

Eye-Tracking in Practice: A First Analysis of a Study on Human Postures

Jürgen Symanzik ^{*} Chunyang Li [†] Boyu Zhang [‡] Breanna Studenka [§]
Eric McKinney [¶]

Abstract

In this article, we present our first results of a study that tries to determine where people are looking when ranking the stability of an actor holding certain postures. A mobile eye-tracker is used to record the original video data of people looking at human postures. Image processing methods are employed to extract statistical information from the video data. The statistical analyses are based on visualization and machine learning approaches.

Key Words: Eye-Tracking Data; Image Processing; Visualization; Machine Learning; Software; EyeTrackR R Package

1. Introduction

An individual's planning and execution of actions is guided by many things, one of which is the anticipated posture at the end-state of an action. Studies on action planning indicate that future postural states can influence the choice of previous postures (Cohen and Rosenbaum, 2004; Herbort and Butz, 2010; Rosenbaum et al., 2006). For example, when grasping an overturned glass with the intent to pour water into it, the glass is typically grasped with a thumb-down posture so that the glass can be turned upward and the hand can end in a thumb-upward posture. This thumb-upward posture is also rated to be more comfortable than a thumb-down posture, and is therefore one example of the end-state-comfort effect — the observed phenomenon whereby most actions end with a comfortable posture (Rosenbaum et al., 1990).

The aim of our planned posture study will be to determine whether judgment of other's action capabilities is based on one's own action experiences. We will examine the stability judgments of individuals viewing an actor holding different postures. Two groups of participants will be examined: one group with extensive yoga experience, and one group with minimal experience with actions that require stability (e.g., yoga, gymnastics, ballet dancing, etc.). We will equip each participant from each group with mobile eye-tracking equipment that will enable us to see what they are looking at as they make stability judgments. We hypothesize that perceptions of others' stability will be influenced by the unique experiences of participants. More specifically, we hypothesize that those with extensive yoga experience will judge an actor to be more stable than those without stability-specific experience. Furthermore, we hypothesize that the visual information used to judge stability will differ between different groups of individuals with unique action experiences.

^{*}Department of Mathematics and Statistics, Utah State University, Logan, UT 84322–3900, USA. E-mail: symanzik@math.usu.edu

[†]Department of Mathematics and Statistics, Utah State University, Logan, UT 84322–3900, USA. E-mail: lichunyang1990@hotmail.com

[‡]Department of Computer Science, Utah State University, Logan, UT 84322–3900, USA. E-mail: boyu.zhang@aggiemail.usu.edu

[§]Department of Kinesiology and Health Science, Utah State University, Logan, UT 84322–3900, USA. E-mail: Breanna.studenka@usu.edu

[¶]Department of Mathematics and Statistics, Utah State University, Logan, UT 84322–3900, USA. E-mail: ericmckinney77@gmail.com

This article is structured as follows: In Section 2, we provide further details on the planned posture study. In Section 3, we present preliminary results from two participants that were used to evaluate the final setup of the full posture study and to assess the functionality of the data recording and data processing. All of our visualizations and analyses are conducted with the R statistical computing platform (R Core Team, 2017). In Section 4, we provide an outlook at the upcoming steps in our full posture study to be conducted in the 2017/2018 academic year.

2. The Posture Study

“Does judging the action capabilities of another person depend on one’s own experiences?” This is the primary research question for our planned posture study. For many human interactions, action anticipation must be present when interacting with others (e.g., to avoid collisions, pass something on to someone, etc.). This study is motivated by research conducted in the Kinesiology and Health Science Department at Utah State University (USU).

Most literature examining ones perceptions of others’ action capabilities involves asking subjects to estimate the performance of another individual, e.g., “can Sally reach the block or not?” (Rochat, 1995). The visual information used by the participant in making their judgments has been limited to making a participant to identify in what direction another person might be looking or what they might be looking at (Michelon and Zacks, 2006). We are unaware of any study looking at the use of eye-tracking to determine what particular visual information about an actor might be used for judgments about action capabilities. Therefore, no literature exists to describe how one might set up and conduct such a human posture study, in particular when incorporating mobile eye-tracking equipment as a major component of the data recording, data processing, and data analysis. Numerous preliminary test participants participated to help evaluate and refine the setup of the study. Data and results from two of these preliminary test participants are presented here. Both participated under conditions similar to the final setup of the posture study.

There will be two groups of participants in the full posture study. In Group 1, there will be 20 students with minimal experience with actions that require stability (e.g., yoga, gymnastics, ballet dancing, etc.) from the undergraduate Psychology pool at USU. In Group 2, there will be 20 students with extensive yoga experience from advanced yoga classes at USU. We anticipate that those with extensive yoga experience will judge an actor to be more stable than those without stability-specific experience. Moreover, we expect that the visual information used to judge stability will differ between different groups of individuals with unique action experiences.

Each of the 40 participants will be shown 22 pictures of a single actor holding a posture. Twenty-two different postures were derived by systematically developing one posture per axis of rotation and within each axis of rotation, choosing three postures of increasing difficulty. Three postures were chosen for sagittal, frontal, and transverse plane rotation, for the combination of sagittal and frontal, sagittal and transverse, and frontal and transverse, and for the combination of all three planes of rotation.¹ Therefore, 21 postures were developed with the addition of a posture with no rotation. The difficulty of the three postures within each rotation was arbitrarily determined by the experimenter. Figure 1 shows six of these postures.

All postures were shown to each participant in random order, projected to a wall, one at a time. Participants were required to judge the stability of each posture (i.e., how long the person could hold the posture). Possible answers were < 1sec, 1 – 10sec, 11sec – 1min,

¹The sagittal plane divides the human body into left and right halves, the frontal plane into front and back halves, and the transverse plane into superior and inferior halves.



Figure 1: Postures 1 to 6 (out of 22) used in the study.

