600	10	626.87	756.41	699.73	834.98
600	40	517.88	588.67	542.26	691.66
600	80	502.52	587.42	539.58	611.70
620	20	552.60	656.87	621.15	795.50
682	12	676.11	794.47	740.45	887.82
682	16	637.03	710.01	673.63	809.26
Grand Mean (SD)		517.70 (70.11)	580.31 (99.49)	558.95 (86.14)	670.06 (117.25)

Table 2: Mean motion time for each target configuration in the experimental task, for each condition.

Distribution of Touch points

The input performance of the participants is not only reflected by motion time, actual touch point performance needs to be considered as well. For each condition, Table 3 shows the touch point distribution, SD_{xy}, of the finger input calibration tasks and of each target configuration of the experimental task. Before complementing the calculations of throughput with the distribution of the touch points, the touch point performance in each condition is subjected to analysis.

As was done for motion time, a two-way analysis of variance with repeated measures was used to analyze potential difference in SD_{xy} in the experimental tasks between the four conditions (see Figure 4). Mauchly's test of sphericity indicated no violation. SD_{xy} was significantly affected by participant motion in the experiment, $F(1, 23) = 55.50$, $p < .001$. The SD_{xy} in the running conditions ($M = 66.90$, 95% CI $[62.20, 71.60]$) was significantly higher compared to the stationary conditions ($M = 33.05$, 95% CI $[28.34, 37.75]$). SD_{xy} was also significantly affected by device placement, $F(1, 23) = 30.95$, $p < .001$. SD_{xy} in the conditions with the device worn on the arm ($M = 56.27$, 95% CI $[53.93, 58.61]$) were significantly higher compared to the conditions with the device held in hand ($M = 43.68$, 95% CI $[41.34, 46.02]$). These main effects were qualified by a significant interaction, $F(1, 23) = 18.33$, $p < .001$. Analysis of simple main effects revealed that device placement had an effect in both stationary and running conditions $F(1, 23) = 6.16$, $p = .021$; $F(1, 23) = 26.9$, $p < .001$.

Adjusted Index of Difficulty

As mentioned, input performance cannot be adequately quantified with only measurement of motion time. To take the touch point performance into account, Soukoreff *et al.* advise using an alternative to the Shannon formulation of ID, ID_e. This formulation calculates an effective target width using the distribution of touch points, $W_e = \sqrt{2\pi e}\sigma$. Using W_e results in the following formulation of index of difficulty:

$$ID_e = \log_2\left(\frac{D}{\sqrt{2\pi e}\sigma} + 1\right)$$

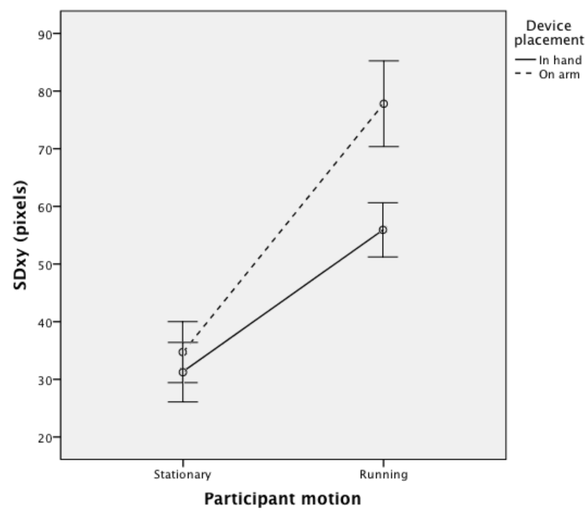


Figure 4: Mean SD_{xy} in the experimental tasks, for each condition. Error bars represent 95% confidence interval.

