2015

Visual Search and Target Selection Using a Bounded Optimal Model of State Estimation & Control

Brandon S. Perelman

Christopher W. Myers

VISUAL SEARCH AND TARGET SELECTION USING A BOUNDED OPTIMAL MODEL OF STATE ESTIMATION & CONTROL

Brandon S. Perelman
Michigan Technological University
Houghton, MI
Christopher W. Myers
Air Force Research Laboratory
Wright-Patterson Air Force Base, OH

Visual attention and motor control are tightly coupled in domains requiring a human operator to interact with a visual interface. Here, we integrate a boundedly optimal visual attention model with two separate motor control models and compare the predictions made by these models against perceptual and motor data collected from human subjects engaged in a parafoveal detection task. The results indicate that humans use an optimal motor control policy limited by precision constraints – humans executed ballistic movements using near-optimal velocity (i.e., bang-bang control), but imprecision in those movements often caused participants to overshoot their targets, necessitating corrective action. Motor movements did not reflect response hedging, but rather a perceptual-motor policy permitting ballistic movements to a target only after localization confidence exceeded a threshold. We conclude that a boundedly-optimal perceptual-motor model can predict aspects of human performance visual search tasks requiring motor response.

Introduction

Visual search is conducted in nearly everything we do, from the mundane to dangerous operational domains. For example, while shopping online we search crowded visual displays to find the correct item. Similarly, sensor and radar operators must discriminate targets from noise and foils. These tasks require that an agent identify the target through visual search, and then select that target. In this study, we integrated a boundedly-optimal state estimation model of visual search with models of motor control, and tested them against human performance in a parafoveal detection task (PDT). The long-term goal of this line of research is a model capable of generating performance ceiling predictions and automatic interface evaluations. In the following sections we introduce visual search and oculomotor control, followed by manual motor control.

Visual Search and Eye Movements

Efforts to model visual search fall largely into two categories (see Kowler, 2011 for a comprehensive review). *Map-based* approaches (e.g., Itti, 2006; Itti & Koch, 2000; Pomplun, 2003; Wolfe, 2007) use bottom-up processing to subdivide raw images into saliency or activation maps. These models predict that the agent will produce a saccade to the most salient or active areas of a map derived from bottom-up processes over feature and spatial information (e.g., color, rotation, distance, etc.). Conversely, *visibility models* (Geisler, 2011; Myers, Gray, & Sims, 2011; Myers, Lewis, & Howes, 2013; Najemnik & Geisler, 2008; Baron & Kleinman, 1969) are top-down models of visual attention, predicting that eye movements are made in the service of maximizing information gain. Such models, often labeled *ideal observers*, have demonstrated much success in accounting for human saccades during search. In the present study, we use a visibility model to derive predictions in the PDT. Our visibility model optimally estimates the state of a presented stimulus given known bounds of the human visual system. We refer to this model as boundedly-optimal, and combine boundedly-optimal state estimation with near optimal oculomotor control (i.e., the near-optimal saccadic selectivity, given a boundedly-optimal estimation of the state of the display).

Manual Motor Control

Evidence in the motor literature suggests that motor control reflects a dynamic decision-making process (Freeman, Dale, & Farmer, 2011; Wolpert & Landy, 2012) whereby participants' motor trajectories reflect cognitive phenomena, such as confidence. These effects have been investigated in the Iowa gambling task (Koop & Johnson, 2011), memory tasks (Papesh & Goldinger, 2012), and item selection from an interface (Bailly, Oulasvirta, Brumby, & Howes, 2014). In ambiguous situations where the operator must distinguish among multiple potential targets, high probability locations often attract the operator's cursor even when they are not ultimately selected (Farmer, Cargill,

& Spivey, 2007). Therefore, motor control does not involve merely converting visual information to motor coordinates, as saccades occurring during a motor movement can cause immediate changes in destination and trajectory (Thompson, Byrne, & Henriques, 2014). It is important to distinguish among input devices, as motor control trajectories derived using a mouse (e.g., Bailly et al., 2014; Koop & Johnson, 2011; Papesh & Goldinger, 2012) differ from those generated using touchscreen inputs (e.g., Parhi, Karlson, & Bederson, 2006).

