A Model of Error in 2D Pointing Tasks

Keenan R. May  
Rice University  
krm2@rice.edu

Michael D. Byrne  
Rice University  
byrne@rice.edu

ABSTRACT

Fitts’s law describes one of the most studied relationships in HCI. Conceived as a model of aimed pointing, variants of Fitts’s law have since been demonstrated to predict movement time and task difficulty of pointing tasks across an astonishing array of contexts and environments. While some aspects of task performance can be predicted well, there is still not a defined model of systematic, nonuniform pointing error. The present study addresses the nature of said error and its potential causes, and produces a model that can be employed to predict the shape of the endpoint variance distribution produced during two dimensional aimed pointing tasks.

INTRODUCTION

Fitts’s law is fundamentally a model of information transfer with the goal of uncertainty reduction. (Fitts, 1954). Fitts’s crucial innovation was to adapt the Shannon-Hartley theorem of signal bandwidth, which relates the maximum theoretical capacity of an electrical channel to the available bandwidth, the signal amplitude, and the amount of noise present, to the human perceptual-motor loop. He theorized that the difficulty and time taken to complete a pointing task could be modeled by treating the human perceptual motor loop as a signal processor with limited bandwidth, thus relating the difficulty of the task to the magnitude of the movement necessary (called Amplitude or Distance, which is analogous to the signal strength in the Shannon-Hartley theorem), and the depth/error tolerance of the target (called Width, which is analogous to signal noise). In its own terminology, Fitts’s law tells us about movement time (MT), task difficulty (ID), and task performance (IP). Fitts (1954) describes the relationship between these variables and the task conditions with the following equations:

\[ ID = -\log_2 \left( \frac{W}{2D} \right) \]  
(1)

\[ IP = -\log_2 \left( \frac{1}{MT} \frac{W}{2D} \right) \]  
(2)

THE ORIGINS OF POINTING ERROR

The index of performance (IP) here refers literally to bits/second. What is worth noting with regard to the present study is that the law is rooted entirely in information theory, and assumes that describing the pointing task as purely about balancing speed versus accuracy given limited bandwidth. Fitts & Peterson rephrased the equation, and added an explicit equation for predicted movement time (MT):

\[ ID = \log_2 \left( \frac{2D}{W} \right) \]  
(3)

\[ MT = a + bID \]  
(4)

In order to get two new terms, the observed movement times are regressed (generally using a least-squares regression) relative to the index of difficulty in order to obtain empirical constants \( a \) and \( b \), which represent base movement time and device-specific difficulty, respectively. These constants allow the law to be adjusted for different tasks and devices. The constant \( a \) is generally taken to represent the preparatory (non-movement) components of the task— for example, the time required to visually acquire the target and decide that it is indeed the desired target. The constant \( b \) has to do with the task difficulty introduced by the device. The throughput (TP) is thus defined as:

\[ TP = \frac{1}{b} \]  
(5)

The relationship described by equations 3 and 4 has been tested at length, and has proven robust with a few specific exceptions. The experimental task as originally conducted by Fitts (1954) was to have subjects tap back and forth between two vertically infinite but horizontally narrow (term W) strips with a stylus. The law was first applied to computing tasks by Card, English and Burr (1978), in which subjects used a mouse, joystick, and keyboard to move from a fixation point to a highlighted block of text. Throughput (5) represents the bandwidth of the operator-device system. Throughput varies both primarily by device and to a lesser extent by operator skill with the device. Card, English and Burr (1978) demonstrated that the mouse tends to have the highest throughput. Mackenzie (1992) provides an excellent review of Fitts’s law studies that speaks to its robustness in a variety of scenarios. MacKenzie (1992) proposed a revised version of law that is more directly derived from the Shannon-Hartley theorem and its bandwidth metaphor that has since been widely (but not universally) adopted, with the following form:

\[ MT = a + b(log_2(1 + \frac{D}{W})) \]  
(6)

Fitts’s law has been tested across various human limbs and maintained its descriptive power. It works well for nonballistic horizontal movements over 200ms (Gan and Hoffmann, 1988). The explanation for this is that the law is describing the bandwidth of the governing visual/motor transmission pathway, and that such factors are generally universal regardless of the limb and device. As such, the reason why ballistic movements are not explained by the law is thought to be because they don’t allow time for this feedback loop to
offer corrective input (Gan and Hoffman 1988). There have been several attempts to adapt the law so that it can effectively describe two dimensional pointing tasks with irregular approach angles. The main issue is that it is unclear what to consider the target width in these cases. Crossman and Goodeve (1983) noted that the height of the target (which the original model assumes is effectively infinite) had a significant effect on movement time. They proposed the following model in response to this observation:

$$\text{MT} = a + b(\log_2(\frac{D}{W} + 1)) + c(\log_2(\frac{D}{H} + 1)) \quad (7)$$

