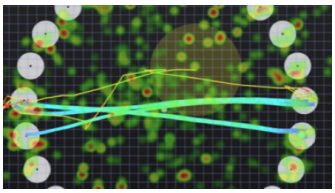
Watching You Watching You: Visualization of Visual-Motor Behavior

Leslie M. Blaha, Joseph W. Houpt, Mary E. Frame, and Jacob A. Kern



Abstract—Understanding the relationship between where people look and where people reach has implications for the designs of interactive visualizations, particularly for applications involving touch screens. We present a new Visual-Motor Analytics Dashboard for the joint study of eye movement and hand/finger movement data. Our Dashboard combines real-time playback of gaze and finger-dragging behavior together with statistical models for the dynamics profiles. To aid in visualization and inference with these data, we introduce Gaussian process models in order to capture the similarities and differences between eye and finger movements, while providing a statistical model of the observed data. Smooth estimates of the dynamics are included in the Dashboard to enable visual-analytic exploration of visual-motor behaviors under differing task difficulty conditions.

Index Terms—Eye tracking, Fitts' Law, Gaussian process models, Information processing efficiency, Visual-motor processing.

1 Introduction

Development of effective interactive visualizations for touch screens requires an understanding of the ways the visual features and interaction difficulty influence both eye movements and finger movements. Evaluating the efficacy of design choices for interactive visualization requires an understanding of how those choices influence the cognitive processes supporting the visual analytics process and how those processes manifest in measurable behaviors [8]. Toward this end, the present work developed a Visual-Motor Visualization Dashboard in which we can jointly examine the timing, trajectories, and dynamics of hand/finger (motor) and eye movements. The purpose of the Dashboard is to facilitate exploration of two key questions about eye-hand interactions: (1) When executing a touch screen movement, do you look at your finger or do you look where you want your finger to go? and (2) Under what task constraints do finger and eye movements look similar or deviate?

Our approach is based on visualizing raw movement data and traditional summary statistics together with Gaussian process (GP) models for estimates of position, velocity, and acceleration values and uncertainty. Previous work on visualizing eye-hand coordination has provided some means of plotting the x-y position over time of eye fixations and mouse locations on a screen using space-time cubes [2, 3, 7] or utilizing traditional fixation heat maps. Both of those approaches have emphasized position and fixation duration. We developed the Dashboard to capture those previously explored properties of movement trajectories and augment them with visualizations of velocity and acceleration estimates in a single visualization platform. In addition to estimating the position, velocity and acceleration for visualization purposes, our approach yields estimates of the uncertainty in each of those those measures. Our approach models both eye and finger movements in a common statistical framework that allows for direct comparison of their behavioral profiles in order to address our key question about the relationship between each of those trajectories.

- *Leslie Blaha is with Air Force Research Laboratory. E-mail: leslie.blaha@us.af.mil.*
- *Joseph Houpt is with Wright State University. E-mail: joseph.houpt@wright.edu.*
- *Mary Frame is with Miami University. E-mail: frameme@miamioh.edu.*
- *Jacob Kern is with Wright State University. E-mail: jacob.kern@outlook.com*

2 Sample Visual-Motor Interaction Data

We used a basic target dragging task representative of all touch screen interactions that involve dragging icons between positions as a test data set to explore the GP models and the Dashboard. In this data set, target size and distance were manipulated in order to vary the visual properties and physical constraints on the motor demands. The task was chosen so that it should not require any complex visual search or decision-making, instead emphasizing the eye-finger coordination dynamics.

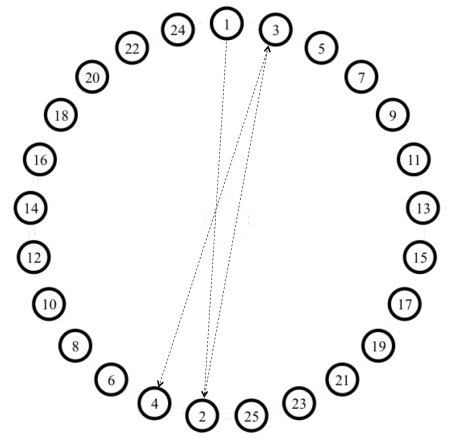
2.1 Equipment

The dragging task was performed on an 82-inch, wall-mounted, rear-projection Perceptive Pixel-brand touchscreen (native resolution 4096×2400 horizontal by vertical pixels, 60 Hz refresh rate). The Perceptive Pixel recorded touch inputs with Frustrated Total Internal Reflection [5]. Screen touch calibration was performed with the native software calibration routine. The task was programmed in Python with Kivy library version 1.8.0 and run with the Anaconda Scientific Python Distribution 1.9.1.

