10.4 EYE-TRACKING APPLICATIONS TO ASSESS IMPACTS OF PHASED ARRAY RADAR DATA ON FORECASTERS' COGNITIVE PROCESSES

*Katie Bowden, Pamela Heinselman, and Ziho Kang

1. INTRODUCTION

The research described in this paper was completed as part of the 2015 Phased Array Radar Innovative Sensing Experiment (PARISE; Heinselman et al. 2016). With Phased Array Radar (PAR) being considered as a future replacement technology to the current WSR-88D network (Zrnić et al. 2007), the goal of PARISE is to understand what impacts higher-temporal resolution radar data will have on National Weather Service (NWS) forecasters' warning decision processes.

To achieve this goal, we have assessed impacts from both a performance measurement (i.e., lead time and verification) perspective and in a qualitative and contextual sense (e.g., Heinselman et al. 2015, Bowden et al. 2015, and Bowden et al. 2016). The latter approach uses a retrospective recall method to obtain detailed information about forecasters' warning decision processes. Once forecasters have completed working a weather event in simulated real time, they are asked to watch a replay video of their onscreen activity while recalling their past thought process. While watching the video, forecasters recall what they were seeing, thinking, and doing at that time. The qualitative data collected using this method has proven to be insightful and essential to our understanding of why forecasters make warning decisions. However, capturing and quantifying the intricacies of forecasters' cognitive processes using retrospective recall methods is challenging. We have therefore begun exploring the use of eye-tracking methods to deepen our understanding of forecasters' warning decision processes and their related cognition.

Eye-tracking methods have been applied in a variety of research fields, including air traffic control (e.g., Kang et al. 2014), medicine (e.g., Wood et al. 2013), and more recently in meteorology (Drost et al. 2015 and Sherman-Morris et al. 2015). In these studies, eye-tracking has been used to understand how air traffic controllers track multiple aircraft on a radar screen, how medical professionals detect and diagnose skeletal fractures, and how the impact of gesturing during weather broadcasts or the choice of color in graphics impacts the effectiveness of communicated weather information. The types of questions that these studies have sought to answer suggest that eye-tracking methods have potential to help address the research goals of PARISE.

2. METHODOLOGY

2.1 Experimental Design

During six weeks in August and September 2015, 30 NWS forecasters from 25 Weather Forecast Offices located in the Great Plains participated in the 2015 PARISE. The eye-tracking experiment formed a small portion of the tasks completed by participants each week. A randomized two-independent group design was chosen, such that forecasters were randomly assigned to either a control group or an experimental group. Control group participants viewed 5-min volumetric PAR updates, while experimental group participants viewed 1-min volumetric PAR updates. The eye-tracking experiment was completed for one case.

2.2 Case

All participants worked the 2230–2330 UTC 8 July 2014 event (Fig. 1). The PAR 90° sector tracked slow-moving pulse thunderstorms that developed in central Oklahoma. While the eastern storm cell was not associated with any severe weather, the central storm cell was associated with both severe hail and wind reports from 2304 UTC until after the case had ended (https://verification/nws.noaa.gov).

2.3 Data Display

Participants viewed both base reflectivity and velocity data using the Waring Decision Support System–Integrated Information display software (Fig.

* *Corresponding author address*: Katie Bowden, OU School of Meteorology/ OU CIMMS, 120 David L. Boren Blvd., Norman, OK, 73072; email: katie.bowden@noaa.gov

2; Lakshmanan et al. 2007). When interrogating the radar data, forecasters could step back and forth in time, sample values, pan, and zoom in and out. In addition to having fixed areas dedicated to radar data, the screen had two fixed areas for control icons, and one fixed area for the warning generation tool (WarnGen).

