

A WEB-BASED EVALUATION TOOL TO PREDICT LONG EYE GLANCES

Ja Young Lee

John D. Lee

University of Wisconsin-Madison

Madison, WI, USA

Email: jayoung.lee@wisc.edu, jdlee@engr.wisc.edu

Summary: We present a web-based evaluation tool that simulates drivers' eye glances to interface designs of in-vehicle information systems (IVISs). This tool computes saliency of each location of a candidate interface and simulates eye fixations based on the saliency, until it arrives at the region of interest. Designers can use this tool to estimate the duration of drivers' eye glance needed to find regions of interest, such as particular icons on a touch screen. The overall goal of developing this application is to bridge the gap between guidelines and empirical evaluations. This evaluation tool acts as an interactive model-based design guideline to help designers craft less distracting IVIS interfaces.

OBJECTIVE

Digital displays are now widely embedded into cars to provide drivers with information and entertainment. For instance, Tesla Motors has installed a 17" touch display on the center stack to serve as a master control interface for the car. Both Apple CarPlay and Android Auto have been developed to provide seamless connection between smartphones and vehicles. Such systems offer drivers information in a form that might be less distracting than accessing that information on a smartphone. However, interacting with such In-Vehicle Information Systems (IVISs) might shift drivers' attention away from the roadway and increase crash risk.

Naturalistic driving data has shown that instances where drivers' eyes were off the forward roadway for more than 2.0 seconds accounted for 18% of crashes and near crashes, producing an odds ratio of 2.19 (Klauer, Dingus, Neale, Sudweeks, & Ramsey, 2006). Similarly, NHTSA's driver distraction guidelines (2014) suggest that mean duration of glance away from the road should not exceed 2.0 seconds. Long eye glances can occur when drivers search for an object of interest on a cluttered display (Tsimhoni & Green, 2001). Saliency, or conspicuity, of objects is one factor that guides visual attention and influences the time to locate objects (Wickens, Goh, Helleberg, Horrey, & Talleur, 2003). If the object of interest is salient then search times can be short. However, if the object of interest is not salient and is surrounded by highly salient then search times can be long. Long search times associated with such misplaced salience can lead to long off-road glances and heightened crash risk.

Empirical evaluation using eye tracking is an expensive and time-consuming method to assess the distraction potential of candidate IVIS interfaces. Recruiting participants, collecting data, and analyzing the eye tracking data can take weeks or months. On the other hand, modeling driver glances represents a promising approach to reduce the cost and time to assess the distraction potential of displays. N-SEEV is a computational model that was developed to predict a time to detect visual changes on displays (Steelman-Allen, McCarley, & Wickens, 2009). The model is

based on four attentional influences: saliency of the signal, effort required to attend to the signal, expectancy of the signal, and value of the signal. Although it considers multiple attentional influences, N-SEEV does not fit perfectly with driving situations, as it focuses on detecting changes while monitoring a display. Also, it requires experts to determine the perceived value of design features, making the model less accessible to ordinary designers. The tool introduced in this paper is simpler than N-SEEV, with focused functionality: it predicts search time associated with saliency of the visual objects. We used markedly advanced saliency model compared to the one implemented in N-SEEV model.

We launched a web-based tool that predicts the distraction potential of design alternatives based on the saliency of the display elements by estimating search time. The purpose of this tool is to support designers with a simple model of visual attention that can be used as an interactive guideline to identify misplaced saliency that might otherwise go unaddressed.

METHODS

User Interface

Rather than empirical results, we present a prototype web tool that can serve as an interactive design guide. The program is written in Python and Django and deployed on web (<http://distraction.engr.wisc.edu/>). Figure 1(a) shows how designers can upload two image files representing different design options. On the first page, designers set five input parameters: screen resolutions (width and height, in pixels), screen dimensions (width and height, in millimeters), and the distance from the driver to the image (in millimeters). These parameters are used to determine the size of foveal area of the display and region of interest (e.g., an icon). On the next page, the original image and the image after saliency calculation are displayed. Designers can select region of interest (ROI) for each of two images (Figure 1(b)). The ROI identifies the object that the driver would be searching for. The selected ROI is immediately shown on the corresponding saliency map. On pressing the 'Simulate' button, the simulation executes 1000 search trials where a series of eye movements are produced to estimate the number of fixations and associated glance time required to locate the ROI. The trials can be thought of as instances where drivers look to the display trying to find a particular icon. The right side of Figure 1(b) shows the results as Pareto graphs, showing the distribution of glances and highlighting the proportion of glances longer than two seconds.