To model participants' velocity profiles, we used two theories of motor control (Kelso, 1982) – an open-loop theory (*bang-bang* control, which produces the optimal acceleration and deceleration between starting and target cursor positions assuming equal rates of acceleration and deceleration), and a closed loop theory (velocity proportional to distance, *vProp*, which begins at maximum acceleration and reduces its speed proportionally with feedback based on the decreasing distance to the target). In the integrated visuomotor model, we represent the decision making process' influence on motor trajectories using closed-loop control that permits the model to dynamically update the cursor destination to the current highest probability target location. Finally, to model the effect of motor control parameters on response times, we implemented a crude version of Fitts' law (Fitts, 1954) whereby the model's ballistic cursor movements were perturbed by noise that scaled with that movement's distance.

Experiment

All participants completed a PDT that required target detection and localization within a pair of items on opposing sides of initial fixation crosshairs (see Figure 1). Participants were instructed to fixate on crosshairs located at the center of the screen (see Stage 1, Figure 1), and on fixation, to click the mouse to initiate stimulus onset. A stimulus consisted of red and green X and O characters in four locations on opposite sides of the crosshairs, but on the same plane (Stage 2, Figure 1). On stimulus onset, the mouse was positioned 334 pixels below the crosshairs, and participants were instructed to respond by clicking on the perceived target location. If the target was not detected, participants clicked on the fixation crosshair (target-absent; Stage 3, Figure 1).

The PDT was programmed using the Psychology Experiment Building Language (Mueller, 2014). The stimulus was presented on a screen (1024 x 768 pixels resolution) viewed from a distance of 20 inches Pairs were separated by 1° of visual angle. Pairs varied in eccentricity from the crosshairs by 8°, 12°, 16°, and 20°. Pairs on each side of the crosshairs were always separated by the same eccentricity. .

Of the 189 ecologically possible displays, 32 were selected that maximized prediction differences between map models and those from our boundedly-optimal state estimation model. Half of these displays contained targets and half did not. Participants completed 640 trials, broken into five blocks of 128 trials. Each block was a randomized set of all 32 displays at each of the four eccentricities.

Modeling

The model was previously fit to a version of the PDT that required no mouse responses, only target-present or target-absent response through keypresses. Best parameter fits were determined by investigating a space of 3,200,000 parameter combinations using www.mindmodeling.org (Harris, Gluck, Mielke, & Moore, 2009). Results indicated best-fitting parameters: saccade threshold = 0.21, response threshold = 0.84, spatial noise = 10, and feature noise = 6. For a detailed description of the boundedly-optimal state estimation model, see Myers et al. (2013). These same parameter values were then used in modeling a version of the PDT requiring point-and-click responses.

To issue point-and-click responses, the model executed motor movements from the starting mouse location to its current target (i.e., highest probability location once the motor threshold is reached. The cursor's destination was perturbed as above, then updated at the model's sampling rate of 25 ms, while unperturbed (direct) trajectories represent the optimal motor response against which to compare human data. Motor control parameters included maximum acceleration and deceleration (2 pixels per sample; though the proper setting of this parameter could be determined empirically using Fitts' Law) and a motor movement initiation threshold (probability of target present or absent >= 0.51). This threshold parameter permitted the model to initiate "early" motor movements, guiding the cursor during stimulus presentation before the model had committed to a particular target location. The cursor update loop ends when the cursor is within the target location's clickable field, at which point the response time is appended with a manual response time, intended to simulate a mouse click, drawn from a gamma distribution (shape = 11.11, scale = 9), which produced a mean manual response time of roughly 100 ms (M = 99.93, SD = 29.77).