1min – 1hour, and > 1hour. To avoid ambiguity, the intervals will be slightly modified in the full study: $\leq 1\text{sec}$, $> 1\text{sec} - 10\text{sec}$, $> 10\text{sec} - 1\text{min}$, $> 1\text{min} - 1\text{hour}$, and $> 1\text{hour}$. The participants wore a mobile eye-tracking device from Applied Science Laboratories (ASL), Bedford, MA, for the entire study. This device uses one camera that continuously records the (changing) scene that is visible for the participant at a frequency of 30 Hz. Moreover, the focal point of the right eye is recorded by a second camera of this device and is overlaid as a cross-hair on each video frame of the recorded scene video. Details on these eye-tracker video recordings and how to extract the focal point of the eye in each video frame and translate it to the coordinates of a static image of the posture via the EyeTrackR R package can be found in Li and Symanzik (2016, 2017).

In practice, a static eye-tracker could have been used as well for this posture study. However, we decided to use the mobile eye-tracker from ASL for two reasons. First, and most importantly, Fawcett (2016) reported that the static eye-tracker available at USU is not very precise for some areas of the computer screen. Thus, fixations from the posture study may not be matched with the proper body parts. Second, the posture study likely is a precursory step to future, more complex studies that will require the participants to do some physical movements in combination with recording the eye movements, e.g., where does a participant look when picking up an item. Such studies can't be conducted with a static eye-tracker and a mobile eye-tracker must be used instead. Therefore, gaining some practical insights from a simpler experiment (such as the posture study) may be valuable for more complex future studies.

The data obtained from the eye-tracker allows us to answer several secondary research questions: At which body parts (such as head, shoulders, hands, elbows, hips, knees, and feet) do the participants look for each posture? In the context of eye-tracking, these body parts are the areas of interest (AOIs) in each posture and have to be specified once for each posture, using the EyeTrackR R package. How long and how often do the participants look at the body parts in each posture? Are the viewing patterns of the participants associated with the stability assessment? Methods from machine learning may be used to answer

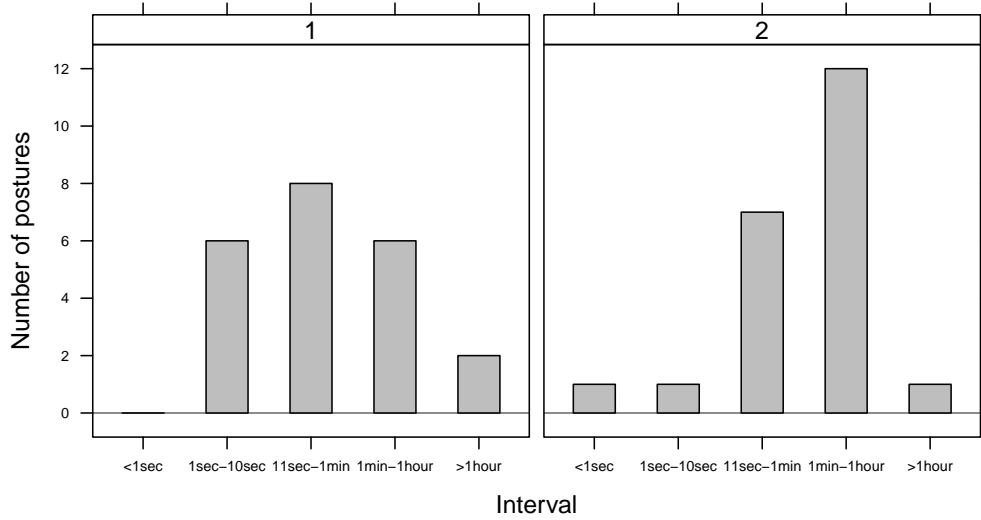


Figure 2: Stability assessment distributions for Participant 1 (left) and Participant 2 (right).

these questions.

The final data collection will consist of 880 answers to the stability assessment (22 postures \times 20 participants \times 2 groups) and 40 recordings, i.e., videos, from the eye-tracker (one for each participant). For the preliminary assessment of two test participants, presented in the next section, 44 answers to the stability assessment (22 postures \times 2 participants) and two video recordings have been analyzed.

3. Exploratory Data Analysis

For the 40 participants of the full study, we expect to be able to answer questions such as:

- What are **within groups** similarities / differences (if any) of the viewing patterns for each posture / for all postures?
- What are **between groups** similarities / differences (if any) of the viewing patterns for each posture / for all postures?

For the analysis of the data of the two test participants, we employed various exploratory data analysis (EDA) methods commonly used for eye-tracking data. These methods are (i) eye scatterplots where the participant's focus points are overlaid on the posture; (ii) attention maps (or hot spot maps) overlaid on the posture; (iii) scanpath maps that show the fixations, i.e., the state when the eye remains stable for a (short) period of time, and the saccades, indicating rapid movement of the eye between two fixations; and (iv) various types of linked microposter plots, introduced in Li and Symanzik (2016, 2017).

The stability assessment distributions for the two participants are shown in Figure 2. While the mode for Participant 1 falls into the 11sec – 1min interval (8 of the 22 postures), the mode for Participant 2 falls into the 1min – 1hour interval (12 of the 22 postures). We expect to see such different distributions for participants from the two groups in our full study as we hypothesize that those with extensive yoga experience (from Group 2) will judge a posture to be more stable than those without stability-specific experience.

