Better Definition and Calculation of Throughput and Effective Parameters for Steering to Account for Subjective Speed-accuracy Tradeoffs

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@lycorp.co.jp

Anil Ufuk Batmaz
Concordia University
Montreal, QC, Canada
ufuk.batmaz@concordia.ca

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

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

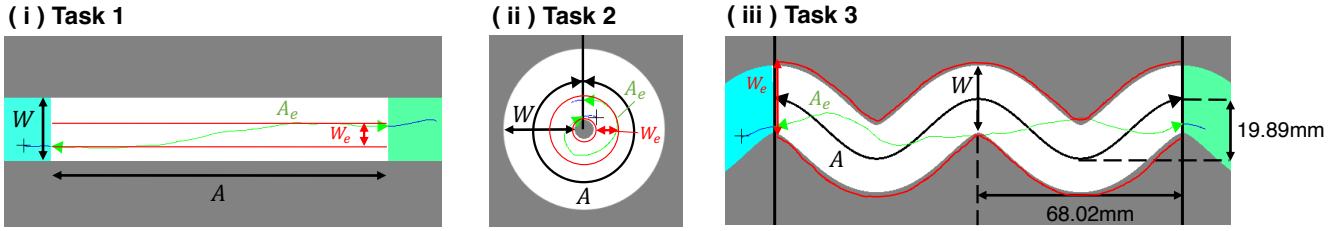


Figure 1: We conducted three steering experiments with (i) linear, (ii) circular, and (iii) sine-wave path shapes. The path width W is the distance perpendicular to the center-line of the path. The path amplitude A (= path length) is the length of the center-line of the path. W and A vary depending on the experimental condition. To prevent differences in the values for (only) the sine-wave conditions from independently affecting the experimental results, the *wave-length* was fixed at 68.02 mm and *wave-amplitude* was fixed at 19.89 mm. The cursor was shown as a black cross beneath the tip of the stylus. The cursor left a green trajectory inside of the path, but a red trajectory when it was outside of the path. Effective Width W_e is calculated as the width that encompasses 96% of the stroked trajectory. Effective Amplitude A_e is calculated as the total length of the stroked trajectory.

ABSTRACT

In Fitts' law studies to investigate pointing, throughput is used to characterize the performance of input devices and users, which is claimed to be independent of task difficulty or the user's subjective speed-accuracy bias. While throughput has been recognized as a useful metric for target-pointing tasks, the corresponding formulation for path-steering tasks and its evaluation have not been thoroughly examined in the past. In this paper, we conducted three experiments using linear, circular, and sine-wave path shapes to propose and investigate a novel formulation for the effective parameters and the throughput of steering tasks. Our results show that the effective width substantially improves the fit to data with mixed speed-accuracy biases for all task shapes. Effective width also smoothed out the throughput across all biases, while the usefulness of the effective amplitude depended on the task shape. Our study thus advances the understanding of user performance in trajectory-based tasks.

CCS CONCEPTS

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

KEYWORDS

trajectory-based interaction, effective parameters, performance metric, throughput, ISO9241-411, understanding people

ACM Reference Format:
Nobuhito Kasahara, Yosuke Oba, Shota Yamanaka, Anil Ufuk Batmaz, Wolfgang Stuerzlinger, and Homei Miyashita. 2024. Better Definition and Calculation of Throughput and Effective Parameters for Steering to Account for Subjective Speed-accuracy Tradeoffs. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24), May 11–16, 2024, Honolulu, HI, USA.* ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3613904.3642084

*This work was conducted when the first author was interning at LY Corporation.
†Currently affiliated with LY Corporation.

CHI '24, May 11–16, 2024, Honolulu, HI, USA

ACM ISBN 979-8-4007-0330-0/24/05…$15.00
https://doi.org/10.1145/3613904.3642084

1 INTRODUCTION

Graphical user interfaces (GUIs) are often designed to involve pointing to targets, but other interaction tasks also exist, such as trajectory-based tasks, including crossing [3] and steering [1]. A steering task requires users to pass through a constrained path, e.g., for the selection in a hierarchical menu, drawing a curve, or moving around in 3D worlds [1]. Research on trajectory-based tasks has thus the potential to contribute to the further development of novel GUIs, for example, by enabling more efficient command selection techniques [5, 34].

In addition to developing novel interaction techniques, understanding and modeling the performance of GUI operations is a core topic in human-computer interaction (HCI) research. A well-studied example of such a model is Fitts' law, which models the movement time *MT* of pointing and crossing operations [1, 13, 22]. Further, steering tasks are modeled by the steering law, which was established on the basis of Fitts' law as a series of crossing tasks [1].

For the pointing task, *throughput TP* based on effective parameters is used as a performance evaluation metric [17], which is calculated from the index of difficulty *ID* of Fitts' law and the experimentally measured *MT*, as it combines speed and accuracy into one value. *TP* has been claimed to represent the performance of input devices [10] and user groups [31, 37], and to be independent of task difficulty [23] and the speed-accuracy tradeoff [24].

The robustness of the *throughput* measure has been studied for pointing tasks by computing *TP*s for multiple task difficulties, i.e., *ID*s [23]. However, for different speed-accuracy biases, *TP* calculated by the nominal *ID* using the target width W and its amplitude A was found to be unstable [27]. Therefore, previous work used effective parameters (i.e., the *effective width* W_e and *effective amplitude* A_e), which take into account the user's actual movement but not the task, to calculate *ID*, which smooths *TP* across different speed-accuracy biases [24, 44]. The *TP* calculated using effective parameters allows for fairer comparisons between different users groups or input devices [24, 31, 37], even if speed-accuracy biases vary across the participants and devices. This benefit was also confirmed to hold for crossing operations [18].

Previous work has also systematically applied W_e to steering tasks [19, 46, 47], but the theoretical validity and the empirical advantage of this approach in terms of smoothing speed-accuracy biases have both not been evaluated in detail. For example, in previous studies [19, 46, 47], the nominal path width W was simply replaced by W_e in an ad-hoc manner, based on the experimental results. However, considering how the steering law was originally modeled as a series of crossing tasks [1], the theoretical derivation of effective width from those for crossing put forth in recent work [18] strengthens the rationale for replacing nominal width with effective width. In addition, these studies have investigated the impact of W_e on the *ID*-*MT* model, but did not investigate A_e and *TP*. Thus, the exploration of the effects of effective parameters in steering tasks remains incomplete in the field of HCI. This may lead researchers to use an inappropriate metric, such as that a previous study [36] used the nominal *TP*, which is strongly affected by subjective speed-accuracy biases, as we later show.

These previous studies demonstrate that effective width for steering tasks have been used without a theoretical basis, and we have no knowledge on how using effective parameters normalizes the speed-accuracy biases and yields stable *TP*s. Therefore, to examine the applicability of *TP* and effective parameters for steering, we conducted three experiments with different path shapes in this work, each under three speed-accuracy biases (Figure 1). Our main motivation is to establish an appropriate performance evaluation metric to compare the steering performance of devices or users in a better way. Thus, we focused on a fairer performance comparison by smoothing the implicit speed-accuracy bias that participants have when they perform path-steering experiments.

Since the steering task is modeled as a continuation of crossing tasks [1], it is possible that the trends of *TP* stability, model fit, or effect of task shape may be similar to those for crossing [18]. On the other hand, for pointing and crossing, the effective parameters are calculated from the distribution of the endpoints, whereas for steering, the distribution of the entire trajectory between the strokes could be used, which might lead to different conclusions. Therefore, it is worthwhile to verify the results experimentally.

Our main novelty and contributions are as follows.

- This study theoretically applies Fitts' *throughput TP* and effective parameters, W_e and A_e, to steering tasks and we analyze how the inclusion of W_e and A_e affects the model fit and *TP* under different speed-accuracy biases. The verification of *TP* for steering in our study is novel.
- We experimentally demonstrate that W_e smooths the speed-accuracy bias for three path shapes, in terms of better model fits and more stable *TP*s. In contrast, A_e improves the model fit and stability of *TP* between task difficulties at each individual speed-accuracy bias only in the circular steering task. Identifying how W_e and A_e affect *TP* for steering is a novel contribution of this study.
- We present implications for future steering-based user experiments and for the evaluation of interaction techniques. For example, to most fairly compare input devices, we suggest using a sine-wave steering task, as our results show that this task can yield most stable *TP*s across different speed-accuracy biases.

2 RELATED WORK

2.1 Fitts' Law and Steering Law

Fitts' law [13, 22] describes the relationship between the nominal parameters of the pointing task (target width W and amplitude between targets A) and the movement time *MT*. In HCI MacKenzie's Shannon capacity formulation is most frequently used [21]:

$$MT = a + bID_n,\ ID_n = \log_2\left(\frac{A}{W} + 1\right), \tag{1}$$

where a and b are empirical constants and ID_n designates the nominal index of difficulty of the task. Although MacKenzie's formulation has been criticized from a theoretical point of view [11, 14, 16], we use this model because it fits the experimental data well and has also been validated in a crossing experiment [18] on which our work is based.

In crossing tasks, the *MT* to select a target by passing through it is also modeled by the same Fitts' law [1]. Leveraging this observation,

many crossing-based interaction methods have been proposed [3, 5, 26, 29, 33, 34].

In contrast, a steering task requires users to make a stroke inside a constrained straight path, which is modeled in the steering law through a sequence of crossing tasks [1]. By applying Fitts' law multiple times, this yields the following relationship:

$$MT = a + bID_n,\ ID_n = \frac{A}{W}, \tag{2}$$

where A is the path length and W is its width. The steering law was later extended to predict the task difficulty of various path-steering operations in GUIs, such as widening and narrowing tunnels [39], passing through corners [28], and sequential linear path segments [41], with many different input devices [2, 45].

Importantly, previous studies have demonstrated that MT and error rate ER heavily depend on the input device and task, e.g., finger touching showed the shortest MT for target pointing compared to stylus and mouse, while a stylus was the best for dragging tasks [7]. Note that these "best" devices in terms of MT still exhibited also the highest ERs. Similarly, Accot and Zhai demonstrated that mice had the shortest MT in a linear-path steering task, while the stylus was the fastest for circular paths [2]. For example, for a circular steering task, MTs for all trials including errors were 2532 ms for the mouse and 2193 ms for the pen tablet, indicating that the pen tablet was faster. However, the error rate was 14.0% for the mouse and 22.9% for the pen tablet, indicating that the mouse was more accurate. Thus, it is not possible to determine directly which device is "better", as their error rates are unequal. As demonstrated by such work, and because different devices are superior for various GUI tasks, evaluating their performance only on the basis of the TP for pointing is not sufficient, which motivated us to derive an appropriate TP for steering.

2.2 Effective Parameters

Fitts' law is based on two nominal parameters, W and A, which do not represent the user's actual behavior. In other words, the task difficulty does not change even if users change their speed or accuracy. For this reason, it is not possible to predict MT when multiple speed-accuracy biases co-exist. To incorporate this bias, Fitts' law can be rewritten in accordance with the user's behavioral parameters as follows [8, 32]:

$$MT = a + bID_e,\ ID_e = \log_2\left(\frac{A_e}{W_e} + 1\right), \tag{3}$$

where W_e and A_e are effective width and amplitude, respectively, calculated as follows [8, 24, 32]:

$$W_e = 4.133 \times \sigma, \tag{4}$$

$$A_e = \text{mean (movement distance)}, \tag{5}$$

where σ is the standard deviation of the cursor endpoints [32]. A_e represents the actual average movement distance and W_e is adjusted so that 96% of clicks fall inside the target. Both measures are computed for movements along the task-axis [32].

Zhai et al. revealed that the ID_e model showed a better fit in terms of R^2 when analyzing the data from multiple speed-accuracy biases in a mixed manner (called the Mixed analysis condition) than using the ID_n model in pointing [44]. However, for each individual bias (such as emphasizing either Accurate, Neutral, or Fast movements), the ID_n model showed a better fit. This indicates that the ID_e model improves the fit for mixed conditions at the expense of higher prediction accuracy of the ID_n model for each individual bias. Kasahara et al. revealed that these tendencies identified by Zhai et al. also apply to crossing [18].

The usefulness of applying effective width for steering tasks has also been demonstrated [19, 46, 47]. For steering, the effective path width W_e is defined as the standard deviation of the cursor coordinate in the width axis, which is perpendicular to the movement direction of the main task. Kulikov et al. showed that the use of W_e improves the accuracy of MT prediction over W [19]. Zhou et al. conducted a study of time-constrained steering tasks and identified that W_e varied with the speed-accuracy tradeoff [46]. Zhou and Ren showed that, with verbally instructed speed-accuracy biases, using W_e improves the prediction accuracy of MT for data with a mixture of different speed-accuracy biases over the use of W alone [47]. They also showed that the model fit with W_e was improved over using W for some individual bias conditions. In contrast to pointing and crossing tasks, in which W_e reduced the model fit for each individual bias [18, 44], steering tasks exhibited better model fits in a single-bias experiment by using W_e [19] and some individual bias conditions [47]. Therefore, W_e in steering might be a more robust parameter than W_e in pointing and crossing.

We acknowledge that previous work proposed an effective width that dynamically changes in a single stroke, such that slower speeds are interpreted as a narrower effective width [20]. However, the purpose of our study is to establish a performance metric for device comparison, when performing a steering task given by W and A. That is, we intend to replace W and A in the steering law with other values that take users' speed-accuracy biases into account. Thus, we calculate W_e and A_e as appropriate for the tasks we investigate here, i.e., where A and W do not change within a single stroke.

2.3 Throughput

For pointing tasks throughput TP is standardized by ISO9241-411 and is used as a performance metric of a user group and an input device [17]. It is calculated as follows:

$$TP = \frac{ID}{MT}. \tag{6}$$

Ideally, the performance of a user group or device should be independent of the difficulty of the task. TP is mostly independent of IDs [23]; Typically when the ID changes, the MT also changes accordingly, which means that the TP then remains almost constant.

