*DHM2020*
*L. Hanson et al. (Eds.)*

*doi:10.3233/ATDE200010*

# Automatic Selection of Viewpoint for Digital Human Modelling

Erik BILLING [a,1], Elpida BAMPOUNI [a], and Maurice LAMB [a]
[a] *Interaction Lab, University of Skövde, Sweden*

**Abstract.** During concept design of new vehicles, work places, and other complex artifacts, it is critical to assess positioning of instruments and regulators from the perspective of the end user. One common way to do these kinds of assessments during early product development is by the use of Digital Human Modelling (DHM). DHM tools are able to produce detailed simulations, including vision. Many of these tools comprise evaluations of direct vision and some tools are also able to assess other perceptual features. However, to our knowledge, all DHM tools available today require manual selection of manikin viewpoint. This can be both cumbersome and difficult, and requires that the DHM user possesses detailed knowledge about visual behavior of the workers in the task being modelled. In the present study, we take the first steps towards an automatic selection of viewpoint through a computational model of eye-hand coordination. We here report descriptive statistics on visual behavior in a pick-and-place task executed in virtual reality. During reaching actions, results reveal a very high degree of eye-gaze towards the target object. Participants look at the target object at least once during basically every trial, even during a repetitive action. The object remains focused during large proportions of the reaching action, even when participants are forced to move in order to reach the object. These results are in line with previous research on eye-hand coordination and suggest that DHM tools should, by default, set the viewpoint to match the manikin's grasping location.

**Keywords.** Cognitive modelling, Digital Human Modelling, Eye-hand coordination

## 1. Introduction

With the shift towards Industry 4.0 and an increased use of simulation and visualization software in all design phases, Digital Human Modelling (DHM) has become an important asset allowing engineers to evaluate ergonomics of new products and workplaces early in the design phase, long before the first physical prototype is built. While most DHM tools were primarily developed to evaluate aspects of physical ergonomics, there has recently been increased interest in modelling various aspects of human cognition [1]. Many DHM tools today allow modelling of view cones or volumetric projections. These tools are commonly used to evaluate driver environments by identifying blind spots around the vehicle [e.g. 2].

[1] Corresponding Author: Erik Billing, University of Skövde, Högskolevägen, 541 28 Skövde, Sweden; E-mail: erik.billing@his.se.

While these tools can effectively determine what is visible, a realistic estimate of what a DHM manikin sees also requires simulation of where they are looking [3]. Currently, a manikin's viewpoint is either manually specified by the designer or the simple result of other task related movements, e.g. the viewpoint moves forward and down as the manikin leans forward to reach for an object. However, from research on visual attention, it is well known that human beings display specific eye-gaze behaviors during daily activities [4], including reaching actions [5]. As a result, accurate manual specification of viewpoints for individual tasks is a challenging and potentially impossible process, but one that can be supported by both designer training and automated software features grounded in vision research.

In the present work, we take the first steps towards automating selection of viewpoint, with the goal of improving the quality of DHM simulation while also reducing the time and effort needed to create these simulations. Since object manipulation in the form of reaching and grasping are very common elements of tasks simulated using DHM software, the present study investigates eye-gaze behavior in a pick-and-place (PAP) task. This is done by recording eye-gaze of 10 participants executing a PAP task in a virtual reality (VR) environment. Collected data is analyzed and validated against existing literature. Based on these findings we present 4 principles for automatic selection of realistic viewpoints in DHM software.

The rest of this paper is organized as follows. Previous research on visual attention, eye-hand coordination, and modelling is presented in Section 2. The procedure for the experimental study is presented in Section 3, followed by analysis and presentation of results in Section 4. Finally, the paper concludes with a discussion in Section 5.

## 2. Background

Our selection of where to look is tightly linked to motions of the rest of our body and highly consistent between individuals [6, 7]. Visual attention is typically discussed in terms of *overt attention*, referring to the act of selecting fixation points, and *covert attention*, which refers to a psychological shift of focus without moving the eyes. In a DHM context, we are primarily interested in the former, often not only in eye-motion, but also in head motion and relevant shifts in posture [7, 8]. Thus, realistic simulation of overt attention is not only relevant from a cognitive perspective, but may also improve ergonomic simulations.

Overt visual attention is mostly studied as a response to stimuli in the form of an image. In this context, bottom-up control of eye-gaze refers to the effect that the image has on eye motion, while top-down control refers to cognitive influences and task context [e.g. 9]. In these studies, participants are typically seated in a relatively static position, though recent advances in eye-tracking allow for more naturalistic participant activity.

