Available online at www.sciencedirect.com

ScienceDirect

Procedia Manufacturing 3 (2015) 4342 – 4349

Procedia MANUFACTURING

6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences, AHFE 2015

Applying Fitts' law to gesture based computer interactions

Rachael A. Burno[a,]*, Bing Wu[b], Rina Doherty[a], Hannah Colett[a], Rania Elnaggar[a]

[a]*Intel Corporation, 2200 Mission College Boulevard Santa Clara, CA, USA*
[b]*Arizona State University, 7271 E. Sonoran Arroyo Mall, Mesa, AZ 85212, USA*

Abstract

As gesture interfaces become more main-stream, it is increasingly important to investigate the behavioral characteristics of these interactions – particularly in three-dimensional (3D) space. In this study, Fitts' method was extended to such input technologies, and the applicability of Fitts' law to gesture-based interactions was examined. The experiment included three gesture-based input devices that utilize different techniques to capture user movement, and compared them to conventional input technologies like touchscreen and mouse. Participants completed a target-acquisition test and were instructed to move a cursor from a home location to a spherical target as quickly and accurately as possible. Three distances and three target sizes were tested six times in a randomized order for all input devices. A total of 81 participants completed all tasks. Movement time, error rate, and throughput were calculated for each input technology. Results showed that the mean movement time was highly correlated with the target's index of difficulty for all devices, providing evidence that Fitts' law can be extended and applied to gesture-based devices. Throughputs were found to be significantly lower for the gesture-based devices compared to mouse and touchscreen, and as the index of difficulty increased, the movement time increased significantly more for these gesture technologies. Error counts were statistically higher for all gesture-based input technologies compared to mouse. In addition, error counts for all inputs were highly correlated with target width, but little impact was shown by movement distance. Overall, the findings suggest that gesture-based devices can be characterized by Fitts' law in a similar fashion to conventional 1D or 2D devices.

Peer-review under responsibility of AHFE Conference

Keywords: Fitts' law; Gesture interaction; 3D

* Corresponding author. Tel.: 480-554-7786.
E-mail address: rachael.a.burno@intel.com

Peer-review under responsibility of AHFE Conference
doi:10.1016/j.promfg.2015.07.429

1. Introduction

As computing devices continue to evolve, so does our method of interacting with them. Today, we are accustomed to interacting with our devices using input methods like the Mouse, Touchpad, Touchscreen, and the Stylus. Looming on the computing horizon are more natural interaction methods, like air gesture. Air gesture allows users to use their hands, fingers and body to interface with the computing devices through the use of cameras. While currently considered a novelty for gaming based applications, gesture has the potential to assist us in many more computing applications like medical systems and human-robot interaction [1]. As we move further into the field of gesture, we need to understand and be able to articulate the usability of air-gesture inputs. Gesture based input modalities need to afford a level of usability that is acceptable by users.

Today, engineers and designers of input modalities typically refer to ISO standards as the means for determining "good usability" [2]. Specifically, ISO standard 9241-411 [2] establishes a means for highlighting how to evaluate task precision with various input modalities using components of the target acquisition task and measures of index of difficulty, throughput, and predicted movement time established by Paul Fitts. This is known as Fitts' Law and its effectiveness has been demonstrated in hundreds of studies from its establishment in 1954 for both one- and two-dimensional operations. Within the field of Human Factors Engineering, Fitts' law is a prominent mathematical model used to evaluate the effectiveness of pointing devices. It states that the time required to move from one target area to a second target area is a function of the size of the target and the distance to the target [3]. Since the introduction of the computer, Fitts' law has been used to test input devices for pointing and selecting.

1.1. Fitts' law

Fitts' law's first component, the target acquisition task, can be discrete, where the subject moves from a "home" to a "target", or serial, where the subject moves quickly between two targets, much like Fitts' original experiment. [4] Targets can be circular or rectangular; however, circular targets allow for a consistent width regardless of direction of approach to target [5].