The performance of a user group or device should also be independent of the user's subjective speed-accuracy bias. MacKenzie et al. showed that TP is constant for different speed-accuracy biases. As demonstrated by their results, if the biases changed MTs more than 10% compared to the result under the Neutral instruction, there was no significant difference in TP between different biases [24]. On the other hand, Olafsdottir et al. showed that TP cannot be smoothed between different speed-accuracy biases, if the bias is extreme enough to force participants to ignore accuracy or to ignore speed [27].

Kasahara et al. showed that TP for crossing has different stability between IDs depending on the task shapes, such that TP tended to

be large for low *ID*s and a fast bias condition [18]. The reason is that even though in their a *crossing with directional constraint* task the crossing task could be completed very quickly, the endpoint distribution was very concentrated (i.e., both σ and MT became small). This suggests that, as the MT prediction model for steering was developed by summing a series of crossing tasks, the stability of TP might differ between steering tasks and pointing tasks.

By directly applying the TP definition of pointing to a linear-path steering task, Wiese et al. compared the performance under various input latencies that were artificially generated [36]. However, they used ID_n and MT to calculate TP in their study, and when MT was smaller as a result of a higher ER, TP became larger, which does not consider the effect of accuracy. To validate such task performance, it is thus important to theoretically establish the TP measure, which then considers the effect of accuracy for steering appropriately.

3 DEFINITION OF EFFECTIVE PARAMETERS FOR STEERING

3.1 Deriving the Steering Law with Effective Parameters

The purpose of effective parameters is to more accurately describe human behavior as people's behaviors vary with the speed-accuracy tradeoff [8, 32]. Since Fitts' law with effective parameters holds for a crossing task [18], we apply it to successive crossings (Figure 2 (i)). The time of a single crossing with effective parameters, MT_1, is

$$MT_1 = a + bID_1, \quad ID_1 = \log_2\left(\frac{A_e}{W_e} + 1\right). \tag{7}$$

If a target is added in the center to divide the distance into two equal parts (Figure 2 (ii)), the nominal distance for each crossing motion becomes $A/2$. So, if there are $N+1$ targets (Figure 2 (iii)), the nominal distance for each crossing motion is A/N. Thus, the entire difficulty, ID_N, using the effective parameters is

$$ID_N = N \times \log_2\left(\frac{A_e}{N \times W_e} + 1\right). \tag{8}$$

As N approaches ∞, this task is equivalent to steering (Figure 2 (iv)). Several previous studies have hypothesized that the number of movement corrections is finite [12, 15, 39, 43]. However, since effective parameters must be calculated from the variance of the trajectory over the entire path, we used the $N \to \infty$ formulation, which also has a theoretical justification [1]. We then obtain the following model by applying a first-order Taylor series expansion.

$$ID_\infty = \frac{A_e}{W_e \ln 2}. \tag{9}$$

As the constant of $\ln 2$ can be absorbed by the slope b in the steering law, we arrive at the following final model.

$$MT = a + bID_e, \; ID_e = \frac{A_e}{W_e}. \tag{10}$$

The calculations of W_e and A_e are described below.

3.2 Effective Width

The W_e is calculated using the standard deviation σ of the coordinates distribution (Equation 4). In pointing and crossing, the distribution of endpoint coordinates in the task-width direction of a set of trials is used to calculate σ [8, 32]. In the same manner, for steering, it is straightforward to use the distribution of coordinates along the task-width direction for calculating σ. Since a single steering task trial includes many crossing task trials according to Accot and Zhai's derivation, the coordinates from all of the sample points of a stroked trajectory can be used for this calculation (Figure 2). However, due to the difference between steering and multiple individual crossing trials, there are three possible derivations of σ for steering.

- Calculate σ from a single trajectory from each trial (σ_{trial}).
- Calculate σ for all trajectories for a single participant for each condition ($\sigma_{participant}$).
- Calculate σ from all trajectories of all participants for each condition ($\sigma_{condition}$).

The $\sigma_{condition}$ approach cannot smooth the user's subjective speed-accuracy biases that varies among participants, since it is not calculated for each participant. Also, it was not used in previous studies in pointing and crossing even though it is theoretically possible [8, 18, 24, 32], so we also did not adopt this for our work on steering. The $\sigma_{participant}$ measure is calculated in the same way as for pointing and crossing [8, 18, 24, 32]. For pointing and crossing, the σ_{trial} measure cannot be calculated from a single trial. However, since a single steering trial contains multiple crossing tasks, σ can be calculated from a single trial. Still, to maintain consistency with previous work in this area [46, 47], we decided to adopt σ_{trial} as the definition of σ for steering. In addition, σ_{trial} could potentially incorporate the effect of participant's speed-accuracy bias, which varies across trials. From this definition of σ and the calculation of Equation 4, W_e for steering becomes a width that encompasses 96% of the stroked trajectory for single trial. Even if a curved path is involved, the method of deriving W_e can be kept the same (Figure 2 (v)–(viii)). This is because W_e is calculated from the variance of the trajectory perpendicular to the direction of task movement within a single stroke. Hence, our definition of σ yields a consistent computation of effective width for steering tasks. For all steering tasks examined in this paper, W_e is thus the width that contains 96% of the stylus variance perpendicular to the direction of task movement of a single stroke.

We point out that the value of 96% is essentially immaterial, because this value can be changed by multiplying Equation 4 by a constant. In the steering law (Equation 2), the value is multiplied by the variable b, which can absorb any constant. Also, the performance of user groups or devices can be compared by TP calculated by Equation 6 as long as the same value of 96% is used.

3.3 Effective Amplitude

The A_e is calculated as the length of the path actually used by the user. In previous work on pointing and crossing, A_e is defined as the average distance along the task axis [18, 32]. That is, the dimensional information is reduced to be along the task axis. However, for steering (especially when involving curved paths), the total distance should be used as A_e, without reducing the dimension on the task axis (Figure 2 (v)–(viii)). The reason for this is that, when stroke is made close to the inner diameter of a curve, the actual stroked distance is obviously shorter than A. This aligns

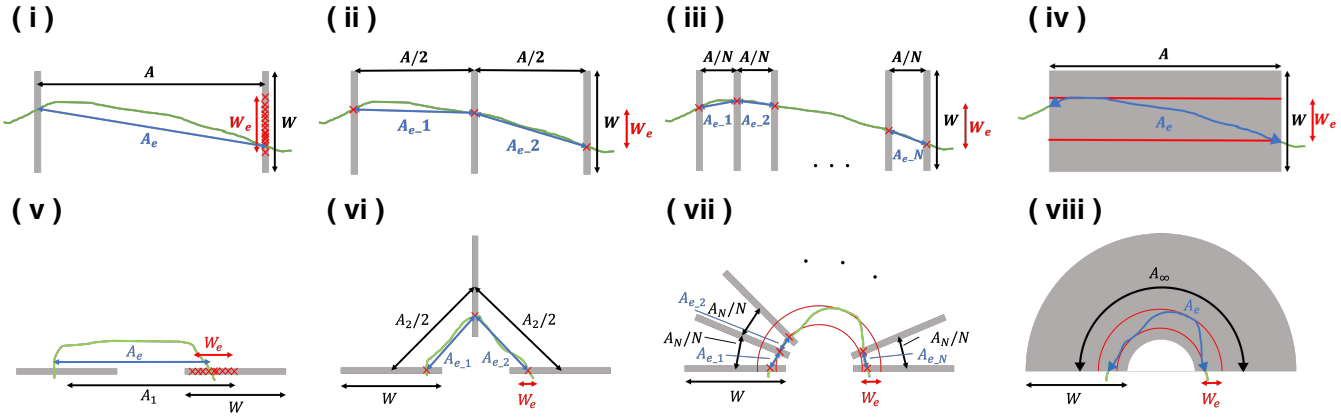


Figure 2: Definition of W_e and A_e for steering. The steering law is modeled as an infinite sequence of crossing tasks [1]. The W_e and A_e for crossing are defined as shown in (i) and (v), respectively [18]. Following these definitions, the W_e and A_e become (ii) and (vi) by adding one target between two targets. Then they become (iii) and (vii) by setting the number of splits to N. Finally, by letting N approach ∞, the definition of W_e and A_e for steering becomes (iv) and (viii). Thus, W_e for steering was defined as the width that encompasses 96% of the stroked trajectory. A_e for steering was defined as the total distance of the stroked trajectory in one trial. These definitions hold whether the path is straight or curved. This is because the definition of W_e is based on the center-line and A_e is based on the total length. Even if curves are involved, the width based on the distance from the center-line and the total length are uniquely determined, so our definition is consistent. When curves are involved, the total length of the task A (the sum of the distances from the center of the start line to the center of the goal line) varies depending on the number of divisions (v, vi, vii, viii). Let A_N be the total length when there are N divisions, then $A_N = A_\infty$ for $N \to \infty$ ((iii) $\to$ (iv) and (vii) $\to$ (viii)). The linear path is a particular case where the A is always the same length, even if the number of targets changes ($A_1 = A_2 = \cdots = A_N = A_\infty = A$) (i, ii, iii, iv).

with Accot and Zhai's definition of A_e for crossing as "the average actual movement amplitude" [3]. Therefore, we interpreted A_e in a single crossing task as the linear distance between the actual start and end points (Figure 2 (i)). This definition and a formulation through a limit $N \to \infty$ (Equation 9) lead us to adopt the total distance of the trajectory for A_e in steering (Figure 2 (iv)). This definition of A_e is consistent even if the path involves a curved path (Figure 2 (v)–(viii)).

4 EXPERIMENTS

4.1 Overview

To verify whether W_e and A_e are applicable to steering tasks with various path shapes, and to determine a better way to calculate TP for steering tasks, we conducted three independent experiments, each of which involved a different task and each conducted on a different day. In our derivation of W_e and A_e, they are calculated consistently for any path shape. W_e is calculated based on the variance of the stroked trajectory perpendicular to the direction of task movement, and A_e is calculated based on the total length of the stroked trajectory.

For all task shapes, participants were given three different speed-accuracy biases (Bias). Through this, we test whether effective parameters can smooth the effect of the speed-accuracy biases, and what the best way to calculate TP is.

4.2 Tasks

Since our effective parameters were derived based on linear steering in section 3, we decided to first examine a linear path shape (Linear) (Figure 1 (i)). For Linear, the task was to draw strokes from right to left within a linear path[1]. The task began when the tip of the stylus entered the white path area from the green area on the right and ended when the tip of the stylus reached the blue area on the left.

Then, because the steering law derived from the linear steering is directly applicable to circular steering [1, 2, 15], we added an experiment with a circular steering task (Circular) (Figure 1 (ii)). We assume that the effect of A_e should be more pronounced in circular steering, as many participants would try to shorten the movement distance by making a circle close(r) to the inner diameter in the Fast condition, resulting in $A \gg A_e$. For Circular, the task was to draw clockwise circular strokes inside a circular path. The task began when the tip of the stylus crossed the black line displayed at the center top of the circular path from left to right. The task ended when the tip of the stylus went around the path and crossed the line once again.

Finally, we added a sine-wave steering task (Sine-wave) (Figure 1 (iii)), which requires more frequent directional changes, in contrast to the Circular task, which is only clockwise. Further, as the Sine-wave condition would prevent the participants from occluding any part of the future path by their hand, participants

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

would need to adhere more strictly to the bias instruction and thus we could be able to more accurately examine the potential of using effective parameters. For Sine-wave, the task was to draw sine-wave strokes from right to left inside the path, and thus all participants could see the future path without occlusion by their hand. Paths were generated so that they always had the same width perpendicular to the direction of the task movement (i.e., perpendicular to the center-line). Thus, the inner curve of the sine-wave overlaps, which yields a sharp feature. Such features do not cause issues for steering along a sine-wave-shaped path, as they do not interfere with the path width itself, only with the visual appearance. The task began when the tip of the stylus entered the white path area from the green area on the right and ended when the tip of the stylus reached the blue area on the left.

For all task shapes, participants had to keep the tip of the stylus on the screen during the task. If the tip of the stylus deviated from the path, the participant had to return to the path as quickly as possible. During the task, the trajectory of the tip of the stylus was displayed in green inside the path and in red outside the path.

4.3 Participants

Twelve right-handed university students participated in the study (4 females and 8 males, mean age 21.0, standard deviation 0.707 years). All participants performed all three experiments, except that we had to replace one participant for the Sine-wave experiment for reasons unrelated to the experiment. They received the equivalent of 13 USD for compensation.

4.4 Apparatus

We used a laptop (Intel Core i7-11800H, GeForce RTX 3070 Laptop, 16GB RAM, Windows 10), LCD tablet (Wacom Cintiq 22, IPS, 344 × 193 mm, 1920 × 1080 pixels), and stylus (Wacom Pro Pen 2). The system was made with Unity and displayed in full screen mode.

4.5 Design

For each task, we used a $3_{\text{Bias}} \times 4_A \times 4_W$ repeated-measures design. The within-subjects factors were Bias, A, and W.

For Bias, we used three different speed-accuracy biases, Accurate, Neutral, and Fast. Previous pointing studies had used more extreme biases, such as *max accuracy* to ask participants to "try to bring the cursor exactly to the target (zero pixel error)" or *max speed* to ask them to "terminate the movements on average in the vicinity of the target" [27]. Such instructions to maximize the speed, i.e., to ignore the target almost completely, seem excessive for device/participant evaluation studies that explore typical user operations/movements. As our motivation is to examine if the *TP* and effective parameters can smooth the influence of participants' subjectively arising speed-accuracy biases for typical user movements, performing the task "as fast and as accurate as possible," as well as one speed- and one accuracy-emphasized instructions should sufficiently explore our current research objective. We did not provide quantitative speed-accuracy bias information to participants, but we encouraged a change in subjective speed-accuracy bias through verbal instructions. The reason for this is that the main purpose of our research is to establish a performance metric that smooths the effects of implicit and subjective bias in user experiments.

We used 3.968, 5.456, 8.928, and 22.32 mm for task width[2] W. By including very small to very large task widths, we verified the usefulness of W_e and A_e more extensively.

