A Model of Multi-touch Manipulation

Jian Zhao
Department of Computer Science
University of Toronto
jianzhao@dgp.toronto.edu

R. William Soukoreff
Department of Computer Science
University of Toronto
will@dgp.toronto.edu

Ravin Balakrishnan
Department of Computer Science
University of Toronto
ravin@dgp.toronto.edu

ABSTRACT
As touch-sensitive devices become increasingly popular, fundamentally understanding the human performances of multi-touch gestures is critical. However, there is currently no mathematical model for interpreting such gestures. In this paper, a novel model of multi-touch interaction is derived by combining the Mahalanobis distance metric and Fitts' law. The model describes the time required to complete an object manipulation task that includes translocation, rotation, and scaling. Empirical data is reported that validates the new model (R^2>0.9). Linear relationship between the difficulty and time elapsed is revealed indicating that the model can provide guidelines for interface designers for empirically comparing gestures and devices.

General Terms
Theory, Human Factors.

Keywords
Multi-touch, Human Performance Modelling, Movement Time.

1. INTRODUCTION
Recent advances in technology have made touch-sensitive displays affordable and widely available. In so doing, these advances have brought direct manipulation and multi-touch interaction to the general population for the first time. This has expanded the interface capabilities of modern computers and mobile computing devices, enabling interface designers to use more expressive gestures, for example, flicking, pinching and twisting. However, a comprehensive usability model has not yet been developed for this new set of interactions, and consequently there currently exists no means to comparatively evaluate, model, or predict human performance for the latest generation of interfaces.

Our long-term goal is to develop a performance model for the range of multi-touch interactions that are emerging today, although in this paper we will focus on a common subset of manipulation gestures. Specifically, the objective is to construct a mathematical model of multi-touch interaction that relates: (1) the time required to complete a given multi-touch task, (2) the accuracy with which the task is completed, and (3) the nature and the physical geometry of the task. Such a model will allow researchers to: comparatively evaluate multi-touch hardware and software enabling their improvement, to predict the time required to perform multi-touch interactions which will be of use to interface designers wishing to make multi-touch software easier and faster to use, and to quantify the relative difficulty of these gestures, and so to reveal which gestures are best suited to specific applications leading to multi-touch gesture design guidelines.

GRAND 2011, May 12-14, 2011, Vancouver, BC, Canada

This paper presents a first step toward this goal. A model of multi-touch manipulation is presented that accommodates the gestures pertaining to the translocation, rotation, and scaling, of virtual objects. This new model is empirically validated by the analysis of human movement data gathered during an experiment.

Next we briefly introduce Fitts' law – a model that describes the translocation of a single point, an activity that is a subset of the multi-touch manipulation handled by our new model.

1.1 Movement Modelling With Fitts' Law
Fitts' law [3][4][5] is a movement model that can be used to analyse the performance of pointing devices (mouse, touchpad, stylus, or finger-tip). A rapid aimed movement task is defined by a distance *D*, and a target width *W*, corresponding to the start and destination of a rapid aimed movement (Figure 1).

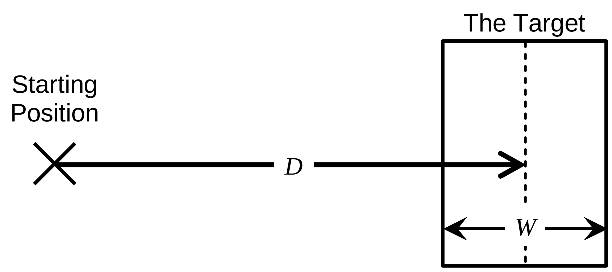

Figure 1 - The Fitts' Law movement paradigm

The difficulty of a movement task is defined,

$$ID = \log_2\left(\frac{D}{W}+1\right), \tag{1}$$

with the units *bits*. The predicted completion time for a movement task is,

$$MT = a + b \cdot ID \tag{2}$$

where *a* and *b* are empirically determined values, and *MT* is measured either in *seconds* or *milliseconds*. Throughput is often defined as $^1/_b$, and measured in *bits per second*.

