Proceedings of the 9th ACM International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '17), September 24–27, 2016, Oldenburg, Germany .

Visual Attention During Simulated Autonomous Driving in the US and Japan

Yumiko Shinohara
Kyoto Institute of Technology,
Kyoto, Japan
yumiko.shinohara.918@gmail.com

Rebecca Currano
Stanford University
Stanford, CA, USA
bcurrano@stanford.edu

Wendy Ju
Stanford University
Stanford, CA, USA
wendyju@stanford.edu

Yukiko Nishizaki
Kyoto Institute of Technology,
Kyoto, Japan
yukikon@kit.ac.jp

ABSTRACT

To explore cultural differences in driver behavior for the purposes of vehicle automation, we used eye tracking to measure fixation patterns of Japanese and US participants (*N* = 98) viewing video simulations of automated driving through San Francisco and Osaka. After each drive, we asked participants questions about objects and events from the video.

Japanese participants showed higher fixation counts and durations than US participants for salient foreground objects in the traffic scene, and answered questions about those objects more correctly. US participants showed higher fixation counts than Japanese participants on visually prominent background features, and gave more correct answers about those. Consistency between fixation count and fixation duration on an object or feature, and recall of that object or feature, suggests that situation awareness during simulated automated driving proceeds from patterns of visual fixation on elements of the traffic scene.

Author Keywords

Eye movements; eye tracking; cross-culture; visual perception; gaze; situation awareness; automated driving.

CCS Concepts

Human centered computing---Interaction design;500 and Human centered computing---Empirical studies in interaction design;500.

INTRODUCTION

Roughly 90 percent of automotive accidents are attributed to human drivers rather than vehicle defects or environmental causes [15,24]. Moreover, about 41 percent of driver-related critical reasons have been found to be recognition errors [15]. Automated vehicles should support safer driving and reduce car accidents caused by human errors, especially human recognition errors. However, the development of automation in vehicles is likely to come in phases; most predict that drivers will need to cooperate with automated driving systems through some levels of automation before vehicles reach a level of full automation [23].

Figure 1. Screen shot from San Francisco drive video, showing the clock tower (framed in white) and the double decker bus.

Figure 2. Screen shot from Osaka drive video, showing a store with a big red sign (framed in white).

Safe cooperation between drivers and automated vehicles, especially during transitions from automated to manual driving, will depend on both the driver's situation awareness, and the automated system's assessment of the driver's situation awareness. Endsley and Kirit [6] found that situation awareness decreases with shifts from active to passive mental processing associated with increasing levels

ACM ISBN 978-1-4503-5150-8/17/09…$15.00.
https://doi.org/10.1145/3122986.3122991

of automation in expert systems. As automated driving systems become more capable, the driver's role will become more passive, resulting in decreased attention and reduced situation awareness. Models of driver attention and situation awareness, however, are often assumed to be universal, and this assumption may be flawed. Previous studies have shown that cognitive characteristics of visual perception can be influenced by culture [3,14,16]; Researchers found that Westerners tended to focus more on focal foreground objects in a scene, whereas East Asians were likely spend more time focused on contextual or background information.

Cross-cultural studies in driving often focus on road safety with respect to cultural differences in driving behavior [12,18,28]. Other research has used eye-tracking in driving studies to examine the relationship between driving performance or behavior and cognitive load [9,20,22,27].

In this paper, we present the first study that examines cross-cultural differences in gaze patterns for drivers in the US and Japan, in automated driving scenarios through San Francisco and Osaka. We also examined the dimension of experience on gaze patterns. This study shows evidence of differences in gaze pattern that need to be accounted for in designing systems that assess drivers' situation awareness.

RELATED WORK

Endsley [5] models situation awareness in three levels: perception, comprehension, and projection, or the ability to perceive surroundings, understand the current state, and project the upcoming state. Situation awareness is measured by several different methods, including subjective ratings, post-questionnaire, eye tracking, and Endlsey's SAGAT method, which stops the activity at various points in order to ask questions in the moment [4]. Post-questionnaires introduce time lag between participants' experiences and the questions that can be problematic. Physiological measures such as eye tracking capture data in-the-moment, but require inferring cognitive awareness. While the SAGAT method avoids problems related to time gaps between activity and post-questionnaires, the intermittent pauses in simulation during the activity can adversely affect sense of presence in the simulated world, and erode ecological validity [11,26]. It can also influence participants' eye gaze, such that they scan the scene differently than they might do without interruption.