Bi *et al.* proposed an alternative to the generally accepted ID_e formulation. The authors describe a formulation specialized in modeling experimental data from $f$inger-based human-computer interactions on touch screens, hence their choice of model name for FFitts law and ID_f. The proposed model is based on the hypothesis that the distribution of touch points is the sum of two independent, normally distributed, variables. The authors define the two components as (1) a relative component reflecting the speed-accuracy trade-off and (2) an absolute component reflecting the absolute precision of the motor system and the actual medium (e.g. the finger). The authors arrive at the following formulation:

$$ID_f = \log_2 \left(\frac{D}{\sqrt{2\pi e(\sigma^2 - \sigma_a^2)}} + 1 \right)$$

In the ID_f formulation, σ^2 reflects the distribution of touch points SD_{xy} in the experimental task. The σ_a^2 reflects the maximum precision, or lowest SD_{xy}, on pointing tasks. The authors report finding that ID_f shows a better model fit

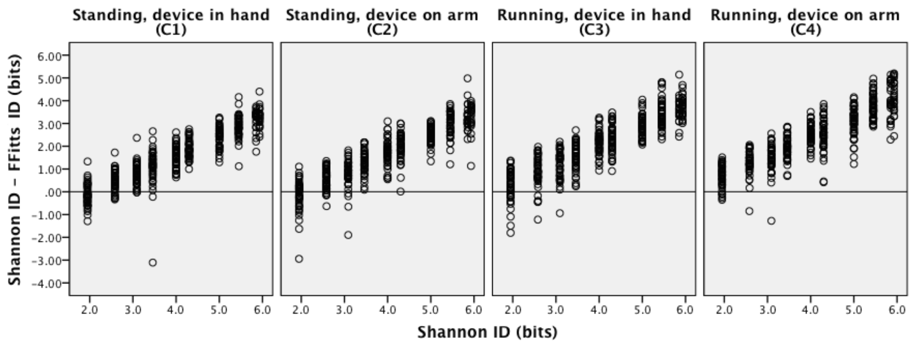
Dimensions (pixels)		*Touch point distribution SD_{xy} (pixels)*			
D	*W*	*C1*	*C2*	*C3*	*C4*
random	40	15.18	17.05	31.59	41.52
200	40	25.79	27.84	47.44	67.13
200	70	27.25	29.19	51.09	71.84
260	26	29.13	29.71	46.91	61.62
300	16	25.42	30.68	51.97	72.61
300	20	26.12	32.64	54.54	72.43
300	40	31.48	31.38	51.29	70.43
400	80	30.42	33.01	57.50	72.75
400	140	30.59	34.02	53.13	75.58
428	10	30.32	34.24	58.73	87.13
496	16	31.13	33.92	57.86	73.19
500	50	33.01	35.09	55.94	81.44
560	30	33.36	36.77	50.55	74.89
600	10	33.57	36.20	57.85	100.89
600	40	32.70	39.41	58.09	84.07
600	80	34.97	37.43	61.34	79.26
620	20	35.88	38.69	54.89	75.00
682	12.00	36.13	44.77	63.34	93.61
682	16.00	37.22	40.10	75.35	90.14
Grand Mean (SD)		31.36 (3.62)	34.73 (4.34)	55.99 (6.59)	78.00 (9.90)

Table 3: Mean SD_{xy} for each target configuration in the experimental task, for each condition.

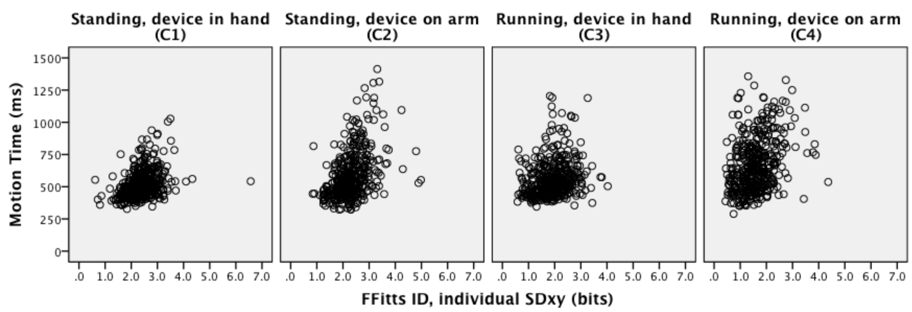
compared to both ID_e and the uncorrected Shannon ID, which would make ID_f especially suitable for this study. The ID_f implementation described by Bi *et al.*, however, pools the SD_{xy} across the target configurations and participants. Soukoreff *et al.* specifically advise against this approach as it fails to make use of within-subject variability.