1.2 Applications of Movement Modelling
The application of Fitts' law to interaction design has had a large impact upon the interface of every desktop computer, cell-phone, pager, and other mobile computing device in use today. The Fitts' law throughput statistic has made it possible to evaluate and compare the efficiency of pointing devices [11]. The ability of Fitts' law to predict movement time has been used to improve the efficiency of software interfaces (for example via "GOMS modelling" [2]). Fitts' law has also made possible models of text entry [10][9], that have allowed several groups to design new

high-efficiency soft-keyboards [6][8]. In short, movement models enable researchers to improve existing user interfaces, and to create novel interaction techniques. However, it is precisely this sort of model that has yet to be developed for touch and multi-touch sensitive interface technologies.

2. MULTI-TOUCH MANIPULATION

Our goal is to determine the speed and precision with which a multi-touch task can be completed, with the goal of formulating a human performance model analogous to Fitts' law but extended so as to encompass the more generalised multi-touch interaction of interest here. In Fitts' law (Equations 1 and 2) distance is a single-dimensional concept used to represent two quantities, a magnitude D, and a tolerance W. However the multi-touch manipulation that we wish to model pertains to three separate quantities: position, rotation, and scale, each having very different units and ranges. To clarify, if the problem was simply to extend Fitts' law to two or three dimensions, then in theory we could simply apply the 2 or 3 dimensional Euclidean distance metric, as the additional dimensions all have the same units (distance). In the case of our new model however, we must consider a definition of distance that can accommodate (1) position (units: metres, range: 0 to 2 metres), (2) rotation (units: degrees, range: -90° to +180°), and (3) scale (units: a unit-less ratio, range: 0.067 to 15, or *one fifteenth* to *fifteen times*)[1]. And further, note that the scale quantity does not behave linearly as distance and rotation do; doubling the length of a movement from 25 to 50 cm, intuitively feels like it will double the effort and time required to complete the movement. However, the two actions of scaling an object to twice its size (scale = 2), and then back again (scale = ½), require approximately the same effort. But the magnitude of the quantities suggests that it is more difficult to enlarge an object than to shrink it, because scale is larger for increasing the size than for decreasing the size, which is not correct. Consequently, the higher-dimensional Euclidean distance cannot be applied.

2.1.1 Mahalanobis Distance

The Mahalanobis distance [7] is a generalisation of the concept of distance that allows multiple factors to be combined into a single statistic, accounting for the individual means, variances, and interdependencies of the factors. The Mahalanobis distance is best described using matrices. Given a position vector $\vec{p}$, vector of means, $\vec{\mu}$, and a covariance matrix, C,

$$\vec{p}=\begin{pmatrix}p_1\\p_2\\\vdots\\p_n\end{pmatrix},\qquad \vec{\mu}=\begin{pmatrix}\mu_1\\\mu_2\\\vdots\\\mu_n\end{pmatrix},\qquad C=\begin{pmatrix}c_{1,1}&c_{1,2}&\cdots&c_{1,n}\\c_{2,1}&c_{2,2}&&\\\vdots&&\ddots&\\c_{n,1}&&&c_{n,n}\end{pmatrix},$$

the Mahalanobis distance is defined,

$$D(\vec{p})=\sqrt{(\vec{p}-\vec{\mu})^T C^{-1}(\vec{p}-\vec{\mu})}\,. \tag{3}$$

2.1.2 Applying the Mahalanobis Distance

Figure 2 illustrates the object manipulation task to be modelled. The experiment task is described in detail in the *Methods* section below, but briefly, an object that may be manipulated by the user is displayed at its starting position, along with a second fixed object that represents the goal or target position. The two objects are differentiated by colour (the text "Starting Position" and "The Target" do not actually appear). Both the manipulatable object and the target have a location defined by their coordinates (x, y), size S, and rotational angle θ.

