Interruptibility for In-vehicle Multitasking: Influence of Voice Task Demands and Adaptive Behaviors

AUK KIM, KAIST, South Korea
JUNG-MI PARK, Samsung Research, Samsung Electronics, South Korea
UICHIN LEE*, KAIST, South Korea

As a countermeasure to visual-manual distractions, auditory-verbal (voice) interfaces are becoming increasingly popular for in-vehicle systems. This opens up new opportunities for drivers to receive proactive personalized services from various service domains. However, prior studies warned that such interactions can cause cognitive distractions due to the nature of concurrent multitasking with a limited amount of cognitive resources. In this study, we examined (1) how the varying demands of proactive voice tasks under diverse driving situations impact driver interruptibility, and (2) how drivers adapt their concurrent multitasking of driving and proactive voice tasks, and how the adaptive behaviors are related to driver interruptibility. Our quantitative and qualitative analyses showed that in addition to the driving-task demand, the voice-task demand and adaptive behaviors are also significantly related to driver interruptibility. Additionally, we discuss how our findings can be used to design and realize three types of flow-control mechanisms for voice interactions that can improve driver interruptibility.

14

CCS Concepts: • **Human-centered computing → User interface management systems**; **Ubiquitous and mobile computing**.

Additional Key Words and Phrases: Interruptibility; In-vehicle information system; Human-vehicle interaction; Auditory-verbal interface

ACM Reference Format:
Auk Kim, Jung-Mi Park, and Uichin Lee. 2020. Interruptibility for In-vehicle Multitasking: Influence of Voice Task Demands and Adaptive Behaviors. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 1, Article 14 (March 2020), 22 pages. https://doi.org/10.1145/3381009

1 INTRODUCTION

New technologies in intelligent vehicles and the popularity of auditory-verbal interfaces (or voice interfaces) provide new opportunities for serving drivers with system-initiated or proactive voice services [40, 46]. This opens up a whole new realm of possibilities for drivers to be connected to various service domains and receive proactive personalized services, ranging from information delivery to decision-making, according to their needs [50]. While proactive voice interactions benefit drivers, prior studies also warned the potential risk of *cognitive distractions* [22, 62]. This is due to the nature of concurrent multitasking with a limited amount of cognitive resources.

*This is the corresponding author

Authors' addresses: Auk Kim, KAIST, Daejeon, South Korea, kimauk@kaist.ac.kr; Jung-Mi Park, Samsung Research, Samsung Electronics, Seoul, South Korea, jungmi.park@samsung.com; Uichin Lee, KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon, 34141, South Korea, uclee@kaist.ac.kr.

2474-9567/2020/3-ART14 $15.00
https://doi.org/10.1145/3381009

For their safety, drivers are required to perform concurrent multitasking [66], by simultaneously executing a primary task (i.e., driving) and secondary tasks (i.e., proactive services) [62] and sharing their limited cognitive resources for multitasking [71, 72]. If *concurrent multitasking* presents greater demands than the available cognitive resources, the performance of either one or both tasks can be degraded (i.e., dual-task interference) [71, 72]. Drivers can reallocate their cognitive resources from one task to another; however, there is a *performance trade-off* between competing tasks [54, 55, 71], and *adaptive behaviors in multitasking* (e.g., speed reduction) are likely to be observed [58, 64]. This *contextual uniqueness of in-vehicle multitasking* highlights the importance of examining and finding in-vehicle opportune moments (or driver interruptibility) in which the negative effects of concurrent multitasking can be minimized (e.g., when drivers can safely engage in proactive voice services successfully).

In the domains of ubiquitous computing and human-computer interaction, there has been an active research area that specifically focuses on understanding and identifying *opportune moments.* To find such moments, research has attempted to quantify to what extent users are interruptible (or being able to engage in an interruption) for a given context. The term *interruptibility* refers to the quality of being interruptible. Traditionally, these studies have considered mostly desktop computers and mobile device environments [1, 4, 5, 12, 67]. In these environments, users can sequentially perform multitasking for their ongoing and interrupting tasks; users can fully switch their visual and/or manual operations between tasks. While these studies offer important insights, their findings are not directly applicable to vehicular environments because, in driving contexts, task-switching induces various unfavorable impacts on drivers (e.g., off-road glance behavior [30, 34]). Technology usage that requires demanding visual-manual operations of drivers is generally considered to be unsafe and has been banned in many countries (e.g., mobile phone use during driving) [11]. For these reasons, voice interactions as an alternative to visual-manual interactions have recently received significant attention.

Research involving concurrent multitasking in driving contexts, for example, driver distraction research, often asks the drivers to prioritize and maintain one task over the other while performing multitasking (see Section 2.2) [8]. While this approach is useful to assess the dual-task interference and the workload of drivers in controlled scenarios, the generalizability of the findings to interruptibility research is limited. According to a well-known driver behavior model [42], in reality, drivers actively involved in *conflict management* in multitasking scenarios dynamically reallocate their attentional resources between driving and secondary tasks based on the changes in each task's performance and driving environments (see Section 2.1). It is important to observe how a driver *naturally* performs multitasking while interacting with conversational agents and performing conflict management in-situ. We hypothesize that by observing naturalistic conflict management behaviors, we can identify driver behaviors that are related to driver interruptibility. Answering this hypothesis lays a foundation for building an intelligent system that can automatically infer interruptibility by monitoring the driver's behaviors with sensors and interaction log data.

In this study, we considered proactive voice interactions, and quantitatively and qualitatively examined (RQ1) *"How do the varying demands of proactive voice tasks under diverse driving situations impact the driver interruptibility?"* and (RQ2) *"What kinds of adaptive behaviors do the drivers show, and how are the adaptive behaviors related to interruptibility?"* This work builds upon the research by Kim et al. [31], which introduced the composite model of interruptibility (i.e., driving safety, auditory-verbal performance, and overall perceived difficulty). For interruptibility measures, we consider the composite model of interruptibility [31] that captures "the quality of being safe, successful, and easy (to interact with) while engaging in voice interaction with in-vehicle systems.'' This work differs from that of Kim et al. [31] in that they focused on building predictive modeling, and fine-grained and descriptive analyses, as well as theoretical grounding, are lacking. In addition, their work did not address the aforementioned research questions (RQ1 and RQ2).

Our results indicate that driver interruptibility significantly varies across the driving maneuver types and the secondary-task demands. Furthermore, we discovered that drivers employ diverse adaptive behaviors to manage their interruptibility as well as managing their driving safety. Our findings highlight that driver interruptibility research should carefully consider the potential impact of dynamic or adaptive multitasking behaviors under diverse secondary-task demands and driving maneuver conditions. Based on these findings, we discuss practical in-situ measures for real-world deployments and design implications, such as the concept of flow control mechanisms for proactive voice interactions that can potentially improve driver interruptibility for both user-initiated and system-initiated voice interactions.

This paper is organized as follows. Section 2 describes the background and related work. Section 3 describes the driver interruption dataset. The analysis method is presented in Section 3.6. The results for RQ1 are presented in Section 4 and the results for RQ2 are delivered in Section 5. This paper reports the main findings in Section 6 and the conclusion is delivered in Section 7.