Simulating Eye Fixations

We used Boolean Map based Saliency model (BMS) (Zhang and Sclaroff 2013), which is currently the most accurate saliency algorithm according to the MIT Saliency Benchmark (<http://saliency.mit.edu/>). BMS uses surroundedness, which is a geometric cue that Gestalt psychology has identified as guiding figure and ground assignment. Regions that are surrounded are seen as the figure. BMS identifies regions separated from the surrounding background to create a Boolean map (Huang & Pashler, 2007) and uses that to compute saliency. A saliency map calculated from BMS is a normalized black-and-white image where the most salient areas are white and least salient are black.

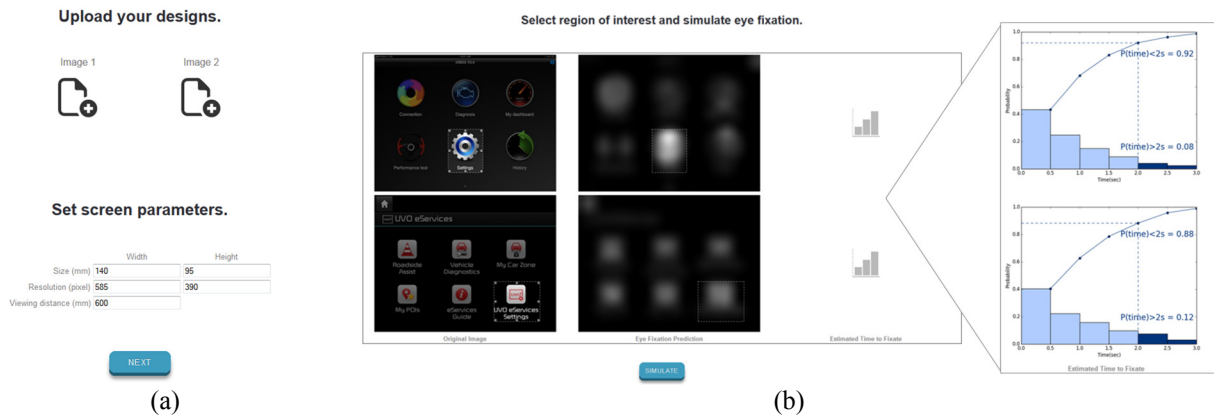


Figure 1. User interface of the glance calculation web application. Designers can upload designs and set parameters on the front page (a). On the next page (b), designers can see original images and corresponding saliency maps. The results appear once ROI is selected for each image and the ‘simulate’ button is clicked

Considering visual attention as a spotlight, when ones’ eyes focus on a certain location, only about 2.0 degrees around the fixation point is seen clearly (Henderson, 2003). To locate a target outside the focal area, eyes saccade rapidly over roughly 2.5 ~ 20 degrees (Land, Mennie, & Rusted, 1999). The present application assumed that eyes fixate on points across the image until they arrives within 2.0 degrees of the ROI. The probability of a saccade landing on a particular location is proportional to its saliency. In the current model, the saliency map was divided into rectangular cells, where each cell corresponds to the 2.0 degrees span of focal vision. The pixel size of the area was calculated based on display size and resolution as well as the distance from display to the viewer. The formula below was used to calculate how many image pixels (p) are within k degrees of visual field. The variable d is the distance from the observer and the image, r is the image resolution in pixel, and l is the image length or width. This implementation used 2 for k . This formula was used to calculate the number of pixels for width and height.

$$p = \frac{2dr}{l} \cdot \tan\left(\frac{k}{2} \text{rad}\right) \quad (1)$$

Average saliency within each cell was then calculated, and this value represented saliency of the cell. Figure 2 (c) shows how the saliency map was divided into cells and was assigned with a saliency value. From this value, the possibility of eye fixation was calculated as the formula (2) below. An assumption here was that the probability of eye fixation is proportional to the saliency of a cell. When there are n total cells, the possibility of eye fixation moving from cell i to cell j (P_{ij}) was calculated based on the saliency of cell i (s_i) and cell j (s_j):

$$P_{ij} = \frac{s_j}{\sum_{k=1}^n s_k - s_i} \quad (2)$$

It was assumed that the fixation starts in the center of the image because people tend to fixate in the center of images (Tatler, 2007). After this initial fixation, the location of the next fixation

was calculated according to the probability associated with the cell saliency, as in formula (2). With each fixation the program added 230ms to the total time elapsed, assuming 230ms between saccades (Henderson & Hollingworth, 1999). A cell that has been visited was not visited again, reflecting the inhibition of return. With each fixation, the system also checked if its destination cell included an ROI. The search terminated once a fixation arrived at the ROI and then the total time elapsed was recorded. The process was repeated 1000 times, and the result was represented as a Pareto chart. Figure 2 summarizes this process.