Extensive experimental research on eye-hand coordination over the last 40 years has established a high correlation between eye-gaze and hand position [e.g. 5, 7, 10, 11]. This correlation is proactive in the sense that the eyes specify the target for the upcoming action. When reaching for an object, we will typically first fixate on that object, and then the hand follows executing the reach. The eye-gaze will remain fixated on the target until the reaching action is completed. As soon as the action is finished, the gaze will shift to the next goal in a highly coordinated fashion [10]. A similar pattern emerges during transportation of objects. People first look at the object being grasped, then shift their gaze towards the goal of the action. During object transport, the gaze typically remains fixated on the goal location until the release of the object occurs [5]. This proactive relationship between eye-gaze and body motion is likely to stem from simultaneous neural activity connecting both eye and body, while the temporal relationship is primarily resulting from differences in inertial properties between eye and body [10].

While this basic proactive eye-gaze pattern appears very stable between individuals, it is affected by several factors. Johansson et al. [5] demonstrate that participants consistently fixate on obstacles during object manipulation, shifting gaze towards the goal only after the hand has passed the obstacle. In cases involving sub-tasks or usage of both hands, earlier gaze transitions may occur in order to visually guide other pending tasks [12, 13].

The dominant approach to modelling visual attention is Koch and Ullman's [14] saliency map method. Models of visual attention is also a big research field in computer vision, see Borji et al. [15] for a review of 65 different models. The vast majority of these are bottom-up models, but there are also mixed models that incorporate both bottom-up and top-down aspects.

Although there is a growing body of literature describing visual attention in more ecological task related terms (see Hayhoe & Ballard [16] for a review), this approach to modelling eye-gaze is less developed. Moreover, careful consideration of eye-gaze in relation to the rest of the body is rare also in this area. One recent initiative was taken by Abboott et al. [11] who proposed a linear autocorrelation model of eye-body coordination. While this model is highly relevant for the present work, it requires training data from real participants executing the task and is thus difficult to apply in a DHM setting.

## 3. Method

The present study investigates a PAP task in VR where participants pick up objects lying on a table and place them in a bin (see Figure 1). Each participant performed 150 trials, in which a single object appeared on the left side of the table at a random distance between 0.2 and 1.4. meters from the participant's starting location. The aim of this layout was to ensure that in some trials the participant was able to pick up the object from the initial starting position, i.e. direct reach (DR), while other trials required the participant to walk before reaching for the object, i.e. relocate and reach (RAR).

During the experiment, participants were wearing a HTC Vive Pro Eye VR headset and two motion trackers (HTC Vive Trackers) placed on the right shoulder and left ankle. Participants also held an HTC Vive controller in their right hand. The virtual environment and task interactions were developed and run using Unity3D 2019.4 LTS. The virtual environment was comprised of a sparse office room (see Figure 1) that contained a table, a bin, and the object to be moved. A physical table of $95x160$cm standing at 87cm height was placed in the middle of the lab space and configured to precisely match the position of its virtual counterpart, allowing participants to touch and lean on the (virtual) table. The bin and objects were solely virtual with the bin located 120cm from the table. Participants were allowed to move around freely within the open physical space. The exact position of headset, hand controller, and motion sensors were tracked using

