Throughput and Effective Parameters in Crossing

Nobuhito Kasahara
Meiji University
Nakano-ku, Tokyo, Japan
k.nobu00714@outlook.jp

Yosuke Oba
Meiji University
Nakano-ku, Tokyo, Japan
bonscow@gmail.com

Shota Yamanaka
Yahoo Japan Corporation
Chiyoda-ku, Tokyo, Japan
syamanak@yahoo-corp.jp

Wolfgang Stuerzlinger
Simon Fraser University
Vancouver, BC, Canada
w.s@sfu.ca

Homei Miyashita
Meiji University
Nakano-ku, Tokyo, Japan
homei@homei.com

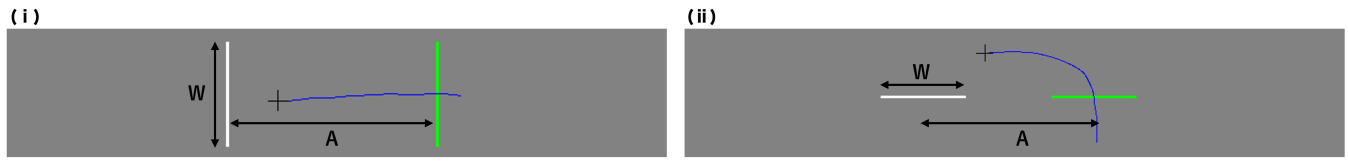

Figure 1: (i) The C/DC task condition used in Experiment 1, where the crossing areas of targets face each other. (ii) The C/AC task condition used in Experiment 2, where the target crossing areas are horizontally aligned. In both tasks, the cursor was shown as a black cross at the same position as the stylus tip. The cursor trajectory was shown as a blue line.

ABSTRACT

In pointing, throughput *TP* is used as a performance metric for the input device and operator. Based on the calculation of *effective parameters* (width W_e and amplitude A_e), *TP* should be independent of the speed-accuracy tradeoff. To examine the validity of *TP* and effective parameters for crossing actions, we conducted two experiments using two established crossing tasks. Our results demonstrate that applying effective parameters to Fitts' law model improves the fit to the data for mixed biases in both tasks. Besides, we observed that effective parameters smoothed *TP*s across biases. However, unlike pointing, *TP* was observed to be unstable across *ID*s in one task, while was stable across *ID*s in the other task. Analyzing speed profiles showed that this was likely due to the fact that one of the tasks could be completed with a ballistic movement at low *ID*s, whereas this was impossible for the other task.

CCS CONCEPTS

• **Human-centered computing → HCI theory, concepts and models**.

KEYWORDS

Pen Crossing, human motor performance, Fitts' law, throughput, effective parameters

CHI EA '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9422-2/23/04.
https://doi.org/10.1145/3544549.3585817

ACM Reference Format:
Nobuhito Kasahara, Yosuke Oba, Shota Yamanaka, Wolfgang Stuerzlinger, and Homei Miyashita. 2023. Throughput and Effective Parameters in Crossing. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23), April 23–28, 2023, Hamburg, Germany.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3544549.3585817

1 INTRODUCTION

Pointing tasks are frequently used to measure the performance of input devices and user groups when operating GUIs. However, in GUIs, there also exist trajectory-based tasks, in which users draw strokes. Crossing is a typical example of such tasks (Fig. 1), and many novel interaction techniques have been proposed for it [4, 10, 11, 15].

The movement time *MT* for crossing can be predicted by the same Fitts' law model as for pointing [1]. In pointing, Throughput *TP* is used as a measurement of performance for input devices and user groups. Theoretically, *TP* should be unaffected by Fitts' law's Index of Difficulty *ID* [17]. However, operating with different speed or accuracy biases (*Bias*), the *TP* calculated with the nominal *ID*, or ID_n, using the nominal target width W and amplitude A between targets has been shown to be different for each *Bias* [20]. To resolve this issue, prior work recommended to smooth the difference of the *Bias* by using the ID_e, which is calculated with the effective width W_e and effective amplitude A_e [18]. This enables a fair comparison of performance even when a user group interacts through various devices with different speed-accuracy balances. However, although Fitts' law predicts the *MT* of crossing, the effect of using effective parameters on crossing tasks has not been empirically verified.

To examine the applicability of throughput and effective parameters for crossing, we conducted experiments with two types of crossing tasks under three speed-accuracy biases. If effective parameters can appropriately normalize the speed-accuracy bias, the model fit when analyzing all biases in a mixed manner should then improve compared to the use of nominal parameters, and the *TP*s across all biases should be close to each other.

In Experiment 1, we investigated continuous crossing with a directional constraint (C/DC) task [5] in which the targets were facing each other (Fig. 1 (i)). In Experiment 2, we looked at continuous crossing within an amplitude constraint (C/AC) task [5] where targets were horizontally aligned (Fig. 1 (ii)). Results showed that in both tasks, applying effective parameters to the Fitts' law model improves the fit when analyzing the data across different biases. Besides, we identified that *TP*s were smoothed across biases by calculating *TP* using effective parameters in both tasks. However, we also confirmed that *TP* was not stable across *ID*s in the C/DC task, while it was stable in the C/AC task. This indicated that the range of tasks where *TP* can be used as a measurement of performance in crossing may be limited. We thus also discuss the conditions where *TP* is stable across *ID*s in crossing.

2 RELATED WORK

Fitts' law [12, 16] can predict the movement time *MT* of pointing on the basis of the index of difficulty *ID*, specifically through the nominal *ID* (ID_n):

$$MT = a + b \cdot ID_n,\ with\ ID_n = \log_2 (A/W + 1)\,, \tag{1}$$

where A is the amplitude between targets, W is the target width, and a and b are empirical constants.

In the process of deriving the steering law [1], Accot and Zhai showed that the *MT* of crossing is predicted by Fitts' law. This finding held strongly for various types of crossing tasks [2, 5].

Fitts' law is based on the task parameters W and A, and does not account for the parameters associated with the user's actual behavior. Therefore, replacing A and W with effective parameters enables more accurate prediction of the *MT* [22]:

$$MT = a + b \cdot ID_e,\ with\ ID_e = \log_2 (A_e/W_e + 1)\,, \tag{2}$$

where A_e is the mean movement distance on the task axis, W_e is $4.133 \cdot \sigma$, and σ is the standard deviation of the cursor endpoints [22]. This definition of W_e ensures that 96% of the endpoints fall inside the target boundary, and A_e is the actual movement distance on average on the task axis; both can more accurately represent the user's behavior [22]. To avoid potential confusion, we denote *ID* using the nominal parameters as ID_n hereafter, and use *ID* as a generic term for ID_n and ID_e.

