

# Interaction with Magic Lenses: Real-World Validation of a Fitts' Law Model

Michael Rohs<sup>†\*</sup>, Antti Oulasvirta<sup>°</sup>, Tiia Suomalainen<sup>°</sup>

<sup>†</sup>Deutsche Telekom Laboratories, TU Berlin

<sup>\*</sup>Ludwig-Maximilians-Universität München, Munich, Germany

<sup>°</sup>Helsinki Institute for Information Technology HIIT, Aalto University and University of Helsinki

## ABSTRACT

Rohs and Oulasvirta (2008) proposed a two-component Fitts' law model for target acquisition with magic lenses in mobile augmented reality (AR) with 1) a physical pointing phase, in which the target can be directly observed on the background surface, and 2) a virtual pointing phase, in which the target can only be observed through the device display. The model provides a good fit ( $R^2=0.88$ ) with laboratory data, but it is not known if it generalizes to *real-world* AR tasks. In the present outdoor study, subjects ( $N=12$ ) did building-selection tasks in an urban area. The differences in task characteristics to the laboratory study are drastic: targets are three-dimensional and they vary in shape, size, z-distance, and visual context. Nevertheless, the model yielded an  $R^2$  of 0.80, and when using effective target width an  $R^2$  of 0.88 was achieved.

## Author Keywords

Target acquisition, magic lens pointing, Fitts' law, human-performance modeling, field experiment, augmented reality.

## ACM Classification Keywords

H.5.2 [Information Interfaces and Presentation]: User Interfaces – input devices and strategies, theory and methods.

## General Terms

Experimentation, Human Factors.

## INTRODUCTION

Mobile augmented reality (AR) improves information navigation on handheld devices by overcoming the limitations of display size. Particularly, *magic lens pointing*, or pointing through the movable camera-view of the device [6], is a promising kind of interaction, because it allows large-scale information presentation with private and up-to-date information on a personal display. While the early prototypes were marker-based and mainly used in close-range interaction, the present-day AR allows markerless recognition of objects and can therefore be utilized for a wider range of

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.

Copyright 2011 ACM 978-1-4503-0267-8/11/05...\$10.00.

visual content, such as real-world scenes (Figure 1), as their background [8]. We use the term AR here in a broad way that encompasses real-world interactions using a camera-display unit to aim at targets in the environment.



Figure 1. Magic lens pointing over a real-world scene. The image illustrates target selection at one of the seven sites in the study with targets marked in red (not visible to the users).

In HCI, camera-based interaction with mobile AR interfaces has been analyzed as an instance of target acquisition performance that consists of rapid precise movements towards a point target or a spatially extended target. According to Fitts' law [1], the duration of such movements ( $t$ ) is systematically dependent on the distance to the target ( $D$ ) and the target's width ( $W$ ) as follows:

$$t = a + b \log_2(D/W + 1) \quad (1)$$

Wang et al. [9] show that Fitts' law is applicable to target acquisition tasks in which the phone UI is controlled using optical flow motion sensing. Rohs and Oulasvirta [6] propose a modified Fitts' law model for target acquisition of external targets that fits data better than the standard model. The modified model supposes two phases in sensorimotor control: 1) a physical pointing phase, in which the target can be directly observed on the background surface, and 2) a virtual pointing phase, in which the target can only be observed through the display. Consequently, it needs a parameter for the size of the display ( $S$ ):

$$t = a + b \log_2(D/S + 1) + c \log_2(S/2/W + 1) \quad (2)$$

Applied to the laboratory data of [6] this model achieves a fit of  $R^2=0.88$ , the original model (Eq. 1) only  $R^2=0.57$ .

However, these models, and all models of AR we know of, have been developed based on data acquired in rigorously controlled laboratory conditions where subjects are standing in front a surface (e.g., a large display) on which targets appear. It is an open question if the model generalizes to AR in the real world where targets 1) can be three-dimensional and 2) vary drastically in shape, size, z-distance, and visual context. Moreover, sequential selections can take place between targets with *different* characteristics. Because even seemingly small changes in task conditions can affect the performance of a Fitts' law task [1,2], the question of generalizability is best addressed empirically.

Therefore, to assess if the model of Eq. 2 generalizes to the abovementioned conditions, this paper presents an experiment using real-world targets. In the study, subjects were taken to balconies, bridges, parking lots, and streets to do reciprocal selection of buildings (see Figure 1). As we will show later, the model of Rohs and Oulasvirta [6] generalizes surprisingly well to this real-world task, achieving an  $R^2$  of 0.80, but increasing to  $R^2$  of 0.88 when effective target width [7] is used. We identified an “averaging behavior” that may explain this improvement.