Fitts' law's second component are measures for index of difficulty (ID), throughput (TP) and predicted movement time (MT). Index of difficulty,

$$I_D = \log_2\left(\frac{distance+width}{width}\right),$$

describes the number of bits of information transmitted within the context of the human motor system. In essence, the further the distance and smaller the target, the higher the ID and the more difficult the task will be. The second measurement is the rate of information transmission known as throughput,

$$TP = \frac{I_D}{MT},$$

Where $I_D =$ Index of Difficulty and $MT = Movement\ Time$. Higher throughputs describe better performance. The Fitts' model predicts movement time as a function of a task's ID. Predicted movement time is calculated as

$$MT = a + b \cdot \log_2\left(\frac{distance+width}{width}\right),$$

where a = intercept, or the constant, and b = slope.

1.2. Gesture technology and devices

Today, there are several options commercially available for vision based gesture interaction devices. Vision based interaction devices use a camera to "see" a user's movements. The most popular example of which is the Microsoft Kinect device, which connects to the Xbox gaming console and uses whole body, arm, and hand tracking to control games and applications on the device. Outside of gaming, gesture based inputs are becoming more prevalent, and each of them typically uses a different technology to track a user's movements and as with most technology, there are benefits and drawbacks to each of them.

A study conducted by Langolf et al. [6] tested Fitts' law using various limbs. The results showed that the throughput was progressively worse as the limb changed from smaller movements (finger) to larger movements (arm). For the current study, three different types of gesture inputs with three different ranges of motion were analyzed. A longer range hand/arm tracking camera, a shorter range hand tracking camera, and a horizontally placed finger tracking camera.

The first camera is considered a longer-range camera as the distance required between the user and the camera is greater than other commercially available devices. This input requires that the user be one foot to five feet from the camera. The camera is placed behind the screen. This camera tracks larger hand and arm movements and has a 57.5×45 field of view. It requires the user to hold their arm up in front or to the side of their body, with their palm facing the camera. This type of movement can become exhausting for users after even a moderate amount of use. This camera is referenced throughout this paper as *Gesture-L*, where the L stands for long-range movements. The second type of gesture input is a closer-range camera. It uses hand gestures and finger articulation to control objects on the computing device. The user sits closer, with palm facing the camera, which is typically placed on the user's computer or monitor. The camera then tracks the user's hand at a distance of approximately 6 inches to 3.25 feet. In addition to the arm exhaustion mentioned with the device above, because the user is closer to the camera, placement of the camera becomes important. If the camera is placed on top of the computer monitor or screen the user's own hand may now become an obstruction to their view of the screen. This camera is referenced throughout this paper as *Gesture-M*, where the M stands for medium-range movements. The final vision enabled gesture input is a camera that sits horizontally in front of the computing device or computer monitor. Here, the user places their hand over the camera, which tracks small finger movements. The range of motion for this camera is much smaller and reduces the potential for visual obstruction and arm exhaustion which is more likely with the other two cameras. This camera is referenced throughout this paper as *Gesture-S*, where the S stands for short-range movements.

2. Hypothesis and study overview

The current literature is rich with support for utilizing Fitts' law to conduct target acquisition tasks and assess input performance using the measures of ID, TP and predicted movement time. Given the vast extant literature, a hypothesis can be derived that as ID increases, movement time will also increase.

2.1. Method

2.1.1. Participants

A total of 83 participants took part in this study. There were 41 male participants and 42 female participants. All were fluent in English, had no musculoskeletal injuries in back, arm, neck, or shoulder that affect range of motion for gesture based activities, and self-reported as having regular computer use.

2.1.2. Apparatus

One testing workstation was setup with a touchscreen laptop connected to the USB notebook optical mouse and the three commercially available gesture based cameras:

- Gesture-S: Leap motion Controller (Leap motion, Inc, San Francisco, CA, USA). Horizontally placed short-range camera set 3 inches in front of the computing device. This camera tracks small finger movement above the camera.
- Gesture-M: Creative Senz3D Camera (Creative Technology Ltd, Jurong East, Singapore). Medium-range camera (6 in. – 3.25 ft.), placed on laptop monitor set between 2 and 3 feet in front of the participant. This camera tracks shorter range hand motions.
- Gesture-L: PrimeSense Carmine 1.09 3D Camera (Primesense Ltd, Tel Aviv, Israel). Long-range camera (1 – 5 feet), placed behind the laptop at five feet from the participant. Tracks larger hand and arm movements.