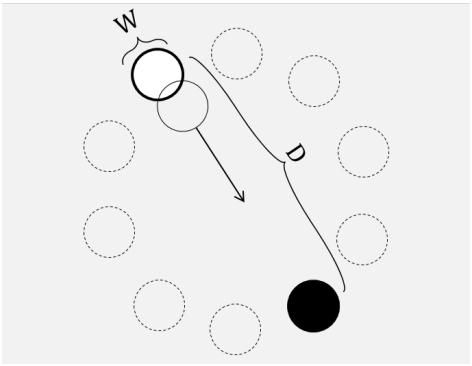
Eye tracking data were collected with the Tobii Glasses 2 Version 1.0.2. This system records a 160 degree horizontal and 70 degree vertical field of view at 50 Hz following a one point calibration procedure. A wide angle camera (1920 x 1080 pixels, 25 Hz) mounted between the lenses, centered on the observer's nose bridge, recorded the user's scene view. The glasses record the eye gaze fixation point using 4 IR cameras, two left and two right, mounted in the lens frame below the eyes. Gaze point data was mapped onto a two dimensional image of the touch screen using Tobii Glasses Analysis Software Version 1.0.

2.2 Visual-Motor Tasks

Eye tracking and finger movement performance were recorded during a predictable target dragging task, designed to be similar to the standardized multidirectional tapping task suggested for pointing device evaluation [11]. In a block of trials, users had to drag a finger along the touch screen between circular start and end targets. The configuration of all 25 targets for a block is shown in Figure 1a; the same configuration and target order was used on each block. Blocks differed in the width of the targets (possible values 64, 128, 192, 256 pixels) and the distance between the targets (or the diameter of the circular configuration; possible values 128, 512, 1024, 2048 pixels; see



(a)



(b)

Fig. 1: (a) The configuration of all targets for the dragging tasks. Each target is numbered according to the order in which it appears in the sequence, starting with 1 at the top. Arrows indicate the progression of the first three trials. (b) Illustration of the width (W) and distance (D) between targets on a single trial. The white circle is the start target and the black circle is the end target. Dragging movement always moves from white to black target, as indicated by the direction of the arrow.

Figure 1b). Target positions were predictable but some target distances were larger than the user's field of view and required looking across the screen to locate the target. Users were instructed to always perform the task accurately but dragging paths were not constrained.

According to Fitts' Law, the width and distance parameters define an *index of difficulty* for each block; computed as $ID = \log_2\left(\frac{D}{W}+1\right)$ [4]. Smaller targets that are further apart are more difficult to drag accurately than larger, closer together targets. Data from this task can be used to examine if finger and eye movement dynamics are influenced by the index of difficulty for a block, or by width or distance alone.

3 Visual-Motor Analytics Dashboard

Figure 2 shows a screenshot of the Visual-Motor Analytics Dashboard. It contains a series of gauges displaying different aspects of the time series data from task performance together with model estimates of performance dynamics and summary analytics resulting from two additional cognitively relevant statistical models. Combined models and frame-by-frame playback enable in-depth review of a user's overt behaviors during task performance.

The Dashboard visualizes and replays data stored in CSV files, which may be raw recordings or pre-processed outside the Dashboard. The Dashboard interface is a web-browser application written in Node.js using common web languages, such as HTML and CSS, and JavaScript libraries (e.g., D3 [1]) to capture different data types within customizable dynamic gauges. We incorporated gauges for replaying motor trajectories, eye fixations, and saccadic eye movements, together with model-based analyses explained in subsequent sections. We also incorporate the eye tracker's scene camera video recording in the upper right. The modular design of the gauges means they are easily exchanged for other models or desired analyses. All time-series data are synced according to the timeline of the experiment itself. Playback controls and a block-by-block timeline are presented at the top for analyst-driven replay. The timeline includes both the blocks of the tasks (blue colors) as well as the between-block rest periods (orange). During the between-block rests, no touch data was recorded but the eye tracker continued recording. The variations in the blue color are according to the index of difficulty for the given task, with the easier blocks being lighter and the more difficult blocks being darker.

