Moving Ahead with Peephole Pointing: Modelling Object Selection with Head-Worn Display Field of View Limitations

Barrett Ens[1], David Ahlström[2], Pourang Irani[1]

[1]University of Manitoba
Winnipeg, Canada
{bens, irani}@cs.umanitoba.ca

[2]Alpen-Adria-Universität Klagenfurt
Klagenfurt, Austria
david.ahlstroem@aau.at

ABSTRACT

Head-worn displays (HWDs) are now becoming widely available, which will allow researchers to explore sophisticated interface designs that support rich user productivity features. In a large virtual workspace, the limited available field of view (FoV) may cause objects to be located outside of the available viewing area, requiring users to first locate an item using head motion before making a selection. However, FoV varies widely across different devices, with an unknown impact on interface usability. We present a user study to test two-step selection models previously proposed for 'peephole pointing' in large virtual workspaces on mobile devices. Using a CAVE environment to simulate the FoV restriction of stereoscopic HWDs, we compare two different input methods, direct pointing, and raycasting in a selection task with varying FoV width. We find a very strong fit in this context, comparable to the prediction accuracy in the original studies, and much more accurate than the traditional Fitts' law model. We detect an advantage of direct pointing over raycasting, particularly with small targets. Moreover, we find that this advantage of direct pointing diminishes with decreasing FoV.

Keywords

Head-worn display; HWD; HMD; field of view; FoV; peephole pointing; Fitts' law; modelling

1. INTRODUCTION

Head-worn display (HWD) technology has advanced rapidly over recent years, and powerful, self-contained systems such as Hololens[1] and Daqri[2] are now entering the market. Advanced sensing systems allow these devices to track their relative position, allowing rich and sophisticated user interfaces to be overlaid on the surrounding environment.

One limitation of current see-through HWD technology is the limited viewing region of the see-through display. Different devices use varying optical technologies that provide a field of view (FoV) ranging roughly from 20°[3] to 90°[4] in width. Researchers have found that FoV restrictions negatively influence spatial orientation and navigation [11], however little is known about the impact of FoV on object selection in spatial interfaces. Research has shown advantages of spatial interaction, using head and body motion, despite limitations on the available FoV [1, 3].

This research explores the effects of FoV size on virtual target selection. In a spatial user interface, any virtual objects that lie outside of the available FoV are cropped from view. To select a hidden object, the user must first turn their head until the object enters the viewing region, and then proceed with the given pointing mechanism. This two-step selection process is analogous to "peephole pointing", a term used to describe target selection on spatially aware mobile devices. Models for two-step selection [2, 13] (Figure 1a), proposed in the context of peephole pointing, have been shown to predict selection time more accurately than Fitts' law [4]. Later work has independently verified these models in the similar context of a handheld projector [7].

Fitts' law

$$MT = a + b\left(\log_2\left(\frac{A}{W}+1\right)\right) \quad (1)$$

CLB

$$MT = a + b\begin{pmatrix} n\log_2\left(\frac{A}{S}+1\right)+ \\ (1-n)\log_2\left(\frac{A}{W}+1\right)\end{pmatrix} \quad (2)$$

RO

$$MT = a + b\log_2\left(\frac{A}{S}+1\right) + c\log_2\left(\frac{S/2}{W}+1\right) \quad (3)$$

(a)

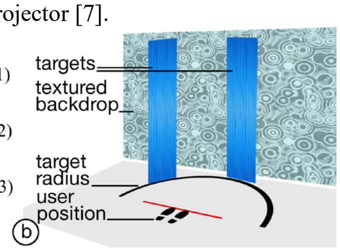

Figure 1. a) Equations for modelling pointing time, including two-component models created to model peephole pointing. b) The setup of our user study in a stereoscopic CAVE environment. Participants do a reciprocal pointing task between targets places on an imaginary arc.

