How we can measure the non-driving-task engagement in automated driving: Comparing flow experience and workload

Sang Min Ko, Yong Gu Ji

Department of Industrial Engineering, Yonsei University, 50 Yonsei-Ro, Seodaemun-gu, Seoul 03722, South Korea

A R T I C L E   I N F O

Keywords:
Automated driving
Flow experience
Take-over time
Eye-tracking
NASA/Tlx

A B S T R A C T

In automated driving, a driver can completely concentrate on non-driving-related tasks (NDRTs). This study investigated the flow experience of a driver who concentrated on NDRTs and tasks that induce mental workload under conditional automation. Participants performed NDRTs under different demand levels: a balanced demand–skill level (fit condition) to induce flow, low-demand level to induce boredom, and high-demand level to induce anxiety. In addition, they performed the additional N-Back task, which artificially induces mental workload. The results showed participants had the longest reaction time when they indicated the highest flow score, and had the longest gaze-on time, road-fixation time, hands-on time, and take-over time under the fit condition. Significant differences were not observed in the driver reaction times in the fit condition and the additional N-Back task, indicating that performing NDRTs that induce a high flow experience could influence driver reaction time similar to performing tasks with a high mental workload.

1. Introduction

Since the introduction of adaptive cruise control (ACC) in the 1990s, a variety of technologies have been introduced to increase the level of automation in vehicle driving (Bishop, 2005). With the addition of various functions related to driving safety in the current advanced driver assistance system (ADAS), automated driving at Society of Automotive Engineers (SAE) Level 2 is now possible. The purpose of ADAS was to support driving such that driver error would be reduced or even eliminated, while enhancing efficiency during traffic and transport (Brookhuis et al., 2001). The concept of driver support has often been associated with bypassing human control inputs in an effort to eliminate driver error (Banks and Stanton, 2016). The early functionality of tools such as ACC took the form of a system taking over some of the driver's tasks; the functionality of such tools has since evolved into an automated system that takes over multiple driving tasks (NHTSA, 2013). In automated driving, the human role shifts from that of an active driver to that of a system supervisor who monitors the situation and takes over control of the vehicle in certain situations. When the vehicle becomes fully automated, the driver will be expected to assume the role of a passenger (Diels and Bos, 2016). Therefore, as the level of vehicle automation increases, the driver will be excluded from the primary role of driving.

The introduction of automated driving has also changed the situation for drivers performing NDRTs. In manual driving, the driver only performed additional secondary tasks as needed while mainly focusing on the primary task of driving. During this process, driver distractions occurred, causing problems in reaction time and driving performance (Merat and Jamson, 2009; Young and Stanton, 2007). However, with automated driving, the driver can focus primarily on performing NDRTs (Carsten et al., 2012; Saxby et al., 2013). Thus, fundamental changes are expected in human–vehicle interaction (HVI), apart from the safety and convenience that can be achieved by the introduction of automated driving.

Previous studies related to manual driving have mainly analyzed the effects of driver performance of secondary tasks on the driver's mental workload, response time, and situation awareness. Several studies have shown that the driver's mental workload negatively affects the driver reaction time and situational awareness. On the other hand, studies related to automated driving have mainly focused on the performance and safety aspects of adopting this technology. For example, studies have been conducted on changes in driving performance with the adoption of automated driving functions (Merat et al., 2014). One comparative study examined the changes in the driver's state, such as changes in driver's attention and situation awareness, for manual and automated driving at SAE Levels 2 to 4 (Endsley and Kaber, 1999). However, in the initial studies related to automated driving, the experiments were conducted with drivers monitoring the driving situation without performing any particular secondary tasks. Therefore, these previous studies lacked realism and were limited by employing artificial
or standardized tasks as part of NDRTs. In fact, when a naturalistic task is used as a secondary task, a difference in the results of experiments has been identified (Shinar et al., 2005). Thus, research on driver's state when performing naturalistic NDRTs in automated driving situations is lacking.

In manual driving, the driver simultaneously performs both driving and secondary tasks. Evaluation of the driver's mental workload in the context of these dual tasks has been an important research topic. Performing dual tasks affects the driver's mental workload and causes driver distraction. However, in a high-level automated driving situation where there are fewer restrictions on the driving task and active driving situation monitoring is not required, the driver can concentrate on performing NDRTs. In this case, the question arises as to whether analyzing the driver's state during automated driving using the same technique used in manual driving would be appropriate; in other words, would analyzing only the mental workload or situation awareness of the driver be insufficient?

The situational characteristics of automated driving, which can be accomplished through the absorption in a particular task, can be explained by the term “flow.” Csikszentmihalyi (1975) proposed the term flow to describe intense engagement or complete absorption in a task (McQuillan and Conde, 1996). Studies on flow and its benefits have been conducted in a variety of contexts, such as reading, media use, and leisure time (Csikszentmihalyi and LeFevre, 1989; McQuillan and Conde, 1996; Sherry, 2004). Tozman et al. (2015) reported that flow could occur when driving a vehicle via a driving simulator. They used experimental settings that manipulate a difficulty level to induce boredom, flow, and anxiety. In their experimental design, a fit condition was set to induce flow through the demand level, which was designed to fit the participant's skill level. The other levels were low (boredom condition) and high (anxiety condition) demand levels. Thus, this study was planned to analyze the driver's state in automated driving by evaluating the flow experience and mental workload of a driver performing NDRTs, and also analyze the relationship between the reaction time of the driver and take-over requests (TOR).

