Using Fitts’ Law for a 3D Pointing Task on a 2D Display: Effects of Depth and Vantage Point

Erik Prytz, Michael Montano and Mark Scerbo

The self-archived postprint version of this journal article is available at Linköping University Institutional Repository (DiVA):
http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-141026

N.B.: When citing this work, cite the original publication.

Original publication available at:
https://doi.org/10.1177/1071181312561396

Copyright: SAGE Publications (UK and US)
http://www.uk.sagepub.com/home.nav
Laparoscopic surgery requires surgeons to make judgments about three-dimensional movements using a two-dimensional display. This arrangement reduces the available visual feedback information, such as certain depth cues. The current study used Fitts’ (1954) law to examine the relationship between psychomotor movement time, target size and target distance for a psychomotor pointing task using a laparoscopic instrument in three-dimensional space projected on a two-dimensional display from different vantage points. Analyses demonstrate an effect for depth of target on accuracy, internal consistency and movement time. The results demonstrate that Fitts’ law can be utilized to detect differences in conditions when a three-dimensional task must be completed with the visual feedback presented on a two-dimensional display. No reliable results of vantage point were found. Thus, the location of a two-dimensional display may not be critical for the type of laparoscopic pointing tasks examined in the present study.

A second issue with the presentation of a 3D scene on a 2D display is the observer’s vantage point. Previous research (Todorovic, 2008; Prytz & Scerbo, 2012) has shown that the discrepancy between the vantage point of the observer and image projection center influences spatial judgments of the scene. For laparoscopic surgery, the monitor placement is not standardized. Thus, the monitor displaying the video feed may be placed in such a way that the projection center of the image does not correspond to the surgeon’s vantage point. Such a setup may lead to increased movement times or reduced accuracy of movement. Fitts’ law could again serve as a useful metric to compare the effect of different vantage points on both these variables.

Further, previous research on the effect of vantage points on spatial judgments has focused on static images and perceptual judgments made through aligning a physical object with features of a presented scene. In contrast, the current study used a live video feed and a psychomotor task performed in 3D but presented on a 2D display. Thus, it attempts to bring the previous basic research closer to the more applied context of laparoscopic surgery.

Fitts’ Law

Fitts’ law is a mathematical model that describes the relationship between psychomotor movement time, target size, and distance (Fitts, 1954). It has been used to describe and predict a variety of tasks including how fingers move in relation to the time needed to find and press keys on a keyboard or moving a mouse cursor to a target icon. There are a number of studies that have applied Fitts’ law to tasks that utilize similar psychomotor skills as those used in laparoscopic surgery. For instance, a recent study showed that Fitts’ law held for pointing tasks under restricted visual conditions (Wu, Yang, & Honda, 2010). A number of other studies have found similar results supporting Fitts’ law in relation to pointing tasks (Hoffman, Drury, & Romanowski, 2011; Gery, Vogel, Balakrishhana, & Cockburn, 2008). MacKenzie (1992) outlines this formulation of Fitts’ law:

\[ T = a + b \times ID_e \] (1)
where $T$ is the movement time, $a$ and $b$ are tool and task specific constants, and $IDe$ is the effective index of difficulty. $IDe$ is given by:

$$IDe = \log_2 \left( 1 + \frac{D}{We} \right)$$

where $D$ is the distance from the starting position to the target, and $We$ is the effective target width. The effective target width is particularly useful in experimental research where there is no guarantee that participants will actually hit the target (MacKenzie, 1992). $We$ is calculated by:

$$We = 4.133 \sigma$$

where the standard deviation $\sigma$ is calculated on the accuracy for each participant. That is, $We$ is a measure of the spread, or internal consistency, of the endpoints for each participant. Another derivation of Fitts’ law is a throughput measure (Soukoreff & MacKenzie, 2004). Throughput simultaneously combines both speed and accuracy, effectively combining the intercept $a$, and slope $b$, into one single dependent measure by dividing $IDe$ by $MT$ (Soukoreff & MacKenzie, 2004). In addition to the throughput measure, the average movement times and accuracy, or $We$, for each condition should also be reported to show how the conditions differ.

### The Effect of Vantage Point on Perception

In 2D displays, depth is created with a single projection center that corresponds to the vantage point of the depicted scene. This projection center may be determined by the camera used to capture the scene, or by the algorithms used to render computer-generated scenes. However, individuals may observe the display from a perspective other than the one represented by the projection center.

Two hypotheses have been posited to explain what role an observer’s vantage point plays in the perception of 3D space. According to the vantage point compensation hypothesis, the perception of perspective will not change in relation to the observer’s vantage point (Hagen, 1974; Pirenne, 1970; Shepard, 1992; Vishwanath, Girshick, & Banks, 2005). Other researchers, however, have argued that some distortions cannot be compensated for by human perception (Kubovy, 1986; Todorovic, 2005). Additional research seems to agree that different vantage points result in changes in the perception of an image (Prytz & Scerbo, 2012; Todorovic, 2008).

