Taylor & Francis
Taylor & Francis Group

International Journal of Human–Computer Interaction

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hihc20

Advanced Investigation of Steering Performance with Error-Accepting Delays

Takuma Hidaka, Yusuke Sei, Naoto Nishida, Shota Yamanaka & Buntarou Shizuki

To cite this article: Takuma Hidaka, Yusuke Sei, Naoto Nishida, Shota Yamanaka & Buntarou Shizuki (2023): Advanced Investigation of Steering Performance with Error-Accepting Delays, International Journal of Human–Computer Interaction, DOI: 10.1080/10447318.2023.2192586

To link to this article: https://doi.org/10.1080/10447318.2023.2192586

Published online: 03 Apr 2023.

INTERNATIONAL JOURNAL OF HUMAN–COMPUTER INTERACTION
https://doi.org/10.1080/10447318.2023.2192586

Advanced Investigation of Steering Performance with Error-Accepting Delays

Takuma Hidaka[a], Yusuke Sei[a], Naoto Nishida[b*], Shota Yamanaka[c], and Buntarou Shizuki[a]

[a]Department of Computer Science, University of Tsukuba, Tsukuba, Japan; [b]Graduate School of Interdisciplinary Information, The University of Tokyo, Tokyo, Japan; [c]Yahoo! JAPAN Research, Yahoo Japan Corporation, Tokyo, Japan

ABSTRACT

The Steering Law is a robust model to predict the movement time (*MT*) for steering through a constrained path, and the most representative example in human-computer interaction (HCI) is navigating cascaded menus. In typical implementations of cascaded menus, however, users can deviate from the path for a short time; we call this error-accepting delay, or T_{delay}. Yamanaka modified the Steering Law to predict *MT* under several T_{delay} conditions, and our goal is to investigate the reproducibility of his model with more various T_{delay} values. In addition, HCI researchers have recently formed a consensus that the goodness of models should be judged by the prediction accuracy for future (untested) task conditions. Thus, for the sake of completeness, we conducted two analyses: a shuffle-split cross-validation and leave-one-T_{delay}-out cross-validation. The results showed that, regardless of the all-data and cross-validation analyses, Yamanaka's modified model outperformed the baseline Steering Law, which strengthened his original experimental report.

KEYWORDS

Graphical user interfaces (GUI); user performance modeling; Steering Law; replication study

1. Introduction

Deriving new models to predict human performance is one of the core topics in human-computer interaction (HCI) research. In addition, the importance of replication studies has been repeatedly claimed in HCI (Banovic, 2016; Cockburn et al., 2020; Hornbæk et al., 2014; Wilson et al., 2014). In this context, re-evaluating the validity of an existing model (i.e., reproducibility) should deepen the understanding of human motor behaviors and thus contribute to various scientific fields.

In this article, the model we specifically investigate is Steering Law (Accot & Zhai, 1997; Drury, 1971; Rashevsky, 1959), which predicts the movement time (*MT*) to navigate a constrained path (see Figure 1a). In HCI, navigation in cascaded menus has been considered a typical application of Steering Law (Accot & Zhai, 1997; Ahlström, 2005; Dennerlein et al., 2000).

Contrarily, in many cascaded-menu implementations, a submenu appears *shortly after* the cursor hovers over a parent menu item. For example, as shown in Figure 1b, the submenu related to "Computers" is kept open even while the cursor accidentally enters the item labeled "Movies" momentarily. In this case, although the user successfully accomplishes this path-navigation task, it is unclear whether Steering Law is applicable in its as-is formulation, because Steering Law theoretically holds only when the cursor passes through the path without deviating (Accot & Zhai, 1997; Drury et al., 1987; Montazer et al., 1988).

Such a duration from when the cursor enters a parent menu item to when its submenu opens is typically configured by means of the `setTimeout` (JavaScript), `delay` (jQuery), or other similar programming functions. We call this delay "error-accepting delay" as a general term and use "T_{delay}" to describe it as an independent variable in the experiment.

Note that, as the error-accepting delay increases, users have to pay less attention to the path boundaries, and thus the cursor-movement speed increases, resulting in a shorter *MT*. If one can predict the *MT* for a given configurations of path length (A), width (W), and T_{delay}, designers and engineers can develop appropriate cascaded menus so that, for example, users can pass through each item within 1 s.

Recently, Yamanaka (one of the authors of this article) measured users' performance on steering tasks in which users are allowed to deviate the cursor from a path for a short duration (Yamanaka, 2019). He evaluated the effects of T_{delay} values of 0, 100, 200, 400, 600, and 800 ms on *MT*, and then he derived a modified formulation of Steering Law to predict *MT* accurately even when T_{delay} changed. However, his model's validity was examined only once, and the experimental design had limited T_{delay} values. Therefore, as described above, it is worth testing its reproducibility with a different participant group and more varied T_{delay} values.

More specifically, we should investigate denser and larger values of T_{delay} than those in the original study by

CONTACT Shota Yamanaka syamanak@yahoo-corp.jp Yahoo! JAPAN Research, Yahoo Japan Corporation, Kioi Tower, 102-8282, Tokyo Garden Terrace Kioicho, 1-3, Kioi-cho, Chiyoda-ku, Tokyo, Japan
*Present Affiliation: ValueCommerce Co., Ltd.

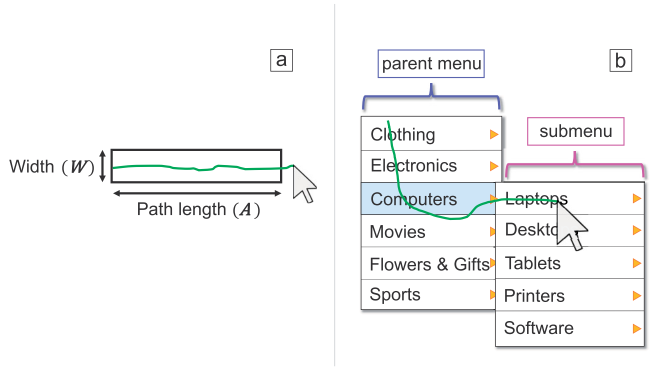

Figure 1. (a) Steering Law predicts the *MT* to pass through a constrained path with length *A* and width *W*. (b) An example in which users can deviate from the path boundaries for a short time.

Yamanaka (2019). The reasons behind this claim are as follows.

- Keele and Posner (1968) reported that the human reaction time is 190–260 ms on average, and because the delay in our experimental system was 57.9 ms, the ability of participants to react to the cursor's deviation from the path boundaries changes around 300 ms, which was not tested in the original study.
- According to Lin and Hsu (2014), the human reaction time in their experiment was 273 ms, but this was the average value for all participants. They reported that individual participant's reaction times ranged from 87 and 441 ms. As the reaction times of the participants in our experiment may have been biased towards shorter or longer, the slowest one may take approximately 500 ms to react to the path deviation (reaction time of 441 ms and system delay of 57.9 ms; 498.9 ms in total), so this value was newly examined in our experiment.
- The longer the T_{delay}, the more participants should be able to complete the task without causing errors of straying outside the path boundaries for longer than T_{delay}, even if they deviated from the path. Hence, the task requirement would turn into "Just clicking in the start and end areas," which is no longer modeled by Steering Law. However, if the participants still pay attention to pass boundaries under a long T_{delay} condition, Yamanaka's modified model of Steering Law may be suitable. Therefore, to better validate the modified model, it is better to experiment with long T_{delay} values. We thus set the maximum value of T_{delay} to 1000 ms, which is 25% larger than the maximum value in the original study, thus strengthening the validation of the modified model.

Accordingly, this article is an extended version of the conference paper (Yamanaka, 2019) with an entirely new user study. The purposes of the study are twofold.

- First is to confirm the reproducibility of the superior performance of Yamanaka's model to the baseline (i.e., the original Steering Law) to predict *MT* accurately. To do so, we collected another group of participants (none of them joined the original experiment) and used more T_{delay} conditions.
- Second is to examine the prediction accuracy for future (untested) task conditions for the sake of completeness to evaluate the model performance. In the original study, to evaluate the prediction accuracy of the candidate models, Yamanaka computed the values of adjusted R^2 and Akaike Information Criterion (AIC) (Akaike, 1974) by using all 48 fitting points, i.e., $2A \times 4W \times 6T_{delay}$ conditions (Yamanaka, 2019). However, recently, HCI researchers have become aware that a model's capability should also be discussed on the basis of the prediction accuracy for untested data, and thus cross-validations have been conducted, e.g., for target-pointing (Ko et al., 2020) and path-steering studies (Yamanaka et al., 2020). From this standpoint, the analysis in the original study (Yamanaka, 2019) went only halfway: Yamanaka's modified model fitted well when all the obtained data points were used, while the robustness of the model for new path conditions remained unclear.

We tackle the second issue by running a shuffle-split cross-validation and leave-one-T_{delay}-out cross-validation. This approach is also applied to the dataset of the original study (Yamanaka, 2019) for reanalysis. These results show that Yamanaka's model again outperforms the baseline for all-data and cross-validation analyses, which emphasizes a stronger robustness of his modified model than claimed in the original study.

When extending the original paper (Yamanaka, 2019), we reconstructed the Introduction and Related Work sections. The subsequent sections from User Study onward are newly added, which accounts for $\sim$62% (approximately 3,680 words out of 5,970) in this article.

2. Related work

2.1. Steering Law

To understand how Yamanaka (2019) modified Steering Law to take error-accepting delay into account, we first describe the derivation of Steering Law. To steer through a constant-width straight path whose length is A and width is W as shown in Figure 1a, Accot and Zhai (1997) proposed the following model, the baseline Steering Law:

$$MT = a + b\frac{A}{W}, \qquad (1)$$

where MT is the movement time, and a and b are empirically determined constants. The A/W ratio is called the index of difficulty (ID) of Steering Law.

