

# Effects of Motor Scale, Visual Scale and Quantization on Small Target Acquisition Difficulty

OLIVIER CHAPUIS, LRI - Univ Paris-Sud & CNRS; INRIA  
PIERRE DRAGICEVIC, INRIA

13

Targets of only a few pixels are notoriously difficult to acquire. Despite many attempts at facilitating pointing, the reasons for this difficulty are poorly understood. We confirm a strong departure from Fitts' Law for small target acquisition using a mouse and investigate three potential sources of problems: motor accuracy, legibility, and quantization. We find that quantization is not a problem, but both motor and visual sizes are limiting factors. This suggests that small targets should be magnified in both motor and visual space to facilitate pointing. Since performance degrades exponentially as targets get very small, we further advocate the exploration of uniform, target-agnostic magnification strategies. We also confirm Welford's 1969 proposal that motor inaccuracy can be modeled by subtracting a "tremor constant" from target size. We argue for the adoption of this model, rather than Fitts' law, when reflecting on small target acquisition.

Categories and Subject Descriptors: H.5.2 [Information Interfaces and Presentation]: User Interfaces

General Terms: Human Factors, Experimentation, Performance, Theory

Additional Key Words and Phrases: Fitts' law, small targets, quantization, C-D gain, pointing techniques, target-agnostic.

## ACM Reference Format:

CHAPUIS, O., and DRAGICEVIC, P. 2011. Effects of Motor Scale, Visual Scale and Quantization on Small Target Acquisition Difficulty. *ACM Trans. Comput.-Hum. Interact.* 18, 3, Article 13 (July 2011), 32 pages. DOI = 10.1145/1993060.1993063 <http://doi.acm.org/10.1145/1993060.1993063>

## 1. INTRODUCTION

Common computer tasks such as window resizing and text selection require pointing at items of only a few pixels. In mobile devices, small items are also common and notoriously difficult to acquire. Such problems have repeatedly been mentioned in the past, and a variety of approaches have been proposed to ease the acquisition of small targets [Kabbash and Buxton 1995; Worden et al. 1997; Ren and Moriya 2000; Albinsson and Zhai 2003; Cockburn and Firth 2003; Cockburn and Brewster 2005; Ramos et al. 2007; Chapuis and Roussel 2010]. Each of them has advantages and drawbacks and new solutions are still actively being explored. However, surprisingly little is known about the reasons why very small targets are so difficult to acquire. Answering that question can not only better guide future research and inform future designs, but can also help choosing among existing approaches and refining them.

Since Fitts' law [Fitts 1954] is routinely used to motivate and inform research on pointing facilitation techniques, we first discuss its suitability as a conceptual framework in the context of small

---

Authors' addresses: LRI Bat. 650, Univ Paris-Sud, F-91405 Orsay, France; email: chapuis, dragice@lri.fr.

Authors version of: Olivier Chapuis & Pierre Dragicevic, Effects of Motor Scale, Visual Scale and Quantization on Small Target Acquisition Difficulty. *ACM Trans. Comput.-Hum. Interact.* 18, 3, 13:1 – 13:32 2011. <http://doi.acm.org/10.1145/1993060.1993063>

target acquisition. We also show why taken together, previous studies on pointing do not provide a clear picture of why small targets are difficult to acquire.

We then present a study in which three potential sources of problems are investigated in mouse pointing: motor accuracy, legibility and quantization (i.e., the fact that the mouse cursor moves by single-pixel increments). Most small target acquisition tasks considered in the Human-Computer Interaction literature combine at least these three problems, and we show how new insights can be gained from a separate assessment of these three factors.

### 1.1. Fitts' Law and Small Target Acquisition Techniques

In addition to being a powerful predictive theory, Fitts' law has often been employed in HCI as a prescriptive and generative theory in order to guide, motivate and justify designs [Kabbash and Buxton 1995; Worden et al. 1997; Zhai et al. 2000; Balakrishnan 2004; Blanch et al. 2004; Grossman and Balakrishnan 2005; Kobayashi and Igarashi 2008]. However, it is also understood that Fitts' Law only holds within certain limits [MacKenzie 1992; Plamondon and Alimi 1997; Accot and Zhai 2001; Guiard 2009].

Fitts' Law has often proved to be an extremely good predictor of pointing performance under a wide range of conditions and one could question the need for more complex models [Van Galen and De Jong 1995]. In this section we discuss why Fitts' law is actually not an adequate paradigm for justifying and guiding the design of small target acquisition techniques, even when used as a first approximation. We then list the benefits one can expect from more systematic investigations into that topic.

*Fitts' Law is not an Adequate Model for Small Target Acquisition Techniques.* It seems natural to appeal to Fitts' law to motivate a new small target acquisition technique. Fitts' law does predict that small targets are harder to acquire than larger ones. But Fitts' law is first of all a law about *scale invariance*, stating that acquisition time depends solely on  $D/W$ , the ratio between target distance ( $D$ ) and target width ( $W$ ).  $D$  is hence as important as  $W$  and nothing in Fitts' law justifies a particular focus on target size.

Fitts' law additionally predicts that acquisition time increases approximately logarithmically with  $D/W$ , which implies that increasing target size should not have a particularly strong impact on movement time. For example, if  $W = 1$  and  $D = 64$ , the target theoretically needs to be expanded by 9 to yield a movement twice as fast<sup>1</sup>. Conversely, shrinking a target should not strongly penalize movement time.

However, a number of observations suggest that this is not the case when targets are already small. Performance seems to degrade very rapidly when the target size falls below a certain threshold, typically below 4 or 5 pixels [Ren and Moriya 2000; Albinsson and Zhai 2003; Ramos et al. 2007]. Sometimes, acquisition time and error rate explode: in one study [Ramos et al. 2007], users missed a 1-pixel target more than ten times in a row in 25% of the trials.

Such observations suggest a strong scale effect by which small-scale pointing tasks are more difficult than large-scale ones. It is this observed violation of Fitts' law – and not Fitts' law itself – which best justifies the design of small target acquisition techniques.

*Fitts' Law does not Provide Adequate Guidelines for Small Target Acquisition Techniques.* When used as a prescriptive and generative model for designing pointing facilitation techniques, Fitts' Law exclusively advocates the use of *target-aware* pointing facilitation approaches and is not able to discriminate between them. We explain why, and why this is problematic.

---

<sup>1</sup>This assumes the Shannon formulation [Mackenzie 1991]. In the case of a non-negligible positive intercept  $a$ , the target expansion required can be much larger.

Among the many possible approaches for facilitating pointing, there has been a large amount of research on target-aware pointing techniques, i.e., techniques that use a-priori knowledge about potential targets. Most of these techniques reduce  $D/W$  by manipulating target widths and/or distances in the motor and/or the screen space [Balakrishnan 2004]. Some of them, like sticky targets [Cockburn and Firth 2003; Cockburn and Brewster 2005] and area cursors [Kabbash and Buxton 1995; Worden et al. 1997; Chapuis et al. 2009], have been specifically proposed as a solution to the small target problem.

Although Fitts' law provides theoretical support for the idea of reducing  $D/W$  (by decreasing  $D$ , increasing  $W$ , or both), it does not adequately capture the problem with small targets. First of all, Fitts' Law might actually underestimate the improvement obtained when expanding very small targets. Recall that it only predicts a weak effect of target expansion on movement time, no matter the original target size. It does not say anything about the effects of expanding a target above a "critical threshold". Similarly, Fitts' law is indifferent to the method used to reduce  $D/W$ . It does not predict that reducing  $D$  might not be as efficient as increasing  $W$ , in case  $W$  is very small.

Finally, according to Fitts' Law, the only way to reduce movement time is to reduce  $D/W$ , i.e., to adopt a target-aware strategy. But these strategies are limited [Balakrishnan 2004]: they essentially consist in redistributing targets to have their  $D/W$  ratios better match their probability of acquisition — a special case being removing intervening spaces with zero click probability [Guiard et al. 2004]. But often there is little or no space to remove, e.g., when using a color picker in an image authoring application.

Simple target-agnostic approaches are nonetheless possible that might facilitate acquisition despite leaving  $D/W$  unchanged. They consist in increasing all  $W$ 's and  $D$ 's in the same proportions, i.e., a uniform *magnification* solution. As in target-aware techniques, magnification can be performed in the screen space, in the motor space, or in both. Examples of screen-space magnification techniques are software screen magnifiers [Kline and Glinert 1995] and fisheye views [Furnas 1986]. Lowering the control/display (C-D) gain yields a uniform motor-space magnification [Casiez et al. 2008]. Zooming magnifies both screen and motor space [Bederson et al. 1996; Ramos et al. 2007; Roudaut et al. 2008]. But since they all leave  $D/W$  unchanged, none of these techniques can be adequately motivated or designed using Fitts' Law as a conceptual framework.

## 1.2. Key Questions

We argue that the following three questions regarding small targets are worth investigating due to their potential implications for design:

- (1) *Is there any small-scale effect?* That is, when  $D/W$  is fixed, are tasks where  $W$  is very small more difficult? If the answer is no, it would imply that any target-aware approach could be the way to go. If it is true, it would imply that a) among the possible target-aware strategies, expanding  $W$  should work better than reducing  $D$  and b) target-agnostic magnification strategies deserve more attention.
- (2) *What are the causes of the small-scale effect?* Assuming there is a small-scale effect, it is still unclear whether target expansion should be performed in the screen space, in the motor space, or both. Answering these questions requires identifying, isolating and testing the potential sources of difficulty in small target acquisition tasks.
- (3) *What would be a good model of small target acquisition?* Assuming effective pointing facilitation strategies can be identified, it is still unclear how to tune them. For example, will a magnification of 4x be enough? With such a magnification, what gain in performance can be expected? Being able to predict user performance with small target acquisition tasks will help answering these questions.

With respect to our second question above, we investigate three potential sources of problem with small targets:

- (a) *Motor scale*: Targets that are small in motor space could be difficult to acquire because they demand too much motor precision.
- (b) *Visual scale*: Targets that are small in screen space are difficult to see, which could make their acquisition more challenging.
- (c) *Quantization*: By quantization we refer to the motion of the mouse pointer by discrete (usually single-pixel) increments. Quantization is much more noticeable when moving 4 pixels to click on a 1-pixel target than when moving 400 pixels to click on a 100-pixels target, and this could also be a source of difficulty.

Note that other sources of problems in small-scale target acquisition tasks have been mentioned in the literature:

- (d) *“Fat finger” problem*: On touch devices, the relatively large contact area of the finger makes it difficult to acquire small targets because it is not clear for the user where the cursor’s hot spot exactly lies within this area [Holz and Baudisch 2010].
- (e) *Parallax*: On some touch screens and screen tablets, parallax – i.e., systematic deviation between the location of the device or finger and the cursor’s hot spot – is also known to impede pointing on small targets [Ramos et al. 2007].
- (f) *Physical occlusion*: On touch screens, the finger can occlude small targets and make their acquisition challenging [Sears and Shneiderman 1991].
- (g) *Landing and take-off imprecision*: On pen-based devices, landing the stylus on the surface of the tablet or taking off the stylus in order to acquire a small target is error-prone [Ren and Moriya 2000].

These last four issues are device-dependent and none of them affects standard desktop computer pointing, i.e., pointing with a computer mouse and a computer display<sup>2</sup>. In this article, we choose to solely focus on the first three potential sources of problem listed above as they are intrinsic to all small target pointing tasks. We believe that taken together, they form a general and rather complete definition of the concept of “small target”.

Note that C-D gain is absent from our list, despite the fact that targets that are small only in the motor space or only in the visual space will yield high or low C-D gains, which could in turn impact performance. We omitted C-D gain because C-D gain only captures the *ratio* between visual and motor scales and says nothing about task scale. For example, if high C-D gain *per se* is believed to impact performance, then it should impact performance both for small-scale and large-scale pointing tasks in motor space. Conversely, if a high C-D gain is found to impact performance only for small motor scales, then the problem should be attributed to a small-scale effect with motor scale, not to C-D gain.

Motor scale, visual scale and quantization have not received the same amount of attention in HCI. Motor scale is often seen as the only limiting factor and visual scale has been largely ignored [Kabbash and Buxton 1995; Worden et al. 1997; Accot and Zhai 2001; Albinsson and Zhai 2003; Cockburn and Firth 2003; Cockburn and Brewster 2005; Chapuis et al. 2009]. The third problem (quantization) is rarely raised, as it is often assumed that a computer screen provides a smooth and continuous feedback on pointing. We however chose to test its effects, because the continuity assumption does not hold anymore when acquiring targets of only a few pixels in size: jumps of

---

<sup>2</sup>It is possible that the mouse cursor can occasionally occlude small targets. We do not address this potential issue here and control for it in our experiments.

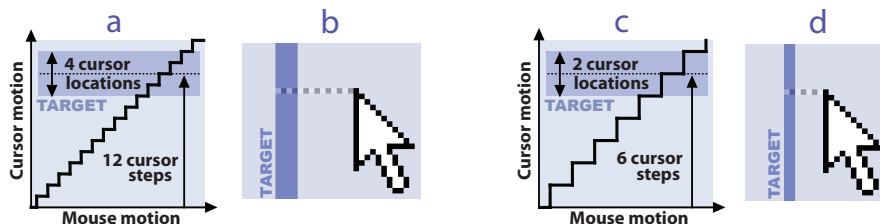


Fig. 1. Quantization in computer pointing tasks. (a) observed motion of the cursor on the screen as a function of physical mouse motion (stepped curve). The target is 12 steps away from the cursor's origin and has 4 slots to accommodate it ; (b) illustration with a right-to-left pointing of 1-pixel steps. (c) same function for a pointing task with double quantization (or half resolution) ; (d) illustration.

the mouse cursor start to get very noticeable near the target and are not easy to predict. It is thus reasonable to assume that they might introduce uncertainty and affect pointing performance.

In the next section we clarify the notion of task scale and the terms *visual scale*, *motor scale* and *quantization*.