Throughput, *TP* was standardized by ISO9241-9 and is calculated as follows [23].

$$TP = ID/MT \tag{3}$$

In pointing, *TP* is considered to be independent of *ID* [17]. In other words, even if the *ID* changes, the *MT* also changes accordingly and thus *TP* remains (more or less) constant.

Mackenzie and Isokoski conducted a pointing experiment with three *Bias* conditions: speed-emphasis, accuracy-emphasis, and nominal (i.e., neutral) [18]. In the speed- or accuracy-emphasis condition, the participants were asked to change their *MT*s by 10% compared with the nominal condition. There was no significant difference in *TP* using ID_e across different *Bias*es, and thus *TP* could smooth out differences in *Bias*. However, Olafsdottir et al. showed that when the bias is larger than 10%, the invariance of *TP* was not observed [20].

Zhai et al. revealed that using ID_e showed a higher fit when analyzing the data from multiple biases in a mixed manner (called the *Mixed* analysis condition) than using ID_n in pointing [26]. However, for each individual *Bias* (such as Accurate, Neutral, and Fast), the ID_n model showed higher fit than the ID_e model. This indicates that the ID_e model improves the fit for *Mixed* at the expense of a higher prediction accuracy of the ID_n model in each individual *Bias*.

Luo and Vogel tested the applicability of W_e to crossing with direct finger input, found that Finger-Fitts law showed a good fit, and compared it with the results using W and W_e [14]. To our knowledge, this is the only study in which W_e was applied to crossing. The purpose of their study was to compare the fits of ID_n, ID_e, and Finger-Fitts *ID* with a single *Bias*. Still, previous work has not verified whether the effective parameters can smooth the effect of *Bias* on the fit and *TP*.

3 EXPERIMENT 1

We conducted a study of the C/DC crossing task where targets face each other (Fig. 1 (i)). Ten university students joined (3 females and 7 males, mean age 20.7, standard deviation 1.06). All participants were right-handed.

3.1 Apparatus

We used a laptop PC (Intel Core i7-11800H, GeForce RTX 3070 Laptop, 16GB RAM, Windows 10 education), LCD tablet (Wacom Cintiq 27QHD, IPS, 569.7×335.6mm, 2560×1440 pixels), and Wacom stylus. The system was made with Unity and displayed in full screen mode.

3.2 Design

For this experiment, we used a $3_{Bias} \times 3_A \times 5_W$ repeated-measures design. The within-subjects factors were *Bias* (Accurate, Neutral, Fast), amplitude A (46.60 mm, 128.2 mm, 205.0 mm), and width W (1.864 mm, 3.495 mm, 6.524 mm, 11.65 mm, 23.30 mm). A task *set* comprised the 15 combinations of A and W presented in random order, and 21 such sets were performed in each *Bias* condition. The ten participants were randomly divided into two groups of five. Group 1 was tested in the order of Neutral, Fast, and Accurate. Group 2 was tested in the order of Neutral, Accurate, and Fast. This ordering, i.e., the Neutral condition as the first condition, allowed the participants to perform the task more rapidly/slowly in the remaining two *Bias* conditions relative to the first one, which is the same design as used in a previous study [25].

3.3 Procedure

First, the task was explained to the participants, i.e., that they had to perform a stroke from the start target (right) to the end target (left)[1]. The task was then completed by passing the start target from right to left and then passing the end target (also from right to left). During the task, participants had to keep the stylus tip on the screen. While the stylus tip was on the screen, a blue trajectory was displayed, and when the stylus tip was lifted off the

[1]To prevent the target from being occluded by the hand during the task, we restricted participants to right-handed ones and used right-to-left strokes.

screen, the trajectory disappeared. Before starting each *Bias* condition, participants were instructed to perform the task either "as fast and as accurately as possible" for Neutral, "as fast as possible without worrying about mistakes" for Fast, and "as accurately as possible without worrying about time" for Accurate. Once participants passed through the start target, the task was considered to be started; if participants initially passed outside it, they had to consequently try crossing it again. After the task started, participants were required to cross the end target without lifting the stylus from the screen. If the stylus crossed the end target, we recorded a success; otherwise, an error was recorded. Subsequently, appropriate audio and visual feedback was presented depending if the trial was successful or erroneous. If the stylus was lifted in the middle of a trial, the trial had to be restarted by crossing the start target. After each trial ended, releasing the stylus from the screen displayed a button labeled "Next", and the participants needed to tap it to proceed to the next trial.

3.4 Measurement

All positions of the stylus tip during task execution were recorded with a time stamp. The dependent variables were *ER* (Error Rate), *MT* (Movement Time), σ (Standard Deviation of Endpoints), and *TP* (Throughput). *ER* was the percentage of trials in which the cursor passed outside the end target. *MT* was the time taken for each trial to complete. For σ, we computed the deviation from the target's midpoint when the cursor passed through that target (or its vertical extension). W_e was calculated by multiplying σ by 4.133, and 96% of the crossing points at the end targets fell within this W_e [22]. *TP* was calculated by MT/ID, and two types of *TP* were analyzed: TP_n derived by ID_n calculated from W and A, and TP_e derived by ID_e calculated from W_e and A_e.

4 RESULTS OF EXPERIMENT 1

The first set consisting of 15 trials in each *Bias* condition was considered as practice, and the remaining 9000 trials ($3_{Bias} \times 3_A \times 5_W \times$ 20 sets $\times$10 participants) were analyzed. Since ANOVA is robust against normality violation [9, 19], we analyzed all 9000 trials by mean-of-means calculation via RM-ANOVA with a Bonferroni post-hoc test. The independent variables were *Bias*, *A*, and *W*, and the dependent variables *ER* (error rate), *MT*, σ (the deviation from the midpoint of the target at the end of the stroke), and *TP*. Throughout this paper, error bars in the graphs indicate 95% CIs. ***, **, and * in the graphs indicate $p < .001$, $p < .01$, and $p < .05$, respectively.