Equation (1) indicates that if W is large, MT becomes short because the user can move the cursor roughly through the path without deviating from it. In previous studies, researchers showed that Equation (1) is applicable under the following conditions: multiple kinds of input devices, such as a mouse or touchpad (Accot & Zhai, 1999; Senanayake & Goonetilleke, 2016), small- and large-scale displays (Accot & Zhai, 2001), dominant and nondominant hands (Hoffmann, 1997), various steering directions (Thibbotuwawa et al., 2012; Zhou et al., 2008), and different priorities on speed or accuracy (Zhou & Ren, 2010).

Accot and Zhai proposed another form, which showed that the average speed V throughout the path (i.e., $V = A/MT$) is proportionally related to the path width:

$$V = bW. \tag{2}$$

Other studies also support this simple relationship between V and W (Drury, 1971; Montazer & Drury, 1989; Rashevsky, 1959).

We have $W = V/b'$ from Equation (2) where b' is a new coefficient for distinction. This can be substituted into Equation (1), and we obtain

$$MT = a + b\frac{A}{W} = a + b\frac{A}{V/b'} = a + b''\frac{A}{V} \quad (\text{let } \ b'' = b \cdot b'). \tag{3}$$

Hoffmann also validated this model (Hoffmann, 2009).

In summary, for a constant-width straight path, although there is an option to predict MT or V on the left side of an equation, there are clear consensuses that the movement speed V proportionally increases with the path width W and that the time MT required to navigate the path decreases inversely.

2.2. Modifications of Steering Law

Numerous studies have evaluated path shapes other than horizontal straight paths of constant width. Examples include circular paths (Accot & Zhai, 1999, 2001; Hoffmann, 2009), curved paths (Montazer et al., 1987; Nancel & Lank, 2017; Yamanaka & Miyashita, 2019), linearly narrowing paths (Accot & Zhai, 1997; Yamanaka & Miyashita, 2016), widening spiral paths (Accot & Zhai, 1997), paths with a corner (Pastel, 2006), successive path segments (Yamanaka et al., 2017, 2018), and diagonal direction paths (Thibbotuwawa et al., 2012). Typically, researchers modified Steering Law to predict MT under these specific conditions and demonstrated that the novel model outperformed the baseline Steering Law in terms of the prediction accuracy (i.e., model fit, such as R^2). In contrast, in our study, because we consider an application of Steering Law to menu operations to help designers and engineers set appropriate error-accepting delays, we investigate only a straight path of constant width.

Kulikov et al. proposed a model that included trials in which the cursor deviated from the path (Kulikov et al., 2005). They corrected the path width by an amount of *effective width* in accordance with the cursor-coordinate variability perpendicular to the direction of movement, regardless of whether the cursor was in/outside the path tolerance. Hence, their task was a kind of path steering with error-accepting conditions. However, they instructed the participants to "pass through as quickly as possible without deviating from the path." Therefore, the participants had to move the cursor so that it would not deviate from the path, and thus this result is not suitable for designers who want to know the average operating time under a certain acceptance for deviation from a given path. In contrast, Yamanaka (2019) instructed the participants that they could purposefully deviate from the path for a given error-accepting delay, which should affect their cursor-operation speed.

2.3. Modified Steering Law model with error-accepting delay

In Yamanaka's experimental results, Steering Law in its speed-prediction form (Equation (2), $V = bW$) showed adjusted $R^2 = 0.882$ (Yamanaka, 2019). Because he found statistical main effects of W and T_{delay} on V, he examined the following form to model V:[1]

$$V = bW + cT_{delay}, \tag{4}$$

where b and c are regression constants. This model showed adjusted $R^2 = 0.986$, and W and T_{delay} were significant contributors ($p < 0.001$). These results indicate that the inclusion of T_{delay} significantly improved the prediction accuracy for V.

By substituting Equation (4) for V in Equation (3), and then merging some constants, Yamanaka proposed the following model (Yamanaka, 2019):

$$MT = a + b\frac{A}{W + cT_{delay}}. \tag{5}$$

Because Yamanaka's model uses three free parameters while the baseline Steering Law uses two (Equation (1)), he compared the prediction accuracies by AIC, which penalizes using more free parameters (Akaike, 1974). Yamanaka's model gave $AIC = 528$ and the baseline Steering Law gave 647. A model with smaller AIC is better, and a difference greater than 10 is statistically significant (Burnham & Anderson, 2003). Thus, Yamanaka concluded that his new model significantly improved the prediction accuracy of MT. We will first follow this procedure when analyzing our data.

2.4. Effects of delay on graphical user interface operations

A concern with adding a delay in displaying submenus is that the user then has to wait to view the submenu items, which could negatively affect their subjective feelings. In addition, a latency or lag in reacting to a user's action directly increases the task completion time.

For mouse-pointing tasks, researchers have proposed models modified from Fitts' law to capture the negative effect of lag (Hoffmann, 1992; MacKenzie & Ware, 1993). Tochioka et al. (2019) investigated the effect of visual latency when steering with a finger while using a tablet computer. Similarly, Friston et al. (2016) conducted a Steering Law experiment with transmission latency.

Note that these studies added a lag from the mouse or finger movement to the cursor displacement, whereas under our condition, an error-accepting delay is added only in exposing submenus when the cursor hovers over the related parent menu item.

2.5. Menu-navigation techniques

There exist numerous techniques to help users to navigate cascaded menus in a short time and/or with fewer operation errors. For example, expanding each menu item's height is a direct approach to ease the cursor-movement difficulty (Cockburn & Gin, 2006; Tanvir et al., 2011; Tanvir et al., 2008). This strategy is theoretically supported by Steering Law: expanding an item tolerance (i.e., W) enables users to move the cursor less carefully, which increases V and decreases MT. Shortening the path length A of Steering Law is also promising to reduce MT, e.g., moving the cursor for a short distance rightwards in a cascaded drop-down menu item opens submenus immediately (Kobayashi & Igarashi, 2003), or using Jumping Menu with which the cursor jumps to the related submenus (Ahlstrom et al., 2006).

Blocking the cursor from deviating from an item tolerance is also effective for menu navigation. Previous studies have proposed using virtual gravity (or *force-fields* (Ahlström, 2005)) and a physical force (Dennerlein et al., 2000) to pull the cursor to the path center. Each of these techniques is a kind of error-accepting technique in motor space, similar to expanding the path tolerance.

Using Marking Menu (Kurtenbach & Buxton, 1993) and its variations (Bailly et al., 2008; Zhao et al., 2006; Zhao & Balakrishnan, 2004) or using Pie Menu (Callahan et al., 1988) enables users to select an intended item without precise cursor positioning. A thorough survey by Bailly et al. (2017) encompasses menu-selection techniques. Still, except for the original paper by Yamanaka (2019), no work has been conducted to evaluate the effects of temporal support for cascaded-menu navigation, although such a delay is widely used in many cascaded-menu implementations.

3. User study

We conducted a user study to measure the performance of steering tasks accepting a deviation within a short duration. This study's apparatus, task, design, and procedure were almost the same as in the original study (Yamanaka, 2019).

A remarkable difference from the original study was the levels of T_{delay}; Yamanaka used 0, 100, 200, 400, 600, and 800 ms to sufficiently cover times shorter than and longer than human's reaction time in path-steering tasks (approximately 273 ms (Lin & Hsu, 2014)). The tested T_{delay} values were acknowledged as being slightly sparse and limited (Yamanaka, 2019). Hence, in our experiment, we investigate the reproducibility of Yamanaka's model for denser T_{delay} values of 0 to 1,000 ms with intervals of 100 ms.

3.1. Participants

We used G*Power (Faul et al., 2007) to determine the sample size. Regarding our main interest of the effect of T_{delay} on MT, the previous study by Yamanaka (2019) showed $\eta_p^2 = 0.423$, which is considered a *large effect size* (van den Berg, 2022). As it is unclear whether a similarly large effect size exists in our current experiment, a smaller (but still *large*) effect size of $\eta_p^2 = 0.05$ was set here (Heidel, 2022). The remaining parameters are $\alpha = 0.05$ and Power $= 0.8$ as the conservative values recommended by Cohen (1988). With these settings, we found that we needed 15 participants.

As we had to balance the order of 11 T_{delay} values, we recruited 22 participants (21 males; age: $M = 21.8$, $SD = 0.96$ years). All had normal or corrected-to-normal vision and were right-handed. Twenty participants were familiar with mouse operations, and ten of them were daily mouse users. This research was approved by the Institutional Review Board at Research Ethics Committee, Faculty of Engineering, Information and Systems, University of Tsukuba (the approval number is 2022R608). Informed consent was obtained from each participant.

3.2. Apparatus

We used a Sony Vaio Z (Core i7-6567U, 3.30 GHz, four cores; 16 GB RAM; Windows 10) as a PC. The display was a Dell 2407WFPb (24-in diagonal, 1920×1200 pixel resolution, 518.4×324.0 mm display area, 3.70 pixels/mm; 16-ms response time; 60 Hz refresh rate). The input device was an iBUFFALO BSMBU05 optical mouse (81.6 g, 1,000 dpi). We used the default cursor speed and enabled pointer acceleration to enable the participants to perform mouse operations with higher ecological validity. We also used a Sanwa MPD-NS3-LL as a large mousepad (350 mm $\times$ 260 mm).

3.3. Task

A trial was to click on a blue start area, pass through a white path, and then click on a green end area (Figure 2). The movement direction was always to the right. Between the two clicks, if the cursor entered the gray out-of-path areas and then stayed there longer than the given T_{delay}, the operation was recognized as a steering error. In this case, a beep sounded to inform the participants of the steering error, and the cursor position was marked with a cross (i.e., "$\times$"). On the other hand, if the cursor returned to the path within T_{delay}, a steering error was not marked. Even when participants caused a steering error or clicked outside the end area (designated as a click error), they had to complete the trial. After each trial, a large circular button labeled "Next" appeared, and the participants clicked it to display the next path condition.