Hypotheses and Model Predictions

Using the aforementioned parameters, we ran the model on 25 trials for each of the 128 display combinations, for a total of 3,200 runs, to produce a dataset against which to compare the human subjects' performance. Model velocity profiles using both the bang-bang and vProp algorithms provide a baseline against which to compare participants' performance. Straight paths between the cursor starting and target locations provide optimal motor trajectories against which to compare the human data. Finally, a crude implementation of Fitts' law (see above) permitted the model to generate motor response times for comparison. Because the visual attention model applies feature and spatial noise that increases with distance from the point of fixation, it predicts an effect of eccentricity on all aspects of task performance. Therefore, we expect that,

1. Motor Velocity: We expect that humans will exhibit one of the experimental (bang-bang or vProp) motor velocity profiles given there is evidence for each in the literature, However, because of the static nature of the task and only a required straight movement to reach a location, then bang-bang is a better candidate.
2. Motor Trajectories: Dynamic decision making theory suggests that error in participants' motor response trajectories should increase with increasing difficulty, therefore for humans we expect increased response times and increased motor trajectory error (greater divergence from the optimal trajectory), measured by pathmapping) with increasing eccentricity. Furthermore, we expect less motor control error on trials where participants respond correctly.
3. Response Initiation: The model builds evidence toward a decision more slowly as task difficulty increases, therefore the proportion of trials on which the agent initiates an *early motor movement* > 25 pixels from cursor starting location *during* the 500 ms stimulus presentation window should similarly decrease.
4. Response Accuracy: The model predicts that target identification (distinguishing target present from target absent displays) and localization (determining the specific location of the target) performance should decrease with increasing eccentricity

Data Analysis

Motor data were subdivided into two components – trajectory and velocity. To analyze the velocity profiles, we split human and model trajectories into two halves, then divided the average velocity in the second half of the trajectory by the average velocity in the first, creating a *split-half velocity ratio*. We expected that participants would exhibit one of three velocity profiles, each distinguishable by the velocity ratio. Because the bang-bang algorithm uses maximum acceleration and deceleration before and after a halfway point, its velocity profile predicts a split-half velocity ratio around one (i.e., the average velocity in both trajectory halves are equal). Conversely, vProp predicts a split-half velocity ratio of either greater or less than one, depending upon whether the cursor accelerates or decelerates as it approaches its destination.

To analyze error in the human data, we used the 'pathmapping' package (Mueller & Perelman, 2013) built for the R statistical computing language. This package creates a polygon from two arbitrary paths (i.e., the empirical trajectory, and the optimal trajectory, a straight line from starting position to that target), the area of which is the error in pixels, and holds an advantage over traditional measures of motor error (Koop & Johnson, 2011).

Results

Hypothesis 1: Velocity Profiles and Optimal Control

Across all eccentricities, and for all target locations, the model produced mean split-half velocity ratios of 1.03 ($SD = 0.19$) using the bang-bang algorithm, and 6.68 ($SD = 1.82$) with vProp. Participants' trajectories consisted of an initial ballistic trajectory toward the target location, which often carried the cursor past the target location, followed by a corrective trajectory, which brought the cursor back to the target. The ballistic trajectory was operationalized as all sampled trajectory points to the cursor's farthest distance from the target location, with corrective trajectory accounting for remaining points. Split-half velocity ratios indicated that the ballistic trajectory was very similar to bang-bang style movement ($M = 0.99$, $SD = 0.08$), whereas the corrective trajectory functioned similarly to a vProp control method ($M = 0.85$, $SD = 0.35$), owing largely to the requirement for the participant to change direction of travel nearly 180 degrees back to the target location, and the relatively short travel time. Participants tended to overshoot more distant targets by a margin that increased with eccentricity, $F(3, 3928) =$

23.70, $p < .001$ (see Table 1), an effect which holds implications for modeling motor performance in higher fidelity. These results indicate that humans exhibit optimal bang-bang control in this task.