We conduct a user study that emulates an HWD's FoV restriction in a CAVE environment (Figure 1b), to investigate whether these models retain their accuracy when head motion, rather than handheld device motion, is used to locate the target in the first step of the selection. We also intend to shed light on the effects of FoV width on target selection, to provide insights that may be useful for designing interfaces to work across a range of platforms with varying FoV widths. Unlike prior work with these models, we run our study using two different input techniques. We run a between-subjects study with raycasting and direct pointing, two potentially useful techniques for future HWD interfaces.

2. TWO-COMPONENT POINTING MODELS

Fitts' law [4] is an established model to estimate the average time required to select a target of width W at distance (amplitude) A. The movement time, MT, is predicted by the formula (1), as shown in Figure 1a, where a and b are experimentally determined constants. This model is commonly applied to 1D target selection tasks, but can be applied or adapted to selection in 3D virtual environments [5, 10, 15]. However, this model assumes that the target being selected is visible to the user. More recent models handle situations where the virtual workspace is larger than the available display space [8], such as with "peephole pointing" on a mobile device. Two similar models, shown in Figure 1a, were introduced

[1] http://www.microsoft.com/microsoft-hololens/en-us
[2] http://daqri.com/
[3] http://www.epson.com/MoverioBT200
[4] http://www.metavision.com/

independently: (2) by Cao, Li and Balakrishnan [2] (hereafter CLB), and (3) by Rohs and Oulasvirta [13] (hereafter RO).

Both models are derived from Fitts' law, but contain an additional parameter *S*, which represents the available screen size. Both models also give a sum of two components: the first component models the time required to search for an off-screen target until it becomes visible, and the second component represents the time needed to then select the target. In this paper, we generalize these using the term *two-component models.*

CLB tested their model by emulating a 1D peephole environment on a desktop computer, and found that targeting performance drops with a decrease in *S*. As *S* increases, the model converges to a standard Fitts' pointing task, however, their two-component model fits markedly better than Fitts' law when *S* is a variable.

RO tested their model using a spatially aware mobile device with a 2D selection task. Contrary to CLB, they only examined one value of *S* and found that Fitts' law is a good predictor in the "peephole" context, when the target exists only in the virtual workspace, but not for a "magic lens" context, when the actual target can be seen in the real world. Their model provided a good fit in both contexts.

Kaufmann and Ahlström [7] (hereafter KA) later revisited these models, in the context of a "spotlight" metaphor, where the virtual workspace is explored with a handheld projector. They verified both models, but found that they substantially outperform Fitts' law only when users do not have prior knowledge of the target location, and when *S* is sufficiently small. This explains to some extent the findings of RO, as they provided prior knowledge of the target layout. KA noted that Fitts' law is a good predictor within each individual value of *S*, and also found some interesting interaction effects, which we follow up on in our analysis.

Our study builds on these works by testing these models in a new context, where the search component is driven by head motion, independent of the selection technique. Head motion [6], direct pointing [17] and raycasting [9] have all been shown to follow Fitts' law, but it remains to be seen whether head motion combined with either selection technique can be predicted using existing two-component models. Beyond verifying these models, we also shed light on the effects of FoV limitation on target selection, and reveal how these effects vary between direct pointing and raycasting.

3. USER STUDY

We ran a study with 24 experienced computer users. All were males between 20 and 27 years old (mean 22.2, *SD* 1.7), and two were left-handed. Participants had varying previous experience with viewing 3D displays. Participation lasted approximately 1 hour.

We conducted our study in a CAVE environment to emulate the FoV limitation of a see-through HWD [3]. The CAVE combines a floor and wall display (Figure 1b) into a unified 3D image. Each stereo projector pair has a 1920×1080 resolution. Participants stood 1m from the wall, giving an apparent display width of roughly 130°. Textured targets were placed in front of a textured backdrop to assist stereo convergence [14]. Head and pointing motion were tracked using a high-precision, low-latency Vicon system.