Participants were instructed to minimize the time between the two clicks and to avoid steering errors and click errors. The time from crossing the start line to crossing the end line was measured as MT.

3.4. Design

This study was an $11 \times 2 \times 4$ repeated-measures design with three independent variables: T_{delay}, the path length A, and

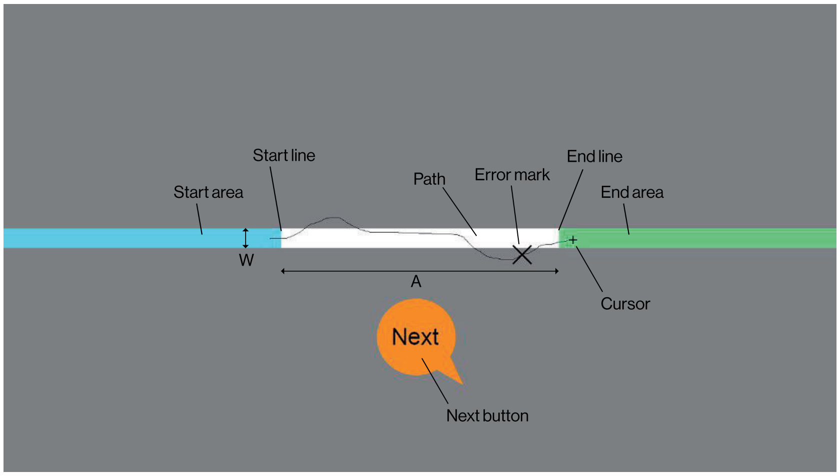

Figure 2. Steering Law task application for our user study.

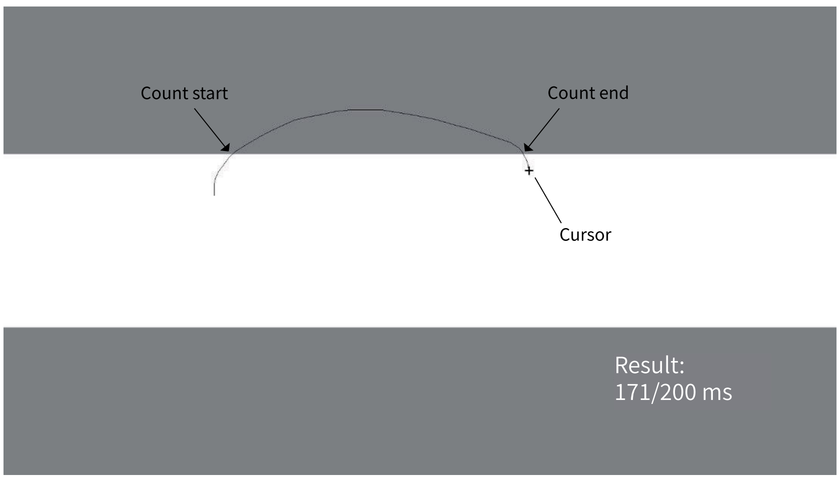

Figure 3. Application for learning the error-accepting delay.

the path width W. We used 11 levels for T_{delay} (0, 100, 200, 300, 400, 500, 600, 700, 800, 900, and 1,000 ms), which gave more variety than in the original study (Yamanaka, 2019).

We used two values for A (480 and 640 pixels, or 130 and 173 mm on the display, respectively) and four values for W (15, 23, 33, and 45 pixels, or 4.05, 6.21, 8.91, and 12.2 mm, respectively), which were all the same as in the original study. The A/W ratios (10.7–42.7) were greater than 10, and thus the participants had to perform continuous visually controlled steering movements under the $T_{delay} = 0$ ms condition (Senanayake & Goonetilleke, 2016; Senanayake et al., 2013; Thibbotuwawa et al., 2012).

We measured five dependent variables: the steering error rate (ER_{steer}), the temporal ratio of deviation from the path ($Ratio_{out}$), the average count of deviations from the path per trial ($Count_{out}$), MT, and V. The click error rate (ER_{click}) was not included as a dependent variable, because our main focus was the analysis of steering operations.

3.5. Procedure

In the study, one *block* consisted of a random order of $2A \times 4W \times 7$ repetitions $= 56$ trials with a fixed T_{delay} value. The first repetition was considered practice. Before each block, the participants used our exercise application, shown in Figure 3, to learn the T_{delay} of the next block except for the $T_{delay} = 0$ ms condition. The experimental procedure for an example participant is shown in Figure 4.

In that exercise application, the participants moved the cursor into the top or bottom gray area from the center white area and then returned the cursor to the white area. A beep sounded if the cursor stayed in the gray area longer than T_{delay}, and the cursor position there was marked with a cross (i.e., "$\times$"). Through this task, the participants learned to return the cursor within T_{delay} after deviating from the path. The participants repeated this until they felt they had sufficiently learned T_{delay}, which typically required 30–60 s.

The order of the 11 T_{delay} values was balanced among the 22 participants using a Latin square pattern. In total, we recorded $2A \times 4W \times 6$ repetitions $\times\ 11T_{delay} \times$ 22 participants $= 11,616$ data points. This task took about 40 min per participant.

4. Results

We analyzed the experimental results by using repeated-measures ANOVA, which can be robust regardless of the data distribution (Dixon, 2008; Mena et al., 2017).

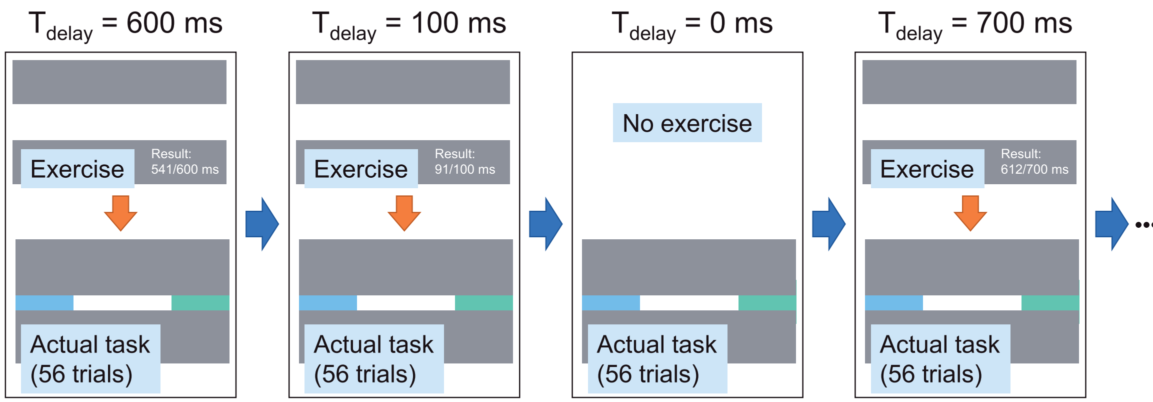

Figure 4. Experimental procedure of an example participant. Except for the $T_{delay} = 0$ ms condition, the participants first experienced the exercise application to get used to the new T_{delay} condition and then performed the actual data-collection block. The order of the 11 T_{delay} conditions was balanced among the 22 participants.

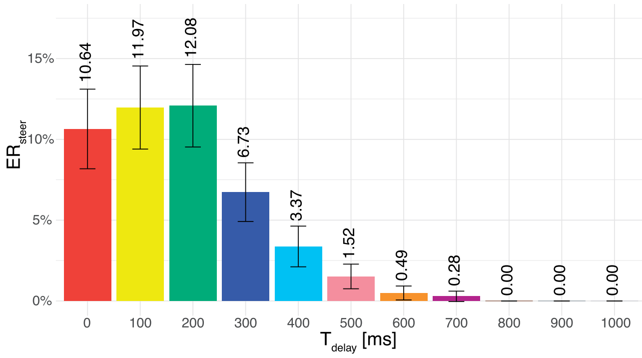

Figure 5. Graph of ER_{steer} on T_{delay}, with error bars representing 95% confidence intervals.

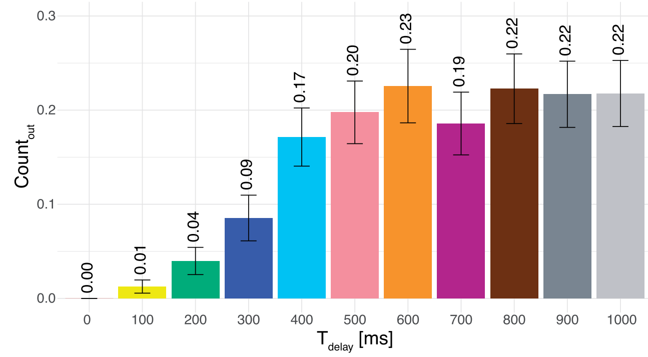

Figure 6. Graph of $Count_{out}$ on T_{delay}, with error bars representing 95% confidence intervals.

4.1. Error rate

Figure 5 shows the results for ER_{steer}. We found significant main effects of T_{delay} ($F_{10,210} = 25.336$, $p < 0.001$, $\eta_p^2 = 0.547$), W ($F_{3,63} = 31.811$, $p < 0.001$, $\eta_p^2 = 0.602$), and A ($F_{1,21} = 7.633$, $p < 0.05$, $\eta_p^2 = 0.267$) on ER_{steer}. A significant interaction was found for $T_{delay} \times W$ ($F_{30,630} = 5.465$, $p < 0.001$, $\eta_p^2 = 0.207$).