Chua et al. [3] studied object recall and gaze patterns of American and Chinese participants while viewing photographs of a single focal object on a complex background. Both groups had higher fixation count on the backgrounds than on objects, and longer duration fixations on objects than on backgrounds. However, Chinese participants made more fixations on backgrounds (and more fixations overall) than Americans, while Americans had longer fixation durations overall, and showed significantly longer duration fixations on objects than backgrounds, compared to Chinese. Additionally, Americans fixated on objects sooner than Chinese, and Chinese were less likely than Americans to recall having seen a focal object when it was presented on a new background. The authors noted that differences in recall are influenced by differences in visual attention.

Rayner et al. [21] compared fixations for Chinese and American participants while viewing scenes with up to six points of interest on uncluttered backgrounds. They found that Chinese participants demonstrated fixations of higher numbers and shorter durations, overall, than Americans, while viewing scenes, however, they did not see differences in time spent on foreground vs. background between the 2 groups. They noted that their inability to replicate the findings of Chua et al. with respect to foregrounds and backgrounds may be due to Rayner's stimuli having more numerous foreground objects, and proportionally less background area. It is also possible that in such scenes, the difference between foreground (which they defined as any focal object), and background were not as clearly distinct as in Chua's study. Rayner et al. defined foreground as any points of interest. In one case, the top of a mountain was considered focal, while in Chua's study, a mountain top prominent in one picture was considered part of the background. These discrepancies indicate possible difference in definition for foreground and background, which could also impact their findings differently. However, when analyzing a subset of the data having stimuli more comparable to Chua's, Rayner et al. found that Chinese participants spent less time viewing foreground objects and made more fixations on the background than American participants. This suggests that Chua's findings may apply only to cultural differences in viewing scenes with low numbers of foreground objects.

Goh et al. [8] studied eye movements of Chinese Singaporean and Caucasian US participants when viewing pictures in which foreground objects and backgrounds were selectively changed. They found that US participants had longer mean fixation durations for both objects and backgrounds than Singapore participants. Also, US participants' eye movements remained more within the focal object, while Singapore participants' gaze moved between objects and backgrounds more.

Masuda and Nisbett [14] compared Japanese and Americans descriptions of short animated underwater scenes, containing focal objects and contextual features. They found that Americans first described focal objects more than Japanese, and Japanese more often began by describing contextual objects. Additionally, Japanese included 60% more contextual information in their descriptions. This suggests greater prioritization of, and potentially greater awareness of contextual features by Japanese than American participants.

We believe that cultural differences will similarly impact visual attention, and in turn, situation awareness, during the highly visual task of driving. For example, Japanese drivers may attend more to, and therefore be more aware of contextual or background features in the traffic environment than American drivers. Background features include non-mobile objects such as buildings, trees and light posts, and people or objects not in the roadway or in near proximity to the participant's vehicle. Conversely, Americans may have higher degrees of awareness of more salient foreground objects and features in the traffic environment. By salient, we mean people or objects in the roadway, whose mobility, proximity to the participant's vehicle, and presence in the lanes of traffic render them potential hazards and/or vulnerable to harm.

While most of the prior studies used static images as stimuli, or simple animated scenes, shown from a third person perspective, drivers experience complex dynamic traffic scenes, which they view from a first-person perspective. It is not clear whether the same cultural effects seen in prior studies will manifest in the task of driving. Cultural differences in viewing patterns that do emerge during driving, however, will likely indicate differences in attention and situation awareness in the driving scenario. Understanding these differences, therefore, is important in order to design automation systems that effectively evaluate and support situation awareness across different cultures.

STUDY DESCRIPTION

The current study explores gaze patterns and situation awareness between drivers in Japan and in the US to determine whether and how previous findings translate to the context of driving. We used eye tracking to look for cultural differences in visual attention to elements of the driving scene. We used post-drive questionnaires to assess cognitive awareness of these elements through recall. And finally, we looked for correlations between gaze patterns and situation awareness during simulated automated driving. Thus we combined methods to examine situation awareness through both physiological and recall measures, without the drawbacks of interrupting the driving simulations.

Hypothesis

Based on previous research, we hypothesized that US participants would show higher fixation counts and longer fixation times on salient foreground objects, which we defined as objects in the roadway, in immediate proximity to the driver's car, and which could present safety concerns, either to themselves or to the participant's vehicle. We also hypothesized that longer or more fixations on particular objects would correlate positively with greater awareness of those objects, as measured by more correct answers to questions about them in the situation awareness questionnaires.

METHOD

Participants

Fifty-one people (23 females and 28 males) in the US and 47 people (23 females and 24 males) in Japan participated in the study. US participants were between the ages of 18 and 58 (*M*=26.9, *SD*=9.09), and Japanese participants were the ages of 21 and 59 (*M*=32.3, *SD*=12.8). All of them agreed to the treatment of protection of personal data. This study was approved by Stanford University's Human Subjects Research and IRB (eProtocol# 30016), and the Kyoto Institute of Technology Institutional Review Board (Document No. 2016-21).