Hoffman and Sheikh (1994) observed that the target height should only matter when the normally distributed error cloud would be taller than the vertical target area. Mackenzie and Buxton (1992) had taken this property into account, and, after testing several models (each of which take the target height into account in different ways), found that the following equation was the best fit for the data:

$$\text{MT} = \log_2(\frac{D}{W} + 1) \quad (8)$$

In this equation, W refers to the ‘apparent width,’ which is defined as the magnitude of the target along the approach vector, calculated geometrically. This version holds up well for bivariate pointing tasks, and implies a normally distributed error distribution along both axes. In addition, Hoffman and Sheikh noted that the relationship that this formulation predicts between height, width, and MT is too simplistic to be accurate, since whichever is smaller among the Height and Width becomes completely dominant. The data indicates more of a graded contribution to effective target size. Accent & Zhai (2003) developed a formulation that still uses a Euclidean calculation of effective target width (W’), but that also allows target height and width to have different contributions to the index of difficulty through the inclusion of a weight n:

$$\text{MT} = a + b(\log_2(\sqrt{(\frac{D}{W})^2 + n(\frac{D}{H})^2} + 1)) \quad (9)$$

This is currently the most common formulation of Fitts’s law to be employed on bivariate tasks. In order to avoid the issue of 2D target width, the present study employs circular targets. Future studies as to the best way to model error distribution given a rectangular- or perhaps even highly irregular might increase the accuracy of the model developed here. However, the present investigation was primarily concerned with the nature of the error cloud given the simplest possible target.

Many Minimizations
Of particular note to the present investigation is that none of the two dimensional versions of Fitts’s law explicitly tell us anything about error. They do imply that error distribution will be gaussian about the target center along both the on and off axis (Wobbrock et al., 2011b). In order to understand why this might or might not be the case, it is helpful to examine the underpinnings of neuromotor error. The underlying question is what sort of strategic minimization is being made to contain the effects of signal noise, and what this says about the shape of the endpoint error cloud.

While Fitts’s law is generally unconcerned with the details of the physiology that realizes a information processing pathway being modeled, in recent years neuroscience research has provided insight into what neuromuscular noise actually is and how it should influence the bandwidth needs of operators. Gomez et al. (1986) measured motor neuron activity and found that firing rate variance increases proportionally with average firing rate. This leads to the conclusion that higher magnitude signals should have more error. Jones et al. (2002) noted that aggregate motor unit recruitment patterns lead to increased error based on the magnitude of the command signal. A similar conclusion was reached by Slifkin and Newell (1999), who studied isometric force production and found evidence that noise in the output force is related to the amount of transmitted information.

Churchland et. al (2006) compared several competing hypotheses on the source of neuromotor noise. The first hypothesis is that noise arises due to variability in the late-process transmission to the muscles, and the second is that it arises at the incipience of the process during the preparation of a movement due to the selection of a less than perfect motor command ‘plan.’ The investigators had monkeys perform a pointing task, and recorded the activity of particular neurons in the dorsal premotor cortex and primary motor cortex. They attributed about half of the variance in reach speed (which should have been inversely related to task error) to activity in the premotor cortex, which is generally associated with movement planning. Other studies, such as Johnson et al. (2002) provide evidence that motor planning has a negligible impact on pointing velocity and error, instead proposing that noise is introduced primarily during downstream neuromuscular execution. Specifically, Johnson et al. (through a review of existing literature) attribute such transmission variation to post-cerebellum motor noise, sensory noise (both transmission noise and situational signal ambiguity), and conjunctive noise in the sensorimotor feedback loop, which includes prediction of limb position based of sent motor commands.

Frameworks
The supposition has been made that the information transfer being described Fitts’s law is constitutes visual feedback-mediated responses to the result of submovement that can be seen as iterative corrections (Crossman and Goodeve, 1983). While Fitts’s original account relied on information theory and parallels to signal bandwidth, looking at pointing as a series of iterative corrective submovements provides room for a mechanical account of how random noise may be introduced to the system, and what the consequence of such noise should be. Van Beers et al. (2004) describes a framework known as TOPS (task optimization in the presence of signal-dependent noise), which models how the brain integrates sensory and motor information in order to minimize the effects of noise.
TOPS is based on the assumption that error arises due to physiological signal noise which scales in response to the magnitude of the movement and associated movement command (as opposed to constant temporal noise), and that the central nervous system aims to minimize error thus derived by integrating and comparing various proprioceptive, motor and sensory data sets. Movement trajectories in a variety of paradigms are predicted with this sort of optimization in mind. The implication is that movements of the sort predicted by Fitts’s law tend to have an error rate of around 4% because the central nervous system makes a pathing optimization that minimizes endpoint variance, as predicted by the summed magnitude of the correction commands, and that it simply cannot do so perfectly due to ambiguity in afferent and efferent signals. Other models have different ideas on the nature of this optimization. Flash and Hogan (1985) explain observed movement smoothness with a framework that is based around the need to minimize mean-squared jerk. Meyer et al., 1988 applied the assumption that errors may be the effect of neuromotor noise introduced to each of a series of corrective submovements, and that such noise may be modeled stochastically, and produced an optimal submovement model. Fitts actually mentioned this possibility in his 1954 paper, instead opted for macro-level signal-based description. Harris and Wolpert (1998) developed the minimum variance (MV) model, which postulates that the brain acts to minimize the projected endpoint variance either by simulating the desired movement and selecting the best motor program, or by looking the desired movement up in a large, experience-derived internal store. Tanaka et al (2006) developed a revised version of Harris and Wolpert’s minimum variance model. It incorporates the assumption that subjects actually satisﬁce when it comes to allowable error, and thus select the program that can be executed in the shortest time within a certain error threshold, instead of trying to minimize endpoint variance entirely. The important difference between the MV model and the TOPS framework is what the framework considers the cost to be of a motor program. TOPS considers predicted movement error due to signal noise in the motor commands to be the cost of a motor program, while the revised MV model consider time taken within an allowable threshold— to be the cost. The satisﬁcing MV model predicts the data of time-constrained Fitts tasks such as that used by Fitts and Peterson (1964). Adoption of the satisﬁced minimum variance paradigm leads to the conclusion that the 4% error rate generally observed in pointing tasks is due to a deliberate choice of motor program that allows for time optimization given a certain error tolerance.