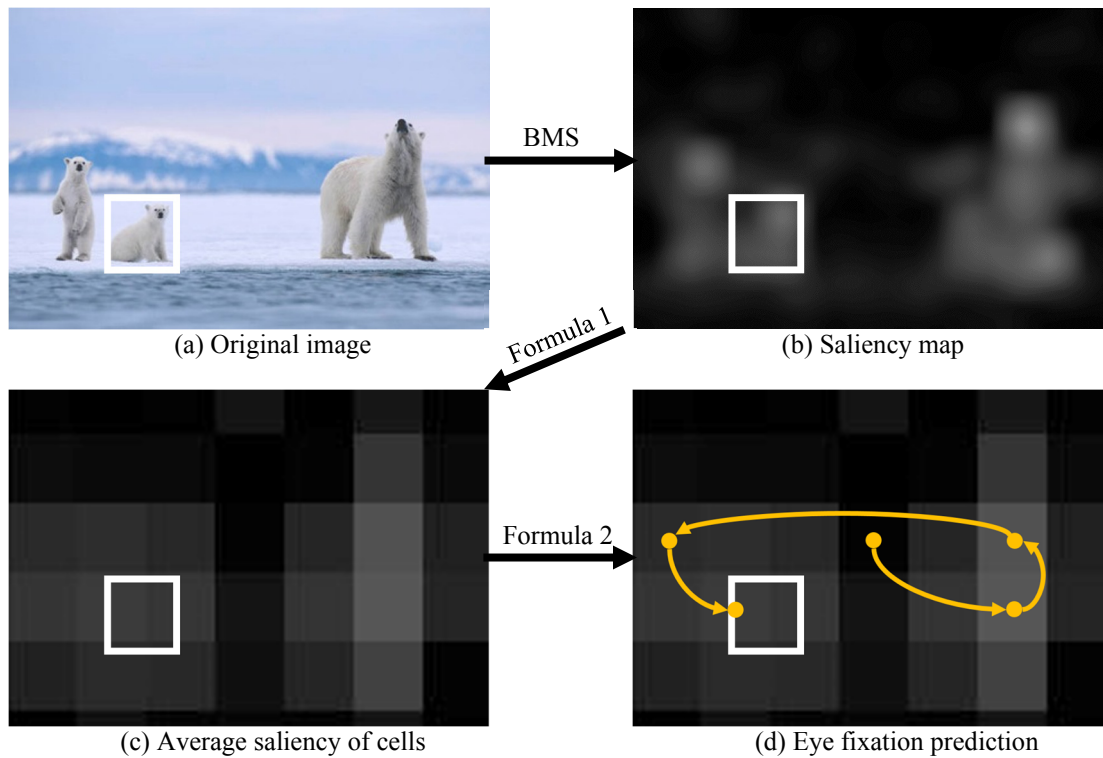


Figure 2. Example of an eye fixation simulation. Image (d) is an example of one run with four saccades and fixations after initial central fixation. This run of simulation is recorded as $230 \times 4 = 920\text{ms}$

RESULTS

The application allows designers to compare expected percentage of long glances (over 2.0 seconds) for multiple ROIs and multiple design concepts. For example, a designer can test distraction potential of two different icons in one screen. The first two results from Figure 3 shows the eye glance predictions for two different icons: 'Maps' and 'Settings'. Pareto plots show that 'Maps' is easier to locate than 'Settings'. The software estimates that 4% of trials will fail to locate 'Maps' within 2.0 seconds, but 39% will fail to locate 'Settings'. A designer can also predict eye glance durations for the same function 'Settings' in two different design alternatives. It would more likely to take longer to find 'Settings' icon in the second design ($P(\text{time} > 2\text{s}) = 0.55$) than in the first design ($P(\text{time} > 2\text{s}) = 0.39$). Such comparison can help explore trade-offs between design parameters.