Since HWD FoV is commonly measured in degrees, we designed our study to use model parameters *A*, *W* and *S* based on angles, measured in degrees of apparent width. Angular measures are commonly applied in immersive environments [12] and Fitts' law has been successfully expressed in terms of angular width [9]. Because the models are affected by the ratios between parameters, and not the absolute units, we expect that two-component models will predict angular units equally well to linear units.

3.1 Task, Selection Techniques & Study Design

Task: For purposes of comparison, we closely followed the study design of KA, and used a similar non-conventional, reciprocal 1D pointing task. Each sequence consists of two target selections, the first without prior knowledge of target location, and the second with. This knowledge is given by placing the second target in the same location as an initial starting target. Targets appear as bars the height of the display wall (Figure 1b). Targets of identical angular width *W* are placed at angular distance *A* (Figure 2a), however, each target appears only after the preceding target is selected to prevent accidental knowledge about their locations.

To begin a sequence, the start target must be located through the available viewing frusta (Figure 2a). The start target's placement to the left or right of centre informs the participant in which direction to search for the second target. This location is randomly chosen to be left or right, and is randomly offset to prevent predictability. The participant proceeds scanning until the first target is located, at which point they make a selection, then return to the initial start location to select the second target. Success is indicated with visual and audio feedback. Trials proceed whether or not a selection is successful, however, all targets must be selected successfully for a sequence to be valid. Invalid sequences are re-queued.

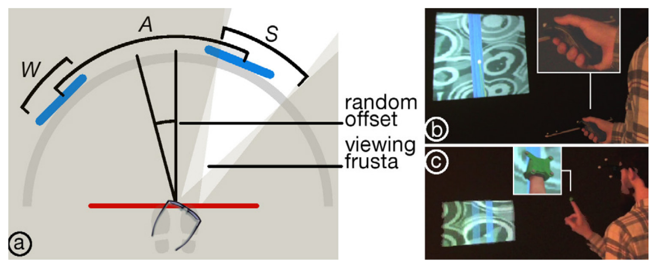

Figure 2. a) Our study layout, showing the relationships of the model parameters. A user making a selection using the b) raycast (*Ray*) and c) direct pointing (*Tap*) techniques.

Selection Techniques: We investigated two selection techniques for this study, raycasting (*Ray*) and direct pointing (*Tap*).

In the *Ray* technique, participants use a spatially tracked handheld mouse (Figure 2b). A virtual ray extends from the mouse and selection is trigger by pressing either top button. If the button is pressed when the ray is not intersecting the target area, then the selection is recorded as a miss.

In the *Tap* technique, the participant's dominant index finger is instrumented with tracking markers (Figure 2c). Selection is triggered when the finger crosses the target plane. If the finger crosses the plane outside of the target area, then the selection is recorded as a miss.

Targets are positioned 1m away from the participant for *Ray*, and 0.5m for *Tap*. This puts targets within reach for the *Tap* technique, but farther away in *Ray* to allow effective raycasting. In both techniques, the stereo frusta overlap at the target distance, so that any visible part of the target will be seen equally by both eyes.

Both of these techniques are "decoupled", in that pointing motion is independent from head motion. CLB's study compared a coupled cursor (fixed at the centre of the peephole) with a decoupled cursor (controlled by a secondary input device), and found the coupled technique to be faster. RO and KA studied coupled techniques only.

Study design:
We used a between-subjects design with four independent variables: target distance *A* (24, 48, 72, 96°), target width *W* (2, 4, 8, 16°), FoV width *S* (8, 16, 32, 64, 128°, SNR=No Restriction),

and prior knowledge of target position, *PK* (NoPK, PK). Each participant completed four blocks of 96 sequences. Each block contained one sequence for each *A-W-S* combination in a random order. Half of the participants used each technique (*Ray*, *Tap*). Participants were asked to select targets as quickly and accurately as possible, and were given adequate breaks between blocks. We collected data from a total of 10,242 sequences (20,484 selections).