4.1 Error Rate (*ER*)

We observed 1187 erroneous trials, where uses passed outside the final target (13.2%). Significant main effects were found on *Bias* ($F_{2,18} = 39.0, p < .001, \eta_p^2 = .813$), A ($F_{2,18} = 10.1, p < .01, \eta_p^2 = .529$), and W ($F_{4,36} = 85.9, p < .001, \eta_p^2 = .905$). Also, we identified significant differences for all *Bias* pairs (Fig. 2 (i)). Further, we found significant interactions on $Bias \times A$ ($F_{4,36} = 7.27, p < .001, \eta_p^2 = .447$), and $Bias \times W$ ($F_{8,72} = 23.4, p < .001, \eta_p^2 = .723$).

4.2 Movement Time (*MT*)

We analyzed the *MT* for all 9000 trials (because W_e normalizes the *ER* to 4%). Significant main effects were found on *Bias* ($F_{2,18} = 43.7, p < .001, \eta_p^2 = .829$), A ($F_{2,18} = 110, p < .001, \eta_p^2 = .924$), and W ($F_{4,36} = 104, p < .001, \eta_p^2 = .920$). Significant differences were found on all *Bias* pairs (Fig. 2 (ii)). Significant interactions were found on $Bias \times A$ ($F_{4,36} = 34.3, p < .001, \eta_p^2 = .792$), $Bias \times W$ ($F_{8,72} = 32.7, p < .001, \eta_p^2 = .784$), $A \times W$ ($F_{8,72} = 38.9, p < .001, \eta_p^2 = .812$), and $Bias \times A \times W$ ($F_{16,144} = 2.55, p < .01, \eta_p^2 = .221$).

4.3 Standard Deviation at the Endpoint

We analyzed the standard deviation of the endpoint scatter data (σ). Significant main effects were found on *Bias* ($F_{2,18} = 24.7, p < .001, \eta_p^2 = .733$), A ($F_{2,18} = 52.2, p < .001, \eta_p^2 = .853$), and W ($F_{4,36} = 106, p < .001, \eta_p^2 = .922$). Significant interactions were found on $Bias \times A$ ($F_{4,36} = 9.67, p < .001, \eta_p^2 = .518$), $Bias \times W$ ($F_{8,72} = 2.58, p < .05, \eta_p^2 = .223$), and $A \times W$ ($F_{8,72} = 3.79, p < .001, \eta_p^2 = .296$).

4.4 Model Fitting

For each *Bias*, the baseline Fitts' law model (ID_n) showed strong fits (Fig. 2 (v)). Using ID_e with ${W_e}^2$ showed poorer fits for each *Bias* (Fig. 2 (vi)). Since the number of free parameters is 2 for both ID_n and ID_e, we used non-adjusted R^2 values in this paper. In addition, to analyze the fits in a comparative manner, we used the *AIC* measure [3]. The lower the *AIC*, the better the fit, and a difference of 2 or more is considered to be significant [7]. The differences in the *AIC*s between ID_n and ID_e at each *Bias* were 16 for Neutral, 3 for Fast, and 12 for Accurate, i.e., all larger than 2. Therefore, the ID_n model better predicts the *MT* for each *Bias*.

For a mixed *Bias* (*Mixed*), the fit of the ID_e model was better than the ID_n model (Fig. 2 (v, vi)). The difference in terms of *AIC* was approximately 58 and thus clearly significant. Therefore, the ID_e model is recommended to predict the *MT* from data that may contain multiple biases.

4.5 Throughput (*TP*)

We analyzed the *TP* calculated using ID_n (TP_n) and ID_e (TP_e).

4.5.1 $\mathbf{TP_n}$. Significant main effects were found on *Bias* ($F_{2,18} = 54.5, p < .001, \eta_p^2 = .858$), A ($F_{2,18} = 84.6, p < .001, \eta_p^2 = .904$), and W ($F_{4,36} = 16.2, p < .001, \eta_p^2 = .922$). Significant differences were found on all *Bias* pairs (Fig. 2 (iii)). Significant interactions were found on $Bias \times A$ ($F_{4,36} = 50.8, p < .001, \eta_p^2 = .849$), $Bias \times W$ ($F_{8,72} = 6.25, p < .001, \eta_p^2 = .410$), $A \times W$ ($F_{8,72} = 14.3, p < .001, \eta_p^2 = .613$), and $Bias \times A \times W$ ($F_{16,144} = 6.39, p < .001, \eta_p^2 = .415$).

4.5.2 $\mathbf{TP_e}$. Significant main effects were found on *Bias* ($F_{2,18} = 53.0, p < .001, \eta_p^2 = .855$), A ($F_{2,18} = 103, p < .001, \eta_p^2 = .919$), and W ($F_{4,36} = 69.4, p < .001, \eta_p^2 = .885$). Significant differences were found on all *Bias* pairs (Fig. 2 (iv)). Significant interactions were found on $Bias \times A$ ($F_{4,36} = 34.8, p < .001, \eta_p^2 = .795$), $Bias \times W$

[2] In Experiment 1, we did not use A_e because the actual cursor trajectory distance along the task axis was always the same as A.