What Influences Error Distribution?

While axis differences in error have been noted in the past, they have not historically been subject to concerted analysis. Wobbrock et al. (2011b) noted that the distribution of points tends to vary more widely along the on-axis than the off-axis. Gordon et al. (1994) had subjects perform a bivariate pointing task on a tablet, and noted that on-axis error was generally larger than off-axis error. The investigators observed that error clouds tended to be elliptical, with the main axis aligned along the axis of movement and eccentricity decreasing with distance. However, in this task the subjects could not see the cursor during the movement, and as such subjects were not able to make course corrections (where costly, high-magnitude corrective motor commands could be enacted). Thus, Gordon et al. attributed endpoint variability to errors in task planning as opposed to noise introduced during iterative course correction. Another potential component in endpoint variance distribution is approach angle. Wobbrock et al. (2011b) makes the claim that angle effects on error distribution and percentages are insignificant, citing ISO 9241-11 (1998). However, both Van Beers et al. (2002) and Harris and Wolpert (1998) provide model-based reasons why angle may have notable effects. Both TOPS and the minimizing/satisficing variants of the minimum variability model do not actually predict straight movement lines as being optimal when in reducing corrective motor command magnitude; instead a curved approach is predicted. However, in all cases of feedback-corrected movements where angle effects were studied, angle has been found to be insignificant (Gordon et al., 1994). However, Van Beers (2004) had subjects perform a bivariate pointing task that did not allow for mid-course corrective feedback, and noted significant variance in endpoint distribution ellipsoid properties as a function of angle. The aspect ratio and main axis orientation for 95% confidence interval ellipses varied significantly with direction, using Rayleigh’s test for directional data. Additionally, aspect ratio/ eccentricity tended to be higher in the 22-45 degree range, and again in the 225-270 degree range. Using a model derived from the TOPS framework, Van Beers et al. achieved a good model fit for these observations by incorporating signal dependent noise, constant noise, and temporal noise. The investigators also noted that trajectories- both observed and predicted- actually tended to be slightly but significantly curved, and that the endpoint cloud actually tends to orient it itself along the last part of the approach line, especially at longer amplitudes. It should be noted that this study did not allow for corrective movements; however, it still provides evidence for angle effect in the planning stage of a pointing movement. Harris and Wolpert (1998) also predicted curved approach paths using the basic minimum variance model. Both the TOPS and MV models explain the gentle curvature as a consequence of providing minor course corrections (either from visual or proprioceptive feedback, depending on the task) to an initially flawed command signal, since providing drastic corrections would introduce a high volume of signal dependent motor noise. Accordingly, the present study was aimed at identifying angle effects on endpoint distribution, in addition to width/distance effects.

Existing Error Models

Recently, an error model specific to pointing tasks was derived by Wobbrock et al. (2008) with the classical speed-accuracy optimization in mind. In order to create this model, the investigators first noted that in order for the speed-accuracy tradeoff predicted by Fitts’s law (which is obtained by solving the ID equation for target width) to hold true, the error rate must be about 4%, and that on-axis errors should be Gaussian in their distribution. In this context, errors are points that are outside of +- (W/2). The investigators then de-
derived an equation for the z score associated with a 4% error rate, and multiplied this by
\[
\frac{W(\text{predicted})}{W(\text{observed})}
\]
(10)
They next inserted the speed/accuracy equation for W(observed), thus obtaining the z score of the error threshold in terms of variables used by Fitts’s law. Subsequently, they integrated the standard function of a normal curve in terms of this z score, and took one minus this value to be the error range. Finally, they substituted the z score equation into this normal integral, thus obtaining a function that predicts the area outside the normal curve for a given set of parameters. The resulting equation, which is particularly sensitive to constants \(a\) and \(b\), describes the probability of a given point being in error in the following way:
\[
P(E) = 1 - \text{erf}\left(\frac{2.066 \times W}{\sqrt{2}} \times \left(\frac{\text{MT}_{\text{bs}} - a}{b} - 1\right)\right)
\]
(11)
Here, \(\text{erf}\) refers to the Gauss error function. The investigators noted that if predicted movement time equals observed movement time, then this equation predicts a 3.88% error rate, which is close to the expected 4%. The investigators found that the equation had good predictive power in a one dimensional metronome-mediated tapping task.