Consistent with the original study (Yamanaka, 2019), as T_{delay} increased, the participants tended to make less ER_{steer}. This result suggests that adding error-accepting delays helped users steer through a narrow constrained path more easily. The smaller ER_{steer} for longer T_{delay} should thus contribute to increasing the movement speed as the participants did not have to pay attention to deviating from the path, which will be proven later.

4.2. Average count of path deviations ($Count_{out}$)

We measured the number of times the cursor left the path per successful trial as $Count_{out}$, and the results are shown in Figure 6. The measurement method was the same as in the original study (Yamanaka, 2019). The figure shows that for T_{delay} values up to 600 ms, $Count_{out}$ increased with T_{delay}. On the other hand, when the T_{delay} value was 700 ms or more, $Count_{out}$ changed only slightly.

We found significant main effects of T_{delay} ($F_{10,210} = 21.770$, $p < 0.001$, $\eta_p^2 = 0.509$), W ($F_{3,63} = 53.496$, $p < 0.001$, $\eta_p^2 = 0.718$), and A ($F_{1,21} = 40.806$, $p < 0.01$, $\eta_p^2 = 0.660$) on $Count_{out}$. Significant interactions were found for $T_{delay} \times W$ ($F_{30,630} = 6.465$, $p < 0.001$, $\eta_p^2 = 0.235$) and $T_{delay} \times A$ ($F_{10,210} = 2.305$, $p < 0.05$, $\eta_p^2 = 0.099$). The result showed that increasing T_{delay} tended to induce more path deviations, which would help the participants steer through the path more rapidly.

4.3. Ratio of path deviations ($Ratio_{out}$)

We defined $Ratio_{out}$ as $[(\text{out of path time})/MT] \times 100\%$, as in (Yamanaka, 2019). This metric was to analyze whether a longer error-accepting delay caused users to move the cursor less precisely. As shown in Figure 7, $Ratio_{out}$ increased with T_{delay} when $T_{delay} \leq 600$ ms, similarly to the result of $Count_{out}$.

We found significant main effects of T_{delay} ($F_{10,210} = 10.566$, $p < 0.001$, $\eta_p^2 = 0.335$) and W ($F_{3,63} = 18.636$, $p < 0.001$, $\eta_p^2 = 0.470$) on $Ratio_{out}$, but not for A ($F_{1,21} = 0.436$, $p = 0.516$, $\eta_p^2 = 0.020$). A significant interaction was found for $T_{delay} \times W$ ($F_{30,630} = 2.472$, $p < 0.001$, $\eta_p^2 = 0.105$).

4.4. Movement time (MT)

We found significant main effects of T_{delay} ($F_{10,210} = 11.334$, $p < 0.001$, $\eta_p^2 = 0.351$), W ($F_{3,63} = 28.848$, $p < 0.001$, $\eta_p^2 = 0.579$), and A ($F_{1,21} = 55.549$, $p < 0.001$, $\eta_p^2 = 0.726$)

on MT. Significant interactions were found for $T_{delay} \times W$ ($F_{30,630} = 10.497$, $p < 0.001$, $\eta_p^2 = 0.333$), $T_{delay} \times A$ ($F_{10,210} = 9.448$, $p < 0.001$, $\eta_p^2 = 0.310$), $A \times W$ ($F_{3,63} = 24.551$, $p < 0.001$, $\eta_p^2 = 0.539$), and $T_{delay} \times A \times W$ ($F_{30,630} = 2.228$, $p < 0.001$, $\eta_p^2 = 0.096$).

As shown in Figure 8, MT decreased as T_{delay} increased. However, the effect of T_{delay} plateaued at approximately 400 ms. Because the human's corrective reaction time is ∼270 ms and the system's latency from the mouse movement to the cursor displacement was 57.9 ms,[2] we assume that the participants could begin to return the cursor in 327.9 ms after noticing the cursor deviated from the path. Therefore, giving $T_{delay} = 400$ ms was sufficient to help the participants increase the movement speed, and using $T_{delay} > 400$ ms did not significantly contribute to reducing MT any more.

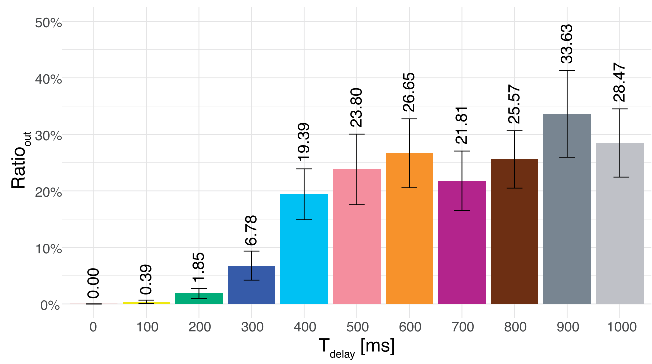

Figure 7. Graph of $Ratio_{out}$ on T_{delay}, with error bars representing 95% confidence intervals.

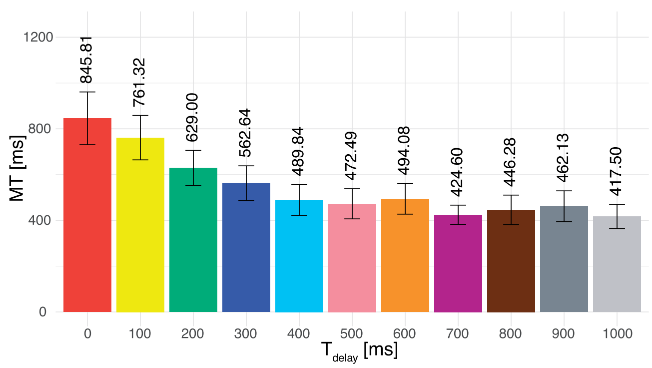

Figure 8. Graph of MT on T_{delay}. The error bars represent 95% confidence intervals.

4.5. Average movement speed (V)

Finally, we found significant main effects of T_{delay} ($F_{10,210} = 3.060$, $p < 0.01$, $\eta_p^2 = 0.127$) and W($F_{3,63} = 89.368$, $p < 0.001$, $\eta_p^2 = 0.810$) on V, but not for A ($F_{1,21} = 2.463$, $p = 0.132$, $\eta_p^2 = 0.105$). A significant interaction was found for $T_{delay} \times W$ ($F_{30,630} = 1.768$, $p < 0.05$, $\eta_p^2 = 0.071$).

4.6. Model fitting

We here used all 88 fitting points (i.e., $11 T_{delay} \times 2A \times 4W$) for regression expressions, as the same procedure used in the original study (Yamanaka, 2019). The candidate models for predicting V and MT were also the same as in the original study.

The fit of the baseline formulation of Steering Law to predict the average speed ($V = bW$; Equation (2)) is shown as Model #1 in Table 1 and Figure 9. To evaluate how the additional factor T_{delay} affected the speed, we show the result of the refined model with T_{delay} of Equation (4) as Model #2 (see Figure 10). In Figure 10, as the optimized coefficient values are $b = 0.0308$ and $c = 0.000595$, we obtain $V = bW + cT_{delay} = b[W + (c/b)T_{delay}]$, and thus the equation in the figure is noted as $y = 0.0308x$ and the x-axis label is $W + (c/b)T_{delay} = W + 0.0194T_{delay}$.

Table 1 also lists the fitting results for predicting MT. The baseline Steering Law is denoted as Model #3 (Figure 11) and Yamanaka's modified formulation is Model #4 (Figure 12). As seen in the table, the modified model had a better adjusted R^2 and AIC than those for the baseline model. Therefore, for these all-data analyses, we conclude that Yamanaka's report was appropriately reproduced; i.e.,

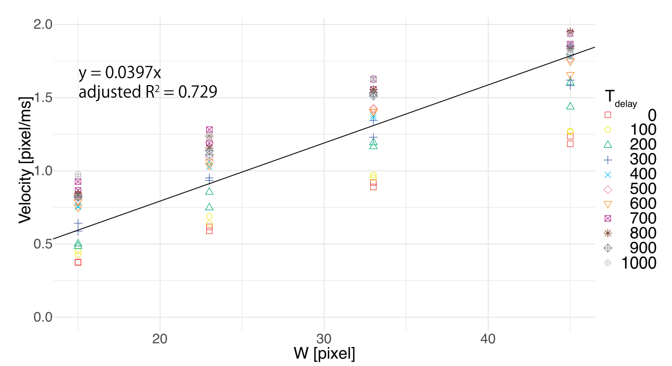

Figure 9. Steering Law fitness of Model #1 for all data points.

Table 1. Model fitting results with adjusted R^2 (higher is better) and AIC (lower is better) values for the candidate models.

Model	a	b	c	Adjusted R^2	AIC
(#1) $V = bW$		0.0397		0.729	1.83
		$p < 0.001$			
(#2) $V = bW + cT_{delay}$		0.0308	0.000595	0.942	−150
		$p < 0.001$	$p < 0.001$		
(#3) $MT = a + b\frac{A}{W}$	63.2	21.7		0.603	1166
	$p = 0.182$	$p < 0.001$			
(#4) $MT = a + b\frac{A}{W + cT_{delay}}$	−73.3	40.5	0.0238	0.964	953
	$p < 0.001$	$p < 0.001$	$p < 0.001$		

The significance results (p-values) are listed at the bottom of each row in the middle column.