This study is based on the assumption that the mental workload and flow state affect the reaction time of the driver. Mental workload refers to the amount of attentional resources required to complete a task (Wickens, 2002; Young and Stanton, 2004). Mental workload can impact the driver's attentional resource capacity and lead to a decrease in performance. Therefore, increasing levels of difficulty in mental tasks will result in performance deterioration (Wickens, 2008). Flow is a state of pleasantness, where a person feels in control and focused, and a balance exists between the demands of a task and the skills of the person. The flow state requires similar attentional resources as mental workload (Connolly, 2007). Based on this definition, we attempt to analyze the relationship between the flow experience and driver reaction time.

This study investigates how the drivers' subjective states affect the driver's reaction time upon a TOR for the control of a vehicle. Thus, the objectives of this study are as follows: 1) to assess the flow experience and mental workload of a driver performing tasks with different demand levels, 2) to analyze the reaction time of the driver performing NDRTs in simulated automated driving, 3) to investigate the relationship between the flow experience and reaction time of driver, and 4) to investigate the difference between the N-Back task and other tasks.

2. Experiment 1

Experiment 1 was designed to assess the flow experience and the mental workload of a driver performing NDRTs and additional N-Back tasks.

2.1. Participants

Thirty-two participants (male = 19, female = 13) with ages ranging from 23 to 39 (M = 28.22, SD = 4.434) participated in Experiment 1. All participants had a driver's license and drove regularly (M = 2.78 times per week, SD = 2.362). No participants had previous experience with automated driving. In particular, we recruited participants with corrected visual acuity of 0.1. For the participants who wore glasses, only those with a corrected visual acuity of more than 0.1 when wearing contact lenses were qualified for the study. The limits on the corrected visual acuity were designed to eliminate risk factors that might hinder the experiment, especially during the use of eye-tracking devices in Experiment 2.

2.2. Materials and apparatus

2.2.1. Driving simulator

A fixed-based driving simulator was designed to provide the participants with an environment similar to the driver's seat of a real vehicle (Fig. 1). The driving simulator consisted of a desktop computer, simulation software (City Car Driving v1.5), driver's seat, steering wheel, and foot pedals (Logitech Force Feedback Racing Wheel). Three liquid crystal display (LCD) monitors (27 inches) were arranged in front of the driver's seat to provide a surround panoramic view for the driver during the task. The driver's seat, which was connected to a power supply (12 V), was adjusted to fit the participant's body. Moreover, a tablet (iPad Air 2, 9.7 inches) was placed at the center fascia on the right side of the steering wheel so that the participants could perform the tasks required to be done during the experiment. This driving simulator was set to provide conditionally automated driving at SAE Level 3: the vehicle will fully take over the driving responsibilities under restricted conditions, but the human driver is expected to take over when the automated driving system asks for it (Committee, 2014). Therefore, participants can have their hands off the steering wheel and eyes off the road when performing tasks.
2.2.2. Flow experience

Flow experience was assessed using the flow short scale (FKS) (Engeser, 2012; Rheinberg et al., 2003). The FKS has been successfully validated and utilized in various research areas (Engeser and Rheinberg, 2008; Schüler, 2009; Tozman et al., 2015). This questionnaire consisted of ten items (Cronbach’s alpha = 0.90), including (1) fluency of performance (e.g., “My thoughts/activities run fluidly and smoothly,” “I feel that I have everything under control”) and (2) absorption by activity (e.g., “I do not notice time passing,” “I am totally absorbed in what I am doing”) on a 7-point Likert scale (1 = “Not at all”; 4 = “Partly”; 7 = “Very much”) that were answered according to the participant’s degree of agreement with the statements. The mean value of the 10 items was calculated as the flow score in which higher values indicated stronger flow experiences.

Rheinberg et al. (2007) argued that, in addition to a balance between the challenge and the skill, the perceived difficulty level or demands of the task should also be assessed because challenge combines the difficulty and skill. For example, an “easy” task can be perceived as very challenging to a novice, and a “difficult” task can seem unchallenging to an expert (Moneta, 2012). Thus, the FKS measured the perceived demand level of the activity (e.g., “For me personally, the current demands are ...”) on a 9-point Likert scale (1 = “too low”; 5 = “just right”; 9 = “too high”) (Engeser, 2012; Engeser and Rheinberg, 2008). We used this item for a manipulation check for the experimental condition. All participants completed the questionnaire after each of the experimental tasks.

2.2.3. Mental workload

The NASA task load index (NASA-TLX) is the most common subjective mental workload assessment tool used in experimental studies. Compared to other scales, the NASA-TLX is multidimensional and easy to administer. The NASA-TLX has been supported by many studies in laboratory conditions, and it has been used to show the significant impact of workload changes (Basshel, 2012). The NASA-TLX rates perceived workload on six different scales: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration. Participants first evaluated the six scales according to how much they contributed to the workload required for the task being studied. Then, each of the six scales was weighted. The weightings were determined by the participant’s answers to 15 pair-wise comparisons and were designed to significantly enhance the sensitivity of the overall workload score while reducing between-rater variability (Hart and Staveland, 1988).