We divided the group from each location into two subgroups based on driving experience. Participants who had driven a car at least once a week for more than 2 years (in whole years) were defined as expert drivers (9 females and 7 males in the US, 11 females and 11 males in Japan). Those with less experience driving were defined as novice drivers (13 females and 20 males in the US, 12 females and 13 males in Japan). All Japanese participants had experience driving in Japan, but not in the US, and all US participants had experience driving in the US but not in Japan.

Apparatus and materials

Driving simulation

To test our hypotheses, we measured eye movements during two three-minute drives, using a Tobii X2-60 eye tracker, mounted just below on a 27-inch screen (*GW2255, Ben Q Corporation, Japan*). A T80 Thrustmaster steering wheel, mounted on the table in front of the screen, was not operational, but served to support the sense of being in the driver's seat of a vehicle. In order to get accustomed to the setup, participants first watched a one-minute driving video taken on a divided highway in Canada. This practice video showed a single direction of traffic along a rural route, and did not include any culturally specific characteristics. Participants were told, *"Imagine that you are sitting in the driver's seat of an autonomous car, and watch the video. You won't be able to control the steering, but please keep your hands on the wheel."*

Following the practice video, they watched two three-minute driving videos, one of a drive through San Francisco, and the other of a drive through Osaka, both at standard city speeds (approximately 25mph). Each of these was selected from a longer drive recorded by the researchers on a clear day in October between 1:30pm and 3:00pm, in their respective time zones. Both videos were recorded using a GoPro 4 camera mounted in the center of the windshield below the rear-view mirror. The 3-minute clips were selected to contain comparable key features, such as 2-way traffic, traffic density, at least one right and one left turn, stop lights, oncoming and preceding vehicles, similar urban settings, and presence of pedestrians, buildings, and other common urban environmental features.

Situation Awareness Questionnaire

Following the driving portion, each participant answered a situation awareness questionnaire (see Appendix). To ensure comparability between the two questionnaires, and because we did not know in advance which objects or features would be more likely to elicit different cultural responses, each contained a total of 18 questions and included at least one regarding each of the following features:

A. an object on a right or left lane
B. a pedestrian at a corner
C. a preceding vehicle

D. a crossing pedestrian or vehicle
E. a traffic sign or signal
F. a focal object

Seven to eight questions (in each) asked if the participant noticed an object ("Did you see…?", or "Was there a…?"). If the person responded yes, they were then presented with one or two sub-questions regarding that object or feature. Each questionnaire also contained one or two questions asking the participant to recall details about a traffic indicator (color of the light, speed limit, turn signal of a vehicle ahead). And one question in each asked the participant to reason whether their car could have performed an action. Two questions in each asked about an action taken by the participant's car and/or another vehicle (what direction it turned, or from what direction it came).

Each questionnaire also included two questions (a primary question and a sub-question) asking about something that was not present in the video, so as not to make participants feel as if they should answer "yes" to all questions.

Figures 1, 3, and 4 show objects asked about in the following sample questions from the San Francisco drive questionnaire:

- Did you see a clock tower? What direction did your car turn just before the clock tower? (Figure 1)
- Did you see a person riding a bicycle at the first intersection? Where was the bicycle? (Figure 3)
- Which direction did the double-decker bus turn at the last intersection? (Figure 4)

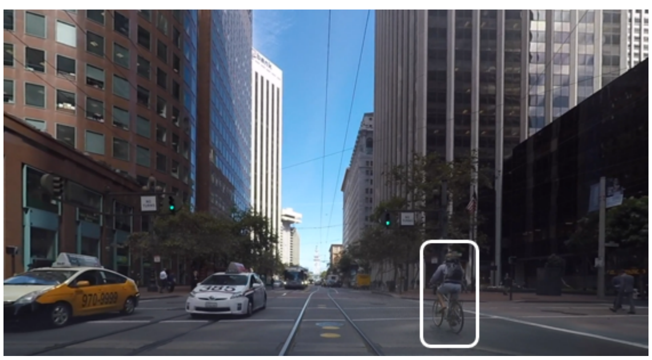

Figure 3. Screen shot from San Francisco drive video, showing the person riding a bicycle at the first intersection.