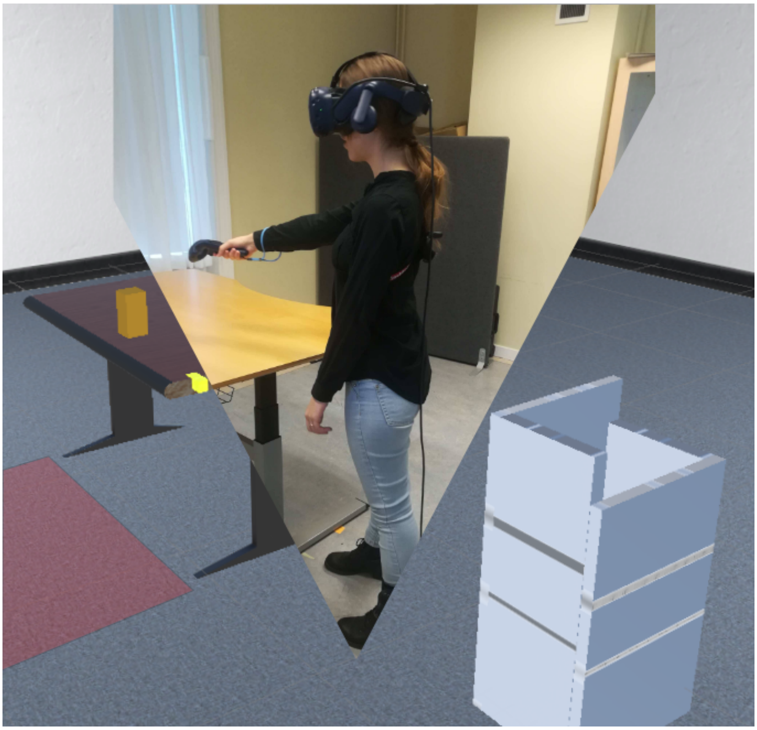

**Figure 1.** Example image of the experimental setting in VR and the physical room (overlay). Second author demonstrating. When participants moved into the red region on the floor, their action was classified as RAR, otherwise the action was classified as DR, (for analysis only, not visible to participants).

the SteamVR 2.0 tracking system. In order to avoid accidents, the limits of the physical space were indicated in the virtual environment using the VR system's built-in boundary system. The task object was a dark yellow box of size $10x10x20$cm and the participant could see their hand position represented by a virtual version of the hand controller.

### 3.1. Procedure

A detailed script had been constructed and was followed for each participant in order to ensure similar treatment. First, written and verbal consent was obtained from each participant after a brief description of the task expectations. After the consent process a more detailed explanation of what their task entailed was provided. This included verbal and visual instructions regarding the use of VR equipment in order to complete the task. Participants were reminded that they could take breaks or withdraw consent and cease their participation at any time without having to explain themselves and without losing any compensation.

An identification number was assigned to ensure data de-identification prior to eye tracking calibration and the beginning of the task. Upon entering the virtual lab room, participants were given the instruction *"You will be picking up boxes and placing them into the bin. Some of the boxes might be too far to reach and so you may move around the table to grasp them"*. After all the trials were finished, participants would sit next to the experimenter and verbally answer a questionnaire regarding the task, demographics, and previous VR experience. Physical measurements of height, arm and leg length were also recorded. Participants then received a cinema ticket as compensation and were provided the opportunity to ask study-related questions during the debriefing. The complete procedure took about 35 minutes out of which approximately 16 minutes were spent within the virtual environment.

### *3.2. Participants and Data collection*

Data was collected from 12 healthy right handed individuals of age (19 - 29) with full vision or corrected to full vision. Two participants were excluded as a result of not following task instructions and problems with the data collection. The final sample comprised 10 participants with a mean age of 22 years. Four identified as female, 5 as male, and 1 as non binary. All of them were students and all but one had minimal exposure to VR in daily life.

The position and orientation of headset, hand controller, motion trackers and the object were recorded with a temporal resolution of 90 Hz. In addition, eye-gaze vectors, viewpoint in the virtual environment, and fixated object were recorded with a temporal resolution of 120 Hz. Eye gaze data was pulled from the the headset through HTC's SRanipal plugin[2]. An eye gaze vector and origin were used along with the headset's position and orientation to calculate the participant's current focus point in the virtual environment. These values were calculated in real-time in the Unity environment, allowing focus points to be calculated with respect to the first virtual object that the headset relative eye gaze vector intersected in the scene. SRanipal provides processed eye values relative to the headset with a delay from eye record to provided data due to processing of approximately 16.6 ms. As a result, given Unity's 90hz framerate for the HTC Vive, eye tracking values have an expected latency of up to 33.2 ms.

The project was submitted for ethical review to the Ethical Review authority of Sweden (#2020-00677, Umeå) and was found to not require ethical review under Swedish legislation (2003:615).

## 4. Results

For the purpose of analysis, the data from each trial is divided into two phases. The first phase begins with the appearance of the object on the table and ends with the participant grasping the object. We refer to this phase as *appear-to-grasp (ATG)*. The second phase starts with the grasp and ends with the object landing in the bin. This object transport phase will not be analysed in this paper.

On any given trial participants executed one of two action strategies during the ATG phase; DR or RAR. Trials where the participant moved such that the left foot was located at the side of the table during the ATG phase were defined as RAR, otherwise the trial was labeled as a DR (c.f., Figure 1). Out of 1500 trials in total, 620 trials (41%) were labeled as RAR.