2.3. Experimental tasks

With increasing levels of automation, drivers are more likely to engage in NDRTs (de Winter et al., 2014). Clark and Feng (2017) observed how drivers voluntarily engaged in NDRTs during simulated driving with conditional automation; drivers engaged in various non-driving-related activities, such as using an electronic device or listening to music. Zeeb et al. (2015) examined the influence of engagement in NDRTs on take-over time and quality. The NDRTs included “writing an email,” “reading the news,” and “watching a video.” Zeeb et al. (2015) expected that all tasks could exert a comparatively high level of visual and cognitive demand and might be attractive for the driver. The actual effect of NDRTs (secondary tasks) on driving behavior depends on the type of the task (Naujoks et al., 2016); specifically, the effect depends on how long the drivers take their eyes off the road (NHTSA, 2012). Based on previous studies, the two experimental tasks of “viewing a video” and “reading an article,” which a driver can naturally perform in an automated driving environment, were chosen for this experiment. In addition, we used the N-back task to artificially induce mental workload as an additional task.

According to the original flow theory (Csikszentmihalyi, 1975), flow is best achieved if the demands of a task are in balance with the skills of a person (challenge–skill balance). Rheinberg and Vollmeyer (2003) demonstrated that flow could be induced experimentally by presenting tasks with different difficulty levels or different demands. The highest flow experience was induced during the tasks with optimal (moderate) difficulty levels or demands (flow state). Moreover, the lowest flow experience was induced during tasks that were too easy (boredom) or too difficult (anxiety) (Rheinberg and Vollmeyer, 2003).

Sherry (2004) argued that enjoyment of media results from the flow experience realized when the media message content balances with an individual’s ability to interpret that message; thus, enjoyment of media has many of the same aspects of flow. First, media use clearly provides an enjoyable experience. Second, the gratifications of media use are indicative of the intense focus and loss of self-consciousness experienced in a flow state. Third, many people have experienced temporal distortion and losing track of time while using media, such as when watching a movie or reading a novel. Finally, media use is, at least in part, intrinsically motivating. Flow state in media use can be achieved if the message difficulty is balanced with the usage skills of an individual (difficulty–skill balance) (Sherry, 2004). Furthermore, Towey (2000) reported that people frequently reach a flow state while reading self-selected materials. About seventy percent of materials that provide a flow experience was self-selected. Self-selected materials do not guarantee which people can experience a flow state. However, the previous interest in the selected topic is important in predicting an achievement of a flow experience (Towey, 2000).

The demand levels for experimental tasks were adjusted by media difficulty (Sherry, 2004) and self-determination in the selection of topics (Towey, 2000). Each experimental task consisted of three conditions: boredom, flow, and anxiety. The participants had been provided assigned materials with no specific topic as experimental tasks in the boredom condition. Having no specific topic means that participants had no difficulty in understanding and proceeding with the task, regardless of their background knowledge and interest. We used a story without content and a calm natural scenery as the experimental task. The participants were allowed to choose self-selected topics of interest for experimental tasks in the fit condition. They chose various topics such as movie, sports, music, game, and social issues. Similar to the boredom condition, in the anxiety condition, the participants had been provided assigned materials on a specific topic as experimental tasks. Having a specific topic means that participants had difficulty in understanding and proceeding without adequate knowledge or interest in the assigned materials. We used a scientific theory as the experimental task.

Radlmayr et al. (2014) compared visual tasks and a cognitively demanding task (2-back) within an experiment on highly automated driving. They showed that cognitively demanding NDRTs can lead to similar results on the take-over time and quality compared with visual tasks (Gold et al., 2016; Radlmayr et al., 2014). In this study, we employed the N-Back (2-back) task as an additional task to induce mental workload. It presented a single-digit number (0–9) to the participants with an interstimulus interval of 2 s via a tablet screen. The participants needed to decide whether or not to touch a tablet screen if the current number was the same as that presented two trials ago.

2.4. Procedure

Participants received instructions on the task procedures and were asked to complete a preliminary survey that consisted of questions about their demographical information, driving experience, and automated driving. The participants watched a video that enhanced their understanding of automated driving, as most participants had only simple knowledge of automated driving, but had never experienced it firsthand. Then, the participants were given a detailed explanation of the information used in the experiment, such as the feature configurations on the NASA-TLX and FKS, as well as the N-Back tasks. Moreover, a short practice task was performed in as little as 30 s so the participants could smoothly perform the N-back task. Lastly, the participants were asked to sit in the driver’s seat of the driving simulator to adjust the seat.
to the best driving position for their body.

The experiments were conducted in a fixed-based driving simulator. The experiment adopted a 2 × 3 within-subject design with the task type and level of demand in tasks. We presented six NDRTs with different demand levels to induce boredom, flow, and anxiety conditions. Furthermore, we presented the N-Back (2-back) task as an additional task. All participants were exposed to the tasks in random order. Each task took 5–6 min to complete. Tozman et al. (2015) reported that they were able to induce boredom, flow, and anxiety in a 6-min experimental task. After completing each task, participants filled in the NASA-TLX and FKS evaluation forms and took a 5-min break. After all experiments in Experiment 1 were completed, participants took a 20-min break.