3.2 Results

Errors: In 1,026 sequences, participants missed either one or both targets. In total 1,230 target misses were recorded (6% of all target selections). A Mann-Whitney U test showed that the number of missed selections did not differ between the two techniques (*Ray* 504, *Tap* 726; U=52.0, p=.266). As expected, and similar to what CLB and KA report, our analyses showed a significant effect of target width (Friedman test: $\chi^2_{(3,N=24)}$=65.78, p<.001) with an increasing number of misses as the target width decreases (*W16* 38, *W8* 95, *W4* 243, *W2* 854; Bonferroni adjusted post-hoc Wilcoxon Signed-Rank tests showed that all widths differed, all p's<.0083). Also in line with CLB and KA, we found no significant differences between the four amplitudes (*A24* 291, *A48* 306, *A72* 299, *A96* 334) or between selections with and without prior knowledge of the target location (*PK* 634, *NoPK* 596). Contrary to CLB and KA, who report significant effects of their peephole size, we did not find any significant effect of FoV width. However, we see similar trends to those reported by CLB and KA: fewer misses with smaller FoV widths and more misses with the two largest widths (*S8* 180, *S16* 177, *S32* 179, *S64* 172, *S128* 227, *SNR* 295). CLB speculate that with a small peephole the user is very careful, whereas with a large one the user may be too confident and relaxed, thus making more mistakes. Our data support this speculation and suggest that this may be the case in HWD-situations too.

***Movement time*:** The following analyses are based on error free sequences only (18,432 selections in 9,216 sequences). The movement times (*MT*) were right skewed and we performed a logarithmic transformation (which resulted in distributions close to normal) before analyzing the data. A mixed ANOVA with *technique* as between-subject factor and *A*, *W*, *S*, *and PK* as within-subject factors showed the same significant main effects for *W* ($F_{3,66}$=233.0, p<0.001, η^2=0.91), *A* ($F_{3,66}$=476.4, p<0.001, η^2=0.96), *S* ($F_{5,110}$=1064.9, p<0.001, η^2=0.98) and *PK* ($F_{1,22}$=29.5, p<0.001, η^2=0.57) as CLB and KA report.

As expected from Fitts' law literature, and confirming CLB's and KA's results, we found that close targets took shorter time to select than targets further away: *A24* 1.09s, *A48* 1.31s, *A72* 1.50s, *A96* 1.72s (we use the geometric mean, i.e., the antilog of the mean of the log-transformed data to report time measures and we use Bonferroni adjustment for post-hoc tests. Post-hoc: all *A*s differed, all p's<0.0083). Likewise, small targets took longer time to select than larger targets: *W2* 1.72s, *W4* 1.50s, *W8* 1.29s, *W16* 1.11s (post-hoc: all *W*s differed, all p's<0.0083).

Our results regarding prior target location knowledge and influence of FoV width on target selection performance with head motion confirm CLB's and KA's peephole findings; We also found better performance in selections with prior knowledge than without (*PK* 1.29s, *NoPK* 1.49s) and that performance increased with increasing FoV width (*S8* 2.24s, *S16* 1.80s, *S32* 1.46s, *S64* 1.18s, *S128* 1.03s, *SNR* 0.99s, post-hoc: all *S*s differed, all p's<0.0033).

We did not find a main effect for *technique* (*Ray* 1.43s, *Tap* 1.34s) but did notice a significant *technique*×*W* interaction ($F_{3,66}$=9.57, p<0.001, η^2=0.30), plotted in Figure 3a. We attribute *Ray*'s low performance with small targets to the well-known jitteriness of raycasting techniques [14], which makes selecting small targets difficult. Technique also interacted with *S* ($F_{5,110}$=8.08, p<0.001, η^2=0.27), as shown in Figure 3b. It seems *Tap* loses its advantage over *Ray* as FoV becomes more restricted. We suspect this is because participants were more careful with a narrower FoV, as noted by the decrease in errors. However, we are surprised to learn that *Ray* does not perform even worse with decreasing *S*, as the virtual ray becomes increasingly cropped by the view frusta.

