Adaptive multi-modal interface model concerning mental workload in take-over request during semi-autonomous driving

Weiya Chen, Tetsuo Sawaragi & Toshihiro Hiraoka

Published online: 11 Mar 2021.

Submit your article to this journal

Article views: 1112

View related articles

View Crossmark data

SICE

Taylor & Francis
Taylor & Francis Group

Adaptive multi-modal interface model concerning mental workload in take-over request during semi-autonomous driving

Weiya Chen[a], Tetsuo Sawaragi[a] and Toshihiro Hiraoka[b]

[a]Graduate School of Engineering, Kyoto University, Kyoto, Japan; [b]Institute of Industrial Science, The University of Tokyo, Tokyo, Japan

ABSTRACT

With the development of automated driving technologies, human factors involved in automated driving are gaining increasing attention for a balanced implementation of the convenience brought by the technology and safety risk in commercial vehicle models. One influential human factor is mental workload. In the take-over request (TOR) from autonomous to manual driving at level 3 of International Society of Automotive Engineers' (SAE) Levels of Driving Automation, the time window for the driver to have full comprehension of the driving environment is extremely short, which means the driver is under high mental workload. To support the driver during a TOR, we propose an adaptive multi-modal interface model concerning mental workload. In this study, we evaluated the reliability of only part of the proposed model in a driving-simulator experiment as well as using the experimental data from a previous study.

1. Introduction

With the development of automated driving technologies, human factors involved in automated driving are gaining increasing attention for a balanced implementation of convenience brought by the technology and safety risk in commercial vehicle models. One influential human factor is mental workload (MWL), which is defined as the proportion of a human operator's mental capabilities occupied when performing a given task [1]. The driver's MWL is crucial for driving safety; if the driver is under too high an MWL, he/she cannot execute proper action in a timely manner. During a take-over request (TOR) from autonomous to manual driving at level 3 of International Society of Automotive Engineers' (SAE) Levels of Driving Automation in particular, the time window for the driver to have full comprehension of the driving environment is extremely short. To support the driver during a TOR, we propose an adaptive multi-modal interface model concerning MWL. In the proposed model, the MWL is affected by different cognitive channels: visual, spatial, and verbal. We evaluated the reliability of only part of the model in these cognitive channels concerning a driver's MWL in a driving-simulator experiment. Physiological data (pupil diameter), secondary-task performance, and the NASA Task Load Index were used for MWL measurement. The results indicate that during manual driving, the driver's MWL in the visual cognitive channel is the highest, followed by that in the spatial cognitive channel then that in the verbal cognitive channel.

Besides the experiment with visual stimuli [2] to study the relationship among cognitive channels of visual modality, the unused data from Author's former experiment [3] is used to evaluate the reliability of the proposed model.

2. Background

2.1. Human-automation interaction in semi-autonomous vehicles

There are many categories definitions of vehicle automation, among which SAE Levels of Driving Automation is widely recognized and accepted (SAE J3016).

How the SAE levels are distinguished is based on the allocation of the dynamic driving task between the automation system and human driver. If a human driver performs all the driving tasks, the vehicle is said to be at SAE level 0; if the autonomous driving system is in charge of all driving tasks, the vehicle is said to be at SAE level 5.

Different from SAE level 5, levels 0 to 2 (SAE J3016) require human drivers to monitor the driving environment keep them in the loop. When the automation system cannot handle a particular situation, control will be returned to the driver. The human driver needs to comprehend the situation quickly and appropriately to avoid potential risks during take-over. After take-over, situation awareness (SA) requires extra MWL of the driver to perceive his/her surroundings correctly and

comprehend their meanings by projecting them onto his/her status.

This requirement is difficult for a driver to meet when he/she over-trusts the automation system. If automation replaces human drivers in some function blocks, it is unavoidable for drivers to depend on the system and lose their vigilance on monitoring the environment. Even if the driver could react to requests from the autonomous driving system, their SA will be challenged.

In other words, a driver is under higher MWL to figure out what is going on around him/her in the current traffic situation. High MWL can result in bad performance during take-over, which may lead to fatal consequences.

Compared to the MWL decreased by automatic accelerating or braking, such a situation requires extra MWL from a driver. In addition, if the driver is involved in other tasks, such as texting when a Take-Over Request (TOR) occurs, it will increase the MWL compared when the driver is always monitoring the environment. Whether different secondary tasks affect the driver's MWL differently is another crucial issue that needs to be addressed for a better design of a human-machine interface.

2.2. Interface design issue

Before SAE level 5 is achieved, the driver is responsible to react in a reasonable time to serious situations. However, the driver also obtains the option to perform non-driving-related tasks (NDRTs), which means there would be various occasions before an emergency occurs. Before a TOR, the driver is engaged in NDRTs or is monitoring the driving environment; therefore, the proper form of warning information should be provided according to the concurrent state.

2.3. Research motivation

This study's motivation is to find the general solution for the issue mentioned above in Section 2.2; an adaptive multimodal interface for efficient TOR is regarded as a promising alternative. To develop an adaptive interface for efficient TOR, the driver must be continuously monitored by the automation system. One crucial factor for efficient TOR is the MWL. When the driver is under too high an MWL, he/she may not be able to respond on time. Therefore, MWL can be a decent indicator of TOR performance. A qualitative model is helpful to predict an MWL. Based on MWL allocation, which is deduced from such a qualitative model, the remaining accessible mental resources can be distributed to suitable cognitive channels for efficient TOR with lower MWL.

3. Theoretical basis

The introductions of terms (TOR and Mental Workload) in following subsections are beneficial to understand the research objective in detail and MWL measurement methods are described in Section 3.3 as the theoretical basis for MWL measurement in the experiment of this study.

