AMIA
INFORMATICS PROFESSIONALS. LEADING THE WAY.

OXFORD

Brief Communications

Evaluation of eye tracking for a decision support application

Shyam Visweswaran [iD],[1,2] Andrew J. King [iD],[1,3] Mohammadamin Tajgardoon [iD],[2] Luca Calzoni,[1] Gilles Clermont,[3] Harry Hochheiser [iD],[1,2] and Gregory F. Cooper [iD][1,2]

[1]Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, Pennsylvania, USA, [2]Intelligent Systems Program, University of Pittsburgh, Pittsburgh, Pennsylvania, USA and [3]Department of Critical Care Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

Corresponding Author: Shyam Visweswaran, Department of Biomedical Informatics, University of Pittsburgh, The Offices at Baum, 5607 Baum Blvd., Suite 523, Pittsburgh, PA 15206, USA (shv3@pitt.edu)

ABSTRACT

Eye tracking is used widely to investigate attention and cognitive processes while performing tasks in electronic medical record (EMR) systems. We explored a novel application of eye tracking to collect training data for a machine learning-based clinical decision support tool that predicts which patient data are likely to be relevant for a clinical task. Specifically, we investigated in a laboratory setting the accuracy of eye tracking compared to manual annotation for inferring which patient data in the EMR are judged to be relevant by physicians. We evaluated several methods for processing gaze points that were recorded using a low-cost eye-tracking device. Our results show that eye tracking achieves accuracy and precision of 69% and 53%, respectively compared to manual annotation and are promising for machine learning. The methods for processing gaze points and scripts that we developed offer a first step in developing novel uses for eye tracking for clinical decision support.

Key words: eye tracking, electronic medical record system, relevant patient data

LAY SUMMARY

In the context of electronic medical record systems, eye tracking is used extensively to explore attention and cognitive processes. We investigated a novel application of eye tracking to collect training data for machine learning-based clinical decision support. We evaluated several methods for processing gaze points that were recorded using a low-cost eye-tracking device. The methods for processing gaze points and scripts that we developed offer a first step in developing novel uses for eye tracking for clinical decision support.

INTRODUCTION

There is growing interest in leveraging eye tracking to support advanced types of clinical decision support (CDS) tools in electronic medical record (EMR) systems.[1,2] For example, eye tracking could capture which data (eg, vital signs, laboratory test results, medication orders, etc.) in a patient's EMR a physician has viewed in the context of a clinical task.[3] If eye-tracking devices were deployed on EMR computer monitors, then the data that are viewed by many

physicians could be collected, and machine learning models derived from such data can predict which data are likely to be relevant in a given patient. Such predictive models can form the basis of a CDS tool to highlight relevant patient data and draw the physician's attention to them.[4] Further, such a tool has the potential to mitigate the cognitive overload arising from the large amounts of patient data that physicians have to collate and assess in data-rich settings like the intensive care unit (ICU).

We developed in learning EMR (LEMR) system to investigate the feasibility of a CDS tool to predict and highlight relevant data.[5] The LEMR system relies on supervised machine learning models that predict which patient data are likely to be relevant in the context of a clinical task.[6,7] However, a critical barrier to building machine learning models for a LEMR system is the acquisition of training data regarding which patient data are relevant for a clinical task. Such data are not recorded in sufficient granularity in currently deployed EMR systems. Hence, we collected training data in a laboratory setting where physicians reviewed patient cases and provided manual annotations about which data were relevant.[4] However, manual annotation is onerous, expensive, and time-consuming and limits the amount of training data that can be collected. Eye tracking offers an alternative method for capturing training data, and we investigated whether eye-tracking data are as accurate as manual annotations.[2,3] If so, eye tracking can provide a promising, higher-throughput alternative that if deployed on EMR systems will unobtrusively capture which data are viewed by thousands of physicians and provide large volumes of data for machine learning. To our knowledge, we are the first to investigate the use of eye-tracking data to develop machine learning models for CDS.[1–3]

BACKGROUND

In this section, we provide brief descriptions of eye-tracking technology and its application to EMR systems, the LEMR system, and the eye-tracking device used in the LEMR system.