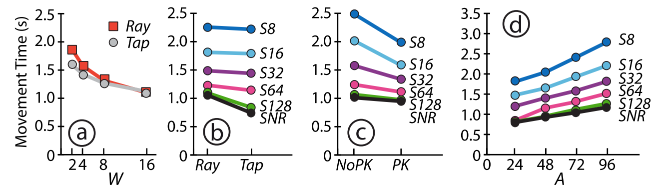

Figure 3. a) *MT*×*W*, b) *Technique*×*S*, c) *PK*×*S*, and d) *A*×*S* interactions found in our user study.

We observe a significant *PK*×*S* interaction ($F_{5,66}$=9.57, p<0.001, η^2=0.30) in Figure 3c, also reported by CLB and KA: the advantage of prior location knowledge is more beneficial with small FoV widths and almost disappears with large widths. A partial reason for this, as pointed out by KA, is that, with the largest FoV widths, some of the targets were visible from the onset of the movement selection (*A24* with *S64*, *S128* and *SNK*, *A48* with *S128* and *SNK*, and *A72* and *A96* with *SNR*). Intuitively, knowledge about target location is less important if the target is already visible.

Finally, we observed a significant *A*×*S* interaction ($F_{15,330}$=16.21, p<0.001, η^2=0.42), also reported by KA but not by CLB. This interaction is visualized in Figure 3d. With all FoV widths but *S64*, we see consistently increasing movement times as *A* increases. With *S64* we see a sudden increase between *A24* and *A48*. This mirrors the effect of having a target visible at onset: with *S64* targets at *A24* were visible, targets at *A48* were not.

***Model fitting*:** We separately analyze the data from the two techniques, and data from *NoPK* and *PK* selections. We obtain the following R^2 values:

Ray+NoPK:	CLB R^2=.932, RO R^2=.877, Fitts R^2=.345
Ray+PK:	CLB R^2=.950, RO R^2=.847, Fitts R^2=.483
Tap+NoPK:	CLB R^2=.941, RO R^2=.916, Fitts R^2=.204
Tap+PK:	CLB R^2=.944, RO R^2=.898, Fitts R^2=.272

These results confirm KA's findings: when combining data from all six FoV widths, we find better fits with CLB's and RO's models than with the Fitts' law model. The reason is obvious in Figure 4, where we see how the regression lines for the separate *S* values are spread wide apart (the plots for *Ray+NoPK* and *Tap+NoPK* look similar but with larger intercept and slope parameters; the corresponding parameters can be obtained from Table 1).

Accordingly, we follow KA's analysis approach and look at model fits for separate FoV sizes (RO collected data from only one

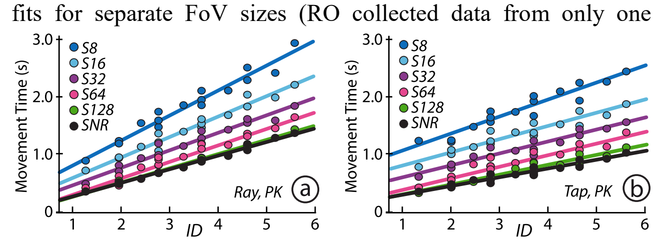

Figure 4. Fitts' regression plots for *Ray* and *Tap*, with *PK*.

peephole size, and CLB do not present separate analyses for their five peephole sizes). These modelling results represent realistic HWD interaction where the FoV width is kept constant during use. However, the two-component models may nonetheless be beneficial to designers, for instance to make an interface consistent across platforms with different FoV widths. There may also be situations where dynamically restricting the FoV is warranted, for example to encourage precise selection in critical instances.