With this in mind, the ID_f formulation was implemented using the SD_{xy} of individual conditions, participants, and target configurations. As each participant, condition, and target configuration has a specific SD_{xy}, many unique values of $\sigma^2 - \sigma_a^2$ arise. This results in as many values of ID_f. For each calculated value of ID_f, the difference between the original ID and the calculated

ID_f is shown in Figure 5a. Figure 5b shows the ID_f values in relation to motion time, revealing a less strong linear relation to motion time than ID ($R^2_{C1} = .127$; $R^2_{C2} = .196$; $R^2_{C3} = .051$; $R^2_{C4} = .102$).



(a) Difference between ID and ID_f, for each value of ID.



(b) Motion time from the experimental task in relation to ID_f, for each condition.

Figure 5: The relation between ID, ID_f, and motion time.

Throughput

With the calculated ID_f values, the index of difficulty is now adjusted based on the actual touch point performance. Average throughput TP may then be calculated for comparison of input performance between the four conditions using the following equation[5]:

$$TP = \frac{ID_f}{MT}$$

As was done for touch point distribution, a two-way analysis of variance with repeated measures was used to analyze potential difference in throughput in the experimental tasks between the four conditions (see Figure 6). Mauchly's test of sphericity indicated no violation. Throughput was significantly affected by participant motion in the experiment, $F(1, 23) = 99.88$, $p < .001$. Throughput in the running conditions ($M = 3.02$, 95% CI $[2.87, 3.16]$) was significantly lower compared to the stationary conditions ($M = 4.43$, 95% CI $[4.28, 4.58]$). Throughput was also significantly affected by device placement, $F(1, 23) = 43.138$, $p < .001$. Throughput in conditions with the device worn on the arm ($M = 3.31$, 95% CI $[3.18, 3.44]$) was lower compared to conditions with the device held in hand ($M = 4.13$, 95% CI $[4.00, 4.27]$). These main effects were qualified by a significant interaction, $F(1, 23) = 6.92$, $p = .015$. Analysis of simple main effects revealed that device placement had an effect in both stationary and running conditions $F(1, 23) = 13.71$, $p = .001$; $F(1, 23) = 56.11$, $p < .001$. The running condition with the device worn on the non-dominant upper arm, C4, showed the least throughput ($M = 2.50$, 95% CI $[2.33, 2.66]$).

Running speed

The baseline running speed of participants ($M = 2.92$, $SD = 0.59$) was used to create relative values, in percentages, for both the running conditions. One-way analysis of variance with repeated measures was used to analyze potential difference in running speed during the experimental tasks between the two running

[5]The equation was implemented using the mean-of-means approach, as described by Soukoreff *et al.* [17].

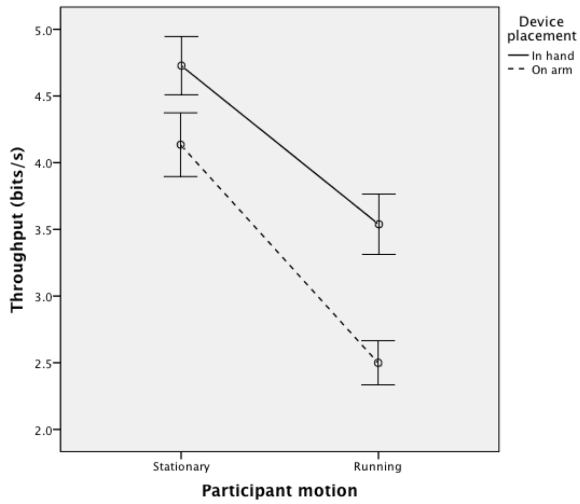


Figure 6: Mean throughput in the experimental tasks, for each condition.

conditions and the baseline. Mauchly's test of sphericity indicated no violation. Running speed was significantly affected by interacting with the device, $F(1,46) = 39.04$, $p < .001$. Pairwise comparison with Bonferroni adjustment showed that running speed (percentage relative to individual baseline) with the device on the arm ($M = 74.06$, 95% CI $[70.82, 77.30]$) was lower compared to the running speed with the device held in hand ($M = 83.67$, 95% CI $[80.52, 86.83]$) ($p = .001$).