The center Target Plot gauge gives an animation of the task overlaid with both finger and eye movement data. The targets for the task are shown as the circular configuration of light gray circles. These vary between the blocks scaled to the actual targets in the task, appearing similar to the user's view in the upper right scene video. Finger dragging movements are shown as a blue trace line; the green-blue circle at the head of the line shows the current finger position. The color of the line is graded according to velocity, with brighter greens corresponding to faster speeds, and darker blues corresponding to slower speeds. Similarly, saccadic eye movements are shown in the red-yellow trace lines. The foveal area is given by the yellow circle, centered at the head of the eye trace line. Again, the line color is graded according to velocity, with darker red being slower speeds and brighter orange-yellow being faster speeds. The Target Plot contains an optional eye fixation heat map overlay, which can be toggled on/off. The fixation heat map is colored according to the length of fixation in any position on a green-to-red scale, with green being short fixations and red being longer. The heat map is reset for each block of trials.

To the right and left of the Target Plot gauge, we include phase-space gauges to capture the full position, velocity, and acceleration dynamics in path-to-target coordinates based on GP models of the finger and eye movement data. The Target-Referenced Finger (Eye) Dynamics gauges illustrate the profiles for a single trial at a time. The end position of the trial is given by the dot at the tail of the curves, with the trace line fading in brightness from the end to the beginning of the trail (beginning being the least bright). We will return to the patterns of these plots later after describing the GP models.

Below the Target Plot gauge are three gauges illustrating the raw dynamics and position data. In each of these gauges, the vertical gray line indicates "now" in the time series, and the data slide from right to left over time during playback. The horizontal gray line is the zero or threshold value for the data presented. The left most gauge is the Eye-Finger Delta gauge, which illustrates the difference between the position of the finger and the eye in touch screen coordinates. The lines are disjoint, because the finger is not always in contact with the screen (lifted in between trials). The line only has a given position while the finger is actively touching the screen. The horizontal threshold is set to 64 pixels, and lines dropping below threshold occur when the eye and finger are within 64 pixels of each other. These data often show a "v" shape, where the finger and eye move apart, then closer, and then apart during a single target-dragging movement.

The Target Delta for Completed Trials (center gauge) shows two lines tracking the distance between each of the finger (blue) and eye fixation (orange) and the current target (horizontal threshold). Over time, these data generally show a larger delta at the beginning of the trial, and delta decreases as the eye and finger move toward the target location. What is highlighted well in this data is the pattern of the eye moving toward the target before the finger and that the slope of the eye movement is often steeper (faster saccadic movement) than the finger. The eye data is continuous, with the sharp rising slopes occurring when the target position changes in the tasks. The finger data continue to be disjoint, reflecting the finger lifted from the screen between movements.

Third, the right tile shows the Finger Velocity gauge, giving the raw velocity of the finger in both the X (blue) and Y (orange) dimensions.