The perspective transformation hypothesis describes how the projection center of the display changes as the observer alters his or her vantage point, resulting in the perception of a geometrically transformed scene (Todorovic, 2008). Todorovic (2008) and Prytz & Scerbo (2012) have demonstrated that when the vantage point of observers was shifted laterally from left to right, their perception of the object displayed in the 2D display shifted counterclockwise.

Further, research has demonstrated that shifting the participant results in the same geometrically transformed scene as that which occurs when the image is shifted while the participant remains stationary (Prytz & Scerbo, 2012).

### Summary

Fitts’ law (1954) has been shown to be a reliable model that describes the relationship between psychomotor movement time, target size, and distance. Additionally, previous research has demonstrated how the representation of 3D space on 2D surfaces can lead to deviations in spatial judgments through the perspective transformation hypothesis.

The current study expanded on the previous basic research and simultaneously emphasized the real world application of results from the vantage point research. While previous research has used Fitts’ law for 2D tasks, this experiment used a real 3D pointing task that, but represented on a 2D display. Thus, the primary purpose of the current study was to use Fitts’ law to describe the performance of a 3D pointing task shown on a 2D display. The second purpose was to investigate how differences in a participant’s vantage point might affect his/her performance on the pointing task. Theory on vantage point effects predicts that the observer will undershoot the target due to a failure of adequately compensating for the vantage point shift. That is, as the vantage point is shifted to either side of the participant, it is predicted that their performance will deteriorate.

### METHOD

A pointing task was used to measure throughput in order to demonstrate Fitts’ law. Participants were instructed to use a laparoscopic tool to point to targets distributed in the X, Y, and Z planes, using a video feed as visual feedback. The video feed was then moved among six different vantage points (VPs). A pilot test was conducted to determine whether differences existed between images that presented targets with a linear perspective and those without. No differences were found between the two image types and as a result target images without perspective were used in the current study.

### Participants

Eighteen undergraduate students, six males and twelve females, from the ODU subject pool participated for research credit. The average age was 20.33 with a SD of 1.28. All had normal or corrected-to-normal vision.

### Experimental Setup

The experimental setup and layout can be viewed in Figure 1. A custom laparoscopic simulator was used for the experiment. It consisted of a clear plastic box (C) that held a camera (Logitech C270) (B) and a drawing tablet (Wacom Bamboo Create) (H). A stylus paired with the tablet was
attached to the end of a laparoscopic instrument (F). The tablet displayed the five targets. A button (U-HID Nano programmable circuit board) (G) was fixed at a distance from the box, such that when the laparoscopic tool handle was placed on the button, the tip of the stylus extended 11 cm into the box. The drawing tablet could be moved between three slots, one 8 cm (Depth C) from the stylus tip, one 14 cm (Depth B) from the stylus tip, and one 20 cm (Depth A) from the stylus tip. The button and drawing tablet were connected to a laptop (A) that recorded the timing of the button presses and releases, and where the participants hit the tablet.

![Figure 1. Physical layout of experimental apparatus.](image)

A projector (D) was used to back-project the video feed from the camera placed inside the box onto a large screen (E) 3.42 meters in front of the participants divided into two rows of three columns. The center column was oriented directly in front of the participant. The other two columns were located 1.22 meters (19.63°) to the right and the left of the center column (see Figure 2). Each cell could independently display the video feed (1.22 x 1.24 m) allowing for the image to be displayed at different vantage points. Given the distance between the participant and the screen, the size of the video feed, and the size of the target inside the box, the visual angle of the target was approximately 1° for the middle depth condition at vantage point E.

![Figure 2. Layout of cells (A-F) projected on the screen and targets (1-5) from the tablet presented within the active cell.](image)

**Procedure**

The participants completed an informed consent form and received instructions in how to perform the task. Participants completed 180 trials. They were told to point to 1 of 5 targets (see Figure 2). The participants were told to hold the handle of the laparoscopic tool to press the button. After a random interval between one and three seconds, an auditory signal was presented and the participant would point to the designated target as fast as possible. Next, the tablet would be moved to a new depth, the video feed would be moved to a new position, and the participant would be told to point to a new target. Participants pointed to each target at each depth at each video position twice. The presentation of targets, depths, and video positions was randomized into six unique orders of 90 trials. Each successive trial had a different target, depth, and video position. The six orders were balanced across participants.

**Variables**

The independent variables are the video position, target location (along the X and Y axes), and target depth (along the Z axes). The dependent variables are response time and accuracy. The response time measure represented the time from the release of the button to striking the target (i.e., physical movement time from start to end position). The accuracy of the primary task was initially measured in pixels from the center of the target to the position of contact and then translated to mm. The movement time and accuracy were combined into the throughput measure using Fitts’ law.