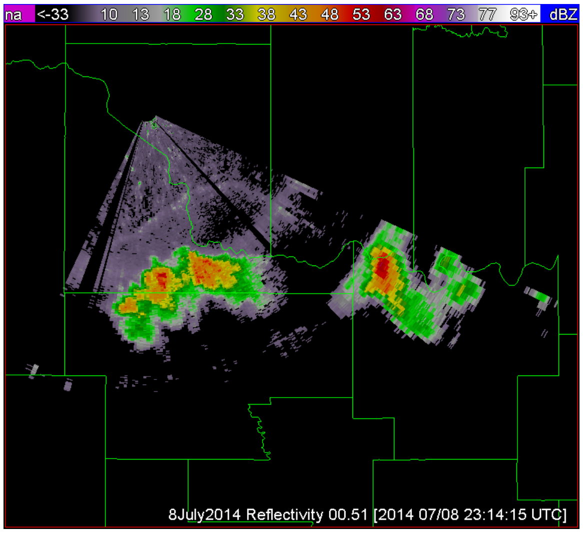

Figure 1. 0.51° reflectivity snapshot of pulse thunderstorms at 2314 UTC on 8 July 2014

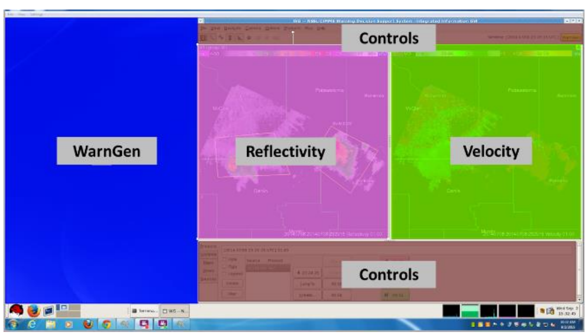

Figure 2. Participants screen display was set up into fixed areas for warning generation tools, reflectivity data, velocity data, and two control panels.

2.4 Procedure

The eye-tracking experiment took place in a room separate to the rest of the 2015 PARISE activities. Only the participant and assisting researcher were present during the experiment. Participants first watched a pre-briefing video that shared environmental data relevant to the event as well as satellite and radar images prior to case start time. Forecasters used information from the pre-briefing video to form expectations for how they thought the case would evolve.

Following the pre-briefing video, forecasters were asked to find a comfortable sitting position in front of the computer so that we could calibrate their eyes to the Tobii TX300 eye-tracker. During calibration, we completed a test run to ensure that the forecaster's eye gaze could be tracked successfully. Following the calibration and test run, the forecaster began working the weather event in simulated real time. Given the unobtrusive nature of the eye-tracking device (which was positioned underneath the computer monitor), forecasters were asked to maintain a steady position throughout the simulation. Forecasters were instructed to interrogate the radar data and make warnings decisions if and when they thought they were necessary. Storm reports were provided to the forecaster verbally, and forecasters were informed by the researcher when the simulation had come to an end. Following the simulation, forecasters watched a replay video of their onscreen activity and completed a retrospective recall. All verbalizations from the retrospective recall were recorded into a timeline and provide a detailed account of what forecasters were seeing, thinking, and doing, throughout the simulation.

3. FIXATION ANALYSIS

In the analysis of this eye-tracking data, each fixed area on the screen display is considered a separate area of interest (Fig. 2; AOI). The two control fixed areas are combined to represent a single Controls AOI in the analysis. Within each AOI, the forecaster's eye fixation behavior is of interest. Eye fixations define times when a person's gaze momentarily focuses on a specific location. Humans are able to process information during eye fixations, and therefore fixation behavior can give insight into one's cognitive processes (Liversedge and Findlay 2000).

Two common measures of eye fixations are count (i.e., the number of times a person fixates within an AOI for a given time) and duration (how long a fixation lasts). While higher numbers of fixation count indicate that information is either more noticeable or important, longer fixation durations indicate that information is either more engaging or requires greater mental effort to process (Poole and Ball 2006). The results discussed in this paper focus on these two fixation measures. Of the 15 participants in each group, 12 participants' eye gaze data were collected successfully and are used in this analysis.

3.1 Big Picture

From a big picture perspective, it was of interest whether differences in fixation behavior existed between the control and experimental groups. The first, most simple approach is to assess the distribution of fixation count and mean fixation duration of the control and experimental groups for the entire 1-hour simulation (Fig. 3). Although the distribution of total fixation count for the control group is wider than that of the experimental group, the median value of total fixation count differs only very slightly between the two groups (Fig. 3a). This slight difference is likely because the forecasters were actively looking at the screen for the same period. Both groups fixated more frequently in the Reflectivity AOI than the Velocity AOI, suggesting that both groups found the reflectivity data to be of overall more importance during this simulation. Both groups fixated least frequently in the Controls AOI, followed by the WarnGen AOI.

