Applied Ergonomics 60 (2017) 293–304

Contents lists available at ScienceDirect

Applied Ergonomics

journal homepage: www.elsevier.com/locate/apergo

How much time do drivers need to obtain situation awareness? A laboratory-based study of automated driving

Zhenji Lu*,[1], Xander Coster [1], Joost de Winter

Department of BioMechanical Engineering, Faculty of Mechanical, Maritime and Materials Engineering, Delft University of Technology, Mekelweg 2, 2628 CD, Delft, The Netherlands

ARTICLE INFO

Article history:
Received 6 June 2016
Received in revised form
28 November 2016
Accepted 4 December 2016

Keywords:
Scene perception
Spatial memory
Eye tracking
Hazard perception

ABSTRACT

Drivers of automated cars may occasionally need to take back manual control after a period of inattentiveness. At present, it is unknown how long it takes to build up situation awareness of a traffic situation. In this study, 34 participants were presented with animated video clips of traffic situations on a three-lane road, from an egocentric viewpoint on a monitor equipped with eye tracker. Each participant viewed 24 videos of different durations (1, 3, 7, 9, 12, or 20 s). After each video, participants reproduced the end of the video by placing cars in a top-down view, and indicated the relative speeds of the placed cars with respect to the ego-vehicle. Results showed that the longer the video length, the lower the absolute error of the number of placed cars, the lower the total distance error between the placed cars and actual cars, and the lower the geometric difference between the placed cars and the actual cars. These effects appeared to be saturated at video lengths of 7–12 s. The total speed error between placed and actual cars also reduced with video length, but showed no saturation up to 20 s. Glance frequencies to the mirrors decreased with observation time, which is consistent with the notion that participants first estimated the spatial pattern of cars after which they directed their attention to individual cars. In conclusion, observers are able to reproduce the layout of a situation quickly, but the assessment of relative speeds takes 20 s or more.

1. Introduction

Over the past few decades, an increasing number of automated driving systems have become available, both for research and consumer purposes. Automated driving systems in which the driver can opt to be 'out-of-the-loop' by engaging in non-driving tasks such as reading or resting are expected within a decade (Kyriakidis et al., 2015; Underwood, 2014). When the automation malfunctions or reaches its functional limits, control has to be handed back to the driver. In such cases, the automation typically issues a warning signal (also called a take-over request, see Lu et al., 2016 for a review) after which the driver has to resume the driving task (SAE International, 2014).

A critical design parameter in the development of automated driving system is the available time for taking over control, sometimes referred to as 'lead time', 'time buffer', or 'time budget' (Gasser and Westhoff, 2012; SAE International, 2014; Zeeb et al., 2016). If drivers do not have sufficient time to assess the situation prior to taking control, an accident may result (Mok et al., 2015). Drivers may prefer long lead times to prepare for the upcoming transition of control, but in reality, this is not always technologically feasible. For example, limitations in sensors (e.g., the limited range of a forward-facing radar) pose barriers regarding the maximum lead time that is feasible. In summary, it is important to understand how much time drivers need for gaining situation awareness (SA), because this sets demands on the automated driving technology.

Various studies have previously examined the effects of lead time on drivers' behaviour after resuming control (Clark and Feng, 2015; Gold et al., 2013; Mok et al., 2015). For example, a driving simulator study by Gold et al. (2013) found that the less time is available until colliding with a stationary object (5 s vs. 7 s), the more abrupt are the drivers' braking and steering inputs after receiving a take-over request. This study reported an average gaze reaction time (i.e., the time between the take-over request and the eye-gaze moving away from the non-driving task) of 0.5 s, an

* Corresponding author.
E-mail address: z.lu@tudelft.nl (Z. Lu).
[1] Joint first authors.

http://dx.doi.org/10.1016/j.apergo.2016.12.003

average hands-on-steering-wheel time of 1.5 s, and an average mirror-scan time of about 3 s (for similar results see Kerschbaum et al., 2015). Van den Beukel and Van der Voort (2013) found a decrease in the number of accidents and higher self-reported SA scores when more time was available. Mok et al. (2015) found that only few participants in a 2 s lead-time condition safely negotiated a hazardous situation, while the 5 s and 8 s conditions yielded considerably safer driver behaviours. A driving simulator study by Samuel et al. (2016) compared 4, 6, 8, and 12 s lead times, and found that participants needed a lead time of at least 8 s in order to detect a latent pedestrian hazard with the same accuracy as they did when being in control of the vehicle. Driving simulator research by Merat et al. (2014) and by Desmond et al. (1998) suggests that it may take up to 20 s or 40 s before vehicle control is fully stabilised after reclaiming control. Although the above studies provide valuable knowledge, they do not offer much insight into the cognitive processes of *how* drivers are able to build SA of a traffic situation as a function of the available time.