2.5. Results

2.5.1. Perceived demand level

We tested whether different participants perceived the demand level of the experimental tasks differently. No significant effect of gender was observed (F(1, 191) = 1.629, p = 0.203). Further, we investigated whether different participants perceived the demand levels of the task type and the experimental conditions differently. We performed a two-way repeated measures analysis of variance (ANOVA) using the task type and the experimental condition as the within-subjects repeated measure. Participants rated the perceived demand level of the viewing tasks (M = 4.281, SD = 0.106) to be significantly lower than that of the reading tasks (M = 5.010, SD = 0.100) (Wilks’ λ = 0.461, F(1, 31) = 36.270, p < 0.001, η² = 0.539). In addition, participants rated the perceived demand level during the experimental conditions significantly differently (Wilks’ λ = 0.092, F(2, 30) = 147.404, p < 0.001, η² = 0.908). Post-hoc tests were performed to specifically contrast the experimental conditions based on the perceived demand level. All experimental conditions were found to be statistically significant (p < 0.05). The fit condition (M = 5.063, SD = 0.132) required a higher perceived demand level than the boredom condition (M = 2.594, SD = 0.136). The anxiety condition (M = 6.281, SD = 0.193) required a higher perceived demand level than the fit condition (M = 5.063, SD = 0.132). This result suggested that experimental tasks were successfully manipulated.

2.5.2. Flow experience

No significant gender effect was observed (F(1, 191) = 0.036, p = 0.850). Further, to investigate the effect of the task type and the experimental conditions on the flow score, we performed a two-way repeated measures ANOVA. There was a statistically significant interaction in the flow score between the task type and the experimental condition (Wilks’ λ = 0.378, F(2, 30) = 24.640, p < 0.001, η² = 0.622). The result showed a significant effect of the experimental condition (Wilks’ λ = 0.278, F(2, 30) = 39.028, p < 0.001, η² = 0.722); however, the task type did not show a significant effect (Wilks’ λ = 0.994, F(1, 31) = 187, p = 0.669, η² = 0.006). Post-hoc test results showed that the difference between the fit conditions and other conditions were statistically significant (p < 0.05). However, the difference between the boredom condition and the anxiety condition was not statistically significant. The flow score was the highest in the fit condition (M = 6.042, SD = 0.092), and lowest in the boredom condition (M = 4.761, SD = 0.177). In addition, the flow score in the anxiety condition (M = 4.952, SD = 0.181) was higher than that in the boredom condition (M = 4.761, SD = 0.177). However, the difference between the boredom condition and the anxiety condition was not significant.

2.5.3. Mental workload

No significant effect of gender was observed F(1, 191) = 0.625, p = 0.430). Further, we performed a two-way repeated measures ANOVA using the task type and the experimental conditions as the within-subjects repeated measure. There was a statistically significant interaction in the workload score between the task type and the experimental condition (Wilks’ λ = 0.706, F(2, 30) = 6.260, p = 0.005, η² = 0.294). The result of the task type was significant (Wilks’ λ = 0.234, F(1, 31) = 101.730, p < 0.001, η² = 0.766), and the experimental condition was also significant (Wilks’ λ = 0.223, F(2, 30) = 52.356, p < 0.001, η² = 0.777). The workload scores of the reading task were higher than those of the viewing task for each experimental condition: 1) reading task (M = 34.427, SD = 14.126) was higher than the viewing task (M = 14.261, SD = 8.997) in the boredom condition, 2) reading task (M = 38.218, SD = 15.587) was higher than the viewing task (M = 20.458, SD = 13.048) in the fit condition, and 3) reading task (M = 51.906, SD = 13.287) was higher than the viewing task (M = 42.458, SD = 14.338) in the anxiety condition.

2.5.4. Summary

We first examined whether the demand level of the task type and the experimental condition were perceived differently. There was a significant difference in the perceived difficulty of each task type. For the viewing task, the average perceived demand level was 4.281 (SD = 0.106) while that for the reading task was 5.010 (SD = 0.100). These results suggested that the task type had an effect on the perceived demand level. In addition, there was a significant difference in the perceived difficulty of each experimental condition. For the boredom condition, the average perceived demand level was 2.594 (SD = 0.136); that for the fit condition was 5.063 (SD = 0.132), and that for the anxiety condition was 6.281 (SD = 0.193) (Fig. 2). These results suggested that the experimental condition had an effect on the perceived demand level. Specifically, these results indicated that manipulating the demand level of the experimental condition was successful.

We then tested the effect of the task type and experimental condition on the flow experience. We assessed the flow experience using the FKS. There was a significant interaction in the flow experience between the task type and the experimental condition. These results suggested that each task had an effect on the flow experience. In addition, there was a significant effect between the flow experience and the experimental condition. For the boredom, fit, and anxiety conditions, the average flow experiences were 4.761 (SD = 0.177), 6.042 (SD = 0.092), and 4.952 (SD = 0.181), respectively. These results suggested that the experimental condition had an effect on the flow experience. Specifically, the flow experience was the highest in the fit condition, and was significantly different from the flow experience in other conditions (Fig. 2).

We measured the mental workload using the NASA-TLX. There was a significant interaction in the mental workload between the task type and the experimental condition. These results suggested that the mental workload of each task was rated differently. There was a significant effect on the mental workload for each task type. For the viewing task, the average mental workload was 25.726 (SD = 1.508), while that for the reading task was 41.517 (SD = 2.025). These results suggested that task type had an effect on the mental workload. There was a significant effect on the mental workload for the experimental condition. For the boredom, fit, and anxiety conditions, the average mental workload values were 24.344 (SD = 1.474), 29.338 (SD = 2.087), and 47.182 (SD = 2.275), respectively. These results suggested that the experimental condition had an effect on the mental workload. Specifically, the mental workload was the highest in the anxiety condition, and was significantly different from the mental workload in other conditions. Moreover, the difference between the fit condition and boredom condition with regards to mental workload was statistically significant (p < 0.05) (Fig. 2).
3. Experiment 2

Experiment 2 was conducted to analyze the driver reaction time upon TOR for the control of a vehicle when the driver was performing tasks with different demand levels during simulated automated driving.