3.1. Take-over request (TOR)

We divide a TOR into a series of phases: when the driving automation system of the vehicle detects a forthcoming difficult situation (Trigger); the vehicle system informs the driver of potential danger (TOR); the driver takes control (Take-Over); and the vehicle becomes stable and safe after the driver has sufficient SA and properly operates the vehicle. The TOR process is illustrated as a time sequence in Figure 1. For example, at SAE Level 3, in the first phase from "Trigger" to "TOR," the driver may be engaged in NDRTs. The MWL of the driver is only related to the NDRT type. In the second phase, after the TOR from the automated system, the driver's SA will drastically change, and he/she will be more focused on the emergency. The reaction time of the driver is defined between the TOR and take-over action (hands-on or pedal press). This reaction time is influenced by the form of the alert information. How sensitive the driver is to the alert modality corresponds to how long until the driver notices the alert. At this phase, the driver's MWL is affected by the NDRT and the TOR-signal presented modality. In the third phase, the driver's SA-gaining phase starts at the same time as the second phase, but it may end before or after the second phase. The period of SA completion depends on the alert-information content, i.e. how much of the necessary points to describe the driving environment the
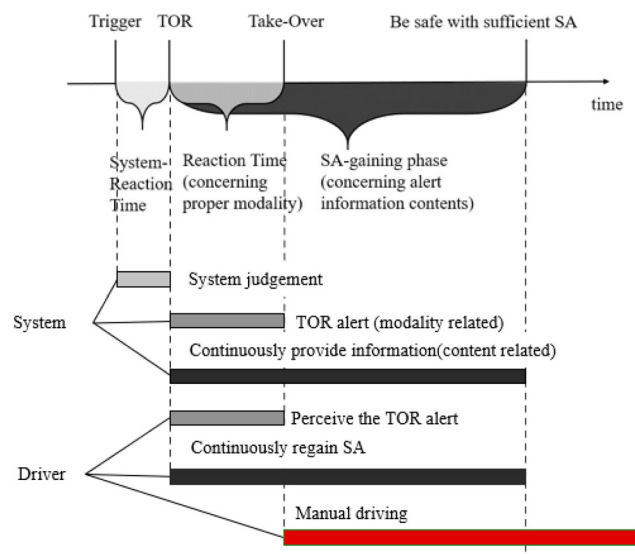

Figure 1. TOR process as time sequence.

alert information includes. The driver may take different strategies during a TOR; some drivers tend to take over after totally comprehending the situation. In contrast, others may first follow the TOR instruction and try to figure out the situation afterward. Traffic situation as an objective factor and the alert-information content are the primary influencing factors of a driver's MWL during the third phase. The more complicated the traffic scenario, or the less efficient the alert describing the situation, the larger the MWL put onto the driver.

This is clearer when this process is divided into two parties: System and Driver. The time from Trigger to TOR is required for System to make proper judgement; from TOR to take-over, System should select the proper modality to alert the driver through a multi-modal interface and the final part is providing continuous information to support the driver to regain his/her SA as soon as possible. Before TOR, Driver is engaged in certain NDRTs and does not monitor the driving situation. When the TOR is presented, the driver starts to understand the alert and the concurrent driving situation simultaneously. After the driver takes action, which is the take-over, the alert-perception period ends and the manual-driving periods starts. As shown in the lower part of Figure 1, the continuous-SA-regaining phase and manual-driving phase overlap, which indicates that Driver is multi-tasking in which he/she relies on the information content provided by the multi-modal interface to shorten the SA-regaining phase.

The driver's MWL changes throughout the TOR process, and the TOR should be completed in a short-bounded time interval. The interface of informing the driver of a TOR must be carefully designed so that it can adapt to the driver's changing MWL. This adjustability makes an HMI be more efficient and safer.

3.2. Mental workload

Human MWL is an important design concept for exploring the interactions between people and technological devices [4]. The principal reason for MWL assessment is to quantify the cognitive cost associated with performing a task for predicting operator or system performance. However, it has been extensively reported that mental underload and overload can negatively influence performance. On one hand, during information processing, when MWL is at a low level, individuals may frequently feel frustrated or annoyed. On the other hand, high MWL can lead individuals to confusion, decrease their performance in processing information, and increase the chances of mistakes.

3.3. Mental-workload measurement methods

There are several methods of measuring MWL: primary-task measurement, secondary-task measurement, and subjective evaluation. Physiological indicators of MWL have also recently been identified.

Primary task measurement is a task-oriented method for computing the MWL imposed by the task through the operator's performance. Its disadvantage is that the result depends on the type of task, which makes it difficult to compare two different tasks. Furthermore, the performance-measurement indicators cannot directly determine MWL.

Secondary-task measurement is one of the most widely used methods for accessing operator MWL. An operator needs to perform the primary task under specific conditions and use spare resources and capacity to complete the required secondary tasks. How well the operator performs the secondary tasks is an indicator of how much MWL the operator is under at that time point. Secondary-task measurement has several advantages. First, it is a sensitive measure of operator capacity and distinguishes between alternative equipment configurations that are indistinguishable with primary-task measurement. Second, it provides a sensitive index of task impairment due to stress. Third, it provides a common metric for comparing various tasks. However, this method has one major disadvantage: intrusion on the performance of the primary task. Furthermore, operators can use different strategies when performing secondary tasks.

Subjective evaluation is by far the easiest and most popular method of measuring MWL [5]. Operators are required to determine and report on the MWL put onto them when performing secondary tasks. This method is more practical since it is easy to implement without instrumentation and has higher sensitivity. The NASA Task Load Index (NASA-TLX) and Subjective Workload Assessment Technique (SWAT) are commonly used subjective rating scales. Mental fatigue is also a problem that may occur. There are several advantages of subjective evaluation, i.e. they are inexpensive, unobtrusive, easily administered, and readily transferable to large-scale vehicles and to a wide range of tasks. However, there are also several limitations according to O'Donnell and Eggemeier [6], i.e. potential confounding of mental and physical workload; difficulty in distinguishing external demand or task difficulty from actual workload; unconscious processing of information that the operator cannot rate subjectively; dissociation of subjective ratings and task performance; requirement of well-defined questions; and dependence on short-term memory.

With physiological indicators, it is assumed that MWL can be measured by means of physiological changes [7]. Indicators such as heart rate and P300 wave (an event-related potential (ERP) component elicited in the process of decision making) were found to correlate with the degree of novelty of the information people received [8]. The advantage of physiological indicators is their continuity throughout an experiment and

they do not affect the performance of primary tasks. However, special equipment is necessary, and these physiological indicators are sensitive to other irrelevant factors.

4. Research target

4.1. Previous research

The research conducted by Chen et al. in 2018 [3] confirmed the importance of keeping the driver in the loop through a dual-task driving simulation. They measured the MWL of a driver when he/she is requested to perform a variety of secondary tasks at three partial-automated-driving levels. They found that the driver's MWL decreases from level 0 to level 1, which proves that one-dimensional automated control is beneficial for a driver's quick reaction. However, the driver's MWL increases unexpectedly from level 1 to level 2. Although the MWL at level 2 is not as high as that at level 0, these experimental results indicate that over-trust occurs at level 2. Because over-trust occurs at level 2, the driver does not constantly monitor the driving environment, reducing his/her SA. In that experiment, when potential emergencies occurred, e.g. a pedestrian crossing the road, the driver was under higher MWL than usual because he/she had to regain her/his SA to know what was happening.

Another conclusion from Chen et al.'s study is that secondary tasks involving motor operation put a higher MWL onto the driver, and secondary tasks without any working memory requirement puts less MWL onto the driver.

