DRIVER'S SITUATION AWARENESS DURING SUPERVISION OF AUTOMATED CONTROL – Comparison between SART and SAGAT measurement techniques

Arie P. van den Beukel, Mascha C. van der Voort

ABSTRACT: Systems enabling to drive automatically are being introduced on the market. When using this technology, drivers are in need for interfaces which support them with supervision of the automated control. Assessment of Situation Awareness (SA) which drivers are able to gain while using such interfaces, is important. Based on comparison between SART and SAGAT measurement techniques within a simulator study, the test set-up presented in this paper suggests to be successful in providing a coherent test-bed with relevant situations to assess the level of SA drivers gain when involved in supervision of automated control and while using different types of feedback.

1   INTRODUCTION

Automotive industry has started implementation of automated driving for the consumer market through introduction of driver assistance which allow both lateral and longitudinal system control during specific situations within existing infrastructure (e.g. motorway cruising). The systems introduced are based on semi-automation meaning that automation is only possible when specific boundary conditions are being met, like detection of road lines and driving on motorways. This requires human (driver) readiness to act as a back-up in case automation fails or exceeds her boundary limits. The role of the driver therefore changes from actively operating the vehicle to supervising the system during automation. However, performing supervisory tasks is related to low vigilance, causing e.g. slower reaction times and misinterpretation when intervention is needed [1]. Carefully designed driver-interfaces are therefore needed to support drivers in their additional role to supervise the automation. During this development, a difficulty is to assess the contribution potential interfaces have in supporting drivers with their supervisory task. Although it is commonly recognised by researchers that measurement of Situation Awareness (SA) is relevant to assess driver's

ability to take back control, there is limited consensus on the appropriate technique to measure SA. Two techniques are most common: SART (a self-assessment method) and SAGAT (a probe-taking method). The reliability and validity of both techniques are subject to discussion [2]. Also an earlier experiment by the author intended to measure SA in circumstances relevant for semi-automated driving (i.e. taking back control) showed contrary results between SART and SAGAT [3]. Although most existing studies show results in favour of SAGAT, by e.g. showing better face validity [2], the result of the author's earlier experiment indicated that SAGAT was producing false scores. Therefore, the goal of this research is to renew the test set-up, update the scenarios and evaluate whether these changes help in establishing a more coherent framework for SA-assessment when using both SART and SAGAT techniques for the assessment of interfaces which support supervision of automated control.

2   MEASURING SITUATION AWARENESS

Endsley defines Situation Awareness as the "perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future" [4]. This definition is well accepted within the research community [2]. However, ambiguity exists on how to measure SA. Two rating techniques are most popular: SAGAT and SART.

The Situation Awareness Global Assessment Technique (SAGAT) involves the administration of queries during 'freezes' in a simulation. The queries relate to probes and need to be tailored to represent 3 levels of SA, in line with Endsley's definition, i.e.: level 1 Perception, level 2 Comprehension and level 3 Projection. An example of a level 2 question is: "What vehicle's manoeuvre is currently (i.e.: during 'freeze') causing a dangerous situation?". Applying SAGAT requires intensive preparations. Nonetheless, the objectivity of this technique, while using predefined probes which are representative for the relevant elements to comprise Situation Awareness, is its main advantage.

The Situation Awareness Rating Technique (SART), on the other hand,

involves self-assessment of SA by participants based on standardized queries and is typically administrated post-trial [5]. The technique accounts for individual differences in attention and available cognitive resources to achieve SA: the standardized questions encompass three groups: (1) "Demand", referring to variability and complexity of a situation; (2) "Supply", referring to applied cognitive recourses and (3) "Understanding; referring to quantity and quality of understood information. After taking cumulative group scores, a total score is calculated according to SA-SART = U – (D – S).

Validation studies have only found moderate correlation between sub-scores of SA, i.e.: between SAGAT level 1 and overall SART [5] and between SAGAT level 1 and SART-Supply [2]. According to Salmon [2] no studies have reported significant correlation between overall scores of both methods, leading to the conclusion that the SAGAT and SART are actually assessing different aspects of SA. SAGAT, essentially measures the extent to which a participant is aware of pre-defined elements in the environment and their understanding of these elements. SART, on the other hand, provides a measure of how generally aware participant's perceive themselves to be without referring to specific elements within the environment. Several studies have shown significant correlation between overall SAGAT scores and overall performance whereas SART did not show this relation [2],[6]. Therefore SAGAT is regarded the more reliable technique for assessing SA. As explained in the introduction, a previous study of the author, showed contrary result. Due to the test set-up it was presumed that SAGAT produced false scores. Therefore we decided to compare again both techniques within a renewed set-up.