Wobbrock at al. (2011a) applied this model to a bivariate pointing task, and found that the model fit the observed data fairly well using both aggregated bivariate error (\(r^2 = .948\)), and univariate on-axis error (\(r^2 = .944\)). In this context, an ‘error’ was a point that fell more than \(W/2\) in any direction from the target. However, this error model has not yet been evaluated with regards to its ability to predict on and off-axis error separately.

**Systematic Undershoot and Energy Consumption**

All of the models discussed thusfar predict an error cloud normally distributed about a target centroid, either with a flat 4% rate or a rate mediated by other parameters. There is actually evidence that brain performs automated strategic management of the probable direction of error. This paradigm provides an alternative to the idea that the neuromotor system is simply trying to minimize the magnitude of noise, but some of the same principles apply. Engelbrecht et al., 2003 reviewed literature on manual aiming tasks and noted that endpoints along the primary movement axis were more likely to fall short of the target. Engelbrecht et al. also performed an experiment in which subjects were asked to perform a pointing task during various levels of stochastic noise, and observed that subjects tended to undershoot by a default amount under each noise condition. Over time, subjects were able to adapt to the level of noise- employing smaller strategically implemented undershoot for low-noise conditions and larger undershoot for high-noise conditions. This was primarily achieved through varying the magnitude of the initial control command. The investigators proposed two explanations for why undershooting is optimal in conditions of high or uncertain noise. The first is that the control command to adjust the gain of the movement should have a smaller magnitude (and thereby less cost in terms of introduced signal noise) than the command to change direction entirely and return to the target after an overshoot. The second is that the brain is making an effort to reduce the effort expended in the movement, and that the act of countering the inertia of a moving pointing device in order to move it back into the target area is extremely high; therefore, scenarios where this is necessary are avoided by the motor system.

Harris (1995) used Monte-Carlo simulations to generate a model of optimal primary motor command gain based off of the level of stochasticity, given the goal of minimizing eye saccade time in order to maximize clear vision time. Given extremely error-prone movements of the sort found in infants, the model predicted an optimal gain of .6 (meaning the initial command would take the eyes about 60% of the desired total travel distance), while for an adult gain was predicted to be around .95 (Harris, 1995). These predictions agree with empirical data on saccadic undershoot. Berthier (1996) extended the minimum-jerk model to infant reaching movements, and made very similar predictions. The model assumes that infants are unsure what their current motor capabilities are since they are changing so rapidly (of particular note here is the level of stochasticity of movements), and must plan and execute movements without this knowledge. The model predicted an optimal (in terms of jerk minimization) initial command gain of .5 for high-stochasticity movements, and a gain close to 1 when stochasticity is low. This model was fitted to infant pointing data; differences in predicted and observed mean and standard deviation for error distributions were not significant. Hansen et al. (2006) found evidence that performers who are forewarned of the presence of visual feedback as to the accuracy of a pointing movement spent significantly less time in the planning stage (and, accordingly, more on the actual movement stage). In addition, they actually spent most of their ‘movement time’ after peak movement velocity has been achieved. This support the supposition that the general behavioral pattern for aimed pointing is to quickly get close to the target with a high-magnitude, noisy command, then use the bulk of the movement time to perform minor corrections. This sort of a methodology would support the notion that overshoots are costly, since an overshoot would cause this near-target correction time -which is actually the majority of the movement time- to be significantly larger.

Of particular note to the present study is the predictions made by these models regarding task learning. When stochasticity is uncertain, it may become strategically advantageous to ‘play it safe’ and lower initial movement gain, thus increasing undershoot. Even in adults, who presumably are familiar with their motor systems, for a given task several parameters always begin with some amount of uncertainty: the capabilities of the device, the operator’s ability to manipulate the device, etc. As such, one might expect the undershoot to start off around a certain threshold and then adjust based off of the throughput of the device. Within each learning level, there should also be a linear increase in undershoot with movement amplitude, since systematic undershoot is likely due to the ap-
plication of a linear gain function. Elliott et al. (2009) found evidence that operators tend to modulate behavior over time in order to optimize energy expenditure. Of particular note is the prediction that energy expenditure will be reduced; traditionally, in tasks to which Fitts’s Law is applied, accuracy remains relatively constant and, since speed is generally constrained by desired accuracy, it remains constant as well Elliott et al. (2009).