Over the last 25 years, the topic of SA has been extensively investigated (Endsley, 2015). The Situation Awareness Global Assessment Technique (SAGAT) is one of the standard instruments for measuring SA (Garland and Endsley, 1995). In this method, the screens of a simulator are temporarily blanked, and participants subsequently have to answer queries about objects and unfolding events in the simulation. Although SAGAT has been criticized for the fact that it measures SA intermittently rather than continuously, and for relying heavily on memory skills (for discussion see Durso et al., 2006; Gutzwiller and Clegg, 2013; Stanton et al., 2015), there is now a sound body of literature showing that SAGAT scores exhibit criterion validity with respect to task performance (Durso et al., 1995; Gardner et al., 2015; Loft et al., 2015; Salmon et al., 2009). Various promising alternative methods have been proposed for measuring SA, such as real-time probing (e.g., Loft et al., 2013; Martelaro et al., 2015; Pierce, 2012) and physiological techniques (e.g., Crundall et al., 2003; Gugerty, 2011), but at present SAGAT still appears to be the most widely applied and validated SA assessment tool (see also Endsley, 2000, 2015).

The answer categories in SAGAT are usually discrete or discretised values of the state of the virtual environment (e.g., Salmon et al., 2006; Loft et al., 2015). Gugerty (1997) used a similar technique as SAGAT for measuring participants' dynamic spatial memory by means of continuous values. Specifically, participants watched animations of traffic situations that lasted 18–35 s, and after each video, they indicated the positions of surrounding cars from a top-down view. Participants' level of SA was operationalized by comparing the positions of the placed cars with the actual positions of the cars in the animation. Gugerty found that the more cars are to be recalled, the poorer the performance on the SA task. Furthermore, he found that when the number of cars was larger, participants showed a prioritization effect whereby the most hazardous cars were remembered best.

In the present research, we refined the method used by Gugerty (1997) for determining the effect of time on SA scores. Specifically, we investigated the effect of viewing time (i.e., video length) with two levels of traffic density, namely 4 or 6 cars in surrounding traffic. The use of 4 and 6 cars is in approximate agreement with Pylyshyn and Storm (1988), who found that people can track up to five moving objects in a perceptual task, and with Gugerty (1997) who used 3 to 8 cars in his research. In our study, six different video lengths were adopted, ranging between 1 s and 20 s. The video lengths were based partly on a pilot study conducted prior to the present study (Coster, 2015). In this pilot, seven participants watched videos of animated traffic scenes and pressed the spacebar when they had assessed the situation to such an extent that they would feel safe to take over control. The results showed that a viewing time of 12 s was generally deemed sufficient, with an overall minimum of 3 s. In visual processing research, it has been found that participants can recognize the gist of a scene when having viewed it for only 20 ms (Thorpe et al., 1996). Oestmann et al. (1988) found that radiologists were able to detect ‘subtle cancers’ and ‘obvious cancers’ from a radiograph in 0.25 s with true positive rates of 30% and 70%, respectively (cf. 74% and 98%, respectively, for unlimited viewing times). However, sub-second viewing times are probably too short for processing dynamic traffic scenes that require visual search by means of multiple fixations and saccades (see Rayner, 2009 for a review on eye movements and visual search). Lead times that are typically used in driving simulator research range between 2 s and 12 s (Gold et al., 2013; Körber et al., 2015, 2016; Melcher et al., 2015; Mok et al., 2015; Samuel et al., 2016). In summary, our range of video lengths encompasses lead times that are commonly used, and range from extremely short (1 s) to longer than has been studied before (20 s).

The dependent measures in this study were: (1) self-reported task difficulty and time sufficiency, (2) the absolute error between the number of placed cars and the actual number of cars, (3) the error between the positions and indicated speeds of the placed cars relative to the actual positions/speeds of the cars, making use of an algorithm that globally selects a match between placed and actual cars by minimizing the positional difference, (4) the geometric difference between the positions of the placed and actual cars, and (5) eye-gaze activity. We expected that when the viewing time is longer, participants would find the task easier and have a better reproduction performance. Our corresponding aim was to explore at which video length these effects would saturate.

The geometric difference method is an innovation in SA assessment. It is based on a method for comparing polygons previously introduced by Arkin et al. (1991), which was said to be “invariant under translation, rotation, and change of scale, reasonably easy to compute, and intuitive” (p. 209). We applied this technique to obtain a generic index of difference that avoids the use of arbitrary parameters, such as correction factors related to the fact that people have a tendency to underestimate the distance to objects in virtual environments.

Eye tracking is widely used in studies of hazard perception, a term often equated with SA (Horswill and McKenna, 2004; Hosking et al., 2010; Underwood et al., 2002, 2013). We used eye tracking to gain a deeper understanding of how participants build SA as a function of time. It is well known that eye movements are correlated with bottom-up and top-down attention (Borji and Itti, 2013; Henderson, 2003; Itti and Koch, 2001) and memory of visual objects (Irwin and Zelinsky, 2002; Moore and Gugerty, 2010). For example, using a SAGAT method, Moore and Gugerty (2010) found that the more participants fixated on an aircraft in an air traffic control task, the higher their SA (i.e., responses to state queries) for that aircraft. Unema et al. (2005) and Over et al. (2007) found that in visual search tasks, participants exhibit a course-to-fine eye-movement strategy, whereby the first fixations and saccades had a short duration and large amplitude, respectively, and later fixations became longer-lasting with smaller-amplitude saccades in between the fixations. In this paper, we measured whether participants glanced at the road or at the mirrors, and how frequently they glanced at the mirrors, as a function of observation time. We explored whether these measures of attention distribution and glance frequency exhibit similar saturation profiles as the objective task scores.