For A, we used four different task amplitude (= path length). For Linear and Circular, we adjusted the value of A so that the ID_n values were spread out as much as possible, and we used 89.28, 119.0, 173.6, and 210.8 mm for task amplitude[2] A. However, for Sine-wave, to prevent the potentially confounding effects of different trial-end position in a sine-wave cycle on the task outcomes, we set As so that the start and end positions were located at the peaks of the waves. Thus we used 80.60, 161.2, 241.8, and 322.4 mm for A in Sine-wave[2]. The *wavelength* and *amplitude* parameters specific to Sine-wave, were fixed at 68.02 mm and 19.89 mm for all conditions (Figure 1 (iii)), which meant that the path length A varied.

A task *set* comprised all combinations of 4A and 4W presented in random order, with 16 such *set*s for Linear and 11 such *set*s for Circular and Sine-wave, all of which were performed for each Bias condition (with the first one considered as practice). As circular-path and sine-wave steering takes considerably longer than a linear task [2], we changed the number of *set*s from 16 to 11 to reduce participant fatigue, which could negatively affect performance.

The 12 participants were randomly divided into two groups of six. 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., Neutral 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 previous studies [18, 38, 42].

4.6 Procedure

First, and for each of the three tasks, we explained the task to participants, i.e., to draw strokes without mistakes (without deviating from the path). Participants sat in a chair in a comfortable pose and operated the tablet tilted approximately 19 degrees from the desk. Before starting each Bias condition, instructions for each Bias were given to the participants. Participants were instructed to complete the task "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. In addition, the current Bias label was always displayed in the upper right corner of display during the task. To signal success or an error in the task, audio feedback was presented when the stylus was lifted. Previous study had shown that the type of feedback (visual, auditory, or tactile) affects the results in *MT* and *ER* only slightly in steering tasks [35], and thus we provided auditory error feedback when the participant deviated even by one pixel. Since the feedback in this study is identical across all task shapes, the effects of feedback on user performance or speed-accuracy biases should be consistent throughout the experiments. If the stylus was

[2]To reduce the effect of round-off errors, we chose round numbers for the pixels values for the task dimensions. To still enable others to replicate our work on different hardware, we list the As and Ws with the necessary precision in absolute units.

lifted during a trial, the trial had to be restarted from the right start area. When the stylus was lifted after reaching the left end area, whether the task was successful or not, a button labeled "Next" was displayed, as well as the time taken to complete the trial and the number of deviations from the path. By pressing the *Next* button, participants then proceeded to the next trial. Participants were encouraged to take a break at any time between trials, whenever they felt tired.

4.7 Measurement

The position of the tip of the stylus was recorded with a time stamp during the trials at a 240 Hz sampling rate. The dependent variables were *ER*, *MT*, σ, A_e, and *TP*. *ER* was the percentage of trials in which the tip of the stylus deviated from the path at least once. *MT* was the time taken for each trial to complete. For σ, we computed the standard deviation from the center line of the path. W_e was then calculated by multiplying σ by 4.133. Thus, W_e represents the width containing 96% of the stroked trajectory. A_e was the total distance of the stroked trajectory. *TP* was calculated by MT/ID, and three types of *TP* were analyzed: TP_n derived via ID_n as calculated from W and A, TP_{we} from ID_{we} as calculated via W_e and nominal A, and TP_e derived via ID_e as calculated from W_e and A_e.

5 RESULTS AND DISCUSSIONS

To make our results easier to understand, we present the analysis of all three experiments as one here, and only distinguish among the tasks in the following. For all task shapes, the first set for each Bias condition was considered as practice, and the remaining trials were analyzed. For Linear 8640 trials ($3_{\text{Bias}} \times 4_A \times 4_W \times 15$ sets $\times$12 participants), and for Circular and Sine-wave steering tasks 5760 trials ($3_{\text{Bias}} \times 4_A \times 4_W \times 10$ sets $\times$12 participants) were used for analysis. For all measures, the mean value for each participant was calculated and analyzed. All of the trials including errors were used for all analysis, similar to previous pointing and crossing research [18, 23, 24]. The reason for including all trials (including erroneous ones) is that W_e enables the calculation of the overall effective task difficulty. Since ANOVA is robust against mild normality violation [9, 25], we analyzed the data via RM-ANOVA. For the post-hoc test, we analyzed data via pairwise tests with Bonferroni p-value adjustments.

5.1 Error Rate *ER*

In total, we observed 922 errors (*ER* = 10.7%) for Linear, 1236 (*ER* = 21.5%) for Circular, and 1331 (*ER* = 23.1%) for Sine-wave. The corresponding ANOVA results are presented in Table 1. We found significant main effects for all independent variables (Bias, A, and W). Significant interactions were observed for almost all combinations (except Bias $\times$ A $\times$ W for Circular). Also, for all task shapes, we identified significant differences for all Bias pairs for *ER*s (Figure 3). This indicates that participants varied their accuracy depending on the Bias condition. Comparing *ER*s across task shapes[3], tasks involving curves (Circular and Sine-wave) showed higher *ER*s in the Fast condition. However, the difference of *ER* in the Accurate condition was not as large as that in the Fast condition. These suggest that even tasks including curves can be completed accurately if instructed, but are more challenging to complete at high speeds without errors.

[3]Since the three tasks are not directly comparable due to different experimental conditions, we just compare the trends.

5.2 Movement Time MT

For each task shape, the mean *MT* was 762.4 ms for Linear, 1710 ms for Circular, 3155 ms for Sine-wave. The corresponding ANOVA results are presented in Table 2. Significant main effects and interactions were found for all independent variables and interactions. For all task shapes, we identified significant differences for all Bias pairs (Figure 4). That indicates that participants varied their speed depending on the Bias condition.

5.3 Standard Deviation σ of Stroked Trajectory to Compute W_e

The σ of each trial was averaged over each condition for analysis, i.e., we used 48 ($3\text{Bias} \times 4A \times 4W$) σ values for each participant. The corresponding ANOVA results are presented in Table 3. Significant main effects were found for all independent variables, but not all interactions were significant.

In addition, we validated the index of utilization (I_u) to verify the utilization of path width. The I_u is an metric proposed by Zhai et al. [44] and is calculated as follows:

$$I_u = \log_2\left(\frac{W_e}{W}\right). \tag{11}$$

By this calculation, I_u for pointing can represent the ratio of the target width utilized. For steering, I_u can represent the ratio of the path width utilized. If I_u is greater than 0, it means that an excess width over the path width is being used; if I_u is less than 0, it means that a width less than the path width is being used. The corresponding ANOVA results are presented in Table 4. Significant main effects were found for all independent variables, but not all interactions were significant.

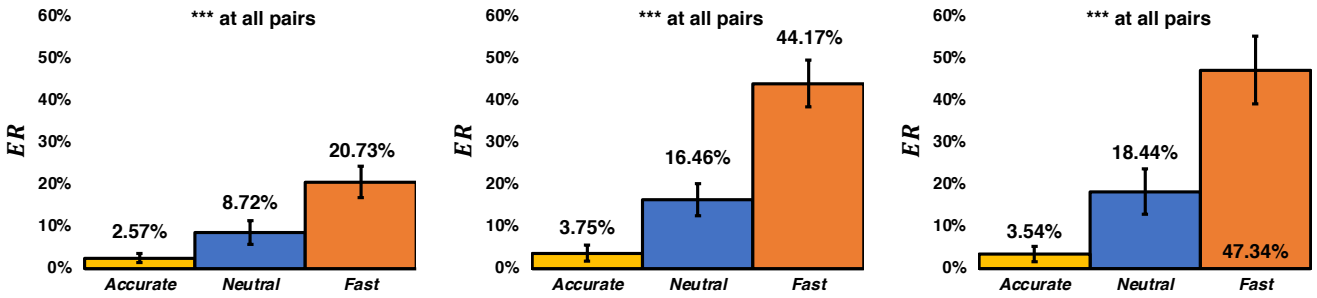
For all task shapes, the differences in I_u between Biases were statistically significant for all combinations and I_u tended to be larger when users emphasize speed more (Figure 5). In addition, I_u tended to increase as ID_n increased and the I_u differences between Bias increased as ID_n increased (Figure 6). These results indicate that the utilization of the paths increases as the task tended to be more difficult, and this trend strengthens as the user emphasizes speed. Only for Linear, the utilization of path width was very small in the low-ID_n conditions. This suggests that easy Linear path strokes, with low-ID_n condition, do not require a directional change, unlike Circular or Sine-wave strokes, and thus can complete the task with small positional variations of the stylus. Furthermore, in most conditions, the value of I_u was negative, suggesting that a width smaller than 96% of the path width was actually used in these steering tasks.

5.4 Total Distance of the Stroked Trajectory (A_e)

The A_e of each trial was averaged over each condition for analysis, i.e., we used 48 ($3\text{Bias} \times 4A \times 4W$) A_e values for each participant. The corresponding ANOVA results are presented in Table 5. Significant main effects were found for all independent variables.

Table 1: The ANOVA results for *ER*. Throughout this paper, *, **, and * in the tables and graphs indicate $p < .001$, $p < .01$, and $p < .05$, respectively. In the Table 1–5 and 8, yellow cells indicate $p < .05$.**

	Linear			Circular			Sine-wave		
Factor	F	p	η_p^2	F	p	η_p^2	F	p	η_p^2
Bias	$F_{2,22} = 67.2$	$4.26 \times 10^{-10}_{***}$	.859	$F_{2,22} = 194$	$1.09 \times 10^{-14}_{***}$	.946	$F_{2,22} = 82.3$	$6.11 \times 10^{-11}_{***}$	.882
A	$F_{3,33} = 68.7$	$2.83 \times 10^{-14}_{***}$	.862	$F_{3,33} = 18.3$	$3.59 \times 10^{-7}_{***}$	.625	$F_{3,33} = 102$	$9.29 \times 10^{-17}_{***}$	.903
W	$F_{3,33} = 121$	$7.12 \times 10^{-18}_{***}$	.917	$F_{3,33} = 174$	$2.61 \times 10^{-20}_{***}$	.941	$F_{3,33} = 110$	$2.83 \times 10^{-17}_{***}$	.909
Bias × A	$F_{6,66} = 17.3$	$6.33 \times 10^{-12}_{***}$	.612	$F_{6,66} = 4.07$	$1.56 \times 10^{-3}_{**}$	.270	$F_{6,66} = 10.7$	$2.85 \times 10^{-8}_{***}$	.493
Bias × W	$F_{6,66} = 41.0$	$2.07 \times 10^{-20}_{***}$	.788	$F_{6,66} = 72.1$	$4.77 \times 10^{-27}_{***}$	.868	$F_{6,66} = 47.3$	$5.02 \times 10^{-22}_{***}$	.811
$A \times W$	$F_{9,99} = 16.6$	$2.43 \times 10^{-16}_{***}$	.602	$F_{9,99} = 3.77$	$4.12 \times 10^{-4}_{***}$	.255	$F_{9,99} = 8.70$	$1.59 \times 10^{-9}_{***}$	.442
Bias × $A \times W$	$F_{18,198} = 4.05$	$3.79 \times 10^{-7}_{***}$	.269	$F_{18,198} = 1.52$	8.48×10^{-2}	.122	$F_{18,198} = 2.88$	$1.61 \times 10^{-4}_{***}$	.207

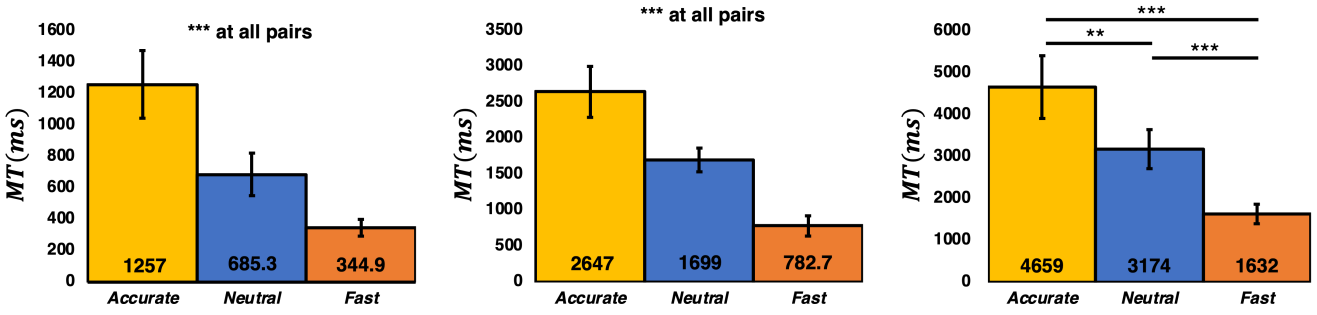


(a) Linear (b) Circular (c) Sine-wave

Figure 3: The error rate (*ER*) for the three tasks. Throughout this paper, the error bars in the graphs indicate 95% CIs. Accuracy of participants varied with their given speed-accuracy biases.