2 BACKGROUND AND RELATED WORK

2.1 Hierarchical Driver Behavior Model and Conflict Management

Lee and Strayer [42] proposed a control theory-based hierarchical driver model (see Chapter 3 in Regan et al. [62]). In this model, drivers control or make decisions related to their driving and secondary tasks based on their observation of changes in the ongoing performance and the overall demand (workload) of their driving and secondary tasks. A modified conceptual diagram with decision-making criteria for interruptibility is presented in Figure 1. The hierarchical driver model involves three levels of driver decision making: (1) strategic decisions (e.g., which route to choose), (2) tactical (navigational) decisions (e.g., which lane to choose, whether to engage in a secondary task), and (3) operation (motor control) decisions (e.g., speed control in response to particular stimuli such as road signs).

At the tactical level (see ① in Figure 1), a driver decides whether to engage in a secondary task. Once the driver decides to engage, he or she makes decisions for speed, lane choice, and resource allocation based on the previous observation results. Next, in the operation level (see ② in Figure 1), according to the tactical-level decision, the driver controls their vehicle and performs resource investment. At the same time, the driver observes the changes in the ongoing performances (e.g., headway-distance, lane-keeping, and secondary task performance) and the driving and secondary task demands. When observing any conflicts (e.g., performance decrements and/or expecting an increase in demand levels), the driver makes a new tactical decision (i.e., conflict management; see ③ in Figure 1). For the new tactical decision, the driver typically considers the following approaches: e.g., speed/lane choices, secondary task engagement/disengagement, and resource allocation policy selection. These imply that the driving task performance, secondary task performance, and overall workload (or the overall demands of both tasks) are important criteria for drivers to make a decision to engage or continuously engage in a

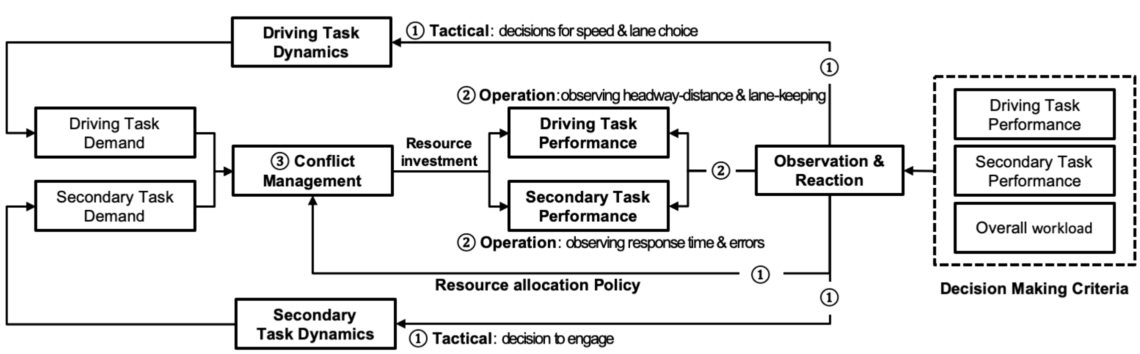

Fig. 1. Control theory-based hierarchical driver model with decision making criteria for interruptibility.

secondary task while driving. For this reason, we considered Kim et al. [31]'s composite model of interruptibility as decision-making criteria for interruptibility. The dimensions include driving safety, auditory-verbal performance, and overall perceived difficulty. The details of the dimensions can be found in Section 3.4.

2.2 Existing Experimental Frameworks

Existing experimental frameworks mostly aim to assess a driver's mental workload for concurrent multitasking of two tasks, by measuring the performance of one task whereas the driver is asked to prioritize and maintain the other task [8]. This approach is known as *secondary-task methodology* because the secondary task workload is manipulated intentionally to create mentally overloaded situations such that the driver's mental workload can be indirectly measured. The use of two concurrent tasks, or concurrent multitasking, is mandatory because in under-loaded situations, single-tasking cannot show any performance variation (or degradation) according to the multi-resource theory [71]. Depending on which task is prioritized, two different approaches were considered. When a driver is asked to maintain a given driving task, the secondary task performance (e.g., target tracking) can be measured to estimate the mental workload (known as the subsidiary task framework). When the driver is asked to maintain a given secondary task, the primary task performance (e.g., lane-keeping) can be measured to estimate the mental workload (known as the loading task framework).

Unfortunately, these well-known experimental frameworks are less adequate for examining a driver interruptibility. The primary reason is that these frameworks artificially manipulate a driver's natural decision-making process. To be precise, a driver's decision-making policy for conflict management is pre-determined by the experimenters (e.g., prioritizing either a primary or secondary task). Therefore, it is difficult to observe a natural decision-making process, which will then be translated into driver behaviors. Prior studies hinted that observing naturalistic driver behaviors is important. This is because drivers have difficulty in prioritizing one task over another task [44, 51]. In addition, driver behaviors are not always rational since drivers suffer from various cognitive biases (e.g., capability overestimation, risk-taking behaviors) [24]. In reality, conflict management is likely to happen in a dynamic fashion, and interdependence between primary and secondary tasks exists. To summarize, the major departure from existing studies is that our goal is not about workload assessment in a controlled setting; instead, we focused on interruptibility assessment in a naturalistic setting.

2.3 Multitasking Demands and Driver Interruptibility

Despite the importance of concurrent multitasking in driving contexts and the popularity of voice secondary tasks such as proactive services, only a few studies have examined in-vehicle opportune moments towards voice secondary tasks [7, 60, 68]. These studies have considered the effect of primary-task demands (or driving demands) on driver interruptibility. However, prior studies did not consider the secondary-task demand (e.g., the workload of the proactive voice services) [7, 60, 68]. For example, Semmens et al. examined how the driving demand affects interruptibility by measuring the vehicle status and environmental contexts (e.g., steering wheel angle and speed) [68]. However, they did not vary the demand for secondary tasks; thus, their work did not examine the effects of the variation on driver interruptibility (or availability, in the context of this study).

The contextual uniqueness of in-vehicle multitasking implies that secondary-task demand also affects the workload of drivers [71, 72] and can thereby impact the driver interruptibility. Studies on driver distractions have also considered the demand for both tasks because the impact on the driving performance highly depends on the aggregate workload of the tasks [62]. These findings imply that the demand for secondary tasks can affect interruptibility. However, the effects of varying the demand levels of voice secondary tasks in diverse driving contexts on the driver interruptibility have not been examined in detail. Therefore, we attempt to answer the following question (RQ1): *How do the varying demands of proactive voice tasks under diverse driving situations impact driver interruptibility?*

2.4 Adaptive Behaviors and Driver Interruptibility

In the driving context, concurrent multitasking with limited cognitive resources and prioritization of driving safety influence drivers to employ adaptive behaviors, such as speed reduction [64]. In particular, drivers change their driving behavior to maintain their driving performance and to compensate for any additional workload that is imposed by a secondary task. This is especially true when the cognitive demand for driving is high [58]. Research on driver behavioral adaptation has shown that drivers display various types of adaptive behaviors, such as speed reduction [27, 49, 58], maintaining a longer headway [49], reducing lane changes or overtaking [6, 23], and minimizing the attention for secondary tasks [20, 27].

Although these studies generally examined the relationship between adaptive behaviors and the accident likelihood or driving performance, these findings also suggest that adaptive behaviors may help to increase interruptibility towards voice-based secondary tasks. This is because adaptive behaviors may reduce the workload of the primary task while performing secondary tasks [15, 27, 64]. For example, let us consider an adaptive behavior where drivers reduce their driving speed when performing a secondary task [27, 49]. While drivers reduce the speed, the time to make a decision and react to changes in the driving environment increases, and driving tasks require lower cognitive demands. Thus, drivers can balance workload levels between conflicting tasks since the remaining cognitive resources can be used for a secondary task. These imply that a speed reduction would benefit drivers in terms of interruptibility. However, at the same time, it needs to be determined what kinds of adaptive behaviors drivers typically show and whether they can still employ adaptive behaviors sufficiently in the real world. This is because prior research has observed and examined adaptive behaviors mostly in simulator-environments that are less demanding than the real world [9, 62].