Hypothesis 2: Motor Trajectories and Optimal Control

Human motor control error (i.e., divergence from optimal measured via pathmapping) increased with eccentricity, $F(3, 5115) = 74.94$, $p < .001$. Trajectories produced during correct responses were roughly twice as close to optimal as incorrect response trajectories, $t(560.46) = 9.20$, $p < .001$. One potential criticism of this approach is that longer trajectories leave more room for potentially producing error, due to the cursor travel distance. To address this problem, we scaled the error values at each eccentricity by the cursor's distance from the target location. Using these scaled error terms, effects of eccentricity, $F(3, 5115) = 16.7$, $p < .001$, and accuracy, $t(562.49) = 8.72$, $p < .001$, persisted even when controlling for cursor distance of travel.

Hypothesis 3: Motor Movements and Dynamic Decision Making

Humans and the model produced the expected effect of eccentricity on response time during target trials, however only humans exhibited this effect when the target was not present (see Figure 3). Humans were faster by 185 and 198 ms on the target-absent and target-present trials, respectively. The model, given a motor threshold of 0.51, and participants produced fewer early motor movements with increasing eccentricity (see Table 2).

Participants exhibited mean motor velocities that varied with signal detection and correctness. Hits ($M = 9.19$, $SD = 3.02$) and correct rejections ($M = 8.85$, $SD = 3.37$) produced faster mean motor velocities than misses ($M = 8.09$, $SD = 3.37$) or false alarms ($M = 7.69$, $SD = 4.42$). A 2 (Correct vs. Incorrect) x 2 (Trajectory: Ballistic vs. Corrective) factorial ANOVA revealed a significant interaction effect, $F(1, 9037) = 85.98$, $p < .001$, whereby correct answers produced faster ballistic trajectories ($M = 12.68$, $SD = 6.93$) than incorrect answers ($M = 11.08$, $SD = 7.69$), but slower corrective trajectories ($M = 1.55$, $SD = 2.54$) than incorrect answers ($M = 3.47$, $SD = 4.56$). These results indicate that confident response selection produces more precise and expedient motor movements.

Hypothesis 4: Target Identification and Localization

Both the model, $F(3, 3196) = 5.56$, $p < .001$, and participants, $F(3, 5116) = 38.81$, $p < .001$, exhibited target identification performance that degraded with increasing eccentricity. In addition, localization degraded with increasing eccentricity in participants, $F(3, 2556) = 27.57$, $p < .001$, and the model, $F(3, 3196) = 10.36$, $p < .001$ (see Figure 2). To further evaluate model and participant accuracy in target identification, we applied signal detection theory (see Table 2). Humans' target discriminability (D') degraded with increasing eccentricity. The model demonstrated a similar trend, with the exception that the model's D' at 12 degrees was higher than at 8 degrees of eccentricity. Given the aforementioned effect of eccentricity on response accuracy, this difference lies in the false alarm rate and may reflect a tradeoff whereby the model adopts a more conservative strategy than humans.

Conclusions and Future Directions

The human subjects and modeling results, taken together, indicate that a bounded optimal state estimation model of visual attention, coupled with bang-bang motor control, produces similar effects to those seen in humans in the PDT. Specifically, the model predicts decreasing response confidence, and performance, as measured using identification and localization accuracy and response time, with increasing eccentricity from the point of fixation.

One proximal goal of future research is to further validate the motor control system. Analyzing model overshoots and trajectory divergence would permit direct comparison with the human data. Furthermore, it is not currently clear as to whether the differences in RT between humans and the model were due to differences in motor speed or accuracy, and trial-by-trial analysis should elucidate this in future research.