The overall distributions of mean fixation duration also did not differ between the two groups substantially. However, there is a slight tendency towards longer mean fixation durations in the experimental group for all AOIs, most notably in the Velocity AOI and the Controls AOI. However, no differences were found to be statistically significant.

3.2 Temporal Trend

It is possible that some differences in fixation measures when considered with respect to the entire simulation are masked. To assess group fixation measures on a shorter temporal scale, the group median 5-min fixation counts and 5-min mean fixation durations were computed (Fig. 4). In each 5-min period, the experimental group's median value for fixation count in the Reflectivity AOI was lower than that of the control group's. This difference is more noticeable in some 5-min periods (e.g., 5 and 10) than others (e.g., 2 and 9). In many instances the experimental group also fixated less frequently in the Velocity AOI than the control group. In the WarnGen AOI, the peaks in group median fixation counts were higher for the experimental group than the control group. For both groups, the fixation counts were very similar in the Controls AOI for all 5-min periods.

The experimental group's median value for mean fixation duration within the Velocity AOI was longer in every 5-min period than the control group, suggesting that as a whole, the experimental group devoted more mental effort to processing the velocity data than the control group. On several occasions the experimental group's median value for mean fixation duration also peaked higher in the Reflectivity AOI than the control group. On the whole, the experimental group trended towards longer mean fixation duration values in both the Controls and WarnGen AOIs.

3.3 Forecaster Variability

Forecaster fixation behavior should also be assessed on an individual level. Regardless of the temporal resolution that forecasters received, it is natural to expect that forecasters will interact with radar data differently. Understanding the variability between forecasters both within the same group and across different groups will be necessary for developing a deeper understanding of the reasons behind forecasters' different styles of interrogation. This deeper understanding will also aid in identifying specifically how the temporal resolution of radar data impacted forecaster behavior and cognition.

Figure 5 provides examples of four forecasters' fixation count and mean fixation duration for each 10-min period during the 1-hour simulation. These measures are provided for each of the AOIs. Both differences and similarities in trends of the forecasters' fixation behavior exist. For example, control participants C6 and C15 behavior similarly in terms of fixation count trends within the Reflectivity AOI. However, within the Velocity AOI, C15 fixates much less frequently. C15 attended to Reflectivity AOI more intently than C6, as also supported by his longer mean fixation duration in the Reflectivity AOI. Of the four forecasters' fixation behavior, E6 fixated most frequently within the Reflectivity AOI (Fig. 5). Compared to E9, E6 fixated less often in the Velocity AOI, but overall E6's and E9's mean fixation durations were comparable (Fig. 5). Trends in fixation count within the WarnGen AOI show that C6 and C15 interacted with the warning tools earlier than E6 and E9 (Fig. 5), suggesting that E6's and E9's first warning decisions came later than C6's and C15's.

5. NEXT STEPS

The results of this research are preliminary. We have explored group differences in fixation measures for the full 1-hour scenario, and explored possible differences in fixation measures between groups when examined on finer temporal scales as well as on an individual basis. Our preliminary results indicate that an examination of differences in fixation measures on even smaller temporal scales (i.e., 1 min) will be necessary for identifying specifically when in the scenario the temporal resolution of radar data impacted forecasters' warning decision processes. Additionally, the preliminary results support that individual differences both within groups and between groups exist. We wish to better capture what those differences are, and understand what drove those differences in forecaster cognitive behavior.

Future work will also analyze forecaster fixation behavior for specific phases of the warning decision process. For example, we are interested in analyzing forecasters' fixations immediately (e.g., 1 min) prior to and after the issuance of a warning product. Using eye gaze data to understand how forecasters use their attention during this important stage of the warning decision will be helpful for determining what impact the temporal resolution of radar data had at those times.

Finally, we will be using additional sources of data to enhance our analysis and interpretation of each forecaster's warning decision process. During the retrospective recall portion of this experiment, forecasters provided cognitive workload ratings for every 5-min period in the simulation. We will analyze trends in forecaster's cognitive workload and assess how they correspond to their fixation behavior. The retrospective recall data will also provide contextual support for each forecaster's eye gaze data, and will ensure a more complete interpretation of their warning decision processes.