Eye tracking and its applications to EMR systems

Eye tracking is a method to track and record eye movements and gaze points across time and task, and it is commonly used for observing and measuring the allocation of visual attention.[8] Eye-tracking devices record a sequence of gaze points with a regular sampling rate. Gaze point data form the basis of a variety of analyses of visual attention, such as the characterization of fixation and saccadic eye movements in an area of interest (AOI)[9] and dwell time, which is the total amount of time spent looking within an AOI.[10] Eye tracking is widely used to study cognitive processes that underlie a variety of tasks such as visual search, comprehension, and judgment and decision-making.[11]

A range of methods have been developed to measure visual attention and dwell time in AOIs from gaze point data. Simple methods calculate the dwell time or the time spent looking at an AOI by summing the time that the gaze points were within the AOI.[12] More sophisticated methods identify fixations within an AOI based on the assumption that visual attention occurs only during fixations and calculate the dwell time by summing the time that the fixations were within the AOI.[13]

In the context of EMR systems, eye-tracking research has focused on understanding users and their interactions with the systems. For example, investigators have used eye tracking to understand clinical reasoning,[14] track information search patterns,[15]

evaluate usability,[16] measure time use,[17] and investigate visual and cognitive processes while performing tasks.[18] Recent reviews of the literature have described the application of eye tracking in clinical decision-making[19] and usability of EMR systems.[1] However, little work has been done to investigate the use of eye tracking in EMR systems to enable CDS tools. In previous work, we described the application of eye tracking to CDS for the collection of training data for deriving machine learning models of relevant patient data.[2,3]

The LEMR system

We developed the LEMR system to highlight relevant patient data by using machine learning models to identify such data in the context of a clinical task for a specific patient.[4,6] The LEMR system can also be used to collect training data for machine learning. The LEMR interface, shown in Figure 1, enables the collection of data that is judged to be relevant for a clinical task by a physician in two ways: (1) manual annotation, when the physician annotates relevant data by clicking on checkboxes (see Figure 3) and (2) eye gaze, when an eye tracking device records gaze points while the physician is reviewing the patient's record (see Figure 2). Our goal in developing the LEMR system was not to replicate an entire modern EMR but to have a useful prototype for displaying patient data and for capturing gaze point data and annotations to support studies in a laboratory setting.

Eye-tracking device

The LEMR system is outfitted with Tobii EyeX, an inexpensive portable eye-tracking device and software package, which is primarily marketed for developing computer gaming and virtual reality applications.[20] The hardware component, the Tobii EyeX Controller, is mounted at the bottom edge of a computer monitor (see Figure 1) and samples eye-gaze point coordinates at approximately 60 Hz. The software component, the Tobii EyeX Software Development Kit, records and outputs x-axis and y-axis gaze point coordinates for each eye.

METHODS

In this section, we describe the patient data and physician reviewers, the clinical task, data collection, the methods that we applied to process the gaze point data, and the evaluation measures we used.

Patient data and physician reviewers

We randomly selected 178 patients who were admitted to an ICU between June 2010 and May 2012 at the University of Pittsburgh Medical Center and had a diagnosis of either acute kidney failure (ICD-9 584.9 or 584.5; 93 patients) or acute respiratory failure (ICD-9 518.81; 85 patients). Eleven critical care physicians, including fellows and attending physicians, reviewed the EMRs of the selected patients in the LEMR system with eye tracking in a laboratory setting.

Clinical task

We chose the clinical task of identifying relevant patient data that have accumulated during the past day to present a summary of the patient's clinical status at morning rounds in the ICU. This task is performed daily and is typically time-consuming, with the physician painstakingly searching the EMR to identify and retrieve relevant data. Each patient record was loaded into the LEMR system as

Figure 1. A computer monitor displaying the LEMR interface as it appears during the familiarization and preparation tasks (see Methods section). From left to right, the system displays patient data on vital signs, ventilator settings, intake and output, medication administrations, laboratory test results, and free-text notes and reports. The eye-tracking device mounted at the bottom is used to capture gaze points during the preparation task (see Methods section).
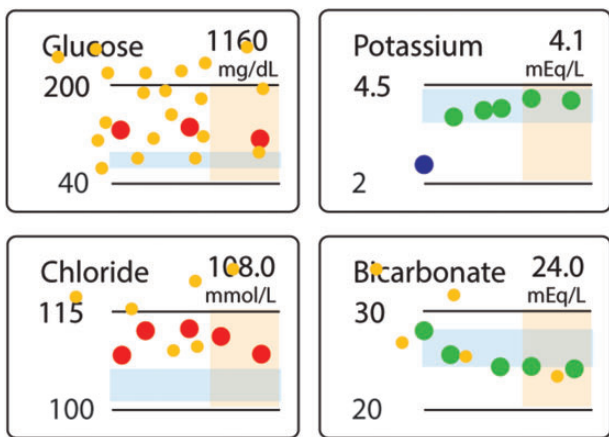