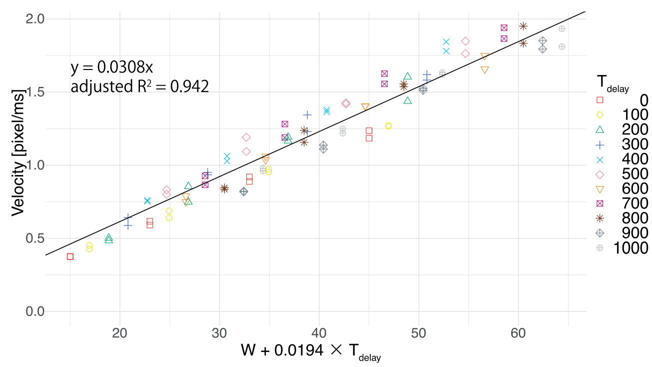

Figure 10. Steering Law fitness of Model #2 for all data points.

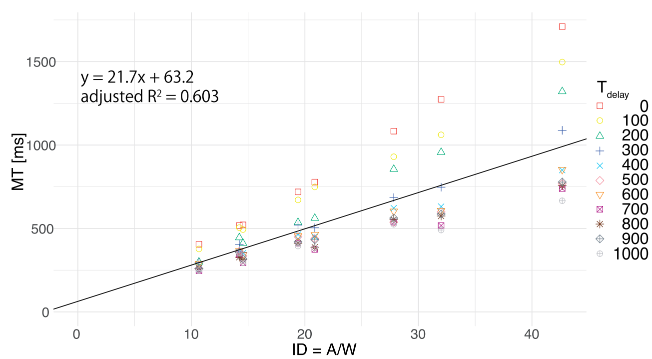

Figure 11. Steering Law fitness of Model #3 for all data points.

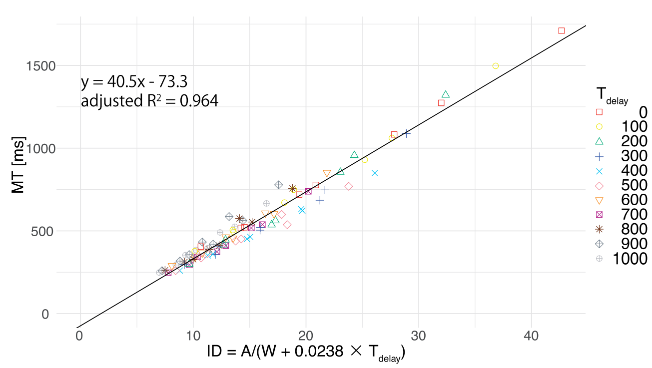

Figure 12. Steering Law fitness of Model #4 for all data points.

adding the T_{delay} term significantly improved the fits in both V and MT models.

5. Prediction accuracy for unknown data

The analyses in Section 4.6 used all *known* MT data. However, model validity in terms of the prediction accuracy should be judged for the future (unknown) data. For example, by using the coefficients for the best formulation to predict MT, Model #4 in Table 1, how accurately can we predict the MT under an untested condition such as $(T_{delay}, A, W) = (375\,\text{ms}, 500\ \text{pixels}, 17\ \text{pixels})$, compared with the baseline Steering Law (Model # 3)?

To investigate this, we ran shuffle-split and leave-one-T_{delay}-out cross-validations to compare these two candidate models. For the shuffle-split cross-validation, we randomly split the 88 fitting points into two groups (train and test datasets) and then compute the MT values for the test dataset on the basis of the model coefficients obtained by the train dataset.

In this process, the data-size proportion for each T_{delay} is not considered, and thus, for example, all eight fitting points $(= 2A \times 4W)$ under the $T_{delay} = 500$ ms condition may be included in the train dataset in frequent iterations. This allows us to not predict MT values for $T_{delay} = 500$ ms, which might be undesirable to evaluate the candidate models' prediction accuracies.

In contrast, the leave-one-T_{delay}-out cross-validation does not have such a problem, but this method cannot evaluate the prediction accuracy when we have to use a small dataset for training. Running both cross-validations complements these pros and cons.

5.1. Shuffle-split cross-validation

When the ratio of (train:test) datasets is (80%:20%), the data-processing steps are as follows.

a. Randomly selecting 80% (rounded up) of the data points for training, i.e., 71 out of 88 path conditions.
b. For the 71 fitting points of the train dataset, regressing Steering Law, $MT = a + b(A/W)$, to obtain coefficients a and b, and also regressing Yamanaka's model to obtain coefficients a–c.
c. Predicting MT values for the remaining 17 test-data points by using these coefficients a and b of the baseline Steering Law and by using a–c of the modified model.
d. Checking the R^2, mean absolute error MAE, and root mean square error $RMSE$ values between the predicted and observed MT values of the 17 test-data points for each of the two models.

To handle the randomness when splitting the whole data into train and test datasets, we performed this process over 100 iterations and obtained the mean values of R^2, MAE, and $RMSE$. Because the prediction performance can change depending on the sizes of the train and test datasets, we report five ratios: (train:test) = (90%:10%), (80%:20%), (70%:30%), (60%:40%), and (50%:50%).

The results are shown in Figure 13. Note that the error bars (95% confidence intervals) were often small for Yamanaka's model and not visible in some cases. Regardless of the train and test data sizes, Yamanaka's model gave, on average, $R^2 > 0.94$, $MAE < 42$ ms, and $RMSE < 53$ ms. Even though the train data became smaller, Yamanaka's model outperformed the baseline Steering Law in all the three metrics (R^2, MAE, and $RMSE$). The two models' error bars do not overlap for the three metrics, and thus we can fairly conclude that the baseline Steering Law has almost no chance of outperforming the modified model.

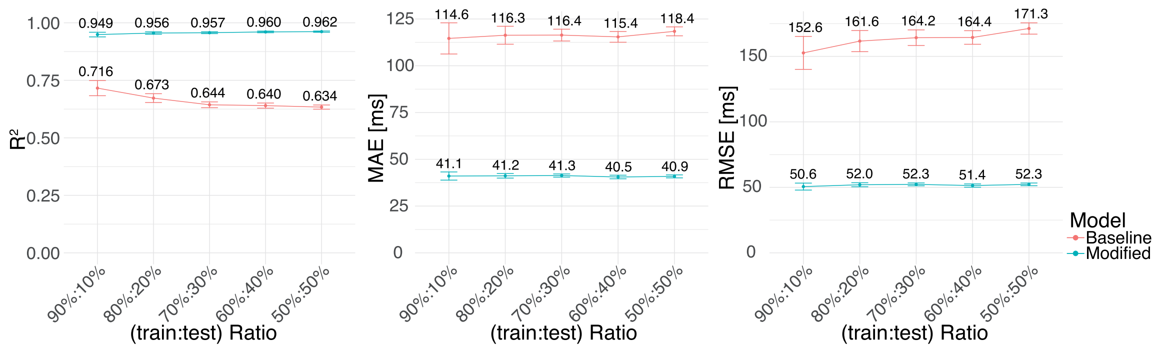

Figure 13. Mean and 95% confidence interval of the shuffle-split cross-validation for our user study.

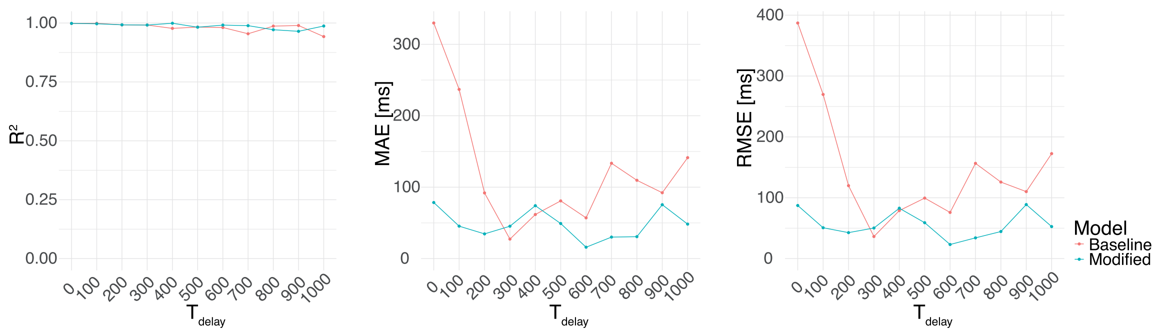

Figure 14. Results of the leave-one-T_{delay}-out cross-validation for our user study.

5.2. Leave-one-T_{delay}-out cross-validation

For this analysis, we first remove the eight data points under the $T_{delay} = 0$ ms condition, and the remaining 80 data points are used for computing the coefficients for both models. We then predict the MT values for the removed eight data points by using the obtained coefficients. The R^2, MAE, and $RMSE$ were calculated by comparing the predicted vs. observed MT values. This process was iterated for each of the 11 T_{delay} values, and then we obtained the average values for each metric for each model.

The results for each T_{delay} are shown in Figure 14. Yamanaka's modified model outperformed the baseline Steering Law in most cases for R^2, but there are several counter-examples. For example, for $T_{delay} = 800$ ms, the observed and predicted MT values are shown in Figure 15. The baseline Steering Law had a higher R^2 of 0.9868, but this metric only indicates the *correlation* between the observed and predicted values.

In contrast, to discuss how accurately a model can predict the task outcome, we should also look at MAE and $RMSE$. Figure 15 shows that Yamanaka's modified model outperformed the baseline for these two metrics. In particular, the modified model predicted MTs more accurately than the baseline when MT was small (< 500 ms).

Figure 14-Middle and Right show that, for $T_{delay} = 300$ and 400 ms, MAE and $RMSE$ values for the baseline Steering Law were lower (better) than those for the modified model. However, for $T_{delay} = 300$ ms, the differences in MAE and $RMSE$ were 18.0 and 14.0 ms, respectively; for $T_{delay} = 400$ ms, those were 12.4 and 3.97 ms. Given our apparatus's refresh rate (60 Hz, i.e., 16.7-ms loop), these differences are too small to determine whether the baseline Steering Law was a better model.