3.1. Participants and experimental tasks

The same group of 32 participants from Experiment 1 participated in Experiment 2. Only one female participant wore contact lenses to proceed with the experiment. The same experimental tasks used in Experiment 1 were used in Experiment 2.

3.2. Apparatus

3.2.1. Driving simulator

The experiment was conducted in the fixed-based driving simulator that was the same as that used in Experiment 1. Additionally, Experiment 2 used an eye-tracking system to track the movement of the participant’s eyes, and a video camera to observe the participants’ behavior.

3.2.2. Recording of eye movements

To capture eye gaze movements, we used eye-tracking glasses (ETG) and the iView ETG v2.7.1 controlling software produced by SensoMotoric Instruments (SMI, Berlin, Germany) (Fig. 3). The SMI ETG recorded eye gaze at a sampling frequency of 60 Hz and scene video at 30 Hz with integrated audio recording (SMI, 2016). To obtain accurate eye gaze movements, we utilized the three-point calibration mode. The iViewETG software recorded the eye tracking data using SMI ETG.

We used the behavioral and gaze analysis (BeGaze) v3.6 (SMI: Berlin, Germany) to analyze the eye-tracking data. The SMI BeGaze software analyzed and collected information from the participants’ eye gaze movements, produced meaningful graphs to visualize the eye-tracking data, and exported the eye-tracking statistics. Participants wearing the SMI ETG performed NDRTs within the simulated automated driving. After a TOR, participants were required to redirect their gaze from the NDRTs to the roadway and return their hands to the steering wheel and feet to the pedals. In Table 1, we summarize the definition of the variables of the driver reaction time.

3.3. Procedure

The participants were instructed on the procedures of the task, the duration of the task, and how to handle additional TOR. The driver adjusted the position of the seat to feel comfortable during the simulations. Participants put on the SMI ETG, and gaze calibration was conducted for each participant. Then, we validated the gaze calibration with the live feedback on the scene video. The experiment adopted a 2 × 3 within-subject design, with task type (watching a video, reading an article) and level of demand in tasks (boredom, fit, and anxiety condition). As in Experiment 1, the participants performed six NDRTs with different demand levels and performed the N-Back task as an additional task. The driving tasks were chosen from the PC game package City Car Driving v1.5.

Audio guidance notified the driver about the switch to automated driving while the vehicle was stopped. After the notification, the vehicle started to drive automatically. According to the surrounding road conditions and traffic, the vehicle was driving at fixed speeds of 80 km/h (49.71 mi/h). The participants were instructed to perform NDRTs when they judged that driving was stabilized. As in Experiment 1, participants conducted six NDRTs and the N-back (2-back) task as an

Fig. 2. Analysis results of (a) perceived demand level, (b) flow score, and (c) workload score for experimental conditions.

Fig. 3. Eye-tracking equipment and software: (a) SMI eye-tracking glasses (ETG) and (b) iViewETG.
additional task. All participants were exposed to the tasks in random order. The participants performed the given tasks during simulated automated driving. The auditory sound was issued to the participants with the TOR signal. The TOR was randomly presented at simulated driving times between 5 min and 6 min. The participants were instructed to stop performing the ongoing task, take the steering wheel, and proceed with manual driving. After completing each task, participants took a 10-min break. The eye-tracker device was removed after all the experiments were completed (Fig. 4).

3.4. Results

3.4.1. Driver reaction time

As Table 1 shows, the driver reaction time consists of four variables. All variables could be expected to correlate with each other. Therefore, to analyze the difference in the vectors of means, we utilized a multivariate analysis of variance (MANOVA). Before conducting the MANOVA, we performed correlation analysis to test the MANOVA assumption that the dependent variables would be correlated with each other in a reasonable range (Meyers et al., 2006). All measured variables had significant correlations with each other (Table 2).

No significant effect of gender was observed (Wilks’ Λ = 0.780, F (12, 51) = 1.201, p = 0.308, η^2 = 0.220). A two-way MANOVA was conducted to analyze the main effects and interactions of task type and experimental condition on the driver reaction time. Not all interaction effects were statistically significant (Wilks’ Λ = 0.950, F(8, 366) = 1.183, p = 0.308, η^2 = 0.025). The results for the experimental condition were significant (Wilks’ Λ = 0.596, F(8, 255) = 13.489, p < 0.001, η^2 = 0.228). However, those for task type were not significant (Wilks’ Λ = 0.967, F(4, 183) = 1.581, p = 0.181, η^2 = 0.033).

A one-way MANOVA was conducted to analyze the effects of the experimental condition on the driver reaction time. There was a statistically significant multivariate effect for the experimental condition (Wilks’ Λ = 0.599, F(8, 372) = 13.594, p < 0.001, η^2 = 0.226). Given the significance of the overall test, the univariate main effects were examined. Significant univariate effects of the experimental condition were obtained for the gaze-on time (F(2, 189) = 5.312, p = 0.006, η^2 = 0.053), road-fixation time (F(2, 189) = 32.638, p < 0.001, η^2 = 0.257), hands-on time (F (2, 189) = 17.069, p < 0.001, η^2 = 0.153), and take-over time (F (2, 189) = 39.370, p < 0.001, η^2 = 0.294). Post-hoc test results showed that the difference between the fit condition and other conditions was statistically significant (p < 0.05) for each dependent variable. However, the difference between the boredom condition and anxiety condition for each dependent variable was not (Fig. 5).