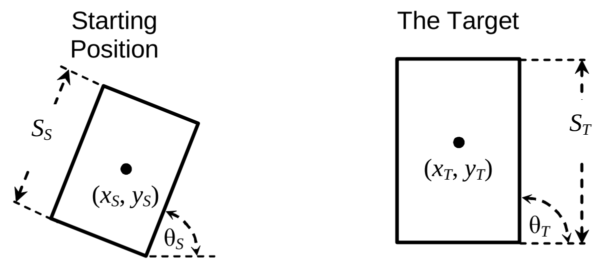

Figure 2 - The multi-touch manipulation paradigm

The Mahalanobis position is defined as the difference between these positional quantities, about a mean of zero (so, $\vec{\mu}=\vec{0}$). Specifically,

$$\vec{p}=\begin{pmatrix}x_S-x_T\\y_S-y_T\\\theta_S-\theta_T\\\log_2(S_S)-\log_2(S_T)\end{pmatrix}. \tag{4}$$

And because the individual components of distance are not inter-correlated, the covariance matrix is diagonal,

$$C=\begin{pmatrix}d^2&&&\\&d^2&&\\&&e^2&\\&&&f^2\end{pmatrix}, \tag{5}$$

where the quantities d^2, e^2, and f^2 represent the variance of the respective components of distance: position (x and y presumed to have identical variances), angle and scale.

Note that we have addressed the non-linearity of scale by applying a logarithm. This approach achieves reasonable values for the relative difficulty of adjusting the scale. For example, leaving an object's size unchanged ($\Delta S = S_{After}/S_{Before} = 1$) results in a difficulty of $\log_2(\Delta S)=0$, and for the example provided above, doubling the size of an object ($\Delta S=2$; $\log_2(\Delta S)=1$) and halving the size of an object ($\Delta S=½$; $\log_2(\Delta S)=-1$) both yield intuitive results.

Because the form of the covariance matrix is diagonal, some simplification can be brought to the formulation, by substituting Equations 4 and 5, into 3,

$$D=\sqrt{\frac{\Delta x^2}{d^2}+\frac{\Delta y^2}{d^2}+\frac{\Delta\theta^2}{e^2}+\frac{\Delta S^2}{f^2}}\,, \tag{6}$$

where $\Delta x^2=(x_S-x_T)^2$, $\Delta y^2=(x_S-x_T)^2$, $\Delta\theta^2=(\theta_S-\theta_T)^2$, and $\Delta S^2=(\log_2(S_T) - \log_2(S_S))^2$.

[1] Note that the ranges depend upon the operator's arm length, flexibility, and maximum span from the tip of their thumb to the tip of their index finger.

2.1.3 *A Model of Multi-touch Manipulation*

The final form of our proposed model is found by inserting the Mahalanobis distance into Equation 1 (and assuming that the width parameter is unity, $W = 1$), and substituting the result into Equation 2, yielding,

$$MT = a + b \cdot \log_2\left(\sqrt{\frac{\Delta x^2}{d^2} + \frac{\Delta y^2}{d^2} + \frac{\Delta\theta^2}{e^2} + \frac{\Delta S^2}{f^2}} + 1\right) \qquad (7)$$

The assumption that the width parameter, *W*, of Equation 1, should be unity, is reasonable given that (1) the Fitts' law width parameter is commonly interpreted as a measure of the standard deviation of movement endpoint accuracy, for example when the adjustment for accuracy is applied [4][5][11], and (2) the variance parameters d^2, e^2, and f^2, subsume the role that *W* plays in Equation 1, in effect normalising the distance metric to have a standard normal distribution, wherein the variance is 1. The values of the variance parameters will be determined empirically.

3. COLLECTION OF EMPIRICAL DATA

The purpose of the study reported here is to examine single-handed dual-finger object manipulation. The dependent variables of interest are the task completion times, and the precision with which three common physical manipulations, translocation, rotation, and size scaling, can be performed. These manipulations are accomplished via pinching, dragging and twisting actions of one hand, simulating the task one would face in adjusting the arrangement and layout of photos in a photo album, or items in a document.

3.1 Participants

Four volunteers participated in this experiment; two were female, two were male. The participants were members of the university community (graduate students and researchers). All four of the participants were right-handed. The participants were allowed to use the hand of their choice to complete the experiment; all used their dominant hand. All participants had used a multi-touch device before; two were daily users of a multi-touch device.