Figure 2. A portion of the LEMR interface as it appears during the preparation task (see Methods section) showing four laboratory test results. The horizontal light blue band indicates the normal range for the corresponding laboratory test and the vertical light orange band indicates the most recent 24-h period. The larger green circles, red circles, and purple circles denote normal, high, and low values of the corresponding laboratory values. The smaller orange circles denote the location of gaze points recorded by the eye-tracking device; these are shown for illustrative purposes only and are not visible on the interface.
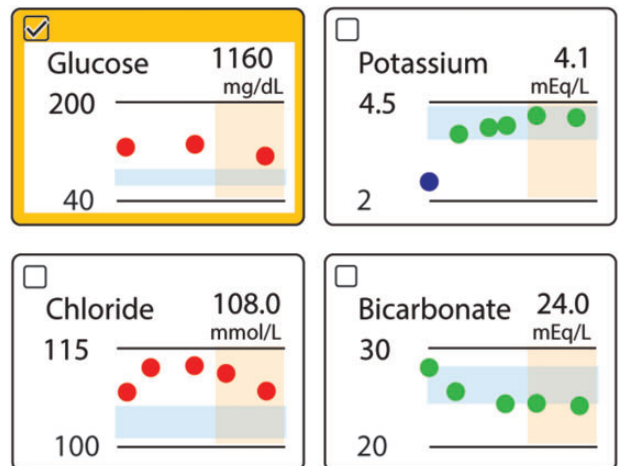

Figure 3. A portion of the LEMR interface as it appears during the annotation task (see Methods section) showing four laboratory test results with checkboxes. Physicians indicate which patient data are relevant by clicking on the corresponding checkboxes. The glucose laboratory test is surrounded by a yellow margin to indicate that its checkbox has been clicked.

shown in Figure 1. A physician reviewed a record by completing three tasks sequentially. In the *familiarization task*, the physician was shown patient data from the time of ICU admission up to 8:00 AM on a random ICU day between day 2 and the day before discharge from the ICU (inclusive). The physician was asked to review the data, become familiar with the patient, and understand the clinical course. After becoming familiar with the patient, the physician switched to the *preparation task* and was shown an additional 24 h of patient data with instructions to review the new data for the task of summarizing the clinical status for presentation at morning rounds. During this task, eye tracking was used to record the physician's gaze points. After reviewing the new data, the physician turned to the *annotation task* and indicated which patient data were relevant by clicking on checkboxes.

Data collection

From gaze points recorded during the preparation task, we estimated which patient data were considered to be relevant by measuring the dwell times within AOIs; from the manual checkbox annotations recorded during the annotation task, we derived a reference standard of patient data that was specified as relevant by the same physician. We compared the performance of the gaze-derived relevant patient data with the manual annotation reference standard.

Methods for processing gaze data

We evaluated five methods for processing gaze points to infer relevant patient data. We focused only on vital signs, ventilator settings, intake and output measurements, laboratory test results, and medication administration, and excluded free-text notes and reports (see Figure 1). These data are presented in rectangular areas in the LEMR interface and are the AOIs for which we applied five methods to calculate dwell times. We assumed that AOIs with longer dwell times indicate relevant patient data. We selected three fixation identification algorithms that include the velocity-threshold identification (I-VT), the dispersion-threshold identification (I-DT), and area-of-interest identification (I-AOI) methods. In addition, we developed two simple gaze point algorithms called the gaze point (GP) and distributed gaze point (DGP) methods. The fixation algorithms measure the dwell times of fixations while the gaze point algorithms measure the dwell times of all gaze points within a ROI. We provide brief descriptions of the five methods next.

The I-VT method is a velocity-based method that computes a velocity for each gaze point based on the distance between the gaze point and the previous gaze point. Based on a velocity threshold, a point is classified as a fixation point if its velocity is below the threshold or as a saccade point if its velocity is above the threshold. I-VT identifies a fixation as a collection of consecutive fixation points and discards the saccade points. Thus, I-VTI requires a velocity threshold as an input parameter. We fixed the velocity threshold to 20 degrees per second.