#### MAGIC LENS POINTING: APPLYING THE MODEL OF ROHS AND OULASVIRTA TO REAL-WORLD SCENES

In the benchmark model [6], pointing consists of two phases. The first phase is denoted as *physical pointing*: The target can be directly observed in the physical world. At some point during the movement towards the target, the target falls below the display and can no longer be observed directly, but only through the magic lens. With a screen width of  $S$ , the split point is located at a distance of  $S/2$ . At this point, the second phase—*virtual pointing*—begins: The target can now only be observed through the display. As soon as a target falls below the lens, the characteristics of the *mediation* by the camera-display-unit come into play, such as delay and display update rate.

For the outdoor situation with buildings and landmarks as targets, the model has to be slightly adapted. The real sizes of the objects do not help in characterizing the situation, because objects can be located at a wide range of z-distances from the user. A smaller object might thus cover a wider field of view than a larger object if it is located closer to the user. The angular size of a target from the viewpoint of the user is more meaningful. Hence, we measure  $D$  and  $W$  in angular degrees, as proposed in earlier work [4].

Another complication lies in the fact that target shapes can be arbitrary polygons (see Figures 1-3) rather than squares as typical in Fitts' law studies. We defined the centers of the real-world targets as the centroid of the polygon area. Since the targets are arbitrary polygons, the target width is not constant, but depends on the direction from which the target is approached. To handle this case for rectangular targets, we follow MacKenzie [5] by considering the extent

of the target along an approach vector through the center. But we extend this by defining the movement vector to lie on a line through the centers of the previous target and the next target as in Figure 2. The two intersection points of this line of movement and the next target's polygon outline define the angular width of the target. Thus, in the real-world pointing case  $W$  of next target does not only depend on that target alone, but also on the movement vector from which the target is approached. Grossman et al. [3] derive a probabilistic model for rectangular targets that describes 2D pointing from different directions.

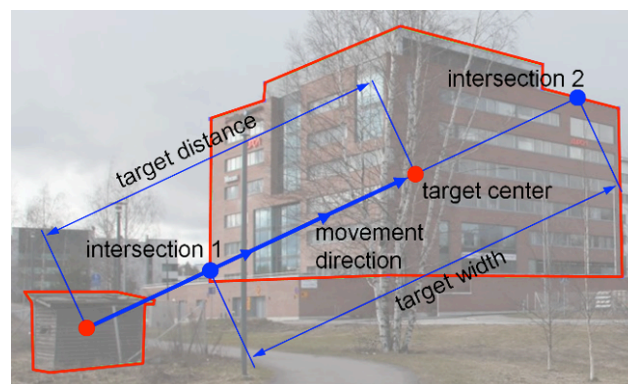


Figure 2. Target polygons are outlined in red. The movement axis goes through the target center points, from the left target to the right target. Target width depends on movement direction (intersection of movement vector and polygon).

#### EXPERIMENT

In the experiment, we utilized a non-conventional *reciprocal pointing task*, in which users have to move back and forth between two targets and the two targets have different properties (size, z-distance, shape, visual context). An interesting aspect of the experiment was to find out how users adapt to this situation. In particular because of widely differing widths, we wondered whether users would change their movement speed according to Fitts' law or whether they would apply some kind of averaging behavior. We therefore measured the effective target width  $W_e$ , which is computed as the width of the target based on the scatter of selection points around the target center point [7]. We also applied the model to averaged target widths, i.e., for each target pair we computed the average width and used this as the input to the model.

##### Participants

Twelve subjects (6 M, 6 F, age 21-26 years) were recruited from a local technical university. Seven were frequent users of the camera functionality on their phones. All subjects had normal or corrected-to-normal vision. Subjects were paid a small incentive for participation.

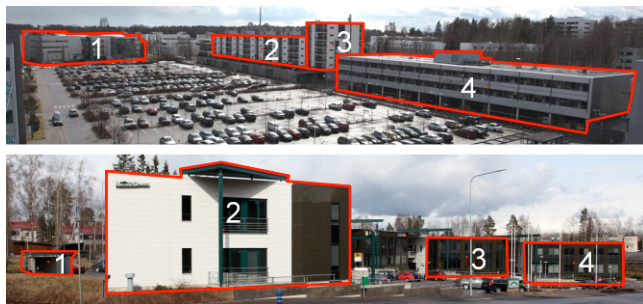
##### Experimental Platform

The experiment was conducted on a Nokia N95 8GB Symbian phone. The angular screen width of the phone camera is  $42.5^\circ$ . The test application shows a viewfinder image

with a crosshair in the center of the screen to facilitate precise selection. The subject has to press the joystick button to make a selection. Since we dealt with arbitrary outdoor scenes and changing lighting conditions, no attempt was made to automatically recognize the image in real-time. This would have led to a noticeable processing delay with a detrimental effect on pointing performance [10]. Therefore, upon selection, the last visible frame is stored as a PGM image together with meta data.