### 1.3. Task Scale and Quantization

Since task difficulty depends on both  $W$  and  $D$ , it is helpful to think in terms of *task scale* rather than target size per se. Task scale is difficult to define in absolute terms [Guiard 2009] but relative task scale is a straightforward concept. The scales of different pointing tasks can be expressed relatively to each other if they share the same  $D/W$  ratio. For example, the scale of a task can be said to be twice the scale of another task if both the target size and distance are twice as large.

In configurations involving a mouse and a digital display, we must specify which scale we are talking about. There are three different measures of scale: visual scale, motor scale and resolution (the inverse of quantization). Visual and motor scale are well-understood concepts but the notion of quantization calls for a more careful examination.

In this article *quantization* refers to the subdivision of cursor movements into discrete increments. Since *single pixels also count as increments*, there is always quantization in mouse pointing tasks, unless both the mouse sensors and the display have infinite resolution. Note that although quantization is observed on the screen, it does not matter if it originates from the screen, from the input sensors, or from both. The result is always a cursor moving in discrete steps as a response to continuous motions of the mouse.

Like visual scale and motor scale, defining quantization as an absolute measure is difficult, but it can easily be expressed in relative terms. We define relative quantization and its inverse, relative resolution as follows:

- The *quantization* of a pointing task is  $k$  times the quantization of another task when both the number of increments required to reach the center of the target and the number of possible cursor locations inside the target are *divided by  $k$*  (see Figure 1).
- The *resolution* of a pointing task is  $k$  times the resolution of another task when both the number of increments required to reach the center of the target and the number of possible cursor locations inside the target are *multiplied by  $k$* .

Our definition of quantization departs from two previous publications on C-D gain [Jellinek and Card 1990; Casiez et al. 2008], where quantization was defined as a problem that appears when “*the maximum resolution of the control device together with a high C-D gain prevents every pixel from being addressable on the display*” [Casiez et al. 2008]. But this definition captures a different issue

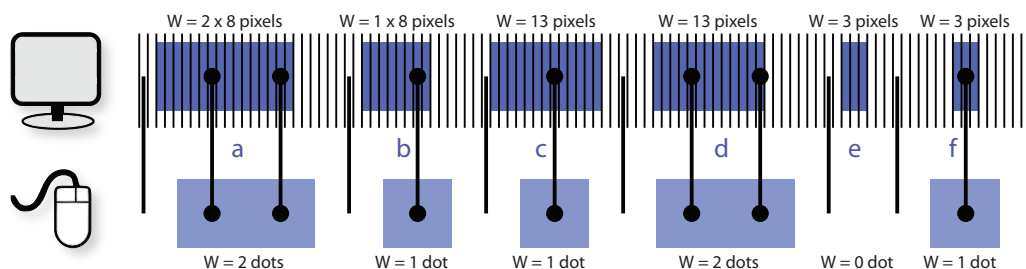


Fig. 2. Target misalignment between visual and motor spaces. On the top, thin lines show boundaries between pixels (only the horizontal dimension is considered). Thick lines show all pixels addressable by the mouse cursor (C-D gain of 8 *pixel/dot*), and those with a round dot fall within a target. Rectangles below reveal actual targets in motor space. a) and b): the targets in screen and motor spaces are offset but their sizes match. c) to f): there is a mismatch between target sizes. We avoid these situations in our experiments.

– i.e., a possible misalignment of targets between the screen and motor spaces – that is more of a design problem than a fundamental property of pointing tasks.

This is illustrated in Figure 2. Here pointing can be difficult not only because of quantization as we define it, but also because the user is not being shown the actual targets in the motor space. Furthermore, if a Fitts experiment is conducted with IDs computed in the screen space, serious drops in performance might be observed not because of quantization or erroneous target display, but simply because the IDs are wrong. Taking Figure 2e as an extreme example, the target cannot be addressed at all, so a pointing experiment would yield infinite movement time for a 3-pixel target. Although such a bad performance could be explained in terms of mouse quantization problems, it is simpler to attribute it to the fact that the target has zero size in motor space and task ID is hence infinite.

These issues vanish if targets on the screen are always perfectly aligned with targets in the motor space. In other terms, target boundaries should fall between two cursor increments. In Figure 2, this would mean aligning the targets above to the targets below. In the rest of this article, we will assume this is always the case.

Note that our simple definition of quantization is consistent with the general meaning of the term<sup>3</sup>. It also has the advantage of avoiding conflation with other factors. Quantization is a separate concept from C-D gain because we can keep C-D gain – defined as a ratio of metric quantities – constant while varying pointing resolution and vice-versa. Pointing resolution can also be manipulated while keeping  $W$  and  $D$  – as expressed in metric units – constant in the visual and in the motor space. With respect to visual and motor scale, resolution is just another way of measuring task scale.

## 2. RELATED WORK

In this section, we briefly recall previous work on small target acquisition techniques. Also relevant to our three key questions are previous investigations into scale effects in Fitts' law and some of the proposed alternatives to Fitts' law.

### 2.1. Small Target Acquisition Techniques

There has been considerable interest for pointing facilitation techniques, especially for helping with the acquisition of small targets. These techniques have been designed and tested on different types of hardware.

<sup>3</sup>From Merriam-Webster's online dictionary, *to quantize*: to subdivide (as energy) into small but measurable increments.

*Desktop and laptop interfaces.* Motor-space target-aware magnification of small targets is probably the most popular approach, and a number of studies have confirmed its usefulness [Cockburn and Firth 2003; Cockburn and Brewster 2005; Kabbash and Buxton 1995; Worden et al. 1997; Blanch et al. 2004; Mandryk and Gutwin 2008]. Another strategy is providing cursor-in-target visual or auditory feedback, but it has been shown to yield marginal improvements at best [Cockburn and Brewster 2005; Cockburn and Brock 2006].

*Pen-based devices.* Ren et al. compared different landing and take-off strategies for selecting targets on a digitizing tablet and found 5 pixels (1.8 mm) to be a critical target size [Ren and Moriya 2000]. Ramos et al. showed that pressure-activated zooming lenses dramatically facilitate pointing for targets less than 4 pixels (1.1 mm) [Ramos et al. 2007].

*Touch screens.* By suppressing finger occlusion, the take-off technique was shown to reduce errors on touch-screens, but tiny targets were still extremely slow to acquire [Sears and Shneiderman 1991]. Albinsson et al. proposed the use of discrete taps or a levering scheme that magnifies motor space [Albinsson and Zhai 2003]. They found both methods to be useful for targets of 1 pixel (0.4 mm), but zooming was found to be the fastest of all. The authors still advocated motor-space approaches in accordance with the belief that, in pointing, “the limitation is in control, not visual resolution” [Albinsson and Zhai 2003], head of the second column of page 106.

Taken together, these studies provide enough anecdotal evidence for a small-scale effect and suggest ways to overcome it, with a strong focus on motor-space target-aware strategies. Other than pointing out a critical target size threshold, however, they have little explanatory power and generalizability. Fitts’ law analyses are either absent or deemed inconclusive — with the exception of two studies that exhibited a good fit but with a single target size condition<sup>4</sup> [Cockburn and Firth 2003; Cockburn and Brewster 2005]. Also, it is not clear what exact problems each of the techniques solves, especially on pen and touch devices, where the “fat finger” problem, occlusion, parallax and landing/take-off imprecision can also affect pointing. Finally, most techniques have several parameters — e.g., magnification factor — whose values seem to be chosen arbitrarily and vary across studies.

## 2.2. Small-Scale Effects in Fitts’ Law Studies

Although scale effects have been studied for very small or very large movement amplitudes [Langolf et al. 1976; Zhai et al. 1996; Balakrishnan and MacKenzie 1997; Kostakos and O’Neill 2008; Casiez et al. 2008], here we primarily focus on scale effects caused by small target widths.

Fitts’ Law studies have been initially conducted on physical apparatuses with equal visual and motor scale and no quantization [Fitts 1954]. Under these conditions, several early analyses observed disparities in the relative contributions of  $D$  and  $1/W$  to movement time, suggesting a problem with targets that are small in both the motor and the visual spaces (for reviews, see [MacKenzie 1992; Plamondon and Alimi 1997]). In particular, it was observed that “reductions of target width cause a disproportionate increase in movement time compared to similar increases in movement amplitude” [MacKenzie 1992]. Similar observations have been made with error rates [MacKenzie 1992; Wobbrock et al. 2008]. However, these studies do not consider the separate effects of visual size, motor size and quantization.

Langolf et al. studied target acquisition under a wide range of motor scales using microscope magnification [Langolf et al. 1976]. The pointing throughput was found to depend on movement amplitude, but when  $D$  was fixed, acquisition time followed Fitts’ law, suggesting no motor accuracy issue even with very small targets (0.076-1.07 mm). Similarly, Guiard et al. used a double-scale visual magnification scheme to test high IDs and found the performance to degrade on very small

<sup>4</sup>Note that this is consistent with Welford’s tremor model (Section 2.3, Equation 4).

targets (0.06 mm) using a puck but not using a stylus [Guiard et al. 1999]. They postulated that with a precision grip “the likely limiting factor for tolerance [ $W$ ] is vision, not motor control”, which contradicts other intuitions [Accot and Zhai 2001; Albinsson and Zhai 2003].

Studies on C-D gain typically manipulate the scale of pointing tasks in the motor space while leaving the visual space unchanged. Despite inconsistent findings (for a review, see [Casiez et al. 2008]), human performance has often been found to be an inverted U-shaped function of C-D gain [Accot and Zhai 2001; Casiez et al. 2008]. Since most studies increase C-D gain by reducing motor pointing scale, the drop in performance observed for very large C-D gains might be due to a motor accuracy problem. Some postulated it could be due to the fact that not all pixels are addressable [Jellinek and Card 1990; Casiez et al. 2008]. But as discussed in Section 1.3, it is not clear whether the cause would be the misalignment of targets between visual and motor spaces, the artificial reduction of targets in the motor space, or the quantization of cursor movement as we define it. With an ultra high-resolution mouse, Casiez et al. still observed a slight decrease in performance for high C-D gains that they attributed to a motor accuracy issue [Casiez et al. 2008]. Similarly, a slight drop in performance was observed for steering tasks that are small in the motor domain and was attributed to motor noise [Accot and Zhai 2001].

Taken together, these studies suggest that motor scale, visual scale and quantization might all contribute to the small-scale effect. However, the relative importance given to these factors varies considerably across studies. As observed before [Accot and Zhai 2001], scale effects are not well-documented in the literature and results are “scattered and controversial”. This is partly due to the use of different input devices (e.g., pen vs. mouse) and/or different muscular groups [Langolf et al. 1976; Zhai et al. 1996; Balakrishnan and MacKenzie 1997; Guiard et al. 1999] across studies. It is arguably difficult to study all possible hardware configurations. That being said, a study that would measure the separate effects of motor scale, visual scale and quantization on a single hardware configuration would be very helpful in understanding the small-scale effect. To the best of our knowledge, there has been no such study before.

### 2.3. Alternatives to Fitts’ Law

A number of modifications to Fitts’ law have been proposed to improve its fit with observed data (for reviews, see [Plamondon and Alimi 1997; Schedlbauer 1997]). While most of them still express movement time ( $MT$ ) as a function of  $D/W$ ,  $D$  and  $1/W$  must be given asymmetric roles in order to account for small-scale effects. We are aware of only four such formulae<sup>5</sup>.

One well-known formula is by Welford and assumes different throughputs for the ballistic and the homing phases of movement [Welford et al. 1969]. This formula has been occasionally used to correct for asymmetric contributions of  $D$  and  $1/W$  to  $MT$  [MacKenzie et al. 1987; Graham and MacKenzie 1996; Parker et al. 2005]:

$$MT = a + b \cdot \log_2(D) + c \cdot \log_2(W) \quad (1)$$

Kvålseth also proposed a power model that exhibited a better fit to the original Fitts’ data [Kvålseth 1980]:

$$MT = a \cdot D^b \cdot W^c \quad (2)$$

<sup>5</sup>We added the missing intercept  $a$  to the original Equations 1 and 4, and replaced 0.5 by 1 in Equation 4 to allow easier comparison with the widely-used Shannon formulation of Fitts’ law [MacKenzie 1992].

Note that this is simply an exponential version of Equation 1. In order to account for the interaction between  $W$  and  $D$ , Oel et al. argued for the following refinement [Oel et al. 2001]:

$$MT = a \cdot D^{b+d \cdot \log_2(W)} \cdot W^c \quad (3)$$

Finally, the first part of Welford's 1969 article contains an alternative model, which to the best of our knowledge has never been mentioned in HCI [Welford et al. 1969]:

$$MT = a + b \cdot \log_2\left(\frac{D}{W-c} + 1\right) \quad (4)$$

Where  $c$  is an experimentally-determined constant attributed to hand tremor. In contrast to the previous formulae this model is easier to interpret and as we later show, it best accounts for targets that are small in the motor space. Note however that the term “tremor” should not be taken literally, and this will be later discussed in section 6.4.

### 3. EXPERIMENT

We conducted a user study in order to confirm the existence of a small-scale effect on a standard mouse-display configuration and to investigate its causes. We asked participants to perform 1-D target acquisition tasks and independently manipulated visual scale, motor scale and quantization. We first introduce our experiment design and our use of *scaling methods* as a way to manipulate these three factors. We then discuss our main findings and their implications.

#### 3.1. Scaling Methods

As explained in the introduction, desktop pointing tasks are characterized by three different measures of scale: *visual scale*, *motor scale* and *quantization* (or its inverse, *resolution*). Recall these quantities are all *relative*, e.g., a visual scale of 2 means magnifying  $D$  and  $W$  by  $2\times$  on the screen according to a nominal pointing task. Also recall these quantities are *independent* of each other. Thus, they form a three-dimensional space that can be used to characterize the scale of pointing tasks with respect to an arbitrary nominal task with the same  $D/W$ .

An ideal experiment would evenly sample this three-dimensional space for different values of  $D/W$ . However, there are two difficulties. First, some points cannot be measured due to hardware constraints: if a pointing task involves a 1-pixel target, it is not possible to scale it visually by a factor of 0.5 or 1.5, nor is it possible to reduce quantization. Second, a full factorial design would either yield too many conditions for a single experiment or prevent us from testing enough scale values. We therefore devised an approach based on scaling methods (see Figures 3 and 4).