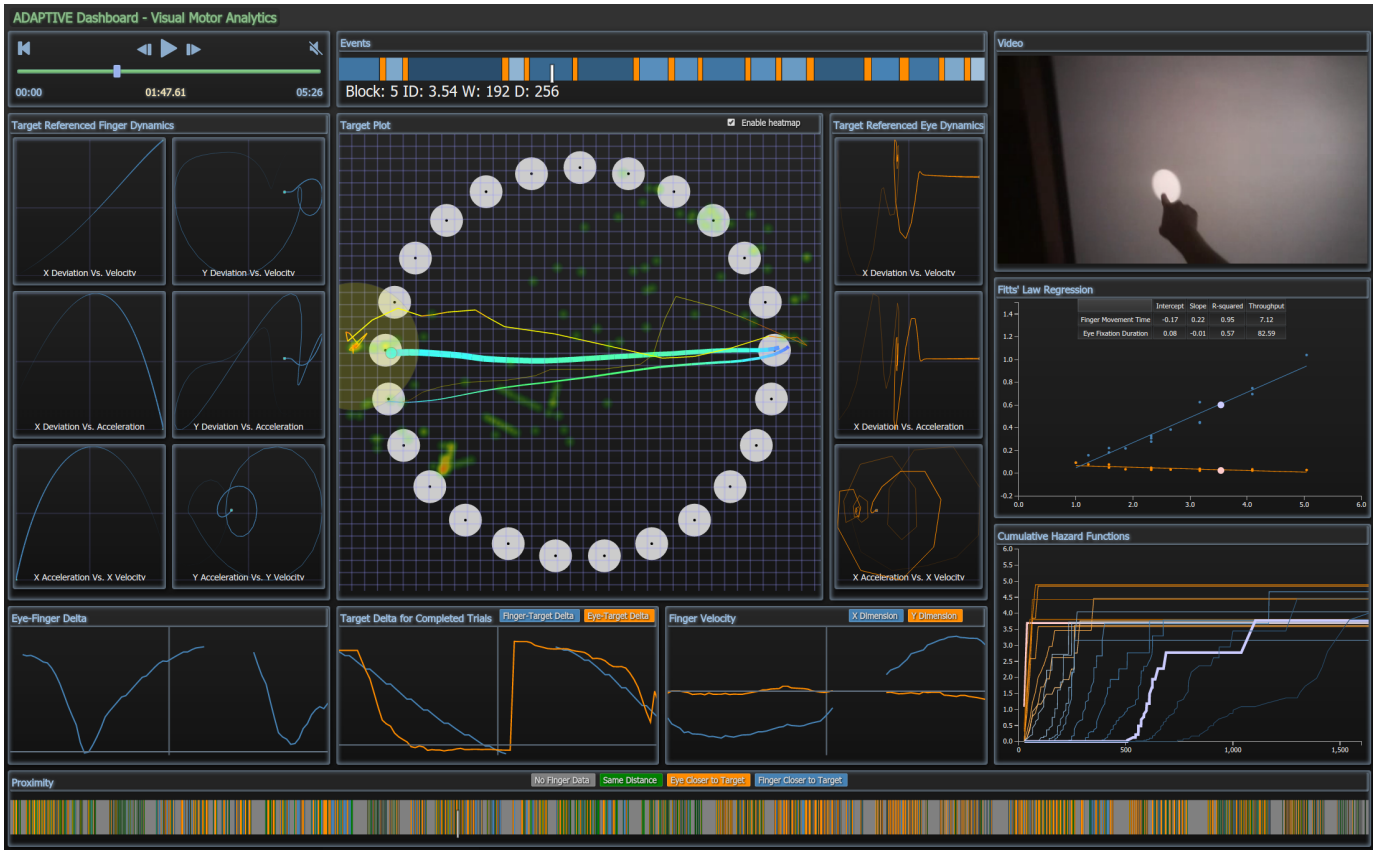


Fig. 2: Visual-Motor Dashboard for combined eye and finger movement data. The central gauge illustrates the target configuration for a given block overlaid with finger paths (blue lines), eye scanpaths (yellow-red lines), and a fixation heat map. The remaining plots capture various aspects of movement dynamics which are summarized in the text.

The horizontal line here is zero velocity. The center of the touchscreen is position zero in both X (right-left) and Y (up-down) dimensions; positive velocities are dragging movements up and/or to the right, and negative curves reflect movements down and/or to the left. Thus, variations in these lines reflect the direction of the movements and some natural biomechanics that constrain arm/wrist/finger movements. The total deviation from zero reflects total speed of the movement and how that changes from start to end of a trial. Flat blue lines at zero are vertical movements, and flat orange lines at zero are horizontal dragging movements. Notable in this gauge is relative differences in velocities as people move across their bodies, where biomechanics constrain more the larger arm movements for more distant targets. For the right-handed user shown in the present data, the movements are often faster from left-to-right than right-to-left (moving toward versus away from his dominant side). However, not much difference is seen for different movement types when the targets are close to the screen center requiring shorter overall arm movements. Interestingly, the screen camera shows that in some of the furthest finger movements, the user does not always turn his head to look at the starting point of the far movements. Consequently, the head constrains the eye movements to a smaller overall area and do not mirror the motor dynamics in this gauge.

In the study of eye-finger movement dynamics in a task where the targets are constantly changing, the areas of interest (AOIs) are not fixed on the screen. Rather, the AOIs are the relative position of the finger and the eye foveal region. We define three key AOIs: eye-finger in the same location (defined as both being within 64 pixels of each other), eye leading finger (eye closer to target), and finger leading eye (finger closer to target). The scarf plot in the Proximity gauge illustrates the AOIs over the course of the experiment, similar to plots used in [10]. In the scarf, same location is coded in green, eye-leading-finger is coded in orange, and finger-leading-eye is coded in blue. Gray indicates times when no dragging movements were happening. The pattern of AOIs is dominated by orange, with the eye leading the finger most of the time. Second, we see times of green, indicating both are in the same position, often occurring at the end of the trial when people may be verifying they correctly reached the end target location. There are few instances of the finger leading the eye in this data set.