3   RELEVANT DRIVING SITUATIONS TO TEST SA

The presumed false scores within the previous study are most likely due to the test set-up, involving quite many relatively short trials with limited variation in the accompanying driving situations, making the experiment to be

like a reaction test based on impulsive reactions without much cognitive throughput. Moreover, the time duration between probe occurrence and probe taking seemed to have caused misunderstanding to what situation in time the probes were referring. Therefore we wanted to renew the situations. Based on systemboundaries we therefore defined six scenarios, which differed in hazardous and critical situations. The hazardous situations required attention, without direct necessity of intervention. A hazardous situation could develop into a critical situation which would require the driver to intervene. System boundaries for semi-automated driving depend on available technology (e.g. performance of sensors and algorithms) and on choices in system design (e.g. defining a boundary speed). Within this study the concept of congestion assistance is taken as a reference: the system operates only with a maximum speed of 50 km/h, if lines are being recognised, if a target vehicle is being recognised and if driving on a motorway without roadwork. In line with these system boundaries, we have defined three critical scenarios which involve accident avoidance. These scenarios are:

- **Emergency Brake** (EB) - While driving automatically, the target vehicle makes an emergency brake and comes too close, violating minimum distances. This causes the system to warn and requires the driver to take over control. Without intervention a collision would occur.

- **Merge Out** (MO) - While driving automatically, the target vehicle merges out to the left lane. As there is no new target vehicle on the own lane, the ego vehicle terminates automation and requires the driver to take over control. Without intervention the ego vehicle would drift out of lane with the danger to collide with neighbour vehicles.

- **Cut-in** (CI) - Just before an exit and while driving automatically, a vehicle from the left lane cuts in closely in an attempt to take the exit. With this manoeuvre the vehicle comes too close, violating minimum distances. This causes the system to warn and requires the driver to take over control. As the cut-in vehicle continues to brake, reluctance to intervene would lead to collision.

As we want to assess support drivers are provided with to execute their supervisory task, we also included three rudimentary interface-types which differed in their way to offer feedback. The characteristics of these feedback-types are: Type A provides only audible feedback. The system's detection of

an hazardous situation was announced by an alerting one-tone sound, while a critical situation used an alarming 3-tone sound (both exceeding the simulated engine and road roar with about 12 kHz). Type B provides in addition to the same audible feedback a simple textual feedback to indicate whether the audible warning is for a hazardous or critical event. Apart from the audible warning (which was again the same as for type A), type C also provide detailed visual feedback on system status, like successfulness of detecting a target vehicle. The belief was not that these types of feedback would be particularly good, but the intention was to serve as an input to have something to compare during measurements.

Fig. 1   Driving simulator used for the experiment

4   DRIVING SIMULATOR EXPERIMENT

4.1   Task and Simulator Environment

Participants were seated in a mocked-up vehicle, which was placed in a simulated motorway environment, as shown in figure 1. Every participant drove 6 test trials with different driving situations. Within each trial, participants drove automatically, but remained responsible for safe driving. Their main task was to supervise system operations and to intervene when required. As described in the previous section, an interface supported the drivers with their supervisory tasks, by either requesting extra attention (so

called 'soft warning') or requesting intervention (so called 'hard warning' for critical situations). In order to include realistic circumstances, participants had functionality at their disposal from a smartphone and were invited to read mails and review a calendar. As participants remained responsible for safe driving, they were advised to divide their attention appropriately. Judgement whether it would be necessary to intervene, was at the driver. Common automobile control interfaces, including a physical steering wheel and physical gas and brake pedals, allowed participants to take full control of the vehicle if necessary. Other vehicles drove in front and behind the simulated vehicle, as well as on the neighbouring lanes. All vehicles drove with time headways between 1 and 1,5s. at about 50km/h, as to simulate jammed traffic. Between experiments the position of the neighbouring vehicles was identical per situation to ensure that every participants got the same chance of resolving the situation.