However, the influence factor of various secondary tasks on MWL has not been verified. The main conclusion is only about the total MWL concerning different autonomous driving levels; secondary task type has not been discussed in the previous paper. Although various secondary tasks caused differently MWL on the driver in the simulation, the qualitative relation has not been found out. The MWL allocations for various secondary tasks have not been identified, which is essential for the total MWL projection.

4.2. Research purpose

To develop an efficient interface for autonomous driving, especially when handling take-over, MWL allocation for driving is the first issue to be addressed. Usually, before the driver has sufficient SA, she/he will first choose to take action and regain SA after control is pass to her/him (e.g. "Take-Over" before "be safe with sufficient SA" in Figure 1). In this case, manual driving, i.e. SAE level 0, starts before the end of the SA-gaining phase, which means the SA-regaining phase will overlap with the manual-driving process. During this period, a driver has to tackle the MWL needed to drive manually

as well as that incurred to make a smooth transfer from autonomous to manual driving. To discuss the MWL during this safe time buffer more accurately, we have to analyze how MWL is incurred for fully manual driving assumed to follow the SA-gaining phase. The experiment we conducted attempted to determine the MWL allocation for the driver, namely at SAE level 0.

The results of the experiment are the fundamental part of the whole proposed multi-modal interface model. The proper adjustment of TOR presentation is possible based on the driver's concurrent state (e.g. engaged in NDRTs) and the MWL allocation for manual driving.

This paper only concerns the second phase of a TOR, namely which modality should the interface select for the TOR. Note that content of alert information is not considered.

5. Proposed model for multi-modal interface

5.1. Revising Wicken's multiple resource theory

Multiple Resource Theory (MRT) proposed by Wickens [9] has a practical application in predicting the interference level of concurrently performed tasks, i.e. the four-dimensional multi-resource model. MRT states that the conflict between two time-sharing completed tasks will be higher when the tasks share stages (perceptual/cognitive vs. response), sensory modalities (auditory vs. visual), codes (visual vs. spatial), and channels of visual information (focal vs. ambient).

However, MRT is not suitable for underloading situations because task interference is less relevant when the total MWL is low. Regarding the take-over issue, in which mental overload circumstances potentially may occur, MRT can predict performance breakdowns. In the perceptual-modalities dimension, only two levels are considered: visual and auditory. There should also be a third perceptual modality: tactile. A haptic interface is promising to enhance the operator's performance of dealing with unexpected changes in a complicated scenario in which large amounts of data exist [10,11]. In MRT, stage perception and stage cognition are not separated. This combination may need further discussion. Perception is concerned with sensory stimuli reception and interpretation, while cognition is more concerned with understanding the information gathered through the understanding by requiring mental resources, knowledge, and experience. These two stages involve different mental resources, which should lead to necessitating an unusual amount of MWL.

According to the above problems with MRT, our proposed model extends MRT for the autonomous driving. This model, illustrated in Figure 2, separates the perception and cognition stages and adds the tactile perceptual modality. One more processing code channel is added as well: visual. Our experiment
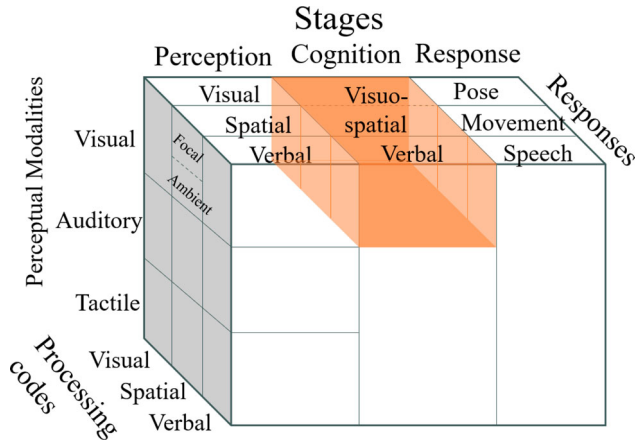
Figure 2. Three-dimensional representation of proposed model's structure.

Figure 3. Graphical interpretation of hypothesis.

only focused on the part of the model highlighted in Figure 2. In the cognition stage, working memory is much more involved. Based on the multicomponent workspace of working memory in working memory theory [12], working memory has temporary limited-capacity memory systems for verbal (articulatory loop) and non-verbal (visuospatial scratchpad) material that connects with the content from long-term memory. However, according to the research conducted by Sanada [13] in 2015, shape and spatial working-memory capacities are mostly independent. Therefore, we needed to test whether visual and spatial channels in the cognition stage should be separated as well.

5.2. Conflicts in cognitive channel

Total manual driving requires a different amount of mental resources in different channels, which means that two simultaneously performed tasks will incur different amounts of MWL in each channel. In autonomous driving, the main task is driving, and different types of secondary tasks will influence the total MWL differently. Depending on which channel the interference occurs in, the amount of MWL will also differ. For example, a visual stimulus requires visual, spatial, or verbal cognition or any combination of these three channels if another stimulus is in audio form; the interference will not be as high as two visual stimuli occurring at the same time, as shown in Figure 3. The concrete example will be comparing driving while watching the news on a tablet and driving while listening to the radio news. The latter should be easier for the driver to stay safe and keep up with the news content. Furthermore, if two visual stimuli simultaneously occur but require different cognitive channels, for example, one for a visuospatial resource and another for a verbal resource, it will incur less MWL than two stimuli both occupying the visual cognitive channel. The concrete example will be the comparison between watching two pieces of news both in video form on
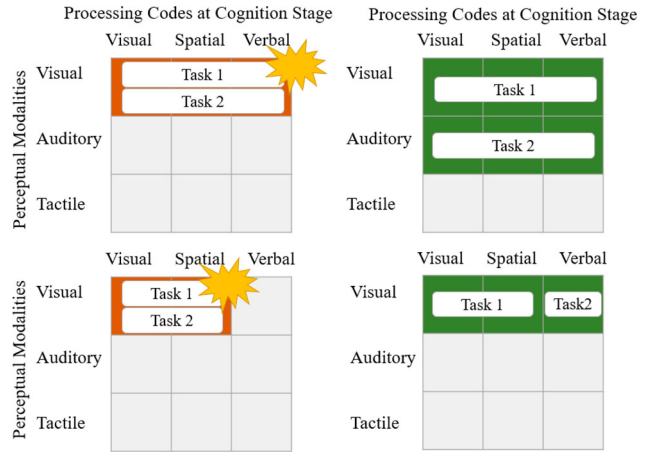
one screen and watching one piece of news presented in the video while another presented in scrolling text under the video. The latter should cost the audience less effort to grasp the content of both news. The MWL put onto the driver during TOR should be minimized, but not too low in the underload zone, for safe transition from autonomous driving to manual driving. For example, when the driver is texting using a smartphone at SAE level 3, where he/she does not expend any mental resources on the primary driving task, his/her MWL is mainly located in visual perception, visual and verbal cognition, and movement response. If the TOR comes from the autonomous driving system at this moment, it should use the alert information located in the unoccupied auditory or tactile perception, in the spatial cognition stage rather than use the visual perceptual modality or required visual/verbal cognition information.