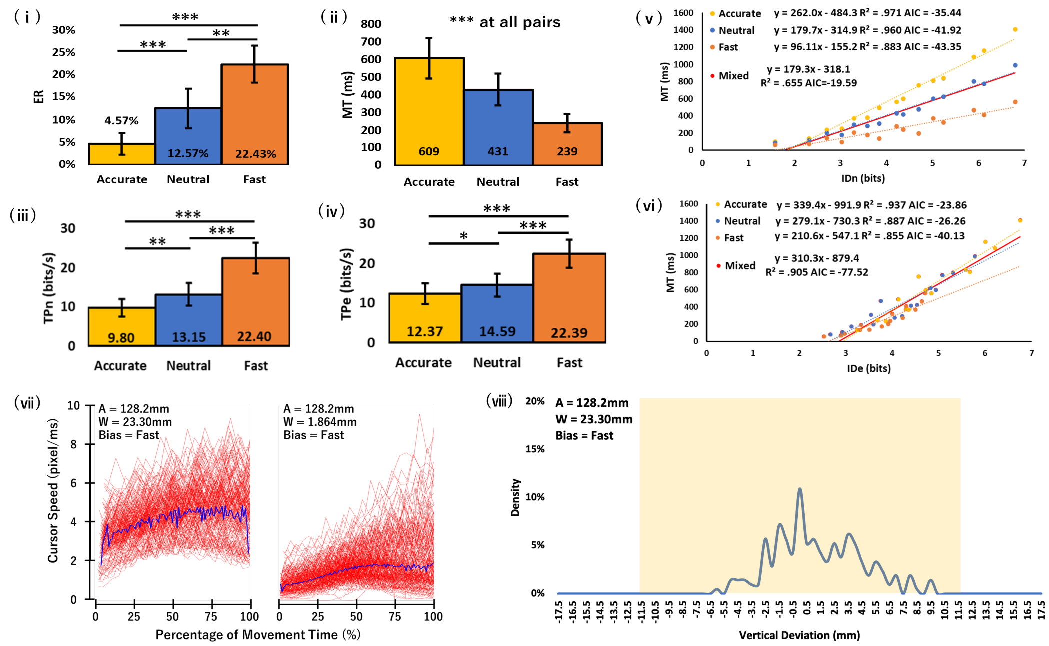

Figure 2: Results of Experiment 1. (i) Error rate across *Bias*. (ii) Movement time across *Bias*. (iii) TP_n across *Bias*. (iv) TP_e across *Bias*. (v) ID_n vs MT. (vi) ID_e vs MT. (vii) Cursor speed against task progress in %. The blue lines represent the average cursor speed. (viii) Cursor endpoint distribution. Yellow rectangle indicates the task success area.

($F_{8,72} = 2.37, p < .05, \eta_p^2 = .209$), $A \times W$ ($F_{8,72} = 52.4, p < .001, \eta_p^2 = .853$), and $Bias \times A \times W$ ($F_{16,144} = 2.03, p < .05, \eta_p^2 = .184$).

5 DISCUSSION OF EXPERIMENT 1

5.1 Model Fitting

The strong fit of the ID_n model for each *Bias* indicates that the ID_n model can be used as a model for predicting the MT of the C/DC crossing task (Fig. 2 (v)). This result matches previous studies that investigated only the Neutral condition [2, 5], but extends it for multiple *Bias* conditions. The ID_n model had strong fits for each *Bias*, and the ID_e model had a strong fit for the *Mixed* bias. This result is similar to a previous study on pointing [26] and demonstrates the ability of the ID_e to normalize the effect of different *Bias* conditions.

5.2 Throughput

We found that TP was not stable across different ID values for each *Bias* (Table 1). To analyze this objectively, we computed the standard deviation SD of TP_n across all ID_n values. For Accurate, the mean TP_n was 9.804 bits/s and SD was 4.342 bits/s. Thus, the ratio of SD to the mean TP_n was 44.29%, while this value should be zero if TP_n values for all IDs were the same (as the SD is 0 in this case). Similarly, the values were 36.77% for Neutral, and 34.32% for Fast. For TP_e, we obtained 65.16% for Accurate, 56.83% for Neutral, and 50.80% for Fast (Table 1). In contrast, a previous study on pointing had identified that this metric was 13.2% in the Neutral condition [17]. These results indicate that, in the C/DC task, TP is not a robust metric in terms of the invariance across IDs.

For the invariance across different *Bias* conditions, TP_e is smoother than TP_n (Fig. 2 (iii, iv)). In TP_n, the difference between Fast and Accurate was 56.27% of the value for Fast, whereas with TP_e, the difference decreased to 44.74%. This value is almost the same as the difference between max-speed and max-accuracy conditions (42%) reported in a previous study on pointing [20]. This confirms the effectiveness of smoothing the effect of *Bias* conditions through TP_e.

5.3 Cursor Velocity and Endpoint

In low-ID conditions, the cursor velocity was not stable across trials nor participants (Fig. 2 (vii)). Also, there was no clear slowdown of the cursor velocity near the end of task, which is inconsistent with pointing movements [6, 21]. Furthermore, in several condition (16/45), the participant's average MT was smaller than the average human reaction time of 240 ms (Table 1) [8]. These results suggested that participants probably completed the task as a purely ballistic movement under low-ID conditions [13]. Still, the cursor endpoint distribution was concentrated (reasonably well) near the center of the target (Fig. 2 (viii)).

These results point to a potential reason behind the high TPs compared with those for pointing (e.g., the maximum value of TP was 11.54 bits/s for Neutral for a pointing task according to MacKenzie [17]). Our results identify a maximum value of 36.63 bits/s for TP_n for Fast (Table 1). Our high TPs were probably due to the fact that the C/DC task allowed passing near the center of the target with a purely ballistic movement, which made both W_e and MT small and thus TP high.

Table 1: *TP* across *ID*s for the *C*/*DC* task (Experiment 1)

Bias	A[mm]	W[mm]	W_e[mm]	ER[%]	ID_n[bits]	ID_e[bits]	MT[ms]	TP_n[bits/s]	TP_e[bits/s]
Neutral	46.60	23.30	8.910	0.00	1.585	2.666	83.39	21.32	35.35
		11.65	7.246	0.00	2.322	2.926	110.6	23.87	29.79
		6.524	6.245	7.50	3.026	3.120	174.3	19.83	20.14
		3.495	4.590	15.0	3.841	3.542	308.2	14.06	12.78
		1.864	3.960	27.5	4.700	3.752	471.7	11.84	9.152
	128.2	23.30	11.83	0.00	2.700	3.590	197.8	15.09	19.84
		11.65	8.378	0.50	3.585	4.049	278.7	14.22	16.07
		6.524	6.567	6.50	4.368	4.406	416.2	11.16	11.13
		3.495	4.957	19.5	5.235	4.821	619.5	9.253	8.343
		1.864	4.239	36.0	6.124	5.080	772.3	8.647	6.993
	205.0	23.30	12.22	0.00	3.293	4.210	296.0	12.20	15.52
		11.65	9.568	2.50	4.217	4.529	427.0	10.79	11.42
		6.524	7.235	10.5	5.019	4.936	599.7	9.049	8.733
		3.495	5.685	25.0	5.899	5.308	800.5	8.160	7.176
		1.864	4.127	38.0	6.794	5.781	990.1	7.752	6.411
						Mean	436.4	13.15	14.59
						SD	271.4	4.835	8.290
Fast	46.60	23.30	10.15	0.00	1.585	2.527	58.03	29.33	46.15
		11.65	7.357	1.00	2.322	2.897	70.06	36.11	44.65
		6.524	7.028	6.50	3.026	2.972	90.13	36.63	35.87
		3.495	5.582	28.5	3.841	3.299	133.5	31.61	26.78
		1.864	5.674	43.5	4.700	3.326	191.6	28.54	19.81
	128.2	23.30	12.54	0.00	2.700	3.508	136.4	20.81	26.97
		11.65	10.74	6.00	3.585	3.773	175.6	21.89	22.92
		6.524	9.083	19.0	4.368	3.953	239.1	19.97	17.91
		3.495	9.235	43.0	5.235	3.999	322.0	18.08	13.36
		1.864	7.203	60.5	6.124	4.299	409.7	16.99	11.65
	205.0	23.30	14.83	0.00	3.293	3.913	203.9	17.01	20.27
		11.65	11.98	8.00	4.217	4.220	274.7	16.59	16.46
		6.524	10.24	18.0	5.019	4.496	369.5	14.75	12.97
		3.495	9.069	42.0	5.899	4.637	465.2	13.89	10.74
		1.864	8.812	60.5	6.794	4.709	560.5	13.75	9.346
						Mean	246.7	22.40	22.39
						SD	151.3	7.687	11.37
Accurate	46.60	23.30	7.237	0.00	1.585	2.923	95.08	18.50	33.72
		11.65	5.735	0.50	2.322	3.239	135.8	19.31	26.55
		6.524	4.179	2.00	3.026	3.677	248.1	14.24	17.14
		3.495	3.033	5.50	3.841	4.132	491.0	8.836	9.215
		1.864	2.142	15.5	4.700	4.566	755.3	6.872	6.554
	128.2	23.30	9.612	0.00	2.700	3.857	235.3	12.72	18.15
		11.65	7.164	1.00	3.585	4.296	376.5	10.47	12.40
		6.524	5.197	1.00	4.368	4.725	598.0	8.052	8.620
		3.495	2.954	5.50	5.235	5.588	835.1	6.705	7.002
		1.864	2.257	15.5	6.124	6.004	1159	5.728	5.472
	205.0	23.30	10.65	0.00	3.293	4.355	369.2	9.771	12.93
		11.65	7.650	1.00	4.217	4.841	562.2	8.094	9.166
		6.524	4.300	1.00	5.019	5.669	809.8	6.751	7.543
		3.495	2.955	4.00	5.899	6.214	1086	5.785	6.002
		1.864	2.068	12.0	6.794	6.756	1408	5.216	5.102
						Mean	611.0	9.804	12.37
						SD	393.5	4.342	8.061