Prior studies on driver interruptibility have not examined driver behaviors in concurrent multitasking contexts as well as the impact of adaptive behaviors on interruptibility [7, 52, 68]. For example, Bellet et al. [7] examined how environmental contexts affect the driver's capability to receive vocal information. However, they did not examine the changes in driving behaviors while engaging in the tasks (or adaptive behaviors) and their relationship with interruptibility. Accordingly, we attempt to answer the following question (RQ2): *What kinds of adaptive behaviors do the drivers show, and how are they related to interruptibility?*

3 DRIVER INTERRUPTIBILITY DATASET

To address the research questions, we used the in-vehicle interruptibility dataset collected from the study by Kim et al. [31]. The data were collected from 29 drivers while they were engaging in proactive auditory-verbal (or voice) secondary tasks with various interruption contexts such as three varying secondary tasks and diverse driving contexts.

3.1 Voice Secondary Task

As shown in Figure 2, the procedure of the task consists of three stages: 1) asking, 2) interacting, and 3) measuring. To induce three levels of cognitive demand, the auditory-verbal *n*-back task [48] was considered in the interacting stage. This is because it 1) has a similar cognitive engagement when a driver engages in an externally paced (or system initiated) voice interactions, 2) induces systematically structured levels of cognitive demand, and 3) has been widely used as a reference task to study and examine in-vehicle voice interfaces in a large number of real-road driving studies [48]. This includes projects carried out by the U.S. National Highway and Transportation Safety Administration (NHTSA) [61] and the International Standard Organization [10].

At the asking stage, a driver is asked if he is willing to engage in the remaining stages. If the driver answers *"Yes"* within 2.5 s, the remaining stages are presented. Otherwise, the current task ends immediately. The asking stage was introduced to ensure that any interactions are performed under safe conditions.

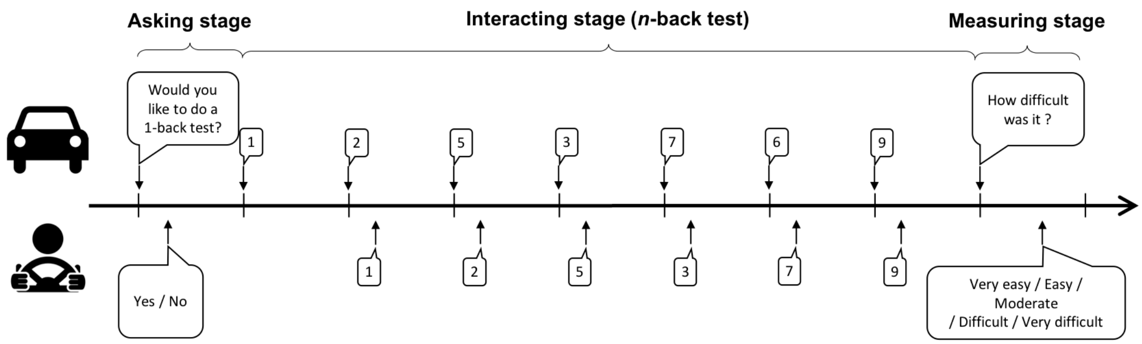

Fig. 2. Procedure of a voice secondary task. *n*-back types = 1-back (from Kim et al. [31])

At the interacting stage, the driver is asked to perform the *n*-back task. The *n*-back task sequentially presents seven randomly selected numbers from 0 to 9 to the driver at 2.25-second intervals. After each number is presented, the driver is required to verbally repeat back the *n*-th digits by performing one of the three following tasks (for more details, see [48]):

- *0-back (very mild task demand)*: repeating each number after it is presented
- *1-back (moderate task demand)*: repeating the number one item back in the sequence
- *2-back (high task demand)*: repeating the number two item back in the sequence

At the measuring stage, the driver is asked to verbally indicate the overall perceived difficulty of concurrent multitasking during the interacting stage. Asking the overall multitask difficulty immediately after completing the task does not interfere notably with the primary task performance [17].

3.2 Driving Course and Task Triggering Method

Figure 3 shows Route A and B of the round-trip driving course (18.92 km) that were employed for the data collection. Route B (9.21 km) is the return route and slightly differs from Route A (9.71 km) due to differences in traffic rules. The course was designed to reflect various driving-environment factors. The factors include various driving maneuvers (for each route, at least four left turns, four right turns, four lane changes, and one U-turn), differing levels of traffic density (traffic-heavy roads: 3.82 km, yellow zone), and other potential factors (high pedestrian traffic: 0.5 km, green zone; school zones: 1 km, red zone). These have been suggested as the most important factors to determine the driving environment complexity and its effects on a driving task's demands [22].

The green and red flags indicate the starting and end points for Route A. The points are opposite for Route B. Both routes require approximately 30 minutes of driving time. The light-orange flags show the 20 predetermined locations where secondary tasks were consistently presented for each route; in total, there were 40 locations for the round-trip driving course. This design ensured that a reasonable number of tasks were presented across the maneuver types. In addition to the presentation at 40 predetermined locations, secondary tasks were also presented at random intervals between 30 and 90 s when the distance to the next predetermined location was sufficiently long. This ensured that the distribution of tasks reflected realistic road situations, such as traffic jams.

3.3 On-Road Data Collection and Exit-Interview

Twenty-nine drivers participated in the data collection. The average age of the drivers was 43.2 (SD = 11.4, range = 19–64). Fifteen drivers were females (males: n = 14). All drivers had a valid driving license, at least one year of