6. Hypothesis on mental workload allocation of driver during TOR

If combining the above-mentioned conflict principle with the TOR time phases, we can determine the driver's MWL allocation as a time sequence throughout the TOR process. Continuous support from the automation system is necessary, and especially beneficial for the third time phase (from "TOR" to "be safe with sufficient SA"). For example, at SAE level 2, the driver is required to monitor the driving environment thus, is only allowed to make a phone call through wireless equipment, such as Bluetooth earphones, without any hand operation. At SAE level 3, however, the driver is allowed to perform any secondary task, so she/he can operate a phone by hand. The MWL allocation is illustrated as a time sequence (only concerning cognition stage) in Figure 4 to distinguish the TOR process when the driver is making a phone call at SAE levels 2 and 3.

The driver's MWL allocations at the cognition stage before TOR are shown in Figure 4. The horizontal axis represents three different cognitive channels: visual,
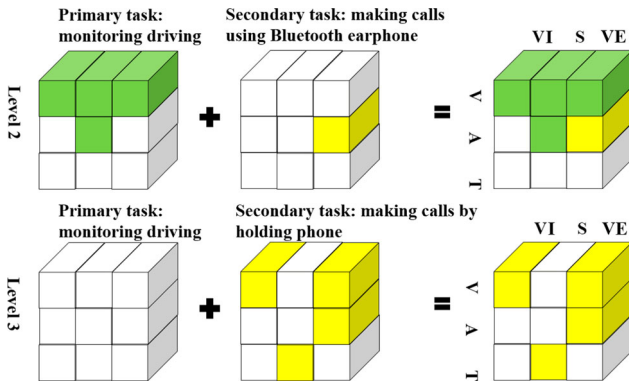
Figure 4. Graphical interpretation of MWL at SAE levels 2 and 3.

spatial, and verbal; the vertical axis represents three perceptual modalities: visual, auditory, and tactile. At SAE level 2, the driver is still required to monitor the driving situation. The driver needs to gather information concerning visual cognition (e.g. colour of the traffic light), spatial cognition (e.g. other vehicles' locations), and verbal cognition (e.g. words on traffic signs) using visual perception. The sounds from other entities can also be used to judge their location (e.g. car horns). Therefore, the driver's MWL for monitoring is located in the visual-visual, visual-spatial, visual-verbal, and auditory-spatial cognition channels (four green cubes in Figure 4). When the driver is making a phone call using Bluetooth earphones, only the auditory-verbal cognition channel is occupied because the driver only communicates through voice and receives auditory input from the smartphone. Because the driver performs both tasks simultaneously, when we combine the MWL allocations for the primary monitoring and secondary calling tasks, we can determine the total MWL allocation of the driver at SAE level 2 for this specific occasion. The same principle can also be applied to SAE level 3, i.e. the driver has no monitoring task, so there is no cognitive channel being occupied. For the secondary task of making a phone call through hand operation, the mental resources in the four cognition channels are used: visual-visual (e.g. shape and colour on smartphone screen), visual-verbal (e.g. words on screen), auditory-verbal (e.g. sound response from the smartphone) and tactile-spatial (e.g. typing). Because there is no monitoring task required, the driver's MWL before a TOR depends only on the secondary task in which the driver is engaged.

As previous research results indicated [3], the driver's MWL at SAE level 1 is higher than that at SAE level 2, and that at level 0 is the highest among these three levels. The MWL at SAE levels 0 and 1 can be decomposed in the same manner as at SAE level 2, as shown in Figure 5. The difference is the amount of mental resources at each cognitive channel rather than the location of occupied cognitive channels. At SAE level 3, however, the total MWL is equal to that caused by secondary tasks because the driver does not need

to perform the monitoring task. Once we determine the driver's MWL allocation, the unoccupied cognitive channels can be potential candidates for the TOR. A tactile TOR is promising at SAE level 2. The driver's MWL allocation of the primary driving and secondary tasks will be kept for a short time during the driver's reaction period, so the MWL allocation is the same as in Figure 4. The TOR will also consume extra mental resources according to its presented modality. In this case, the tactile-spatial cognitive channel will be occupied, for example, the vibration of the steering wheel or seat belt to inform suggestive turning direction. This is shown in Figure 5 in red. At the same time, the driver starts regaining her/his SA. The mental allocation influenced by the TOR content will not be further discussed, just use the broad common way, by continuously supporting the driver by using visual and auditory information. The MWL allocated for SA is in orange in Figure 5. One crucial criterion is that the MWL allocation for alert-information content should be decided according to the allocation of the phase before, namely, the allocation of "reaction time ." The content should be mainly presented in the form in which no mental resources are consumed in the "reaction time" phase, which is in red in this specific case: "auditory-visual"
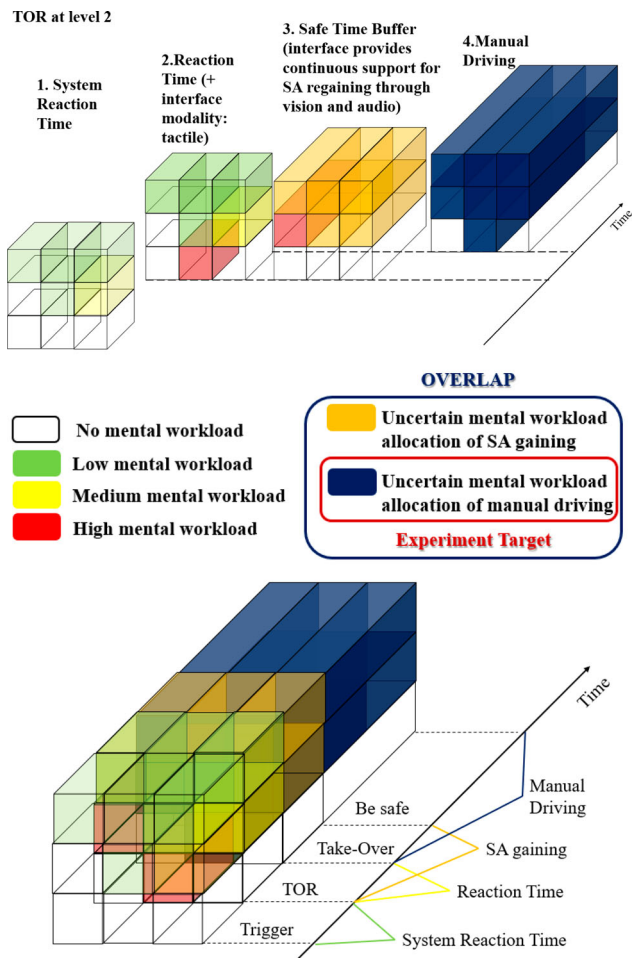

Figure 5. Driver's MWL allocation at SAE Level-2 during TOR.