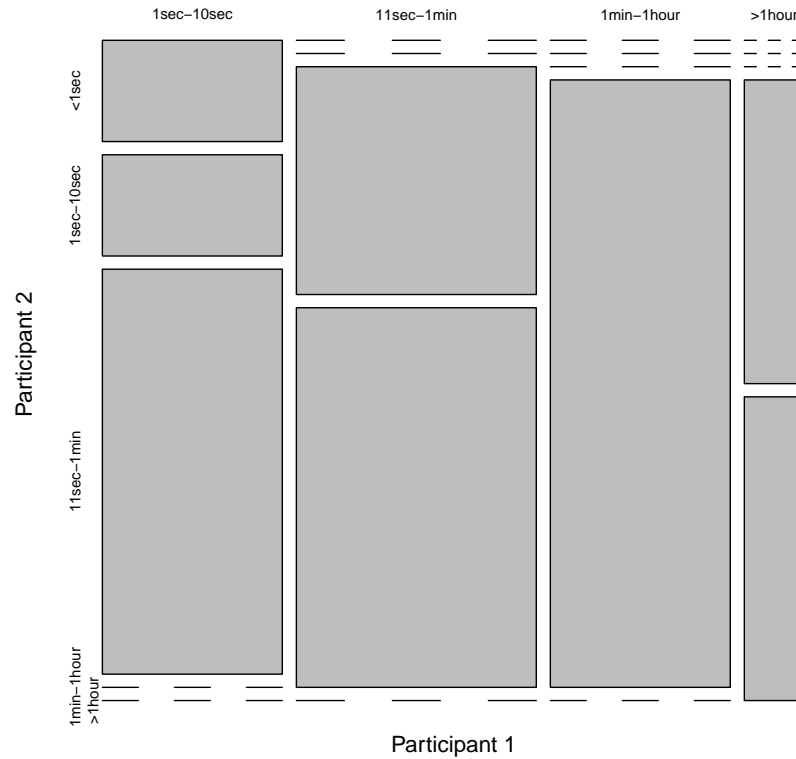


Figure 3: Mosaic plot for the stability assessment comparison of Participant 1 and Participant 2.

For the two test participants, there was agreement on the stability assessment in 50% of the postures (11 out of 22). For the remaining 50% of the postures, there was a one-level difference, as shown in Figure 3. Most notable, whenever Participant 1 indicated that a posture could be kept for 1min – 1hour, Participant 2 came to the same assessment. Eventually, in our full study, we expect to see that in a comparison of participants from Group 1 (on the horizontal axis) and Group 2 (on the vertical axis), more differences will occur in the triangular area below the diagonal representing identical assessments than in the triangular area above the diagonal. This is because of our hypothesis that those with extensive yoga experience (from Group 2) will judge a posture to be more stable than those without stability-specific experience.

For the two test participants, and also for the participants of the full study, we will use EDA methods to determine whether the eye-tracking device is properly calibrated before the participant looks at the 22 postures and remains properly calibrated throughout the entire experiment. This is done by asking the participants to look at four dots outside the posture area, shown one-by-one for about two seconds before the 22 postures. The posture used for the calibration assessment is an electronically modified version of one of the actual postures and does not show up in the actual study. The same four dots also are shown in the same way after the 22 postures. A simple scatterplot of the focal points, overlaid on this additional posture is then used to assess the quality of the calibration, here shown for Participant 1 in Figure 4. Here, before and after the experiment, the coordinates extracted from the eye-tracking device scatter in a mostly random way around the four dots. Considerable mismatches between the coordinates and the dots before the experiment are



Figure 4: Scatterplot showing the focus points, confirming the proper calibration of the eye-tracking device at the start of the experiment (left) and end of the experiment (right) for Participant 1.

a likely indicator of improper calibration of the eye-tracking device. Mismatches between the coordinates and the dots after the experiment could be an indicator of a physical shift or misalignment of the eye-tracking device during the experiment. Unless some geometric adjustments of the coordinates can be made, such recordings may have to be treated carefully and may have to be redone under the same setup with an additional participant from the group. Ultimately, more than 40 participants, may have to participate in the study until 20 participants from each of the two groups with meaningful recordings have been observed.

Theoretically, eight possible combinations of the stability assessment of a posture, the fixations, and the saccades exist when comparing the answers of two participants:

- (i) identical assessments, similar fixations, similar saccades
- (ii) identical assessments, similar fixations, different saccades
- (iii) identical assessments, different fixations, similar saccades
- (iv) identical assessments, different fixations, different saccades
- (v) different assessments, similar fixations, similar saccades
- (vi) different assessments, similar fixations, different saccades
- (vii) different assessments, different fixations, similar saccades
- (viii) different assessments, different fixations, different saccades

We consider two saccades to be similar if they are based on similar fixations in about the same order. It may be possible to define some quantitative similarity measures in the future, based on the data from the 40 participants. The combinations (iii) “identical assessments, different fixations, similar saccades” and (vii) “different assessments, different fixations, similar saccades” are impossible to occur in practice because different fixations automatically result in different saccades. Several of the remaining six combinations were observed for the two test participants presented in this article.