In service of the project goal of predicting human performance, a more accurate implementation of Fitts' Law would provide a more realistic account of the motor data, and should also impact response time distributions. Finally, testing the model against human data in a more complicated task, such as a computer interface, would speak to external validity of this approach for predicting performance in naturalistic tasks.
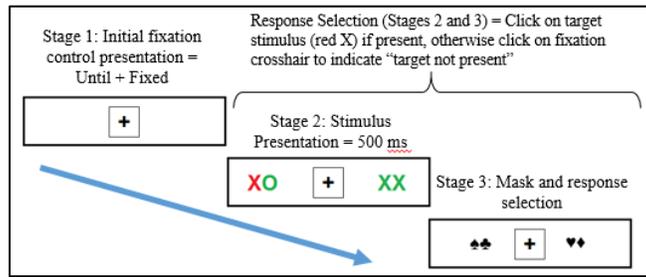
Tables and Figures

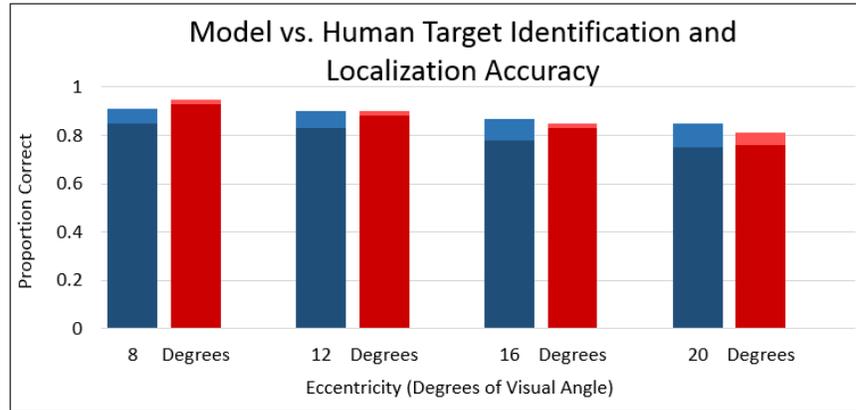*Figure 1.* Parafoveal detection task time course and instructions.

Figure 2. Model and human performance in identifying and localizing targets across all experimental eccentricities in target present trials. Blue and red bars correspond to model and human performance, respectively. Bars indicate the proportion of trials in which the agent successfully identified the target, while darkened portions indicate the proportion of those trials in which they also successfully localized the target.
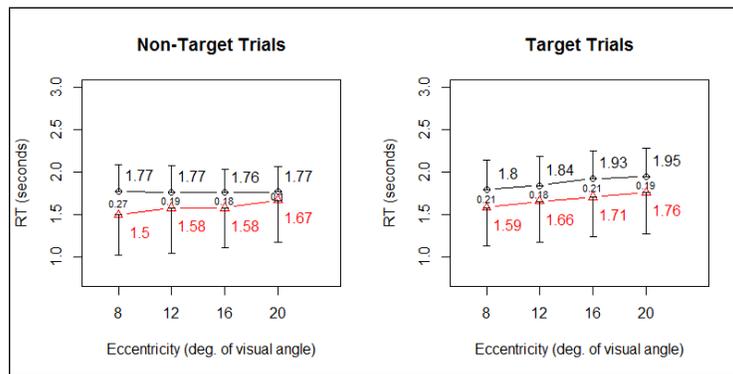

Figure 3. Model (black circle) and human (red triangle) response times in target and non-target trials by inner eccentricity. Difference scores are shown between the two data series. Errors bars reflect standard deviation.

Table 1.
Corrective trajectory split-half velocity ratios and target overshoot by eccentricity.

Eccentricity	Split-Half Velocity Ratio	Target Overshoot (Pixels)
8 Degrees	0.83	39
12 Degrees	0.85	52
16 Degrees	0.86	61
20 Degrees	0.87	73

Table 2.
Proportion of trials where the agent initiated an early motor movement (left), and D' calculations (right) at each experimental eccentricity.