**RESULTS**

**Data Treatment**

Outliers that were 2.5 standard deviations or more from the mean of each trial were removed and replaced by 2.5 SD from the mean. This included 39 data points for movement time (1.2%) and 35 data points for accuracy (1.1%). The accuracy outliers were all cases where the position of contact was a far distance (mean plus 2.5 SDs) from the target measured in mm. One likely reason for such outliers would be that the participant accidentally aimed at the wrong target. However, this cannot be determined with certainty. Thus, the outliers were truncated and retained to not unduly influence the data distribution.

**Analysis**

Three repeated measures ANOVAs were conducted on throughput, movement time, and accuracy. To calculate the effective width, $W_e$, for each participant, the standard deviation of each participant’s accuracy score (distance from target) was obtained across targets within one depth and video position. This provided a unique effective width, and thus also unique $I_{del}$ through equation (2), for each participant at each depth and video position. As this calculation was performed across targets, the repeated measures ANOVA for throughput was run as a 6 (Video Position; A, B, C, D, E, F) by 3 (Depth; A, B, C) design.
There was a significant main effect of depth on throughput, $F(2, 34) = 31.830, p < .001$, partial $\eta^2 = 0.652$, and a significant VP by depth interaction, $F(10, 170) = 2.484, p = .008$, partial $\eta^2 = 0.127$, but no main effect of VP, $F(5, 85) = .829, p = .532$, partial $\eta^2 = 0.047$. Sidak post hoc tests showed that depth A $(M = 1.978, SE = .079)$ was significantly different from depth B $(M = 2.608, SE = .105)$, $p < .001$, and C $(M = 2.749, SE = .138)$, $p < .001$. Depths B and C were not significantly different, $p = .495$.

![Figure 3](image-url) Interaction of depth (A, B, and C) and VP on throughput.

As noted earlier, one of the main objectives was to examine the effect of vantage point by comparing straight ahead (center) perspectives with shifted (side locations) perspectives. Thus, the video positions were collapsed to form center video positions (B, E) and side video positions (C, D; E, F). Accordingly, to analyze the interaction three Bonferroni-corrected contrasts between the VPs located to either side of the participant (A, C, D, F) and the VPs located in front of the participant (B and E) at the three different depths showed a significant difference for depth B $(Side M = 2.686, Center M = 2.450)$, $p = .006$, but not for depths A $(Side M = 1.966, Center M = 2.001)$, $p > .783$, or C $(Side M = 2.764, Center M = 2.717)$, $p = .578$. Given the differences in throughput, further analyses of the effective width, $We$, and movement time are needed.

A 6 (VP) by 3 (Depth) ANOVA on $We$ showed a significant main effect of depth, $F(2, 34) = 7.040, p < .001$, partial $\eta^2 = 0.293$, but no effect of VP, $F(5, 85) = .386, p = .857$, partial $\eta^2 = 0.022$, or VP by depth interaction, $F(10, 170) = 1.239, p = .270$, partial $\eta^2 = 0.068$. Sidak post hoc analyses showed that depth B $(M = 11.603, SE = .976)$ had significantly lower $We$ than both depth A $(M = 15.727, SE = 1.525)$, $p = .006$, and C $(M = 14.595, SE = 0.906)$, $p = .012$. Depths A and C were not significantly different, $p = .792$.

A 6 (VP) by 3 (Depth) ANOVA on movement time MT showed a significant main effect of depth, $F(2, 34) = 174.071, p < .001$, partial $\eta^2 = 0.911$, and a VP by depth interaction, $F(10, 170) = 3.042, p = .001$, partial $\eta^2 = 0.152$, but no main effect of VP, $F(5, 85) = 1.285, p = .278$, partial $\eta^2 = 0.070$. Sidak post hoc tests showed that depth C $(M = 1.226, SE = 0.060)$ had a lower MT than depth B $(M = 1.630, SE = 0.084)$, $p < .001$, and C $(M = 2.137, SE = 0.114)$, $p < .001$, and depth B had a significantly lower MT than C, $p < .001$. Three Bonferroni-corrected contrasts were again conducted comparing the VPs on either side to the centrally located VPs, as these differed on the throughput measure. These analyses showed a significant difference for depth B $(Side M = 1.197, Center M = 1.696)$, $p < .001$, but not for depths A $(Side M = 2.153, Center M = 2.104)$, $p = .077$, or C $(Side M = 1.212, Center M = 1.256)$, $p = .109$. Thus, the lower throughput at depth B for the centrally located VPs compared to the peripheral VPs is due to an increase in movement time.