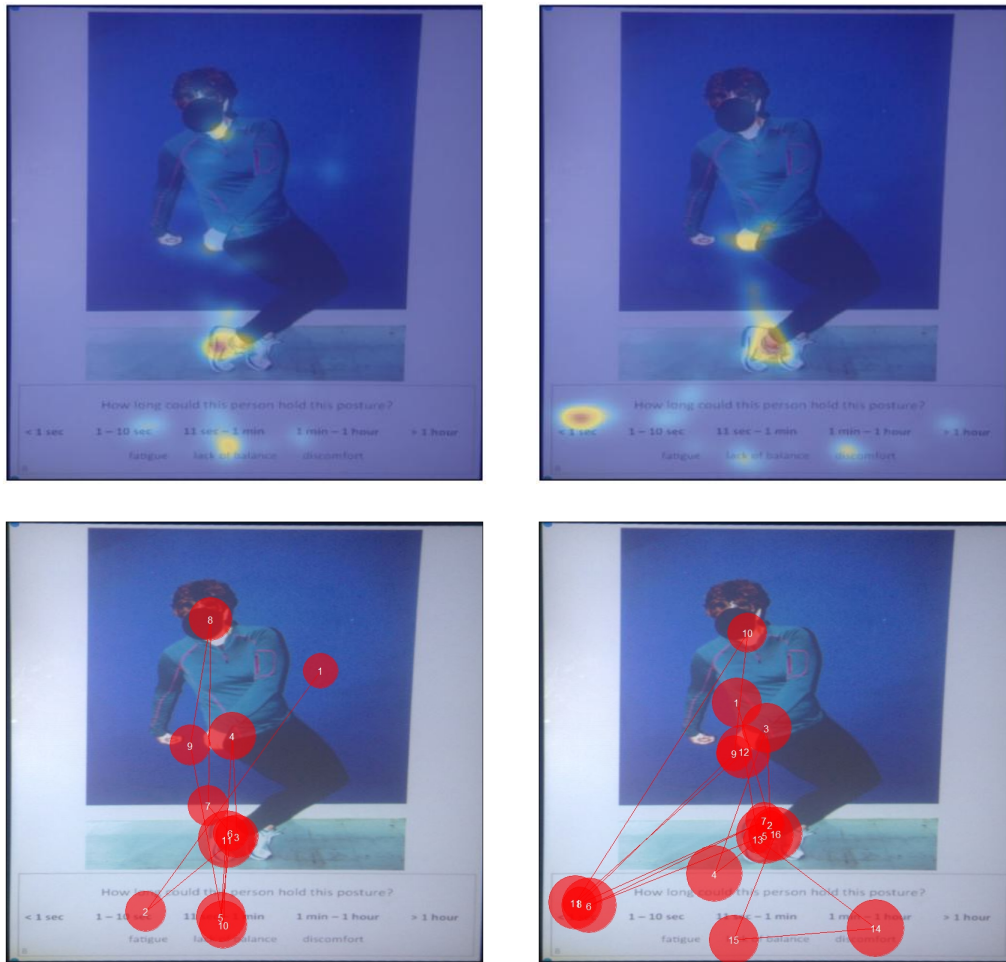


Figure 5: Posture 8: Different assessments (1sec–10sec vs < 1sec), but similar fixations (as shown in the attention maps in the top row) and similar saccades (as shown in the scanpaths in the bottom row) for Participant 1 (left column) and Participant 2 (right column).

For Posture 8, combination (v) occurred. The two participants indicated that this posture could be kept for 1sec – 10sec and < 1sec, respectively. In fact, this was the only posture where one of the two participants indicated that such a posture was highly unstable and could only be kept for less than 1sec. While the assessments are different, the fixations (see Figure 5, top row) and also the saccades (see Figure 5, bottom row) are similar.

For Posture 17, combination (viii) occurred. The two participants indicated that this posture could be kept for 11sec – 1min and 1min – 1hour, respectively. Moreover, the fixations (see Figure 6, top row) and also the saccades (see Figure 6, bottom row) were different.

For Posture 14, combination (i) occurred. Both participants indicated that this posture could be kept for 11sec – 1min. In addition to the assessments, the fixations (see Figure 7, top row) and also the saccades (see Figure 7, bottom row) were similar. However, one might as well argue that there are some additional fixations from Participant 2, looking at the right foot of the actor in this posture. These fixations and related saccades are only visible in the scanpath, but not in the attention map. Some further work remains to be done to determine what constitutes identical (or similar) fixations and scanpaths in the final study.

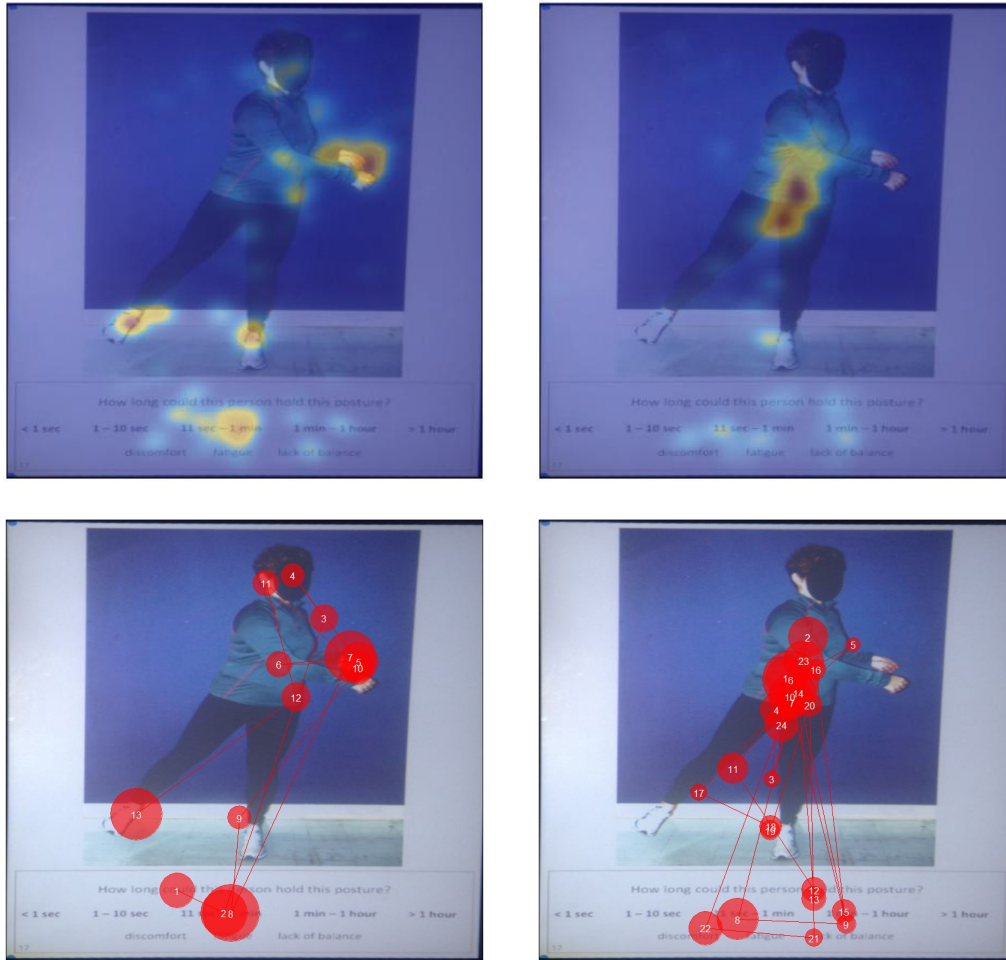


Figure 6: Posture 17: Different assessments (11sec–1min vs 1min–1hour), and also different fixations (as shown in the attention maps in the top row) and different saccades (as shown in the scanpaths in the bottom row) for Participant 1 (left column) and Participant 2 (right column).