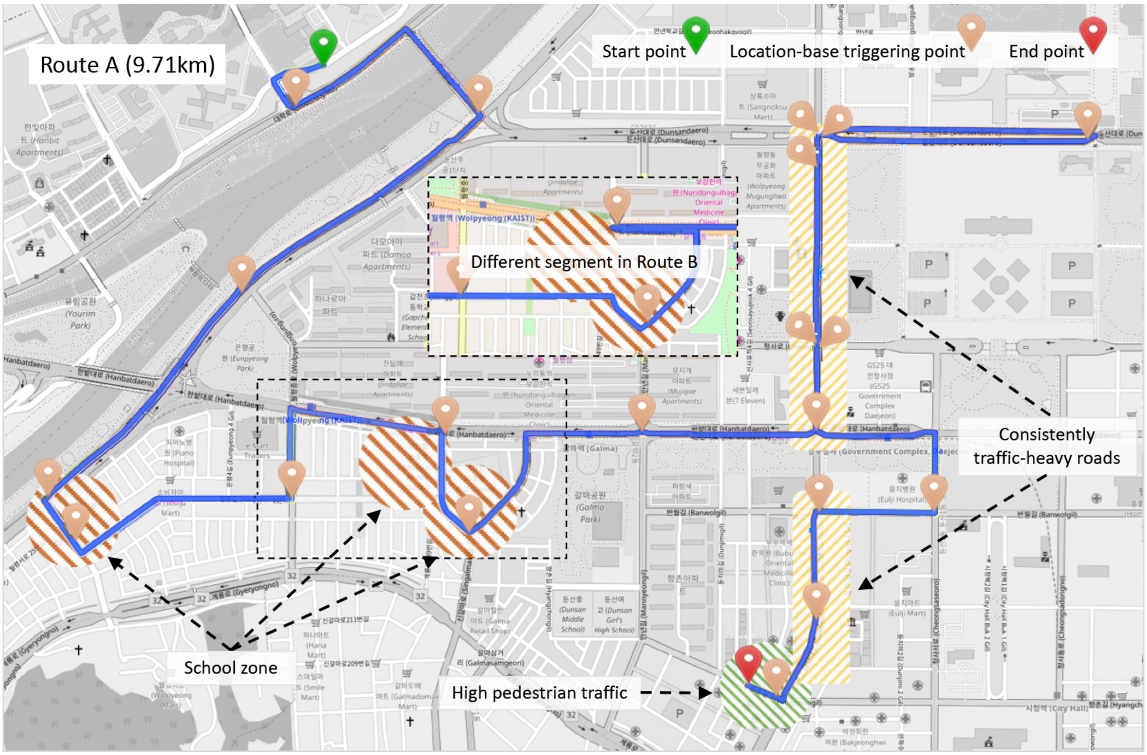

Fig. 3. Route A and B in the round-trip driving course. Route B is slightly different from Route A. See the dotted box for the different part (from Kim et al. [31])

driving experience (M = 13.4, SD = 8.7, range = 1–39), and they drive at least 30 minutes every day. They drove a similar sized car (Kia K3, a popular midsize sedan in Korea) that was used for data collection. Approximately 60% of the driving course consisted of well-known roads; the remaining roads were not familiar to the drivers.

For the data collection, drivers were asked to drive a round-trip driving course twice (two driving sessions): once for the baseline session and once for the secondary-task session. The baseline session did not involve secondary tasks, and it was used as a baseline to compare the driving performances. Prior to each session, the participant drove to the campus and adapted to the driving. For the secondary-task session, a training session of n-back tasks was additionally conducted.

During driving sessions, one researcher was seated in the rear seat of the vehicle to ensure safe vehicle operation and experimental instrument checking. The open-source navigation app (OsmAnd [69]) was modified to guide the next directions of the driving course at least 10 s before or after presenting the secondary tasks to avoid interruption of the turn-by-turn navigation instructions to the secondary tasks. In-car audio systems were used as the main speakers.

After two driving sessions, an exit interview was conducted. During the interview, the drivers were asked (1) what the most/least difficult circumstances were, (2) how they engaged in secondary tasks, and (3) why they had answered incorrectly for a given case after reviewing which circumstances they were in.

3.4 Ground Truth: Composite Model for Driver Interruptibility

The contextual uniqueness of in-vehicle multitasking and its impacts on drivers emphasize the importance of considering both objective and subjective measures in the definition of driver interruptibility. The composite model of driver interruptibility considers both objective and subjective measures of driver interruptibility: 1) driving safety, 2) auditory-verbal performance, and 3) overall perceived difficulty. In this work, for the sake of simplified analyses, we assume that each dimension indicates interruptibility as a binary outcome (i.e., interruptible or uninterruptible). A summary of each dimension is provided as follows:

- *Driving safety* measures 'how safely a user drives a vehicle.' It indicates "interruptible" if, while concurrently executing driving and secondary tasks (*n*-back task), and the driving performance is not decreased in comparison with that when driving without secondary tasks (baseline). As a driving performance indicator, the steering wheel reversal rate (SRR) [62] was used because it has been widely used in a real-road driving setting as opposed to other metrics, such as the brake jerks, which are largely influenced by a small variation in the traffic conditions [57]. The SSR measures the frequency of the steering wheel reversal (or steering wheel correction) events whose angles exceed a certain minimum angular value (or the gap size). The visual load and cognitive load induce different patterns of reversal angles [38, 57]. Small reversals (gap size = 0.1 degree) were used to achieve the highest sensitivity for the cognitive load of voice secondary tasks [57].
- *Auditory-verbal performance* measures 'how well a user performs a voice secondary task' based on the accuracy of the *n*-back task. It indicates "interruptible" if all of the items in a given *n*-back task are answered correctly.
- *Overall perceived difficulty* measures 'how difficult it is to perform a dual-task.' This is based on the answer (very easy: 1; very difficult: 5) when rating a task during the measuring stage (or after the *n*-back task) of the voice secondary task. It indicates "interruptible" if the normalized value of the rating was less than or equal to 0.5.

3.5 Maneuver Type Labeling and Descriptive Statistics

The data consisted of 1,388 secondary-task cases (M = 47.86, SD = 6.83 per driver). Since each case often involved more than one maneuver type, we labeled four representative maneuver types for each case based on the vehicle GPS trajectory and street view: STOP, STRAIGHT, TURN, and LANE CHANGE. Three human annotators labeled the maneuver types according to the criteria in Table 1. we removed 28 cases (0-back = 9, 1-back = 9, 2-back = 10) that did not fall in any category. After the removal, six drivers had missing cases for one of the *n*-back type conditions that were associated with either the TURN or LANE CHANGE because the entire cases for the conditions were removed as they included both TURN and LANE CHANGE. The number of drivers for each missing case are as follows: *TURN* $\times$ *0-back* = 2 , *TURN* $\times$ *2-back* = 1, *LANE_CHANGE* $\times$ *0-back* = 2, *LANE_CHANGE* $\times$ *2-back* = 1. In total, 1,360 remained. The number of interruptible cases across the interruptibility dimensions were 1,099, 1,168, and 1,218 in terms of overall perceived difficulty, auditory-verbal performance, and driving safety, respectively. The frequency of the remaining cases across the *n*-back types and maneuver types for a driver are listed in Table 2. The percentage of interruptible cases across the *n*-back types and maneuver types for a driver are presented and discussed in Section 4.

3.6 Analysis Method

In order to answer the research questions, statistical and content analyses were conducted. For the *statistical analysis*, we performed Repeated Measures ANOVAs. The dependent and independent variables differ across each analysis. The details are presented in each corresponding section. For the ANOVAs, we adjusted the degree of freedom according to the significance of Mauchly's Test of Sphericity. For repeated tests in the post-hoc analysis, this study adjusted the p-values based on the Bonferroni correction [26]. For the effect size, η^2 (Eta squared) was

Table 1. Labeling criteria for maneuver types

	Being stationary for the entire period of the secondary task	Percentage of stationary time < 25% of total time	Do not include following maneuver types
STOP	Yes	-	Straight, Turn, Lane change
STRAIGHT	No	Yes	Turn, Lane change
TURN	No	Yes	Lane change
LANE CHANGE	No	Yes	Turn

Table 2. Average frequency of task cases across the *n*-back types and maneuver types for a driver.

Mean (SD)		**Overall**	**Maneuver type**			
			Stop	**Straight**	**Turn**	**Lane change**
Overall		46.90 (6.96)	12.41 (3.91)	19.14 (4.76)	8.66 (2.57)	6.69 (2.51)
***n*-back types**	**0-back**	15.52 (4.45)	4.24 (2.03)	5.45 (2.40)	3.00 (1.83)	2.52 (1.35)
	1-back	16.10 (3.87)	3.69 (1.69)	7.03 (2.49)	2.86 (1.27)	2.21 (1.40)
	2-back	16.24 (3.57)	4.48 (1.77)	6.66 (2.61)	2.79 (1.52)	1.97 (1.21)

reported [63]. For missing cases, we used a well-known method called regression imputation, which produces the estimates closest to those of the original variables [53]. For the imputation, we used a hierarchical linear regression (multilevel model), while applying maximum likelihood estimations and grand-mean centering to prevent convergence issues by reducing the multicollinearity [35, 56]. For the *content analysis*, we transcribed the interview records and used affinity diagramming to uncover how participants performed concurrent multitasking (RQ1) and what types of adaptive behaviors occurred (RQ2) [47].