To analyze these results further, a 3 (Depth) by 2 (VP; Center v. Side) ANOVA was conducted on the accuracy scores. The analysis showed a significant main effect of depth, $F(2, 22) = 16.645, p < .001$, partial $\eta^2 = 0.495$, but no significant effect of VP, $F(1, 17) = 4.195, p = .056$, partial $\eta^2 = 0.198$, nor a VP by depth interaction, $F(2, 34) = 1.376, p = .266$, partial $\eta^2 = 0.075$. Sidak post hoc tests showed higher accuracy at depth B $(M = 6.715, SE = 0.641)$ than depth A $(M = 9.273, SE = 0.846)$, $p < .001$, and C $(M = 7.756, SE = 0.522)$, $p < .001$, and depth C had a significantly higher accuracy than A, $p < .001$.

**DISCUSSION**

This present experiment applied Fitts’ law to a 3D pointing task guided by a 2D display. Thus, participants were deprived of natural binocular depth cues that would otherwise aid them in aiming the instrument. Further, the experiment manipulated the vantage point of the visual display, as previous research has shown that vantage point can moderate spatial judgments of 3D space represented on 2D surfaces.

Overall, these results show that Fitts’ law can be used in a meaningful way to detect differences among conditions when a 3D task is presented on a 2D display in addition to the more traditional approaches of true 3D pointing tasks (Fitts, 1954) and 2D pointing tasks (Soukoreff & MacKenzie, 2004) as well as more recent developments such as virtual 3D pointing tasks (Cha & Myung, 2010).

Further, the results derived from Fitts’ law demonstrate that the depth of the pointing task influences the performance in different ways. As for the depth effects, aside from the obvious effect that movement time increases as distance to the target increases, accuracy and internal consistency (i.e., the spread of points of contact with the target plane), are affected by the depth of the target. Both accuracy and internal consistency were best at a moderate distance (depth B condition). At the extreme ends, close to the operator (depth C) and far away from the operator (depth A), the accuracy and internal consistency were both lower. There are two possible explanations for this effect; an error in depth perception or an influence of dynamic anthropometrics.

Regarding depth perception, it is possible that participants overestimated the distance from the tip of the stylus to the targets when the tablet was placed at the closest position (depth C), and thus undercompensated their lateral and vertical...
movements in relation to their forward motion. Similarly, they may have underestimated the distance to the targets when the tablet was placed at the furthest position (depth A), and thus overcompensated their lateral and vertical movements in relation to their forward motion. This explanation supports the notion that 3D depth judgments are difficult when presented on a 2D display.

Another explanation, based on dynamic anthropometrics, is that the movements for the close targets were constrained by the hand and arm position, while the far targets required the participants to involve more of their body to reach the targets with the stylus (i.e., their shoulders, chest and waist). This would also result in lower accuracy and internal consistency for the close and far targets as compared to the moderate ones. This explanation suggests that motor performance on this particular task may have been affected more than difficulties in depth perception. Further research is needed to support either explanation.

The second purpose of this experiment was to determine whether the vantage point, that is the placement of the visual feedback, had any impact any on the measures. The results showed some interaction effects of VP and depth, such that the participant’s movement time was increased for the moderate depth condition when the video feed was placed in front of the participant. While this may seem counterintuitive, as it implies poor performance for the central VP, there are some possible explanations. First, this finding may be explained by a speed-accuracy trade-off. That is, when the vantage point is shifted to either side, the participant would emphasize speed, but when the vantage point was in front of them, they emphasized accuracy. The analyses of the accuracy measure were not significantly different, even though the distance from target center measurements showed a difference with 8.110 mm (SE .656) for the side VPs and 7.719 mm (SE .624) for the central VPs. Thus, this does not exclude the possibility that the participants attempted to be more accurate, thus sacrificing speed, but failing to increase their accuracy to a statistically significant degree (p = 0.056). Further, previous research has typically included a physical object used to judge the spatial relations with respect to the 2D display (e.g., Deregowski & Parker, 1992; Todorovic, 2008; Prytz & Scerbo, 2012). However, in this experiment, all judgments were contained within the 2D display. Thus, it is not certain that the vantage point manipulation was effective. Overall, it is clear that further research is needed to examine the effects, if any, of the vantage point on psychomotor tasks.

Conclusions

This research extended previous findings on perceptual issues associated with 2D representations of 3D scenes to a psychomotor pointing task. Fitts’ law was applied to measure performance. It was found that performance on moderate depth was superior to close and far depths. However, it is not clear if these effects occur due to psychomotor issues with the task design or difficulty of perceptual depth judgments. Most importantly, however, it demonstrates the usefulness of Fitts’ law under conditions of 3D tasks represented on 2D displays. Further, this research varied the vantage point but found no reliable differences. One potential impact of these findings in real world settings would be that the placement of laparoscopic monitors may not affect the performance of simple movements such as pointing. Further research is needed on more complex tasks, as well as to determine what underlying factors, e.g. perception or anthropometrics, leads to performance differences across different depths.

REFERENCES