3.2 Apparatus

Participants were asked to manipulate virtual objects on a Microsoft *Surface* multi-touch sensitive display (the original model of Surface). The display size was 30 inches diagonal, and the resolution, 1024×768 pixels. The Surface was physically raised by placing it on a stand 40 centimetres high, so the participants stood while operating it (this made it easier for the participants to reach the entire display, and avoided the problem of where the participants should put their knees while using the original model of Surface).

Custom experiment software was written using the Microsoft Developer Studio development environment and the C# language. This software presented the manipulation tasks to the participants, and recorded the time that the participants took to perform the tasks. Particular care was taken in the software to accurately measure the participants' movement times.

3.3 Method

The participants were observed as they completed a representative series of multi-touch interactions. The participants performed the manipulation tasks using the thumb and index finger of their preferred hand. Clutching (lifting the hand off of the surface) during a manipulation was not allowed; the object manipulations had to occur in a single motion. The participants' accuracy and manipulation times were recorded by the experiment software.

3.3.1 *Conditions*

Three levels were employed for each of three movement parameters of interest. Specifically,

Distance conditions: 0, 300, 500 pixels ×
Angle conditions: -30, 0, 30 degrees ×
Scale conditions: 0.6, 1.0, 1.6 times = 27 conditions total.

Each condition was repeated 8 times, for a total of 216 trials performed by each participant. The order of presentation of the 216 trials was randomised. For each trial, the direction of movement (i.e., the vector from the manipulatable object's initial position to the target) was randomly chosen from the eight directions (N, S, E, W, NE, NW, SE, and SW). The size and rotational angle of the target of each trial was uniform: 150 pixels width × 150 pixels height, with 0 degrees of rotation.

3.3.2 *Procedure*

The experimenter demonstrated the manipulation task to the participant, and the participants performed 32 practice trials before the study trials began. (Data from the practice trials was ignored.) Each trial began with the display of the "Start" button. After clicking the button, the manipulatable object (a blue bordered square) and target (a white bordered square) were displayed (Figure 3). Both squares were filled with a yellow-to-red colour gradient to indicate their orientation. The manipulatable object was semi-transparent so that it did not occlude the target near the end of the trial as it overlapped it. Once the object was close enough to the target that the sum of the distances between the corresponding corners was less than a predefined threshold (150 pixels), the border of the object turned green to indicate that the trial was successful (i.e., not an error).

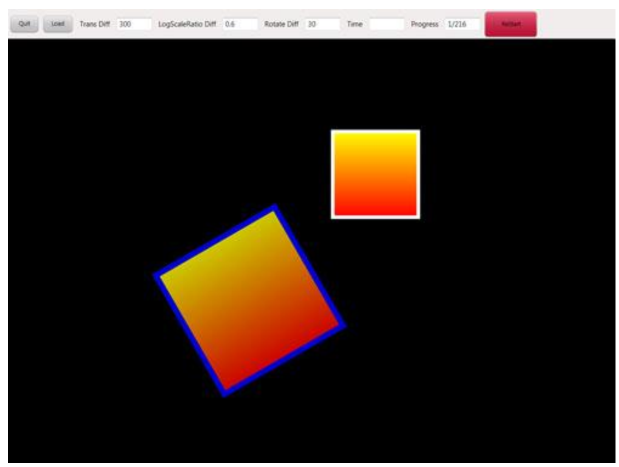

Figure 3 - Snapshot of the experiment software

Once the Start button had been clicked, the software measured the time beginning with when the participant first touched the display surface until the fingers were lifted at the completion of the trial. Errors occurred if the user failed to complete the manipulation task to within the indicated accuracy (i.e., the boarder of the object had not yet turned green). Trials in which errors occurred were presented to the participants again.

4. RESULTS

The data was gathered and regression analysis was performed on the non-error trials. The results appear in Table 1 and Figure 4. The *d*, *e*, and *f*, values represent a single standard deviation for the distance, rotation, and scale axes of manipulation respectively.