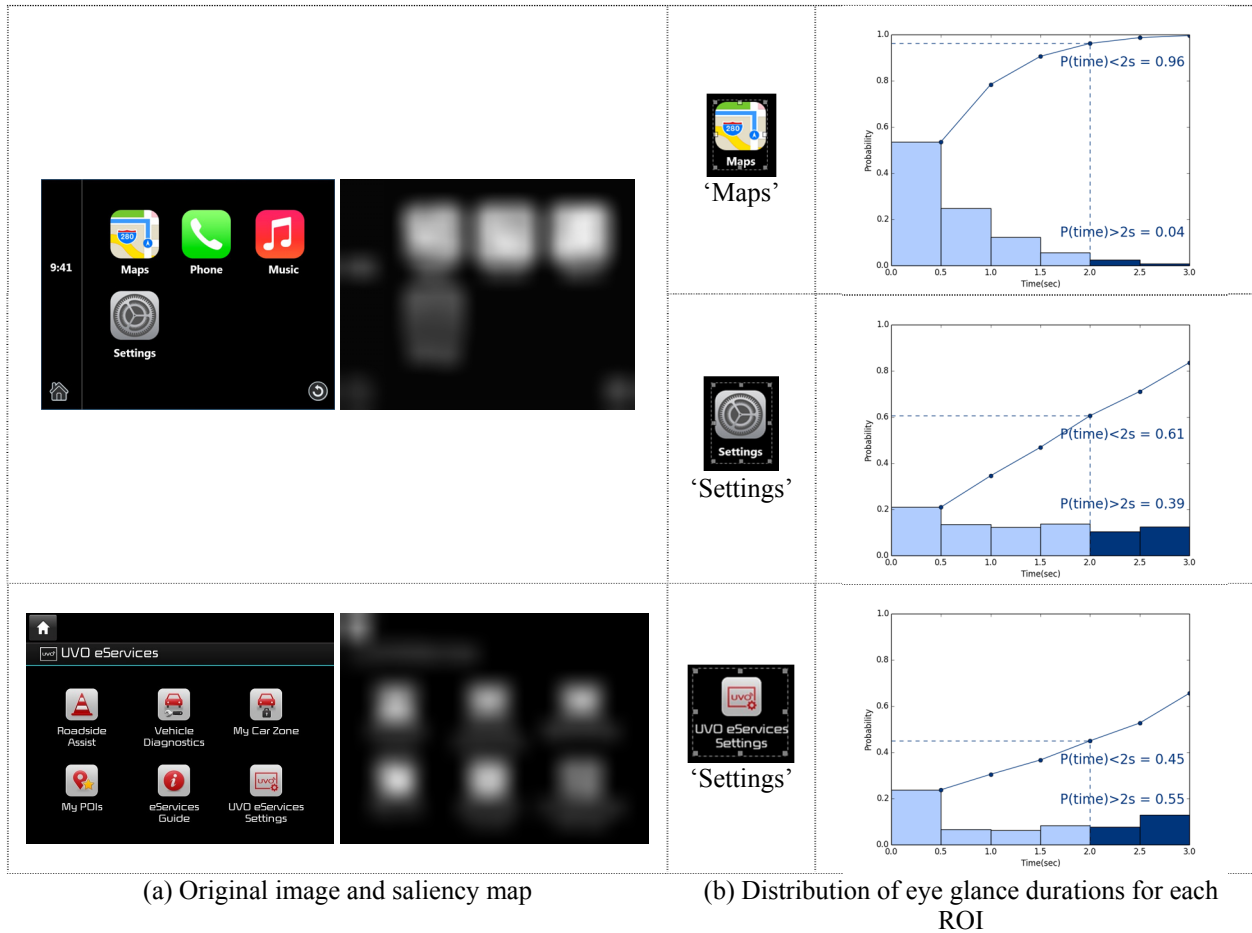


Figure 3. Sample runs of the application with different designs

CONCLUSION

A fundamental challenge in evaluating the distraction potential of vehicle displays is the costly and time-consuming data collection with human participants. Written design guidelines help address this issue, but they rarely provide quantitative results to guide design. They often fail to support tradeoff analysis in comparing alternate designs. Computational models, such as the one described in this paper can help bridge this gap. Such models can improve the design workflow by providing designers with immediate and measurable feedback regarding their design alternatives. This standalone web-based application makes it easy for designers to reflect on human factors considerations at the start of the design process and thus reduce the cost of redesigning interfaces at later stages. In addition, hundreds of design alternatives can be quickly iteratively assessed, helping to producing less distracting interface designs.