A *scaling method* is a specific way of scaling up a pointing task. Suppose the task is to acquire a target of distance and size  $D_v, W_v$  on the screen and  $D_m, W_m$  in the motor space. Let  $D_s$  be the number of discrete steps required for the cursor to progressively reach the middle of the target and  $W_s$  the number of possible cursor locations inside the target. These are typically equal to  $D_v$  and  $W_v$  as expressed in pixels. Finally, let  $C_v$  be the cursor's visual size and  $G$  the C-D gain, i.e., the ratio between the cursor's and the mouse's travel distances, both expressed in metric units.

Such a target acquisition task can be scaled up by a positive integer  $S$  using five possible methods (see Figure 3):

- *MotorMag* multiplies  $D_m$  and  $W_m$  by  $S$  and leaves the rest unchanged, except for  $G$  which is multiplied by  $1/S$ . In other words, this is a motor-space magnification.
- *VisualMag* multiplies  $D_v, W_v$  and  $C_v$  by  $S$  and leaves the rest unchanged, except for  $G$  which is multiplied by  $S$ . In other words, this is a visual-space magnification.
- *FullMag* multiplies  $D_m, W_m, D_v, W_v$  and  $C_v$  by  $S$ , and leaves the rest unchanged. This amounts to increasing the size of the pixel and of the mouse dot.

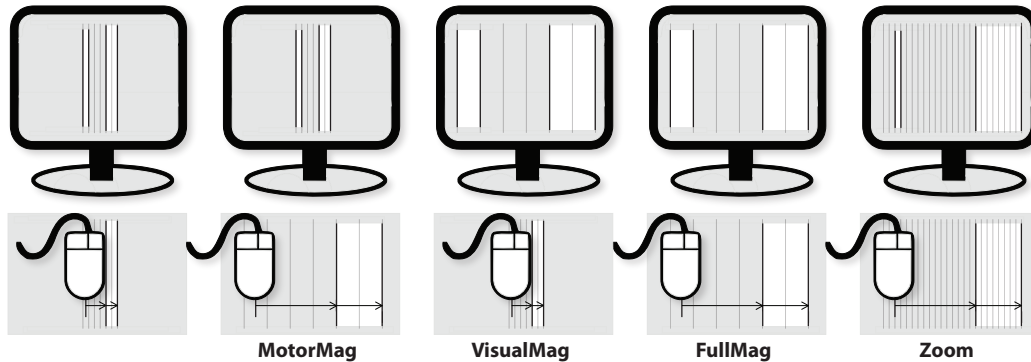


Fig. 3. Scaling up a task by  $\times 2$  in the motor space (MotorMag); in the visual space (VisualMag); in the motor and visual spaces (FullMag); in the motor, visual and resolution spaces (Zoom). The left rectangle on the screen is the mouse cursor.

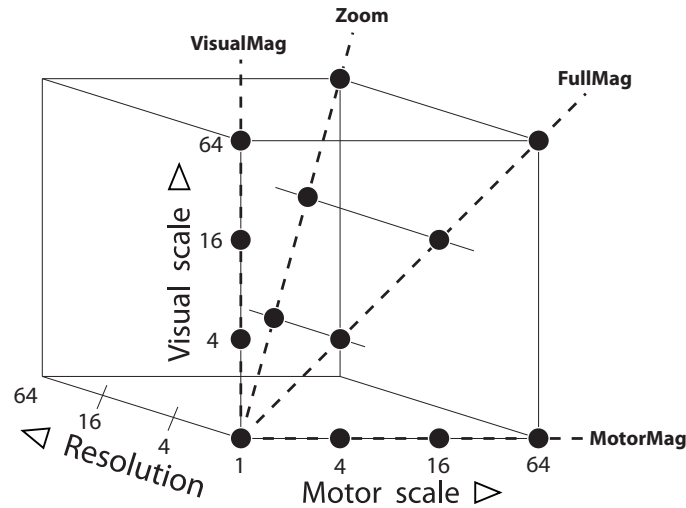


Fig. 4. Each of the scaling methods covers a specific line in the three-dimensional scale space. The black dots show the measures included in our experiment. Each dot corresponds to a scaled-up version of a nominal pointing task (the black dot at the origin) that is small-scale in all respects.

- *Zoom* is the same as FullMag except that  $C_v$  is left unchanged and  $W_s$  and  $D_s$  are multiplied by  $S$ . This amounts to enlarging the task in both spaces while keeping the original pointing resolution<sup>6</sup>.
- *ZoomBis* is the same as Zoom, but in addition, the cursor's size is multiplied by  $S$ .

We introduced the ZoomBis variant because one effect of FullMag is to increase the cursor's size, while the cursor remains visually small for Zoom — hence a possibility of confound (see Figure 3). In ZoomBis and FullMag, the cursor is the same. However, since participants still have to bring the

<sup>6</sup>This method can cause pixels to be misaligned. For example, on Figure 3, the nominal task  $W = 2, D = 5.5$  scales up to  $W = 4, D = 10$ . For the Zoom condition, this requires the cursor to initially lie between two pixels and for the mouse sensor to lie between two dots. For the mouse, we use the extra resolution available to achieve this. For the screen, we arbitrarily choose between the left and the right pixel and counterbalance across trials.

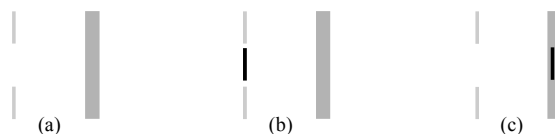


Fig. 5. A sample trial: (a) the target and the future location of cursor are shown; (b) the cursor appears after a random foreperiod; (c) the user selects the target.

whole cursor into the target,  $W_v$  is enlarged by the same amount as  $C_v$  in order to preserve  $W_s$  and  $W_m$ .

Figure 4 illustrates the way these scaling methods sample the three-dimensional scale space. The sampling does not fully cross the three axes but still allows us to assess the relative effect of each of the axes on pointing performance. For example, FullMag and Zoom only differ by their amount of quantization. The effects of quantization can hence be examined by comparing these two conditions at different scales. Similarly, if we assume the C-D gain to have little or no influence [Casiez et al. 2008], the only important difference between VisualMag (resp. MotorMag) and FullMag is the task's motor (resp. visual) scale.

### 3.2. Apparatus

We conducted the experiment on a high-end workstation running X Window. The display was a 22" ultra-high-resolution LCD monitor with a native resolution of  $3840 \times 2400$  and 0.125 mm pixel size (about 200 dpi). The pointing device was a ultra-high-resolution gaming mouse of 83.5 dots per mm (about 2000 dpi), polled at 500Hz and whose cursor was refreshed at 60Hz. The mouse was teflon-coated and was used on a varnished plywood surface.

We set the display to half its native resolution and programmatically divided the resolution of the mouse by 5, thus approaching the resolutions and C-D gain of a standard desktop computer with mouse acceleration disabled, while virtually eliminating potential confounds due to display quality and input sensor accuracy. Interactions between color and geometry present on LCD displays were also eliminated by the exclusive use of shades of gray [Messing and Daly 2002].

The experiment software was written in Java using the SwingStates [Appert and Beaudouin-Lafon 2008] and TouchStone [Mackay et al. 2007] libraries, and uses the JInput library for polling mouse data at a low-level, i.e., in pure dot units and without acceleration.

### 3.3. Participants

Twelve unpaid volunteers, ten male and two female, all right-handed, participated in the experiment. Two additional participants failed the vision test (see below) and thus did not proceed with the experiment. Participants were experienced mouse users with ages ranging from 21 to 32 (mean  $25.33 \pm 3.80$ , median 25).

### 3.4. Task and Procedure

Each trial was a 1-D discrete target acquisition task decomposed as follows. First, a gray target and two gray markers symbolizing the location of the cursor appeared at the center of the screen (Figure 5a). Participant were instructed not to move the mouse until the actual (black) cursor appeared. After a short random foreperiod, the black cursor was shown (Figure 5b) and the participant had to bring it inside the target (Figure 5c) and press the left mouse button.

To better resemble a typical computer pointing task, a trial had to end successfully even if it included mouse presses outside the target. The next trial then started after the participant had released the mouse button. We recorded movement times from the first mouse movement to the first mouse

press, as well as to the first successful mouse press. The participant was allowed to rest every 22 trials. Note that we used a discrete movement protocol that ensured that the measured movement time is the duration of a pure execution process, avoiding interpretation problems [Guiard 1993; Fernandez et al. 2006].

Before starting the experiment, participants were given written instructions telling them to be as fast and as accurate as possible and to avoid mouse clutching. To allow them to do so, mouse drifting was eliminated by grouping trials by pairs with the same target distance and width, but with opposite movement direction.

Upon reading the instructions, participants were asked to sit in a comfortable position and the screen was moved to a distance of 70 cm from their eyes. A string was then placed in front of their chest to remind them not to lean forward, as mentioned in the written instructions.

Participants were then given a vision test involving 12 labeled representations of a target and a cursor, where the cursor was 1 pixel and the targets 1 or 2 pixels. The cursor was either inside the target or just next to it. Participants were asked to tell when the cursor was inside the target. The test was considered successful if all targets were properly identified and there was no false positive.

### 3.5. Design

The experiment used a within-participants design with the following factors:

- five scaling methods *METHOD*: Zoom, ZoomBis, FullMag, MotorMag and VisualMag;
- four scales *SCALE*: 1, 4, 16 and 64;
- four nominal widths *W*: 1, 2, 4 and 8 pixels/dots;
- four nominal distances *D*: 2, 4, 8 and 16 pixels/dots.

We included all the possible combinations of  $D = \{8, 16\}$  and  $W = \{1, 2, 4, 8\}$  plus the combinations (2, 1), (4, 1) and (4, 2), for a total of 11 (*D*, *W*) couples. (*D*, *W*) was fully crossed with *METHOD* and *SCALE*, giving a total of 220 conditions. Fitts' IDs remained constant across *METHOD* and *SCALE* and ranged from 1 to 4.09<sup>7</sup>.

As mentioned before, movement directions were alternated to prevent mouse drift. For each *METHOD* × *SCALE* condition, participants were presented with 5 blocks of *Direction* × *W* × *D* = 5 × 2 × 11 = 110 trials. The presentation order of *W* × *D* was randomized within each block. The first block was for training and the remaining four were recorded.

A pilot study suggested that large variations in C-D gain were the most difficult to accommodate (the nominal gain was 4.15 and ranged from 0.065 to 4.15 for MotorMag and from 4.15 to 266 for VisualMag). We hence blocked by *METHOD* and sub-blocked by *SCALE*. For MotorMag and VisualMag, we presented the *SCALE* conditions monotonically and informed participants of changes in mouse sensitivity through textual messages. For the other methods, where the C-D gain was constant, the presentation order of *SCALE* was randomized.

Additionally, in order to avoid important changes in C-D gain when transitioning between MotorMag and VisualMag, we always presented these two methods at the second and at the fourth position. We computed a Latin square for (Zoom, ZoomBis, FullMag) and crossed it with the two possible orderings of (MotorMag, VisualMag), yielding six different orderings for *METHOD*, each of which was presented to two participants.

<sup>7</sup>These rather low IDs stem from the fact that we are testing a wide range of scales, and the limited screen size and mouse footprint prevent us from including larger values of *D*. Small-scale effects are nonetheless most likely to occur near targets and should be therefore more measurable if small distances are used. We also chose to fully cross *D* and *W* as proposed by [Guiard 2009] in the case of a non Fitts' law experiment.

Table I. Results of the ANOVA for  $MT \sim \text{METHOD} \times \text{SCALE} \times W \times D \times \text{Random}(\text{PARTICIPANT})$ .

| Factors                                     | All SCALE |       |          | SCALE $\geq 4$ |        |          |
|---|-----------|-------|----------|----------------|--------|----------|
|   | DF,Den    | F     | p        | DF,Den         | F      | p        |
| METHOD                                      | 4,44      | 71.2  | < 0.0001 | 4,44           | 71.2   | < 0.0001 |
| SCALE                                       | 3,33      | 35.7  | < 0.0001 | 2,22           | 16.9   | < 0.0001 |
| W   | 3,33      | 166.9 | < 0.0001 | 3,33           | 560.0  | < 0.0001 |
| D   | 1,11      | 234.5 | < 0.0001 | 1,11           | 1441.4 | < 0.0001 |
| METHOD $\times$ SCALE                       | 12,132    | 36.3  | < 0.0001 | 8,88           | 15.9   | < 0.0001 |
| METHOD $\times$ W                           | 12,132    | 27.3  | < 0.0001 | 12,132         | 27.3   | < 0.0001 |
| METHOD $\times$ D                           | 4,44      | 23.0  | < 0.0001 | 4,44           | 23.0   | < 0.0001 |
| SCALE $\times$ W                            | 9,99      | 17.7  | < 0.0001 | 6,66           | 16.0   | < 0.0001 |
| SCALE $\times$ D                            | 9,33      | 2.4   | 0.0829   | 2,22           | 11.8   | 0.0003   |
| W $\times$ D                                | 3,33      | 0.7   | 0.5509   | 3,33           | 2.0    | 0.1365   |
| METHOD $\times$ SCALE $\times$ W            | 36,396    | 8.0   | < 0.0001 | 24,264         | 2.4    | 0.0004   |
| METHOD $\times$ SCALE $\times$ D            | 12,132    | 2.9   | 0.0013   | 8,88           | 0.6    | 0.7727   |
| METHOD $\times$ W $\times$ D                | 12,132    | 1.6   | 0.0857   | 12,132         | 1.6    | 0.0857   |
| SCALE $\times$ W $\times$ D                 | 9,99      | 1.6   | 0.1226   | 6,66           | 1.9    | 0.0908   |
| METHOD $\times$ SCALE $\times$ W $\times$ D | 36,396    | 1.3   | 0.1262   | 24,264         | 1.2    | 0.1951   |

Note that in the case of SCALE 1, all the METHOD conditions are equivalent (see Figure 4). We hence decided to present the condition only once and arbitrarily assigned it to the Zoom METHOD. This removed 4 (METHOD)  $\times$  5 (Blocks)  $\times$  11 (W  $\times$  D)  $\times$  2 (Direction) = 440 redundant trials.