There is also reason to believe that the magnitude of undershoot bias might be influenced by the energy costs associated with correcting for an overshoot. Elliott et al. (2004) crafted an experiment in which subjects had to complete two kind of pointing tasks: one in which it was more energy efficient to overshoot the target (unassisted movement) and one in which it was more efficient to undershoot the target (assisted condition), and found that the assisted condition produced an overshoot bias, while the unassisted condition produced an undershoot bias, lending credence to the theory that the desire to minimize unnecessary energy expenditure accounts for at least part of the observed on-axis error. Lyons et al.(2006) had subjects perform a three-dimensional pointing task, and found that the magnitude of initial target undershoot was significantly greater for conditions in which subjects were reaching below home position than reaching above home position. The investigators explained this observation with the supposition that subjects were sensitive to the fact that correcting for an overshoot against the pull of gravity (in the 'reaching down' conditions) was more costly than correcting for an undershoot with the pull of gravity (as in the 'reaching up' conditions).

STUDY

The first goal of the present study was both to test whether error varies systematically by axis, and to disentangle some of the varied factors that could be effecting the extent and direction of such variation. Subsequently, we built a simple predictive model of endpoint distribution that was sensitive to relevant task features, and strove to provide an explanation for why such a model might work and what this might mean for HCI.

EXPERIMENT 1

Participants

Participants were 25 Rice University undergraduates who were familiar with the use of a mouse.

Apparatus

We used a 400 dots-per-inch Logitech mouse. The screen used was a 13x9 inch, 1024x768 LCD display (80.95 pixels per inch).

Procedure

We used an experimental paradigm similar to that described in ISO 9241-9. Subjects were presented with a 3 second fixation cross, and then tasked with moving a mouse cursor from the center of the screen to click on a circular target presented somewhere onscreen for an unlimited amount of time. We had subjects re-center the mouse after every pointing task, and there was no metronome or back-and-forth movement of any sort. This is in contrast with several earlier described studies of aimed pointing to which Fitts’s law has been applied. In addition, subjects were not instructed to optimize for time or accuracy; they were simply told to click on the targets as they appeared. Subjects were not instructed to hit the exact center of the targets; they were simply told to hit them. Subjects were given no feedback as to whether they had hit a target or not, and a new target was displayed either way. Targets were presented as blue circles at distances of 100, 200, and 300 pixels from the center of the screen. Target angle varied between 0, 45, 90,130, 180, 225, and 270 degrees. Target size varied between 43, 50, 57 and 64 pixels in diameter. Subjects sat about 24 inches away from the screen, and held the mouse their preferred hand on a table beneath and to the side of the screen. With this setup, the target sizes presented approximately 1.32, 1.54, 1.76, and 1.97 minutes of visual angle. Each of these conditions was replicated three times in random, nonsequential fashion.

Results

In accordance with Fitts’s law, observed movement time increased linearly with ID. There was a clear main effect of target size on MT (F = 12.132, P = .000) and of target diameter on MT (F = 14.202, P = .000). All conclusions were drawn and stated using a Greenhouse-Geisser corrected Sphericity Assumption test to assess variance within subjects unless otherwise noted. The formulation for Fitts’s law used here was equation (7) from Mackenzie and Buxton (1992), with the diameter of the target treated as W. From equation 7, the average intercept constant was 12.5 Ms, and the average slope constant was 311.81. These constants were initially calculated on a per-subject, per-condition basis, and then averaged across subjects within each condition. Effects of target diameter (F = 1.859, P = .154) and distance (F = 2.187, P = .140) on hit percentage were not significant, which indicated that the speed-accuracy tradeoff being made was essentially the same across conditions. The raw (nondirectional) magnitude of observed error increased linearly with width (F(2,24) = 7.75, p = .001).

![Figure 1. Error Magnitude vs. Target Diameter](image)

Raw error magnitude also varied significantly by distance (F(3,24) = 27.93, p = .000).
Of particular interest to the present study was the finding that average error differed significantly with axis ($F(1,24) = 13.21$, $p = .001$).

On-axis error was not only consistently larger than off-axis error; it was also consistently negative, providing support for an undershoot bias. The magnitude of the undershoot preference varied significantly with target size ($F(2,24) = 10.678$, $p = .000$) and target distance ($F(2,25) = 9.250$, $p = .000$). Larger target sizes and longer distances led to larger undershoots; however, target width had a notably larger effect on the mean undershoot magnitude. There was also evidence for an effect of target angle ($F(7,25) = 2.69$, $p = .011$), but it was unclear what the nature of this effect might be.

Figures (4) and (5) are plots of observed movement endpoints for the smallest and largest targets. Of note is the observation that with the larger target size, the distributions tend to be more elliptical in shape, and tend to center closer to the inner rim of the target than with the smaller target size. For our next experiment, we wanted to see if this effect held even larger and smaller targets.