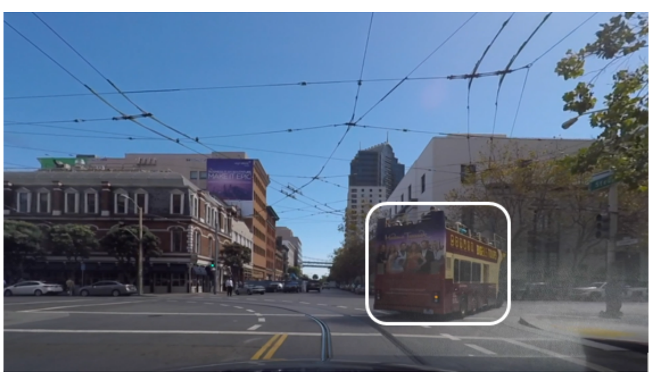

Figure 4. Screen shot from the San Francisco drive vide, showing the double-decker bus turning at the last intersection.

Figures 2, 5, and 6 show objects asked about in the following sample questions from the Osaka drive questionnaire:

- Did you see a store with a big red sign? What kind of store was it? (Figure 2)
- Did you see a pedestrian walking on the left side after the first intersection? What was the pedestrian doing? (Figure 5)
- Did you see a taxicab pull out in front of your car? Which direction did the taxicab come from? (Figure 6)

Figure 5. Screen shot from the Osaka drive video showing the pedestrian walking on the left side after the first intersection.

Figure 6. Screen shot from the Osaka drive video showing a taxicab pulling out in front of the participant's car.

Procedure

Before beginning the experiment, participants completed a questionnaire about their driving experience and cultural background. The experimenter provided a brief introduction and then participants viewed a practice video, followed by two three-minute videos of a drive through Osaka and San Francisco. Immediately after each drive video, the participant answered a situation awareness questionnaire about that video.

After finishing the first situation awareness questionnaire, participants were told, *"Please try to watch [the second video] in the same manner that you watched the first video. Although we asked you questions about what you saw in the last video, this is not a memory experiment, and it's important to maintain consistency in how you watch the videos, so we don't want you to modify the way you watched* *in order to remember things."* We recognize that despite their best efforts, participants may be influenced in how they watched the second video, so we counterbalanced the order of the two videos to deal with order effects.

The eye tracker was calibrated before the practice video and before each of the 3-minute videos. While the participant was watching the videos, the experimenter waited outside the experiment room.

Analysis of eye movements

We first used Olsson's [17] classification algorithm to automatically generate total fixation count, and total fixation duration across the entire video of each drive. From these we calculated mean fixation duration, and then looked for differences in overall fixation patterns. Additionally, we measured fixation count and fixation duration on key objects or events that were included in the situation awareness questionnaire. Figures 1-6 show screen shots from each of the drive videos with objects from the situation awareness questionnaire framed in white.

RESULTS

One US participant in the expert driver group did not watch the Osaka driving video due to experimenter error.

Eye Tracking (Overall)

We divided our data and ran two separate 2x2 (culture x experience) ANOVAs, one for each drive, to see if effects due to culture and experience can be seen in both an American and a Japanese city, despite the many differences that are inherent between the two cities.

Results are shown in Table 1. In the San Francisco driving video, main effects for culture were significant for mean fixation time: $F(1,93)= 9.04$, $p<.01$, and total fixation time: $F(1,93)=7.36$, $p<.01$, with US participants fixated for longer mean times and total times than Japan participants. Experience effects were significant for mean fixation time only: $F(1,93)=5.04$, $p<.05$, with novice participants having greater mean fixation times than expert drivers. In the Osaka driving video, main effects for culture and driving experience were significant for mean fixation time: $F(1, 92)=9.74$, $p<.01$, and total fixation time $F(1, 92)=3.35$, $p<.05$, respectively. A marginally significant interaction was seen for mean fixation time, $F(1, 92)=3.35$, $p<.10$. No significant effects were seen for total fixation count in the both videos.

For each drive, we adjusted the *p*-values according to the Benjamin & Hochberg [2] method (see adjusted *p*-values in parentheses in Table 3). This method was chosen as a middle ground, to balance the risk of increased Type 1 errors when running several tests, with the risk of increasing Type 2 errors associated with highly conservative adjustments such as Bonferroni correction [1]. Three of the four significant culture effects remained after adjusting the *p*-values.

Osaka Drive Video	Driving Experience	Japanese *M* (*SD*)	American *M* (*SD*)	Main Effects (Adjusted) Culture	Experience
Total Fixation Count	Novice	386.5(132.4)	381.4(119.4)	p=0.64 (p=0.64)	p=0.12 (p=0.18)
	Expert	408.0(92.69)	429.6(78.03)		
Mean Fixation Time (s)	Novice	0.21(0.06)	0.30(0.13)	**p=0.002 (*p=0.012)	*p=0.04 (p=0.08)
	Expert	0.20(0.07)	0.23(0.07)		
Total Fixation Time (s)	Novice	84.31(41.09)	111.7(44.52)	*p=0.02 (p=0.06)	p=0.44 (p=0.528)
	Expert	83.53(35.82)	98.89(38.65)		
San Francisco Drive Video					
Total Fixation Count	Novice	330.6(149.9)	382.6(119.1)	p=0.36 (p=0.432)	p=0.65 (p=0.65)
	Expert	365.9(115.1)	369.9(121.0)		
Mean Fixation Time (s)	Novice	0.20(0.06)	0.27(0.10)	**p=0.003 (*p=0.018)	*p=0.03 (p=0.06)
	Expert	0.18(0.06)	0.21(0.10)		
Total Fixation Time (s)	Novice	71.80(44.69)	106.3(43.99)	**p=0.008 (*p=0.024)	p = 0.16 (p=0.24)
	Expert	66.88(33.50)	84.46(48.54)		