A participant hence performed 5 (METHOD)  $\times$  4 (SCALE)  $\times$  5 (Blocks)  $\times$  11 (W  $\times$  D)  $\times$  2 (Direction) – 440 = 1760 pointing tasks, 1408 of which were recorded (the others being training trials). We obtained 96 measures for a full condition and a total of 16 896 measures. The experiment lasted approximately 65 minutes.

## 4. RESULTS

We first perform a full factorial analysis by removing the data for D = 2 and D = 4 and by duplicating the data from Zoom METHOD at SCALE = 1 to the other methods. Recall only D = {8, 16} and W = {1, 2, 4, 8} are fully crossed. The remaining data, e.g., the (D, W) couples (2, 1), (4, 1) and (4, 2), will be dealt with later on.

### 4.1. Movement Time

In this analysis, we consider MT to the successful mouse press, as this measure has the advantage of including penalties caused by errors. We performed a *nominal* factorial repeated measures analysis of variance for the model  $MT \sim \text{METHOD} \times \text{SCALE} \times W \times D \times \text{Random}(\text{PARTICIPANT})$ . Table I shows the results of this ANOVA for the data with and without SCALE = 1<sup>8</sup>.

Note that in this analysis we consider all our factors – in particular SCALE, W and D – as nominal (categorical) factors. This is common practice and has the advantage that the analysis does not assume linear correlations between SCALE, W and D vs MT and can thus detect “any kind” of effect.

**4.1.1. Scaling Methods.** We first observe a significant effect of METHOD and SCALE on MT and a significant interaction METHOD  $\times$  SCALE. Figure 6<sup>9</sup> shows the effect of SCALE on each METHOD. We observe that Zoom and ZoomBis follow a similar pattern: MT rapidly decreases as scale increases, up to a scale of 16. A post-hoc Tukey test for difference in means confirms these observations.

<sup>8</sup>Recall SCALE = 1 is a special case that corresponds to the nominal pointing tasks. At SCALE = 1 all methods leave the nominal tasks unchanged and are therefore equivalent. Also recall these tasks have been presented only once and duplicated across methods to allow full-factorial analysis.

<sup>9</sup>Error bars in all the figures represent the 95% confidence limits of the sample mean ( $mean \pm StdErr \times 1.96$ ).

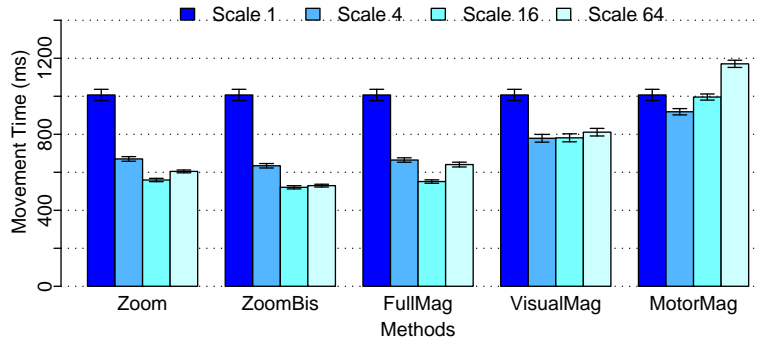


Fig. 6. MT as a function of SCALE, grouped by METHOD. The five darkest bars represent the same data.

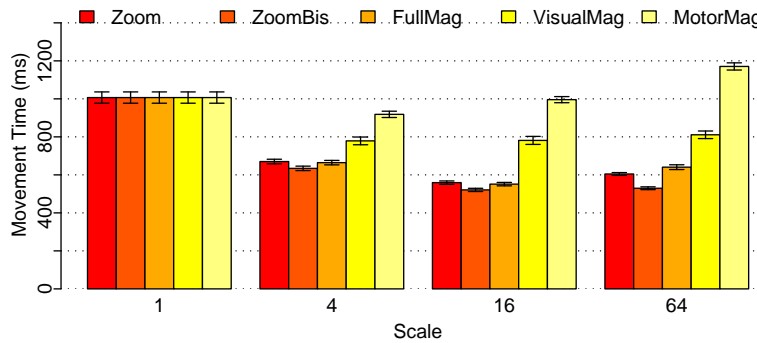


Fig. 7. MT as a function of METHOD, grouped by SCALE. The five leftmost bars represent the same data.

FullMag is similar but we observe a drop in performance for SCALE = 64. With MotorMag, we see a small performance improvement from scale 1 to 4 and then a degradation. For VisualMag, we also find an improvement from scale 1 to 4, but higher scales seem to have no effect. All these observations were confirmed by post-hoc tests.

Figure 7 shows the effect of METHOD on each SCALE. An important observation is that Zoom, ZoomBis and FullMag seem very close. Indeed, the only difference shown by a post-hoc Tukey test is between ZoomBis and FullMag for SCALE = 64.

These results confirm the existence of a strong small-scale effect: when the task scale is small, increasing pointing scale using either Zoom or ZoomBis largely improves performance, despite the  $D/W$  ratios being the same. That FullMag would yield the same improvements was quite unexpected, and implies that quantization has no or very little negative effect on performance. Another surprising result was the large initial improvement obtained by increasing visual scale (VisualMag from SCALE 1 to 4), suggesting that the cause of the small-scale effect is partly visual. An initial improvement is also observed with MotorMag, confirming a motor cause to the small-scale effect. But surprisingly, increasing the scale further degrades performance. Since Zoom did not exhibit this problem despite involving the exact same targets in the motor space, it is very likely that this degradation is due to the low C-D gains involved (1.04, 0.26 and 0.065).

**4.1.2. Width and Distance.** We now analyze the factors W and D. Unsurprisingly, we found simple effects on MT. Figure 8 illustrates the effect of D as well as the interaction  $D \times \text{METHOD}$ . In

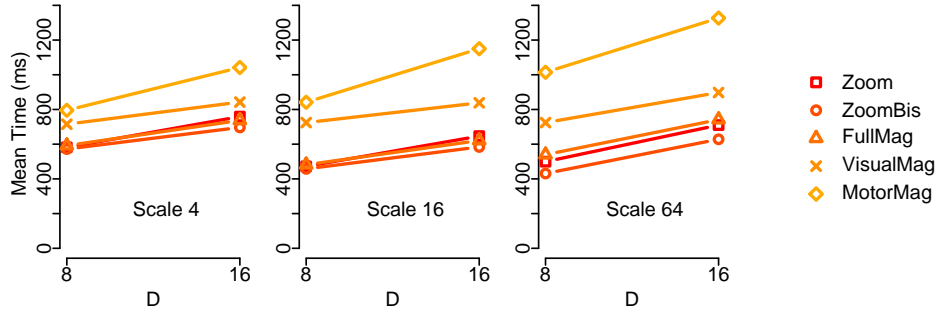


Fig. 8. MT as a function of D by METHOD for each SCALE  $\geq 4$ .

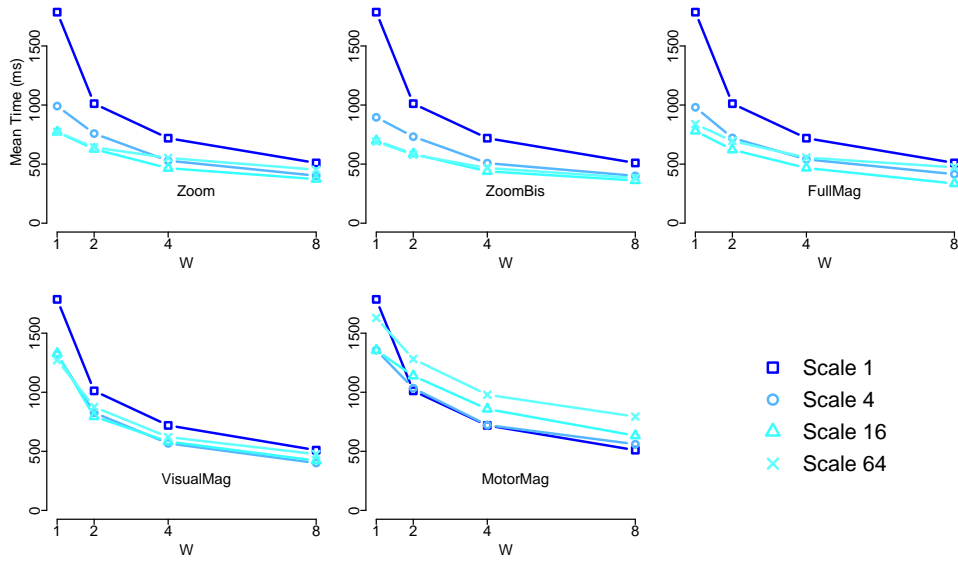


Fig. 9. MT as a function of W by SCALE for each METHOD.

particular, one can observe (and post-hoc Tukey test can show) that the impact of D is weaker with VisualMag than with the others methods.

Figure 9 illustrates the interactions between W and SCALE for each METHOD. We can see that for Zoom, the difference between SCALE = 1 and the others scales decreases as W increases. ZoomBis, FullMag and VisualMag exhibit a similar trend, but for VisualMag and SCALE  $\geq 4$  the lines are almost confounded. However, as can be observed in Figure 9, the situation is completely different with MotorMag: for all W,  $\geq 2$  the difference between scales is almost constant, but with SCALE = 1, a clear degradation can be observed at W = 1. These observations show the cause of the significant interaction METHOD  $\times$  SCALE  $\times$  W. This together suggests that scale has little or no effect as long as it is large enough, except for motor-space magnification where increasing scale – and reducing C-D gain – keeps degrading performance.

#### 4.2. Error Rate

We found an overall error rate of 7.02% (5.80% for SCALE  $\geq 4$ ). Repeated measures analysis of variance for the model ERROR RATE  $\sim$  METHOD  $\times$  SCALE  $\times$  W  $\times$  D  $\times$  Random(PARTICIPANT) reveals a significant effect of METHOD ( $F_{4,44} = 10.1, p < 0.0001$  both) and W ( $F_{3,33} = 32.9, p <$

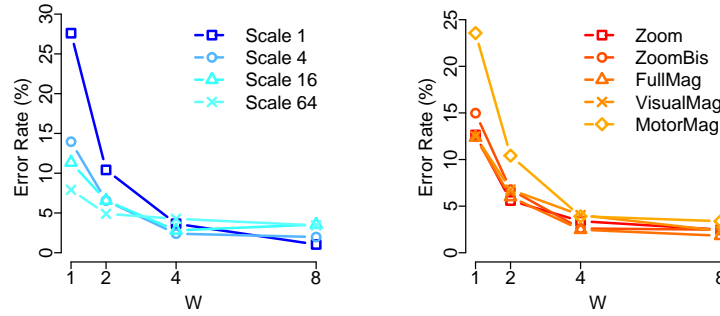


Fig. 10. Error rate as a function of  $W$ , for each SCALE (left) and each METHOD (right).

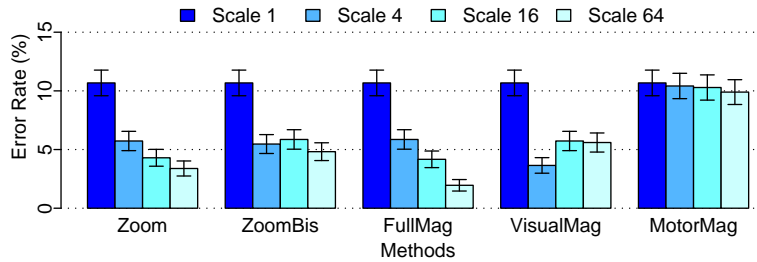


Fig. 11. Error rate as a function of SCALE grouped by METHOD

0.0001 and  $F_{3,33} = 49.1, p < 0.0001$ ) for both datasets under consideration (all SCALE and  $SCALE \geq 4$ ) and a significant effect of SCALE for the data with all SCALE ( $F_{3,33} = 7.0, p = 0.0009$ ) but not for  $SCALE \geq 4$  ( $F_{2,22} = 2.7, p = 0.0885$ ). No effects were found for  $D$  ( $F_{1,11} = 2.9, p = 0.1116$  and  $F_{1,11} = 3.9, p = 0.0718$ ).

However, we observed significant interactions of  $METHOD \times W$  ( $F_{12,132} = 6.1, p < 0.0001$  both) and  $SCALE \times W$  ( $F_{9,99} = 8.3, p < 0.0001$  and  $F_{6,66} = 10.6, p < 0.0001$ ) for both datasets, as shown in Figure 10. For  $SCALE = 1$  in particular, we have a higher error rate than other scales when  $W = 1$  but this difference vanishes as  $W$  increases. The same can be said when comparing MotorMag with other methods. For the remaining METHOD, we only see very small differences in the error rates.

Note that we did not find a significant interaction  $METHOD \times SCALE \times W$  as we did for *MT*. However, we found a significant interaction of  $METHOD \times SCALE$  ( $F_{12,132} = 3.9, p < 0.0001$  all SCALE), which was no longer significant when restricting the data to  $SCALE \geq 4$  ( $F_{8,88} = 1.9, p = 0.0750$ ). Figure 11 shows the error rate as a function of SCALE grouped by METHOD. Increasing SCALE using MotorMag fails to reduce error rate, but with the other METHOD, increasing SCALE from 1 to 4 drastically reduces error rate.

These results confirm our hypotheses that target size is the main cause of errors and that “small sizes” can lead to very high error rates. Interestingly, we found no effect of target distance (and no interaction between distance and other factors), and thus we measured a small target size effect relatively to Fitts’ Law – as in [Wobbrock et al. 2008]. Another interesting observation is that visual magnification alone is enough to drastically reduce error rate (even if it leads to huge C-D gains) while motor magnification has no effect.