Acknowledgments. *We thank the 30 NWS forecasters and their respective Weather Forecast Offices for their participation in the 2015 PARISE. Additionally, we thank Chris Carter and Robert Coggins for their ongoing technical support during the eye-tracking experiment, Greg Schoor for participating in the eye-tracking pilot, and Gabe Garfield for preparing the 2015 PARISE pre-briefing videos.*

REFERENCES

Bowden, K. A., P. L. Heinselman, D. M. Kingfield, and R. P. Thomas, 2015: Impacts of phased array radar data on forecaster performance during severe hail and wind events. *Wea. Forecasting*, **30**, 389–404.

Bowden, K. A., and P. L. Heinselman, 2016: A qualitative analysis of NWS forecasters' use of phased array radar data during severe hail and wind events. *Wea. Forecasting*, *in press*.

Drost, R., J. Trobec, C. Steffke, and J. Libarkin, 2015: Eye tracking: Evaluating the impact of gesturing during televised weather forecasts. *Bull. Amer. Meteor. Soc.*, **96**, 387–392.

Heinselman, P. L., D. LaDue, D. M. Kingfield, and R. Hoffman, 2015: Tornado warning decisions using phased array radar data. *Wea. Forecasting*, **30**, 57–78.

Heinselman, P. L., K. A. Bowden, D. M. Kingfield, and C. M. Kuster, 2016: 2015 Phased array radar innovative sensing experiment. Preprints, *32nd Conf. on Environmental Information Processing Technologies*, New Orleans, LA, Amer. Meteor, Soc., 12A.6.

Kang, Z., and S. J. Landry, 2014: Using scanpaths as a learning method for a conflict detection task of multiple target tracking. *Human Factors*, **56**, 1150–1162.

Lakshmanan, V., T. Smith, G. Stumpf, and K. Hondl, 2007: The warning decision support system–integrated information. *Wea. Forecasting*, **22**, 596–612.

Liversedge, S. P., and J. M Findlay, 2000: Saccadic eye movements and cognition. *Trends in Cognitive Sciences*, **4**, 6–14.

Poole, A., and L. J. Ball, 2006: Eye tracking in HCI and usability research: Current Status and Future Prospects. Encyclopedia of Human-Computer Interaction, C. Ghaoui, Ed., Idea Group Inc., 211–219.

Sherman-Morris, K., K. B. Antonelli, and C. C. Williams, 2015: Measuring the effectiveness of the graphical communication of hurricane storm surge threat. *Wea. Climate Soc.*, **7**, 69–82.

Wood, G., K. M. Knapp, B. Rock, C. Cousens, C. Roobottom, and M. R. Wilson, 2013: Visual expertise in detecting and diagnosing skeletal fractures. *Skeletal Radiology*, **42**, 165–172.

Zrnić, D. S., and Coauthors, 2007: Agile beam phased array radar for weather observations. *Bull. Amer. Meteor. Soc.,* **88**, 1753–1766.