Table 2: The ANOVA results for *MT*

	Linear			Circular			Sine-wave		
Factor	F	p	η_p^2	F	p	η_p^2	F	p	η_p^2
Bias	$F_{2,22} = 55.4$	$2.59 \times 10^{-9}_{***}$	.834	$F_{2,22} = 101$	$8.27 \times 10^{-12}_{***}$	.902	$F_{2,22} = 46.7$	$1.22 \times 10^{-8}_{***}$	.809
A	$F_{3,33} = 212$	$1.29 \times 10^{-21}_{***}$	.951	$F_{3,33} = 276$	$1.99 \times 10^{-23}_{***}$	.962	$F_{3,33} = 323$	$1.64 \times 10^{-24}_{***}$	.967
W	$F_{3,33} = 178$	$1.95 \times 10^{-20}_{***}$	.942	$F_{3,33} = 367$	$2.11 \times 10^{-25}_{***}$	.971	$F_{3,33} = 175$	$2.46 \times 10^{-20}_{***}$	.941
Bias × A	$F_{6,66} = 51.0$	$6.60 \times 10^{-23}_{***}$	.823	$F_{6,66} = 74.8$	$1.69 \times 10^{-27}_{***}$	.872	$F_{6,66} = 41.8$	$1.27 \times 10^{-20}_{***}$	.792
Bias × W	$F_{6,66} = 50.3$	$9.88 \times 10^{-23}_{***}$	.820	$F_{6,66} = 90.2$	$7.36 \times 10^{-30}_{***}$	.891	$F_{6,66} = 41.2$	$1.81 \times 10^{-20}_{***}$	.789
$A \times W$	$F_{9,99} = 126$	$3.34 \times 10^{-50}_{***}$	.920	$F_{9,99} = 165$	$1.88 \times 10^{-55}_{***}$	.937	$F_{9,99} = 128$	$1.60 \times 10^{-50}_{***}$	.921
Bias × $A \times W$	$F_{18,198} = 36.4$	$6.37 \times 10^{-53}_{***}$	.768	$F_{18,198} = 36.7$	$3.61 \times 10^{-53}_{***}$	.769	$F_{18,198} = 31.1$	$5.81 \times 10^{-48}_{***}$	.739



(a) Linear (b) Circular (c) Sine-wave

Figure 4: The movement time (*MT*) for the three tasks. The speed of participants varied with their given speed-accuracy biases.

Table 3: The ANOVA results for σ

	Linear			Circular			Sine-wave		
Factor	F	p	η_p^2	F	p	η_p^2	F	p	η_p^2
Bias	$F_{2,22} = 28.2$	$8.53 \times 10^{-7}_{***}$	.719	$F_{2,22} = 91.1$	$2.27 \times 10^{-11}_{***}$	.892	$F_{2,22} = 70.0$	$2.91 \times 10^{-10}_{***}$	.864
A	$F_{3,33} = 326$	$1.37 \times 10^{-24}_{***}$	.967	$F_{3,33} = 18.4$	$3.36 \times 10^{-7}_{***}$	.626	$F_{3,33} = 32.4$	$6.05 \times 10^{-10}_{***}$	.746
W	$F_{3,33} = 150$	$2.77 \times 10^{-19}_{***}$	.932	$F_{3,33} = 180$	$1.65 \times 10^{-20}_{***}$	.942	$F_{3,33} = 273$	$2.32 \times 10^{-23}_{***}$	.961
Bias $\times A$	$F_{6,66} = 21.3$	$9.84 \times 10^{-14}_{***}$	.659	$F_{6,66} = 6.27$	$3.00 \times 10^{-5}_{***}$	.363	$F_{6,66} = 1.92$	9.05×10^{-2}	.149
Bias $\times W$	$F_{6,66} = 1.96$	8.35×10^{-2}	.151	$F_{6,66} = 5.11$	$2.33 \times 10^{-4}_{***}$	.317	$F_{6,66} = 4.76$	$4.35 \times 10^{-4}_{***}$	.302
$A \times W$	$F_{9,99} = 56.7$	$4.16 \times 10^{-35}_{***}$	.837	$F_{9,99} = 4.60$	$4.37 \times 10^{-5}_{***}$	.295	$F_{9,99} = 1.06$	4.01×10^{-1}	.0877
Bias $\times A \times W$	$F_{18,198} = 1.04$	4.18×10^{-1}	.0863	$F_{18,198} = .888$	5.94×10^{-1}	.0747	$F_{18,198} = .333$	9.96×10^{-1}	.0294

Table 4: The ANOVA results for the index of utilization (I_u) of W

	Linear			Circular			Sine-wave		
Factor	F	p	η_p^2	F	p	η_p^2	F	p	η_p^2
Bias	$F_{2,22} = 26.4$	$1.43 \times 10^{-6}_{***}$	.706	$F_{2,22} = 108$	$4.30 \times 10^{-12}_{***}$	.907	$F_{2,22} = 85.9$	$4.01 \times 10^{-11}_{***}$	.887
A	$F_{3,33} = 270$	$2.72 \times 10^{-23}_{***}$	.961	$F_{3,33} = 19.3$	$2.08 \times 10^{-7}_{***}$	.637	$F_{3,33} = 69.3$	$2.47 \times 10^{-14}_{***}$	.863
W	$F_{3,33} = 610$	$5.79 \times 10^{-29}_{***}$	.982	$F_{3,33} = 242$	$1.53 \times 10^{-22}_{***}$	.957	$F_{3,33} = 64$	$7.42 \times 10^{-14}_{***}$	.854
Bias $\times A$	$F_{6,66} = 8.79$	$4.82 \times 10^{-7}_{***}$	.444	$F_{6,66} = 3.92$	$2.07 \times 10^{-3}_{**}$	.263	$F_{6,66} = .403$	8.74×10^{-1}	.0354
Bias $\times W$	$F_{6,66} = 7.36$	$4.74 \times 10^{-6}_{***}$	.401	$F_{6,66} = 24.3$	$5.44 \times 10^{-15}_{***}$	.689	$F_{6,66} = 26.8$	$6.31 \times 10^{-16}_{***}$	.709
$A \times W$	$F_{9,99} = 21.0$	$2.11 \times 10^{-19}_{***}$	.657	$F_{9,99} = 2.49$	$1.30 \times 10^{-2}_{*}$	.185	$F_{9,99} = 2.72$	$7.06 \times 10^{-3}_{**}$	.198
Bias $\times A \times W$	$F_{18,198} = 1.30$	1.91×10^{-1}	.106	$F_{18,198} = .847$	6.43×10^{-1}	.0715	$F_{18,198} = .537$	9.38×10^{-1}	.0465

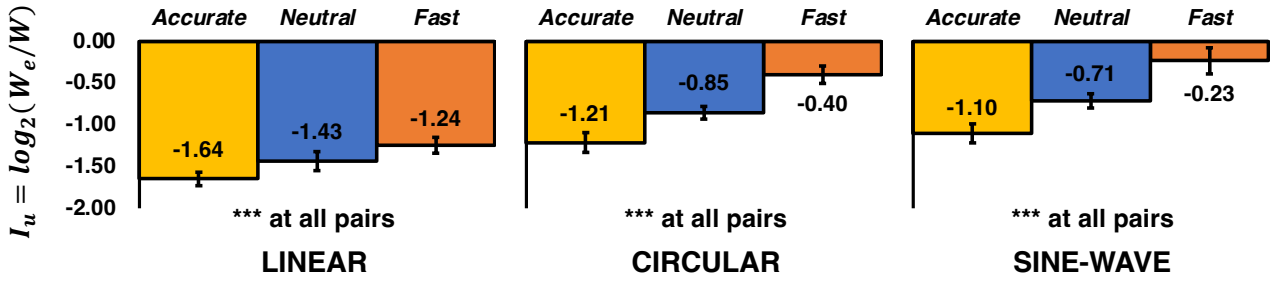


Figure 5: I_u of W between Bias for the three tasks. The I_u consistently tended to be larger when the Bias emphasizes speed more and the differences were statistically significant. These indicating that the wider width were used when Bias emphasize speed.

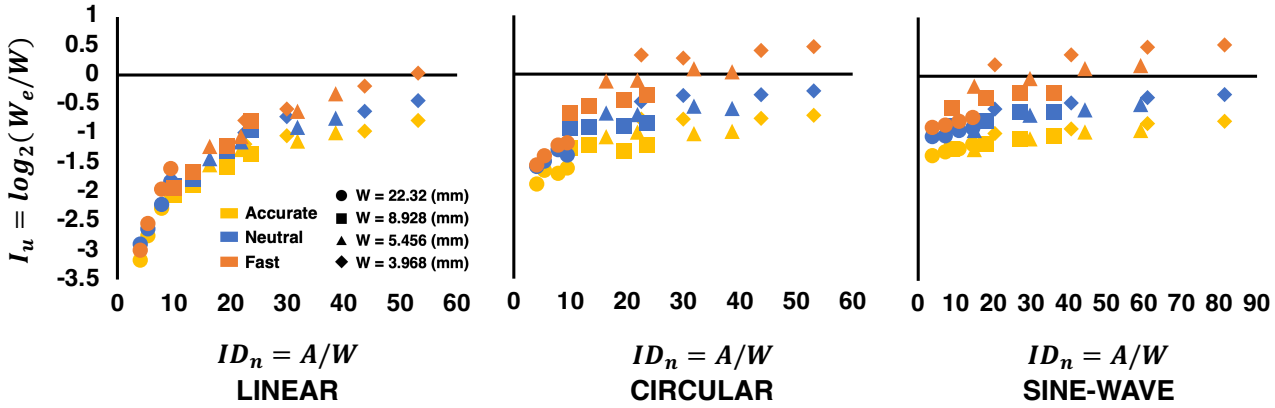


Figure 6: ID_n vs I_u of W for the three tasks. The I_u differences between Biases became larger when speed was emphasized. The I_u values were mostly negative, suggesting that the task width was not fully used in the steering tasks.

Table 5: The ANOVA results for A_e

	Linear			Circular			Sine-wave		
Factor	F	p	η_p^2	F	p	η_p^2	F	p	η_p^2
Bias	$F_{2,22} = 74.0$	$1.70 \times 10^{-10}_{***}$	.871	$F_{2,22} = 7.46$	$3.37 \times 10^{-3}_{**}$	.404	$F_{2,22} = 12.2$	$2.77 \times 10^{-4}_{***}$	.525
A	$F_{3,33} = 5.53 \times 10^6$	$3.97 \times 10^{-94}_{***}$	1.00	$F_{3,33} = 2.33 \times 10^4$	$6.09 \times 10^{-55}_{***}$	1.00	$F_{3,33} = 2.14 \times 10^5$	$7.75 \times 10^{-71}_{***}$	1.00
W	$F_{3,33} = 112$	$2.34 \times 10^{-17}_{***}$	.910	$F_{3,33} = 92.3$	$3.99 \times 10^{-16}_{***}$	.893	$F_{3,33} = 103$	$7.72 \times 10^{-17}_{***}$	.904
Bias × A	$F_{6,66} = .391$	8.82×10^{-1}	.0343	$F_{6,66} = 7.47$	$3.96 \times 10^{-6}_{***}$	.404	$F_{6,66} = 13.1$	$1.12 \times 10^{-9}_{***}$	.543
Bias × W	$F_{6,66} = 6.11$	$3.97 \times 10^{-5}_{***}$	.357	$F_{6,66} = 7.54$	$3.54 \times 10^{-6}_{***}$	.407	$F_{6,66} = 2.44$	$3.42 \times 10^{-2}_{*}$	.182
$A \times W$	$F_{9,99} = 1.58$	1.30×10^{-1}	.126	$F_{9,99} = 2.25$	$2.48 \times 10^{-2}_{*}$	.170	$F_{9,99} = 61.9$	$1.07 \times 10^{-36}_{***}$	.849
Bias × $A \times W$	$F_{18,198} = 1.84$	$2.34 \times 10^{-2}_{*}$	.143	$F_{18,198} = 2.31$	$2.71 \times 10^{-3}_{**}$	.173	$F_{18,198} = 1.68$	$4.54 \times 10^{-2}_{*}$	.132

Table 6: The ANOVA results for the index of utilization (I_u) of A

	Linear			Circular			Sine-wave		
Factor	F	p	η_p^2	F	p	η_p^2	F	p	η_p^2
Bias	$F_{2,22} = 79.9$	$8.19 \times 10^{-11}_{***}$	.879	$F_{2,22} = 10.3$	$6.91 \times 10^{-4}_{***}$	.484	$F_{2,22} = 11.0$	$4.93 \times 10^{-4}_{***}$	.500
A	$F_{3,33} = 206$	$1.89 \times 10^{-21}_{***}$	.949	$F_{3,33} = 6.42$	$1.51 \times 10^{-3}_{**}$	.369	$F_{3,33} = 22.9$	$3.39 \times 10^{-8}_{***}$	.675
W	$F_{3,33} = 116$	$1.31 \times 10^{-17}_{***}$	.913	$F_{3,33} = 68.8$	$2.76 \times 10^{-14}_{***}$	.862	$F_{3,33} = 103$	$8.41 \times 10^{-17}_{***}$	.903
Bias × A	$F_{6,66} = 38.8$	$8.58 \times 10^{-20}_{***}$	.779	$F_{6,66} = 10.9$	$2.25 \times 10^{-8}_{***}$	.497	$F_{6,66} = 1.64$	1.50×10^{-1}	.130
Bias × W	$F_{6,66} = 5.09$	$2.42 \times 10^{-4}_{***}$	.316	$F_{6,66} = 7.26$	$5.64 \times 10^{-6}_{***}$	.397	$F_{6,66} = 2.20$	5.35×10^{-2}	.167
$A \times W$	$F_{9,99} = 43.9$	$1.07 \times 10^{-30}_{***}$	.800	$F_{9,99} = 18.3$	$1.46 \times 10^{-17}_{***}$	.625	$F_{9,99} = 1.68$	1.05×10^{-1}	.132
Bias × $A \times W$	$F_{18,198} = 1.91$	$1.70 \times 10^{-2}_{*}$	.148	$F_{18,198} = 3.73$	$1.95 \times 10^{-6}_{***}$	.253	$F_{18,198} = .899$	5.80×10^{-1}	.0756

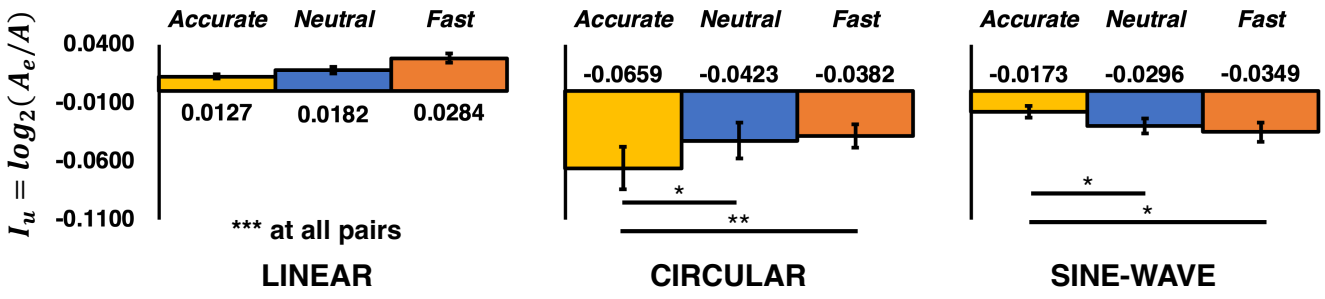


Figure 7: I_u of A between Bias for the three tasks. For Linear and Circular, the I_u tended to be larger when the Bias emphasized speed more. For Sine-wave, the I_u tended to be smaller when Bias emphasized speed more.