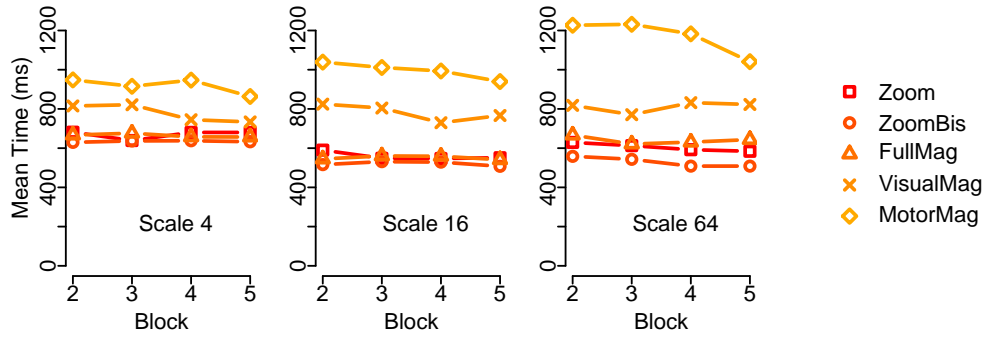


Fig. 12. MT as a function of block by METHOD for each SCALE > 1.

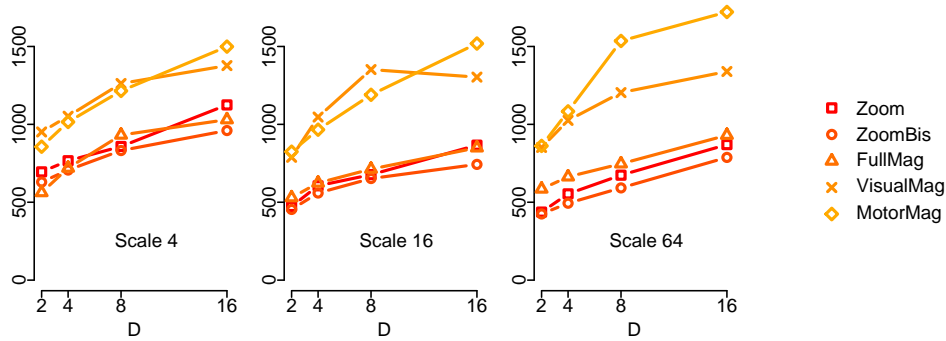


Fig. 13. Movement Time as a function of D by METHOD for each SCALE > 1 for D in {2,4,8,16} and W = 1.

### 4.3. Learning

Regarding learning effects, we performed an analysis of variance for the model  $MT \sim \text{BLOCK} \times \text{METHOD} \times \text{SCALE} \times \text{Random}(\text{PARTICIPANT})$ . We found a significant simple effect of BLOCK for the SCALE  $\geq 4$  data ( $F_{3,33} = 6.4$ ,  $p = 0.0015$ ) but not for the full data set ( $F_{3,33} = 2.3$ ,  $p = 0.0923$ ). However, we found significant interactions for BLOCK  $\times$  METHOD ( $F_{12,132} = 2.5$ ,  $p = 0.0055$ ) and BLOCK  $\times$  METHOD  $\times$  SCALE (and no BLOCK  $\times$  SCALE interaction) for both datasets. In Figure 12 we can observe that the learning effect is essentially caused by MotorMag and that this effect is especially strong for SCALE = 64. Indeed, post-hoc Tukey tests show a significant difference in means (around 100 ms) between block 5 and blocks 2 and 3 for MotorMag but no difference for the other methods. This suggests that participants initially have difficulties with very small C-D gains but are able to improve with practice.

### 4.4. Small Distances and Small Widths

So far we have excluded the trials where  $D = 2$  and  $D = 4$ . To study these cases, we have performed full factorial analyses of variance for (i) the model  $MT \sim \text{METHOD} \times \text{SCALE} \times D \times \text{Random}(\text{PARTICIPANT})$  on the data subset where  $D \in \{2, 4, 8, 16\}$  and  $W = 1$  and (ii) the model  $MT \sim \text{METHOD} \times \text{SCALE} \times W \times D \times \text{Random}(\text{PARTICIPANT})$  on the data subset where  $D \in \{4, 8, 16\}$  and  $W \in \{1, 2\}$ . We obtained results very similar to those discussed previously. The only notable fact is that, for  $D = 2$  and  $D = 4$ , MotorMag performed as well as VisualMag at SCALE = 64 and thus performed better than expected from our previous results. This can be observed in the rightmost graph of Figure 13 and shows that large distances are a problem with MotorMag.

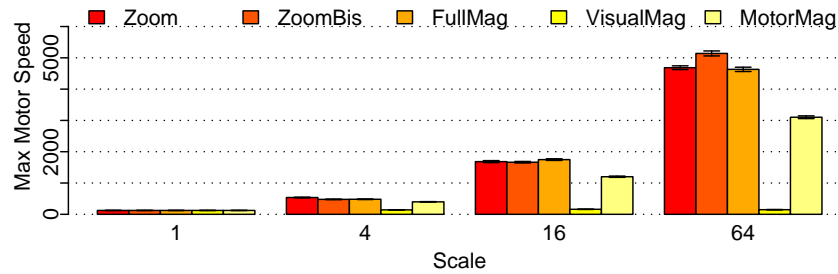


Fig. 14. Maximum motor speed ( $\text{dot.s}^{-1}$ ) as a function of METHOD grouped by SCALE.

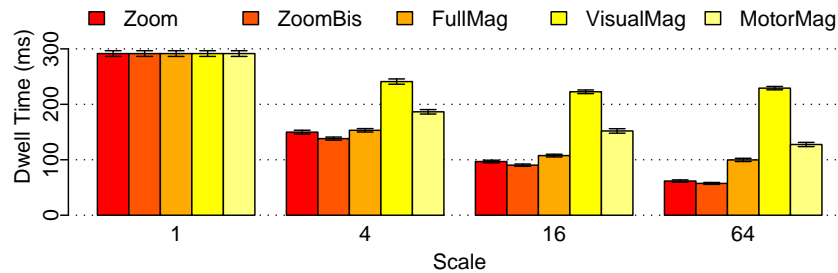


Fig. 15. Dwell time as a function of METHOD grouped by SCALE.

#### 4.5. Maximum Speed and Dwell Time

Figure 14 shows the mean peak velocities of the movements in the motor space as a function of METHOD grouped by SCALE. We observe that MotorMag exhibits lower peak velocities, despite requiring the same motor movement as Zoom, ZoomBis and FullMag (for VisualMag the required motor movement is independent from SCALE). This shows – and statistics confirm – that participants had difficulties performing fast movements with small-amplitude visual feedback. This is consistent with our previous supposition that low C-D gains cause the low performances observed with strong motor-space magnification.

Now we examine the end of the movement by considering dwell time. We define *dwell time* as the time between the last *motor* movement and the click. Figure 15 shows dwell time as a function of METHOD grouped by SCALE. We observe that dwell time decreases as target width increases for Zoom and ZoomBis and also, to a lesser extent, for MotorMag. However, this is not the case for FullMag between SCALE 16 and 64 – we checked that this phenomenon was true for all W and D. Moreover, it can be verified that the difference in MT observed between FullMag and Zoom at SCALE 64 is due to dwell time.

Thus, it seems that the quantization effect that can be measured by comparing the FullMag condition with the zoom conditions at high scales is concentrated at the end of the movement. Zoom and FullMag need the same movement but the feedback provided by FullMag at scale 64 is highly not continuous. This does not disturb the participant during the first part of the movement (the “ballistic” part, see Figure 14), but at its very end.

Also note that when the nominal task is magnified by 4, the reduction in dwell time is much higher for MotorMag than for VisualMag. Moreover, dwell time decreases up to SCALE 16 for MotorMag, while it stops decreasing at SCALE 4 for VisualMag. This suggests that motor accuracy

is the principal source of hesitation at the end of the movement, i.e., when the user has to determine whether the cursor is on the target or not.

#### 4.6. Variants on MT Measures

One may want to stick to a more traditional methodology for measuring movement times and errors [MacKenzie 1992] and either: (i) take movement times up to the first click ( $MT_0$ ) instead of the successful click ( $MT_1$ ); or (ii) remove errors ( $MT_1 = MT_0$ ) and outliers<sup>10</sup> (using  $MT_0$ ). In both cases the analysis yields results that are very similar to those we obtained so far. The results of the ANOVA's yield different  $p$  and  $F$  values but no change in significance. Of course, conditions with high error rates become faster, but the differences are small and do not affect the results.

### 5. MOVEMENT TIME MODELS

As in standard Fitts' law analyses, we consider here the movement time to the first click [MacKenzie 1992]. Moreover, we aggregate the data over METHOD, SCALE, W and D, taking the mean over all participants. Fitts'  $ID$ s are computed from the Shannon formulation [Mackenzie 1991]. Target widths and distances denoted by  $W_m$  and  $D_m$  are expressed in the motor space, in mouse dots (1dot = 0.060mm): although the choice of units does not impact Fitts'  $ID$ s, it will matter when testing alternatives to Fitts' law.

In order to reduce the risk of overfitting our data and following the principle of parsimony, we report adjusted  $r^2$  statistics, a measure of fit that takes into account the number of free parameters in the model. As suggested by one of our reviewers, we additionally report Akaike Information Criterion ( $AIC$ ) values, a measure of fit that penalizes free parameters more strongly and is used for model selection [Burnham and Anderson 1998]. According to this method, the model with the *lowest*  $AIC$  ( $AIC_{min}$ ) is the best one in terms of trade-off between explanatory power and complexity. As a rule of thumb, models for which  $AIC \leq AIC_{min} + 2$  deserve consideration and those for which  $AIC \geq AIC_{min} + 10$  can be safely rejected [Burnham and Anderson 1998].

#### 5.1. Visual Magnification

We begin with a Fitts' Law analysis of all VisualMag tasks. Recall that these tasks involve targets that are *small in the motor space* but have various visual sizes (depending on SCALE). The left plot in Figure 16 shows movement time as a function of Fitts ID and the corresponding regression line. As we already observed in Figure 9, the data departs from Fitts' Law, especially at SCALE 1. The fit is not extremely good ( $r^2 = 0.767$ ), with the smallest targets lying way above the regression line.

Interestingly, correcting IDs using Welford's "tremor" model solves the problem (see also Equation 4):

$$ID_c = \log_2\left(\frac{D_m}{W_m - c} + 1\right) \quad (5)$$

As Figure 16 shows, choosing<sup>11</sup>  $c = 0.79$  dots yields a much better fit ( $r^2 = 0.946$ ). The Akaike information criterion measure confirms that this additional free parameter  $c$  explains the data sufficiently well for justifying its inclusion in the model ( $AIC = 560$  for the  $ID$  model and  $AIC = 495$  for the  $ID_c$  model, recall that smaller values are better).

<sup>10</sup>We follow [Soukoreff and MacKenzie 2004] and define an outliers as a movement time 3 standard deviations away from the mean movement time (for each condition).

<sup>11</sup>Coefficients are computed by doing an exhaustive search with a precision of 0.01 and maximizing the linear fit.

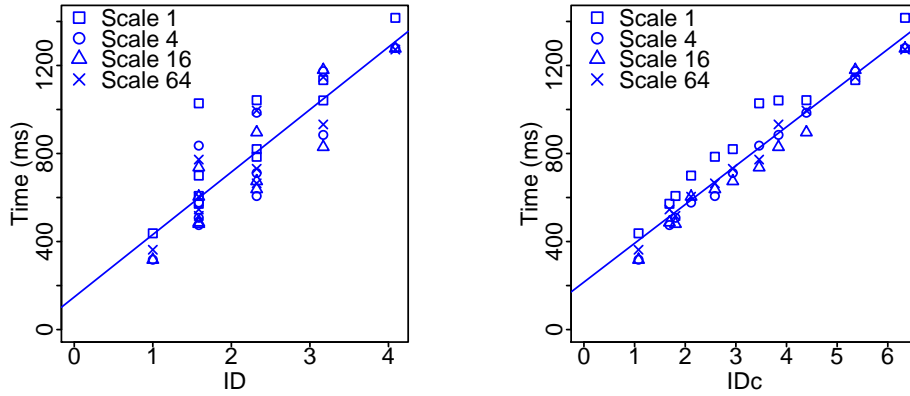


Fig. 16. MT as a function of  $ID$  and  $ID_{c=0.79}$  for VisualMag.

Table II. Adjusted  $r^2$  and  $AIC$  for the various models considered for VisualMag.

|               | All SCALE |       | SCALE $\geq 4$ |       |
|---------------|-----------|-------|----------------|-------|
|               | $r^2$     | $AIC$ | $r^2$          | $AIC$ |
| $ID$          | 0.767     | 560   | 0.799          | 416   |
| $ID_c$        | 0.946     | 495   | 0.982          | 339   |
| Welford Eq. 1 | 0.904     | 522   | 0.931          | 383   |
| Oel et al.    | 0.942     | 497   | 0.978          | 349   |

For SCALE  $\geq 4$ , i.e., excluding tasks where the targets are small in *both* motor and visual spaces, the improvement is even more dramatic, going from  $r^2 = 0.799$  to  $r^2 = 0.982$  with the same best  $c$ . The percentage of unexplained variance  $100(1 - r^2)$  gives a clearer idea of the actual improvements in  $r^2$ : this percentage drops from 20.1% for the  $ID$  model to 1.8% for the  $ID_c$  model. With 33 data points to fit, it is very unlikely that this improvement is due to chance, and this is confirmed by independently testing  $ID_c$  on the 4 different visual scales (yielding  $c = 0.79 \pm 0.04$  and  $r^2 = 0.97 \pm 0.02$ ) and by the Akaike information criterion measure ( $AIC = 416$  for  $ID$  and  $AIC = 339$  for  $ID_c$ ).

Note that the  $ID_c$  model does better than the other (more popular) model from Welford (Equation 1) that also has 3 free parameters but yields  $r^2 = 0.931$  (leaving 6.9% of variance unexplained) and  $AIC = 383$  for SCALE  $\geq 4$ , and  $r^2 = 0.904$  and  $AIC = 522$  for all SCALE. Moreover, the  $ID_c$  model yields a slightly better  $r^2$  than Oel *et al*'s model (Equation 3) that gives  $r^2 = 0.978$ ,  $AIC = 349$  and  $r^2 = 0.942$ ,  $AIC = 497$ , despite having 4 free parameters. These results together suggest that Welford's "tremor" model accurately models pointing for targets that are small in the motor space no matter their visual size, although it does even better when the target is small in the motor space only. See Table II for a summary of the results of this subsection.