The I-DT method identifies tightly clustered groups of gaze points as fixations. More specifically, it identifies a fixation as a collection of consecutive gaze points such that the points are within a maximum distance of one another (called dispersion) and within a period of time exceeding some minimum duration (generally 100 milliseconds).[9] Thus, I-DT requires two input parameters, the dispersion threshold and the duration threshold. Therefore, in our experiments, we explored a range of values for the input parameters. The dispersion threshold was selected from values [50, 80, 100,

150, 200 pixels] and the duration threshold was selected from values [10, 20, 30, 40 data points]. Since the sampling frequency is 60 Hz, the interval between consecutive data points is 16.7 ms; thus 10 data points for the duration threshold translates to a duration of 167 milliseconds.

The I-AOI method identifies fixations in a fashion similar to the I-DT; however, it identifies fixations that occur within one or more AOIs. I-AOI utilizes a duration threshold to distinguish fixations in the AOI from saccades in that area.[9] Thus, I-AOI requires a duration threshold as an input parameter. We selected the duration threshold from values [10, 20, 30, 40, 50, 100, 150, 200 data points].

We developed two simple and computationally efficient gaze point methods that do not rely on fixation identification. The GP method maps all gaze points to AOIs without classifying the points as part of a fixation or a saccade. A higher proportion of the total recorded gaze points that map to an AOI results in a longer dwell time and indicates that more visual attention has been directed there (see Supplementary Appendix for pseudocode).

The DGP method is a probabilistic refinement of the GP method, in which each gaze point contributes to adjacent AOIs in a probabilistic fashion. The fractional contribution of a gaze point to an AOI is equal to the density of a bivariate normal distribution (see Supplementary Appendix for pseudocode). The means of the distribution are located at the center of the gaze point and the variances are derived from the average error of the eye-tracking device in the horizontal and vertical directions that we estimated in a prior study.[3]

Performance measures

We evaluated the performance of the methods with accuracy, precision, and recall. In the context of a gaze-point processing method, a true positive (TP) is an AOI that was identified as relevant by both gaze-point processing and manual annotation, a false positive (FP) is an AOI that was identified as relevant by gaze-point processing but not by manual annotation, a true negative (TN) is an AOI that was identified as irrelevant by both gaze-point processing and manual annotation, and a false negative (FN) is an AOI that was identified as irrelevant by gaze-point processing but as relevant by manual annotation. Accuracy is obtained by dividing the sum of true positives and true negatives by the total number of AOIs (accuracy = (TP + TN)/(TP + FP + FN + TN)). Precision is obtained by dividing the number of true positives by the sum of true positives and false positives (precision = TP/(TP + FP)), and recall is obtained by dividing the number of true positives by the sum of true positives and false negatives (recall = TP/(TP + FN)). We used the binomial test of proportion to perform pairwise statistical comparisons of the methods for each of the performance measures.

RESULTS

The reviewers were physicians trained in critical care medicine and included fellows and attending physicians. Characteristics of the reviewers are summarized in Table 1.

The performance of the five methods for processing gaze data is shown in Table 2. Overall, I-VT and GP had the highest accuracy at 69%, GP had the highest precision at 53% and DGP had the highest recall at 48%. I-VT had slightly lower precision than GP. For each performance measure, the highest value or values are statistically significantly higher than the remaining values on the binomial test of proportion.

Table 1. Characteristics of physician reviewers

Number of physicians	Average number of years spent in ICU	Average number of weeks per year spent rounding in ICU
11	1.8 (0.3–7.0)	34 (26–42)

Table 2. Accuracy, precision, and recall values with standard error of five methods for processing gaze data

Method	Accuracy (%)	Precision (%)	Recall (%)
I-VT	**69 ± 0.03**	52 ± 0.08	33 ± 0.24
I-DT	67 ± 0.03	46 ± 0.05	26 ± 0.23
I-AOI	68 ± 0.03	49 ± 0.09	31 ± 0.26
GP	**69 ± 0.04**	**53 ± 0.10**	38 ± 0.28
DGP	67 ± 0.05	50 ± 0.08	**48 ± 0.25**

The highest values for each performance measure are in bold font.