4 RQ1: HOW DOES THE VARYING DEMAND OF PROACTIVE VOICE TASKS UNDER DIVERSE DRIVING SITUATIONS IMPACT DRIVER INTERRUPTIBILITY?

In this section, to answer the first research question, we analyzed how the varying demands of the proactive voice tasks under diverse driving situations impact the following subjective and objective measures of interruptibility: 1) perceived overall difficulty, 2) auditory-verbal performance, and 3) driving safety. We performed three two-way Repeated Measures ANOVAs. As demonstrated in Figure 4, for the independent variable of each ANOVA, we used the percentage (or empirical probability) of interruptible moments across each interruptibility dimension. For the dependent variables, we consistently used the secondary-task demands and maneuver types. The secondary-task demands indicate the cognitive demand of the *n*-back task in the secondary task: very mild level (0-back), moderate level (1-back), and high level (2-back). For maneuver types, we considered the four representative maneuver types: STOP, STRAIGHT, TURN, and LANE CHANGE. It should be noted that the maneuver type is the major factor that determines the driving workload [22]. The result of each analysis is presented as follows:

4.1 Overall Perceived Difficulty

We first analyzed how the secondary-task demands and the maneuver types affect the percentage of interruptible moments in terms of the overall perceived difficulty. An increase in the percentage indicates that the drivers are more likely to perceive concurrent multitasking (driving and proactive voice interaction) as easy. For convenience and better comprehension, this will be referred to as an increase in the percentage as a decrease in the overall

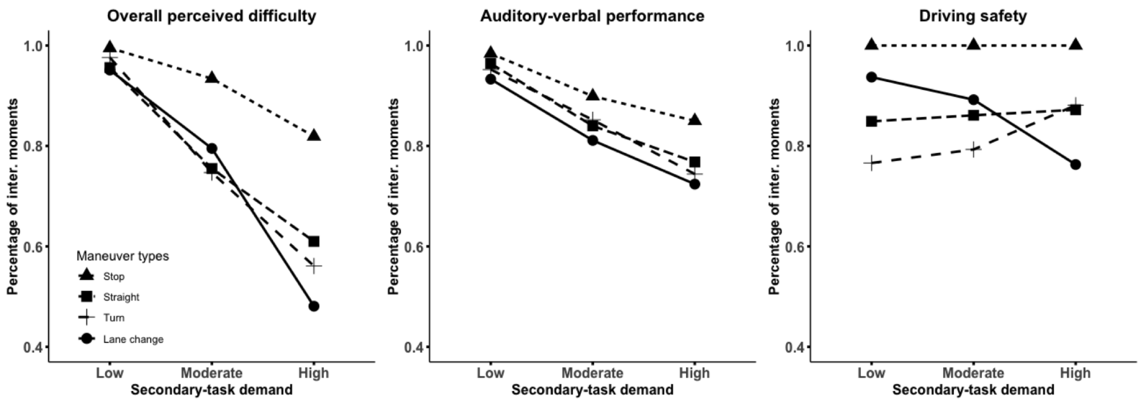

Fig. 4. Percentage of interruptible moments for each interruptibility dimension across secondary-task demands and maneuver types. Low (0-back), moderate (1-back), high (2-back)

Table 3. Statistical results for the percentage of interruptible moments for each interruptibility dimension across the secondary-task demands and maneuver types. Overall perc. diff. = overall perceived difficulty, Audit.-verbal perf. = auditory-verbal performance, F = F-value, p = p-value.

	Overall perc. diff.			Audit.-verbal perf.			Driving safety		
Effects	**F**	**p**	η^2	**F**	**p**	η^2	**F**	**p**	η^2
n-back types	31.157	< 0.001	0.53	10.903	< 0.001	0.28	0.061	0.94	0.00
Maneuver types	12.131	< 0.001	0.30	3.882	< 0.05	0.12	1.439	0.25	0.05
Interaction	3.363	< 0.01	0.11	0.306	0.93	0.01	3.198	< 0.05	0.10

perceived difficulty, whereas a decrease in the percentage is an increase in the overall perceived difficulty. The results show that drivers are more likely to perceive concurrent multitasking as more difficult as the secondary-task demand increases while the amount of difficulty increases for the following order of maneuvers: STOP < STRAIGHT < TURN < LANE CHANGE (most difficult). These suggest that, in addition to the driving demand, that secondary-task demand is also an important factor to determine the opportune moments from the driver's perspective.

As demonstrated in Table 3 (left), the statistical analysis shows that both the main effects and the interaction effect are significant. Furthermore, the post-hoc analysis showed that when either the *n*-back types were 0-back or the maneuver type was STOP, the difficulty did not significantly vary regardless of the changes for the main effects ($p = 0.49 - 1.00$). This result is unsurprising because in these cases, the cognitive demand of either driving or the secondary task was almost zero. These cases resemble those in which the driver performed a single task rather than a multitask. As the cognitive demand of the *n*-back types increased to 1-back, the STOP had a significantly lower difficulty than the rest of the maneuvers ($p < 0.05 - 0.01$), while the rest of the maneuvers had no significant difference between each other, with a p-value of 1.00 calculated for each pair. As the cognitive demand of *n*-back types increased to 2-back, in addition to the significant difference in STOP and resting maneuvers ($p < 0.001$), LANE CHANGE had a significantly higher difficulty than STRAIGHT ($p < 0.05$).

Our interview analysis revealed similar findings. In general, drivers stated that they perceived multitasking to be more difficult as the cognitive demand of the secondary task (*n*-back) increased. For example, D28 said, *"It was more difficult for one-back and two-back (than zero-back)."* D12 stated, *"Actually, up to one back, it wasn't really that hard for me."* In addition, when the secondary-task demand was low, drivers perceived concurrent multitasking to not be difficult regardless of the maneuver types. In contrast, when the demand was high, they perceived concurrent multitasking to be more difficult for the maneuver in following order: STOP < STRAIGHT < TURN < LANE CHANGE (most difficult). For example, as expressed by D8, *"Changing lanes was the most difficult, next in order were the right turn, left turn, and straight."* D7 said, *"Left turn as well as a right turn, when making turns, when changing lanes, zero-back and one-back weren't really what bothered me, but two-back was [different]. I was more tense."*

4.2 Auditory-verbal Performance

Next, we analyzed how the *n*-back types and maneuver types affected the percentage of interruptible moments in terms of auditory-verbal performance. An increase in the percentage indicates that drivers are more likely to maintain their auditory-verbal performance while engaging in a secondary task. For convenience and better comprehension, we will simply refer to the percentage as auditory-verbal performance.

As described in Table 3 (center), the analysis shows that only the main effects are significant. In addition, the post-hoc analysis indicates that, while there was a decreasing trend in the performance toward a higher secondary-task demand, the performance was significantly higher for 0-back than 1-back ($p < 0.05$) and 2-back ($p < 0.001$). However, it was not significantly different between 1-back and 2-back ($p = 0.20$). Furthermore, the performance was significantly higher for STOP than for a LANE CHANGE ($p < 0.05$); however, it was not significantly different for the other comparisons ($p = 0.11 - 1.00$).