Finally, on the right-hand-side below the scene video, we include two plots of summary data in order to capture how changes in the task difficulty influence eye and finger movement behaviors. The Fitts' Law Regression gauge provides a regression summary of mean data over the index of difficulty values for the experiment. Currently shown is a regression model of the mean finger movement times (blue) and mean eye fixation durations (orange). Regression parameters are summarized in the top, together with Fitts' Law throughput, which indicates amount of information (defined as bits/second) pushed through each type of movement system. As the playback progresses, the scatterplot points for the current block playing in the other gauges are highlighted. Other mean summary data, such as mean saccade duration or length, could also be displayed in these plots.

The Cumulative Hazard Function gauge in the lower right gives functional data for the same summary values. Finger movement times are captured in the blue lines, and fixation duration is shown in orange. Darker colors are for higher index of difficulty conditions (more difficult conditions), and the colors grade to lighter as the index of difficulty decreases (the task gets easier). The curves for the current block are again highlighted during replay. These plots illustrate that changing the difficulty of the target dragging task has the effect of ordering entire functions of behavior. Again, we can add distributions of

total numbers of saccade or saccade speed, or other summary data, for further comparison here.

4 Modeling Visual-Motor Dynamics

To visualize time-series data, particularly the first and second derivatives of the estimated time series, the Dashboard makes use of GP regression [e.g., 9]. GP models provide smooth estimates of the functional shapes of both the hand and eye movement trajectories along with an estimate of the uncertainty associated with those estimates. These models are particularly useful as statistical models for the interpolation between observed points in a time series. Under a Bayesian interpretation of the GP model, the estimate is a posterior distribution over possible functions that pass through (or near) the observed points while simultaneously reflecting uncertainty in interpolated time points. This degree of uncertainty increases with increased distance to the nearest observed values. We chose to use the GP modeling framework because it allows for the alignment and display of multimodal information collected from different apparati with unequal sampling rates using the same fundamental framework. Additionally, the GP modeling approach can provide estimates of the velocity, acceleration, jerk, and other higher order derivatives of path information along with indication of the variability of those estimates.

Formally, the distribution of a GP is uniquely defined by its mean and covariance function. In practice, the choice of covariance function dictates properties of the possible estimated paths through the observed data. Much like univariate and multivariate Gaussian random variables, linear combinations of GPs are also GPs. This means a GP model of the difference between the finger location time-series and the gaze-location time-series is implied by the difference between GP models of the individual time-series data.

The most commonly used covariance function is the radial basis function $cov(s,t) = \tau^2 \exp\left(-\frac{(s-t)^2}{2l^2}\right)$, which we used for both the eye and finger path data. The radial basis function constrains the paths to be relatively smooth curves. Although this is sensible for motor path data, it may be less reasonable for eye trajectory data, which is more irregular due to microsaccadic movements. To accommodate for the fact that eye-gaze data includes both small-scale movements (e.g., micro-saccades, drift) and large-scale movements, we used a combination of two radial-basis functions for gaze data. One radial-basis function was constrained to have a small characteristic length scale, which means that in small time intervals the location can change dramatically. The other radial basis function was constrained to have a large-length scale, which only accounts for relatively slower changes in position. To model this combination in the GP framework, we can simply add the two covariance functions, which is equivalent to modeling the gaze trajectory as a sum of two GPs.

We use a GP model to estimate the velocity and acceleration of the finger and gaze trajectories, because the derivative of a GP is also a GP. The covariance function between an GP at time s and the derivative of the GP at time t' is given by the derivative of the covariance function with respect to its second term. For example, the correlation between the location at s and the velocity at time t' assuming a radial-basis covariance function is,

$$cov(s,t') = \frac{\tau^2(s-t')}{l^2} \exp\left(-\frac{(s-t')^2}{2l^2}\right).$$

The Dashboard illustrates target-relative position, velocity, and acceleration of the mean of the posterior from the GP models fit individually to each trial and with τ and l fixed within blocks. Eye gauges are on the right of the Target Plot gauge, and finger gauges are on the left. Over the course of a block of trials, which contain 26 movements, velocity-by-position and acceleration-by-position phase plots show the variations in speed against the direction of movement. For the block depicted in Figure 2, finger velocity shows differences depending on direction across the body, but acceleration is fairly consistent. Eye velocity-by-position does not show this difference. Velocity-by-acceleration phase plots show consistent variations of speed across a whole block, illustrated in the constant cycles in the bottom left phase plane gauges.