(e.g. auditory description of the driving environment). By using the cognitive channel without any overlap with "reaction time," the MWL allocation is optimized for a shorter time for completing the TOR process. After the driver obtains sufficient SA, she/he finishes the TOR process and starts manual driving. At this time, we assume that the mental resources in seven cognitive channels will be used, including all visual and auditory perceptual modalities and tactile-spatial channel (operating the steering wheel). They are in dark blue in Figure 5. There should be conflicts during this transient period, but the MWL allocation for manual driving is not clear, which is the main research target of this paper.

After we assemble the four time phases of MWL allocation (75% transparent colour showed MWL allocation for the driver's state before TOR; 50% transparent colour showed MWL allocation between TOR and driver's response, 50% transparent colour showed MWL allocation for driver regaining SA with the continued support from the automation system, and 15% transparent colour showed MWL allocation for manual driving), the driver's MWL allocation can be accessed, as shown in the lower part of Figure 5.

Figure 5 shows the dynamic change in the driver's MWL, by tracking its change. The interface can allow adjustments for optimal MWL allocation, i.e. different cognitive-channel are used for alert-information presentation at SAE levels 2 in above-mentioned case. The less conflict there is in every channel, the less MWL is put onto the driver; therefore, the driver can finish the entire TOR process in less time.

7. Experiment

7.1. Experimental purpose

As discussed in Section 6, to design the adaptive human-machine interface for TOR, the MWL allocation for SA gaining and MWL allocation for manual driving are the two problems to be solved. We should discuss these based on the SA regaining phase overlap and manual driving phase (shown in Figure 5). In this paper, we will study only the MWL allocation for manual driving. Section 5.1 suggests separating the visual and spatial cognitive channels in the proposed model, which objects to the traditional visuospatial scratchpad theory. Therefore, we also need to ascertain this separation in the experiment.

In summary, there were two purposes for this experiment: to determine whether the visual and spatial cognitive channels should be separated and determine the MWL allocation for manual driving.

7.2. Experimental setup

The experiment we conducted involved using the driving simulation software PreScan on workstations. With
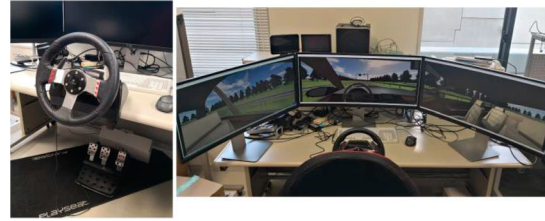

Figure 6. Logitech G29 Driving Force Racing Wheel and driving-simulation environment.

Logitech G29 Driving Force Racing Wheel (Figure 6, left) as the external device, a participant drove in the human-in-the-loop case in this experiment.

Three high-end Dell 34-inch digital monitors (U3415W) were used to enable a surround visual effect. As shown on the right of Figure 6, the participants faced the middle monitor during the driving simulation.

We measured the driver's MWL concerning three cognitive channels: visual, spatial, and verbal, all with visual stimuli. Participants were requested to perform three secondary tasks only involving one cognitive channel during a driving simulation.

7.3. Mental-workload measurement

We used NASA-TLX and an eye tracker to measure MWL. The pupillary response, blink, and eye movement (fixation and saccade) have been proved to be three effective indicators for MWL [14]. Among these, pupil size has the strongest positive correlation with a human MWL. Larger pupil size reflects a higher MWL.

7.4. Choosing secondary tasks

Three secondary tasks were chosen for the driving simulation, i.e. visual task including visual cognition resources (form and colour), spatial task involving spatial cognition resources (spatial and movement information), and verbal task only including verbal cognition resources. All tasks were presented through visual media. We chose a no-nameable-item sequence memory task, location-sequence in 3×3 grid memory task, and number-sequence-memory tasks (refer to Figure 7).

Three tasks should be kept at a similar difficulty level; therefore, a preliminary experiment was necessary to measure the memory span for these three secondary
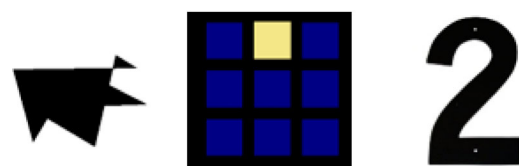

Figure 7. Examples of visual, spatial, and verbal secondary tasks.

tasks. Nine sequences were prepared: nine different non-namable shapes for the visual task, nine locations at a 3 × 3 grid for the spatial task, and one to nine verbal tasks. Ten participants from Kyoto University (eight males and two females aged from 21 to 27. Nine had valid driver's licenses) took part in this preliminary experiment, the average span for these three tasks were 2.9 spans, 7.2 spans, and 8.6 spans correspondingly. We applied different load sizes to equalize the general difficulty of the three secondary tasks: the visual task had three spans, spatial task had seven spans, and verbal task had nine spans. The three secondary tasks were presented using the open source software PEBL [15].

7.5. Participants

Twenty-five participants from Kyoto University (19 males and 6 females aged from 21 to 28. Twenty-two had valid licenses) took part in the formal experiment. Ten of them participated in the experiment in 2019, and the rest 15 participated in the experiment in 2020.

7.6. Experimental procedure

The participants consented to and signed a participation consent form. After a brief introduction to the equipment and explanation of each secondary task, each participant wore an eye tracker (Pupil Core from Pupil Labs for experiment in 2019 and Tobii Pro Glasses 2 for experiment in 2020). When a participant become familiar with the simulator after 5-min driving simulation, e.g. how fast the brake functioned or how sensitive the steering wheel was, the formal experiment was carried out. The participants were required to perform the visual, spatial, and verbal tasks while in fully manual driving mode. The influence of task order is eliminated using counter balancing method. The task procedure is illustrated in Figure 8. Each stimulus was shown on an iPad for 1 s, after all items were presented, an answer sheet with all nine possible items was shown,
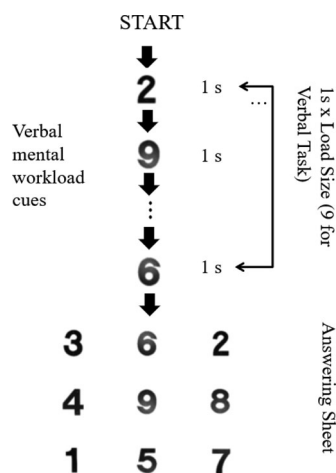

Figure 9. Driving-simulation scenario.

the participants needed to point out the stimuli in the presented order. A video camera was used to record the entire driving simulation. After the experiment, the participants filled out a NASA-TLX questionnaire for each secondary task using the NASA-TLX App on the iPad. The automatic answer-collection system was still under construction during this experiment, so a physical answer sheet was presented to the participants, as shown in Figure 9. The participants were required to point out the order of stimuli they remembered without any oral response.