Discussion and conclusions

The present study set out with the aim of quantifying the effect of running on input performance on mobile devices. To find this effect, a Fitts' law experiment was performed in several conditions. Participants were either stationary or running, and either held a mobile device in hand or wore it on the non-dominant upper arm. Motion time, touch point distribution, throughput and running speed were measured in all conditions.

Motion time was approximately 12% higher in the running conditions, compared to the stationary conditions. Schildbach *et al.* reported similar effects of walking on motion time [14]. Interestingly, Lin *et al.* reported not finding this effect [1]. Though Lin *et al.* did not find an effect of walking on motion time, they did report finding an effect of walking on pointing accuracy which is consistent with findings of Schildbach *et al.* and Bergstrom-Lehtovirta *et al.* [5, 14]. As Lin *et al.* also noted, this may have been due to the nature of the pointing tasks used in that study. The tasks that were used were not continuous pointing tasks, but consisted of separate and brief moments of interaction. This may have allowed for sufficient momentary focus of attention on the pointing task, without increasing the motion time within that pointing task. The present study found an effect on accuracy as well. Bivariate touch point distribution in the running condition approximately doubled, compared to the stationary conditions. These varying reports of either motion time or accuracy being affected by walking, or both, is a reflection of speed-accuracy trade-off. That is, motion times may increase in favor of accuracy. This underlines the importance of using throughput as an encompassing measure for input performance. The present study found that throughput had decreased by approximately 47% in the running conditions compared to the stationary conditions. Throughput was calculated by using the ID_f formulation as proposed by Bi, Li and Zhai, taking both motion time and touch point distribution into account [18]. As motion time and touchpoint deviation were higher in running conditions, and higher still in conditions with the device worn on the non-dominant upper arm, condition four showed the lowest input performance.

The combination of running and interaction with a mobile device did not only affect input performance. Interacting with the device worn on the arm, while running, caused a 26% decrease in running speed compared to baseline.

Running speed was maintained better in the condition with the device held in hand, showing a 16% decrease compared to baseline. Such decreasing speeds, albeit while walking, were also reported by Lin *et al.*, Barnard *et al.*, Schildbach *et al.* and Bergstrom-Lehtovirta *et al.* [1, 4, 5, 14].

The results from the experiment indicated that Fitts' law is suited for prediction of motion time in the demanding conditions that were used, corroborating and extending the findings of Lin *et al.* [1]. However, it must be noted that the classic linear relation between motion time and index of difficulty was only witnessed for the uncorrected values of ID, and Lin *et al.* did not specify which formulation of ID was used in their analysis. Motion time increased as a function of target difficulty, but accuracy suffered as a function of target difficulty too. This caused even greater downward corrections for higher values of ID, resulting in a strong non-uniform compression towards lower values of ID. As a result, contrary to the findings of Bi *et al.*, the corrected values of ID_f showed poor correlation with motion time in all four conditions [18]. The non-uniform compression appears to have taken away much of the predictive value of the corrected index of difficulty, ID_f. The large difference in correlation between the work of Bi *et al.* and the present study may possibly be attributed to the fairly limited range of target configurations and values of ID (1.92 to 3.75 bits) that was used by Bi *et al.* The non-uniform compression witnessed in the present study is fairly limited in that specific range of ID. As ID_f has taken the actual touch point performance into account, the mean throughput is suitable for comparison of conditions nonetheless. However, only the uncorrected values of ID should be used for predictions of motion time.

Though it is the view of the researcher that the experiment design and analysis were adequately robust, it is important to bear in mind that this study did not control for experience with running, nor for experience with use of touch screens. Also, ambient lighting and temperature could not be controlled in this natural set-up, potentially having caused more fatigue-effects in some participants than in others.

Runners' best choice for interaction style would likely depend on the complexity of the interaction. Best input performance requires handheld control, but this would either mean taking off and putting on the running armband or removing the device from the armband and placing it back. Obviously, this also takes time and attention, especially if being stationary is not an option. For simple interactions, the device may well be left in the running armband. Generally, common interactions while running should simply be foreseen by developers and designers of mobile applications. Though larger buttons obviously contribute to higher input performance, this does not mean that all buttons must be extremely large. A balance should be sought between the size of the interaction elements and the number of elements to be shown, given the screen dimensions. The GUI should always be forgiving of user error in the typical context of the user. As long as the limitations on the user's ability to interact with the device are recognized and understood during development, the application may become a pleasurable part of daily life for many people.