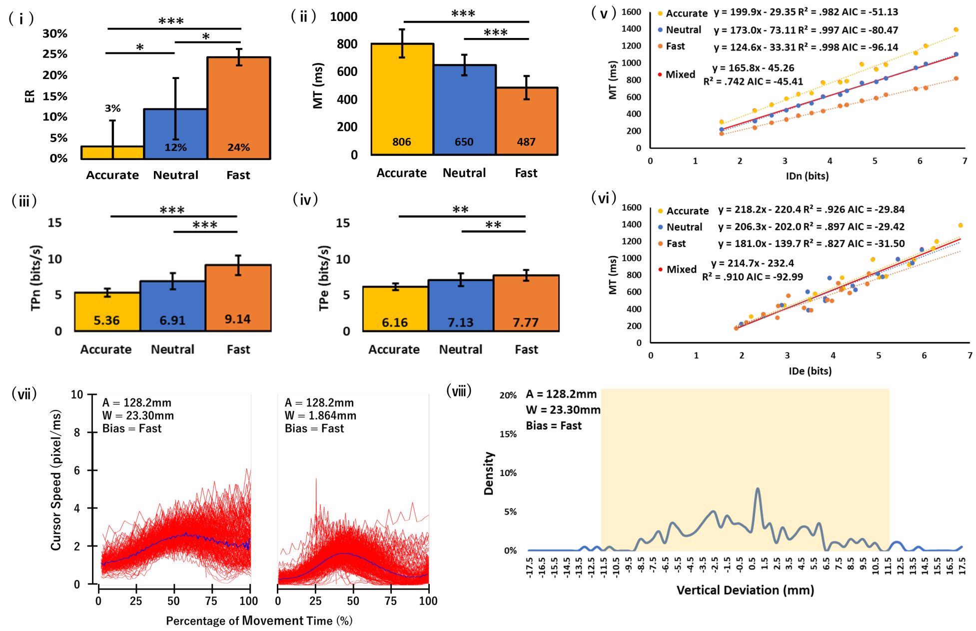

Figure 3: Results of Experiment 2. (i) Error rate across *Bias*. (ii) Movement time across *Bias*. (iii) TP_n across *Bias*. (iv) TP_e across *Bias*. (v) ID_n vs MT. (vi) ID_e vs MT. (vii) Cursor speed against task progress in %. The blue lines represent the average cursor speed. (viii) Cursor endpoint distribution. Yellow rectangle indicates the task success area.

6 EXPERIMENT 2

To further investigate crossing motions, we conducted a study using the C/AC task, where targets are horizontally aligned (Fig. 1 (ii)). The direction of crossing was from bottom to top on the start target (right) and from top to bottom on the end target (left). This task makes it impossible to complete the task by simply moving the stylus in a (more or less) straight line, preventing completion of the task through a simple ballistic movement. The apparatus, design, procedure, and measurements were the same as in Experiment 1. We invited ten new university students (2 females and 8 males, mean age 20.8, standard deviation 1.23). All participants were right-handed. Thus, only the orientation of the target and the participants were changed.

7 RESULTS OF EXPERIMENT 2

The scope of data analyzed, method of analysis, and the independent and dependent variables were all the same as in Experiment 1.

7.1 Error Rate (*ER*)

We identified 1180 error trials (13.1%). Significant main effects were found on *Bias* ($F_{2,18} = 24.2, p < .001, \eta_p^2 = .729$), *A* ($F_{2,18} = 6.31, p < .01, \eta_p^2 = .412$), and *W* ($F_{4,36} = 38.2, p < .001, \eta_p^2 = .809$). Significant differences were found on all *Bias* pairs (Fig. 3 (i)). Significant interactions were found on $Bias \times W$ ($F_{8,72} = 14.0, p < .001, \eta_p^2 = .609$).

7.2 Movement Time (*MT*)

Significant main effects were found on *Bias* ($F_{2,18} = 25.7, p < .001, \eta_p^2 = .740$), *A* ($F_{2,18} = 456, p < .001, \eta_p^2 = .981$), and *W* ($F_{4,36} = 130, p < .001, \eta_p^2 = .935$). Significant differences were found between Neutral and Fast as well as Accurate and Fast (Fig. 3 (ii)). Significant interactions were found on $Bias \times A$ ($F_{4,36} = 14.9, p < .001, \eta_p^2 = .624$), $Bias \times W$ ($F_{8,72} = 8.53, p < .001, \eta_p^2 = .487$), $A \times W$ ($F_{8,72} = 6.51, p < .001, \eta_p^2 = .420$), and $Bias \times A \times W$ ($F_{16,144} = 1.74, p < .05, \eta_p^2 = .162$).