8. Results

The data collected from the eye tracker were combined with the NASA-TLX scores and secondary-task performance to see if there was a significant main effect of cognitive channels. Unfortunately, due to the low accuracy of the eye tracker, only 2 out of 10 participants' datasets could be regarded as valid with stable confidence over 0.6. Therefore, the pupil-diameter data for the first 10 participants in the 2019 experiment are not included in the pupil-diameter analysis. We measured the pupil diameter of simulated driving without any secondary task as the baseline and used the pupil diameter change as the indicator. The one-way repeated-measures analysis of variance (ANOVA) was applied to the NASA-TLX scores and secondary-task performance. The NASA-TLX data were simplified using the AWWL method [16]. The mean number of correct responses [17] was used to determine the correct response rate of secondary-task performance. The ANOVA results (cognitive channels as an independent variable) are listed in Table 1. In Table 1, Psi is an effect size estimator for multiple comparisons, namely root-mean-square standardized effect. The intervals are confident intervals.

Figure 8. Schematic diagram for verbal task.

Table 1. Results from one-way repeated-measures ANOVA ($\alpha = 0.05$).

	NASA-TLX	Correct response rate	Pupil diameter change
Visual/spatial/verbal	$p = 0.0013$	$p = 2.28e{-}04$	$p = 0.1098$
	$\Psi = -7.82$	$\Psi = 0.099$	$\Psi = -0.1143$
	([$-17.271.62$])	([$-0.020.22$])	([$-0.430.21$])

There was a significant main effect of cognitive channels with respect to the secondary-task-performance data concerning NASA-TLX score and correct response rate. Therefore, we can verify the significant differences between the visual and spatial channels and spatial and verbal channels. Accordingly, the visual and spatial channels in the proposed model should be separated into two independent channels instead of merging into one visuospatial channel.

For quantitative comparison the three cognitive channels, NASA-TLX scores and correct response rates of secondary-task are presented in Figure 10.

The means and standard deviations of these two indicators are listed in Table 2. The visual task had the highest score, followed by the spatial task, and the verbal task had the lowest score. The higher MWL is, the higher the NASA-TLX score and the lower the correct rate should be. As shown in Figure 10, the mean correct rates of secondary tasks shared the same tendency with
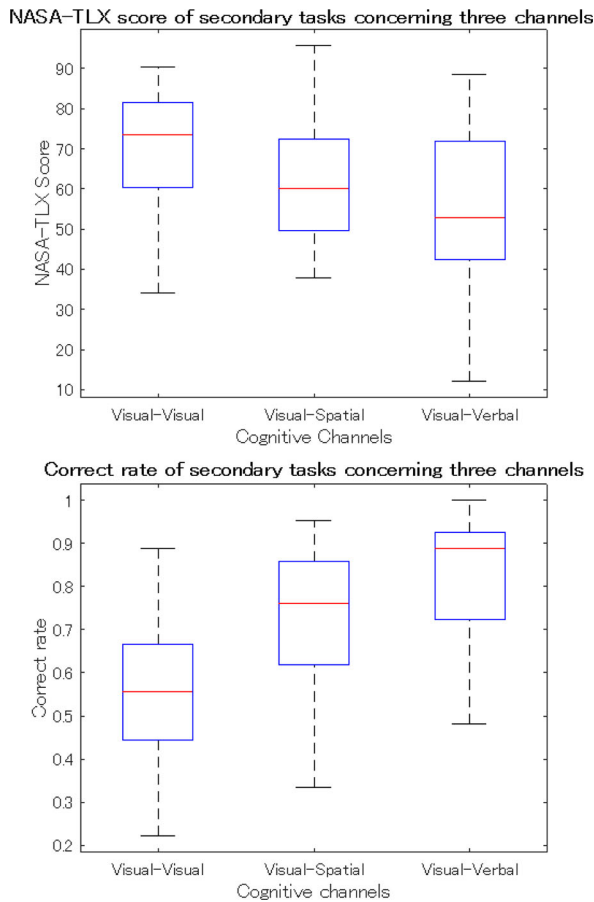
the NASA-TLX scores, MWL induced by the verbal task was the lowest while MWL caused by visual task was the highest.

Because the pupil diameter has a big individual difference, we used the pupil diameter change as the indicator. The baseline is the mean pupil diameter when the participant only performed the driving simulation without any secondary tasks. Although we use the Tobii Pro Glasses 2 to measure the pupil diameter, unfortunately, the pupil-diameter change did not show a significant effect of cognitive channels (in Table 1). Therefore the discussion of the results will not consider the data from pupil diameter change.

9. Discussion

9.1. Mental-workload allocation of manual driving

According to 15 participants pupil-diameter change, 3 cognitive channels' significant effect has not been observed. Thus, we cannot trust this physiological indicator too much. Therefore, we should conclude without the consideration of pupil-diameter data.

Based on the results obtained from the driving-simulation experiment with 25 participants, the overall result was that the visual task put the highest MWL onto the participants, followed by the spatial task. The verbal task increased the participants' MWL the least. As the conflict principle explained in detail in Section 5.2, the increase in MWL indicated conflict with the primary driving task at a specific cognitive channel. We can conclude that visual cognition is mostly required during driving, and the verbal cognitive channel is where conflict may occur with the lowest possibility. Other than this expected conclusion, the separation between visual cognition and spatial cognition could be verified, which leads to the determination of separating visual channel and spatial channel in the proposed model. This fact objects to the visuospatial scratchpad concept in working memory theory.

The reason for the visual cognitive channel being most occupied by manual driving is obvious; the entire driving task requires constant visual signal input. Most of the information gathered through the visual perceptual modality goes to visual cognition, which is in charge of shape and colour. Driving-environment monitoring is the primary consumer of mental resources, which involves visual and spatial cognition. The judgment of location and relative distance with other vehicles or infrastructure uses spatial cognition resources. Verbal cognition is only connected to traffic-sign reading and dashboard monitoring during visual input; the proportion of verbal-resource usage time is generally quite low compared with the other two channels. Therefore, the verbal-cognitive secondary task put the least extra MWL onto the participants. We only focused on

Figure 10. NASA-TLX score and secondary-task performance boxplots.

Table 2. Mean and standard deviation of NASA-TLX scores and secondary tasks' correct response rates.