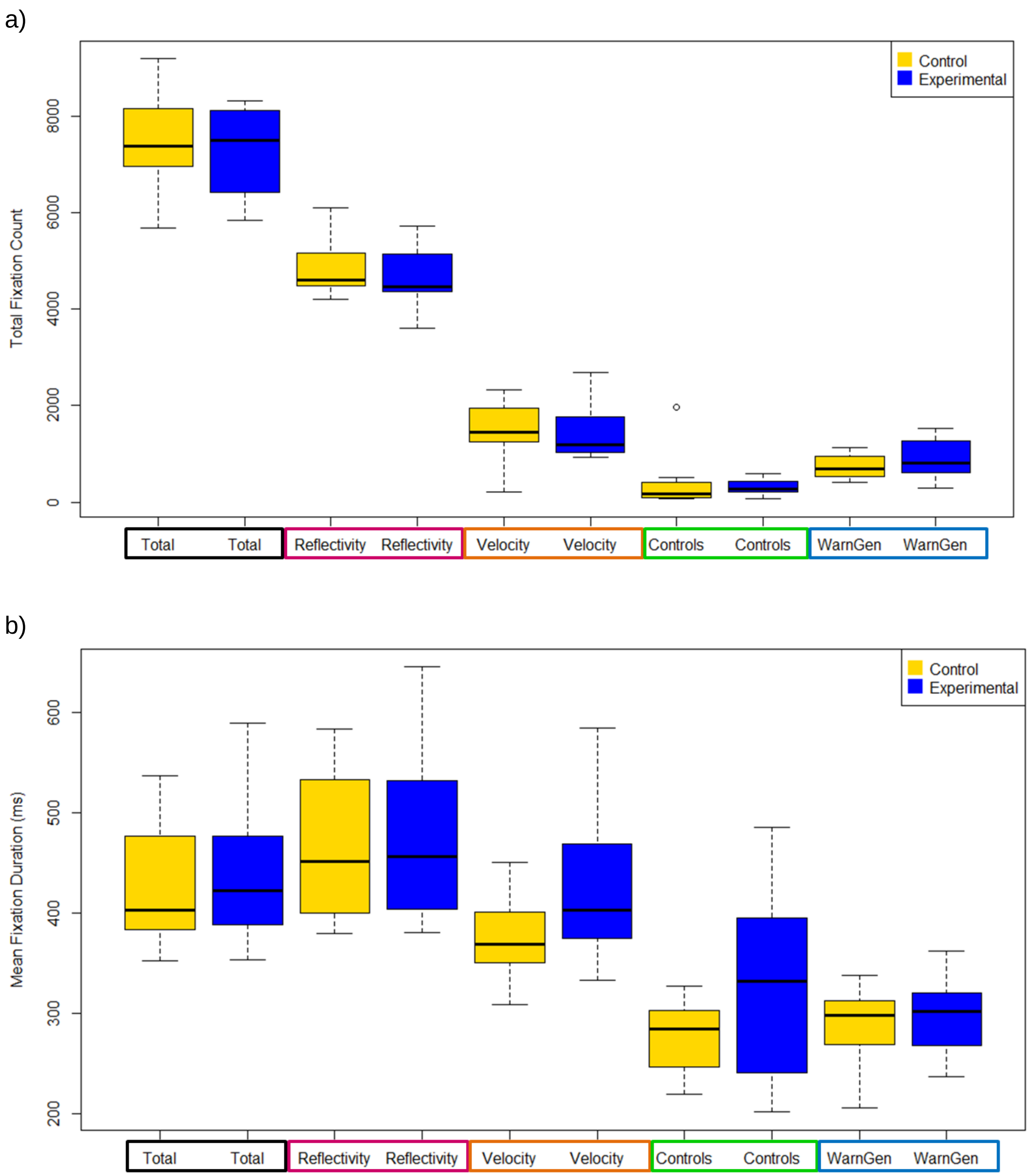

Figure 3. Distributions of (a) fixation count and b) mean fixation duration for the control and experimental groups for the 1-hour simulation.