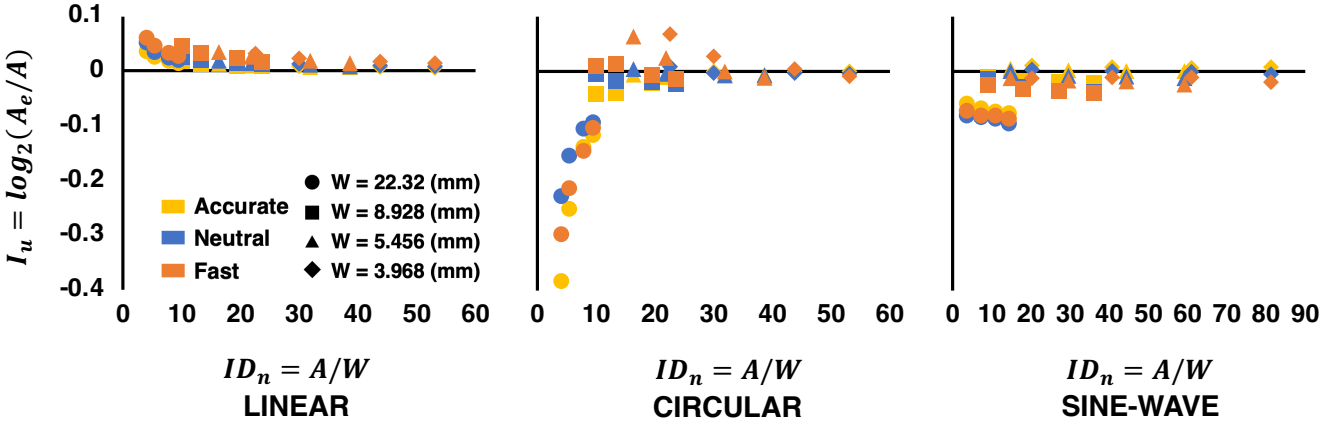


Figure 8: ID_n vs I_u of A for the three tasks. For Linear, the I_u increased as ID_n decreased. For Circular and Sine-wave, the I_u decreased as ID_n decreased. Especially in Circular, A_e can describe the shortening of the trajectory length due to the wider W.

Similar to the I_u for W, we also analyzed the index of utilization (I_u) for A. We calculated the I_u for A as follows:

$$I_u = \log_2\left(\frac{A_e}{A}\right). \quad (12)$$

The corresponding ANOVA results are presented in Table 6. Significant main effects were found for all independent variables and significant interaction were found for all combination in Linear and Circular.

The trend of I_u between Bias depended on the task shape (Figure 7). For Linear and Circular, the I_u tended to be larger when participants emphasize speed more, meanwhile the I_u tended to be smaller for Sine-wave. These trends were statistically significant at all combinations for Linear, but not significant between Neutral and Fast for Circular and Sine-wave. In addition, for Linear, A_e increased as ID_n decreased, and for Circular and Sine-wave, A_e decreased as ID_n decreased (Figure 8). These results suggest that for Sine-wave, when participants emphasize speed more, participants made strokes that shorten the distance, while this did not apply for Linear and Circular.

For Linear, since A is the minimum distance to complete the task, it is likely that the faster the stylus was moved, which increased the positional variability, this increased A_e. The increase in A_e was more pronounced for smaller ID_n or faster Bias, which suggest that the positional variability of stylus increases with faster manipulation.

For Circular, although participants can shorten the distance by stroking near the inner diameter of the path, the path minimization strategy did not seem to have been adopted when Bias emphasized speed more (Figure 7-(middle)). Rather, participants focused on moving the stylus faster, which seemed to lengthen the strokes (i.e., increase the positional variability of the stylus). However, in the widest-W condition, I_u was notably reduced (Figure 8-(middle)). These results suggest that in the wide-W conditions, users tried to reduce the movement time by stroking the inner diameter of the path, but this may not have affected results due to the Bias. Therefore, while A_e for Circular can account for the effects of short length trajectories in wide-W conditions, it likely cannot account for the effects of strategic trajectory changes due to the Bias.

For Sine-wave, the more the Bias emphasized speed, the smaller the I_u and the shorter the trajectories (Figure 7-(right)). In addition, the wider the W and the longer the A, the smaller the I_u (Figure 8-(right)). These results suggest that for Sine-wave, the participants completed the task quickly by strategically shortening the trajectory.

5.5 Model Fitting

We analyzed the model fit by running linear regressions for MT against ID with two approaches. One analyzes the data for the three Bias conditions separately (Accurate, Neutral, and Fast); here each Bias has $4_A \times 4_W = 16$ fitting points for regression. The other analyzes all $3_{\text{Bias}} \times 4_A \times 4_W = 48$ fitting points in a mixed manner (Mixed). The regression analysis with ID was performed via IBM SPSS Statistics 29. Three models were analyzed: an ID_n model using W and A, an ID_{we} model using W_e and A, and an ID_e model using W_e and A_e. Since the number of free parameters is two for all models, we used non-adjusted R^2 values. In addition, to analyze the fits in a comparative manner, we used the AIC measure [4]. The lower the AIC, the better the fit, and a difference of 2 or more is considered to be significant [6].

For Linear, the ID_n model showed the highest R^2 and lowest AIC for each Bias, while the ID_{we} and ID_e models showed higher R^2 and lower AIC for Mixed, with significant differences (Table 7 and Figure 9). Looking at the difference between the ID_{we} and ID_e models, the ID_{we} model consistently showed a higher R^2 and smaller AIC for all Biases and Mixed, but the AIC differences were smaller than 2, which indicates that these are not significant.

For Circular, the ID_n model for Accurate and Neutral, the ID_e model for Fast, and the ID_{we} and ID_e for Mixed showed the highest (or higher) R^2 and lowest (or lower) AIC, with significant differences (Table 7 and Figure 10). Comparing the ID_{we} and ID_e models, for each Bias, the ID_e model consistently showed a higher R^2 and lower AIC, with significant differences for Neutral and Fast, but not for Accurate. For Mixed, the ID_{we} model showed a higher R^2 and smaller AIC, but the difference was not significant.

For Sine-wave, the ID_n model for Accurate and Neutral, the ID_{we} model for Fast, and the ID_{we} and ID_e for Mixed showed the highest (or higher) R^2 and lowest (or lower) AIC, with significant differences (Table 7 and Figure 11). Comparing the ID_{we} and ID_e models, for each Bias, the ID_{we} model showed a higher R^2 and lower AIC for Neutral and Fast, with significant differences, and lower R^2 and higher AIC for Accurate, with no significant difference. For Mixed, the ID_e model showed a higher R^2 and smaller AIC, but this difference was not statistically significant.

To our knowledge, there are no studies that have verified the steering law can be applied directly to the sine-wave path shape. However, the steering law held well with $R^2 = .992$ (Neutral) (Table 7 and Figure 11). Furthermore, the use of effective parameters worked well to fit the data, and showed a high Bias smoothing effect with an $R^2 = .972$ for Mixed, which is substantially higher than for Linear (.890) or Circular (.943).

Consistently across all path shapes, the W_e improved the model fit to Mixed, but reduced the model fit to each individual Bias (except Sine-wave Neutral and Fast). These results largely match the results of previous studies on pointing and crossing [18, 44, 47], but extend these insights to steering. Similar to previous work, W_e seems to smooth out the speed-accuracy bias, at the expense of less robust movement time prediction accuracy for each individual speed-accuracy bias, even for steering. The improved model fit for Sine-wave Fast is similar to previous studies for steering, suggesting that W_e for steering might be even more robust than for pointing or crossing [19, 47]

On the other hand, the use of A_e did not improve the model fit for Mixed. Thus, A_e for steering had no effect on smoothing the speed-accuracy bias. Only for Circular did A_e improve the model fit to each individual Bias. In Circular, participant did not need to comply with the A constraint, and instead performed a small circular stroke close to the inner diameter of the overall path. Thus, the model fit of the ID_n model decreased, and A_e improved the model fit to each Bias.

In summary, the ID_n model is generally (except for Circular Fast and Sine-wave Fast) recommended for predicting the movement time of data with *individual speed-accuracy biases* in steering tasks. For data with multiple speed-accuracy biases, the ID_{we} model

Table 7: The results of model fitting. Blue cells and † indicate the best model with significant differences in terms of *AIC*.

		Linear				Circular				Sine-wave			
Model	Bias	*a*	*b*	R^2	*AIC*	*a*	*b*	R^2	*AIC*	*a*	*b*	R^2	*AIC*
ID_n model	Accurate	−110	62.3	.998	169.8†	−178	129	.999	168.0†	−312	167	.998	214.2†
	Neutral	33.3	29.7	.993	162.3†	17.7	76.6	.996	185.6†	91.8	103	.992	218.7†
	Fast	96.3	11.3	.953	163.4†	255	24.0	.945	190.0	452	39.6	.866	235.7
	Mixed	6.59	34.5	.501	732.7	31.5	76.5	.547	800.3	77.4	103	.637	853.7
ID_{we} model	Accurate	−1562	49.3	.958	215.4	−1104	81.7	.955	239.5	−1075	95.5	.984	245.2
	Neutral	−909	32.5	.939	197.7	−659	66.9	.965	219.1	−544	82.7	.991	220.5†
	Fast	−441	18.6	.705	192.6	−187	39.4	.932	193.2	−171	58.8	.975	208.9†
	Mixed	−1449	44.7	.890	659.9†	−983	76.4	.943	700.7†	−925	90.2	.972	731.1†
ID_e model	Accurate	−1597	49.6	.957	215.7	−885	78.6	.958	238.6	−993	94.4	.985	244.5
	Neutral	−942	32.8	.936	198.3	−577	65.6	.970	216.5	−349	58.9	.990	223.3
	Fast	−454	18.6	.679	194.0	−149	38.4	.963	183.6†	−142	58.9	.968	212.6
	Mixed	−1497	45.1	.888	660.9†	−880	74.8	.943	701.2†	−854	89.5	.972	729.9†

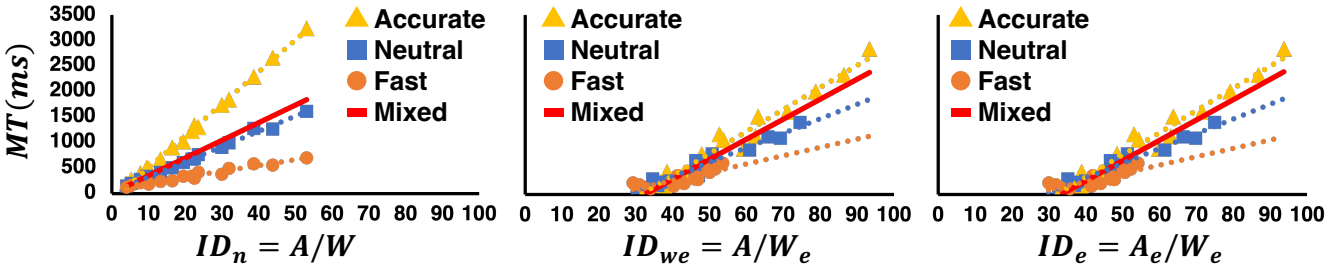


Figure 9: *ID* vs *MT* for Linear. The use of W_e smoothed the effect of speed-accuracy biases, at the expense of a lower model fit for each individual Bias. A_e had no effect on smoothing the speed-accuracy biases, nor on improving model fit at each Bias.

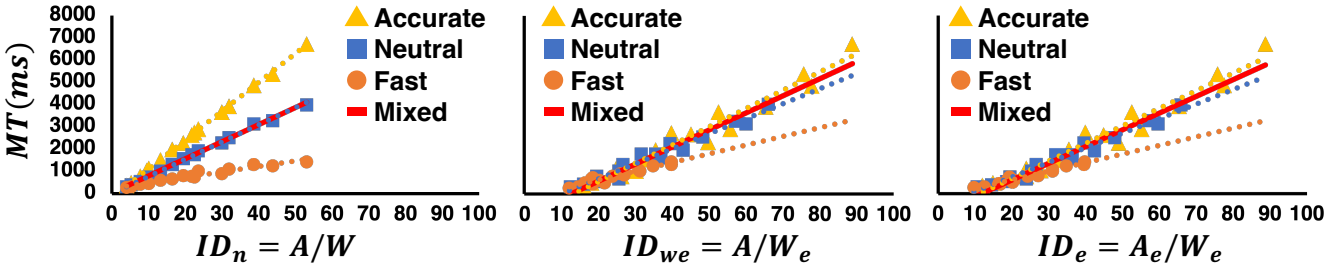


Figure 10: *ID* vs *MT* for Circular. W_e smoothed the effect of speed-accuracy biases at the expense of higher model fit in each Bias. A_e improved the model fit at each Bias, but did not smooth the speed-accuracy biases.