The stepdown analysis involved a multi-step application of univariate linear models involving the dependent variables. The Roy–Bargmann stepdown F-test has been suggested as a post-hoc procedure for significant MANOVA results (Finch, 2007). This method can be used to perform univariate analysis of the significance of the first dependent variable, and then test the second dependent variable control of the first as a covariate, and so on, sequentially rotating the dependent variables to the status of covariates (Stevens, 1996). This testing was performed for each of the dependent variables, with variables higher in the sequence serving as covariates for those lower in importance (Finch, 2007).

To investigate the impact of the experimental condition on the individual dependent variables, the Roy–Bargmann Stepdown F-test was performed on the prioritized dependent variables. As Table 3 shows, there was a statistically significant effect for the experimental condition (Wilks’ Λ = 0.599, F(8, 372) = 13.594, p < 0.001), indicating a difference in the dependent variables between the experimental conditions. Gaze-on time, road-fixation time, and take-over time were judged to be sufficiently reliable to warrant stepdown analysis. Significant effects of the experimental condition were obtained for gaze-on time (F (2, 189) = 5.312, p = 0.006), road-fixation time (F (2, 188) = 34.594, p < 0.001), and take-over time (F (2, 186) = 12.183, p < 0.001). However, a significant effect for the hands-on time was not observed (F (2, 187) = 2.058, p = 0.131).

3.4.2. Comparison with N-back task

We tested the difference between the effect of the N-back task and the experimental condition on the reaction time using a one-way MANOVA. There was a statistically significant multivariate effect for the experimental condition (Wilks’ Λ = 0.580, F(12, 574.42) = 10.942, p < 0.001, η^2 = 0.166). Given the significance of the overall test, the univariate main effects were examined. Significant univariate main effects of the experimental condition were obtained for gaze-on time (F (3, 220) = 7.637, p < 0.001, η^2 = 0.094), road-fixation time (F(3, 220) = 25.577, p < 0.001, η^2 = 0.259), hands-on time (F(3, 220) = 11.549, p < 0.001, η^2 = 0.136), and take-over time (F(3, 220) = 29.422, p < 0.001, η^2 = 0.286). Post-hoc tests were performed to accurately contrast the experimental conditions on the variables of the reaction time. The analysis results showed that the difference between the fit condition and N-back was not statistically significant (p < 0.05) for each dependent variable. Moreover, the

---

Table 1

<table>
<thead>
<tr>
<th>Variables</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaze-on time</td>
<td>The time it takes for a participant to watch the windshield for the first time</td>
</tr>
<tr>
<td>Road-fixation time</td>
<td>The time it takes for a participant to watch the road for the first time</td>
</tr>
<tr>
<td>Hands-on time</td>
<td>The time it takes for a participant to hold the steering wheel for the first time</td>
</tr>
<tr>
<td>Take-over time</td>
<td>The time it takes for a participant to get a control authority (when the steering wheel changed more than 5°)</td>
</tr>
</tbody>
</table>

Table 2

<table>
<thead>
<tr>
<th>Variables</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaze-on time</td>
<td>–</td>
<td>0.748**</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Road-fixation time</td>
<td>0.748**</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Hands-on time</td>
<td>0.613**</td>
<td>0.676**</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Take-over time</td>
<td>0.574**</td>
<td>0.735**</td>
<td>0.890**</td>
<td>–</td>
</tr>
</tbody>
</table>

Note: **: Correlation is significant at the 0.01 level (2-tailed).

---

Fig. 4. A flow chart of the procedure for Experiment 2.
Results of the Roy-Bargmann Stepdown F-test were performed on the prioritized dependent variables. As Table 4 shows, there was a statistically significant effect for the experimental condition (Wilks’ $\lambda = 0.580$, $F(12, 574.42) = 10.942$, $p < 0.001$), indicating a difference in the dependent variables between the experimental conditions. Gaze-on time, road-fixation time, and take-over time were judged to be sufficiently reliable to warrant stepdown analysis. Significant effects of the experimental condition were obtained for gaze-on time ($F(3, 219) = 23.696$, $p < 0.001$), road-fixation time ($F(3, 219) = 23.696$, $p < 0.001$), and take-over time ($F(3, 217) = 10.294$, $p < 0.001$). However, a significant effect of the experimental condition for hands-on time was not observed ($F(3, 218) = 2.321$, $p = 0.076$).

### 3.4.3. Summary

We examined whether there was a difference in the driver reaction time for each task. We measured the four variables of the driver reaction time by analyzing the participants’ eye-gaze movement data. First, we conducted correlation analysis to investigate the relationship between the four variables of reaction time. The four variables of reaction time were found to have significant correlations with each other. A two-way MANOVA was conducted to analyze the effects and interactions of the task type and experimental condition on the driver reaction time. Not all interaction effects were statistically significant. In addition, only a significant effect of the experimental condition was observed. Therefore, we investigated the effects of the experimental condition on the reaction time.