*Model Development*
The next step in our analysis was to come up with a simple model of the overall frequency distribution of endpoint error. This was not intended to be a procedural model of aimed pointing, nor was it necessarily expected to work under conditions of constrained time or accuracy. The model we endeavored to build was crafted with the goal of being as simple and predictive as possible, and to provide a potential guide for future procedural models of pointing. In order to develop the model, we first made the observation that the on and off axis error distributions to be normally distributed along each axis. We thus began with the equation for a normal distribution:

$$\frac{1}{2\pi\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$ (12)

Since the off-axis error was observed to center approximately at zero, we set $\mu$ to zero. For the on-axis error, we observed that the mean of the frequency distribution would shift in accordance with the change in the mean directional error (magnitude of overall undershoot). Thus, $\mu$ was set equal to the result of a multivariate regression performed on the two clearly significant factors of target size and target distance, yielding the following equation for $\mu$:

$$\mu = -.172W - .007D + 7.801$$ (13)

This produced the following as a model for on-axis error:

$$\frac{1}{2\pi\sigma^2} e^{-\frac{(x-(-.172W-.007D+7.801))^2}{2\sigma^2}}$$ (14)

Of note is the constant term, which can be seen as predicting that an extremely close or extremely small target will be overshot by some amount, which is consistent with conventional literature on Fitts’s law MacKenzie (1992). Also of interest is the relative insignificance of the distance term. Part of the impetus behind performing a second part to this study was to find out if the distance term was really necessary.

We provide several methods for generating the standard deviation term $\sigma$, depending on the amount and nature of the subject data available. If a decent amount of subject data is available (enough to get a meaningful observed error rate), the following equations can be used to calculate the standard deviation of the error distribution, if one wishes to assume a 4% error rate:

$$W_e = W - \frac{2.066}{Z(accuracy)}$$ (15)

$$\sigma = \frac{W_e}{4.133}$$ (16)

These relationships come from Soukeroff & MacKenzie (2004), page 756. Both equations derive from the idea that the error rate should be about 4%, leading to the conclusion that the effective width can be computed as 4.133 multiplied by the standard deviation of observations. If this observation is true (it is often close to being accurate), then $\sigma$ can be calculated through equation (12). This is the calculation recommended by ISO 9241-9. This equation requires the effective width, which can be calculated with the following equation, where $MT_e$ is the movement time predicted by Fitts’s law: (Wobbrock et al., 2008).

$$W_e = \frac{D}{2^{MT_e/a} - 1}$$ (17)

If the investigator has reason to believe that the error rate may be far from 4%, the full model proposed by Wobbrock et al. (2008) can be applied. Given $W_e$ from (13), the expected accuracy can be calculated to be:

$$P(H) = \text{erf}\left(\frac{2.066 W_e}{\sqrt{2}}\right)$$ (18)

It is important to know that if the investigator knows constants $a$ and $b$, equations (13) and (14) can be used in conjunction with the $MT_e$ given by a formulation of Fitts’s law (6). There are forseeable scenarios in which $a$ and $b$ are available yet the observed hit rate and movement time are not. Ideally, $a$ and $b$ would be available in a general database, and could be applied in the manner described here. Target width and distance should always be available, at the very least.

However, in the interest of simplicity, a third version of the standard deviation calculation exists that employs three assumptions: first, that the effective target width will be the same as the actual target width; second, that the observed movement time would be the same as the expected movement time predicted by Fitt’s law; third, that the error rate will be around the commonly expected 4%. This is appropriate in scenarios in which there is no observational data, no availability of constants $a$ and $b$. Essentially, this version is useful when the overall error distribution model is being used either to predict behavior in an unknown pointing environment or to inform the construction of a methodological pointing model. In these scenarios, the equation for effective width collapses down to (15):

$$\sigma = \frac{W}{4.133}$$ (19)

While we strove to provide the model with the ability to adapt to the scenario, one question raised by the initial model construction was whether the more complex methods of calculating $\sigma$ were actually necessary, especially given relatively common task conditions.

**Model Fit**

The model fit fairly well across conditions, both using constants derived from the data observed in experiment 1 and without such constants. Before performing a more comprehensive analysis of model quality, we decided to perform an-
other experiment in order to determine with greater certainty whether width and distance were indeed the right factors to be including, and whether using the simpler method of calculating the standard deviation would allow the model to retain its validity under more extreme target conditions.

**EXPERIMENT 2**

**Participants**
Participants (N=27) were taken from the pool of Rice University undergraduates who had not participated in previous pointing studies.

**Apparatus**
In order to test the effect of mouse DPI, and the associated change in overshoot cost, we used two Cooler Master Storm Spawn adjustable DPI mice. DPI was a between subjects variable, with N=15 using the mouse set to 800 DPI and N = 12 using the mouse set to 1800 DPI. These DPI settings were somewhat high for most participants; many verbally commented on the sensitivity during initial trials. This factor, along with the fact that the mice were irregularly shaped (being made for ‘claw grip’ computer gaming) was meant to serve the dual purpose of accentuating any learning effects observed during the increased repetitions of each condition.

**Procedure**
The testing method used for experiment 2 was almost identical to that used for experiment 1. However, this time we used three target diameters (30 pixels, 50 pixels, and 70 pixels), four target distances (83 pixels, 165 pixels, 247 pixels, and 330 pixels), the same eight angles, and five trials per condition.