These results are interesting because, owing to the limited cognitive resources during multitasking, dual-task interference, i.e., the decrement in the performance of either one or both tasks, is more likely to have occurred as the demand of either one or both tasks increases [71, 72]. This implies that the performance is more likely to vary among STRAIGHT, TURN, and STOP. From this, it is likely that the driving demand is higher for STRAIGHT and TURN than STOP. This finding was observed even when the secondary task demand was high (2-back).

4.3 Driving Safety

Finally, we analyzed the effects of *n*-back types and maneuver type on the percentage of interruptible moments in terms of driving safety. According to the driving safety measure, drivers are considered to be interruptible when their driving performance is not decreased while simultaneously executing both driving and secondary tasks compared to while executing the driving task alone. An increase in the percentage of interruptible moments indicates that the drivers are more likely to maintain their driving performance while engaging in a secondary task. For convenience and better comprehension, we will simply refer to the percentage as driving safety.

As illustrated in Figure 4 (right), when the maneuver type was STOP, the safety was unsurprisingly always perfect regardless of the *n*-back types because the drivers were not involved in "real" driving (in motion). It is worth noting that without comparisons, the safety for the STOP cases is always higher than the remaining cases involving other maneuver types. For the analysis, the STOP cases were removed because it violated the normality assumption.

The analysis shows a crossover interaction [45]; i.e., the interaction effect is significant, whereas the two main effects are non-significant. The crossover interaction indicates that the effect of the secondary-task demand on driving safety was opposite depending on the maneuver type. Namely, when the maneuver type was STRAIGHT, driving safety was not affected regardless of the increments in the secondary-task demand ($p = 0.61 - 0.81$). This was also true for TURN ($p = 0.13 - 0.73$). However, when the maneuver type was LANE CHANGE, driving safety decreased as the secondary-task demand increased from 1-back to 2-back ($p < 0.05$).

The outcome of the crossover interaction is surprising because, as previously mentioned, the safety is more likely to decrease as the secondary-task demand increases. In addition, when the maneuver type was TURN, there was an increasing trend in the safety toward a higher secondary-task demand, although the trend was not statistically significant. A possible explanation for these outcomes is a trade-off performance between driving and secondary tasks, which can be considered as an adaptive behavior of the drivers. This is discussed in more detail in the following section (Section 4.4).

4.4 Discussion

Our results highlight that in addition to the driving demand, secondary-task demand also significantly affects driver interruptibility, regardless of interruptibility dimensions. Furthermore, the percentage of interruptible moments (or empirical probability of being interruptible) differs considerably across interruptibility dimensions, especially between driving safety and other interruptibility dimensions. As the cognitive demand of the *n*-back types increased, the overall perceived difficulty significantly increased. The greatest difficulty was experienced when the maneuver types changed in the following order: STOP < STRAIGHT < TURN < LANE CHANGE (most difficult). However, when the maneuver type was STRAIGHT or TURN, the driving safety did not significantly vary or show any difference across the *n*-back types. Similarly, the auditory-verbal performance was not significantly different across the STOP, STRAIGHT, and TURN types.

Our correlation analysis for the interruptible moments across each pair of interruptibility dimensions also concludes that interruptible moments are varied across interruptibility dimensions. When interpreting the coefficient value according to the prior work [37], it was determined that there was a negligible correlation between driving safety and other types; driving safety vs. overall perceived difficulty: $r = 0.07$, $p < 0.05$; driving safety vs. overall perceived difficulty: $r = 0.05$, $p = 0.08$. There was a low correlation between the overall perceived difficulty and auditory-verbal performance: $r = 0.37$, $p < 0.001$. The difference in the percentage of interruptible moments across driver interruptibility dimensions could be attributed to *naturalistic* conflict management behaviors (see Section 2.1), in which drivers dynamically adjust their concurrent multitasking behaviors of driving and secondary tasks and reallocate their attentional resources between the tasks according to the changes in each task's performance and demand [42, 62]. By considering the outcomes, in Section 5, we investigated the adaptive behaviors of the drivers and their relationships with interruptibility.

5 RQ2: WHAT KINDS OF ADAPTIVE BEHAVIORS DO THE DRIVERS SHOW? AND HOW ARE THEY RELATED TO INTERRUPTIBILITY?

We first investigated the adaptive behavior of the drivers by analyzing the exit-interview data. For the commonly reported behaviors, an evaluation was performed to determine how they are related to interruptibility. We presented the interview results, followed by the statistical analysis results.

5.1 Adaptive Behaviors during Proactive Voice Tasks

Based on the exit-interview data, we revealed two major types of adaptive behaviors: reducing speed (primary task) and reducing or ceasing to engage in secondary tasks. Other minor adaptive behaviors include delaying the vehicle start and performing certain driving tasks in advance.

5.1.1 Speed Reduction. Drivers (n = 23, 79%) commonly reported that when they engage in these interactions, they reduced the vehicle's speed to increase driving safety and interruptibility for the interactions. For example, D18 said, *"(During the n-back) I think I was somehow more careful by slowing down."* D12 stated, *"I would drive faster and change lanes straight away if I wasn't on the n-back task, but didn't speed up while I was doing the n-back task."* Similarly, drivers stated that when the vehicle started (i.e., after a traffic light), they slowly increased the vehicle's speed to improve their interruptibility, as explained by D17 who indicated, *"I drove slowly at the beginning, [...] so I could pay more attention to the interaction (secondary task)."*

5.1.2 Secondary Task Adaptation: Reducing Engagement in Secondary Tasks. Drivers (n = 24, 83%) also commonly reported that they reduced their attention or ceased to engage in secondary tasks whenever they needed to pay more attention to the driving task. For example, D17 stated, *"It was kind of cheating though, but, sometimes, I didn't focus much on the n-back task."* Drivers naturally prioritized driving tasks as opposed to the secondary tasks for their driving safety, as explained by D9: *"you know, driving safety is always a top priority, so I wasn't concerned with how well I did on the n-back task [...] I wasn't disturbed while doing that because sometimes I ignored (n-back task) even though I would get a wrong answer."*

This behavior was a common reason for low interruptibility cases in which the drivers had answered incorrectly. When researchers verbally described the approximate location of the vehicle and the driving contexts to remind the drivers in which circumstance they were in when they answered incorrectly, they commonly gave reasons that were similar to what was given by D27, *"Because I was focusing more on driving (than the secondary task)."* As they paid relatively less attention to the secondary tasks, some drivers were not even aware of the correctness of their answers. For example, D23 said, *"I wouldn't even have known (that the answers were incorrect) if you wouldn't have told me that. I was not aware of that at all."* This tendency to not being aware of the correctness could be problematic in real-world scenarios because these drivers would not correct their past decision mistakes on their responses for in-vehicle services. Note that the corresponding quantitative results (e.g., percentage of incorrect cases) are presented in Section 4.2 and Section 5.2.2.