## 5.2. Motor Magnification

Now we perform a Fitts' Law analysis of all MotorMag tasks, which involve targets that are *small in the visual space* but have various motor sizes. We drop SCALE 1 in order to obtain large enough targets in the motor space (we already showed that small motor scale is well-modeled by Equation 4). The left plot in Figure 17 shows the standard Fitts regression. Like VisualMag, the fit is not extremely good ( $r^2 = 0.824$ ), but the alternative models mentioned above do not yield satisfactory

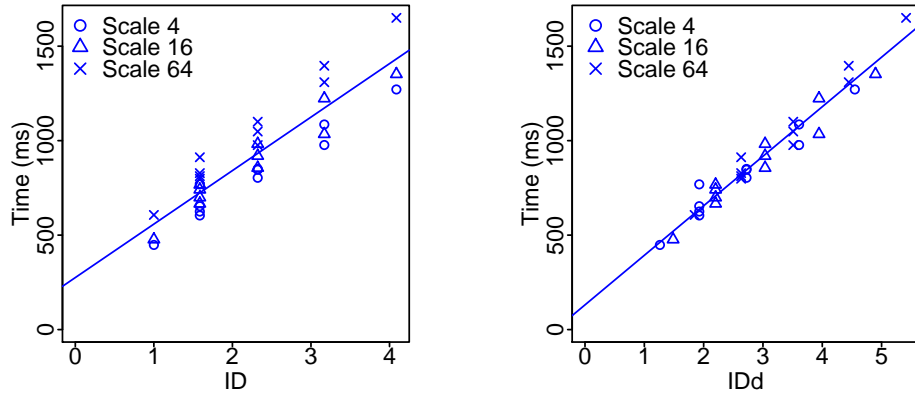


Fig. 17. MT as a function of  $ID$  and  $ID_{d=0.24}$  for MotorMag.

results: the improvements obtained are weak and are most certainly due to over-fitting. This suggests a different phenomenon.

For MotorMag, recall that we observed a large drop in performance as scale grows, which we attributed to a C-D gain problem. One way to correct for the ID is to increase target distance as the scale grows and this led us to consider the following computational model:

$$ID_d = \log_2\left(\frac{D_m + d \cdot \sqrt{\frac{4.15}{CDGain} - 1} \cdot D_m}{W_m} + 1\right) \quad (6)$$

where  $d$  is a constant and  $4.15/CDGain$  is relative gain (4.15 is the gain for  $SCALE = 1$ ).

$ID_d$  improves the fit of the MotorMag data for  $SCALE \geq 4$ , as shown in Figure 17 ( $r^2 = 0.955$  for a best  $d = 0.24$ ). This model also yields an  $AIC = 370$  as opposed to  $AIC = 410$  for the standard Fitts regression. Again our model does slightly better than Welford (Equation 1,  $r^2 = 0.920$ ,  $AIC = 383$ ) and is comparable to Oel et al ( $r^2 = 0.931$ ,  $AIC = 370$ ). We can also use a fourth degree of freedom to combine  $ID_c$  with  $ID_d$ :

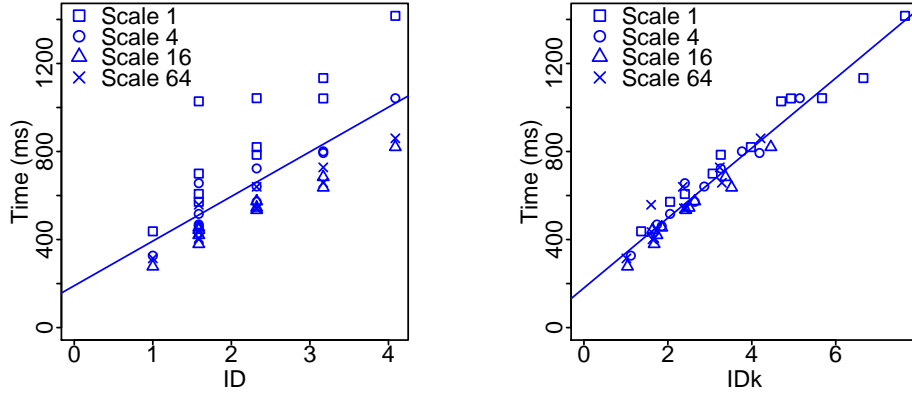
$$ID_{c,d} = \log_2\left(\frac{D_m + d \cdot \sqrt{\frac{4.15}{CDGain} - 1} \cdot D_m}{W_m - c} + 1\right) \quad (7)$$

and fit all MotorMag data including  $SCALE = 1$ . We obtain  $r^2 = 0.941$ ,  $AIC = 496$  with  $c = 0.67$  dots and  $d = 0.28$  vs.  $r^2 = 0.914$ ,  $AIC = 501$  for Oel et al.

Note that previous Fitts' Law modifications involving C-D gain were to correct for high C-D gains, not for very low ones [Casiez et al. 2008]. For example, the formula from [Johnsgard 1994] is  $ID = \log_2(D \cdot CDGain/W + 1)$ . It gives very low indices of difficulty for very low gains, and therefore obviously does not fit our data.

### 5.3. Zooming

Now we analyze all Zoom tasks. Zoom tasks, as well as ZoomBis and FullMag tasks, involve a wide range of target sizes, including targets which are *small in both the motor and visual spaces*, and all have a constant C-D gain of 4.15. The left plot in Figure 18 shows the standard Fitts regression with the Zoom method (see also Table III). The mediocre fit ( $r^2 = 0.534$ ) is essentially caused by the condition  $SCALE = 1$ . Removing this condition improves the fit ( $r^2 = 0.825$ ), which is not surprising since this leaves targets that are magnified in *both* the visual and motor space and we get closer to regular Fitts' Law tasks.

Fig. 18. MT as a function of  $ID$  and  $ID_{k=0.92}$  for Zoom.Table III. Adjusted  $r^2$  and  $AIC$  for the various models considered for Zoom.

|               | All SCALE |       | SCALE $\geq 4$ |       |
|---------------|-----------|-------|----------------|-------|
|               | $r^2$     | $AIC$ | $r^2$          | $AIC$ |
| $ID$          | 0.534     | 577   | 0.825          | 377   |
| $ID_c$        | 0.898     | 513   | 0.933          | 352   |
| $ID_k$        | 0.951     | 483   | 0.921          | 356   |
| Welford Eq. 1 | 0.787     | 544   | 0.858          | 372   |
| Oel et al.    | 0.943     | 486   | 0.959          | 332   |
| $ID_e$        | 0.227     | 600   | 0.768          | 387   |
| $ID_{ec}$     | 0.663     | 565   | 0.940          | 344   |

If we keep SCALE = 1 but use Welford’s “tremor” model instead we also improve the fit, with  $r^2 = 0.898$  and  $c = 0.89$ . The model yields an  $AIC = 513$ , as opposed to  $AIC = 577$  for the standard Fitts regression (all SCALE). This confirms that Welford’s “tremor” model, which was previously shown to capture small-scale motor tasks, also scales well to large-scale motor tasks (where it reduces to Fitts’ Law).

However, a surprising finding is that for SCALE  $\geq 4$ , Welford’s “tremor” model yields a higher “tremor” constant  $c = 2.44$  dots ( $r^2 = 0.933$  instead of  $r^2 = 0.877$  for  $c = 0.89$  dots). This suggests that  $c$  might actually increase with  $D_m$  or with  $W_m$ . We considered several computational models (e.g.,  $c$  linear in  $W_m$  or  $\sqrt[3]{W_m}$  with various powers) and obtained the best results with the following one:

$$ID_k = \log_2\left(\frac{D_m}{W_m - k \cdot \log_2(W_m + 1)} + 1\right) \quad (8)$$

With  $k = 0.92$ , we obtain  $r^2 = 0.951$ ,  $AIC = 483$  for all scales (see Figure 18) which is better than  $ID_c$  and again similar to Oel et al. This model also improves the fits for individual scales, which suggests that Welford’s “tremor” model can be refined by allowing the “constant”  $c$  to vary with  $W$ .

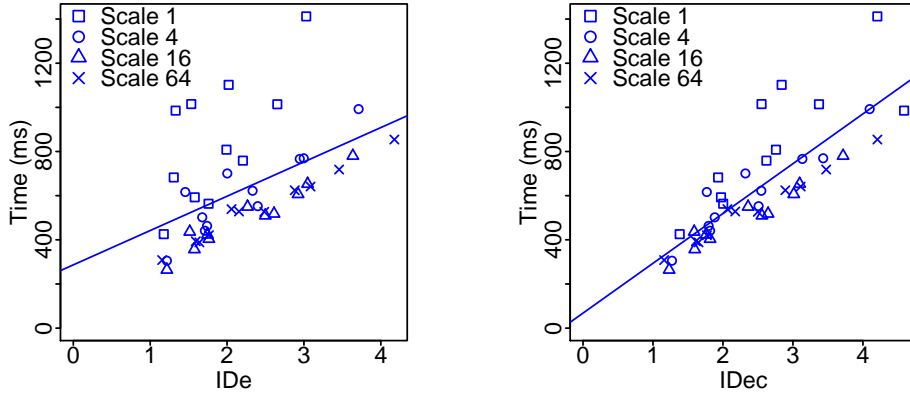
We obtained very similar results for ZoomBis. For FullMag we also obtain relatively good fits, yet not as good. A reason for this is that the  $ID_c$  and  $ID_k$  models do not account for the degradation we measured for scale 64.

#### 5.4. Effective Widths

The idea of effective width, introduced by Crossman in his unpublished doctoral dissertation, is to perform an adjustment for accuracy when error rates (or target utilization) vary across conditions:

Table IV. Average value of effective width ( $W_e$ ) as a function of motor width ( $W_m$ ) for Zoom.

| $W$   | 1    | 2    | 4    | 8    | 16    | 32    | 64    | 128   | 256   | 512    |
|-------|------|------|------|------|-------|-------|-------|-------|-------|--------|
| $W_e$ | 2.09 | 2.83 | 4.73 | 8.27 | 17.38 | 31.81 | 62.72 | 99.09 | 172.9 | 290.19 |

Fig. 19. MT as a function of  $ID_e$  and as a function of  $ID_e c$  ( $c = 1.31$ ) for Zoom.

target widths are corrected so that the data yields an uniform error rate of 4%. We refer the reader to [Fitts and Peterson 1964; Soukoreff and MacKenzie 2004] for the methodological details.

Computing Fitts' regressions from uncorrected widths and from effective widths are both valid, and the appropriate method depends on the goal [Zhai et al. 2004]. Using uncorrected widths – as we did so far – is useful when one needs to reliably assess actual user pointing times in user interfaces, accounting for natural biases in target utilization, and to assess errors separately. Using effective widths instead usually deteriorates the fit but is strongly recommended when one needs to interpret the  $a$  and  $b$  constants, e.g., in order to compare the performance of different input devices [Soukoreff and MacKenzie 2004].

Note that we cannot rely on adjustment for accuracy alone to improve the bad fit we obtained with Fitts' law. Effective widths remove biases in target utilization by enlarging error-prone targets and shrinking underutilized targets while leaving  $MT$ s intact. Our tiny targets yield unusually high error rates (see Figure 11) and will hence be significantly enlarged (see Table IV). However, since  $MT$ s are also unusually high – instead of being unusually low in accordance to the speed-accuracy trade-off principle – the data points will be moved even farther away from the regression line.

Indeed, redoing our standard Fitts analysis ( $ID$ ) with effective widths and effective distances ( $ID_e$ ) does not improve the fit. For instance, the left plot in Figure 19 shows the  $ID_e$  regression with the Zoom method (see also Table III). We obtain  $r^2 = 0.227$  and  $AIC = 600$ , as opposed to  $r^2 = 0.534$  without effective width correction<sup>12</sup> (Figure 18 left). Removing the smallest targets (using  $SCALE \geq 4$ ) yields  $r^2 = 0.768$ ,  $AIC = 387$ , instead of  $r^2 = 0.825$  with uncorrected widths. Thus, the fit deterioration with  $ID_e$  is mainly caused by the smallest targets: since they are both abnormally longer to acquire *and* more error-prone compared to larger-scale pointing tasks of the same  $ID$ , an adjustment for accuracy reinforces their departure from Fitts' Law.

We have also tested the  $ID_c$  “tremor” model on effective widths. For Zoom and all SCALE we obtained  $r^2 = 0.663$  and  $AIC = 565$  for a best  $c = 1.31$  (see right of Figure 19). Thus, using  $ID_c$  rather than  $ID$  yields a non-negligible improvement when fitting corrected widths, a result consistent

<sup>12</sup>Since AIC values cannot be compared on different datasets, we provide them for effective widths only.

Table V. Adjusted  $r^2$  and  $AIC$  for the  $ID$  and  $ID_e$  models for all methods, but Zoom (see Table III).

|                       | All SCALE |       | SCALE $\geq 4$ |       |
|-----------------------|-----------|-------|----------------|-------|
|                       | $r^2$     | $AIC$ | $r^2$          | $AIC$ |
| VisualMag             |           |       |                |       |
| <i>ID</i>             | 0.767     | 560   | 0.799          | 416   |
| <i>ID<sub>e</sub></i> | 0.446     | 600   | 0.532          | 455   |
| MotorMag              |           |       |                |       |
| <i>ID</i>             | 0.807     | 548   | 0.824          | 370   |
| <i>ID<sub>e</sub></i> | 0.603     | 579   | 0.747          | 420   |
| ZoomBis               |           |       |                |       |
| <i>ID</i>             | 0.421     | 587   | 0.816          | 369   |
| <i>ID<sub>e</sub></i> | 0.152     | 604   | 0.739          | 381   |
| FullMag               |           |       |                |       |
| <i>ID</i>             | 0.572     | 572   | 0.870          | 366   |
| <i>ID<sub>e</sub></i> | 0.291     | 593   | 0.664          | 398   |

with what we had previously obtained without width correction. This improvement is not as strong as without width correction, but when fitting the data SCALE  $\geq 4$ , the  $ID_c$  model yields similar fits on corrected ( $r^2 = 0.940$ ,  $AIC = 344$ ,  $c = 3.10$ ) and uncorrected widths ( $r^2 = 0.933$ ,  $c = 2.44$ ).