A one-way MANOVA was conducted to analyze the effects of the experimental condition on the reaction time. There was a statistically significant multivariate effect for the experimental condition. Given the significance of the overall test, the univariate main effects were examined. Significant univariate main effects of the experimental condition were obtained for all variables of the driver reaction time. These results suggested that the experimental condition had an effect on the driver reaction time. Next, the Roy-Bargmann Stepdown F-test enabled examination of the pattern of relationships between the dependent variables (gaze-on time, road-fixation time, hands-on time, and take-over time) and the independent variable (experimental condition). There was a statistically significant effect for the experimental condition. Significant effects of the experimental condition were obtained for gaze-on time, road-fixation time, and take-over time. However, there was no significant effect for hands-on time. These results suggested that the experimental condition had an effect on the driver reaction time, but not on the hands-on time.

### 4. Discussion and conclusions

In flow studies, one of the major challenges to inducing flow is matching the skill level of the participants to the demand level of the task (Keller et al., 2011). Thus, the experimental tasks should be properly chosen so that participants feel self-involved during the experiment, which will adequately induce boredom, flow, and anxiety (Tozman et al., 2015). Results of our analysis showed that the flow experience depended on the experimental condition. The fit condition was rated the highest flow experience, as expected, and the boredom condition was rated the lowest. The anxiety condition was rated to have a higher flow experience than the boredom condition. However, the difference between the boredom condition and the anxiety condition was not significant. Our findings were consistent with those of previous studies on flow (Csikszentmihalyi, 1975; Tozman et al., 2015).

We aimed to investigate the relationship between the flow experience of the drivers and the driver reaction time of TOR. We analyzed reaction time when drivers performed NDRTs during simulated autonomous driving. The driver reaction time consisted of four variables: gaze-on time, road-fixation time, hands-on time, and take-over time. After TOR, drivers redirected their gaze from the NDRTs to the roadway, and they had to return to manual driving. The gaze-on time of the driver in the flow condition was the highest at 1.258 s, while the gaze-on time was 1.030 s in the boredom condition and 1.011 s in the anxiety condition. The gaze-on time in the flow condition was significantly different from those in the other conditions; however, the difference between the gaze-on times in the boredom condition and anxiety condition was not statistically significant. The mean gaze-on time increased with flow experience. Similar to the gaze-on time, the other three driver reaction time variables were the highest in the flow condition, and the differences in their values for the other conditions were not significant. However, the difference between the boredom condition and anxiety condition was not statistically significant.

### Table 3

Results of the Roy-Bargmann stepdown F-test on NDRTs.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Wilks' $\lambda$</th>
<th>StepDown $F$</th>
<th>df</th>
<th>Error df</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaze-on time</td>
<td>0.599***</td>
<td>5.312</td>
<td>2</td>
<td>189</td>
<td>0.006</td>
</tr>
<tr>
<td>Road-fixation time</td>
<td>34.594</td>
<td>2</td>
<td>188</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>Hands-on time</td>
<td>2.058</td>
<td>2</td>
<td>187</td>
<td>0.131</td>
<td></td>
</tr>
<tr>
<td>Take-over time</td>
<td>12.183</td>
<td>2</td>
<td>186</td>
<td>0.000</td>
<td></td>
</tr>
</tbody>
</table>

Note) ***: $p < 0.001$.

### Table 4

Results of the Roy-Bargmann stepdown F-test for the N-back task.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Wilks' $\lambda$</th>
<th>StepDown $F$</th>
<th>df</th>
<th>Error df</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaze-on time</td>
<td>0.580***</td>
<td>7.637</td>
<td>3</td>
<td>220</td>
<td>0.000</td>
</tr>
<tr>
<td>Road-fixation time</td>
<td>23.696</td>
<td>3</td>
<td>219</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>Hands-on time</td>
<td>2.321</td>
<td>3</td>
<td>218</td>
<td>0.076</td>
<td></td>
</tr>
<tr>
<td>Take-over time</td>
<td>10.294</td>
<td>3</td>
<td>217</td>
<td>0.000</td>
<td></td>
</tr>
</tbody>
</table>

Note) ***: $p < 0.001$. 

Fig. 5. Analysis results of the experimental condition on the driver reaction time.
Previous studies reported the first contact with the steering wheel occurred at about 1.2 s–1.8 s (Gold et al., 2013; Zeeb et al., 2015). Gold and Bengler (2014) further reported that take-over time depended on the time budget that was given to a driver. Drivers showed longer first gazes at the scenery when they were given a time budget of 7 s (0.9 s) instead of 5 s (0.7 s). Furthermore, the drivers took longer time to touch the steering wheel when they were given a longer time budget (1.8 s and 1.5 s for a time budget of 7 s and 5 s, respectively). In addition, previous studies reported an influence of NDRT (secondary task) demands on take over time (Merat et al., 2012; Petermann-Stock et al., 2013). Petermann-Stock et al. (2013) reported take-over times between 5.7 s and 8.8 s. Further studies found mean take-over times ranging from 2.1 s to 4.1 s. In this study, the first time the hands touched the steering wheel was 1.851 s (1.657 s–2.219 s), and the whole take-over time was 2.258 s (1.954 s–2.727 s). These results are not far off the findings of previous studies on take-over time in automated driving. In future studies, it is necessary to differentiate the time budget and take-over scenario that is given to participants.

The concept of mental workload relates most strongly to the demand, characterizing the demand imposed by tasks on a person's limited mental resources (Moray, 1979; Wickens, 2008). Results of the analysis showed that the mental workload depended on the experimental condition. The anxiety condition was rated to have the highest mental workload, and the boredom condition was rated to have the lowest mental workload. The flow condition was rated to have a modest mental workload, and the difference between the boredom condition and flow condition was significant. The mean mental workload increased with the demand level of the tasks. These results suggested that the mental workload increased as the demand level of the task increased.