5.1.3 Other Adaptive Behaviors. Other minor adaptive behaviors used by several drivers include a delayed vehicle start and advanced preparation for interaction. Some drivers (n = 8, 28%) reported that they delayed the vehicle start, as explained by D12, *"I was looking at the left side to make a turn while listening to the sound (of a n-back task). I had made a turn after I was done with (n-back) without being aware."* Similarly, D15 said, *"When no one was behind me, I did everything (n-back) and started the vehicle. And even when someone was behind me, it was still okay to start a little later after the car in front of me had started. It was my trick."* Several participants (n = 7, 24%) prepared for the interactions by adjusting their driving behaviors in advance (e.g., advanced lane change) so they could expect another secondary task given that a new secondary task was presented at least for 30 s and at most 90 s after a prior secondary task. For example, D10 said, *"I changed lanes beforehand because once I had done the lane change, I could do (n-back) without any difficulty."*

5.2 Influence of Adaptive Behaviors on Interruptibility

This section analyzes how the major adaptive behaviors (i.e., speed reduction and secondary task adaptation) are related to interruptibility.

5.2.1 Speed Reduction. Drivers commonly reported that they reduced their speed when engaging in secondary tasks. We conducted two statistical analyses to examine 1) how the speed-reduction behaviors varied across the n-back types and the maneuver types and 2) how they are related to driver interruptibility. As the dependent variable for each analysis, we considered the speed difference by subtracting the average speed during concurrent multitasking from that of the 5 s time period before multitasking. For example, consider the average speeds of 45 km/h and 40 km/h before and during multitasking, respectively. In this case, the speed difference is -5 km/h. For the analyses, we removed any cases where the vehicle had not been in motion (0 km/h) within the time period of 5 s before and after multitasking.

As displayed in Figure 5, we first performed a two-way Repeated Measures ANOVA to investigate whether the mean of the speed difference varied across the maneuver types and n-back types. Afterwards, an additional statistical analysis was performed with the average speed as the dependent variable. As demonstrated in the

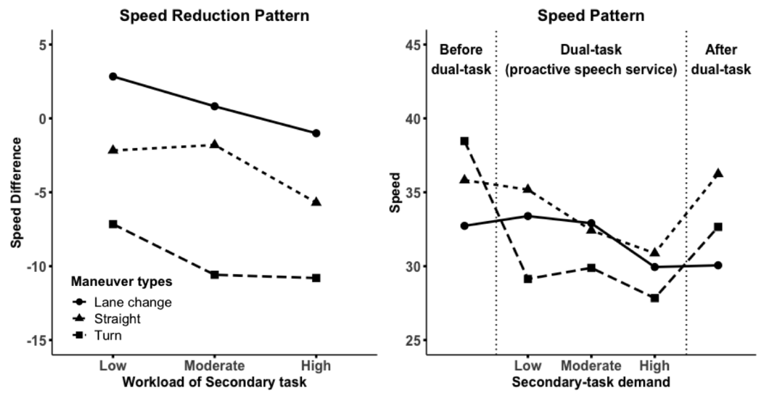

Fig. 5. Speed reduction pattern (km/h) and speed pattern (km/h) across the secondary-task demand and maneuver types

Table 4. Statistical results for speed reduction pattern and speed pattern across the secondary-task demand and maneuver types.

Effects	Speed reduction			Speed		
	F	p	η2	F	p	η2
Stage		N/A		71.063	< 0.001	0.72
Maneuver types	27.083	< 0.001	0.49	2.75	< 0.07	0.09
n-back types	3.920	< 0.05	0.12	0.19	0.93	0.01
n-back x Maneuver	0.425	0.72	0.01	0.239	0.92	0.01
Stage x *n*-back		N/A		5.025	< 0.05	0.15
Stage x Maneuver		N/A		17.996	< 0.001	0.39
Sta. x *n*-ba. x Man.		N/A		0.896	0.47	0.03

speed reduction column of Table 4, the results showed that both of the main effects are significant whereas the interaction effect is non-significant. Namely, the drivers reduced their speed to a greater extent for turns than for driving straight ($p < 0.001$) and changing lanes ($p < 0.05$). In addition, the demand of the secondary task increased (0-back vs. 2-back, $p < 0.05$).

For the additional statistical analysis, similar results were discovered (please refer to the speed column in Table 4). In particular, the average speed was lower during multitasking than before. In addition, the average speed during multitasking was lower for 2-back than 0-back ($p < 0.05$) and 1-back ($p < 0.05$). In addition, it was also lower for TURN than for STRAIGHT ($p < 0.001$) and LANE CHANGE ($p < 0.01$). These findings suggest that, regardless of the maneuver types, drivers are more likely to reduce their speed (or maintain a lower speed) when the secondary-task demand is high.

Next, we explored how the speed-reduction behavior of the drivers is related to their interruptions as depicted in Figure 6. We performed three two-way Repeated Measures ANOVAs with the mean of the speed difference as the dependent variable, and the maneuver types and interruptibility outcome (uninterruptible vs. interruptible) as the dependent variables. The types of interruptibility dimensions were varied across the ANOVAs. We only included cases with 1-back and 2-back, because most of the cases with 0-back were interruptible and produced a considerable number of missing values.

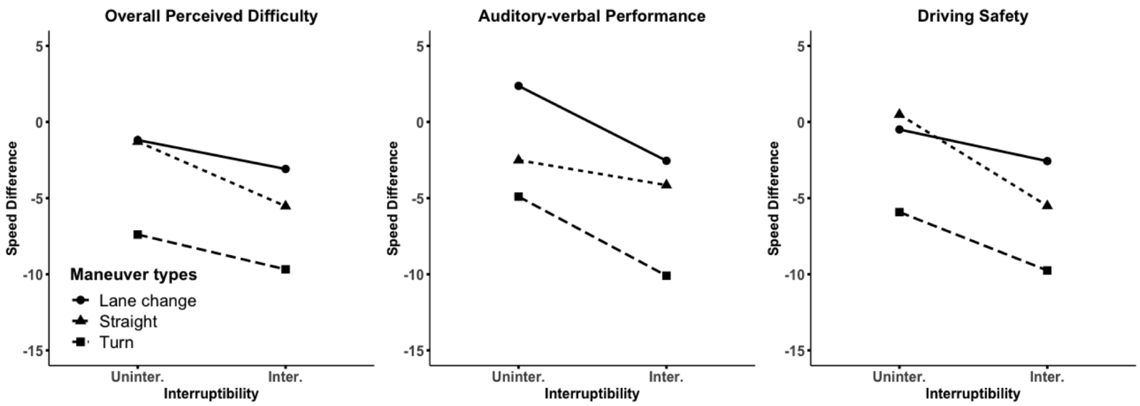

Fig. 6. Speed reduction pattern (km/h) across maneuver types and interruptibility. Uninter. = uninterruptible cases, Inter. = interruptible cases

Table 5. Statistical results for speed reduction pattern across maneuver types and interruptibility. Overall perc. diff. = overall perceived difficulty, Audit.-verbal perf. = auditory-verbal performance, F = F-value, p = p-value.

	Overall perc. diff.			Audit.-verbal perf.			Driving safety		
Effects	**F**	**p**	η^2	**F**	**p**	η^2	**F**	**p**	η^2
Interruptibility	8.470	< 0.001	0.23	10.721	< 0.001	0.28	7.539	< 0.001	0.21
Maneuver types	4.871	< 0.05	0.15	7.198	< 0.05	0.20	7.878	< 0.01	0.22
Interaction	0.240	0.79	0.01	0.804	0.45	0.03	0.927	0.40	0.03

As listed in Table 5, the analyses showed consistent results across the interruptibility dimensions, in which both the main effects are significant; however, the interaction effect is insignificant. The post-hoc results were also consistent with the different p-values. In particular, the drivers reduced their speed to a greater extent for TURN than for STRAIGHT and LANE CHANGE, while the speed reduction was greater for interruptible than uninterruptible cases. These results indicate that, regardless of the maneuver types, the speed reduction behavior improves driver interruptibility.