As the task or display constraints are varied, GP models will flexibly capture the systematic variations in movement dynamics. For example, suppose we add constraints that the user must drag a physical target and the user must keep the target inside a path or set of guidelines. We hypothesize that the eye movements would show smooth pursuit of the finger and object as they move along the screen, rather than the single saccade between target locations ahead of the finger observed in this data. The GP framework can also be used to model dynamic changes in distance between the point of visual fixation and the location of the finger on the screen.

5 Future Directions

One strength of the Dashboard display framework is the capacity to display GP models under various covariance assumptions, such as a rational quadratic or Matérn class, independently or simultaneously for comparison purposes. Thus, we can leverage the Dashboard playback for visual comparison of real and model dynamics to determine best models, not just fit statistics. An additional set of Eye Dynamics plots might be added to highlight the microsaccadic movements at a different scale than the saccade scale currently shown. We might also consider adding uncertainty bounds on the estimated values here, for the full statistical model. Additionally, the GP models implemented in the Dashboard provide parameter estimates needed to populate computational human cognition and motor control models. For example, the velocity and acceleration from the GP model are needed in the bang-bang model of human control theory. This model posits that motor actions follow a trajectory of maximum acceleration in the direction of the target up to the midpoint, then maximum deceleration to come to a complete stop at the target [6]. By populating this model with estimates derived in the Dashboard, it can simulate realistic behavior for novel visual-motor interaction situations. These model predictions can be incorporated into our current Dashboard for streamlined evaluation of the model predictions in the same gauges as real human data.

The interaction of visual and motor systems is known to have measurable effects on human performance, and even in the small sample shown here, we can observe strong differences in control dynamics based on the difficulty of this simple visual dragging task. In future task applications, the Dashboard enables consistent visualizations and statistical interpretations of visual and motor dynamics as both the visual objects and types of interaction are varied. This is critical for developing maximally effective and efficient interactive information visualizations.

Acknowledgments

This work was supported in part by AFOSR LRIR to L.M.B. and AFOSR grant FA9550-13-1-0087 to J.W.H. Distribution A: Approved for public release; distribution unlimited. 88ABW Cleared 08/26/2015; 88ABW-2015-4098.

References

[1] M. Bostock, V. Ogievetsky, and J. Heer. D3: Data driven documents. *IEEE Transactions on Visualization and Computer Graphics*, pages 2301–2309, 2011.

[2] A. Cöltekin, U. Demsar, A. Brychtová, and J. Vandrol. Eye-hand coordination during visual search on geographic displays. In *Proceedings of the 2nd International Workshop on Eye Tracking for Spatial Research (ET4S 2014)*, 2014.

[3] U. Demšar and A. Çöltekin. Quantifying the interactions between eye and mouse movements on spatial visual interfaces through trajectory visualisations. In *Vienna, Austria: Workshop on Analysis of Movement Data at GIScience*, pages 23–26, 2014.

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

[5] J. Y. Han. Low-cost multi-touch sensing through frustrated total internal reflection. In *Proceedings of the 18th annual ACM symposium on User interface software and technology*, pages 115–118. ACM, 2005.

[6] R. J. Jagacinski and J. M. Flach. *Control theory for humans: Quantitative approaches to modeling performance*. CRC Press, 2003.

[7] X. Li, A. Çöltekin, and M.-J. Kraak. Visual exploration of eye movement data using the space-time-cube. In *Geographic information science*, pages 295–309. Springer, 2010.

[8] R. E. Patterson, L. M. Blaha, G. G. Grinstein, K. K. Liggett, D. E. Kaveney, K. C. Sheldon, P. R. Havig, and J. A. Moore. A human cognition framework for information visualization. *Computers & Graphics*, 42:42–58, 2014.

[9] C. E. Rasmussen, , and C. K. I. Williams. *Gaussian processes for machine learning*. The MIT Press, 2006.

[10] D. C. Richardson and R. Dale. Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive science*, 29(6):1045–1060, 2005.

[11] R. W. Soukoreff and I. S. MacKenzie. Towards a standard for pointing device evaluation, perspectives on 27 years of fitts law research in hci. *International journal of human-computer studies*, 61(6):751–789, 2004.