DISCUSSION

We evaluated five methods for processing gaze data obtained to infer what patient data physicians judged to be relevant for summarizing the patient's clinical status at morning rounds in the ICU. Compared to manual annotation, the results support the use of eye tracking and relatively simple methods for processing gaze data to infer data relevance with modest accuracy and precision. We derived machine learning models using gaze data to predict relevant patient data in the LEMR system and found that they performed as well as models that were derived using manual annotation. The detailed results are reported in a separate publication and support further development of the use of eye-tracking data to drive a CDS tool to identify and highlight relevant patient data in EMR systems.[2] The scripts for the eye-tracking methods with accompanying documentation are freely available on GitHub at https://github.com/ajk77/EyeBrowserPy (last accessed July 16, 2021).

The eye-tracking device, Tobii EyeX, was developed as an inexpensive device for gaming applications. It is simple to install on a computer monitor, unobtrusive, and easy to calibrate. The Software Development Kit provides API bindings for several programming languages including Python that are straightforward to use for programming.

There are several limitations to our study. A key limitation is that the methods we used infer visual attention or seeing rather than cognition, and seeing does not imply that the information was cognitively processed. A second limitation is that the eye-tracking device, Tobii EyeX, was not developed explicitly for research applications and the device has modest temporal and spatial resolution and sampling frequency. However, it was adequate for our application that only required monitoring of simple eye movements. Further, the device could not track head movements, and the physician reviewers in our study had to restrain their head movements. However, this limitation may be mitigated with newer devices such as Tobii Eye Tracker 5 that are capable of tracking both head and eye movements and offer the ability to robustly estimate the coordinates of eye-gaze even if the head position changes.[21] A third limitation is that the interface of the LEMR system is significantly different from the vendor EMR systems currently used in clinical care, and furthermore may not be optimal for the review of patient data. Further studies are needed to assess and improve the LEMR interface. A fourth limitation is that given our results of modest accuracy and precision of eye tracking, the performance of the models derived from such data may be imperfect and unreliable to such an extent that it leads to poorer performance and trust in the LEMR system.[22,23] One approach to mitigating this limitation that we plan to investigate in future studies is to examine whether the performance of the models can be improved with a combination of smaller amounts of more accurate, manually obtained data with larger amounts of less accurate eye-tracking data over either type of data alone. A final limitation is that we did not comprehensively investigate a wider range of methods that are available for processing gaze data; we plan to do so in the future.

CONCLUSION

Eye tracking provides an automated and unobtrusive method to capture which patient data physicians judge to be relevant for a specific clinical task. Gaze point data recorded with an inexpensive eye-tracking device have modest accuracy and precision in inferring relevant data and are promising for deriving machine learning models that identify and highlight relevant patient data. The methods for processing gaze points and scripts that we developed offer a first step in developing novel uses for eye tracking for CDS. In addition, to predicting which patient data in EMR systems are relevant, eye tracking may be useful for additional CDS tools that depend on capturing which patient data were viewed by physicians.

SUPPLEMENTARY MATERIAL

Supplementary material is available at *Journal of the American Medical Informatics Association* online.

CONTRIBUTORS

S.V. conceived and designed the study, participated in data analysis and interpretation, drafted the manuscript and approved the final version for submission. A.J.K. conceived and designed the study, participated in data collection, analysis and interpretation, drafted the manuscript, and approved the final version for submission. M.T. participated in data analysis and interpretation, drafted the manuscript and approved the final version for submission. L.C. made critical revisions to the manuscript and approved the final version for submission. G.C. participated in data collection, made critical revisions to the manuscript and approved the final version for submission. H.H. participated in data analysis and interpretation, made critical revisions to the manuscript and approved the final version for submission. G.F.C. conceived and designed the study, participated in data analysis and interpretation, made critical revisions to the manuscript, and approved the final version for submission.

ACKNOWLEDGMENTS

FUNDING

CONFLICT OF INTEREST STATEMENT

None declared.

DATA AVAILABILITY

The data underlying this article cannot be shared publicly due to the privacy of individuals that participated in the study as well as the privacy of the patients whose data were used in the study. The data will be shared on reasonable request to the corresponding author.