**Results**
Results were similar to the previous study, with a few important exceptions.

**Error Frequency**

First, there was a significant effect of target size on error rate (F(2,26) = 75.54, p = .000). There was also a significant effect of trial number on hit rate (F(4,26) = 3.76, p = .007). These findings are interesting for several reasons. First, it appears that for larger target sizes, the accuracy approaches that predicted by Fitts’s law (without any corrections for expected width). However, for smaller targets the accuracy starts out significantly lower (81% for the smallest target size, and around 88% for the middle size) and seems to level out after the third trial. This indicates that the more complex methods for predicting the standard deviation of the endpoint distribution might be preferable, since the assumption of a 4% error rate is not necessarily valid for all conditions. This also indicates that the rate of error is indeed somewhat more complicated and variant than that predicted by Fitts’s law, and that something like the model proposed by Wobbrock et. al. (2008) is indeed preferable if the necessary information is available.

**Absolute Error Magnitude**
Non-directional error magnitude increased significantly with target distance (F(3,26) = 19.10, p = .000) and target size (F(2,26) = 170.65, p = .000). Larger targets had larger average endpoint deviations, presumably due to their larger error tolerances. Interestingly, the effect of trial number on average error was not significant (F(4,26) = 1.52, p = .2), at least over all five trials. It appears that the bulk of task acquisition has occurred by trial 3–4, indicating that data from trial 5 is perhaps a suitable set for describing ‘expert’ behavior.

**Movement Time**
Observed movement time decreased significantly with trial number (F(4,26) = 9.9, p = .000). Again, the effect appears to level off after the 4th trial, indicating that this may be the
amount of practice needed to optimize behavior, or at least reach a local plateau. There was also a significant size interaction effect with trial number on observed movement time $F(8,26) = 2.82, p = .006$. Interestingly, it appears that the learning curve may level out faster for smaller targets than for larger ones. The reason for this is apparent through an analysis of axis-specific error.

**Axis Error**

First, the average error once again varied significantly by axis ($F(1,26) = 7.22, p = .013$). This is in accordance with the observations from experiment 1, and provides further validation for the basic premise of the error model.

Second, the magnitude of the mean undershoot varies significantly with target size ($F(2,26) = 41.92, p = .000$), while the magnitude of off-axis trajectory error (which hovers around the target axis) does not vary significantly with target size. A visualization of these trends can be seen in figures (15) and (16), which are plots of the endpoint scatters at the smallest and largest target diameters. Once again, it should be visually apparent that error clouds are oblong along the target axis, and that this effect is larger for the larger target diameter. Additionally, it should be apparent that the distributions
are centered closer to the inner target rim for the larger target sizes.

Figure 15. Endpoint Distribution Plot for Diameter =30. Blue circles are clicks, green circles are targets, and red icons are 95% confidence ellipses about a centroid.

Figure 16. Endpoint Distribution Plot for Diameter =70. Blue circles are clicks, green circles are targets, and red icons are 95% confidence ellipses about a centroid.

The magnitude of these trends appears to be sensitive to task learning. Axis error was significantly effected by trial number (F(2,26) = 19.10, p = .000, Figure 19). Interestingly, there appear to be two optimizations going on here. Lateral trajectory variance is being reduced over repeated trials, while on-axis undershoot is actually increasing. Presumably, given some number of further trials, the latter trend would level off as the inner edge of the target was approached. What this indicates is that subjects may be producing not only an effective width, but also an effective target position. This latter item is what the distribution mean \( \mu \) of the model is describing. The following plots provide visual indication of this trend. In (20), at Trial = 1, subjects appear to be operating either at a slight ‘default’ undershoot or perform a preliminary adaptation to target size. In (21), at Trial = 5, subjects hit consistently closer to the inner rim of the allowable target area. This explains the reduced movement times, even though accuracy remains relatively constant (indicating that the speed-accuracy tradeoff being made has not changed). Figure 18 shows a plot of on-axis error. The actual size of the interquartile range does not change with increasing trials, but the position of said range shifts to become more negative. The effect was accentuated by large target sizes, whereas at small target sizes was less apparent.

Conversely, a plot of off-axis error across target sizes and trials shows that the median error gets closer to zero over time. However, the interquartile range does not change significantly. This leads to the conclusion that the increase in off-axis accuracy observed over time is likely due either to making a systematic modification in the trajectory target or in a reduction of the number of extremely wide misses. Importantly, both plots provide evidence that there is not a change in the speed-accuracy tradeoff described by Fitts’s law, and that the observed increase in task performance -specifically, accuracy and movement time- over time must come from overall distribution shifts as opposed to improvements in motor behavior or reductions in noise.