[2] HTC's SRanipal plugin for Unity, `https://developer.vive.com/resources/knowledgebase/vive-sranipal-sdk/`

**Table 1.** *The most frequently observed gaze sequences during the ATG phase.*

| Action | Fixation seq. | Frequency | Percentage |
|---|---|---|---|
| DR | table → object | 444 | 71% |
| | table → object → table → object | 76 | 12% |
| | table → object → table | 57 | 9% |
| | object | 42 | 6% |
| RAR | table → object → floor → object | 58 | 9% |
| | table → object → table → object | 46 | 7% |
| | table → object | 44 | 7% |
| | table → floor → object | 44 | 7% |

### 4.1. Fixation points

In order to analyze where people were looking when reaching for objects, a low-pass filter of 15 frames (8 Hz) was applied to the eye-gaze data. Gaze fixations were identified and grouped into 6 categories of eye-focus points: *hand, object, bin, table, floor,* and *surrounding*. Surrounding included gazes towards e.g. walls and ceiling. Gazes within a radius of 15 cm from the center point of the object were labeled as ”object”.

The most frequently observed fixation sequences during ATG are presented in Table 1. In most cases, participants started a trial by looking at the table and thereafter shifting their gaze to the object once it had appeared. During DR trials, the object typically remains fixated upon until grasp, sometimes combined with a gaze shift to the surrounding area on the table. The object was fixated at least once during reach in 99% of all DR trials, and remained fixated during 65% of the reaching phase. These numbers are even higher if we include the immediate surrounding of the object. 82% of all gazes during DR landed within 30 cm from the object.

In a small proportion of all DR trials (6%), the ATG phase begins with the participant looking at the object's appearance location. Relevantly, the object initial position varies randomly along the dimension extending from the participant, but not along the orthogonal dimensions. Thus, in these cases, it is likely that the participant is looking at the table in the task relevant region before the ATG phase, and just happens to be looking in a region that is occluded by the object when it appears.

RAR trials comprise much more variability in fixation sequences but typically involve combinations of the object, table, and the floor. Notably, the four most common fixation sequences only account for 30% of all trials. Similar to DR, the object was fixated at least once in 99% of all RAR trials, and remained fixated during 47% of the reaching phase (60% within 30 cm from the object).

As can be seen in Figure 2, the differences between DR and RAR are also visible in the eye-gaze distribution over the room. While there are fixations around the room in both conditions, gazes during DR are more tightly grouped along the left side of the table (where the object appears). As indicated by the sequential analysis presented in Table 1, the RAR trials frequently (but not always) comprise gaze shifts towards the floor. Here we can see that these gazes are primarily located within the region where the participant is about to go, but also include scanning gazes around the room.

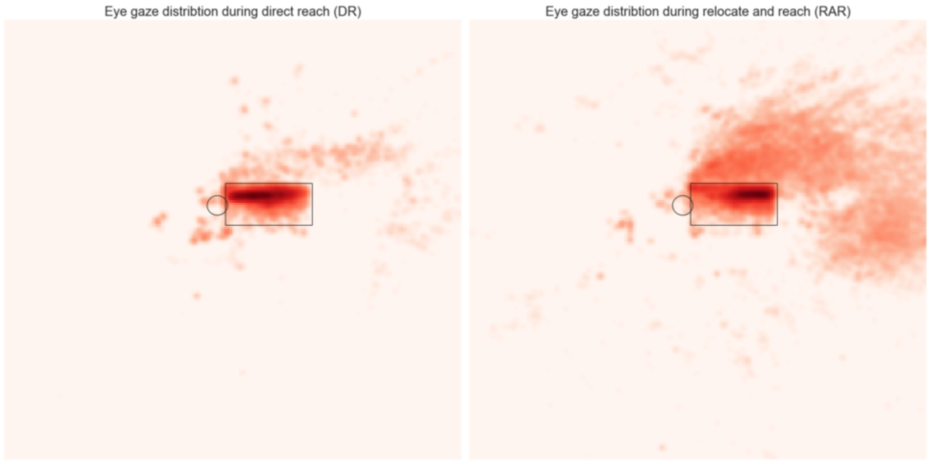

**Figure 2.** Distribution of eye-gaze during the ATG phase, over the room seen from above. Circle and rectangle represents the initial location of the participant and the location of the table, respectively. Intensity of red represents frequency of eye-gaze, scaled logarithmicly to make sparse gaze patterns more visible.