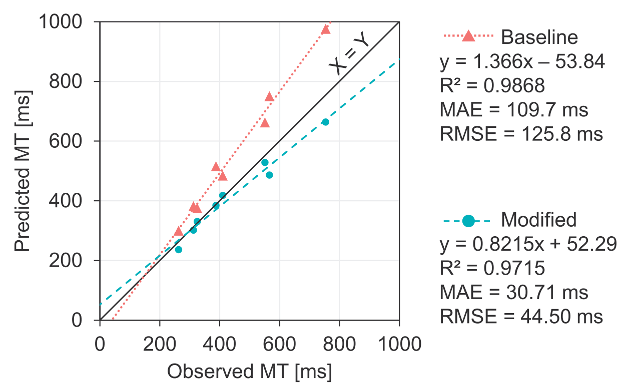

Figure 15. Comparison of predicted and observed MT values under the 800-ms T_{delay} condition in the leave-one-T_{delay}-out cross-validation.

As a result, by averaging the 11 R^2 values in Figure 14-Left, the baseline Steering Law gave $M = 0.9813$ ($SD = 0.01696$), which was slightly worse than that for Yamanaka's model ($M = 0.9876$ and $SD = 0.01040$). In the same manner, for MAE, the baseline Steering Law gave $M = 123.8$ ms ($SD = 83.83$), which was worse than that for Yamanaka's model ($M = 47.97$ and $SD = 19.56$). For $RMSE$, the baseline Steering Law gave $M = 148.4$ ms ($SD = 95.45$), which was worse than that for Yamanaka's model ($M = 56.00$ and $SD = 20.71$). In summary, on the average prediction accuracy, this leave-one-T_{delay}-out cross-validation

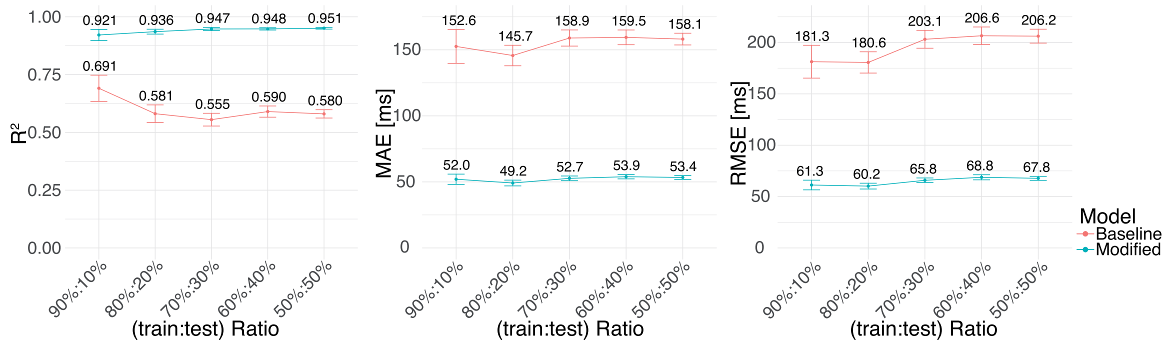

Figure 16. Mean and 95% confidence interval of the shuffle-split cross-validation for Yamanaka's dataset.

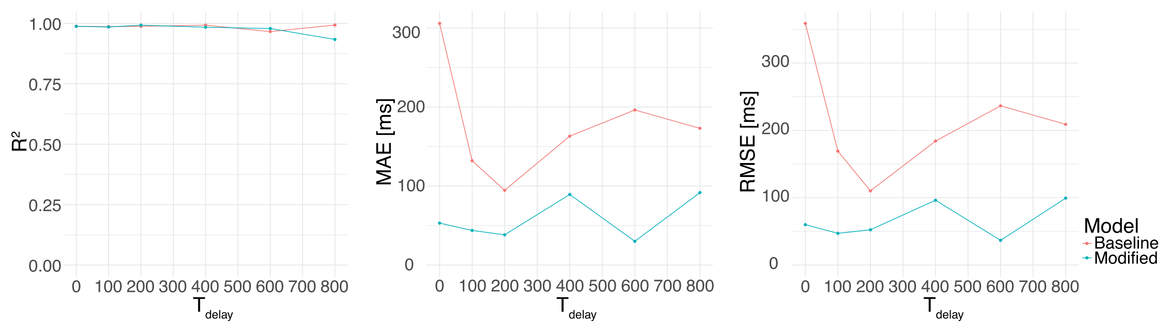

Figure 17. Results of the leave-one-T_{delay}-out cross-validation for Yamanaka's dataset.

showed that the modified model outperformed the baseline Steering Law.

6. Reanalysis of Yamanaka's dataset

We here investigate whether our finding on the superiority of Yamanaka's model to the baseline Steering Law in terms of prediction accuracy for unknown task conditions holds for Yamanaka's original data. As an overview of his experiment, 12 participants joined in the user study in which the same A and W values as in our study were used but T_{delay} values were 0, 100, 200, 400, 600, and 800 ms. In total, $2A \times 4W \times 6T_{delay} = 48$ task conditions were examined. For the all-data analysis to predict MT, the baseline Steering Law gave adjusted $R^2 = 0.577$ and $AIC = 647$, while Yamanaka's model gave adjusted $R^2 = 0.966$ and $AIC = 528$.

6.1. Shuffle-split cross-validation

We ran the same procedure as in Section 5.1. The results are shown in Figure 16. Regardless of the train and test data sizes, Yamanaka's model gave, on average, $R^2 > 0.92$, $MAE < 54$ ms, and $RMSE < 69$ ms. While the train data became smaller, Yamanaka's model outperformed the baseline Steering Law in all the three metrics (R^2, MAE, and $RMSE$).

The two models' error bars do not overlap for the three metrics, and thus, again, we conclude that the baseline Steering Law has almost no chance of outperforming the modified model.

6.2. Leave-one-T_{delay}-out cross-validation

We ran the same procedure as in Section 5.2. The results are shown in Figure 17. By averaging the six R^2 values, the baseline Steering Law gave $M = 0.9857$ ($SD = 0.009134$), which was better than that for Yamanaka's model ($M = 0.9770$ and $SD = 0.01988$). For MAE, the baseline Steering Law gave $M = 177.5$ ms ($SD = 65.82$), which was worse than that for Yamanaka's model ($M = 57.71$ and $SD = 24.19$). For $RMSE$, the baseline Steering Law gave $M = 211.3$ ms ($SD = 76.56$), which was worse than that for Yamanaka's model ($M = 65.30$ and $SD = 24.03$). As a result, overall, this leave-one-T_{delay}-out cross-validation showed that the modified model outperformed the baseline Steering Law.

To sum up, similar to our experimental results, the superiority of the modified model to the baseline Steering Law was consistently observed for Yamanaka's dataset. This supported the robust prediction accuracy of the modified model for unknown task conditions more strongly than testing cross-validations only for our data.

7. Discussion

7.1. Error-accepting delays ease path-steering difficulty

We demonstrated that increasing the error-accepting delay eased the path-steering difficulty. However, tendencies also differed depending on the participant. For example, the results for $Count_{out}$ (Figure 6), $Ratio_{out}$ (Figure 7), and MT (Figure 8) with T_{delay} of 700 ms were lower than the corresponding results for T_{delay} of 500, 600, 800, and 900 ms. The lower results with T_{delay} of 700 ms could have been due to a high ER_{steer} by one participant. This participant experienced the 700-ms T_{delay} condition first and thus might have lacked familiarity with this path-steering task with error-accepting delays. Therefore, these results may be affected by outliers.

The results of V and MT suggested that, overall, a longer T_{delay} helped the participants move the cursor more rapidly and also reduced the ER_{steer}. The only difference in the modified model from the baseline Steering Law is that the T_{delay} term acts to expand the path width. According to the coefficient value $c = 0.0238$ in Model # 4 (Table 1), for example, when T_{delay} is 200 ms, the width is essentially expanded by $\sim$5 pixels ($= 0.0238 \times 200 = 4.76$). Given that the path width was tightly constrained ($15 \leq W \leq 45$ pixels), this increase of five pixels would be noticeable, e.g., a 33% expansion for the $W = 15$ pixels case.

7.2. Robustness of the modified model to all-data and cross-validation analyses

We found that using the modified model improved the fit (adjusted $R^2 = 0.964$) compared with the baseline Steering Law (0.603). In addition, the AIC value of the modified model was significantly lower (953) than that of the baseline (1,166). These results demonstrated that, consistent with Yamanaka's original study (Yamanaka, 2019), taking the error-accepting delay into account enables us to predict MT significantly more accurately than the baseline formulation of Steering Law.

Because of the differences in the participants and experimental design, we cannot directly compare the model fits between our experiment and Yamanaka's study. Still, as a non-rigorous check, there was only a 0.002-point difference in adjusted R^2 values of the modified model between our experiment (0.964) and Yamanaka's study (0.966). We thus did not find a clear drop in model fit in our dataset, which partially supported that the modified model can be robustly used for more various and wider-range T_{delay} values than examined in the original study.

We newly ran cross-validations to evaluate the prediction accuracy for unknown task conditions. The result of shuffle-split one showed that the modified model safely outperformed the baseline in terms of R^2, MAE, and $RMSE$ over 100 random-sampling iterations with no overlaps of error bars regardless of the train-test dataset sizes (Figure 13).

For the leave-one-T_{delay}-out cross-validation, although the baseline model sometimes outperformed the modified one, these counter-examples did not necessarily mean that the baseline model was significantly better, as discussed in Section 5.2. On average, we found the modified model to be superior to the baseline in all the three fit metrics.

These cross-validation procedures were then applied to the dataset from Yamanaka's original study. Again, we confirmed the robust prediction accuracy for unknown task conditions, which filled in a missing piece of his study reporting only the all-data analysis result.