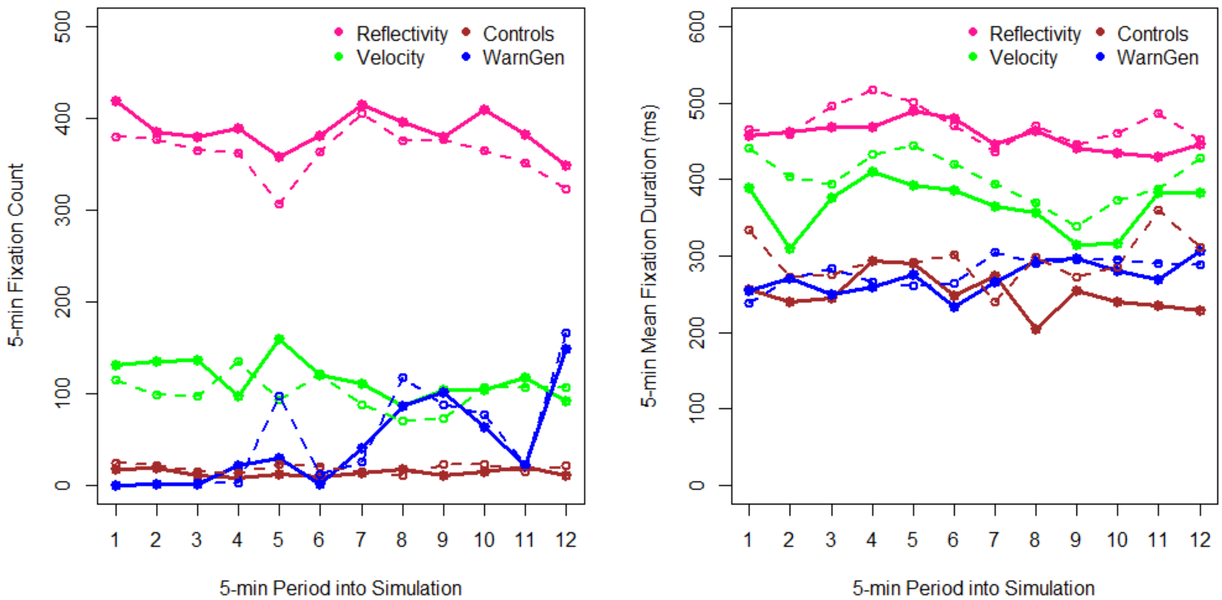

Figure 4. Control group (solid) and experimental group (dashed) a) 5-min fixation count and b) 5-min mean fixation duration for each area of interest for the 1-hour simulation.

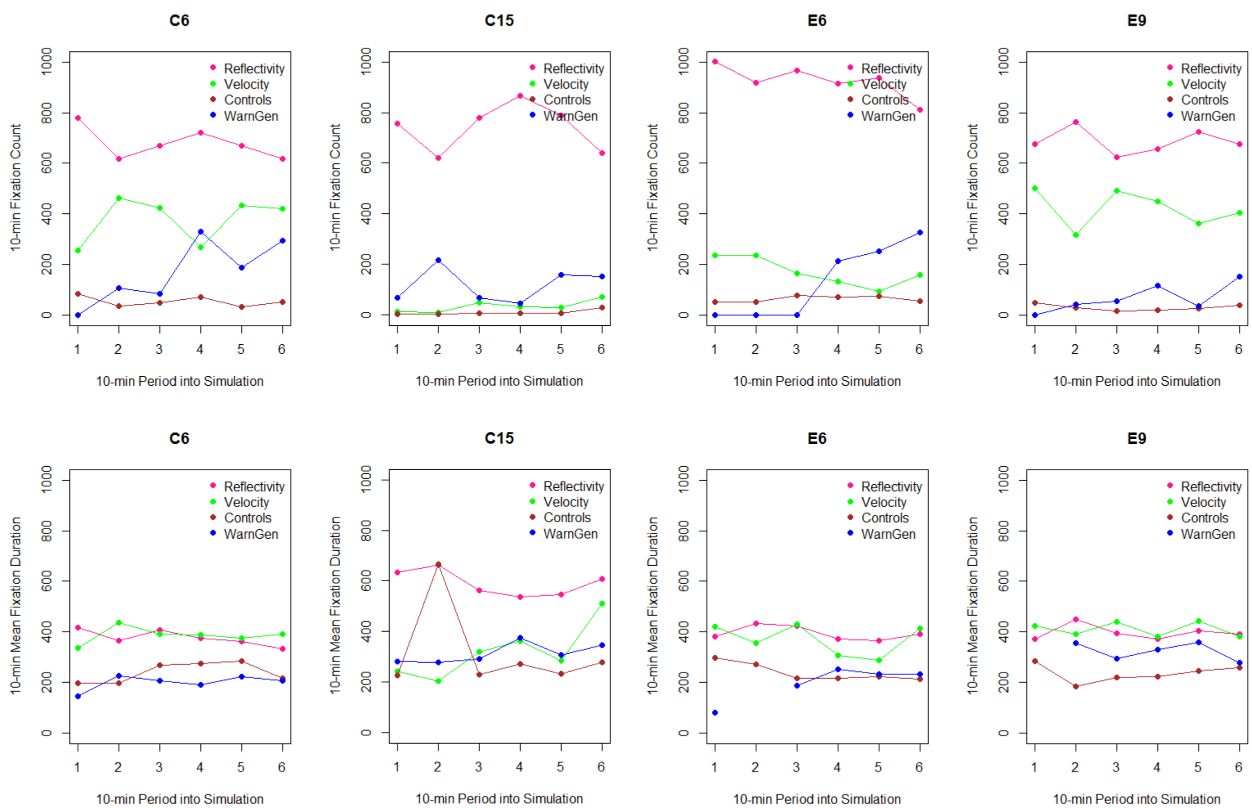

Figure 5. 10-min fixation count (top row) and 10-min mean fixation duration (bottom row) for each AOI for control participants C6 and C15 and experimental participants E6 and E9.