Table 1. Results of 2x2 (Culture x Experience) ANOVA for Total Fixation Count, Mean Fixation, Time and Total Fixation Time over each of the full 3-minute drive videos.

Situation Awareness Questions

We did not approach analysis of the situation awareness questions with a specific hypothesis of which items in a driving scene might show significant cultural differences. Some ideas presented themselves: for example, objects in the middle of the scene (in the immediate vicinity of the driver's car), or objects of a certain type (people and vehicles, vs. signs and traffic signals). However, our intent was to approach the situation awareness questionnaire in an exploratory way rather than to test pre-formulated hypotheses. We therefore recognize that *p*-values resulting from the logistic regression analyses should be viewed as pointers to help us narrow down relationships to study further in future research, rather than evidence of true significant effects. Nonetheless, for the sake of a simplicity in explaining our findings, and to highlight interesting trends observed, we use the word 'significant' here when discussing effects where the *p*-value was less than 0.05, and 'marginally significant' where *p*-values were greater than 0.05 but less than 0.1. Table 2 shows the results from these analyses. For cases showing a culture effect, gray highlighting indicates greater higher rates of correct answers among the Japanese participants, while those not highlighted show higher rates among US participants. In both cases with an experience effect, novice drivers had higher rates of correct answers than experienced drivers.

In most cases, significance appears as an effect of culture, rather than experience. We therefore focus our attention primarily on culture for this part of the study. Interestingly, US participants scored higher on only four of the 14 situation awareness questions that showed significant or marginally significant cultural effects, while Japanese participants scored higher on ten.

Questions in the San Francisco Drive Questionnaire		Effect	Odds Ratio (CI 95%)	*p*
Did you see a person riding a bicycle at the first intersection?	Q1	Culture	4.93 (1.36-24.0)	*p=0.03
Where was the bicycle?	Q2	Culture	2.75 (0.90-9.55)	†p=0.09
What color was it? [Did you see a streetcar?]	Q9	Culture	2.69 (1.01-7.66)	†p=0.05
Did you see a clock tower?	Q10	Culture	0.35 (0.14-0.81)	*p=0.02
Which direction did your car turn just before the clock tower?	Q11	Culture	0.46 (0.20-1.05)	†p=0.07
Which direction did the double-decker bus turn at the last intersection?	Q14	Culture	2.63 (1.14-6.30)	*p=0.03
Did you see a pedestrian standing in the street when your car turned left?	Q15	Culture	0.46 (0.20-1.05)	†p=0.07
Questions in the Osaka Drive Questionnaire		**Effect**	**Odds Ratio (CI 95%)**	***p***
Was there a motorcycle ahead of you on the road?	Q3	Culture	3.58 (1.27-11.4)	*p=0.021
Which direction did the motorbike turn at the first intersection?	Q5	Culture	2.43 (1.06-5.73)	*p=0.039
What was the pedestrian doing?	Q10	Culture	2.98 (0.87-12.1)	†p=0.096
Did you see a store with a big red sign?	Q11	Culture	0.42 (0.18-0.98)	*p=0.047
What kind of store was it?	Q12	Experience	3.93 (1.02-19.2)	†p=0.059
What was the speed limit at the location where the video ended?	Q13	Culture	6.85 (1.01-136)	†p=0.089
Did you see a taxicab pull out in front of your car?	Q15	Experience	2.97 (0.95-11.3)	†p=0.078
Which direction did the taxicab come from?	Q16	Culture	2.87 (1.08-8.29)	*p=0.040
At the moment the video ended, what color was the traffic light?	Q17	Culture	6.38 (2.01-25.0)	**p=0.003

Table 2. Results of Multivariable Logistic Regression Analysis for Post Drive Questions. Rows highlighted in gray indicate Japanese scored higher than Americans. For the remaining culture effect case, Americans scored higher than Japanese. In both cases with an experience effect novice drivers scored higher than experienced drivers.

Eye Tracking (Specific Objects of Interest)

We ran 2x2 ANOVAS (culture, experience), examining fixation duration and fixation count for the objects referenced in the situation awareness questionnaires. Results from these ANOVAS are shown in Tables 3 and 4.