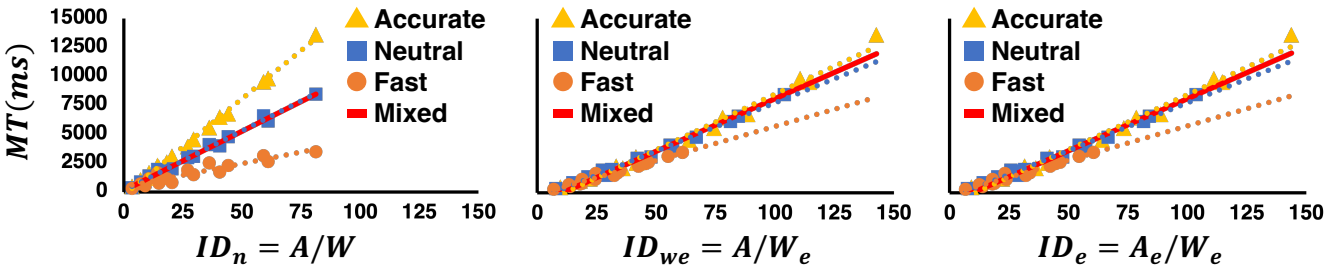


Figure 11: *ID* vs *MT* for Sine-wave. W_e smoothed the effect of speed-accuracy biases and improved the model fit only for Fast. A_e had little effect on smoothing the speed-accuracy biases and on improving the model fit for each Bias.

or ID_e model is recommended for predicting movement time, as this approach smooths the effects of the speed-accuracy biases. Effective amplitude improved the model fit for each individual speed-accuracy bias only for Circular steering.

5.6 Throughput

We analyzed three formulations of *TP*: TP_n calculated by ID_n, TP_{we} calculated via ID_{we}, and TP_e calculated by ID_e. The corresponding ANOVA results are presented in Table 8. In addition to the ANOVA, we analyzed the stability of *TP*s across speed-accuracy biases and task difficulties.

5.6.1 ***Stability across Speed-accuracy Biases.*** To investigate the stability across speed-accuracy biases, we tested with two methods; one used in previous studies [18, 27], and the other through a permutation test. For the first approach, we calculated the averaged *TP* for each Bias, and then analyzed the percentage of the difference between the maximum and minimum *TP* to the maximum *TP*, i.e., $100\% \times (TP_{\max} - TP_{\min})/TP_{\max}$. The smaller this value, the smaller the difference in *TP* between Biases, indicating a smoothing of the effects of speed-accuracy biases. In previous work, this measure was observed to be 42% for pointing [27], 44.74% for *crossing with a directional constraint* and 20.69% for *crossing with an amplitude constraint* [18]. In our study, for Linear the differences between the maximum and minimum *TP*s of each Bias relative to the maximum *TP*s were 68.6% for TP_n, 53.3% for TP_{we}, and 53.8% for TP_e (Figure 12). For Circular, the percentages were 65.7% for TP_n, 34.6% for TP_{we}, and 37.0% for TP_e (Figure 13), and for Sine-wave, 60.4% for TP_n, 24.9% for TP_{we}, and 24.3% for TP_e (Figure 14). These results showed that W_e smooths the effects of Bias for all task shapes. They also showed that TP_{we} and TP_e for Sine-wave are more stable metrics in terms of stability across speed-accuracy biases.

For the second approach, we performed a permutation test, shown in Table 9. To perform this test, we first calculated TP_n, TP_{we}, and TP_e for each participant, each Bias, and each task (i.e., $3Tasks \times 3\text{Bias}es \times 3TPs \times 12participants = 324TP\text{s}$). Then, for each task and *TP*, we extracted two Biases (e.g., Accurate and Neutral for TP_n in Experiment 1) and performed the permutation test. According to the outcome, only for TP_{we} and TP_e in Sine-wave, the difference between Accurate and Neutral were not significantly different, same as pairwise test results (Figures 12–14). These results still suggest that TP_{we} and TP_e for Sine-wave are more stable metrics in terms of stability across speed-accuracy biases.

5.6.2 ***Stability across Task Difficulties.*** To verify the stability across task difficulties, we calculated the coefficient of variation (*CV*) for each individual Bias. The *CV* was calculated by dividing the standard deviation (*SD*) by the mean. The smaller this value is, the smaller the fluctuation of *TP*, indicating that *TP* is stable across task difficulty. In a previous study in pointing, this measure was 13.2% for Neutral (no multiple bias condition was given to participants in said study) [23]. For almost all conditions (except Circular Fast and Sine-wave Fast), TP_n showed the highest stability across task difficulties, and W_e reduced the stability (Figures 15, 16, and 17, and Table 10). For Circular Fast, and Sine-wave Fast, W_e increased the stability. For Linear and Sine-wave, A_e slightly reduced or had little effect on the stability. On the other hand, only for Circular, A_e increased the stability.

For Linear and Circular, TP_n increased as the ID_n increased (Figures 15 and 16). For Sine-wave, TP_n increased as *W* decreased, and *A* did not have a large impact on TP_n (Figure 17). On the other hand, TP_{we} and TP_e decreased consistently across all task shapes as the ID_n increased (Figures 15, 16, and 17). For Circular, in largest *W* conditions, A_e decreased the *TP* and increased the *TP* stability across ID_ns.

5.6.3 ***The Effect of Effective Parameters on TP Stability.*** W_e increased the stability of *TP* across different Biases, but generally (except for Circular Fast and Sine-wave Fast) reduced the stability across ID_ns. The improved stability of *TP* across ID_ns in Circular Fast and Sine-wave Fast is most likely due to the higher *ER*s. The *ER* in Linear was 20.73% for Fast, while it exceeded 40% in Circular and Sine-wave. This indicates that participants did not or were unable to comply with the *W* constraint. As a result, the movement time was unjustifiably small for the small *W* conditions, and TP_n became unstable.

While A_e did not smooth the *TP* across different Biases, it improved the stability of *TP* between ID_ns only in the Circular. This is likely due to the same cause as the effect of A_e on model fitting.

In summary, in input device comparison experiments with steering tasks, if participants exhibit different subjective speed-accuracy biases and one needs to compare device performance by smoothing the effect of different biases, we recommend to use TP_{we} or TP_e. This is because, TP_{we} and TP_e are more stable metrics across different Biases than TP_n for all task shapes (Figures 12, 13, and 14), and can smooth the effects of participants' implicit subjective speed-accuracy bias. If researchers aim to investigate the performance only for different task difficulties, in an experiment with a single device and a specific speed-accuracy bias performed by a group of participants, we recommend using TP_n due to its stability across task difficulties. Only for Circular, effective amplitude should be used to improved the stability of throughput between task difficulties for each individual speed-accuracy bias.

5.7 Stylus Speed and Stylus Positional Variability

To investigate the reason of unstable TP_{we} and TP_e across ID_n, we analyzed the stylus speed and positional variability (Figures 18, 19, and 20). A previous study on crossing tasks had identified that possible causes of *TP* instability are: an unstable cursor speed across trials, a small endpoint distribution despite high cursor speed, and/or no cursor deceleration [18]. Since steering is a series of crossings [1], a similar analysis might be sensible.

For Linear, the stylus speed was very high in low ID_n conditions, and unstable across trials (Figure 18). Meanwhile, positional variability was very small and stable across trials. These results showed that users were able to complete the low-ID_n Linear task with a high speed movement without high positional variability, which may be the reason for the unstable TP_{we} and TP_e across ID_ns, which matches the observations from a previous study on crossing [18].

For Circular, the stylus speed was not very high, lower than in Linear, and stable across trials even in low ID_n condition (Figure 19). In addition, the positional variability was much bigger than in

Table 8: The ANOVA results for *TP*

		Linear			Circular			Sine-wave		
TP	*Factor*	F	p	η_p^2	F	p	η_p^2	F	p	η_p^2
TP_n	Bias	$F_{2,22} = 87.8$	$3.26 \times 10^{-11}_{***}$	.889	$F_{2,22} = 107$	$4.48 \times 10^{-12}_{***}$	.907	$F_{2,22} = 90.0$	$2.55 \times 10^{-11}_{***}$	.891
	A	$F_{3,33} = .473$	7.03×10^{-1}	.0412	$F_{3,33} = 15.8$	$1.56 \times 10^{-6}_{***}$	.589	$F_{3,33} = 30.3$	$1.32 \times 10^{-9}_{***}$	.734
	W	$F_{3,33} = 35.1$	$2.19 \times 10^{-10}_{***}$	.762	$F_{3,33} = 33.6$	$3.87 \times 10^{-10}_{***}$	.753	$F_{3,33} = 55.0$	$6.29 \times 10^{-13}_{***}$	.833
	Bias $\times A$	$F_{6,66} = 1.01$	4.27×10^{-1}	.0841	$F_{6,66} = 26.0$	$1.23 \times 10^{-15}_{***}$	.703	$F_{6,66} = 6.46$	$2.15 \times 10^{-5}_{***}$	.370
	Bias $\times W$	$F_{6,66} = 45.4$	$1.48 \times 10^{-21}_{***}$	.805	$F_{6,66} = 62.2$	$2.93 \times 10^{-25}_{***}$	.850	$F_{6,66} = 50.9$	$6.91 \times 10^{-23}_{***}$	.822
	$A \times W$	$F_{9,99} = 2.97$	$3.59 \times 10^{-3}_{**}$	.213	$F_{9,99} = 4.48$	$6.08 \times 10^{-5}_{***}$	.289	$F_{9,99} = 14.8$	$6.05 \times 10^{-15}_{***}$	.574
	Bias $\times A \times W$	$F_{18,198} = 1.16$	3.00×10^{-1}	.0953	$F_{18,198} = 6.44$	$2.01 \times 10^{-12}_{***}$	.369	$F_{18,198} = 1.46$	1.07×10^{-1}	.117
TP_{we}	Bias	$F_{2,22} = 88.2$	$3.10 \times 10^{-11}_{***}$	.889	$F_{2,22} = 69.7$	$3.03 \times 10^{-10}_{***}$	.864	$F_{2,22} = 33.3$	$2.21 \times 10^{-7}_{***}$	.752
	A	$F_{3,33} = 92.4$	$3.91 \times 10^{-16}_{***}$	.894	$F_{3,33} = 4.56$	$8.88 \times 10^{-3}_{**}$	.293	$F_{3,33} = 89.7$	$6.05 \times 10^{-16}_{***}$	.891
	W	$F_{3,33} = 53.8$	$8.43 \times 10^{-13}_{***}$	.830	$F_{3,33} = 129$	$2.71 \times 10^{-18}_{***}$	.921	$F_{3,33} = 15.0$	$2.46 \times 10^{-6}_{***}$	.577
	Bias $\times A$	$F_{6,66} = 16.1$	$2.59 \times 10^{-11}_{***}$	.594	$F_{6,66} = 4.12$	$1.42 \times 10^{-3}_{**}$	.273	$F_{6,66} = 5.24$	$1.84 \times 10^{-4}_{***}$	.323
	Bias $\times W$	$F_{6,66} = 1.34$	2.54×10^{-1}	.108	$F_{6,66} = 3.48$	$4.78 \times 10^{-3}_{**}$	.240	$F_{6,66} = 17.2$	$7.58 \times 10^{-12}_{***}$	.610
	$A \times W$	$F_{9,99} = 24.6$	$1.32 \times 10^{-21}_{***}$	.691	$F_{9,99} = 2.50$	$1.26 \times 10^{-2}_{*}$	.185	$F_{9,99} = 9.18$	$5.33 \times 10^{-10}_{***}$	.455
	Bias $\times A \times W$	$F_{18,198} = 1.50$	9.37×10^{-2}	.120	$F_{18,198} = 2.37$	$2.00 \times 10^{-3}_{**}$	.177	$F_{18,198} = .751$	7.54×10^{-1}	.0639
TP_e	Bias	$F_{2,22} = 88.8$	$2.90 \times 10^{-11}_{***}$	.890	$F_{2,22} = 76.3$	$1.27 \times 10^{-10}_{***}$	.874	$F_{2,22} = 30.9$	$4.07 \times 10^{-7}_{***}$	.738
	A	$F_{3,33} = 92.2$	$4.01 \times 10^{-16}_{***}$	.893	$F_{3,33} = 2.47$	7.91×10^{-2}	.183	$F_{3,33} = 89.4$	$6.38 \times 10^{-16}_{***}$	.890
	W	$F_{3,33} = 52.8$	$1.10 \times 10^{-12}_{***}$	.827	$F_{3,33} = 115$	$1.48 \times 10^{-17}_{***}$	.913	$F_{3,33} = 8.99$	$1.71 \times 10^{-4}_{***}$	.450
	Bias $\times A$	$F_{6,66} = 17.0$	$8.80 \times 10^{-12}_{***}$	.608	$F_{6,66} = 2.68$	$2.19 \times 10^{-2}_{*}$	.196	$F_{6,66} = 4.99$	$2.89 \times 10^{-4}_{***}$	.312
	Bias $\times W$	$F_{6,66} = 1.21$	3.14×10^{-1}	.0989	$F_{6,66} = 7.79$	$2.36 \times 10^{-6}_{***}$	.415	$F_{6,66} = 16.6$	$1.36 \times 10^{-11}_{***}$	.602
	$A \times W$	$F_{9,99} = 25.1$	$6.84 \times 10^{-22}_{***}$	.695	$F_{9,99} = 1.77$	8.26×10^{-2}	.139	$F_{9,99} = 7.76$	$1.43 \times 10^{-8}_{***}$	.414
	Bias $\times A \times W$	$F_{18,198} = 1.47$	1.05×10^{-1}	.118	$F_{18,198} = 1.63$	5.59×10^{-2}	.129	$F_{18,198} = .725$	7.83×10^{-1}	.0618

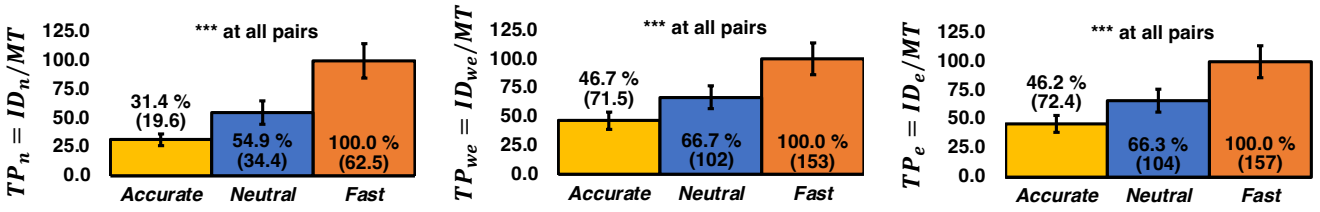


Figure 12: Bias vs *TP* for Linear. Throughout Figures 12–14, the percentages in the graphs represent the ratios when the highest *TP* among the three Bias conditions is 100%. The stability of *TP* across Bias was improved by W_e, but A_e had little effect on it.