While robust effects of size and trial number were noted on the mean axis-specific error, effects of distance and angle were not significant. As such, it was determined that these elements should not be included in the final version of the model.
Revising the Model

Due to the evidence that undershoot optimization continues up until at least the fourth or fifth trial, we decided to use only the data from the fifth trial to construct the model. This is due to the fact that Fitts’s law generally models expert pointing behavior. We considered the possibility of including a modifier to represent the amount of history for the particular task, but decided against it on the grounds that such would be difficult to consistently code and quantify. In addition, since target distance was not a statistically significant factor with regards to directional error magnitude, the distance term was excluded. This allowed us to employ a simple linear regression on the trial 5 on-axis mean undershoots by the trial 5 target sizes, in order to produce the following for $\mu$:

$$\mu = -0.163W + 5.174$$  \hspace{1cm} (20)

For the second iteration of the model we decided not to use the more complicated methods of predicting sigma (although such methods are still available for use). In addition, a correction of .75 was added to the $\sigma$ calculation for the off-axis frequency distribution in order to reflect the reduced variance of off-axis endpoints, especially after a number of trials. It is unknown whether this value would, in fact, be lower given a larger number of trials. Future work could investigate this. Thus, using this correction and the simplified form of the $\sigma$ calculation, we produced the following equations for on-axis and off-axis endpoint frequency, respectively:

$$\frac{1}{2\pi \frac{W}{4.133}} e^{-\frac{(x-\mu)^2}{W^2}}$$  \hspace{1cm} (21)

$$\frac{1}{2\pi \frac{0.75W}{4.133}} e^{-\frac{(x-\mu)^2}{(0.75W)^2}}$$  \hspace{1cm} (22)

Model Fit

The second model was tested against the full set of trial 5 target sizes, as well as the trial 3 target sizes from experiment 1.
The fits achieved were generally very good, especially considering that the generalized standard deviation calculation was used. The following plots illustrate the quality of the fit for on-axis error distribution under each of the three target diameters used in experiment 2. Observed data is in red, while the model is presented in green.

Figure 23. Predicted (green) vs. Observed (red) On-Axis Error (70 px Diameter) ($r^2 = .9942, p= .000$).

Figure 24. Predicted (green) vs. Observed (red) On-Axis Error (50 px Diameter) ($r^2 = .9979, p= .000$).

Figure 25. Predicted (green) vs. Observed (red) On-Axis Error (30 px Diameter) ($r^2 = .9887, p= .000$).

Our model of off-axis error fit similarly well (26) (27) (28).

Finally, we tested the model on the data set from the first experiment in order to gauge how well it would generalize to other conditions, and whether anything significant had been lost by excluding the distance term. This produced good fits overall, as seen in figures (29), (30), (31), and (32), which represent the model fit for target diameters 43, 50, 57, and 64, respectively.

Figure 26. Predicted (green) vs. Observed (red) Off-Axis Error (70 px Diameter) ($r^2 = .9908, p= .000$).

Figure 27. Predicted (green) vs. Observed (red) Off-Axis Error (50 px Diameter) ($r^2 = .9992, p= .000$).

Figure 28. Predicted (green) vs. Observed (red) Off-Axis Error (30 px Diameter) ($r^2 = .9892, p= .000$).

Figure 29. Predicted (green) vs. Observed (red) On-Axis Error (43 px Diameter) ($r^2 = .9887, p= .000$).
Figure 30. Predicted (green) vs. Observed (red) On-Axis Error (50 px Diameter) ($r^2 = .9910$, $p = .000$).

Figure 31. Predicted (green) vs. Observed (red) On-Axis Error (57 px Diameter) ($r^2 = .9802$, $p = .000$).

Figure 32. Predicted (green) vs. Observed (red) On-Axis Error (64 px Diameter) ($r^2 = .9910$, $p = .000$).
DISCUSSION

The preceding work provides clear and repeatable evidence for systematic difference in axis-specific error magnitude. The observed undershoot bias fits the predictions of the earlier studies to some extent; however, the lack of a consistently significant effect of distance coupled with the lack of an effect of task DPI indicates that further research may be necessary to investigate the strategic nature of systematic undershoot. Participants as a whole were remarkably adaptive to the specifics of the task environment, even one as simple and commonplace as using a mouse to click on targets. This has interesting implications for the employment of novel devices, as it implies that there is not full skill transfer in such tasks—perhaps due to technical details of the device, such as dpi, and perhaps due to nuances of the task. Users appear to begin with either a small degree of undershoot or a slight overshoot, depending on the target size (the effect of distance might bear further investigation). Subsequently, they adapt their behavior to target sizes, and move their conceptual endpoint closer to the inner edge of the target while at the same time increasing off-axis precision. This effect appears to level off around trials 4 and 5, indicating that What this means for the science and practice of HCI is that models of task performance applied to an interface can work off of the assumption that users will strategically manage their noise-induced error both initially and as they gain experience with a task, a device, and the way their own perceptual-motor loop handles interaction with a specific environment. The model constructed here is simple, but works well for a fair range of target parameters, and fills

FUTURE WORK

The next step is the development of a procedural model of pointing that will fit the behavioral pattern implied by the observations made by the present study. In addition, while the present model works well, it is very possible that other factors may influence the error distribution given additional task constraints or instructional changes with regards to things like speed or accuracy. It is also likely that scenarios will arise in which the more complex formulations of $\sigma$ become necessary. Finding out exactly what such scenarios are will help researchers and interface designers to understand how best to apply the present model.

ACKNOWLEDGEMENTS

We’d like to thank Professor Marcia O’Malley, students Haley Lindsay, Melissa Gallagher, Nicole Howie, and the rest of the CHIL lab for their help with the study.

References