References

[1] M. Lin, R. Goldman, K. J. Price, A. Sears and J. Jacko, “How do people tap when walking? an empirical investigation of nomadic data entry”, *International Journal of Human-Computer Studies*, vol. 65, no. 9, pp. 759–769, 2007.

[2] A. Sears, M. Lin, J. Jacko and Y. Xiao, “When computers fade: pervasive computing and situationally-induced impairments and disabilities”, in *HCI International*, vol. 2, 2003, pp. 1298–1302.

[3] S. K. Kane, J. O. Wobbrock and I. E. Smith, “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*, ACM, 2008, pp. 109–118.

[4] L. Barnard, J. S. Yi, J. A. Jacko and A. Sears, “An empirical comparison of use-in-motion evaluation scenarios for mobile computing devices”, *International Journal of Human-Computer Studies*, vol. 62, no. 4, pp. 487–520, 2005.

[5] J. Bergstrom-Lehtovirta, A. Oulasvirta and S. Brewster, “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*, ACM, 2011, pp. 143–146.

[6] L. Lundin-Olsson, L. Nyberg and Y. Gustafson, “‘‘stops walking when talking” as a predictor of falls in elderly people”, *The Lancet*, vol. 349, no. 9052, p. 617, 1997.

[7] M. C. Dault, A. C. Geurts, T. W. Mulder and J. Duysens, “Postural control and cognitive task performance in healthy participants while bal-

ancing on different support-surface configurations", *Gait & posture*, vol. 14, no. 3, pp. 248–255, 2001.

[8] G. Andersson, J. Hagman, R. Talianzadeh, A. Svedberg and H. C. Larsen, "Effect of cognitive load on postural control", *Brain research bulletin*, vol. 58, no. 1, pp. 135–139, 2002.

[9] M. Woollacott and A. Shumway-Cook, "Attention and the control of posture and gait: a review of an emerging area of research", *Gait & posture*, vol. 16, no. 1, pp. 1–14, 2002.

[10] M. A. Riley, A. A. Baker and J. M. Schmit, "Inverse relation between postural variability and difficulty of a concurrent short-term memory task", *Brain Research Bulletin*, vol. 62, no. 3, pp. 191–195, 2003.

[11] G. L. Pellecchia, "Postural sway increases with attentional demands of concurrent cognitive task", *Gait & posture*, vol. 18, no. 1, pp. 29–34, 2003.

[12] S. Brewster, "Overcoming the lack of screen space on mobile computers", *Personal and Ubiquitous Computing*, vol. 6, no. 3, pp. 188–205, 2002.

[13] J. Kjeldskov and J. Stage, "New techniques for usability evaluation of mobile systems", *International journal of human-computer studies*, vol. 60, no. 5, pp. 599–620, 2004.

[14] B. Schildbach and E. Rukzio, "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*, ACM, 2010, pp. 93–102.

[15] P. M. Fitts, "The information capacity of the human motor system in controlling the amplitude of movement.", *Journal of experimental psychology*, vol. 47, no. 6, p. 381, 1954.

[16] R. Plamondon and A. M. Alimi, “Speed/accuracy trade-offs in target-directed movements”, *Behavioral and Brain Sciences*, vol. 20, no. 02, pp. 279–303, 1997.

[17] R. W. Soukoreff and I. S. MacKenzie, “Towards a standard for pointing device evaluation, perspectives on 27 years of fitts' law research in hci”, *International Journal of Human-Computer Studies*, vol. 61, no. 6, pp. 751–789, 2004.

[18] X. Bi, Y. Li and S. Zhai, “Ffitts law: modeling finger touch with fitts' law”, in *Proceedings of the 2013 ACM annual conference on Human factors in computing systems*, ACM, 2013, pp. 1363–1372.

[19] J. O. Wobbrock, E. Cutrell, S. Harada and I. S. MacKenzie, “An error model for pointing based on fitts' law”, in *Proceedings of the SIGCHI conference on human factors in computing systems*, ACM, 2008, pp. 1613–1622.