All tasks require thought or mental workload. Therefore, the influence of tasks on driving was examined from the viewpoint of the mental workload. For example, an increase in mental workload due to mental calculations has been shown to increase average reaction time (Makishita and Matsumaga, 2005). Cantin et al. (2009) showed that reaction time increased as the complexity of driving contexts increased. Furthermore, the driver reaction time increased along with task difficulty. Salvia et al. (2016) showed that the mean reaction time increased as the task difficulty increased. According to these previous studies, it could be expected that the increase in mental workload induced by task difficulty would increase the reaction time. Therefore, the driver reaction time was expected to be the highest in the anxiety condition, where the demand level of the task is the highest. However, in contrast to previous studies, all variables of driver reaction time in the flow condition were higher than those in the other conditions. The flow condition has a lower perceived demand level than the anxiety condition and has the highest flow experience. Thus, these results suggested that the flow experience had an effect on the driver's reaction time.

We conducted a Roy–Bargmann Stepdown F-test to investigate the impact of the experimental condition on the individual reaction time variables. Significant effects of the experimental condition were obtained for gaze-on time, road-fixation time, and take-over time; however, no significant effects were observed for the hands-on time. The hands-on time was analyzed using gaze-on time and road-fixation time as covariates. However, no significant effect of the experimental condition was observed. This result substantiated the findings of previous studies on driver motor readiness, which allowed the driver to intervene in vehicle control (Zeeb et al., 2015). When drivers were requested to take over control of the vehicle, they redirected their gaze from the NDRTs to the roadway and returned their hands to the steering wheel. In doing so, drivers established motor readiness (Zeeb et al., 2016). Zeeb et al. (2015) showed that the establishment of motor readiness (time to hands on) was unaffected by driver distraction.

Further, we tested the difference between the effects of N-back and the other experimental conditions on the reaction time. The gaze-on time of the driver in the N-back test was 1.440 s, 1.258 s in the flow condition, 1.030 s in the boredom condition, and 1.011 s in the anxiety condition. Thus, the gaze-on time was the highest in the N-back, and it was not significantly different from that in the flow condition. However, the difference between the gaze-on time in the N-back and those of the other two conditions (boredom and anxiety) was statistically significant. The other three variables of reaction time (road-fixation time, hands-on time, and take-over time) were the highest in the flow condition, and they did not vary significantly from those in the N-back. However, statistically significant differences were observed for these variables between the N-back and the other two conditions. The results of the Roy–Bargmann Stepdown F-test suggested that the N-Back and flow condition have a similar effect on the driver reaction time. In other words, flow experience and mental workload similarly affect the driver reaction time.

Eye-movement measurements consist of different properties of movement during a finite period of time. The properties of movement are direction, amplitude, duration, velocity, and acceleration (Holmqvist et al., 2011) (p. 356). In this study, we used position measures that pertain to the location of the participants' gaze and to the movement properties. The gaze-on time described the first fixation duration after the onset of the stimulus ("How long was the first fixation on the stimulus?"). The gaze-on time coincided with the very first intake and processing of the stimulus, and its duration reflected immediate information processing (Holmqvist et al., 2011) (p. 384). The road-fixation time described the first fixation duration in a given area of interest (AOI) ("How long was the first fixation in an AOI?"). The road-fixation time was reflected by the time required for fast processes, such as recognition and identification (Holmqvist et al., 2011) (p. 385). This study showed that the gaze-on time and road-fixation time were the highest in the fit (flow) condition. These results suggested that the flow experience affected the driver's immediate information processing and recognition.

This study has several practical limitations. First, there are limitations due to the number of participants and their ages. The same group of 32 subjects participated in both Experiment 1 and Experiment 2. As the study was conducted on a small number of participants, it would be premature to expand the results of the study or to generalize the interpretation of the results. In particular, the age group was limited to participants in their 20s and 30s. Compared to the number of participants who were in their 20s (26 people), the number of participants in their 30s (6 people) was relatively small. This limited the impact of age in analyzing the data. The limitations were shown in the results of the analysis on the impact of age-related factors on the perceived demand level, flow experience, and mental workload tested in Experiment 1. Thus, in conducting future studies, it is necessary to expand the age scope to middle-aged and elderly persons who have relatively longer driving experience.

Subjective methods were used in this study to measure the degree of flow experience of the participants performing NDRTs with different demand levels. In many driving-related studies, subjective methods and physiological methods, such as heart rate variability (HRV), electrocardiogram (ECG), and electroencephalogram (EEG), are used to analyze the body conditions of the driver. Recently, Tozman et al. (2015) experimentally showed the relationship between HRV and flow experience. In their experiment, participants were exposed to computer-simulated tasks with different demand levels: a balanced skill–demand level to induce flow, high-demand level to induce anxiety, and low-demand to induce boredom. Their results show that experiencing flow was associated with a decreased low-frequency HRV (LF-HRV). They also found an inverted u-shaped relation between both high-frequency HRV (HF-HRV) and LF-HRV, and the flow experience when participants performed tasks whose demand levels exceeded their skill levels (Tozman et al., 2015). In future studies, physiological methods should be used, along with subjective methods, to measure the degree of the flow experience of a driver more objectively and present more objective results.
References