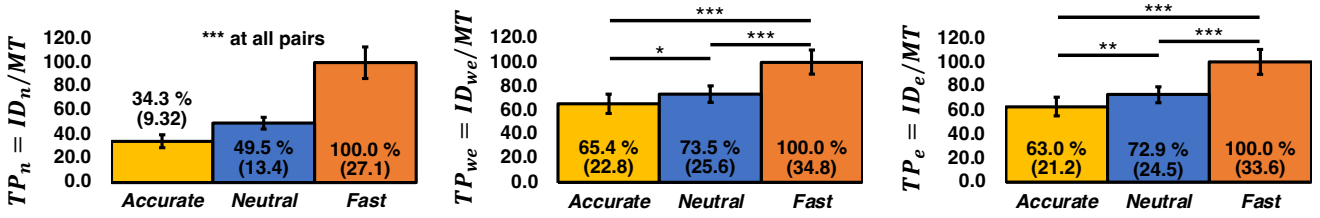


Figure 13: Bias vs *TP* for Circular. The stability of *TP* across Bias was improved by W_e, but A_e had little effect on it.

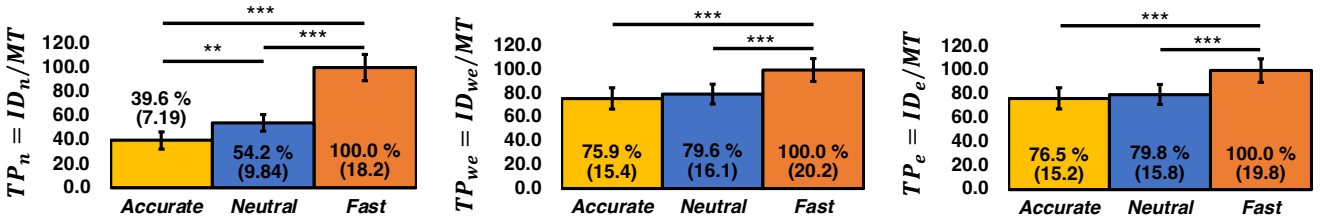


Figure 14: Bias vs *TP* for Sine-wave. The stability of *TP* across Bias was improved by W_e, but A_e had little effect on it.

Table 9: The p-values for the permutation test. Green cells and ‡ indicate $p \geqq .05$.

	Linear			Circular			Sine-wave		
Compared Bias	TP_n	TP_{we}	TP_e	TP_n	TP_{we}	TP_e	TP_n	TP_{we}	TP_e
Accurate Neutral	5.3×10^{-4}	1.0×10^{-3}	9.2×10^{-4}	4.9×10^{-4}	2.5×10^{-3}	9.5×10^{-4}	9.6×10^{-4}	2.9×10^{-1} ‡	3.5×10^{-1} ‡
Neutral Fast	5.0×10^{-4}	5.0×10^{-4}	4.7×10^{-4}	4.6×10^{-4}	4.5×10^{-4}	5.3×10^{-4}	4.7×10^{-4}	5.3×10^{-4}	5.0×10^{-4}
Accurate Fast	5.1×10^{-4}	5.1×10^{-4}	4.8×10^{-4}	4.9×10^{-4}	5.1×10^{-4}	5.0×10^{-4}	4.9×10^{-4}	4.9×10^{-4}	5.1×10^{-4}

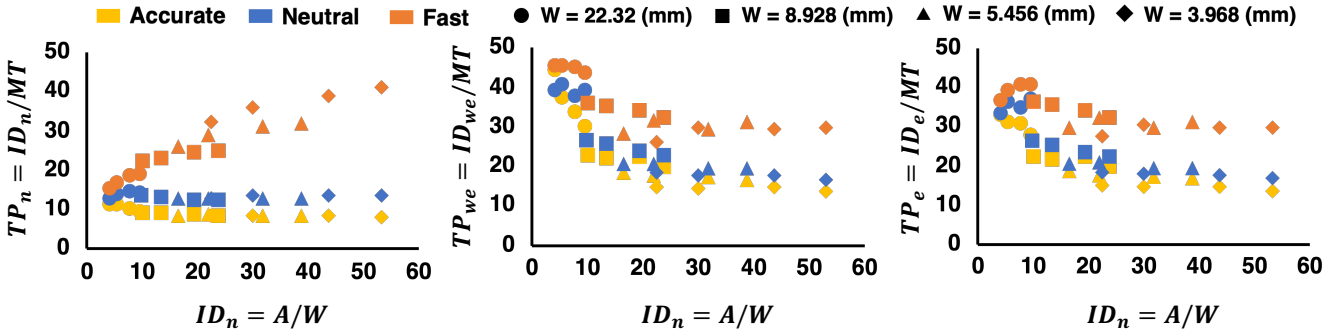


Figure 15: ID_n vs TP for Linear. The stability of TP across ID_n was decreased by W_e for all individual Biases, but A_e had little effect on it.

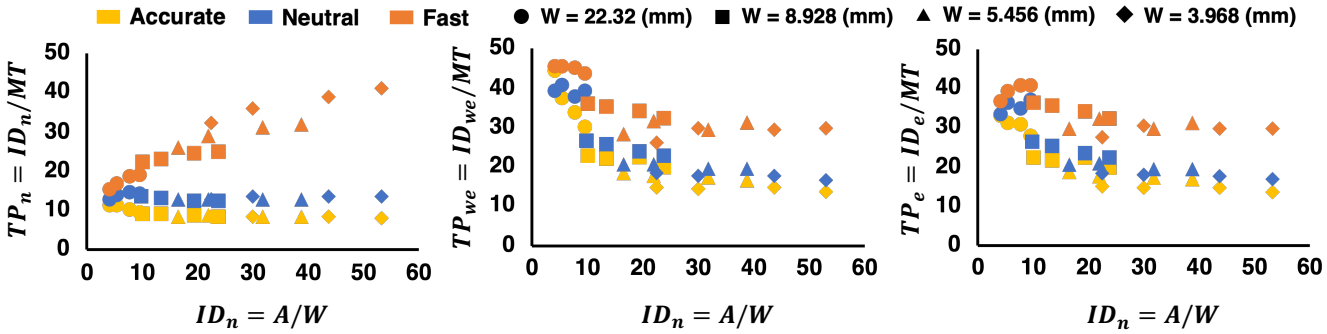


Figure 16: ID_n vs TP for Circular. The stability of TP across ID_n was improved by W_e for Fast, but decreased for Accurate and Neutral. A_e improved the stability for all individual Biases. When W is big, A_e can factor in the change due to strokes near the inner diameter of the circle.

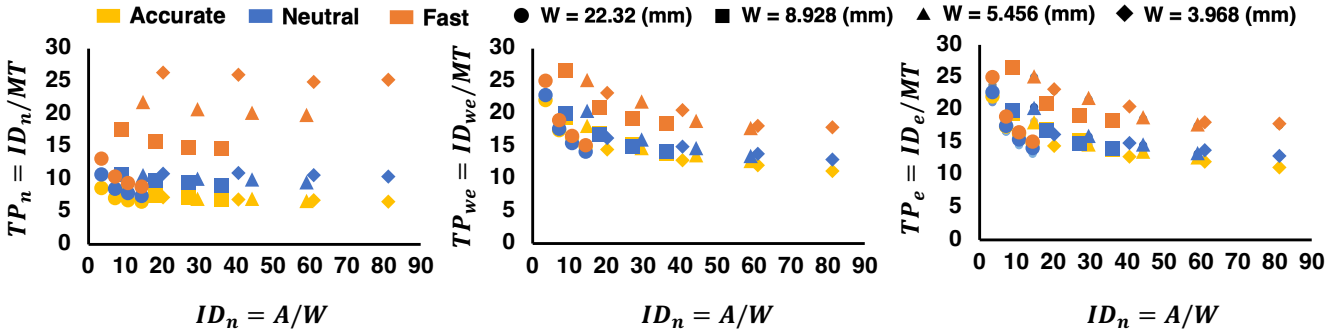


Figure 17: ID_n vs TP for Sine-wave. The stability of TP across ID_n was improved by W_e for Fast, but decreased for Accurate and Neutral. A_e had little effect on it.

Table 10: The stability of *TP*s across task difficulties. *SD* is the standard deviation, and *CV* the coefficient of variation calculated by dividing the *SD* by the *Mean*.

		Linear			Circular			Sine-wave		
TP	Bias	*SD*	*Mean*	*CV*[%]	*SD*	*Mean*	*CV*[%]	*SD*	*Mean*	*CV*[%]
TP_n	Accurate	1.29	19.6	6.59%	0.946	9.32	10.1%	0.537	7.19	7.47%
	Neutral	3.31	34.4	9.63%	0.613	13.4	4.56%	1.05	9.84	10.6%
	Fast	16.8	62.5	26.8%	7.54	27.1	27.8%	5.75	18.2	31.7%
TP_{we}	Accurate	44.4	71.5	62.1%	8.92	22.8	39.2%	2.80	15.4	18.2%
	Neutral	47.9	102	46.9%	8.47	25.6	33.1%	2.69	16.1	16.7%
	Fast	58.1	153	37.9%	6.44	34.8	18.5%	3.18	20.2	15.8%
TP_e	Accurate	45.6	72.4	63.0%	6.22	21.2	29.4%	2.62	15.2	17.3%
	Neutral	49.9	104	48.0%	6.88	24.5	28.1%	2.58	15.8	16.3%
	Fast	61.3	157	39.1%	4.11	33.6	12.2%	3.23	19.8	16.4%

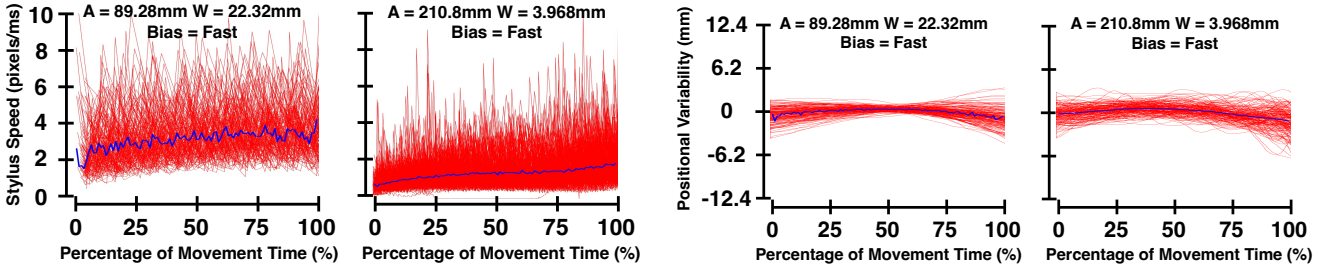


Figure 18: Stylus speed and positional variability against the percentage of task progress for Linear. The blue line indicates the averaged stylus speed and positional variability. For the low ID_n condition, the variability was small despite the high speed, which resulted in too large *TP*s.

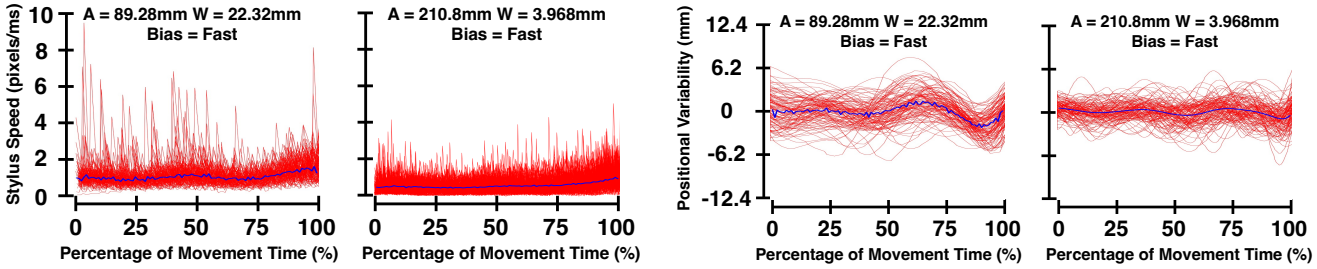


Figure 19: Stylus speed and positional variability against the percentage of task progress for Circular. The blue line indicates the averaged stylus speed and positional variability. The speed was not too fast, and variability increased as *W* increased. The increase of variability in certain parts of the task is likely due to the occlusion problem in Circular tasks.

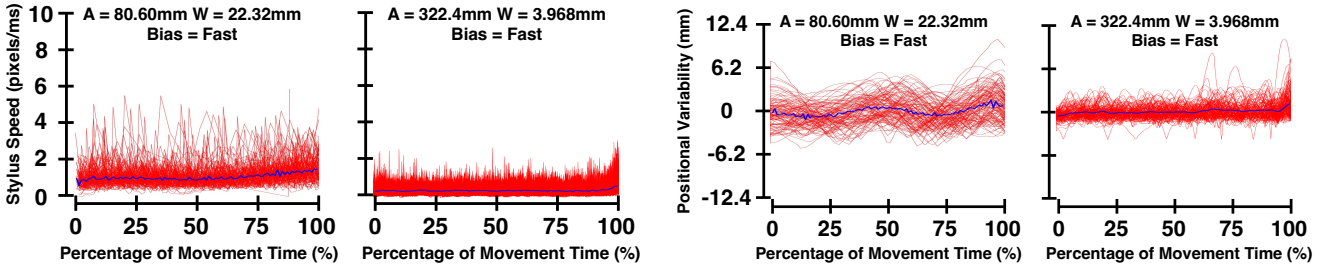


Figure 20: Stylus speed and positional variability against the percentage of task progress for Sine-wave. The blue line indicates the averaged stylus speed and positional variability. The speed was stable, and variability increased as *W* increased. The variability changes during task were also more stable than for Circular.