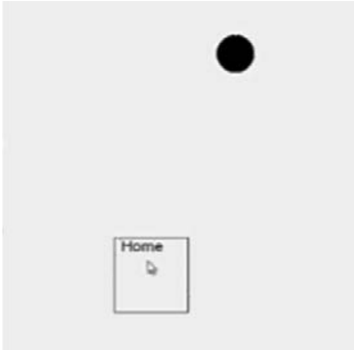


Fig. 1. Fitts' target acquisition task screen.

Customized Fitts' test software was loaded onto the Lenovo for the Fitts' trials. This software was used to complete the Fitts' multi-directional target acquisition tasks and gather user data.

2.1.3. Task Design

A square 50 pixel by 50 pixel home button was presented at a fixed location in the center of the screen. For the movement tasks, there were three defined target width sizes (W) and three defined lengths for target distance (D), allowing for 9 possible combinations and 9 specific Index of Difficulties (IDs). The combinations were as follows:

Table 1. Study Targets. Each target ID displayed twice in random locations on the screen.

Target Distance (D)	Target Width (W)	$ID = \log_2\left(\frac{A}{W}+1\right)$
640 px	70 px	3.34
640 px	30 px	4.48
640 px	50 px	3.79
320 px	70 px	2.48
320 px	30 px	3.54
320 px	50 px	2.89
160 px	70 px	1.72
160 px	30 px	2.66
160 px	50 px	2.07

Each ID displayed twice, for a total of 18 tasks per session. Each participant completed three sessions per input type. To prevent users from learning the ID combinations, nine random or 'dummy' targets with widths varying between 30px – 150px were included. These appeared at random distances and in random locations on the screen. These were not tracked as part of the study metrics. The nine dummy targets were randomly displayed once in each session, interspersed with the 18 defined targets for a total of 27 tasks per session. The within-subjects design resulted in: (3 (d) x 3 (w) x 2 (repeated) + 9 (dummy/distractors)) x 3 (sessions) = 81 tasks.

2.1.4. Procedure

Using a within-subjects design, each participant performed a series of Fitts' target acquisition tasks using all five input types. To account for order bias, participants were assigned the order of the input methods according to a predetermined randomization scheme.

Participants signed the consent form prior to starting the study. Prior to starting the tasks on each input method, the facilitator demonstrated the task several times for the study participant. The movement task consisted of the user selecting a "home button" on the screen and then moving as quickly and accurately as possible to a "target" button. For mouse input, they selected the target by clicking the button on the mouse, for the gesture based input methods they used a button on a wireless presentation clicker, and for the touchscreen they tapped with their finger. The software would start the timer automatically when the user selected the home button and would stop when the participant successfully clicked the target.

Upon completion of the tasks, the Fitts' testing software created a .csv file containing all of the behavioral measures for each participant. This file contained participant number, age range, index of difficulty, gender, response time, input method, throughput, trial number, task number, error count, and target size.

2.2. Data Analysis and results

All dummy trials were removed from data analysis. Error rates were first calculated for each device and each participant. Of a total of 83 participants, two were excluded from data analysis because they had difficulty accomplishing the task and showed high error rates of >20% (much greater than three standard deviations from the mean error rate) even in the mouse condition. Throughput was calculated for each participant and each device as

$$TP = \frac{1}{N}\sum_{i=1}^{N}\frac{I_{D_i}}{MT_I}, [5]$$

where N denoted the combinations of target distances and sizes, using the MT obtained from those trials with no errors.