Table 1 lists the regression results for CLB's and RO's models and the Fitts' law models for the different *S* values with the two selection techniques in trials with and without prior knowledge. As KA found in their model comparison, we also see strong fits with all three models in most cases. The R^2 values with the Fitts' model are slightly lower than those with the other models, particularly so in cases with no prior location knowledge. Nevertheless, in cases with many visible targets (*SNR,* or future HWDs with a very wide FoV) the comparably simple Fitts' law model shows to be a sufficiently good predictor of selection performance.

Table 1. Model fitting results for separate FoV widths.

		CLB model				**RO model**				**Fitts' law**		
		R^2	*a*	*b*	*n*	R^2	*a*	*b*	*c*	R^2	*a*	*b*
Ray + NoPK	***S8***	*.905*	.344	.690	.394	*.898*	-.232	.698	.860	*.853*	.872	.498
	S16	*.929*	.514	.509	.294	*.926*	.101	.570	.543	*.903*	.703	.398
	S32	*.983*	.219	.564	.451	*.990*	-.141	.705	.380	*.937*	.409	.362
	S64	*.978*	.107	.614	.568	*.969*	-.248	.882	.278	*.915*	.250	.317
	S128	*.953*	.327	.344	.348	*.957*	-.081	.814	.220	*.948*	.352	.236
	SNR	*.961*	.315	.461	.504	*.960*	-.571	4.25	.206	*.960*	.322	.233
Ray + PK	***S8***	*.978*	.340	.532	.275	*.969*	-.192	.540	.795	*.956*	.624	.429
	S16	*.988*	.293	.461	.321	*.982*	-.065	.514	.472	*.959*	.480	.351
	S32	*.980*	.306	.413	.347	*.984*	-.005	.536	.330	*.958*	.414	.299
	S64	*.979*	.179	.451	.437	*.971*	-.163	.709	.267	*.955*	.260	.284
	S128	*.989*	.240	.280	.128	*.987*	-.200	.792	.238	*.989*	.248	.248
	SNR	*.987*	.261	.237	.000	*.969*	-.655	3.94	.216	*.987*	.261	.237
Tap + NoPK	***S8***	*.866*	.569	.623	.547	*.846*	.195	.628	.546	*.746*	1.23	.383
	S16	*.872*	.745	.463	.422	*.868*	.493	.508	.403	*.824*	.991	.318
	S32	*.966*	.523	.490	.602	*.976*	.293	.579	.241	*.845*	.744	.256
	S64	*.983*	.209	.669	.737	*.975*	-.024	.847	.184	*.817*	.411	.249
	S128	*.943*	.329	.587	.762	*.946*	.076	.878	.137	*.834*	.423	.182
	SNR	*.952*	.463	1.04	.879	*.963*	-.023	3.25	.112	*.916*	.489	.141
Tap + PK	***S8***	*.948*	.732	.390	.342	*.913*	.392	.395	.512	*.910*	.991	.296
	S16	*.949*	.648	.325	.413	*.944*	.431	.357	.287	*.900*	.817	.225
	S32	*.972*	.493	.338	.484	*.973*	.292	.418	.213	*.916*	.615	.208
	S64	*.977*	.254	.517	.717	*.973*	.057	.665	.154	*.831*	.405	.201
	S128	*.985*	.269	.499	.725	*.986*	.020	.787	.135	*.898*	.345	.172
	SNR	*.990*	.356	1.36	.906	*.988*	-.139	3.58	.114	*.938*	.387	.149

4. CONCLUSION

We conducted a user study to test the prediction accuracy of two-component pointing models with a "peephole" controlled by head motion. Our study compared the performance of direct pointing and raycasting under these conditions.

Our results confirm that two-component (peephole pointing) models apply to target selection when the view is controlled by head motion, as with HWDs. We confirm previous findings from other contexts, that Fitts'law is a good predictor when targets are initially visible, and when *S* (display size, or in our case FoV width) is a known constant. Our comparison shows that direct pointing can be faster than raycasting with a sufficiently wide FoV. Raycasting is slow with small targets [16], however, we find the advantage of direct pointing is negated with reduced FoV.