Linear. These are possible reasons for the observed stable *TP*, compared to Linear. However, positional variability was higher and unstable (Figure 19). More specifically, this variation in the variability occurred even within a single stroke, and the timing at which the variability increased was similar in each trial. This may be due to the occlusion problem in Circular steering. The tip of the stylus is sometimes hidden by the hand during this type of task, depending on the circle diameter, and this may have reduced the stability of the *TP* across ID_n. The ergonomics factor may also have affected this result; the ease of moving the stylus would vary depending on its position within the circle.

For Sine-wave, the stylus speed was not very high, even in the low ID_n condition, and reasonably stable across trials (Figure 20). Stylus speed was more stable than for Linear, and positional variability was more stable relative to Circular. Positional variability was widely spread in low ID_n conditions, but smaller in high ID_n conditions. Similar to the *TP* of *crossing with amplitude constraint* task in a previous study [18], the stylus speed was stable and the positional variability was widely spread, even for easy tasks. These reasons are possible causes of the stable *TP* for Sine-wave.

Although direct comparisons between task shapes cannot be made due to different path lengths, Sine-wave had the highest model fit for Mixed (Table 7), and exhibited stable *TP* across different Biases and ID_ns (Table 10). For Linear, the positional variability was small even when the stylus speed was very high (Figure 18). Therefore, in low ID_n conditions, the task could be completed successfully with a high speed movement, suggesting that *TP* was less stable across task difficulties (Figure 15). In Circular, stylus speed was stable, but in each trial, the positional variability was large in some parts of the circle, likely due the presence of the occlusion problem (Figure 19). With sine-wave steering, the stylus speed and positional variability were stable (Figure 20), and there was no occlusion problem, suggesting stable *TP*. These results suggest that we can recommend the sine-wave steering task for device comparison experiments with trajectory-based tasks, as the task does not suffer from high speed movements nor from occlusion issues.

6 IMPLICATIONS

Our work offers the following implications for future researchers investigating path-steering performance.

- When modeling steering tasks with untested path shapes, we recommend using the ID_n model for estimating *MT* for one device, one group of participants, and a specific speed-accuracy bias.
- In terms of better stability of *TP* across different speed-accuracy biases, we recommend using TP_{we} or TP_e when comparing performances between multiple devices [2], multiple groups of participants [30], or multiple experimental conditions [36].
- For steering performance evaluations, we generally recommend the sine-wave shape due to the stability of *TP*.

7 LIMITATIONS AND FUTURE WORK

The path shape in the real-world steering tasks is not limited to linear, circular, and sine-wave ones. Therefore, additional validation with other shapes, such as corners [28], narrowing and widening tunnels [39], and sequential linear path segments [41], is needed for generalization. Although it is known that task curvature affects movement time [40], the effect of curvature was not investigated in this study. Therefore, additional verification of the effect of curvature is needed. In this study, the *amplitude* of the sine-wave (not the path length *A*) was fixed, but this *amplitude* is likely to affect the movement time. In addition, the effect of the number of waves was not considered. Therefore, additional research on sine-wave steering is needed.

8 CONCLUSION

The usefulness of throughput as a performance metric for steering was demonstrated by three comprehensive tasks, exploring three path shapes, each with three speed-accuracy biases. We also experimentally demonstrated the effects of effective width and effective amplitude measures to better characterize user performance with various path shapes and multiple speed-accuracy biases. Our results provide theoretical support for the use of effective parameters for steering. When effective width was used, the effect of speed-accuracy bias on throughput was smaller than when nominal width was used. When effective amplitude was used, the stability of throughput across different task difficulties was improved in circular steering task. Therefore, we suggest that TP_{we} or TP_e may be better performance metrics than TP_n in terms of smoothing across different speed-accuracy biases. Furthermore, we investigated sine-wave steering task, with our results suggesting that this shape might be most useful for device evaluation experiments. Our results lead to a better understanding of user performance in trajectory-based tasks, and the approach presented here provides an evidence-based performance evaluation method for future researchers.

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. 1999. Performance Evaluation of Input Devices in Trajectory-Based Tasks: An Application of the Steering Law. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Pittsburgh, Pennsylvania, USA) *(CHI '99)*. Association for Computing Machinery, New York, NY, USA, 466–472. https://doi.org/10.1145/302979.303133

[3] 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

[4] 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

[5] 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

[6] 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.

[7] A. Cockburn, D. Ahlström, and C. Gutwin. 2012. Understanding performance in touch selections: Tap, drag and radial pointing drag with finger, stylus and mouse. *International Journal of Human-Computer Studies* 70, 3 (2012), 218–233. https://doi.org/10.1016/j.ijhcs.2011.11.002

[8] Edward RFW Crossman. 1956. *The measurement of perceptual load in manual operations.* Ph. D. Dissertation. University of Birmingham.

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

[10] Sarah A. Douglas, Arthur E. Kirkpatrick, and I. Scott MacKenzie. 1999. Testing Pointing Device Performance and User Assessment with the ISO 9241, Part 9 Standard. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Pittsburgh, Pennsylvania, USA) *(CHI '99)*. Association for Computing Machinery, New York, NY, USA, 215–222. https://doi.org/10.1145/302979.303042

[11] Heiko Drewes. 2023. The Fitts' Law Filter Bubble. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 421, 5 pages. https://doi.org/10.1145/3544549.3582739

[12] Colin G. Drury, M. Ali Montazer, and Mark H. Karwan. 1987. Self-Paced Path Control as an Optimization Task. *IEEE Transactions on Systems, Man, and Cybernetics* 17, 3 (1987), 455–464. https://doi.org/10.1109/TSMC.1987.4309061

[13] 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.

[14] Julien Gori, Olivier Rioul, and Yves Guiard. 2018. Speed-Accuracy Tradeoff: A Formal Information-Theoretic Transmission Scheme (FITTS). *ACM Trans. Comput.-Hum. Interact.* 25, 5, Article 27 (sep 2018), 33 pages. https://doi.org/10.1145/3231595

[15] Errol R. Hoffmann. 2009. Review of models for restricted-path movements. *International Journal of Industrial Ergonomics* 39, 4 (2009), 578–589. https://doi.org/10.1016/j.ergon.2008.02.007

[16] Errol R. Hoffmann. 2013. Which Version/Variation of Fitts' Law? A Critique of Information-Theory Models. *Journal of Motor Behavior* 45, 3 (2013), 205–215. https://doi.org/10.1080/00222895.2013.778815 arXiv:https://doi.org/10.1080/00222895.2013.778815 PMID: 23581725.

[17] ISO. 2012. ISO 9241-411. International standard: ergonomics of human-system interaction — Part 411: Evaluation methods for the design of physical input devices, International Organization for Standardization.

[18] 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* (Hamburg, Germany) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 282, 9 pages. https://doi.org/10.1145/3544549.3585817

[19] Sergey Kulikov, I. Scott MacKenzie, and Wolfgang Stuerzlinger. 2005. Measuring the Effective Parameters of Steering Motions. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems* (Portland, OR, USA) *(CHI EA '05)*. Association for Computing Machinery, New York, NY, USA, 1569–1572. https://doi.org/10.1145/1056808.1056968

[20] Edward Lank and Eric Saund. 2005. Sloppy Selection: Providing an Accurate Interpretation of Imprecise Selection Gestures. *Comput. Graph.* 29, 4 (aug 2005), 490–500. https://doi.org/10.1016/j.cag.2005.05.003

[21] I. Scott MacKenzie. 1989. A Note on the Information-Theoretic Basis for Fitts' Law. *Journal of Motor Behavior* 21, 3 (1989), 323–330. https://doi.org/10.1080/00222895.1989.10735486 arXiv:https://doi.org/10.1080/00222895.1989.10735486 PMID: 15136269.

[22] I Scott MacKenzie. 1992. Fitts' law as a research and design tool in human-computer interaction. *Human-computer interaction* 7, 1 (1992), 91–139.

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

[24] I. Scott MacKenzie and Poika Isokoski. 2008. Fitts' Throughput and the Speed-Accuracy Tradeoff. In *Proceedings of the SIHI 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

[25] 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.

[26] Tomer Moscovich. 2009. Contact Area Interaction with Sliding Widgets. In *Proceedings of the 22nd Annual ACM Symposium on User Interface Software and Technology* (Victoria, BC, Canada) *(UIST '09)*. Association for Computing Machinery, New York, NY, USA, 13–22. https://doi.org/10.1145/1622176.1622181

[27] 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.

[28] Robert Pastel. 2006. Measuring the Difficulty of Steering through Corners. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Montréal, Québec, Canada) *(CHI '06)*. Association for Computing Machinery, New York, NY, USA, 1087–1096. https://doi.org/10.1145/1124772.1124934

[29] Charles Perin, Pierre Dragicevic, and Jean-Daniel Fekete. 2015. Crossets: Manipulating Multiple Sliders by Crossing. In *Proceedings of the 41st Graphics Interface Conference* (Halifax, Nova Scotia, Canada) *(GI '15)*. Canadian Information Processing Society, CAN, 233–240.

[30] Xiangshi Ren and Xiaolei Zhou. 2011. An investigation of the usability of the stylus pen for various age groups on personal digital assistants. *Behaviour & Information Technology* 30, 6 (2011), 709–726. https://doi.org/10.1080/01449290903205437

[31] Ather Sharif, Victoria Pao, Katharina Reinecke, and Jacob O. Wobbrock. 2020. The Reliability of Fitts's Law as a Movement Model for People with and without Limited Fine Motor Function. In *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility* (Virtual Event, Greece) *(ASSETS '20)*. Association for Computing Machinery, New York, NY, USA, Article 16, 15 pages. https://doi.org/10.1145/3373625.3416999

[32] 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

[33] Sven Strothoff, Wolfgang Stuerzlinger, and Klaus Hinrichs. 2015. Pins 'n' Touches: An Interface for Tagging and Editing Complex Groups. In *Proceedings of the 2015 International Conference on Interactive Tabletops & Surfaces (ITS)*. ACM, New York, NY, USA, 191–200. https://doi.org/10.1145/2817721.2817731

[34] Ahmed N. Sulaiman and Patrick Olivier. 2008. Attribute Gates. In *Proceedings of the 21st Annual ACM Symposium on User Interface Software and Technology* (Monterey, CA, USA) *(UIST '08)*. ACM, New York, NY, USA, 57–66. https://doi.org/10.1145/1449715.1449726

[35] Minghui Sun, Xiangshi Ren, and Xiang Cao. 2010. Effects of Multimodal Error Feedback on Human Performance in Steering Tasks. *Journal of Information Processing* 18 (2010), 284–292. https://doi.org/10.2197/ipsjjip.18.284

[36] Jannik Wiese and Niels Henze. 2023. Predicting Mouse Positions Beyond a System's Latency Can Increase Throughput and User Experience in Linear Steering Tasks. In *Proceedings of Mensch Und Computer 2023* (Rapperswil, Switzerland) *(MuC '23)*. Association for Computing Machinery, New York, NY, USA, 101–115. https://doi.org/10.1145/3603555.3603556

[37] Shota Yamanaka. 2022. Test-Retest Reliability on Movement Times and Error Rates in Target Pointing. In *Proceedings of the 2022 ACM Designing Interactive Systems Conference* (Virtual Event, Australia) *(DIS '22)*. Association for Computing Machinery, New York, NY, USA, 178–188. https://doi.org/10.1145/3532106.3533450

[38] Shota Yamanaka, Taiki Kinoshita, Yosuke Oba, Ryuto Tomihari, and Homei Miyashita. 2023. Varying Subjective Speed-Accuracy Biases to Evaluate the Generalizability of Experimental Conclusions on Pointing-Facilitation Techniques. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 317, 13 pages. https://doi.org/10.1145/3544548.3580740

[39] Shota Yamanaka and Homei Miyashita. 2016. Modeling the Steering Time Difference between Narrowing and Widening Tunnels. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 1846–1856. https://doi.org/10.1145/2858036.2858037

[40] Shota Yamanaka and Homei Miyashita. 2019. Modeling Pen Steering Performance in a Single Constant-Width Curved Path. In *Proceedings of the 2019 ACM International Conference on Interactive Surfaces and Spaces* (Daejeon, Republic of Korea) *(ISS '19)*. Association for Computing Machinery, New York, NY, USA, 65–76. https://doi.org/10.1145/3343055.3359697

[41] Shota Yamanaka, Wolfgang Stuerzlinger, and Homei Miyashita. 2017. Steering Through Sequential Linear Path Segments. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 232–243. https://doi.org/10.1145/3025453.3025836

[42] 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

[43] Shota Yamanaka, Hiroki Usuba, Haruki Takahashi, and Homei Miyashita. 2020. Servo-Gaussian Model to Predict Success Rates in Manual Tracking: Path Steering and Pursuit of 1D Moving Target. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '20)*. Association for Computing Machinery, New York, NY, USA, 844–857. https://doi.org/10.1145/3379337.3415896

[44] 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.

[45] Shumin Zhai and Rogier Woltjer. 2003. Human Movement Performance in Relation to Path Constraint - The Law of Steering in Locomotion. In *Proceedings of the IEEE Virtual Reality 2003 (VR '03)*. IEEE Computer Society, USA, 149.

[46] Xiaolei Zhou, Xiang Cao, and Xiangshi Ren. 2009. Speed-Accuracy Tradeoff in Trajectory-Based Tasks with Temporal Constraint. In *Human-Computer Interaction – INTERACT 2009*, Tom Gross, Jan Gulliksen, Paula Kotzé, Lars Oestreicher, Philippe Palanque, Raquel Oliveira Prates, and Marco Winckler (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 906–919.

[47] Xiaolei Zhou and Xiangshi Ren. 2010. An investigation of subjective operational biases in steering tasks evaluation. *Behaviour & Information Technology* 29, 2 (2010), 125–135.