This software also has a potential to be integrated into more comprehensive driver distraction models. Many driver models have been developed to predict drivers' behaviors in using IVIS, but none of them simulate dynamic eye glances guided by saliency of visual objects. For example, the Keystroke Level Model puts multiple perceptual processes together into 'mentally prepare for 1.35 seconds' (Pettitt, Burnett, & Stevens, 2007), and Distract-R allocates 150ms for finding and processing visual objects (Salvucci, Zuber, Beregoiaia, & Markley, 2005). Likewise,

the Queuing Network Model uses a strategy in which eyes fixate on the closest unattended visual field with a target feature (Lim & Liu, 2004). These static values can be replaced with dynamic values that reflect specific display characteristics. For instance, the Itti & Koch saliency model has been combined with Distract-R driver model to predict driver distraction (Lee, 2014), where saliency model captures bottom-up (e.g., visual attention) influences and driver model captures top-down (e.g., goal) influences on driving behaviors. Integrating the present application into these driver models will allow better simulation of driving with secondary tasks. Specifically, such integration makes it possible to account for value- or expectation-driven attention (top-down) that might dominate the search of objects of interest with familiar displays in addition to the saliency of visual objects (bottom-up). It will help address the important fact that saliency is not the only force governing glance duration and visual search.

The current model captures important features of saliency-driven attention, and so is an example of a simple model that represents a specific human function and can address a limited set of design issues (Rasmussen, 1983). It addresses the issue on misplaced saliency based on well-established theory of visual perception and attention and has been validated on several test datasets from other domains. As such, this tool represents an interactive design guideline conveniently available on the web that can help designers address one contributor to visual distraction. Nevertheless, further validation with vehicle display designs is required to fully demonstrate the utility of this tool. Input from potential end-users (i.e., designers) will also help refine the tool to better support the design process.

REFERENCES

- Bylinskii, Z., Judd, T., Durand, F., Oliva, A., & Torralba, A. (2012). MIT saliency benchmark. <http://saliency.mit.edu/>
- Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in cognitive sciences*, 7(11), 498-504.
- Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual review of psychology*, 50(1), 243-271.
- Huang, L., & Pashler, H. (2007). A Boolean map theory of visual attention. *Psychological review*, 114(3), 599.
- Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision research*, 40(10), 1489-1506.
- Judd, T., Ehinger, K., Durand, F., & Torralba, A. (2009). Learning to predict where humans look. *Proceedings of the IEEE international conference on Computer Vision*, 2106-2113.
- Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. D., & Ramsey, D. J. (2006). *The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data. Technical Report No. DOT HS 810 594*. Washington, DC: National Highway Traffic Safety Administration (NHTSA).
- Land, M., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, 28(11), 1311-1328.

- Lee, J. (2014). *Integrating the saliency map with distract-r to assess driver distraction of vehicle displays*. (Doctoral dissertation) Retrieved from ProQuest Dissertations and Theses. (Order No. 3611485)
- Lim, J. H., & Liu, Y. (2004). A Queuing Network Model for Visual Search and Menu Selection. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 1846-1850.
- National Highway Traffic Safety Administration. (2014). *Visual-manual NHTSA driver distraction guidelines for in-vehicle electronic devices*. Washington, DC: National Highway Traffic Safety Administration (NHTSA), Department of Transportation (DOT).
- Navalpakkam, V., & Itti, L. (2006). An integrated model of top-down and bottom-up attention for optimizing detection speed. *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2049-2056.
- Pettitt, M., Burnett, G., & Stevens, A. (2007). An extended keystroke level model (KLM) for predicting the visual demand of in-vehicle information systems. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1515-1524.
- Rasmussen, J. (1983). Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *Systems, Man and Cybernetics*, IEEE Transactions on, (3), 257-266.
- Salvucci, D. D., Zuber, M., Beregoaia, E., & Markley, D. (2005). *Distract-R: Rapid prototyping and evaluation of in-vehicle interfaces*. *Proceedings of the SIGCHI conference on Human factors in computing systems*, 581-589.
- Steelman-Allen, K. S., McCarley, J. S., Wickens, C., Sebok, A., & Bzostek, J. (2009, October). N-SEEV: A computational model of attention and noticing. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 53, No. 12, pp. 774-778). SAGE Publications.
- Tatler, B. W. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision*, 7(14), 4.
- Tsimhoni, O., & Green, P. (2001). Visual demand of driving and execution of display-intensive in-vehicle tasks. *Proceedings of Human Factors and Ergonomics Society Annual Meeting*, 1586-1590.
- Wickens, C. D., Goh, J., Helleberg, J., Horrey, W. J., & Talleur, D. A. (2003). Attentional models of multitask pilot performance using advanced display technology. *Human factors*, 45(3), 360-380.
- Zhang, J., & Sclaroff, S. (2013). Saliency detection: a boolean map approach. *Proceedings of IEEE International Conference on Computer Vision*, 153-160.