7.3 Standard Deviation of Endpoint

Significant main effects were found on *Bias* ($F_{2,18} = 20.8, p < .001, \eta_p^2 = .698$), and *W* ($F_{4,36} = 149, p < .001, \eta_p^2 = .943$). No significant interactions were found.

7.4 Model Fitting

For each *Bias*, the ID_n model exhibited strong fits (Fig. 3 (v)). The model based on W_e and A_e (the ID_e model)[3] showed poorer fits for each *Bias* condition (Fig. 3 (vi)). The differences between the *AIC* of the ID_n model and the ID_e model at each *Bias* were 51 for Neutral, 65 for Fast, and 21 for Accurate, all larger than 2. Therefore, the ID_n model better predicts the *MT* for each *Bias*.

For mixed *Biass* (*Mixed*), the fit of the ID_e model was stronger than the ID_n model (Fig. 3 (v, vi)). The difference of *AIC* was approximately 48, which is clearly a significant difference. Therefore, we recommend the ID_e model to predict the *MT* from data that may contain multiple biases.

[3]In Experiment 2, we used the ID_e model corrected by W_e and A_e because the actual cursor trajectory distance on the task axis was not always the same as *A*.

Table 2: TP across IDs for the C/AC task (Experiment 2)

Bias	A[mm]	W[mm]	A_e[mm]	W_e[mm]	ER[%]	ID_n[bits]	ID_e[bits]	MT[ms]	TP_n[bits/s]	TP_e[bits/s]
Neutral	46.60	23.30	46.67	16.18	0.00	1.585	1.994	222.4	7.577	9.492
		11.65	47.74	10.94	6.00	2.322	2.483	318.5	7.722	8.271
		6.524	47.44	7.945	14.5	3.026	2.874	449.3	7.114	6.729
		3.495	46.67	5.205	17.5	3.841	3.444	608.0	6.729	5.828
		1.864	46.85	3.744	31.0	4.700	3.924	769.5	6.658	5.327
	128.2	23.30	126.5	13.13	0.50	2.700	3.460	389.3	7.330	9.271
		11.65	128.4	10.04	5.00	3.585	3.831	528.1	7.099	7.519
		6.524	128.3	6.668	7.50	4.368	4.435	675.4	6.757	6.717
		3.495	128.3	4.711	16.5	5.235	4.979	822.9	6.697	6.221
		1.864	128.3	3.614	26.0	6.124	5.419	993.7	6.596	5.679
	205.0	23.300	203.1	16.00	0.00	3.293	3.828	500.8	6.896	7.953
		11.650	204.5	9.960	5.00	4.217	4.493	630.5	6.921	7.275
		6.524	204.9	6.747	8.00	5.019	5.085	783.7	6.673	6.636
		3.495	205.3	4.085	14.5	5.899	5.738	948.0	6.487	6.265
		1.864	205.1	3.685	28.0	6.794	5.944	1109	6.463	5.575
							Mean	649.9	6.915	6.984
							SD	247.7	0.370	1.255
Fast	46.60	23.30	45.70	17.86	2.00	1.585	1.882	171.5	9.798	11.41
		11.65	47.62	15.05	15.0	2.322	2.100	243.4	10.04	9.059
		6.524	47.83	10.942	28.0	3.026	2.475	339.8	9.709	7.788
		3.495	47.59	8.985	39.0	3.841	2.795	437.4	9.494	6.667
		1.864	47.41	9.286	50.0	4.700	3.018	558.5	9.488	5.544
	128.2	23.30	126.3	22.02	3.50	2.700	2.780	300.7	9.327	9.593
		11.65	129.0	14.66	13.0	3.585	3.362	414.4	8.944	8.313
		6.524	128.6	10.21	17.0	4.368	3.876	510.7	8.855	7.731
		3.495	128.5	7.907	31.0	5.235	4.181	629.8	8.648	6.797
		1.864	128.1	9.656	52.0	6.124	4.117	708.2	9.134	5.909
	205.0	23.30	202.8	20.01	4.50	3.293	3.520	385.8	8.839	9.363
		11.65	204.7	15.18	12.0	4.217	3.975	498.3	8.735	8.077
		6.524	205.1	11.32	16.5	5.019	4.365	591.5	8.720	7.476
		3.495	205.2	8.433	32.5	5.899	4.749	700.4	8.702	6.896
		1.864	204.9	9.526	49.0	6.794	4.784	821.7	8.645	5.944
							Mean	487.5	9.138	7.771
							SD	178.3	0.453	1.547
Accurate	46.60	23.30	46.85	13.33	0.00	1.585	2.201	311.2	5.380	7.394
		11.65	47.19	7.763	1.50	2.322	2.932	445.4	5.304	6.635
		6.524	47.12	4.484	2.00	3.026	3.606	583.9	5.292	6.244
		3.495	46.83	2.914	4.50	3.841	4.199	773.8	5.103	5.521
		1.864	46.64	1.783	8.50	4.700	4.874	988.6	4.988	5.053
	128.2	23.30	126.8	12.71	0.00	2.700	3.495	513.0	5.371	6.886
		11.65	128.4	7.518	1.50	3.585	4.223	647.0	5.598	6.579
		6.524	128.4	3.790	0.50	4.368	5.164	790.2	5.631	6.619
		3.495	128.3	2.500	3.00	5.235	5.769	984.1	5.500	6.010
		1.864	128.2	1.832	5.50	6.124	6.270	1202	5.335	5.398
	205.0	23.30	203.4	12.21	0.00	3.293	4.186	635.6	5.272	6.669
		11.65	205.0	6.814	0.50	4.217	5.016	777.7	5.506	6.515
		6.524	205.2	4.230	0.00	5.019	5.674	930.4	5.507	6.203
		3.495	205.2	3.115	5.00	5.899	6.197	1118	5.466	5.698
		1.864	205.2	1.969	12.5	6.794	6.792	1394	5.102	5.043
							Mean	806.4	5.357	6.165
							SD	288.4	0.181	0.671

7.5 Throughput (*TP*)

7.5.1 **$\mathbf{TP_n}$**. Significant main effects were found on *Bias* ($F_{2,18} = 25.3, p < .001, \eta_p^2 = .738$), and A ($F_{2,18} = 5.40, p < .05, \eta_p^2 = .375$). Significant differences were found between Neutral and Fast as well as Accurate and Fast (Fig. 3 (iii)). Significant interactions were found on $Bias \times A$ ($F_{4,36} = 7.01, p < .001, \eta_p^2 = .438$).