Note that the use of the effective widths yields larger values of  $c$ , in accordance with the fact that the departure from Fitts' Law is larger and  $ID$ s therefore need a stronger correction. An increase in  $c$  might also be consistent with what would have been observed had all subjects achieved an error rate of 4%: a "tremor" interpretation (see section 6.4) suggests that subjects avoided target edges in order to minimize errors due to uncontrollable "tremor", effectively reducing usable target size. The need to achieve a lower error rate (4%) would have probably required them to leave an even larger "safety margin" inside targets. It is possible, however, that tasks involving the smallest targets would have been impossible to perform with an error rate of 4%, no matter how careful the subjects were. This is suggested by the fact that the "tremor" constant computed from effective widths ( $c = 1.31$  dots) is larger than the smallest target ( $W_m = 1$  dot).

Table V provides the fit statistics for all scale methods other than Zoom. Results for ZoomBis and FullMag are similar to Zoom. For VisualMag, using effective widths also leads to a fit deterioration that is not as severe as with Zoom, but restricting the data to SCALE  $\geq 4$  does not really improve the fit as it was the case for Zoom. For MotorMag, there is still a deterioration but it is the least severe.

## 6. DISCUSSION

Here we build upon our findings to address the three questions mentioned in the introduction. We then discuss the limits of our study and possible future work.

### 6.1. There is a Small-Scale Effect

Our experiment confirms previous results and intuitions about a small-scale effect on mouse pointing tasks. This is shown by Zoom and ZoomBis: at high scales, we have normal pointing tasks. At scale 1, however, we have targets that are small in all respects, i.e., typical "small targets" on desktop computers. Task  $ID$ s do not change. Still, we observe a clear drop in performance as scale approaches 1. This naturally yields a bad fit with Fitts' law.

The drop in performance is progressive yet very fast: it starts to be observable from scale 16 (targets of 16-256 pixels<sup>13</sup>) to scale 4 (targets of 4-32 pixels). At scale 1 (targets of 1-8 pixels), there are more than twice as many errors, with a time almost twice as long. Since we used high-end I/O hardware and controlled for pointer occlusion, these are likely underestimates of what we would obtain on typical computer.

<sup>13</sup>All figures are given according to our experimental conditions, i.e., 0.06 mm mouse dot size, 0.25 mm pixel size and a distance of 70 cm to the screen.

## 6.2. What Causes the Small-Scale Effect?

Our experiment suggests that — at least for mouse pointing — the causes are both visual and motor. Quantization does not seem to significantly impact performance.

*Visual Causes.* For very small targets, the primary causes for the small-scale effect are visual. Data from VisualMag provides strong evidence for this. Visual size has no effect when it is large enough. But we observe an important deterioration from scale 4 (targets of 4-32 pixels) to scale 1 (targets of 1-8 pixels): there are twice as many errors and a reduction in speed of about 25%. Our data suggests important differences for targets up to 4 pixels.

The exact origin of these visual problems is unclear, since we actually controlled for visual legibility (we use a sharp monitor and gave a vision test). However, visual legibility is probably not a binary property: participants might have spent more time processing visual information because of its unusually small size.

*Motor Causes.* We also confirmed the common intuition that motor accuracy is a cause of the small-scale effect. This is supported by our data on MotorMag. There is an observable deterioration from scale 4 (targets of 4-32 mouse dots) to scale 1 (1-8 dots). Data suggest problems start to arise at 4 dots. Since we controlled for the accuracy of electronic sensors, the problem is likely physiological or mechanical (this will be further discussed in Section 6.4).

The role of motor accuracy is further supported by the data from VisualMag and FullMag: we observed that motor+visual magnification performs slightly better than visual magnification alone, and the improvement goes up to scale 8 while it stops at scale 4 for visual magnification alone.

*Quantization.* The quantization present in small-scale pointing tasks seems to be a very secondary cause to the small-scale effect. This is suggested by our data on FullMag and Zoom. Recall the first method simply magnifies pixels and mouse dots, whereas the second one improves pointing resolution as scale increases. We found no significant difference between the two methods overall, except for scale 64 (targets from 64 to 512 pixels).

One possible explanation is that although quantization deteriorates the information normally provided by the cursor feedback, it also provides “visual snapping”. With snapping, the cursor is clearly either inside or outside the target. However, it is not clear whether snapping actually helps as we found dwell time to be higher for FullMag than for Zoom at scale 64. Another possibility is that the human visuo-motor system is able to infer the missing information by interpolating the mouse cursor’s location in space and time.

## 6.3. Findings on C-D gain

The goal of our study was not to examine the effects of C-D gain as this issue has been extensively investigated in the past and because, by definition, C-D gain *per se* cannot be a cause of scale effects (see Section 1.2). However, our study incidentally provided us with interesting findings on C-D gain which partly confirms and partly contradicts previous work.

Recent work suggests that the only limiting factor with very high C-D gains is limb precision [Casiez et al. 2008]. Our data is consistent with these findings, since strong visual magnification was not found to harm performance (VisualMag, C-D gains up to 266). It was also suggested that low C-D gains can be detrimental to performance, and this was confirmed by the degradation we observed with motor-space magnification (MotorMag, C-D gains below 1). However, it was postulated that the origins of the problems were a) clutching and b) an upper limit in motor speed [Casiez et al. 2008]. This was not confirmed by our study, since a) we controlled for clutching and b) the

Zoom condition did not exhibit any problem despite involving the exact same targets in the motor space.

The problem observed with very low gains might originate from asymmetric effects of stimulus-response incompatibility on pointing performance. Participants did not have problems performing accurate movements with high C-D gains, but they did have trouble performing large and fast movements when the visual feedback provided was very small, as we observed lower peak velocities with MotorMag than with Zoom. This could be a matter of training, as suggested by the learning effect we found on MotorMag. One might want to run a study with sufficient training to measure asymptotic performance with low C-D gains. However, asymptotic performance is only of limited relevance for pointing magnification techniques, which are primarily meant to be triggered temporarily [Ramos et al. 2007; Roudaut et al. 2008; Appert et al. 2010].

#### 6.4. Is there a Law of Movement for Small Targets?

One important lesson from our study is that small target acquisition is complex. It involves different phenomena, and there is probably no simple law for small target acquisition that accounts for both motor and visual problems.

Welford's model of human tremor, despite being largely ignored in HCI, was nevertheless quite successful at modeling motor problems when acquiring small targets. Not only it has a good predictive power but we also believe it is useful as a conceptual model because in contrast with Fitts' Law, it captures the non-linear nature of small target acquisition tasks as well as the intuitive notion of a "problematic size threshold" [Ren and Moriya 2000; Albinsson and Zhai 2003; Ramos et al. 2007].

Taking the term "tremor" literally, this model has a simple interpretation: one can think of an area cursor whose hot spot's location changes in unpredictable ways. In order to limit errors, users would have to bring the whole cursor inside the target – or part of it depending on the accuracy demands – thus effectively reducing the target's size. Note that in contrast with effective width [Soukoreff and MacKenzie 2004], such a reduction in target size cannot be directly measured from the spread of hits, but has to be measured by fitting Welford's model to experimental data obtained with different target widths.

Although the term "tremor" makes direct sense for devices such as laser pointers [Pavlovych and Stuerzlinger 2009], it is arguably a simplification of the reality in the case of computer mice, since the hand of healthy subjects holding a mouse and the mouse itself are rarely seen trembling. This is confirmed by the fact that we did not observe rapid reversals of the velocity profile despite having recorded kinematic data at 5 times the resolution of the standard mouse dot. A possibility is that noise is present but is not directly observable as tremor because it is filtered by the human limbs, by the mouse's inertia or by the mouse sensors.

There are several possible origins for this putative noise. One is constant neuromotor noise [Van Beers et al. 2004], not to be confounded with motor noise from the impulse variability model, which linearly increases with task scale [Van Galen and De Jong 1995]. Another candidate is dry friction between the mouse and its support: not only it can vary on an imperfect surface, but it can also make small corrective movements difficult to perform due to static friction being higher than kinetic friction, a problem known as stick-slip [Feeny et al. 1998; Richard and Cutkosky 2000]. Another possible cause of noise is involuntary mouse movement during clicks. But these are all conjectures and more work is clearly needed to understand the origin of the "tremor" constant in mouse pointing.

One difficulty with Welford's model is that  $c$  can vary across conditions. Indeed, we found  $c$  to be larger for Zoom than for VisualMag, which means that tremor increases with average task motor scale. We proposed a modified model where  $c$  depends on  $W_m$ , but this model is essentially

computational and does not explain why  $c$  varies across conditions. It could be that depending on the overall task demands, participants used different mouse grips [Langolf et al. 1976; Balakrishnan and MacKenzie 1997; Guiard et al. 1999] or adapted their level of motor noise by varying their muscle stiffness and/or the normal force they applied to the mouse [Van Galen and De Jong 1995]. A stronger pressure or firmer grip on the mouse might also have inhibiting effects on the stick-slip phenomenon.

Our motor magnification model is also mostly computational, and captures both visual legibility problems and low C-D gain issues. For low C-D gains we observed lower peak velocities, and it hence seemed natural to assume that  $D_m$  will penalize MT more than  $W_m$ . The good fit we obtained after correcting for  $D_m$  suggests that this was the case.

### 6.5. Limits of the Study

*Distance Range and IDs.* We chose to test a wide range of scales rather than large distances. The next step could be to restrict the scale range and explore larger distances in order to further validate and possibly refine our models.

*Intermediate Scales.* We did a sparse sampling of the scale range. For all scaling methods, we observed the largest changes between scales 1 and 4. One direction for further experimentation could be to study intermediate scales.

*Hybrid Scaling.* We only investigated a limited number of axes in the scale space (see Figure 4). Since we now know more about visual limits, we can choose a large enough visual scale (4) and then closely examine the effect of motor scale. Or we can choose a large enough motor scale and try to better understand the effect of visual scale.

*Acceleration.* Since we controlled for task scale in both the visual and the motor space, we had to use constant C-D gains: mouse acceleration would have made target sizes and distances in the motor space dependent on mouse speed. However, well-chosen C-D gain transfer functions might facilitate the acquisition of very small targets.

*Error Rate Models.* Since forcing our subjects to perform all tasks at a 4% error rate would have been artificial and likely impossible for the smallest targets, both movement times and errors must be modeled. So far we provided candidate models for movement times alone, but they still need to be complemented with models for error rates.

*Generalizability.* Our study involves 1-D tasks. With 2-D pointing tasks, some factors — especially quantization — might have behaved differently. Similarly, all our findings concern moused-based desktop computers and the results would have probably been different with, e.g., pen devices or touch-screens.

## 7. IMPLICATIONS FOR DESIGN

Our study shows that when targets are small in either the motor *or* in the screen space, pointing becomes problematic. That is, performance is much lower than with larger-scale tasks having the same ratio  $D/W$  (up to twice the movement time and error rates). Our data suggests that degradation in performance varies continuously with scale, so formally, there is no such thing as a “problematic size threshold”. But since this variation has an exponential form, one can provide rough indicative figures: the “problematic size threshold” is *1 to 4 mm* on a screen at a distance of *70 cm*, and *0.1 to 0.2 mm* in the motor space, for a computer mouse. These values have been measured on high-accuracy input and output devices and might be higher on lower-end hardware.

Naturally, the obvious guideline “make targets as big as you can” still applies. But the UI designer rarely has full control over widget size: for example, there might be no room left on the screen, or relative object size might encode meaningful information (such as in information visualization applications) or might be user-specified (such as in graphic authoring applications). For similar

reasons, dynamically expanding  $W$  (a target-aware pointing strategy) is a limited approach, despite having been heavily researched. Such a strategy is very efficient when targets are isolated but does not scale well to dense populations of targets. Target-aware techniques can also be distracting. It has been shown that performance with such techniques is often lower than what Fitts' Law predicts, i.e., users are not able to fully exploit the expansion of  $W$  [Mandryk and Gutwin 2008; Blanch et al. 2004].

We argue that when the size of clickable objects drops below the above thresholds, the GUI designer should first consider adding support for *zooming* rather than resorting to target-aware strategies. Zooming has several advantages: it is *often* relatively easy to implement, it *solves both visual and motor issues*, it preserves C-D gain, and also it preserves the relative size of objects on the screen. Zooming is especially relevant on dense populations of targets, such as in text editors, image processing applications and vector graphics applications. Since part of visual context is lost when zooming, the best zoom factor is the one that enlarges the smallest targets just above the sizes mentioned above (typically  $4\times$  for one-pixel targets).

There is more to zooming than meets the eye, and there are many ways of supporting it. The traditional approach is to use mouse- or keyboard-controlled zoom & pan. However, this type of technique has been designed for navigating into large documents, not for facilitating pointing. It has several drawbacks: it is difficult to control, and the user rapidly gets lost. For the sole purpose of target selection, a better approach might be to employ an automatic or *spring-loaded zooming* strategy, such as the ones recently proposed by [Ramos et al. 2007; Roudaut et al. 2008; Appert et al. 2010]. The design space of this type of technique has just started to be explored, and we believe that more research should be pursued in this direction.

Zooming combines visual magnification with motor magnification. Alternatively, *visual magnification* alone can also facilitate pointing when targets are only small on the screen. Some windowing systems have native support for visual magnification through screen magnifiers. Although screen magnifiers have been initially designed for low-vision users, our work suggests that users with normal vision might also benefit from them for selecting small targets. Local screen magnifiers such as magnifying lenses and fisheyes can be also useful, but the motion artifacts they produce have been found to be distracting [Gutwin 2002]. *Motor magnification* alone can also be used with low-accuracy input devices, and this simply amounts to reducing the device C-D gain (mouse acceleration allows the reduction of the C-D gain only when the mouse moves slowly [Casiez et al. 2008]). However, our data suggests it is advisable to always keep the *C-D gain above 1*.

Another good thing to know is that visual quantization does not harm. For example, one could hide the mouse cursor inside menus to improve text legibility, and rely on roll-over effects instead. Or instead of hiding the mouse cursor, one could make it smaller or less visible when selecting a menu item. Although the visual feedback of the roll-over effect is visually less smooth and more erratic than the mouse cursor, our study suggests this should not harm performance.