Figure 2 shows the mean MTs as the functions of IDs for all devices. Clearly, MT increased with ID for the use of all devices. Further linear regression analyses confirmed that MT could be predicted very well from ID not only for mouse ($r^2 = 0.996, p < 0.001$) and touchscreen ($r^2 = 0.915, p < 0.001$), but also for the three gesture-based interfaces (r2 values ranged from 0.857 to 0.984, all the p's<0.01). This suggested the Fitts' model could be applied to the gesture-based interfaces and used to characterize three-dimensional movements during interaction.

Figure 3 shows the mean throughputs, averaged across all participants, for the five devices. Of the five devices, the mean throughput was largest for touchscreen, and smallest for GestureS and GestureM. One-way repeated-measures ANOVA confirmed that this difference was significant (F(4,320)= 965.72, p<0.001). Further comparisons among the devices with Bonferroni corrections revealed that the mean throughputs for mouse and touchscreen were significantly larger than those gesture-based devices (paired tests with Bonferroni corrections, ts(80)>18.32, ps<0.001). No statistically significant difference was found between average participant performance with GestureS and GestureM (paired t(80)=0.63, p=0.99 with Bonferroni correction). Interestingly, both throughputs were significantly lower than that for GestureL (paired ts(80)>7.52, ps<0.001 with Bonferroni corrections), which was inconsistent with Langolf et al.'s [6] findings that throughput was worse for larger limb movements. We will further discuss this in the *Discussion* section.

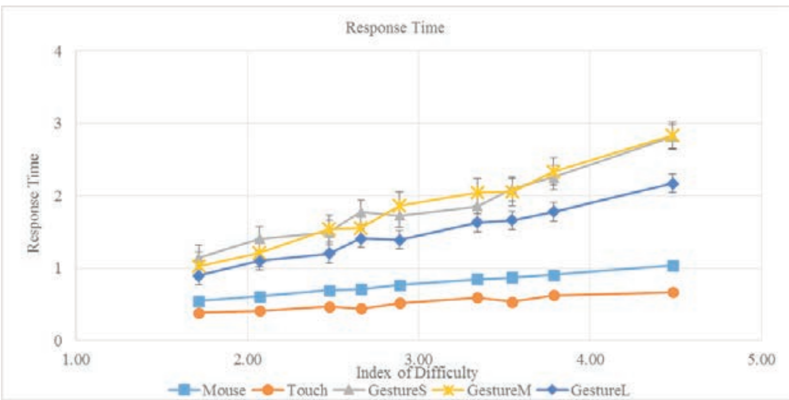


Fig. 2. Relationship between the index of difficulty (ID) and the mean movement time (MT) for 5 input devices.

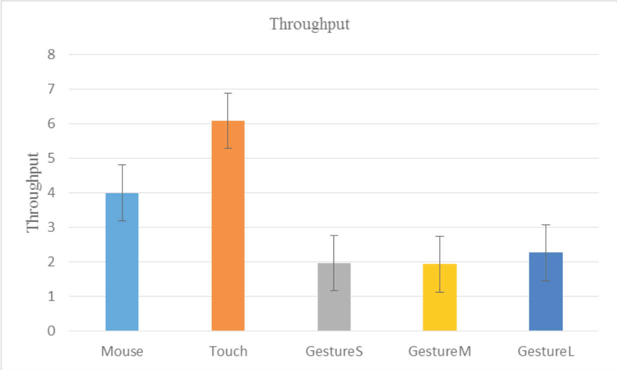


Fig. 3. Average throughputs (TP) for 5 input devices.

In addition to throughput and movement time, error rate was another useful measure of performance. As shown in Figure 4, the lowest error rate was observed in the Mouse condition (paired tests with Bonferroni corrections, ts(80)>9.03, ps<0.001). Participants made considerably more errors with the touchscreen than with the mouse (paired t(80)=13.55, p<0.001 with Bonferroni correction). This partly explained why the movement time with touchscreen was faster than that with mouse. Although the instruction to participants emphasized both speed and accuracy, the participants made trade-offs and their performance might be biased more towards accuracy than speed in the mouse condition because mice are designed for precise pointing. As to the three gesture-based devices, similar error rates were found for using GestureS and GestureM (paired t(80)=1.94, p=0.55 with Bonferroni correction), and the error rate with GestureL were found to be significantly lower as compared to GestureS (paired t(80) =9.30, p<0.001 with Bonferroni correction) and GestureM (paired t(80) =8.20, p<0.001 with Bonferroni correction).