4.2   Experimental Design

The independent variables for the experiment comprised of 'situation' and 'feedback'. 'Situation' was manipulated within subject: Each participant was confronted with three hazardous situations (which required extra attention) and three critical situations in which it was necessary to retrieve control (and avoid an accident). To make the situations non-predictable, the order between situations was arbitrary and also one condition was added in which no extra attention or take-over was required. 'Feedback' was manipulated between subjects and divided over the situations in order to have each feedback-type tested in every situation 8 times. The division of 'feedback' over the situations was randomized for each participant to avoid influence of carry-over effects. Shortly after a hazardous or critical situation occurred, the simulation was paused. Then, the screens were put blank and the experimenter subjected the participant to a SAGAT and SART questionnaire. The order of questionnaires was alternated between the trials. Each SAGAT questionnaire presented three questions based on probes tailored for the specific situation afore. An example is: what caused the system's request for extra attention? Depending on the situation, the correct answer would be "approaching end of motorway", "failure to detect roadlines", etc. After completing both questionnaires, a new trial started.

4.3 Participants and Procedure

24 persons were recruited and had at least one year of driving experience. Participants were either students or university personnel, their age ranged from 20 to 40 years old. Per participant the experiment lasted 1 hour with 15 minutes of instruction and training with the driving simulator and 6 times a 6-minutes trial. Per trial the automated driving lasted between 2,5 and 3 minutes until the simulation was paused to fill in the SA questionnaires. 3 Trials required take-over of control. The experiment was timed to ensure that simulation paused after the ability to retrieve control. The experimenter started each trial manually while the participant was directly driving automatically.

Table I: Comparison between SAGAT and SART scores per feedback-type and depended on situation

Critical situations	SAGAT scores			SART scores		
	Feedback-type			Feedback-type		
	A	B	C	A	B	C
1b; Emergency brake	1,86[1]	1,75	2,00	4,83[1]	5,10	*3,98*
2b; Merge-out	1,88	1,63	*1,38*	**5,59**	5,14	4,54
3b; Cut-in	2,00	**2,50**	2,25	4,50	3,79	4,58
Average all critical situations	1,91	**1,96**	*1,88*	**4,98**	4,68	*4,37*
Possible range low – high SA	0 ("low") to 3 ("high")			-5("low") to 13("high")		

[1]) based on n=7, all other conditions n=8

Note: **highest scores** are highlighted in bold and *lowest scores* with italic and underlined font.

5 RESULTS

Depended on situation, table I shows a comparison between overall SAGAT and SART scores per feedback-type. According to both SAGAT and SART, feedback-type C scores lowest on average over all situations. SAGAT and SART scores differ in indicating the feedback-type with highest scores. According to SAGAT, type B scores highest on average. The highest average SAGAT score of "1,96" for type B indicates that 5 out of 8 participants were able to perceive, understand and predict future states of any situation correctly with feedback-type B. Over all, situation 3b ("Cut-in")

with feedback type B enabled participants to gain highest Situation Awareness according to SAGAT. According to SART, type A scores highest on average. The minimum and maximum values were scored in different situations. This could be explained by the fact that SAGAT is an objective measure and SART a subjective measure, while differences between the critical situations are likely to cause SA perception in one situation to be comparatively lower or higher than in another situation. However, in this study the influence of situation on SA-scores is not included.

Table II: Comparison between subscores SAGAT-level 2 and subscores SART-U per feedback-type and depended on situation

Critical situations	SAGAT-level 2 scores			SART-U scores		
	Feedback-type			Feedback-type		
	A	B	C	A	B	C
1b; Emergency brake	0,86[1]	**1,00**	_0,63_	3,90[1]	4,29	_3,58_
2b; Merge-out	0,75	0,88	_0,63_	**5,21**	4,83	4,50
3b; Cut-in	0,75	0,88	0,75	4,08	4,54	4,46
Average all critical situations	0,79	**0,92**	_0,67_	4,40	**4,56**	_4,18_
Possible range low – high SA	0 ("low") to 1 ("high")			1 ("low") to 7 ("high")		

[1]) based on n=7, all other conditions n=8

Note: **highest scores** are highlighted in bold and _lowest scores_ with italic and underlined font.
Both SAGAT-level 2 and SART-U scores are referring to SA-level 2: Understanding.