Culture Effects

Japanese participants fixated for longer (total) durations on the bicycle from the San Francisco Drive (Q1,2), and the Motorcycle from the Osaka Drive (Q3,4,5). American participants, on the other hand, fixated for longer (total) duration on the clock tower from the San Francisco Drive (Q10,11) and the big red sign from the Osaka drive (Q11,12).

We saw similar differences in fixation count on these objects as well, with Japanese participants fixating more times on the bicycle and the motorcycle, and US participants fixating more times on the clock tower and the big red sign.

Additionally, Japanese participants had higher fixation counts for the Pedestrian wearing a yellow jacket in the San Francisco drive (Q12,13) and for the Man walking between two cars and the Traffic light in the Osaka drive (Q14 and Q17, respectively).

Driving Experience Effects

With respect to driving experience, novice drivers had longer total fixation duration on the motorcycle in the Osaka drive (Q3,4,5), and expert drivers had longer total fixation duration on the big red sign in the Osaka drive (Q11,12).

Expert drivers had higher fixation counts for the bicycle in the San Francisco drive (Q1,2), and for the traffic light in the Osaka drive (Q17), while novice drivers had higher fixation counts for the motorcycle in the Osaka drive (Q3,4,5).

No significant Interaction effects were observed for fixation duration or fixation count in either drive. We found one marginal significant interaction for fixation duration on the motorcycle in the Osaka drive (Q3,4,5).

San Francisco Drive Objects	Driving Experience	Culture		Main Effects	
		Japanese *M (SD)*	American *M(SD)*	Culture	Experience
Bicycle (Q1,2)	Novice	0.22(0.30)	0.12(0.16)	+p=0.07	ns
	Expert	0.30(0.38)	0.17(0.29)		
Clock tower (Q10,11)	Novice	0.02(0.05)	0.05(0.11)	+p=0.08	ns
	Expert	0.03(0.07)	0.08(0.18)		
Osaka Drive Objects					
Motorcycle (Q3,4,5)	Novice	0.41(0.48)	0.14(0.25)	*p=0.05	*p=0.02
	Expert	0.13(0.20)	0.10(0.24)		
Big red sign (Q11, 12)	Novice	0.07(0.11)	0.23(0.33)	*p=0.02	+p=0.09
	Expert	0.18(0.31)	0.35(0.41)		

Table 3. Results of 2x2 (culture x experience) ANOVA for fixation duration on individual objects asked about in the Situation Awareness Questionnaires. Gray highlighting indicates Japanese had longer fixation durations than Americans.

San Francisco Drive Objects	Driving Experience	Culture		Main Effects	
		Japanese *M (SD)*	American *M(SD)*	Culture	Experience
Bicycle (Q1,2)	Novice	1.44(1.96)	0.56(0.73)	+p=0.05	+p=0.08
	Expert	2.14(2.38)	1.35(2.16)		
Clock tower (Q10,11)	Novice	0.12(0.44)	0.44(0.73)	+p=0.06	ns
	Expert	0.27(0.77)	0.71(1.40)		
Pedestrian wearing a yellow jacket (Q12,13)	Novice	1.80(2.90)	1.31(1.40)	+p=0.08	ns
	Expert	2.68(2.95)	1.50(2.02)		
Osaka Drive Objects					
Motorcycle (Q3,4,5)	Novice	1.80(1.87)	0.81(1.17)	*p=0.03	*p=0.03
	Expert	0.86(1.17)	0.52(1.06)		
Big red sign(Q11, 12)	Novice	0.64(0.99)	1.31(1.82)	*p=0.04	ns
	Expert	1.18(2.11)	2.03(2.04)		
Man walking between 2 cars (Q14)	Novice	1.88(1.92)	1.44(1.63)	+p=0.10 (0.0999)	ns
	Expert	1.68(2.06)	0.88(1.82)		
Traffic light (Q17)	Novice	1.00(1.63)	0.50(0.63)	+p=0.10 (0.097)	*p = 0.05 (0.0474)
	Expert	1.91(2.16)	1.15(2.15)		

Table 4. Results of 2x2 (culture x experience) ANOVA for fixation count on individual objects asked about in the Situation Awareness Questionnaires. Gray highlighting indicates Japanese had higher fixation counts than Americans. Rows not highlighted indicate Americans scored higher than Japanese.

DISCUSSION

For each of the full 3-minute drives, US participants demonstrated greater mean fixation and total fixation times than Japanese participants. This is consistent with Chua's [3] findings that Americans have longer fixation durations on both objects and backgrounds.