	Visual	Spatial	Verbal
NASA-TLX	$\mu = 71.58$	$\mu = 61.38$	$\mu = 53.56$
	$\sigma = 13.93$	$\sigma = 15.00$	$\sigma = 20.57$
Correct response rate	$\mu = 0.56$	$\mu = 0.73$	$\mu = 0.83$
	$\sigma = 0.18$	$\sigma = 0.17$	$\sigma = 0.13$

visual perceptual input; thus, the MWL of auditory and tactile perceptual modalities may vary, which we can obtain only from an experiment on auditory/tactile-input secondary tasks.

The higher the MWL for one channel, the more resources the driver will require during driving. From the quantitative results, we could determine the MWL allocation, which is the first step in building an efficient multi-modal interface for a partial autonomous-driving vehicle. Our experiment determined the MWL allocation with visual perceptual modality at SAE level 0 (manual driving). This means that the alert information occupying visual-visual cognition is not preferred in the TOR process at SAE levels 1 to 3. This will incur a high MWL due to the conflict when the driver drives manually and regains her/his SA simultaneously because visual cognition is used more in manual driving. Verbal cognition is the least occupied cognition channel by manual driving with visual perception. Therefore, the verbal stimuli, e.g. "LEFT!" on the display will be a promising candidate for the interface to alert drivers before take-over.

An adaptive interface concerning the driver's MWL should be designed according to the MWL allocation at each SAE level. The alert information should be presented in the cognitive channel with the lowest MWL because of fewer conflicts in that cognitive channel; therefore, the driver will be able to react quicker.

9.2. Evaluating proposed model using previous experiment

The data from a previous experiment [3] were used to evaluate the reliability of the proposed multi-modal interface model. Ten participants took part in a driving simulation, which was similar to that introduced in this paper. The participants were required to carry out the driving simulation and secondary tasks simultaneously, the difference was the secondary tasks choices. This experiment involved nine secondary tasks: [Automated Operation Span Task] (AOSPAN), Time Perception, [Paced Auditory Serial Addition Test] (PASAT), Calling, 2-back task, Texting, Arithmetic, Simple question, and Simple instruction. NASA-TLX was also used.

The secondary-task process and NASA-TLX scores are shown in Table 3.

The MWL allocation of each task is illustrated in Figure 11 based on the proposed model. The MWL was categorized by colour according to Russian Activity Theory [18]:

(1) category one (green): target search
(2) category two (yellow): target search and identification in group
(3) category three (red): logical judgement or memory search
(4) category four (purple): calculation

The MWL allocation for manual driving is illustrated as Figure 12 according to the previous experimental results [18]. Since the MWL allocations of auditory and tactile modalities are still uncertain, the categories of the auditory and tactile modalities are in grey, which means unknown. A different channel is label
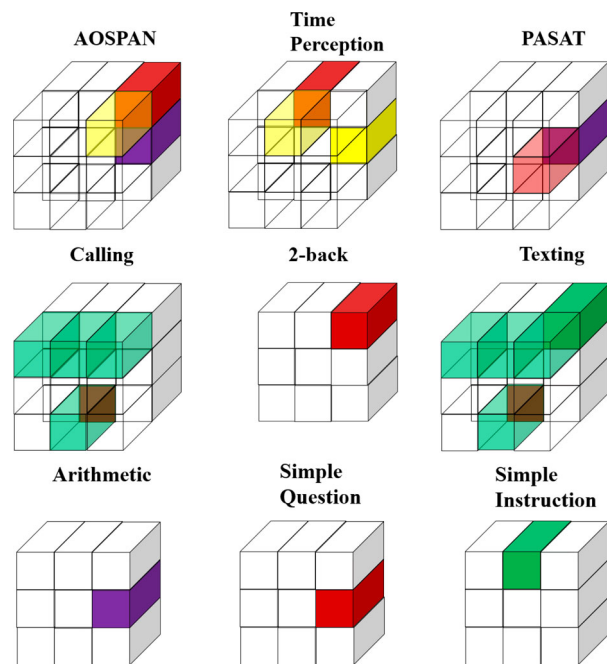

Figure 11. MWL allocation assumptions of nine secondary tasks.

Table 3. Secondary-tasks descriptions and NASA-TLX scores.

| Task	Process description	NASA-TLX
AOSPAN	5 letters in random order were presented on an iPad. After showing them to participants, 5 simple math problems were presented in audio. After solving math problems, investigator asked participants to say 5 letters in order	62.06
Time perception	Estimate time interval. To prevent participants from silently counting seconds, distracting question was asked during each interval to avoid participants counting in mind	62.17
PASAT	Numbers were presented about every 3s in audio and participants' task was to add them continuously	63.56
Calling	Call investigator using a cell phone	75.20
2-back task	10 numbers were shown on iPad one by one, participants had to decide whether current number was same as that presented 2 trials ago	72.92
Texting	Send a message to investigator using cell phone	77.5
Arithmetic	Solving simple math problem in audio	62.13
Simple question	Answer questions about weather, favourite food, etc.	45.11
Simple instruction	Follow instruction of turning at next crossroad	49.33
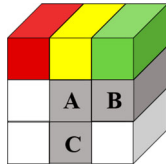
Figure 12. MWL allocation assumption for manual driving.

Table 4. Quantitative verification of proposed model.

Task	Factor	NASA-TLX
AOSPAN	11+B	62.06
Time perception	11+B	62.17
PASAT	7+2B	63.56
Calling	13+2C	75.20
2-back task	4	72.92
Texting	15+2C	77.5
Math arithmetic	4+B	62.13
Simple question	3+B	45.11
Simple instruction	3	49.33

with a letter to distinguish the MWL of three different unknown channels (Figure 12).

When the MWL allocations of both tasks are available, they can be combined to determine which channels have what amount of conflicts. Unit 1, 2, 3, or 4 were assigned to category 1 channel (green), category 2 channel (yellow), category 3 channel (red), and category 4 channel (purple), accordingly. The quantitative calculation principle within each channel is still unclear. Only for a first check of the plausibility of proposed model, when there is a conflict in the channel, add the unit up for the total unit. Therefore, we have the quantitative comparison among different secondary tasks in Table 4.

Although there were three unknown variables: A (unit for Auditory Modality – Spatial Cognition channel), B (unit for Auditory Modality – Verbal Cognition channel) and C (unit for Tactile Modality – Spatial Cognition channel) (value between 1 to 4, highly possible just between 1 and 2 due to less engagement of the auditory and tactile modalities in manual driving activity), the qualitative comparison can be observed in Table 4, except the 2-back and Arithmetic tasks. These two tasks seem to not match the linear relationship between proposed units and measured mental workload, perhaps because cognition dealing with numbers simply requires more resources.