### Materials and Experimental Design

The main goal in choosing target stimuli was to have realistic variety in size, distance between target pairs. In terms of standard Fitts' law, we used index of difficulty (ID) values between 0.72 and 3.91 ( $ID = \log_2(D/W+1)$ ).



**Figure 3: Two scenes from the experiments' 7 sites. The four targets in each site are highlighted in red.**

The route consisted of 7 sites (see Figure 3). It was short enough for the experiment to last about one hour including transitions. Each site consisted of 4 targets, yielding 6 target pairs. With 2 directions in each target pair this results in 84 directed target pairs. The angular target distances were between  $6.8^\circ$  and  $74.8^\circ$  and the target widths  $2.3^\circ$ - $35.3^\circ$ . Target borders were chosen to be clear and easily perceived, not blocked by other buildings or vegetation. Car traffic directly in front of a target was avoided to minimize distraction. Selection always started from the left target in the target pair (see Figure 2). Timekeeping started on the first selection and the next item had to be selected as quickly and accurately as possible. We did not give special instructions beyond that (e.g. how to deal with elongated targets) and did not explicitly check for systematic distortions of selection centers. Within each target pair, 24 selections were made (interleaved 12 left-to-right, 12 right-to-left).

Half of the male and half of the female participants walked the route in opposite direction. Target pairs were run in different random order for each participant. Standing position and orientation at each site were fixed. The total number of selections per subject was 24 selections for practice; and 7 sites  $\times$  6 target pairs  $\times$  24 selections = 1008 for the actual experiment.

### Procedure

All subjects practiced selection of targets indoors before going outdoors. Before starting a selection of a target pair at

a site, the current targets were pointed out to the subject as a photo on paper where their boundaries were highlighted with red. Moreover, they were pointed at in the real world by the experimenter. To select a target, the subjects had to move the crosshair (centered on the screen) on top of the target on the display and press the phone's joystick button.

### Data Analysis

Each user selection resulted in capturing and storing an image on the device. Images were uploaded to a PC and analyzed off-line with a Java tool. We annotated the images, determining the distance between the target center and the point of selection, as well as whether the selection was successful. This data was used to compute the scatter around the target for  $W_e$  and the error rate.

To calculate the angular target distances and widths, a high-resolution image was taken with a DSLR camera at each site. The image was taken from the standing position of the participants and covered all four targets of a site. In addition, a calibration image of a brick wall from a known distance was taken, with the identical camera and lens configuration as with the high-resolution site images. The calibration image enabled us to compute the parameters to convert pixel coordinates in the image to angular coordinates in the real scene. Using this mapping, we computed the angles  $D$  between pairs of target center points. We verified the angular distances using a large-scale local map. To compute the target widths we entered the outlines of each target as polygons into the tool. From the outlines and the target center points, the tool computed the intersection points of the line of approach and the next target (Figure 2), which were then converted into the angular width  $W$  for that target pair.

### RESULTS

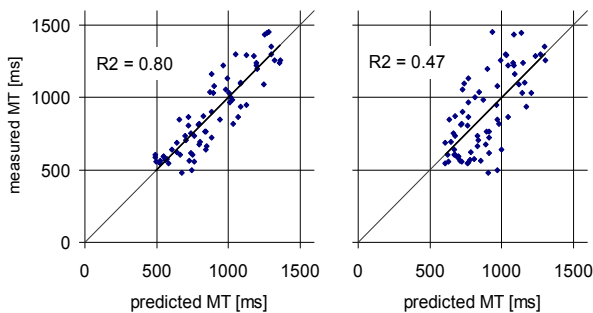
The experiment yielded 11952 data points (one subject completed only 6 sites). The overall error rate (i.e., selection not on target) was low (2%), the mean selection time was 885ms. Preprocessing revealed that two targets (from two different sites) were problematic, because the participants had trouble distinguishing them from their backgrounds. Consequently, the two targets were excluded from further analysis. Including these targets leads to  $R^2 = 0.71$  for our model, and  $R^2 = 0.42$  for standard Fitts' law.

Without the outlier targets, the fit of the measured movement times and the movement times predicted by our model (Eq. 2) is  $R^2 = 0.80$ . The fit for standard Fitts' law (Eq. 1) is  $R^2 = 0.47$ . The results are shown in Figure 4 as predicted time against measured time. Typically ID is shown on the x-axis. With the extended model, ID cannot be used directly. The estimated parameters (all times in milliseconds) for our model were ( $S = 42.5^\circ$ ):

$$t = 185 + 727 \log_2(D/S + 1) + 62 \log_2(S/2/W + 1) \quad (3)$$

and for standard Fitts' law were:

$$t = 447 + 220 \log_2(D/W + 1). \quad (4)$$



**Figure 4: Movement time (MT) prediction by our model (left,  $R^2 = 0.80$ ) and standard Fitts' law (right,  $R^2 = 0.47$ ).**