Further analyses were conducted to examine the influence of movement distance and target size, respectively, on the error rates. As shown in Figure 5(a), the rate was almost uniform across the range of movement for all devices. A two-way (Device x Distance) repeated measures ANOVA followed by Bonferroni post-hoc tests found only a marginally significant difference between the error rates for the shortest and longest movements (t(80)=2.50, p=0.04) and a weak main effect of Distance (F(2,160)=3.97, p=0.02). In contrast, the error rates were highly correlated with target width, as shown in Figure 5(b). The observed error rate was highest for the smallest target and gradually reduced with increasing size of the target. The observed effects of target size were relatively weak for the use of mouse and GestureS, but evident for the remaining three devices. A two-way (Device x Width) repeated measures ANOVA found a significant main effect of Width (F(2,160)=265.33, p<0.001) and also a significant interaction of (Device x Width) (F(8,640)=42.71, p<0.001). Bonferroni post-hoc tests further showed that the error rates were significantly different (ts(80)>7.54, ps<0.001) for all three target sizes.

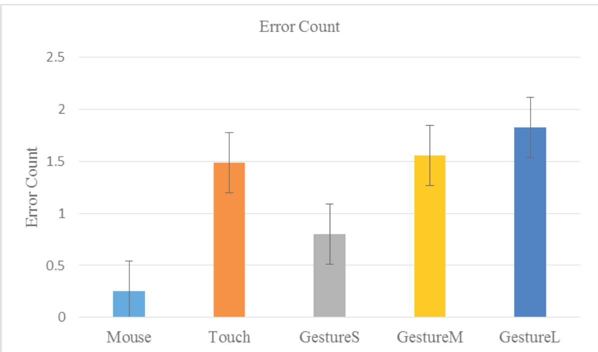
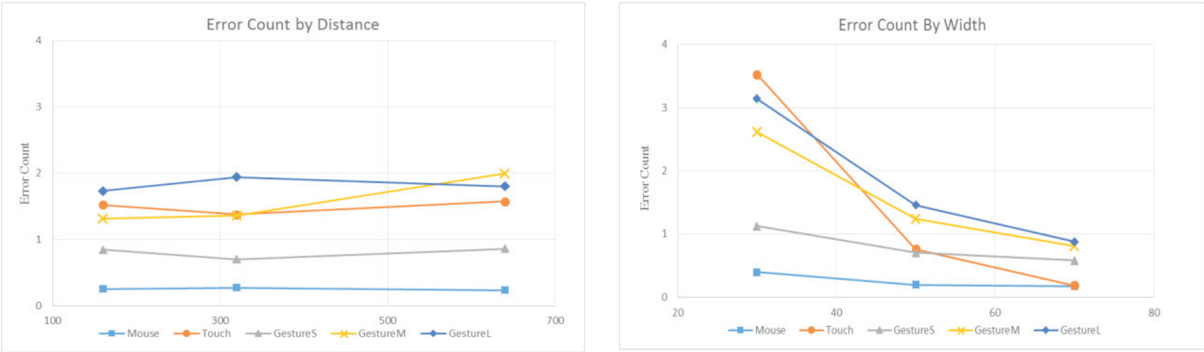


Fig. 4. Average error rates for using 5 input devices.



(a) (b)

Fig. 5. (a) Average error rates as the functions of target width; (b) Average error rates as the functions of target width.

3. Discussion

In this study, three gesture-based interfaces were evaluated in a pointing task and compared to two conventional input devices, namely, mouse and touchscreen. Although the five devices are very different in terms of how users interact with them, a common pattern exists in the users' performance with the devices. The movement time linearly increased with the target's index of difficulty. Such linear relationships provide support for the notion that the Fitts' law can be generalized to 3D and used to characterize the gesture-based devices in a similar fashion like conventional 1D or 2D devices.