3.1 Scanpath Maps vs Linked Scanpath Microposter Plots

In addition to the traditional eye-tracking visualization methods presented so far, it is worthwhile to apply some of the new visualization methods from Li and Symanzik (2016, 2017) to the postures used in this study. These new visualization methods were originally developed for identifying important features of scientific posters. In this section, we will contrast regular scanpaths with linked scanpath microposter plots for Posture 17 (Figure 8) and Posture 14 (Figure 9), both for Participant 2. In a regular scanpath map, all fixations and saccades are overlaid on a single image, i.e., posture in this context. This can often lead to extensive overplotting, making it difficult to identify the labels, sizes, and exact locations of the fixations. In contrast, the fixations and saccades are broken down in several consecutive batches of four to five fixations and saccades in linked scanpath microposter plots. Each of these batches is overlaid individually on an image of the posture. Rather than using the size of circles to distinguish the lengths of the fixations as in scanpath maps, linked scanpath microposter plots contain one (or more) statistical columns where statistical variables can be shown. In this case, there is only one statistical column, showing

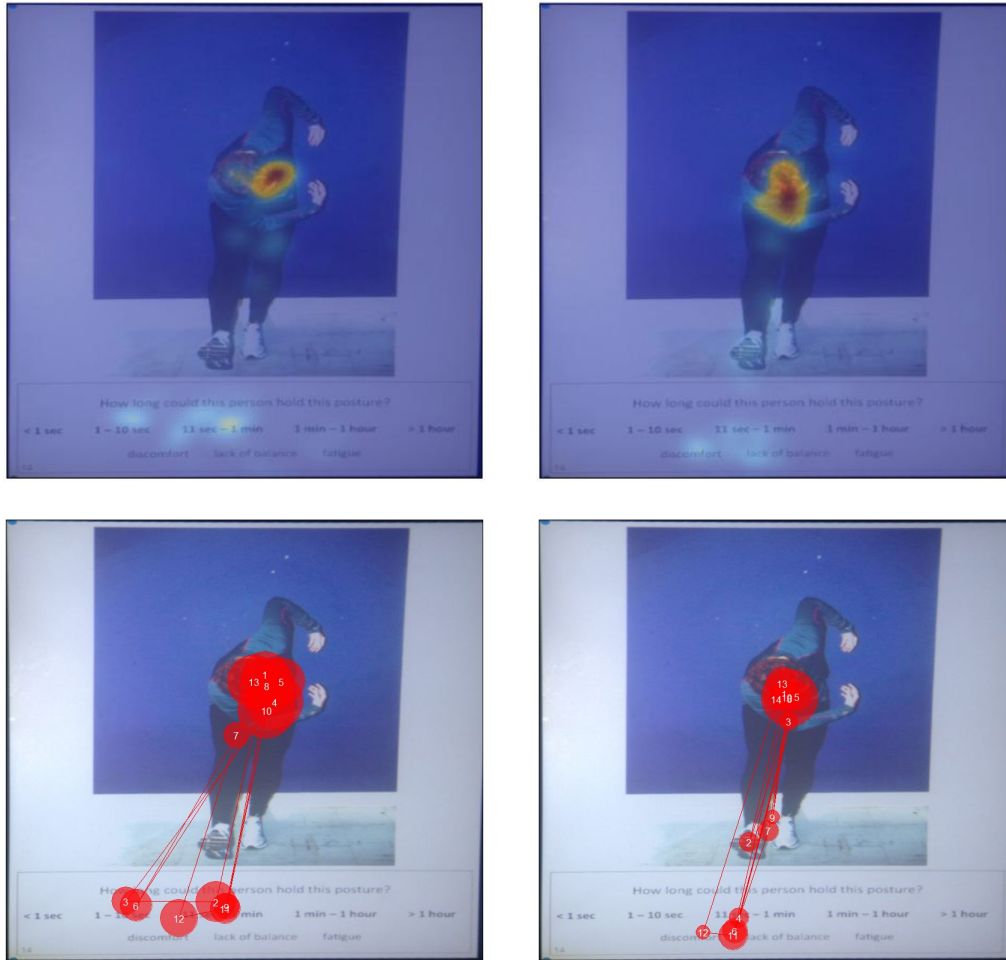


Figure 7: Posture 14: Identical assessments (11sec–1min both), similar fixations (as shown in the attention maps in the top row) and similar saccades (as shown in the scanpaths in the bottom row) for Participant 1 (left column) and Participant 2 (right column).

dotplots for the lengths of each fixation. While most fixations are relatively short (under 1sec), fixation 6 for Posture 17 lasts over 3sec (Figure 8) and fixation 8 for Posture 14 lasts over 4sec (Figure 9).