Furthermore, the small size of both experiments could not guarantee the reliability of the proposed model. The quantitative calculation principle needs more exploration rather than simply addition, and more investigation is necessary to verify the reliability of the proposed model.

10. Conclusion

This paper focused on a driver's MWL concerning cognitive channels and proposed a multi-modal interface model. Through driving simulation experiments,

which required participant to perform secondary tasks during driving, one part of the proposed model was verified, and the MWL allocation of the driver with visual stimulus during manual driving was determined. The driving task occupied the visual cognitive channel most, followed by the spatial cognitive channel, then verbal cognitive channel. The proposed model was also qualitatively evaluated using previous experimental data from different secondary tasks, and the results and predicted MWL using the proposed model were identical with data from previous experiment; therefore, the reliability of proposed model was verified.

Based on both sets of experimental results, we suggest that visual-verbal alert information should be presented when a similar take-over action occurs at SAE level 0 (= manual driving). One candidate is presenting words on the display. To validate the reliability of the entire proposed model, more experiments concerning auditory and tactile perception at SAE levels 1 to 3 should be conducted. With the determined MWL allocation, a suitable human-machine interface concerning take-over action can be designed.

Such an adaptive multi-modal interface can be developed by monitoring driver's engaged in NDRTs and acquire the corresponding MWL allocations of NDRTs (time phase 1 during a TOR). When there is a TOR, the interface will use the least occupied cognitive channel to alert the driver (time phase 2 during TOR). When the MWL allocations of manual driving, SAE levels 1 and 2, are determined (time phase 4 during TOR), the alert-information content (time phase 3 during TOR) enabling the driver to regain her/his SA can be placed in the non-occupied cognitive channels.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Weiya Chen She received her BS degree from Tongji University, China and Ostfalia University of Applied Sciences, Germany and MS degree from Technical University of Berlin University, Germany, in 2016 and 2018, respectively. She is currently a PhD student at Kyoto University, Japan. Her research interests include Human-Machine interaction, Mental Workload and Cognitive Psychology.

Tetuso Sawaragi He received his BS, MS and PhD degrees in Systems Engineering from Kyoto University in 1981, 1983 and 1988, respectively. From 1991 to 1992, he was a visiting scholar in the Dept. of Engineering-Economic Systems of Stanford University. In 1986 he joined Kyoto University, where he is currently a professor in the Dept. of Mechanical Engineering and

Science in Kyoto University's Graduate School of Engineering from 2002. He has been engaged in the researches on Systems Engineering, Cognitive Science and Artificial Intelligence, particularly in the development of human-machine collaborative systems. He is a member of IEEE, SICE, ISCIE, etc.

Toshihiro Hiraoka He received BE and ME degrees in Precision Engineering from Kyoto University in 1994 and 1996 respectively, and a PhD in Informatics from Kyoto University in 2005. He worked at Matsushita Electric Industrial Co. Ltd. from April 1996 to September 1998, the Department of Systems Science, Graduate School of Informatics, Kyoto University as an Assistant Professor from October 1998 to September 2017, and a Designated Associate Professor at the Institutes of Innovation for Future Society, Nagoya University from October 2017 to June 2019. He is currently a Project Professor at the Institute of Industrial Science, the University of Tokyo. His current research interests are the human-machine systems, the advanced driver-assistance systems, and the automated driving systems. He is a member of the Institute of Electrical and Electronics Engineers (IEEE), the Society of Automotive Engineers of Japan (JSAE), the International Association of Traffic and Safety Sciences (IATSS), etc.

References

[1] Boff KR, Kaufmann L, Thomas JP. Handbook of perception and human performance. Vol. 2. Cognitive processes and performance. Signal processing, 42-2. Chiang Mai: Wiley; 1994.

[2] Chen W, Sawaragi T, Hiraoka T. Driver's mental workload measurement concerning cognitive channels. Proceedings of the SICE Annual Conference 2020; 2020. p. 1458–1463.

[3] Chen W, Sawaragi T, Horiguchi Y. Measurement of driver's mental workload in partial autonomous driving. IFAC-PapersOnLine. 2019;52(19):347–352.

[4] Longo L. Designing medical interactive systems via assessment of human mental workload. 2015 IEEE 28th International Symposium on Computer-Based Medical Systems; 2015. p. 364–365.

[5] Shelton-Rayner G, Mian R, Chandler S. Leukocyte responsiveness, a quantitative assay for subjective mental workload. Int J Ind Ergon. 2012;42(1):25–33.

[6] O'Donnell RD, Eggemeier FT. Workload assessment methodology. In: Boff KR, Kaufman L, Thomas JP, editors. Handbook of perception and human performance. New York: Wiley; 1986. p. 1–49.

[7] Stuiver A, Brookhuis KA, de Waard D, et al. Short-term cardiovascular measures for driver support: increasing sensitivity for detecting changes in mental workload. Int J Psychophysiol. 2014;92(1):35–41.

[8] Kumashiro KK, Pinar WF, Graue E, et al. Thinking collaboratively about the peer-review process for journal-article publication. Harv Educ Rev. 2005;75(3):257–285.

[9] Wickens CD. Multiple resources and mental workload. Hum Factors. 2008;50(3):449–455.

[10] Sklar AE, Sarter NB. Good vibrations: tactile feedback in support of attention allocation and human-automation coordination in event-driven domains. Hum Factors. 1999;41(4):543–552.

[11] Bazilinskyy P, Petermeijer SM, Petrovych V, et al. Take-over requests in highly automated driving: a crowdsourcing survey on auditory, vibrotactile, and visual displays. Transp Res. 2018;F(56):82–98.

[12] Logie RH. The functional organization and capacity limits of working memory. Curr Dir Psychol Sci. 2011;20(4):240–245.

[13] Sanada M, Ikeda K, Hasegawa T. Shape and spatial working memory capacities are mostly independent. Front Psychol. 2015;6(581).

[14] Chen S, Epps J. Automatic classification of eye activity for cognitive load measurement with emotion interference. Comput Methods Programs Biomed. 2013;110:111–124.

[15] Mueller ST, Piper BJ. The psychology experiment building language (PEBL) and PEBL test battery. J Neurosci Methods. 2014;222:250–259.

[16] Miyake S, Kumashiro M. Subjective mental workload assessment technique-An introduction to NASA-TLX and SWAT and a proposal of simple scoring methods. Jpn J Ergon. 1993;29(6):399–408.

[17] Mckelvie SJ. Learning and awareness in the hebb digits task. J Gen Psychol. 1987;114(1):75–88.

[18] Bedny G, Meister D. The Russian theory of activity: current applications to design and learning. New York: Psychology Press; 2015. (Applied Psychology Series).