7.3. Contribution of the modified model over the baseline

We have empirically confirmed the contribution of the modified model, i.e., high prediction accuracy for MT. If there is no such a model, when designers want to configure a cascaded menu with a new T_{delay} value, they have no choice but to use the baseline Steering Law. However, based on the result shown in Figure 11, the regression line passes through an area where there are no data points when the task difficulty is high, and MT cannot be estimated accurately.

For example, under the condition where $A = 640$, $W = 15$ pixels, and $T_{delay} = 50$ ms, designers expect that "$MT = 63.2 + 21.7(640/15) = 989$ ms" using the baseline model. However, this baseline model does not include the T_{delay} term, and such a short MT would not be observed when T_{delay} is set to 50 ms. In reality, the MT should probably be between 1,411 and 1,628 ms (under the $T_{delay} = 0$ and 100 ms conditions in our experiment, respectively), which is approximately 50% longer than the predicted MT of 989 ms. In other words, the modified model contributes to preventing wrong decisions in user interface design.

This contribution of the modified model also enables us to lower the cost of conducting additional experiments when changing T_{delay}. Even if the modified model is not used, the MT would be accurately estimated by conducting an experiment with the new T_{delay}. However, this requires additional costs and efforts for both designers and participants. With the modified model, MT can be estimated when T_{delay} is varied; we simulated how the predicted MT changes in accordance with T_{delay} with a 1-ms step as shown in Figure 18. This is not possible with the baseline model and is a clear advantage of the modified model.

7.4. Limitations and future work

We tried resolving limitations in Yamanaka's original study, such as using only six T_{delay} conditions. However, our results are still limited due to the experimental conditions, e.g., testing only two values for A, which should be addressed in the future. Recruiting more numerous and diverse participants (e.g., older adults) will also support the generalizability of the modified model. We used the same input device as in the original experiment (mouse), but given that Steering Law holds for various devices and that input-to-display latencies depend on the device, testing the modified model's validity with non-mouse devices will provide a further contribution.

Our purpose in this study was to examine the reproducibility of Yamanaka's report, and thus further improvements in modeling (e.g., examining a logarithmic or square-root function of T_{delay} when expanding W) to overcome Yamanaka's model were outside the scope. Another untested refinement approach is to combine Steering Law with Fitts' Law (Fitts, 1954). Provided that an expected implication of the modified model is to help designing cascaded-menu configurations, users have to click on an intended submenu item after steering through a parent item (see Figure 1b). This action is called a *targeted-steering motion* and is well-

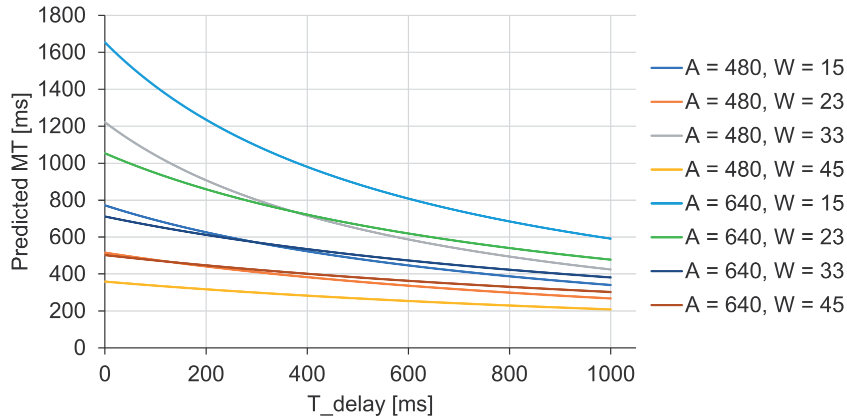

Figure 18. Simulation to predict *MT* when T_{delay} changes precisely with a 1-ms step using the modified model and the parameters from our experiment.

modeled by a combination of the Steering Law and Fitts' Law (Kulikov & Stuerzlinger, 2006; Senanayake et al., 2013). We assume that, to predict *MT* more accurately, the former action (path steering) should be replaced with Yamanaka's model if the menu uses error-accepting delays. These potential refinements of Yamanaka's model will be evaluated in our future work.

8. Conclusion

We investigated human performance in steering tasks with an error-accepting delay, by using a model previously proposed by Yamanaka (2019). Our experiment used 11 T_{delay} conditions instead of six in the original study, resulting in improving the validity of Yamanaka's modified model to predict *MT* across different T_{delay} values. The empirical results showed that the modified model outperformed the baseline Steering Law (Accot & Zhai, 1997) in terms of adjusted R^2 and *AIC* metrics. In addition, we found the modified model to have better prediction accuracy than the baseline under new (untested) task conditions by shuffle-split and leave-one-T_{delay}-out cross-validations. This conclusion also held for the dataset from Yamanaka's original study, which further strengthened the robustness of the modified model. In the future, we plan to conduct user experiments with different experimental apparatuses performed by a larger number of participants to evaluate the modified model's prediction accuracy further.

Notes

1. This modification is justified by the fact that the simplest approach to model a dependent variable is to sum the additional factor (T_{delay}) to the baseline model. This is explained in introductory statistics textbooks or websites, e.g., https://www3.nd.edu/~rwilliam/stats2/l55.pdf, retrieved November 18, 2022.
2. We used the same apparatus as in the original study (Yamanaka, 2019), and this 57.9 ms latency was measured by Yamanaka using a 1000-fps high-speed camera.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Shota Yamanaka http://orcid.org/0000-0001-9807-120X

References

Accot, J., & Zhai, S. (1997). Beyond Fitts' Law: Models for trajectory-based HCI tasks. In *Proceedings of the ACM SIGCHI conference on human factors in computing systems* (pp. 295–302). Association for Computing Machinery. https://doi.org/10.1145/258549.258760

Accot, J., & Zhai, S. (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* (pp. 466–472). Association for Computing Machinery. https://doi.org/10.1145/302979.303133

Accot, J., & Zhai, S. (2001). Scale effects in Steering Law tasks. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1–8). Association for Computing Machinery. https://doi.org/10.1145/365024.365027

Ahlström, D. (2005). Modeling and improving selection in cascading pull-down menus using Fitts' Law, the Steering Law and force fields. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 61–70). Association for Computing Machinery. https://doi.org/10.1145/1054972.1054982

Ahlstrom, D., Alexandrowicz, R., & Hitz, M. (2006). Improving menu interaction: A comparison of standard, force enhanced and jumping menus. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1067–1076). Association for Computing Machinery. https://doi.org/10.1145/1124772.1124932

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Bailly, G., Lecolinet, E., & Nigay, L. (2008). Flower menus: A new type of marking menu with large menu breadth, within groups and efficient expert mode memorization. In *Proceedings of the working conference on advanced visual interfaces* (pp. 15–22). Association for Computing Machinery. https://doi.org/10.1145/1385569.1385575

Bailly, G., Lecolinet, E., & Nigay, L. (2017). Visual menu techniques. *ACM Computing Surveys*, *49*(4), 1–41. 6041. https://doi.org/10.1145/3002171

Banovic, N. (2016). To replicate or not to replicate? *GetMobile: Mobile Computing and Communications*, *19*(4), 23–27. https://doi.org/10.1145/2904337.2904346

Burnham, K. P., & Anderson, D. R. (2003). *Model selection and multi-model inference: A practical information-theoretic approach*. Springer Science & Business Media.

Callahan, J., Hopkins, D., Weiser, M., & Shneiderman, B. (1988). An empirical comparison of pie vs. linear menus. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 95–

100). Association for Computing Machinery. https://doi.org/10.1145/57167.57182

Cockburn, A., Dragicevic, P., Besançon, L., & Gutwin, C. (2020). Threats of a Replication Crisis in Empirical Computer Science. *Communications of the ACM*, *63*(8), 70–79. https://doi.org/10.1145/3360311

Cockburn, A., & Gin, A. (2006). Faster cascading menu selections with enlarged activation areas. In *Proceedings of graphics interface 2006* (pp. 65–71). Canadian Information Processing Society. http://dl.acm.org/citation.cfm?id=1143079.1143091

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum Associates, Publishers. https://doi.org/10.4324/9780203771587

Dennerlein, J. T., Martin, D. B., & Hasser, C. (2000). Force-feedback improves performance for steering and combined steering-targeting tasks. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 423–429). Association for Computing Machinery. https://doi.org/10.1145/332040.332469

Dixon, P. (2008). Models of accuracy in repeated-measures designs. *Journal of Memory and Language*, *59*(4), 447–456. https://doi.org/10.1016/j.jml.2007.11.004

Drury, C. G. (1971). Movements with lateral constraint. *Ergonomics*, *14*(2), 293–305. https://doi.org/10.1080/00140137108931246

Drury, C. G., Montazer, M. A., & Karwan, M. H. (1987). Self-paced path control as an optimization task. *IEEE Transactions on Systems, Man, and Cybernetics*, *17*(3), 455–464. https://doi.org/10.1109/TSMC.1987.4309061

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/bf03193146

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, *47*(6), 381–391. https://doi.org/10.1037/h0055392

Friston, S., Karlström, P., & Steed, A. (2016). The effects of low latency on pointing and steering tasks. *IEEE Transactions on Visualization and Computer Graphics*, *22*(5), 1605–1615. https://doi.org/10.1109/TVCG.2015.2446467

Heidel, E. (2022). *Sample size for repeated-measures ANOVA*. https://www.scalestatistics.com/sample-size-for-repeated-measures-anova.html

Hoffmann, E. R. (1992). Fitts' Law with transmission delay. *Ergonomics*, *35*(1), 37–48. https://doi.org/10.1080/00140139208967796

Hoffmann, E. R. (1997). Movement time of right- and left-handers using their preferred and non-preferred hands. *International Journal of Industrial Ergonomics*, *19*(1), 49–57. https://doi.org/10.1016/0169-8141(95)00092-5