3.2 The Problem of Detecting Meaningful Fixations

Similar to detecting clusters in higher-dimensional data, detecting meaningful fixations highly depends on the underlying method / algorithm selection and choices of the starting parameters. In fact, fixations could be considered to be clusters in a 3-dimensional space–time coordinate system. The *EyeTrackR* R package makes use of the *saccades* R package (von der Malsburg, 2015). According to the description on its help page, the `detect.fixations` function from this package “takes a data frame containing raw eye-tracking samples and returns a data frame containing fixations.” An input argument for this function, called *lambda*, is defined as follows: “a parameter for tuning the saccade detection. It specifies which multiple of the standard deviation of the velocity distribution should be used as the detection threshold.” It turns out that different choices for *lambda* result in rather large variations in the number of fixations that are detected, and also in the

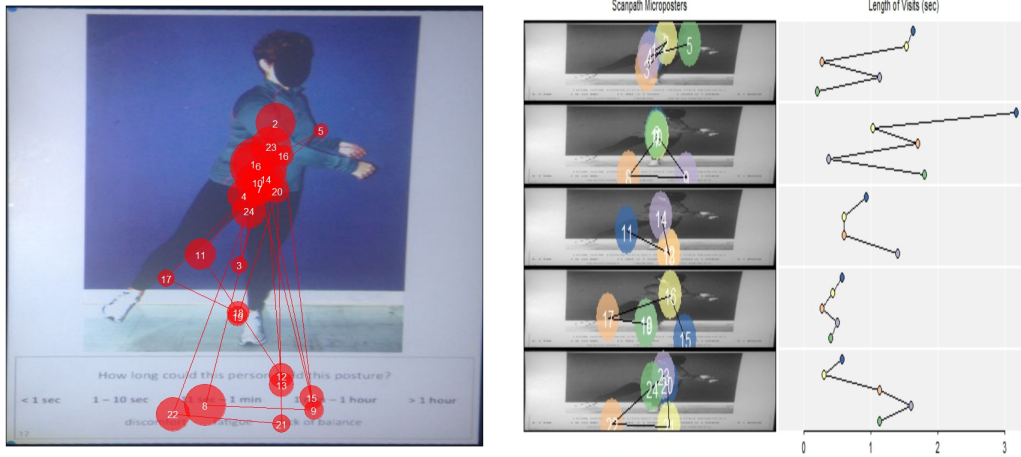


Figure 8: Posture 17: Scanpath map (left) vs linked scanpath microposter plots (right) for Participant 2.

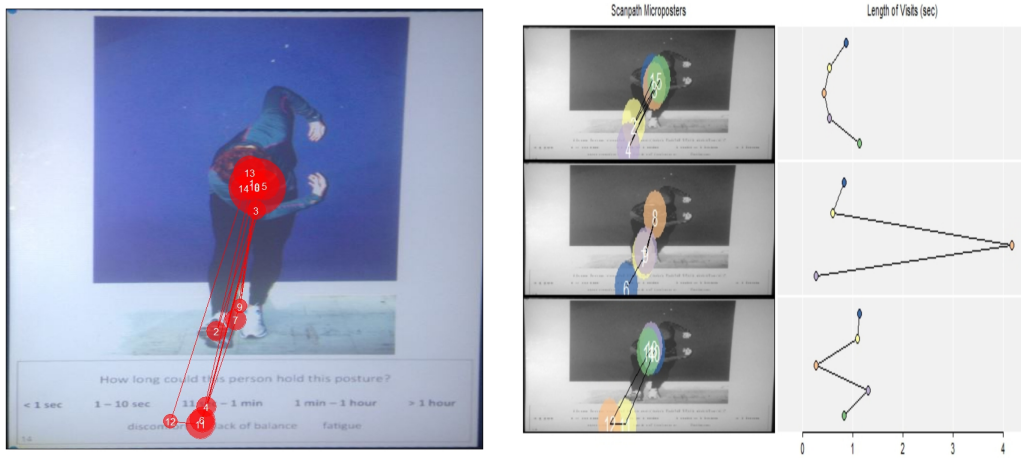


Figure 9: Posture 14: Scanpath map (left) vs linked scanpath microposter plots (right) for Participant 2.

actual locations of these fixations, as outlined in the following three examples.

The `detect.fixations` function from the `saccades` R package uses a velocity-based fixation detection algorithm for saccades, introduced by Engbert and Kliegl (2003). Salvucci and Goldberg (2000) provided a taxonomy of fixation identification algorithms and conducted a qualitative assessment of five different algorithms. They came to the conclusion that “the choice of identification algorithms can dramatically affect the resulting identified fixations.”

Figure 10 shows the fixations and saccades for eight different values of λ for Posture 8 for Participant 2. The number of fixations detected increased from 4 (for $\lambda = 1$) to 16 (for $\lambda = 8$) and then decreased again to 5 (for $\lambda = 15$). A likely fixation at the chin/head area only was detected for $\lambda = 5$ and 8, making these the best choices here. Small values of λ missed most of the fixations in the text area at the bottom. Large values of λ combined several fixations into a single artificial fixation at the spatial center of the combined fixations. This becomes most obvious for $\lambda =$

12 and 15 where no fixation is located at the left hand of the actor.

Figure 11 shows the fixations and saccades for eight different values of λ for Posture 17 for Participant 2. The number of fixations detected increased from 8 (for $\lambda = 1$) to 30 (for $\lambda = 5$) and then decreased again to 11 (for $\lambda = 15$). The smallest value of λ missed most of the fixations in the text area at the bottom. Again, large values of λ combined several fixations. This becomes most obvious for $\lambda = 10$, 12, and 15 where no fixation is located in the lower left of the text area. The best choices again are $\lambda = 5$ and 8.

Figure 12 shows the fixations and saccades for eight different values of λ for Posture 14 for Participant 2. The number of fixations detected increased from 6 (for $\lambda = 1$) to 19 (for $\lambda = 5$) and then decreased again to 11 (for $\lambda = 15$). Small values of λ missed most of the fixations in the text area at the bottom. Once more, large values of λ combined several fixations. This becomes most obvious for $\lambda = 15$ where an additional fixation is located in the upper leg area of the actor. The best choices once more are $\lambda = 5$ and 8.

Based on the observations from Figures 10 to 12, $\lambda = 8$ has been selected for both participants for the scanpath maps and linked scanpath microposter plots from Figures 5 to 9. While this is a reasonable choice here, this needs some careful reassessment in the full study.