Another implication of our work is that Fitts' law does not always model pointing performance perfectly and for some special types of tasks alternatives are sometimes worth considering. When working with small targets, Welford's "tremor" model seems more adapted and might better help answer questions such as which scale factor — or pointing technique, or input device — is the best for a specific target layout. We also think it is a better conceptual framework for thinking about small target acquisition in general, since it accounts for the non-linear relationship between pointing scale and pointing difficulty.

## 8. CONCLUSION AND FUTURE WORK

Our goal was to better understand why small targets are abnormally difficult to acquire, a question that has been only partially addressed by previous work. We had subjects perform small target acquisition tasks on a desktop computer, and manipulated the scale of the tasks by various means. Our results confirm the existence of a small-scale effect that violates Fitts' law and whose causes are both visual and motor. One implication is that more research effort should be directed towards the design of target-agnostic pointing techniques based on zooming [Ramos et al. 2007; Roudaut et al. 2008].

We have also found our data to be consistent with Welford's tremor model of pointing, where IDs can be normalized by removing a constant from target widths in the motor space. We argue for the adoption of this model instead of Fitts' law when designing for small targets. This work is exploratory but proposes testable hypotheses and suggests directions for further investigations. These include testing larger distances and IDs and using a finer sampling of small scales. Two key questions that remain to be addressed concern the interpretation of Welford's tremor constant and the exact causes of visual problems.

Although our work focuses on mouse pointing, we believe that our approach and some of our findings can be relevant to small target acquisition with other pointing devices. There is some evidence for small scale effects on pen-based devices [Ren and Moriya 2000; Ramos et al. 2007]. Touch screens also seem to exhibit small-scale effects, even when "fat finger" and occlusion problems are eliminated [Sears and Shneiderman 1991]. The causes of constant error in touch input have recently started to be understood [Holz and Baudisch 2010] but still little is known about variable error. We hope that our work will encourage researchers to put Welford's tremor model to the test on these devices.

## ACKNOWLEDGEMENTS

Thanks to Theophanis Tsandilas, Tomer Moscovich, Audrey Girouard, Géry Casiez, James Eagan and our reviewers for their useful feedback on this article.

## REFERENCES

- ACCOT, J. AND ZHAI, S. 2001. Scale effects in steering law tasks. In *CHI '01: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 1–8.
- ALBINSON, P.-A. AND ZHAI, S. 2003. High precision touch screen interaction. In *CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 105–112.
- APPERT, C. AND BEAUDOUIN-LAFON, M. 2008. SwingStates: Adding state machines to Java and the Swing toolkit. *Software: Practice and Experience* 38, 11, 1149 – 1182.
- APPERT, C., CHAPUIS, O., AND PIETRIGA, E. 2010. High-precision magnification lenses. In *CHI '10: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 273–282.
- BALAKRISHNAN, R. 2004. "Beating" Fitts' law: virtual enhancements for pointing facilitation. *Int. J. Hum.-Comput. Stud.* 61, 6, 857–874.
- BALAKRISHNAN, R. AND MACKENZIE, I. S. 1997. Performance differences in the fingers, wrist, and forearm in computer input control. In *CHI '97: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 303–310.
- BEDERSON, B. B., HOLLAN, J. D., PERLIN, K., MEYER, J., BACON, D., AND FURNAS, G. W. 1996. Pad++: a zoomable graphical sketchpad for exploring alternate interface physics. *J. of Visual Languages and Computing* 7, 1, 3–32.
- BLANCH, R., GUIARD, Y., AND BEAUDOUIN-LAFON, M. 2004. Semantic pointing: improving target acquisition with control-display ratio adaptation. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 519–526.
- BURNHAM, K. AND ANDERSON, D. 1998. *Model Selection and Inference: A Practical Information-Theoretical Approach*. Springer.

- CASIEZ, G., VOGEL, D., BALAKRISHNAN, R., AND COCKBURN, A. 2008. The impact of control-display gain on user performance in pointing tasks. *Hum.-Comput. Interact.* 23, 3, 215–250.
- CHAPUIS, O., LABRUNE, J., AND PIETRIGA, E. 2009. Dynaspot: Speed-dependent area cursor. In *CHI '09: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 1391–1400.
- CHAPUIS, O. AND ROUSSEL, N. 2010. UIMarks: Quick graphical interaction with specific targets. In *UIST '10: Proceedings of the 23rd annual ACM symposium on User interface software and technology*. ACM, 173–182.
- COCKBURN, A. AND BREWSTER, S. 2005. Multimodal feedback for the acquisition of small targets. *Ergonomics* 48, 9, 1129–1150.
- COCKBURN, A. AND BROCK, P. 2006. Human on-line response to visual and motor target expansion. In *GI '06: Proceedings of Graphics Interface*. Canadian Information Processing Society, 81–87.
- COCKBURN, A. AND FIRTH, A. 2003. Improving the acquisition of small targets. In *Proceedings of the British Computer Society Conference on Human-Computer Interaction*. Springer, 181–196.
- FEENY, B., GURAN, A., HINRICHS, N., AND POPP, K. 1998. A historical review on dry friction and stick-slip phenomena. *Applied Mechanics Reviews* 51, 5, 321–342.
- FERNANDEZ, L., WARREN, W. H., AND BOOTSMA, R. J. 2006. Kinematic adaptation to sudden changes in task constraints during reciprocal aiming. *Hum. Movement Sci.* 25, 695717.
- FITTS, P. M. 1954. The information capacity of the human motor system in controlling the amplitude of movement. *J. Exper. Psych.* 47, 381–391.
- FITTS, P. M. AND PETERSON, J. R. 1964. Information capacity of discrete motor responses. *J. Exper. Psych.* 62, 2, 103–112.
- FURNAS, G. W. 1986. Generalized fisheye views. *SIGCHI Bull.* 17, 4, 16–23.
- GRAHAM, E. AND MACKENZIE, C. 1996. Physical versus virtual pointing. In *CHI '96: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 292–299.
- GROSSMAN, T. AND BALAKRISHNAN, R. 2005. The bubble cursor: enhancing target acquisition by dynamic resizing of the cursor's activation area. In *CHI '05: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 281–290.
- GUIARD, Y. 1993. On Fitts' and Hooke's laws: Simple harmonic movement in upper-limb cyclical aiming. *Acta Psychol.* 82, 139159.
- GUIARD, Y. 2009. The problem of consistency in the design of Fitts' law experiments: consider either target distance and width or movement form and scale. In *CHI '09: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 1809–1818.
- GUIARD, Y., BEAUDOUIN-LAFON, M., AND MOTTET, D. 1999. Navigation as multiscale pointing: extending Fitts' model to very high precision tasks. In *CHI '99: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 450–457.
- GUIARD, Y., BLANCH, R., AND BEAUDOUIN-LAFON, M. 2004. Object pointing: a complement to bitmap pointing in GUIs. In *GI '04: Proceedings of Graphics Interface*. Canadian Human-Computer Communications Society, 9–16.
- GUTWIN, C. 2002. Improving focus targeting in interactive fisheye views. In *CHI '02: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 267–274.
- HOLZ, C. AND BAUDISCH, P. 2010. The generalized perceived input point model and how to double touch accuracy by extracting fingerprints. In *CHI '10: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 581–590.
- JELLINEK, H. D. AND CARD, S. K. 1990. Powermice and user performance. In *CHI '90: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 213–220.
- JOHNSGARD, T. J. 1994. Fitts' law with a virtual reality glove and a mouse: Effects of gain. In *GI '94: Proceedings of Graphics Interface*. Canadian Information Processing Society, 8–15.
- KABBASH, P. AND BUXTON, W. A. S. 1995. The “prince” technique: Fitts' law and selection using area cursors. In *CHI '95: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM/Addison-Wesley, 273–279.
- KLINE, R. L. AND GLINERT, E. P. 1995. Improving GUI accessibility for people with low vision. In *CHI '95: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM Press/Addison-Wesley, 114–121.
- KOBAYASHI, M. AND IGARASHI, T. 2008. Ninja cursors: using multiple cursors to assist target acquisition on large screens. In *CHI '08: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 949–958.
- KOSTAKOS, V. AND O'NEILL, E. 2008. Size matters: performance declines if your pixels are too big or too small. arXiv pre-print 0804.3103.
- KVÅLSETH, T. O. 1980. An alternative to Fitts' law. *Bull. Psychonomic Soc.* 16, 5, 371–373.

- LANGOLF, G., CHAFFIN, D., AND FOULKE, J. 1976. An investigation of Fitts' law using a wide range of movement amplitudes. *J. Mot. Behav.* 8, 113–128.
- MACKAY, W. E., APPERT, C., BEAUDOUIN-LAFON, M., CHAPUIS, O., DU, Y., FEKETE, J.-D., AND GUIARD, Y. 2007. Touchstone: exploratory design of experiments. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 1425–1434.
- MACKENZIE, C., MARTENIUK, R., DUGAS, C., LISKE, D., AND EICKMEIER, B. 1987. Three-dimensional movement trajectories in Fitts' task: implications for control. *The Quarterly journal of experimental psychology. A. Human experimental psychology* 39, 3, 629–647.
- MACKENZIE, I. S. 1991. Fitts' law as a performance model in human-computer interaction. Ph.D. thesis, University of Toronto.
- MACKENZIE, I. S. 1992. Fitts' Law as a research and design tool in human-computer interaction. *Hum.-Comput. Interact.* 7, 91–139.
- MANDRYK, R. L. AND GUTWIN, C. 2008. Perceptibility and utility of sticky targets. In *GI '08: Proceedings of graphics interface*. Canadian Information Processing Society, 65–72.
- MESSING, D. S. AND DALY, S. 2002. Improved display resolution of subsampled colour images using subpixel addressing. In *ICIP (1)*. 625–628.
- OEL, P., SCHMIDT, P., AND SCHMITT, A. 2001. Time prediction of mouse-based cursor movements. In *BHCI-IHM '01: People and Computers XV - Interaction without frontiers (Joint proceedings of HCI 2001 and IHM 2001)*. Springer, 37–40.
- PARKER, J. K., MANDRYK, R. L., AND INKPEN, K. M. 2005. Tractorbeam: seamless integration of local and remote pointing for tabletop displays. In *GI '05: Proceedings of Graphics Interface*. Canadian Human-Computer Communications Society, 33–40.
- PAVLOVYCH, A. AND STUERZLINGER, W. 2009. The tradeoff between spatial jitter and latency in pointing tasks. In *EICS '09: Proceedings of the 1st ACM SIGCHI symposium on Engineering interactive computing systems*. ACM, 187–196.
- PLAMONDON, R. AND ALIM, A. 1997. Speed/accuracy trade-offs in target-directed movements. *Behavioral and Brain Sciences* 20, 2, 279–349.
- RAMOS, G., COCKBURN, A., BALAKRISHNAN, R., AND BEAUDOUIN-LAFON, M. 2007. Pointing lenses: facilitating stylus input through visual-and motor-space magnification. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 757–766.
- REN, X. AND MORIYA, S. 2000. Improving selection performance on pen-based systems: a study of pen-based interaction for selection tasks. *ACM Trans. Comput.-Hum. Interact.* 7, 3, 384–416.
- RICHARD, C. AND CUTKOSKY, M. 2000. The effects of real and computer generated friction on human performance in a targeting task. In *Proceedings of the ASME IMECE 2000 Symposium on Haptic Interfaces for Virtual Environments and Teleoperator Systems*. ASME, 1101–1108.
- ROUDAUT, A., HUOT, S., AND LECOLINET, E. 2008. Taptap and magstick: improving one-handed target acquisition on small touch-screens. In *AVI '08: Proceedings of the working conference on Advanced visual interfaces*. ACM, 146–153.
- SCHEDLBAUER, M. 1997. A survey of human cognitive and motor performance models. Tech. rep. <http://teaching.cs.umd.edu/~heines/techrpts/Papers/TR2007-001.pdf>.
- SEARS, A. AND SHNEIDERMAN, B. 1991. High precision touchscreens: design strategies and comparisons with a mouse. *Int. J. Man-Mach. Stud.* 34, 4, 593–613.
- SOUKOREFF, R. W. AND MACKENZIE, I. S. 2004. Towards a standard for pointing device evaluation: Perspectives on 27 years of Fitts' law research in HCI. *Int. J. Hum.-Comput. Stud.* 61, 6, 751–789.
- VAN BEERS, R., HAGGARD, P., AND WOLPERT, D. 2004. The role of execution noise in movement variability. *J. Neurophysiol.* 91, 2, 1050–63.
- VAN GALEN, G. P. AND DE JONG, W. P. 1995. Fitts' law as the outcome of a dynamic noise filtering model of motor control. *Human Movement Science* 14, 4-5, 539–572.
- WELFORD, A. T., NORRIS, A. H., AND SHOCK, N. W. 1969. Speed and accuracy of movement and their changes with age. *Acta Psychol.* 30, 3–15.
- WOBROCK, J. O., CUTRELL, E., HARADA, S., AND MACKENZIE, I. S. 2008. An error model for pointing based on Fitts' law. In *CHI '08: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 1613–1622.
- WORDEN, A., WALKER, N., BHARAT, K., AND HUDSON, S. 1997. Making computers easier for older adults to use: area cursors and sticky icons. In *CHI '97: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 266–271.

- ZHAI, S., HUNTER, M., AND SMITH, B. A. 2000. The metropolis keyboard - an exploration of quantitative techniques for virtual keyboard design. In *UIST '00: Proceedings of the 13th annual ACM symposium on User interface software and technology*. ACM, 119–128.
- ZHAI, S., KONG, J., AND REN, X. 2004. Speed-accuracy tradeoff in Fitts' law tasks: on the equivalency of actual and nominal pointing precision. *Int. J. Hum.-Comput. Stud.* 61, 6, 823–856.
- ZHAI, S., MILGRAM, P., AND BUXTON, W. 1996. The influence of muscle groups on performance of multiple degree-of-freedom input. In *CHI '96: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 308–315.

Received May 2009; revised June 2010, February 2011; accepted February 2011