7.5.2 **$\mathbf{TP_e}$**. Significant main effects were found on *Bias* ($F_{2,18} = 13.2, p < .001, \eta_p^2 = .594$), and W ($F_{2,18} = 149, p < .001, \eta_p^2 = .808$). Significant differences were found between Neutral and Fast as well as Accurate and Fast (Fig. 3 (iv)). Significant interactions were found on $Bias \times W$ ($F_{8,72} = 8.98, p < .001, \eta_p^2 = .499$), and $A \times W$ ($F_{8,72} = 7.64, p < .001, \eta_p^2 = .459$).

8 DISCUSSION OF EXPERIMENT 2

8.1 Model Fitting

The results were similar to those in Experiment 1. The ID_n model for each *Bias* showed high fits, indicating the capability of the model to predict the *MT* well (Fig. 3 (v)). This result supports previous work [2, 5]. The ID_e model had a stronger fit for the *Mixed* bias than using ID_n, again indicating the normalization capability of this measure for mixed *Bias* data (Fig. 3 (vi)).

8.2 Throughput

In this C/AC task, *TP* is mostly stable across *ID* (Table 2). For each *Bias*, we computed the *SD* of TP_n across all ID_n values. The ratios of *SD* to the mean were 3.378% for Accurate, 5.354% for Neutral, and 4.955% for Fast. For TP_e, we obtained 11.27% for Accurate, 18.60% for Neutral, and 20.60% for Fast. Compared to the results of Experiment 1, these percentages are more similar to those observed in a previous study on pointing (13.2%) [17]. These indicate less variance of *TP* across different *ID*s than for the C/DC task.

TP_e smoothed the *TP* values across different *Bias* conditions (Fig. 3 (iii, iv)). For TP_n, the difference between Fast and Accurate was 41.39% of the value of Fast, whereas for TP_e, the difference decreased to 20.69%. This value is smaller than the difference observed in a previous study on pointing (42%) [20]. This confirms the effectiveness of *Bias* smoothing through the TP_e measure.

8.3 Cursor Velocity and Endpoint

Even in low-*ID* conditions, the cursor velocity was stable across trials or participants (Fig. 3 (vii)). Also, similar to pointing tasks [6, 21], a slowdown of the cursor velocity near the end of task was observed. Further, the participants' average *MT* was only in a few conditions (2/45) smaller than the average human reaction time of 240 ms (Table 2) [8]. These results suggest that participants did (in general) not complete the task by performing only a ballistic movement [13]. In addition, cursor endpoints were distributed over a wide range around the task success area (Fig. 3 (viii)). These results (non-ballistic movements and large σs) are the reasons behind the lower TP_n values (4.99–10.04 bits/s) compared to those in Experiment 1, also resulting in the outcome that *TP* is more stable across *ID*s in the C/AC task.

9 GENERAL DISCUSSION AND CONCLUSION

9.1 Effective Parameters

For both the C/DC (Experiment 1) and C/AC tasks (Experiment 2), the ID_e model consistently improved the fit for a *Mixed* bias, whereas the ID_n model consistently demonstrated a better fit for each individual *Bias* condition. These results indicate that the ID_e term can predict the *MT* for crossing by smoothing the general effect of the *Bias* at the expense of strong fit in each individual *Bias*.

9.2 Throughput

We identified that *TP* for crossing actions is stable across *ID*s under the following conditions.

- Tasks cannot be completed by ballistic movement even under low-*ID* conditions.
- The distribution of cursor endpoints is spread widely over the target width.

These conditions are a direct consequence of the outcomes of our two experiments in which only the task geometry was changed. *TP* was not stable across *ID*s in the C/DC task (Experiment 1), but was stable in the C/AC task (Experiment 2). In the C/DC task, it was possible to complete the task by a ballistic movement, whereas this was impossible in the C/AC task. In addition, the distribution of cursor endpoints was concentrated near the center of the task in the C/DC task, whereas it was much more widely spread over the target width in the C/AC task.

The effective parameters' smoothing effect on TP_e across *Bias*es was more clearly identified in Experiment 2. In the C/DC task (Experiment 1), the difference of *TP*s across *Bias* was 56.27% for TP_n and 44.74% for TP_e. In the C/AC task (Experiment 2), it was 41.36% for TP_n and 20.72% for TP_e. However, the stability of *TP* across different *ID*s at each *Bias* was consistently higher for TP_n. This indicates that TP_e smoothed the differences between *Bias*es at the expense of robust stability for each individual *Bias*.

In conclusion, we empirically confirmed the capability of effective parameters to normalize across different *Bias* conditions. That is, when comparing the performance of multiple input devices or user groups, TP_e smoothes the influence of *Bias*, which is (potentially) different across users (due to their different speed-accuracy tradeoffs), thus enabling a fairer comparison. Therefore, we suggest that in future studies, in addition to reporting *MT* and *ER*, TP_e should also be reported. We believe that this will allow researchers to compare performance across studies.

A limitation of this work is that we cannot compare the results from Experiments 1 and 2 in terms of speed-profiles, because the participant groups were different. This also made it difficult to directly compare the smoothing effect on the effective parameters, which was greater in the C/AC task (Experiment 2) than in the C/DC task (Experiment 1). We plan to conduct a within-subjects study to compare these conditions, as well as to test other types of crossing, such as *discrete crossing* where users lift a stylus between two targets off the surface [5], touch-based crossing [14], and crossing in virtual reality [24].

ACKNOWLEDGMENTS

We greatly appreciate the reviewers for their helpful discussions.