On the other hand, we did not see any significant differences in fixation count between US and Japanese participants, in contrast with Chua's finding that Eastern (Chinese) participants had higher fixation counts than Americans on backgrounds and as a result, overall. Furthermore, Japanese participants' higher fixation counts and longer fixation durations on salient foreground objects in our study stands in contrast to prior research findings that Western participants fixated longer and more frequently than East Asian participants on focal foreground objects.

These discrepancies may be due, in part, to the fact that our driving scenes contained many salient foreground objects, which could introduce ambiguity for both groups between what would be perceived as foreground vs. background in a complex dynamic driving scene. In complex traffic scenes, individual foreground objects may stand out less, and therefore be treated more like background features in Chua's more simple scenes.

Conversely, large, visually prominent background objects (the clock tower and the big red sign) may be treated in much the same way Chua's foreground objects were treated. This is notable, especially since these two objects were very different in key ways – the clock tower stood out in shape (though not color) against the mostly bare sky, while the big red sign stood out primarily in color against a more visually distracting building façade. This indicates that even strong visual differences in cityscape do not preclude the emergence of similar cultural effects on gaze patterns in the two cities.

This observation contradicts what we had expected because we considered the clock tower and the big red sign to be background features, rather than foreground objects. Chua et al. [3] found that Americans fixated for longer durations on focal foreground objects. However, the positioning of these features within the scene (in the center) and visual distinction from other nearby elements in the scene gives them a similarity to the foreground objects in Chua's study. This introduces the possibility that an object's visual prominence and placement within the scene drives fixation, rather than a participant's higher-level assessment of salience or foreground/ background categorization of the object.

Our results showed clear consistency between participants' answers to questions about specific objects and their fixation patterns on those same objects in the eye tracking data. Where Japanese (or US) participants demonstrated significantly longer fixation durations or higher fixation counts on a particular object, they also had higher rates of correct answers to the situation awareness questions which referenced that same object. This suggests that drivers' situation awareness and recall of a given object proceed from greater numbers and durations of fixations on that object.

LIMITATIONS AND FUTURE WORK

We note that our study contained some limitations which should be corrected in future research. First, because Japanese and American drivers drive on opposite sides of the street, it is possible that this difference has some effect on gaze patterns that may confound our findings. Therefore, we plan to conduct our next study with participants from two cultures which drive on the same side of the street (e.g. UK and Japan, or US and China). Second, while we controlled for driving experience in each country, our US participant base was somewhat multi-cultural (primarily, though not exclusively American- and European-born individuals) while the Japanese subjects were all Japanese-born. In future studies, we would more thoroughly screen and control for nation of birth to reduce potential effects due to different levels of homogeneity in each of our participant groups. Finally, participants anticipated the post-drive questionnaire following the second video, but not the first. However, analysis on the first video only showed very similar trends to what we observed from both videos, so it did not seem to be problematic. Despite all the limitations, we were still able to see some clear cultural effects.

Additionally, participants' knowledge of their inability to control steering could be seen as decreasing the realism of the simulation by setting expectations that they would not experience automation failures. This choice, however, was made intentionally, as we hoped to elicit and observe gaze patterns that we might expect when drivers become accustomed to, and comfortable with, reliable automated driving. In this state, drivers may relax their attention, which could present a reduced ability to safely resume control over the vehicle if necessary.

CONCLUSION

Our findings suggest that in viewing complex dynamic driving scenes, culture does impact fixation patterns and as a result, situation awareness of objects and events in the traffic environment. In our study, Japanese participants' gaze patterns put them at an advantage for situation awareness of objects most salient to the driving task, while US participant gaze patterns led to a greater situation awareness of prominent but non-salient background objects. Cross-cultural differences in drivers' eye movements, and subsequent differences in situation awareness, suggest that different interventions might be appropriate for people from different cultures during automated driving, to help establish situation awareness necessary for safe transitions from autonomous to manual driving.

ACKNOWLEDGMENTS

The authors thank Lorin Dole for his guidance and advice regarding statistical analyses, Dylan Moore for his help in recording the San Francisco drive video, and David Sirkin for his help with video and paper editing. We thank the KYOTO Design Lab at the Kyoto Institute of Technology, and Stanford's Center for Design Research for supporting this research study.