Eccentricity	Early Motor Movements		D' Calculations	
	Humans	Model	Humans	Model
8 Degrees	0.17	0.19	3.28	2.40
12 Degrees	0.11	0.15	2.95	2.49
16 Degrees	0.08	0.12	2.62	1.92
20 Degrees	0.08	0.12	1.94	1.61

References

Bailly, G., Oulasvirta, A., Brumby, D. P., & Howes, A. (2014). Model of visual search and selection time in linear menus. In *Proc. of the ACM CHI Conference on Human Factors in Computing Systems,* Toronto, Canada.

Baron S. & Kleinman, D. L. (1969). The human as an optimal controller and information processor. *IEEE Transactions on Man-Machine Systems, 10,* 10-17.

Farmer, T. A., Cargill, S. A., & Spivey, M. J. (2007). Gradiency and visual context in syntactic garden-paths. *J. Mem. Lang., 57,* 570-595.

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

Freeman, J. B., Dale, R., & Farmer, T. A. (2011). Hand in motion reveals mind in motion. *Frontiers in Psychology, 2,* 1-6.

Geisler, W. S. (2011). Contributions of ideal observer theory to vision research. *Vision Research, 51,* 771-781.

Harris, J., Gluck, K., Mielke, T., & Moore, L. R. (2009). Mindmodeling@Home...and anywhere else you have idle processors. In *Proceedings of the Ninth International Conference on Cognitive Modeling*, Manchester, UK.

Itti, L. (2006). Quantitative modelling of perceptual salience at human eye position. *Visual Cognition, 14,* 959-984.

Itti, L. & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40,* 1489-1506.

Kelso, J. S. (Ed.). (2014). *Human motor behavior: An introduction*. Psychology Press.

Koop, G. J. & Johnson, J. G. (2011). Response dynamics: A new window on the decision process. *Judgment and Decision Making, 6,* 750-758.

Kowler, E. (2011). Eye movements: The past 25 years. *Vision Research, 51,* 1457-1483.

Mueller, S. T. (2014). PEBL: The Psychology experiment building language (Version 0.14) [Computer experiment programming language]. Retrieved June 2014 from http://pebl.sourceforge.net.

Mueller, S. T. & Perelman, B. S. (2013). Pathmapping: Software for Determining the Divergence and Mapping Between Two Paths. R package version 1.0.

Myers, C. W., Gray, W. D., & Sims, C. R. (2011). The insistence of vision: Why do people look at a salient stimulus when it signals target absence? *Visual Cognition, 19,* 1122-1157.

Myers, C. W., Lewis, R. L., & Howes, A. (2013). Bounded optimal state estimation and control in visual search: Explaining distractor ratio effects. In *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*, Berlin, Germany.

Najemnik, J. & Geisler, W. S. (2008). Eye movement statistics in humans are consistent with an optimal search strategy. *Journal of Vision, 8,* 4.1-14.

Papesh, M. H. & Goldinger, S. D. (2012). Memory in motion: Movement dynamics reveal memory strength. *Psychon. Bull. Rev., 19,* 906-913.

Parhi, P., Karlson, A. K., & Bederson, B. B. (2006). Target size study for one-handed thumb use on small touchscreen devices. In *Proceedings of the 8th conference on Human-computer interaction with mobile devices and services*, Espoo, Finland.

Pomplun, M. (2003). Area activation: A computational model of saccadic selectivity in visual search. *Cognitive Science, 27,* 299-312.

Thompson, A. A., Byrne, P. A., & Henriques, D. Y. P. (2014). Visual targets aren't irreversibly converted to motor coordinates: Eye-centered updating of visuospatial memory in online reach control. *PLoS One, 9.*

Wolfe, J. M. (2007). Guide Search 4.0: Current progress with a model of visual search. In W. D. Gray (Ed.), *Integrated Models of Cognitive Systems* (pp. 99-119). New York: Oxford University Press.

Wolpert, D. M. & Landy, M. S. (2012). Motor control is decision-making. *Current Opinions in Neurobiology, 22,* 996-1003.