Table II shows a comparison between the subscores SAGAT-level 2 and SART-U. This is important because both subscores refer to the second level of Situation Awareness, i.e. Understanding. With SART-U, participants were asked to give a self-assessment on (a) gained information, (b) quality of understood information and (c) familiarity with the situation. With SAGAT, probes were taken to measure whether the participant understood what aspect required attention in the situation, like approaching end of motorway, or a failure to detect road lines, etc. The results show that the subscores SAGAT-level 2 and SART-U succeed in indicating the same feedback-types with highest and lowest scores. According to both measurements, type B scores best. The highest SAGAT score of "0,92" for type B as average over all situations indicate that on average 7 out of 8 participants were able to understand any situation correctly with feedback type B. The perception of

correct understanding (based on SART) was relatively lower (score "4,56" in a range from 1 "low" to 7 "high"), but according to SART participants also perceived type B overall best.

6   CONCLUDING REMARKS

In comparison with the results from the earlier study [3], giving contrary outcomes of gained driver's SA based on SART and SAGAT scores, the results from this study are encouraging as SART and SAGAT do not show conflicting results. Based on the used SA-measurement techniques, the proposed test set-up seem to be successful in discriminating between the quality with which feedback-types support drivers in their supervisory task. Therefore, we carefully conclude that this renewed set-up does succeed in providing a coherent test-bed with relevant situations to assess the level of SA drivers gain when involved in supervision of automated control and when retrieving control is needed. However, when comparing the results it has to be noted that both for SAGAT and for SART most scores do not differ significantly between conditions. Hence, further assessment with regard to significance and variance between the scores is needed. Moreover, differences in SART-scores between the conditions are low, especially when we acknowledge that these scores could theoretically range between "-5" (low SA) to "13" (high SA) with a median of "4". Our testscores only ranged from "3,98" to "5,59". Maybe this is due to the variety of questions involved in the SART questionnaire. Besides from 'Understanding of the situation', these questions also refer to 'Supply of cognitive resources' and 'Demand of the situation'. It could be that the amount and variety of the questions work as a 'damper' on the scores. Furthermore, it is interesting to mention that it is against expectations that feedback-type C scored worst, while C offers the most 'rich' feedback with both audible and visual information and was therewith expected to offer more support in understanding the circumstances that caused a critical intervention. An explanation for this unexpected result could be that the extra information caused participants to be distracted and therefore less concentrated on the actual traffic situation outside the vehicle.

Concluding that the division of lowest and highest scores were not according to expectations, underlines the necessity to further develop appropriate interfaces for supervisory control of automated driving and underlines the importance of thoroughly testing interfaces in representative situations before making decisions on implementation. For the latter, the results of this research give an important contribution, while providing solutions for assessment of involved levels of Situation Awareness.

References

[1] Martens, M., Beukel A.P. van den. The road to automated driving: dual mode and human factors considerations. In: Proceedings of the 16th International IEEE Annual Conference on Intelligent Transportation Systems (ITSC 2013), The Hague, pp. 2262-2267, 2013

[2] Salmon, P. M., Stanton, N. A., Walker, G. H., Jenkins, D., Ladva, D., Rafferty, L., & Young, M. (2009). Measuring Situation Awareness in complex systems: Comparison of measures study. International Journal of Industrial Ergonomics, 39(3), 490-500.

[3] Beukel A.P. van den, Voort M.C. van der. The Influence of Time-criticality on Situation Awareness when Retrieving Human Control after Automated Driving. In: Proceedings of the 16th International IEEE Annual Conference on Intelligent Transportation Systems (ITSC 2013), The Hague, pp. 2000-2005, 2013

[4] Endsley, M. R., Sollenberger, R., & Stein, E. (2000). Situation awareness: A comparison of measures. Proceedings of the Human Performance, Situation Awareness and Automation: User-Centered Design for the New Millennium, Savannah, GA.

[5] Charlton, S.G., Measurement of cognitive states in test and evaluation. In: Handbook of Human Factors Testing and Evaluation, London, 2002, pp. 115-122

[6] Jones, D. G., & Endsley, M. R. (2004). Use of real-time probes for measuring situation awareness. International Journal of Aviation Psychology, 14(4), 343-367.