REFERENCES

1. R.A. Armstrong. 2014. When to use the Bonferroni correction. *Ophthalmic and Physiological Optics*, *34*(5), 502-508.
2. Yoav Benjamini and Yosef Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B* **57**, 289–300.
3. Hannah F. Chua, Julie E. Boland, Richard E. Nisbett. 2005. Cultural variation in eye movements during scene perception. *PNAS*, 102(35), 12629-12633.
4. Mica R. Endsley. 1995a. Measurement of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 37, 1 (1995), 65–84.
5. Mica R. Endsley. 1995b. Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37(1), 32-64.
6. M.R. Endsley, & E.O. Kiris. 1995. The Out-of-the-Loop Performance Problem and Level of Control in Automation. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *37*(2), 381–394.
7. Karen Gasper and Gerald L. Clore. 2002. Attending to the big picture: mood and global versus local processing of visual information", *Psychological Science*, Vol.13, No.1.
8. Joshua O. Goh, Jiat Chow Tan, Denise C. Park. 2009. Culture modulates eye-movements to Visual Novelty. PLoS ONE 4: e8238.
9. J.L. Harbluk, Y.I. Noy, & M. Eizenman. 2002. The impact of cognitive distraction on driver visual behaviour and vehicle control (No. TP# 13889 E).
10. Shinobu Kitayama, Sean Duffy, Tadashi Kawamura and Jeff T. Larson. 2003. Perceiving an object and context in different cultures: A cultural look at new look. *Psychological Science,* 14, 3.
11. Kwan Min Lee. 2004. Presence, explicated. *Communication Theory* 14, 1 (2004), 27–50.
12. A. Lindgren, F. Chen, P.W. Jordan, & H. Zhang (2008). Requirements for the design of advanced driver assistance systems-The differences between Swedish and Chinese drivers. *International Journal of Design*, *2*(2).
13. Andrew Mackinnon, Anthony F. Jorn, Helen Christensen et al. 1998. A Short form of the positive and negative affect schedule: evaluation of factorial validity and invariance across demographic variables in a community sample. *Personality and Individual Differences* 27 (1999) 405-416.
14. Takahiko Masuda, Richard E. Nisbett. 2006. Culture and change blindness. *Cognitive Science*, 30, 381-399.
15. National Highway Traffic Safety Administration. 2008. National motor vehicle crash causation survey: Report to congress. <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/811059>
16. Richard E. Nisbett, Yuri Miyamoto. 2005. The influence of culture: holistic versus analytic perception. *TRENDS in Cognitive Sciences*, 9(10), 467-473.
17. Pontus Ollson. 2007. Real-time and Offline Filters for Eye Tracking: Master's degree Project. *KTH Electrical Engineering*.
18. T. Özkan, T. Lajunen, J.E. Chliaoutakis, D. Parker, & H. Summala. 2006. Cross-cultural differences in driving behaviours: A comparison of six countries. *Transportation research part F: traffic psychology and behaviour*, *9*(3), 227-242.
19. Robert Paeglis, Kristaps Bluss, Aigars Atvars. 2011. Driving experience and special skills reflected in eye movements. Proc. Of SPIIE, Vol.8155.
20. O. Palinko, A.L. Kun, A. Shyrokov, & P. Heeman. 2010. Estimating cognitive load using remote eye tracking in a driving simulator. In *Proceedings of the 2010 symposium on eye-tracking research & applications*(pp. 141-144). ACM.
21. Keith Rayner, Xingshan Li, Carrick C. Williams, et al.. 2007. Eye movements during information processing tasks: Individual differences and cultural effects. *Vision Research*, 47, 2714-2726.
22. M.A. Recarte & L.M. Nunes. 2000. Effects of verbal and spatial-imagery tasks on eye fixations while driving. *Journal of experimental psychology: Applied*, *6*(1), 31.
23. SAE On-Road Automated Vehicle Standards Committee. 2014. Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems. Technical Report J3016_201401.
24. Paul Salmon, Michael Regan, Ian Johnston. 2005. Human error and road transport: phase one – A framework for an error tolerant road transport system. *Center Report Series*, 256.
25. Atsushi Sato and Asako Yasuda. 2001. Development of the Japanese of positive and negative affect schedule (PANAS) scales. *Japan Society of Personality Psychology*, 9-2, pp.138-139.
26. D. Sirkin, N. Martelaro, M. Johns, & W. Ju. 2017. Toward Measurement of Situation Awareness in Autonomous Vehicles. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 405-415). ACM.
27. T.W. Victor, J.L. Harbluk, & J.A. Engström. 2005. Sensitivity of eye-movement measures to in-vehicle task difficulty. *Transportation Research Part F: Traffic Psychology and Behaviour*, *8*(2), 167-190.

28. H.W. Warner, T. Özkan, T. Lajunen, & G. Tzamalouka. 2011. Cross-cultural comparison of drivers' tendency to commit different aberrant driving behaviours. *Transportation research part F: traffic psychology and behaviour*, *14*(5), 390-399.
29. David Watson, Lee Anna Clark and Auke Tellegen. 1988. Development and validation of brief measures of positive and negative affect: the PANAS scales", *Journal of Personality and Social Psychology*, Vol.54, No.6, pp.1063-1070.