REFERENCES

1. Senathirajah Y Borycki EM Kushniruk A Cato K Wang J. Use of eye-tracking in studies of EHR usability-the current state: a scoping review. *MedInfo* 2019; 1976–7.
2. King AJ Cooper GF Clermont G, *et al*. Leveraging eye tracking to prioritize relevant medical record data: comparative machine learning study. *J Med Internet Res* 2020; 22 (4): e15876.
3. King AJ Hochheiser H Visweswaran S Clermont G Cooper GF. Eye-tracking for clinical decision support: a method to capture automatically what physicians are viewing in the EMR. *AMIA Summits Transl Sci Proc* 2017; 2017: 512.
4. King AJ Cooper GF Clermont G, *et al*. Using machine learning to selectively highlight patient information. *J Biomed Inform* 2019; 100: 103327.
5. King AJ Cooper GF Hochheiser H Clermont G Visweswaran S. Development and preliminary evaluation of a prototype of a learning electronic medical record system. In: AMIA Annual Symposium Proceedings. 2015; American Medical Informatics Association: 1967.
6. King AJ Cooper GF Hochheiser H Clermont G Hauskrecht M Visweswaran S. Using machine learning to predict the information seeking behavior of clinicians using an electronic medical record system. In: AMIA Annual Symposium Proceedings. 2018; American Medical Informatics Association: 673.
7. Tajgardoon M, Cooper GF, Clermont G, *et al*. Modeling physician variability to prioritize relevant medical record information. *JAMIA Open* 2020; 3 (4): 602–610.
8. Carter BT Luke SG. Best practices in eye tracking research. *Int J Psychophysiol* 2020; 155: 49–62.
9. Salvucci DD Goldberg JH. editors. Identifying fixations and saccades in eye-tracking protocols. In: Proceedings of the 2000 Symposium on Eye Tracking Research & Applications; 2000.
10. Albert W, Tullis T. *Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics*. 2nd ed. Waltham, MA: Morgan Kaufmann; 2013.
11. Beesley T, Pearson D, Le Pelley M. Eye tracking as a tool for examining cognitive processes. In: Foster G, ed. *Biophysical Measurement in Experimental Social Science Research*. 1st ed. Cambridge, MA: Academic Press; 2019: 1–30.
12. Vansteenkiste P Cardon G Philippaerts R Lenoir M. Measuring dwell time percentage from head-mounted eye-tracking data–comparison of a frame-by-frame and a fixation-by-fixation analysis. *Ergonomics* 2015; 58 (5): 712–21.
13. Andersson R Larsson L Holmqvist K Stridh M Nyström M. One algorithm to rule them all? An evaluation and discussion of ten eye movement event-detection algorithms. *Behav Res Methods* 2017; 49 (2): 616–37.
14. Blondon K, Wipfli R, Lovis C. Use of eye-tracking technology in clinical reasoning: A systematic review. *Stud Health Technol Inform*. 2015; 210: 90–94.
15. Doberne JW He Z Mohan V Gold JA Marquard J Chiang MF. editors. Using high-fidelity simulation and eye tracking to characterize EHR workflow patterns among hospital physicians. In: AMIA Annual Symposium Proceedings, 2015; American Medical Informatics Association.
16. Asan O Yang Y. Using eye trackers for usability evaluation of health information technology: a systematic literature review. *JMIR Hum Factors* 2015; 2 (1): e5.
17. Rick S Calvitti A Agha Z Weibel N. editors. Eyes on the clinic: accelerating meaningful interface analysis through unobtrusive eye tracking. In: 2015 9th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2015; IEEE.
18. Calvitti A Weibel N Liu L, *et al*. Can eye tracking and electronic health record mouse activity tell us when clinicians are overloaded. *Hum Factors Quart Hum Factors Q* 2014; 8: 1.
19. Al-Moteri MO Symmons M Plummer V Cooper S. Eye tracking to investigate cue processing in medical decision-making: a scoping review. *Comput Hum Behav* 2017; 66: 52–66.
20. Tobii. Specifications for EyeX. 2014. https://help.tobii.com/hc/en-us/articles/212818309-Specifications-for-EyeX Accessed July 16, 2021.
21. Tobii. Tobii Eye Tracker 5. 2016. https://gaming.tobii.com/product/eye-tracker-5/ Accessed July 16, 2021.
22. Parasuraman R Wickens CD. Humans: still vital after all these years of automation. *Hum Factors* 2008; 50 (3): 511–20.
23. Rovira E McGarry K Parasuraman R. Effects of imperfect automation on decision making in a simulated command and control task. *Hum Factors* 2007; 49 (1): 76–87.