### 4.2. *Temporal dynamics*

The analysis presented in the previous section provides insight on *where* participants are looking when reaching for objects. This section will target the *when*. One dominant pattern observed during both DR and RAR trials was an initial shift of gaze from the table to the object. In order to analyse when this shift occurs, the time of the first gaze on the object was calculated relative to action onset. Action onset was here defined as 5% of peak velocity of the right hand, along the X-dimension (forward in relation to the participant's starting position). Results are presented in Figure 3.

We are also interested in the temporal coordination between the eye and the hand motion. In the present task, this is primarily visible in the sideways motion of the participant, when reaching for the object. If previous literature stipulating a strong coordination between eye-gaze and hand motion is correct, we expect to see the hand motion following the eye-focus point, with a certain delay. Results are presented in Figure 4. Here we also include the motion of the left foot, which has been less studied in relation to eye-gaze. Interestingly, results reveal a high correlation between both eye-hand and hand-ankle, with an approximate delay of about 1.0 s and 0.3 s, respectively.

## 5. Discussion

In the present work, we have taken the first steps towards a model for automatic selection of viewpoint. While the present study does not result in a fully developed computational model, we suggest the following principles for automatic viewpoint selection in DHM software. These principles are based on the experimental results from the present study combined with previous literature. Although these formulations do not account for the full variability in human visual behavior, we hope that they will constitute *implementable* formulations of automatic viewpoint selection in DHM tools.

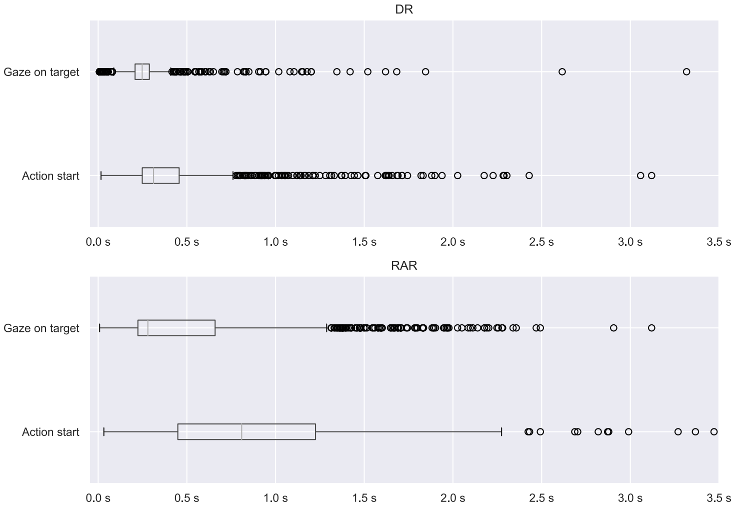

**Figure 3.** Timing of eye-gaze on target object and action onset relative to the appearance of the object on the table.

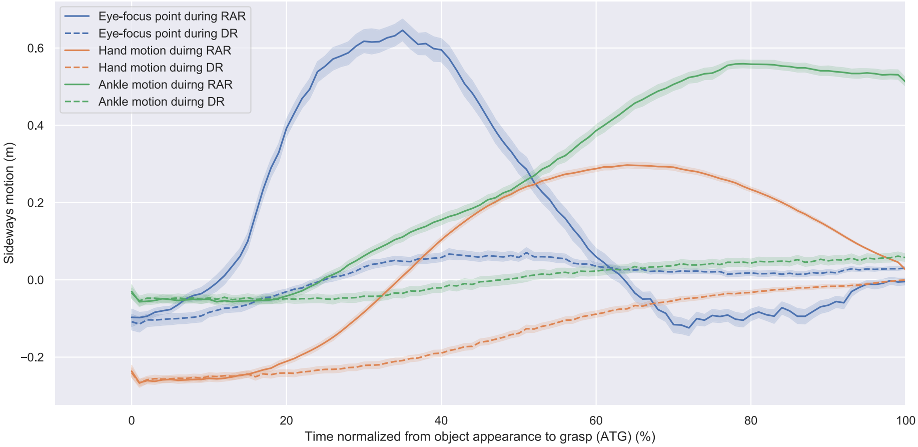

**Figure 4.** Sideways motion of the eye-focus point, hand, and ankle during the ATG phase, relative to the position of the target object. The plot shows mean values from all 10 participants. Envelopes represents the 95% confidence interval of the mean.