4. Outlook on Main Study

The preliminary results for the two test participants presented in this article were obtained after numerous small changes and modifications over the course of several months. Some of the questions we had to answer were how to present the postures, e.g., as printouts or projected on a wall. How far should the participants be located away from the print-out/projection? Should the participants stand or be seated — and should they be allowed to freely move around? Other questions were related to the actual presentation of the postures, e.g., what kind of background or frame should be used around the actual images of the postures and how to separate consecutive postures so that there is no overlap of the eye position from a previous posture when the next posture is shown. Ultimately, after deciding on projecting the postures on a white wall, questions arose which computer and projector to use and how to adjust the lighting in the room (as different computers and projectors have different light intensities and resolutions).

Ultimately, the results from the two participants presented here were obtained from the final setup, except one remaining modification: The text area of the postures still shows a possible reason why a posture has to be given up (with options “fatigue,” “lack of balance,” and “discomfort”). These options will no longer be shown in the full study and the participants won’t be asked any related questions.

The eye-tracking technology provided satisfactory and detailed information about which body parts the participants looked at that hopefully will allow us to draw further quantitative conclusions beyond EDA in the full study. Nevertheless, EDA will remain an important component for the analysis of the data from the full study, in particular to confirm the proper calibration of the eye-tracking device (as seen in Figure 4) and to assess how meaningful automatically identified fixations are (as discussed in Section 3.2). Scatterplots, attention maps, scanpath maps, linked scanpath microposter plots, and other visualization methods will be effective means to summarize and communicate the data and results from the participants in the full study.

At this time, we believe we are ready to start with the data collection for the full study in the 2017/2018 academic year. The necessary approval from the Institutional Review Board

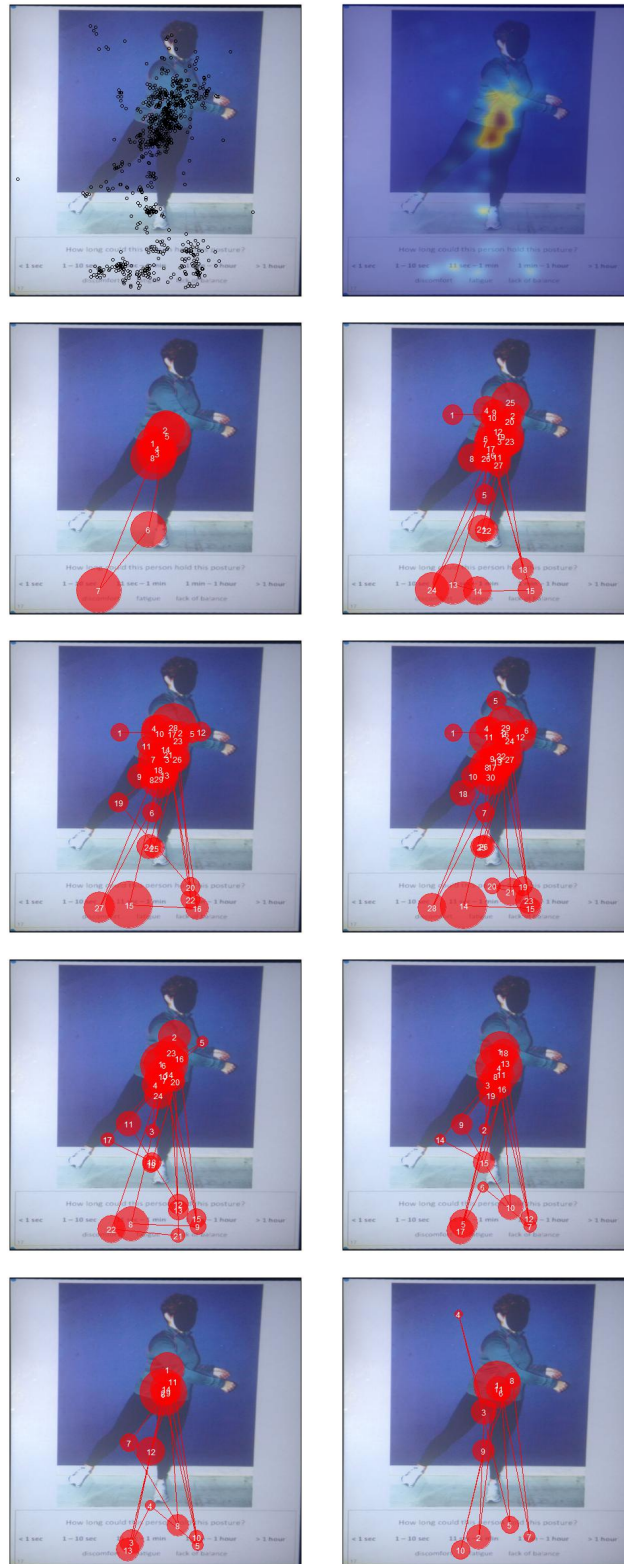


Figure 11: Posture 17: Different fixations detected for Participant 2, based on different values for λ (1, 2, 3, 5, 8, 10, 12, 15 row-wise from top left to bottom right). For comparison, a scatterplot and attention map of the underlying focus points are shown in the top row.