REFERENCES

[1] Johnny Accot and Shumin Zhai. 1997. Beyond Fitts' Law: Models for Trajectory-Based HCI Tasks. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI '97)*. Association for Computing Machinery, New York, NY, USA, 295–302. https://doi.org/10.1145/258549.258760

[2] Johnny Accot and Shumin Zhai. 2002. More than Dotting the i's — Foundations for Crossing-Based Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Minneapolis, Minnesota, USA) *(CHI '02)*. Association for Computing Machinery, New York, NY, USA, 73–80. https://doi.org/10.1145/503376.503390

[3] Hirotugu Akaike. 1974. A new look at the statistical model identification. *IEEE Trans. Automat. Control* 19, 6 (Dec 1974), 716–723. https://doi.org/10.1109/TAC.1974.1100705

[4] Georg Apitz and François Guimbretière. 2004. CrossY: A Crossing-based Drawing Application. In *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology* (Santa Fe, NM, USA) *(UIST '04)*. ACM, New York, NY, USA, 3–12. https://doi.org/10.1145/1029632.1029635

[5] Georg Apitz, François Guimbretière, and Shumin Zhai. 2008. Foundations for designing and evaluating user interfaces based on the crossing paradigm. *ACM Transactions on Computer-Human Interaction (TOCHI)* 17, 2 (2008), 1–42. https://doi.org/10.1145/1746259.1746263

[6] Takeshi Asano, Ehud Sharlin, Yoshifumi Kitamura, Kazuki Takashima, and Fumio Kishino. 2005. Predictive Interaction Using the Delphian Desktop. In *Proceedings of the 18th Annual ACM Symposium on User Interface Software and Technology* (Seattle, WA, USA) *(UIST '05)*. Association for Computing Machinery, New York, NY, USA, 133–141. https://doi.org/10.1145/1095034.1095058

[7] Kenneth P. Burnham and David R. Anderson. 2002. *Model selection and multimodel inference: a practical information-theoretic approach* (2 ed.). Springer, New York City, NY, USA. 1–488 pages.

[8] Stuart K Card, Thomas P Moran, and Allen Newell. 2018. *The psychology of human-computer interaction*. Crc Press, Boca Raton, Florida, USA.

[9] Peter Dixon. 2008. Models of accuracy in repeated-measures designs. *Journal of Memory and Language* 59, 4 (2008), 447–456.

[10] Pierre Dragicevic. 2004. Combining Crossing-based and Paper-based Interaction Paradigms for Dragging and Dropping Between Overlapping Windows. In *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology* (Santa Fe, NM, USA) *(UIST '04)*. ACM, New York, NY, USA, 193–196. https://doi.org/10.1145/1029632.1029667

[11] Leah Findlater, Alex Jansen, Kristen Shinohara, Morgan Dixon, Peter Kamb, Joshua Rakita, and Jacob O. Wobbrock. 2010. Enhanced Area Cursors: Reducing Fine Pointing Demands for People with Motor Impairments. In *Proceedings of the 23Nd Annual ACM Symposium on User Interface Software and Technology* (New York, New York, USA) *(UIST '10)*. ACM, New York, NY, USA, 153–162. https://doi.org/10.1145/1866029.1866055

[12] Paul M Fitts. 1954. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology* 47, 6 (1954), 381. https://doi.org/10.1037/h0055392

[13] Khai-Chung Gan and Errol R Hoffmann. 1988. Geometrical conditions for ballistic and visually controlled movements. *Ergonomics* 31, 5 (1988), 829–839. https://doi.org/10.1080/00140138808966724

[14] Yuexing Luo and Daniel Vogel. 2014. Crossing-Based Selection with Direct Touch Input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 2627–2636. https://doi.org/10.1145/2556288.2557397

[15] Yuexing Luo and Daniel Vogel. 2015. Pin-and-Cross: A Unimanual Multitouch Technique Combining Static Touches with Crossing Selection. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software and Technology* (Charlotte, NC, USA) *(UIST '15)*. Association for Computing Machinery, New York, NY, USA, 323–332. https://doi.org/10.1145/2807442.2807444

[16] I Scott MacKenzie. 1992. Fitts' law as a research and design tool in human-computer interaction. *Human-computer interaction* 7, 1 (1992), 91–139. https://doi.org/10.1207/s15327051hci0701_3

[17] I Scott MacKenzie. 2018. Fitts' law. *Handbook of human-computer interaction* 1 (2018), 349–370.

[18] I. Scott MacKenzie and Poika Isokoski. 2008. Fitts' Throughput and the Speed-Accuracy Tradeoff. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) *(CHI '08)*. Association for Computing Machinery, New York, NY, USA, 1633–1636. https://doi.org/10.1145/1357054.1357308

[19] Blanca Mena, M José, Rafael Alarcón, Jaume Arnau Gras, Roser Bono Cabré, and Rebecca Bendayan. 2017. Non-normal data: Is ANOVA still a valid option? *Psicothema* 29, 4 (2017), 552–557.

[20] Halla B. Olafsdottir, Yves Guiard, Olivier Rioul, and Simon T. Perrault. 2012. A New Test of Throughput Invariance in Fitts' Law: Role of the Intercept and of Jensen's Inequality. In *Proceedings of the 26th Annual BCS Interaction Specialist Group Conference on People and Computers* (Birmingham, United Kingdom) *(BCS-HCI '12)*. BCS Learning & Development Ltd., Swindon, GBR, 119–126.

[21] Jaime Ruiz, David Tausky, Andrea Bunt, Edward Lank, and Richard Mann. 2008. Analyzing the kinematics of bivariate pointing. In *Proceedings of Graphics Interface 2008* (Windsor, Ontario, Canada) *(GI 2008)*. Canadian Human-Computer Communications Society, Toronto, Ontario, Canada, 251–258.

[22] R. William Soukoreff and I. Scott MacKenzie. 2004. Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research in HCI. *International Journal of Human-Computer Studies* 61, 6 (2004), 751–789. https://doi.org/10.1016/j.ijhcs.2004.09.001

[23] I Standard. 2000. Ergonomic requirements for office work with visual display terminals (VDTs)-Part 9: Requirements for non-keyboard input devices. *Iso 2000* 2000 (2000), 54.

[24] Huawei Tu, Susu Huang, Jiabin Yuan, Xiangshi Ren, and Feng Tian. 2019. Crossing-Based Selection with Virtual Reality Head-Mounted Displays. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300848

[25] Shota Yamanaka, Hiroki Usuba, and Homei Miyashita. 2022. Bivariate Effective Width Method to Improve the Normalization Capability for Subjective Speed-Accuracy Biases in Rectangular-Target Pointing. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 211, 13 pages. https://doi.org/10.1145/3491102.3517466

[26] Shumin Zhai, Jing Kong, and Xiangshi Ren. 2004. Speed–accuracy tradeoff in Fitts' law tasks—on the equivalency of actual and nominal pointing precision. *International Journal of Human-Computer Studies* 61, 6 (2004), 823–856. https://doi.org/10.1016/j.ijhcs.2004.09.007