Table 1 - Results of regression analysis

Variable	Value	Units
a	-43.74	ms
b	376.53	ms/bit
d	23.97	pixels
e	4.74	degrees
f	0.14	[scale units]
R (R^2)	0.9689 (0.9388)	

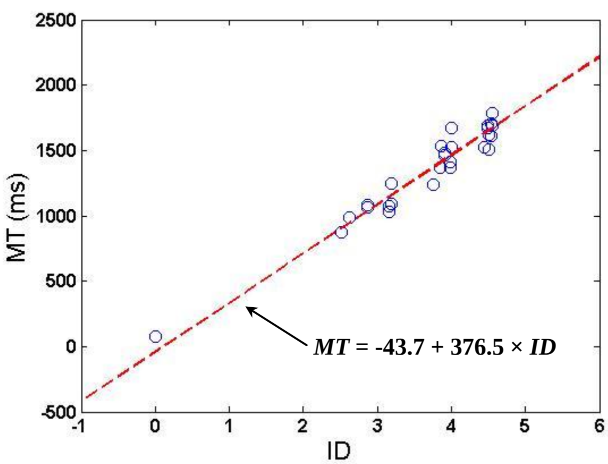

Figure 4 - Movement time versus index of difficulty

5. CONCLUSIONS AND FUTURE WORK

We have proposed a novel model of multi-touch manipulation derived by applying the Mahalanobis distance metric to the index of difficulty equation from Fitts' law (Equations 1 and 2). The result is a relation similar to Fitts' law, but that models a richer set of movements. The model fits empirical data very well (with R and R^2 values well above 0.9).

Our model not only can predict the completion time of a multi-touch task, it can also be used to quantify the effort necessary to complete that task, because difficulty has a linear relationship with elapsed time. Thus the multi-touch model developed here makes it possible to empirically compare devices and gestures in a manner analogous to the way that Fitts' law does. Thus this model can be used to improve multi-touch interfaces, and to develop guidelines that will be of interest to interface designers.

It is our intention to further analyse the empirical data we have gathered. In particular we are interested in determining whether our participants made progress along all of the dimensions (distance, angle and scale) simultaneously, or whether there was any precedence between these dimensions.

Also, we intend to extend this model to more elaborate physical activities. For example, can a higher-dimensional model of steering be derived, in the same way that Fitts' law is the basis from which Accot & Zhai [1] developed their model of steering and path following? Or, can our model be extended to simultaneous bimanual movements? For now, these questions remain unanswered.

6. ACKNOWLEDGMENTS

We would like to express our thanks to those who participated in our experiment and to the reviewers for their time and helpful comments.

7. REFERENCES

[1] Accot J. & Zhai S. (1997) Beyond Fitts' law: models for trajectory-based HCI tasks. *Proceedings of ACM Conference on Human Factors in Computing Systems – CHI 1997*, 295-302. ACM: New York.

[2] Card S. K., Moran T. P. & Newell A. (1983) *The psychology of human-computer interaction*. Hillsdale, NJ: Lawrence Erlbaum.

[3] 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.

[4] MacKenzie I. S. (1989) A note on the information-theoretic basis for Fitts' law. *Journal of Motor Behavior*, 21, 323-330.

[5] MacKenzie I. S. (1992) Fitts' law as a research and design tool in human-computer interaction. *Human-Computer Interaction*, 7, 91-139.

[6] MacKenzie I. S. & Soukoreff R. W. (2002) Text entry for mobile computing: Models and methods, theory and practice. *Human-Computer Interaction*, 17, 147-198.

[7] Mahalanobis P. C. (1936) On the generalised distance in statistics. *Proceedings of the National Institute of Sciences of India*, 2 (1): 49–55. http://www.insa.ac.in/insa_pdf/20005b8c_49.pdf

[8] Silfverberg M. (2007) Historical Overview of Consumer Text Entry Technologies. In MacKenzie I. S. & Tanaka-Ishii K. (Eds.) *Text entry systems: Mobility, accessibility, universality*. Morgan Kaufmann Publishers, San Francisco.

[9] Silfverberg M., MacKenzie I. S. & Korhonen P. (2000) Predicting text entry speeds on mobile phones. *Proceedings of the ACM Conference on Human Factors in Computing Systems – CHI 2000*, pp. 9-16. ACM: New York.

[10] Soukoreff R. W. & MacKenzie I. S. (1995) Theoretical upper and lower bounds on typing speed using a stylus and soft keyboard. *Behaviour & Information Technology*, 14(6), 370-379.

[11] 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, 61*, 751-789.