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

Hornbæk, K., Sander, S. S., Bargas-Avila, J. A., & Grue Simonsen, J. (2014). Is once enough? On the extent and content of replications in human-computer interaction. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 3523–3532). Association for Computing Machinery. https://doi.org/10.1145/2556288.2557004

Keele, S. W., & Posner, M. I. (1968). Processing of visual feedback in rapid movements. *Journal of Experimental Psychology*, *77*(1), 155–158. https://doi.org/10.1037/h0025754

Ko, Y.-J., Zhao, H., Kim, Y., Ramakrishnan, I., Zhai, S., & Bi, X. (2020). Modeling two dimensional touch pointing. In *Proceedings of the 33rd annual ACM symposium on user interface software and technology* (pp. 858–868). Association for Computing Machinery. https://doi.org/10.1145/3379337.3415871

Kobayashi, M., & Igarashi, T. (2003). Considering the direction of cursor movement for efficient traversal of cascading menus. In *Proceedings of the 16th annual ACM symposium on user interface software and technology* (pp. 91–94). Association for Computing Machinery. https://doi.org/10.1145/964696.964706

Kulikov, S., MacKenzie, I. S., & Stuerzlinger, W. (2005). Measuring the effective parameters of steering motions. In *CHI '05 extended abstracts on human factors in computing systems* (pp. 1569–1572). Association for Computing Machinery. https://doi.org/10.1145/1056808.1056968

Kulikov, S., & Stuerzlinger, W. (2006). Targeted steering motions. In *CHI '06 extended abstracts on human factors in computing systems* (pp. 983–988). Association for Computing Machinery. https://doi.org/10.1145/1125451.1125640

Kurtenbach, G., & Buxton, W. (1993). The limits of expert performance using hierarchic marking menus. In *Proceedings of the INTERACT '93 and CHI '93 conference on human factors in computing systems* (pp. 482–487). Association for Computing Machinery. https://doi.org/10.1145/169059.169426

Lin, R. F., & Hsu, C.-H. (2014). Measuring individual corrective reaction time using the intermittent illumination model. *Ergonomics*, *57*(9), 1337–1352. https://doi.org/10.1080/00140139.2014.933268

MacKenzie, I. S., & Ware, C. (1993). Lag as a determinant of human performance in interactive systems. In *Proceedings of the INTERACT '93 and CHI '93 conference on human factors in computing systems* (pp. 488–493). Association for Computing Machinery. https://doi.org/10.1145/169059.169431

Mena, B., José, M., Alarcón, R., Arnau Gras, J., Bono Cabré, R., & Bendayan, R. (2017). Non-normal data: Is ANOVA still a valid option? *Psicothema*, *29*(4), 552–557. https://doi.org/10.7334/psicothema2016.383

Montazer, M. A., & Drury, C. G. (1989). A test of the Beggs' model for self-paced movements. *Ergonomics*, *32*(5), 497–511. https://doi.org/10.1080/00140138908966120

Montazer, M. A., Drury, C. G., & Karwan, M. H. (1988). An optimization model for self-paced tracking on circular courses. *IEEE Transactions on Systems, Man, and Cybernetics*, *18*(6), 908–916. https://doi.org/10.1109/21.23090

Montazer, M. A., Vyas, S. K., & Wentworth, R. N. (1987). A study of human performance in a sewing task. *Proceedings of the Human Factors Society Annual Meeting*, *31*(5), 590–594. https://doi.org/10.1177/154193128703100525

Nancel, M., & Lank, E. (2017). Modeling user performance on curved constrained paths. In *Proceedings of the 2017 CHI conference on human factors in computing systems* (pp. 244–254). Association for Computing Machinery. https://doi.org/10.1145/3025453.3025951

Pastel, R. (2006). Measuring the difficulty of steering through corners. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1087–1096). Association for Computing Machinery. https://doi.org/10.1145/1124772.1124934

Rashevsky, N. P. (1959). Mathematical biophysics of automobile driving. *Bulletin of Mathematical Biology*, *21*(4), 375–385. https://doi.org/10.1007/BF02477896

Senanayake, R., & Goonetilleke, R. S. (2016). Pointing device performance in steering tasks. *Perceptual and Motor Skills*, *122*(3), 886–910. https://doi.org/10.1177/0031512516649717

Senanayake, R., Hoffmann, E. R., & Goonetilleke, R. S. (2013). A model for combined targeting and tracking tasks in computer applications. *Experimental Brain Research*, *231*(3), 367–379. https://doi.org/10.1007/s00221-013-3700-4

Tanvir, E., Bunt, A., Cockburn, A., & Irani, P. (2011). Improving cascading menu selections with adaptive activation areas. *International Journal of Human-Computer Studies*, *69*(11), 769–785. https://doi.org/10.1016/j.ijhcs.2011.06.005

Tanvir, E., Cullen, J., Irani, P., & Cockburn, A. (2008). AAMU: Adaptive activation area menus for improving selection in cascading pull-down menus. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1381–1384). Association for Computing Machinery. https://doi.org/10.1145/1357054.1357270

Thibbotuwawa, N., Goonetilleke, R. S., & Hoffmann, E. R. (2012). Constrained path tracking at varying angles in a mouse tracking task. *Human Factors*, *54*(1), 138–150. https://doi.org/10.1177/0018720811424743

Thibbotuwawa, N., Hoffmann, E. R., & Goonetilleke, R. S. (2012). Open-loop and feedback-controlled mouse cursor movements in linear paths. *Ergonomics*, *55*(4), 476–488. https://doi.org/10.1080/00140139.2011.644587

Tochioka, H., Ikeda, H., Hayakawa, T., & Ishikawa, M. (2019). Effects of latency in visual feedback on human performance of path-steering tasks. In *25th ACM symposium on virtual reality software and technology* (pp. 65:1–65:2) Association for Computing Machinery. https://doi.org/10.1145/3359996.3364726

van den Berg, R. G. (2022). *Spss tutorials – effect size – A quick guide*. https://www.spss-tutorials.com/effect-size/

Wilson, M. L., Chi, E. H., Reeves, S., & Coyle, D. (2014). RepliCHI: The workshop II. In *CHI '14 extended abstracts on human factors in computing systems* (pp. 33–36). Association for Computing Machinery. https://doi.org/10.1145/2559206.2559233

Yamanaka, S. (2019). Steering performance with error-accepting delays. In *Proceedings of the 2019 CHI conference on human factors in computing systems* (pp. 570:1–570:9) Association for Computing Machinery. https://doi.org/10.1145/3290605.3300800

Yamanaka, S., & Miyashita, H. (2016). Modeling the steering time difference between narrowing and widening tunnels. In *Proceedings of the 2016 CHI conference on human factors in computing systems* (pp. 1846–1856). Association for Computing Machinery. https://doi.org/10.1145/2858036.2858037

Yamanaka, S., & Miyashita, H. (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* (pp. 65–76). Association for Computing Machinery. https://doi.org/10.1145/3343055.3359697

Yamanaka, S., Stuerzlinger, W., & Miyashita, H. (2017). Steering through sequential linear path segments. In *Proceedings of the 2017 CHI conference on human factors in computing systems* (pp. 232–243). Association for Computing Machinery. https://doi.org/10.1145/3025453.3025836

Yamanaka, S., Stuerzlinger, W., & Miyashita, H. (2018). Steering through successive objects. In *Proceedings of the 2018 CHI conference on human factors in computing systems* (pp. 1–13). Association for Computing Machinery. https://doi.org/10.1145/3173574.3174177

Yamanaka, S., Usuba, H., Takahashi, H., & Miyashita, H. (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* (pp. 844–857). Association for Computing Machinery. https://doi.org/10.1145/3379337.3415896

Zhao, S., Agrawala, M., & Hinckley, K. (2006). Zone and polygon menus: Using relative position to increase the breadth of multi-stroke marking menus. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1077–1086). Association for Computing Machinery. https://doi.org/10.1145/1124772.1124933

Zhao, S., & Balakrishnan, R. (2004). Simple vs. compound mark hierarchical marking menus. In *Proceedings of the 17th annual ACM symposium on user interface software and technology* (pp. 33–42). Association for Computing Machinery. https://doi.org/10.1145/1029632.1029639

Zhou, X., & Ren, X. (2010). An investigation of subjective operational biases in steering tasks evaluation. *Behaviour & Information Technology*, *29*(2), 125–135. https://doi.org/10.1080/01449290701773701

Zhou, X., Ren, X., Hui, Y. (2008). Effect of start position on human performance in steering tasks. In *2008 International conference on computer science and software engineering* (Vol. 2, pp. 1098–1101). IEEE. https://doi.org/10.1109/CSSE.2008.1310

About the authors

Takuma Hidaka is currently studying for a MSc in Engineering at the University of Tsukuba. He received his BSc in Media Sciences and Engineering from the University of Tsukuba in 2021. His research interests include human-computer interaction.

Yusuke Sei belongs to ValueCommerce Co., Ltd. He received his MEng in Engineering from the University of Tsukuba in 2022. His research interests include human-computer interaction.

Naoto Nishida is currently studying for a MSc in Interdisciplinary Information Studies at the University of Tokyo. He received his BSc in Media Sciences and Engineering from the University of Tsukuba in 2022. His research interests include human-computer interaction.

Shota Yamanaka is a senior chief researcher at Yahoo! JAPAN Research. He received a PhD in engineering from Meiji University in 2016. His research interests include human-computer interaction, graphical user interfaces, and human performance modeling.

Buntarou Shizuki is a professor in the Department of Computer Science at University of Tsukuba. His research interests include human-computer interaction. He received BSc, MSc, and DSc from Tokyo Institute of Technology in 1994, 1996, and 2000.