1. *Eyes on target.* This principle is highly dominating during direct reaching behaviors when the participant is standing still, and aligns well with previous research [e.g. 5, 7]. Somewhat surprisingly, this appears to be also valid when participants relocate, as more than half of all gazes were directed on to the object or in its near surrounding in RAR trials.
2. *Eyes on floor.* Similar to previous research [5], we here observe a shorter fixation on intermediate targets, in this case the floor where participants are about to relocate. This intermediate eye-gaze appears to interrupt the otherwise dominating

principle of looking at the main target for the action.

3. *Coordinated onset.* As visible in Figure 3, the first eye-gaze on target object appears to be highly coordinated with the onset of the action. Here, there is a greater variability in previous literature and there are certainly many situations where individuals look at targets long before initiating an action. However, our results align well with the seminal study of Biguer et al. [10], describing eye-gaze, head motion, and hand motion during reaching as initiated by a single ”neural command”. Although there may be exceptions, we believe this principle constitutes a simple and yet realistic approach to the timing between eye-gaze and reaching actions for DHM tools.
4. *Eyes in the lead.* Although onset of eye movements appear to be synchronous with the action, the eye-gaze is leading motion of the rest of the body. This has been repeatedly expressed in literature, and is also obvious in our data (Figure 4). We believe this constitutes good basis for modelling the relationship between hand motion and eye-gaze, while recognizing the fact that DHM tool developers may want to describe eye-gaze as a product of hand movements; the opposite is likely a more accurate description of the cognitive mechanisms.

## Acknowledgements

This work was financially supported by the synergy Virtual Ergonomics funded by the Swedish Knowledge Foundation, dnr 20180167. `https://www.his.se/sve`

## References

[1] Scataglini S, Paul G, editors. DHM and Posturography. 1st ed. London: Elsevier, Academic Press; 2019.

[2] Summerskill S, Marshall R, Cook S, Lenard J, Richardson J. The use of volumetric projections in Digital Human Modelling software for the identification of Large Goods Vehicle blind spots. Applied Ergonomics. 2016;53:267–280.

[3] Fletcher L, Zelinsky A. Driver inattention detection based on eye gaze - Road event correlation. International Journal of Robotics Research. 2009;28(6):774–801.

[4] Raudies F, Mingolla E, Neumann H. Active Gaze control improves optic flow-based segmentation and steering. PLoS ONE. 2012;7(6).

[5] Johansson RS, Westling G, Bäckström A, Flanagan JR. Eye–Hand Coordination in Object Manipulation. Journal of Neuroscience. 2001 9;21(17):6917–6932.

[6] Land MF, Furneaux S. The knowledge base of the oculomotor system. Philosophical Transactions of the Royal Society of London Series B: Biological Sciences. 1997 8;352(1358):1231–1239.

[7] Land MF, Mennie N, Rusted J. The Roles of Vision and Eye Movements in the Control of Activities of Daily Living. Perception. 1999 11;28(11):1311–1328.

[8] Herst AN, Epelboim J, Steinman RM. Temporal coordination of the human head and eye during a natural sequential tapping task. In: Vision Research. vol. 41. Pergamon; 2001. p. 3307–3319.

[9] Schutz AC, Braun DI, Gegenfurtner KR. Eye movements and perception: A selective review. Journal of Vision. 2011 9;11(5):9–9.

[10] Biguer B, Jeannerod M, Prablanc C. The coordination of eye, head, and arm movements during reaching at a single visual target. Experimental Brain Research. 1982;46(2):301–304.

[11] Abbott WW, Harston JA, Faisal AA. Linear Embodied Saliency : a Model of Full-Body Kinematics-based Visual Attention. PrePrint by bioRxiv. 2020;.

[12] Land MF. Vision, eye movements, and natural behavior. Visual Neuroscience. 2009;26(1):51–62.

[13] Srinivasan D, Martin BJ. Eye-hand coordination of symmetric bimanual reaching tasks: Temporal aspects. Experimental Brain Research. 2010;203(2):391–405.

[14] Koch C, Ullman S. Shifts in selective visual attention: Towards the underlying neural circuitry. Human Neurobiology. 1985;4(4):219–227.

[15] Borji A, Itti L. State-of-the-art in visual attention modeling. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2013;35(1):185–207.

[16] Hayhoe M, Ballard D. Modeling task control of eye movements. Cell Press; 2014.