Even though Eq. 3 yields a better fit when compared to the standard Fitts' law, the fit ( $R^2=0.80$ ) is still relatively low compared to lab studies in the literature. We therefore analyzed whether the large variation in target sizes within a single target pair might have an influence. To this end, we looked at the *actual* speed-accuracy tradeoff the participants chose within a target pair by computing the effective target width  $W_e = 4.133\sigma$ . Here,  $\sigma$  represents the standard deviation of the end-point positions, which here are angular deviations from the target center. Using the effective widths, fit raises to  $R^2 = 0.88$  for our model and  $R^2 = 0.72$  for the standard model. The parameters are now estimated to  $t = -48+743\log_2(D/S+1)+144\log_2(S/2/W+1)$  for our model and  $t = 184+266\log_2(D/W+1)$  for standard Fitts' law.

The difference between actual and effective target width seems to be particularly large if the targets within a condition differ strongly in width. When selecting subsets of target pairs with restricted width ratio  $R^2$  increases. However, also the number of D,W-pairs decreases, leading presumably to a better fit.

The results regarding effective target width let us suspect that participants performed an “*averaging behavior*” within the reciprocal task, i.e., adapting the task for the average of the width of a condition's two targets. We therefore re-evaluated the data using the average (across users) target width. This led to an extremely high fit of  $R^2 = 0.96$  for our model and  $R^2 = 0.88$  for the standard Fitts' law model. Hence we conclude that the users indeed employed an averaging behavior. The parameters for the averaged widths are  $t = -69+750\log_2(D/S+1)+223\log_2(S/2/W+1)$  for our model and  $t = 250+341\log_2(D/W+1)$  for the standard model.

## SUMMARY AND OUTLOOK

The power of Fitts' law lies in its ability to simplify a complex phenomenon in a way that generalizes beyond immediate observations. The question is, whether models of performance in rigidly controlled laboratory conditions generalize to real-world conditions that may have different characteristics [2]. An essential part of research in this field should consist of validation of models in the real world. From this perspective, the results of the present study are highly promising. The laboratory model was found to gen-

eralize to a real-world task that is different in many respects, achieving an  $R^2$  of 0.80. When using effective target widths  $W_e$ , a fit of 0.88 was achieved which is identical to that of the laboratory study [6]. This is remarkable, when considering that the real-world targets had much more complex shapes, were embedded in a rich visual context, and were selected from a wide range of z-distances. Moreover, sequential selections took place between targets with different characteristics.

As future work we intend to investigate the effects of visual saliency as well as the effects of z-distance and perspective on real-world AR pointing tasks. These challenges will require a metric for the visual saliency of a target relative to its background, and we expect that z-distance and in particular the effects of perspective distortion caused by the users position will have a strong influence on performance.

## ACKNOWLEDGMENTS

We thank Antti Nurminen for help with image calibration. AO and TS were funded by the Tekes project Theseus2.

## REFERENCES

1. Fitts, P.M. The information capacity of the human motor-system in controlling the amplitude of movement. *J. Exp. Psychol.* 47 (1954), 381-391.
2. Gillan, D.J., Holden, K., Adams, S., Rudisill, M., and Magee, L. How should Fitts' law be applied to human-computer interaction? *Int. w.Comp.*, 4, 3, 1992, 291-313.
3. Grossman, T. and Balakrishnan, R. A probabilistic approach to modeling two-dimensional pointing. *ACM Trans. Comput.-Hum. Interact.* 12, 3 (2005), 435-459.
4. Kondraske, G. An angular motion Fitt's law for human performance modeling and prediction. *Proc. Annual Int. Conf. of the IEEE, vol. 1. Engineering in Medicine and Biology Society*, 1994, 307-308.
5. MacKenzie, I.S. and Buxton, W. Extending Fitts' law to two-dimensional tasks. *Proc. CHI 1992*, 219-226.
6. Rohs, M. and Oulasvirta, A. Target acquisition with camera phones when used as magic lenses. *Proc. CHI 2008*. ACM Press (2008), 1409-1418.
7. Soukoreff, R.W. and MacKenzie, I.S. Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research in HCI. *Int. J. Hum.-Comput. Stud.* 61, 6 (Dec. 2004), 751-789.
8. Wagner, D., Reitmayr, G., Mulloni, A., Drummond, T., and Schmalstieg, D. Pose tracking from natural features on mobile phones. *Proc. ISMAR 2008*. IEEE, 125-134.
9. Wang, J., Zhai, S., and Canny, J. Camera phone based motion sensing: interaction techniques, applications and performance study. *Proc. UIST 2006*. ACM, 101-110.
10. Ware, C. and Balakrishnan, R. Reaching for objects in VR displays: lag and frame rate. *ACM Trans. Comput.-Hum. Interact.* 1, 4 (Dec. 1994), 331-356.