As compared to mouse and touchscreen, the gesture-based interfaces have lower throughputs. The authors believe this is partially due to the dimensionality of motor behavior involved in using these devices. Our brain carries out actions by coordinating the activities of agonist and antagonist muscles at the joints. The control becomes more complicated when the degrees of freedom in limb movement increase and more muscles and joints are involved [7]. For example, the movement of a mouse can be achieved by slightly rotating hand around wrist, whereas a precise 3D pointing response with Gesture-L, for example, requires the movements of upper-arm, forearm and hand. Not surprisingly, the duration of such movements in 3D space would be affected more by task difficulty than 1D or 2D movements. However, since 3D gesture based input technologies are still relatively new, there are performance limitations that could be impacting the results.

Another interesting finding comes from the comparisons among the gesture-based devices. In contrast to Langolf et al.'s [6] results that throughput became worse as the amplitude of limb movements increased, our results find a slightly but significantly larger throughput (2.25 bits/s) for using Gesture-L with larger limb movements as compared to Gesture-S (1.90 bits/s) and Gesture-M (1.91 bits/s). Note also that the throughputs reported here were considerably lower than 10 bits/s reported by Langolf et al. [6] for arm movements. The low absolute throughputs might be accounted for by the differences in the experimental task. Langolf et al.'s [6] experiment used a reciprocal tapping task and their participants were instructed to tap as quickly as possible with little concern for accuracy. Here the participants were instructed to perform a single movement toward a target on each trial as quickly and also accurately as possible. So the low throughputs obtained in the present experiment were partly due to the accuracy demands. In addition, it is reasonable to expect that throughputs could be limited by some technical aspects of the devices such as temporal and spatial resolutions. The authors also speculate that the relative ranking of the devices might largely be due to the technical differences. This topic will be further examined in future planned research.

This experiment followed the ISO 9241-411 [3] standard and assessed participants' performance with different devices using the behavioral measures such as movement times, throughputs, error rates, etc. The remaining questions are: how are the behavioral data related to the user's subjective experience, can the behavioral data be used to predict the users' acceptance of certain gesture-based input technology, and does the usage model for these alter the application of Fitts'? These questions will also be investigated in future work.

4. Conclusion

In this paper, the ISO 9241-411 [3] standard was applied using a pointing task to evaluate and compare five interaction devices, including three gesture-based interfaces. The results demonstrated that Fitts' law can be generalized to 3D gesture interaction with computing devices, and consequently used to characterize gesture-based devices. As compared to mouse and touchscreen interfaces, gesture-based interface had lower throughputs; this presumably reflects the dimensionality of the motor behavior involved in using these devices. In addition, error counts – a measure of accuracy - for all input devices were found to be highly correlated with target width, but not movement distance.

References

[1] ISO/TS 9241-411. (2012). Evaluation methods for the design of physical input devices. Switzerland: ISO/TS.
[2] Fitts, P. M. (1954). The Information Capacity of the Human Motor System in Controlling the Amplitude of Movement. Journal of Experimental Psychology, 381 - 391.
[3] Murata, A., & Iwase, H. (2001). Extending Fitts' law to a three-dimensional pointing task. Human Movement Science, 791 - 805.
[4] Wachs, J. P., Kolsch, M., Stern, H., & Edan, Y. (2011). Vision-Based Hand-Gesture Applications. Communications of the ACM, 60 - 71.
[5] Soukoreff, R. W., & MacKenzie, I. S. (2004). Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research in HCI. International Journal of Human-Computer Studies, 751 - 789.
[6] Langolf, G., Chaffin, D., & Foulke, J. (1976). An investigation of Fitts' Law using a wide range of movement amplitudes. Journal of Motor Behavior, 113 - 128.
[7] Van Galen, G. P., & De Jong, W. P. (1995). Fitts' law as the outcome of a dynamic noise filtering. Human Movement Science, 14, 539–571.