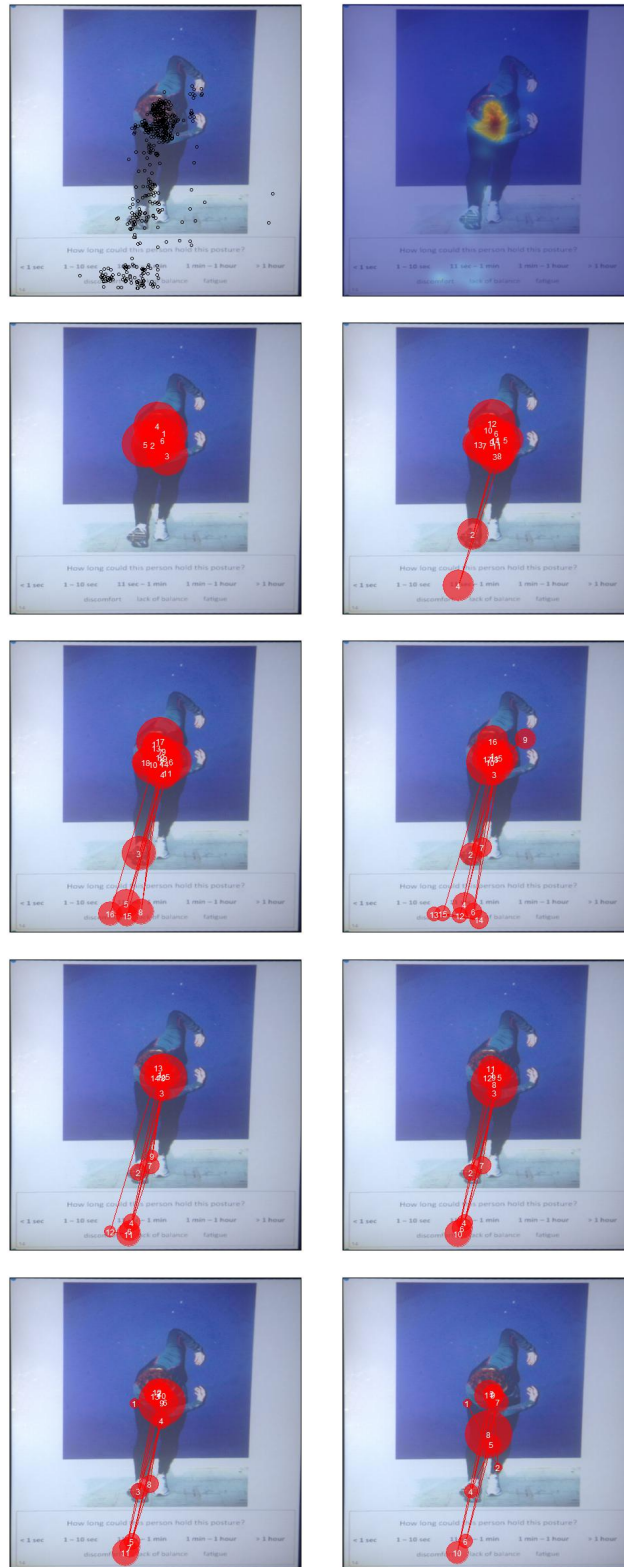


Figure 12: Posture 14: Different fixations detected for Participant 2, based on different values for λ (1, 2, 3, 5, 8, 10, 12, 15 row-wise from top left to bottom right). For comparison, a scatterplot and attention map of the underlying focus points are shown in the top row.

(IRB) at USU has been obtained and fliers that advertise the participation in this study are ready to be posted in buildings and electronically. We are eager to obtain, analyze, and present the data and results from the full posture study, given the promising exploratory results from these two test participants.

References

- Cohen, R. G., Rosenbaum, D. A., 2004. Where Grasps are Made Reveals how Grasps are Planned: Generation and Recall of Motor Plans. *Experimental Brain Research* 157 (4), 486–495.
- Engbert, R., Kliegl, R., 2003. Microsaccades Uncover the Orientation of Covert Attention. *Vision Research* 43 (9), 1035–1045.
- Fawcett, T., 2016. The Eyes Have It: Eye Tracking Data Visualizations of Viewing Patterns of Statistical Graphics. Master's thesis, Utah State University, Department of Mathematics & Statistics, Logan, Utah 84322.
- Herbort, O., Butz, M. V., 2010. Planning and Control of Hand Orientation in Grasping Movements. *Experimental Brain Research* 202 (4), 867–878.
- Li, C., Symanzik, J., 2016. The Linked Microposter Plot as a New Means for the Visualization of Eye Tracking Data. In: 2016 JSM Proceedings. American Statistical Association, Alexandria, VA, (CD).
- Li, C., Symanzik, J., 2017. EyeTrackR: An R Package for Extracting and Visualizing Data from Mobile and Static Eye Trackers. In: 2017 JSM Proceedings. American Statistical Association, Alexandria, VA, (CD).
- Michelon, P., Zacks, J. M., 2006. Two Kinds of Visual Perspective Taking. *Attention, Perception, & Psychophysics* 68 (2), 327–337.
- R Core Team, 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, <http://www.R-project.org/>.
- Rochat, P., 1995. Perceived Reachability for Self and for Others by 3- to 5-Year-Old Children and Adults. *Journal of Experimental Child Psychology* 59 (2), 317–333.
- Rosenbaum, D. A., Cohen, R. G., Meulenbroek, R. G. J., Vaughan, J., 2006. Plans for Grasping Objects. In: Latash, M. L., Lestienne, F. (Eds.), *Motor Control and Learning over the Lifespan*. Springer, New York, NY, pp. 9–25.
- Rosenbaum, D. A., Marchak, F., Barnes, H. J., Vaughan, J., Slotta, J. D., Jorgensen, M. J., 1990. Constraints for Action Selection: Overhand versus Underhand Grips. In: Jeannerod, M. (Ed.), *Attention and Performance 13: Motor Representation and Control*. Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 321–342.
- Salvucci, D. D., Goldberg, J. H., 2000. Identifying Fixations and Saccades in Eye-Tracking Protocols. In: ETRA '00 Proceedings of the 2000 Symposium on Eye Tracking Research & Applications, Palm Beach Gardens, Florida, USA. ACM, New York, NY., pp. 55–97.
- von der Malsburg, T., 2015. saccades: Detection of Fixations in Eye-Tracking Data. R package version 0.1-1 (<http://CRAN.R-project.org/package=saccades>).