Unlike prior research on two-component models, our user study focused on decoupled techniques, where head motion and target selection are controlled independently. In future work, we would like to further explore the degree of coupling between the user's hand and head motion during selections with these and other techniques, and to determine the role that FoV plays in coupling strength. We would also like to confirm our findings on various HWD hardware platforms with 2D and 3D Fitts' pointing tasks. In further work, we would also like to explore design opportunities and benefits of HWD interfaces with dynamically changing FoV.

REFERENCES

[1] Ball, R. and North, C. 2008. The effects of peripheral vision and physical navigation on large scale visualization. *Proc. GI 2008*, 9–16.

[2] Cao, X., Li, J.J. and Balakrishnan, R. 2008. Peephole pointing: modeling acquisition of dynamically revealed targets. *Proc CHI '08*, 1699–1708.

[3] Ens, B.M., Finnegan, R. and Irani, P.P. 2014. The Personal Cockpit: A spatial interface for effective task switching on head-worn displays. *Proc. CHI '14*, 3171–3180.

[4] Fitts, P.M. 1954. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*. 47, 6 (Jun. 1954), 381–391.

[5] Grossman, T. and Balakrishnan, R. 2004. Pointing at trivariate targets in 3D environments. *Proc. CHI '04*, 447–454.

[6] Jagacinski, R.J. and Monk, D.L. 1985. Fitts' Law in two dimensions with hand and head movements. *Journal of Motor Behavior*. 17, 1 (Mar. 1985), 77–95.

[7] Kaufmann, B. and Ahlström, D. 2012. Revisiting peephole pointing: a study of target acquisition with a handheld projector. *Proc. MobileHCI '12*, 211–220.

[8] Kishishita, N., Kiyokawa, K., Orlosky, J., Mashita, T., Takemura, H. and Kruijff, E. 2014. Analysing the effects of a wide field of view augmented reality display on search performance in divided attention tasks. *Proc. ISMAR '14*, 177–186.

[9] Kopper, R. A human motor beh, Bowman, D.A., Silva, M.G. and McMahan, R.P. 2010. avior model for distal pointing tasks. *IJHCI*. 68, 10 (Oct. 2010), 603–615.

[10] MacKenzie, C.L., Marteniuk, R.G., Dugas, C., Liske, D. and Eickmeier, B. 1987. Three-dimensional movement trajectories in Fitts' task: Implications for control. *The Quarterly Journal of Experimental Psychology Section A*. 39, 4 (Nov. 1987), 629–647.

[11] Patterson, R., Winterbottom, M.D. and Pierce, B.J. 2006. Perceptual issues in the use of head-mounted visual displays. *Human Factors*. 48, 3, 555–573.

[12] Poupyrev, I., Weghorst, S., Billinghurst, M. and Ichikawa, T. 1997. A framework and testbed for studying manipulation techniques for immersive VR. *Proc. UIST '97*, 21–28.

[13] Rohs, M. and Oulasvirta, A. 2008. Target acquisition with camera phones when used as magic lenses. *Proc. CHI '08*, 1409–1418.

[14] Stuerzlinger, W. and Teather, R.J. 2014. Considerations for targets in 3D pointing experiments. *Proc. HCI Korea* '14, 162–168.

[15] Teather, R.J. and Stuerzlinger, W. 2011. Pointing at 3D targets in a stereo head-tracked virtual environment. *Proc. 3DUI '11*, 87–94.

[16] Vogel, D. and Balakrishnan, R. 2005. Distant freehand pointing and clicking on very large, high resolution displays. *Proc. UIST '05*, 33–42.

[17] Zeng, X., Hedge, A. and Guimbretiere, F. 2012. Fitts' law in 3D space with coordinated hand movements. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. 56, 1 (Sep. 2012), 990–994.