In Section 4, the analysis of the percentage of interruptible cases in terms of driving safety revealed significant differences in the percentage across the maneuver types (LANE CHANGE vs. STRAIGHT and TURN). One possible explanation for such a difference could be the higher difficulty in speed reduction for LANE CHANGE than STRAIGHT and TURN [65]. In addition, prior studies showed that drivers are more likely to not employ adaptive behaviors when they perceive a higher risk level associated with the behaviors [14, 58, 64]. Accordingly, drivers could be more likely to perceive a higher risk for LANE CHANGE than STRAIGHT and TURN owing to a higher difficulty given that the drivers had a significantly higher overall perceived difficulty for LANE CHANGE than for STRAIGHT and TURN.

5.2.2 Secondary Task Adaptation. Drivers commonly reported that they reduced or ceased to engage in secondary tasks whenever they needed to pay more attention to driving. Overall, we found that 14% of *n*-back tasks (n = 192) exhibited recall errors, in which an average of 2.58 (SD = 1.65) items were incorrect for a driver. Furthermore, in 89% of the cases (n = 171), drivers partially answered incorrectly. These results indicate that the drivers were more

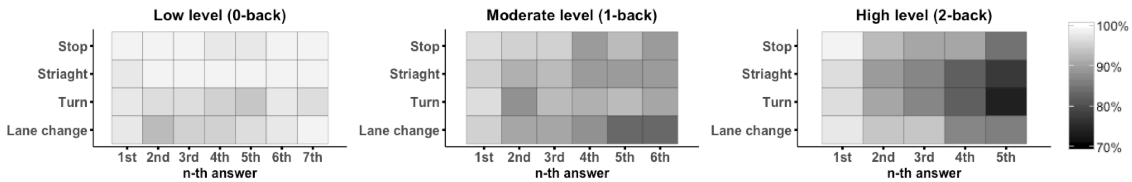

Fig. 7. Heat maps of the percentage of drivers who correctly answered each n-th item in the n-back task across different maneuver types and n-back types. The darker tiles indicate a lower percentage.

likely to temporarily reduce or cease to engage in secondary tasks. Next, we further investigated instances when the drivers temporarily reduced or ceased engaging in secondary tasks. Figure 7 reveals three heat maps that present the percentage of drivers who correctly answered each n-th item in the n-back task across the different maneuver types and n-back types. In this figure, the darker tiles indicate that the drivers are more likely to answer incorrectly for a given item. The maps show that the percentage decreases as the item order (n-th) increases, and decreases further as the demands of the n-back types increase. In addition, when the demands of the n-back types are high, the percentage of the latter items is lower when the maneuver types change in the following order: STOP > STRAIGHT > TURN > LANE CHANGE (lowest percentage). These results indicate that when the secondary-task demand is high, drivers are more interruptible in terms of auditory-verbal performance as the duration of the task decreases. One possible cause of these results is a decrease in the drivers' attention due to the high cognitive demand of concurrent multitasking.

6 DISCUSSION

6.1 Summary of Findings

(RQ1) Impacts of secondary-task demands on driver interruptibility: Our findings highlight that driver interruptibility research should carefully consider the potential impact of the secondary-task demands and the maneuver types on the performance of driving or secondary tasks. The previous studies on driver interruptibility have not considered the effect of secondary task demands [7, 60, 68]. Our results emphasize that in addition to the driving demand, secondary-task demands are also an important factor. As demonstrated in Figure 4, driver interruptibility is significantly varied across the driving maneuver types and the secondary-task demands. For instance, the percentage of interruptible moments (in terms of the "overall perceived difficulty" and "auditory-verbal performance") significantly decreased as the secondary-task demands increased. In addition, this percentage decremented according to the maneuver type in the following order: STOP > STRAIGHT > TURN > LANE CHANGE (lowest percentage). In contrast, the percentage of interruptible moments (in terms of "the driving safety") was significantly varied only when the maneuver was LANE CHANGE. This finding implies that findings for the maneuver type (or a secondary-task demand) may not be applicable to other conditions associated with different maneuver types (or secondary-task demands).

(RQ2) Impacts of adaptive behaviors on driver interruptibility: Prior research on driver interruptibility has not considered how drivers adapt their behaviors while engaging in secondary tasks [7, 52, 68]. Our findings suggest that drivers employ several adaptive behaviors to manage their interruptibility and to ensure driving safety. Two common adaptive behaviors are reducing vehicle speed and secondary-task engagement (e.g., lowering attention or disengaging). In addition, several drivers proactively adjusted various driving tasks to engage in secondary tasks such as delaying the vehicle start and preparing for voice interactions by changing lanes in advance.

Drivers commonly reduce their speed when engaging in secondary tasks. A greater speed reduction was associated with higher interruptibility. Prior research on speed reduction generally examined the relationship between speed reduction and accident likelihood or driving performance [62, 64]. Even though this finding is in line with previous research that demonstrates that a speed reduction improves the driving performance (driving safety in this study) [58], we revealed an additional finding that the speed reduction also improves secondary-task performance (auditory-verbal performance in our study context).

Drivers also commonly alter secondary-task engagement whenever they needed to pay more attention to their driving. The reduction of secondary-task engagement was a common reason for the cases with low auditory-verbal performance. For a given *n*-back task, drivers were more likely to recall information incorrectly toward the end of the given task. This phenomenon was exacerbated further as the demand for either one or both tasks increased. This implies that when the multitask demand is high, drivers are more likely to be interruptible due to the auditory-verbal performance as the duration of a secondary task decreases.

To the best of our knowledge, prior studies on driver interruptibility have neither examined the impact of the length of a secondary task on interruptibility nor considered the length in interruptibility classification. For in-vehicle systems, voice interactions are becoming increasingly popular; hence, there need to be more studies that examine how the duration of voice interactions impact driver interruptibility. Beyond the vehicular contexts, our work also brings new insights into existing receptivity models for mobile just-in-time health intervention with mobile and wearable technologies [13], which requires further studies (e.g., effects of concurrent multitasking and adaptive behaviors for health intervention delivery).

Previous research has observed that the adaptive behaviors of drivers mostly occur in simulator-environments [62]. We examined the adaptive behaviors in naturalistic driving environments. Our results uncovered a considerable number of uninterruptible moments (on average, 14%) in terms of driving safety (see Figure 4). This implies that in the real world contexts, drivers may not always fully employ adaptive behaviors to compensate for an additional workload of voice tasks, and are vulnerable to be distracted by the task. Drivers often show the so-called optimism bias [18, 70] of engaging in any tasks with little conscious evaluation of driving safety risks [28, 62]. These suggest that there needs to be a safety aid (e.g., automatic interruption classification). The following discusses how our findings can be used to develop such an aid scheme

6.2 Potentially Useful In-situ Measures for the Classification of Driver Interruptibility

Driving tasks can be decomposed into multiple sub-tasks [21, 62]. Previous research on driver interruptibility has considered major sub-tasks, such as braking, accelerating, and handling the steering wheel, to classify interruptibility [31, 52, 68]. When analyzing the interviews, we discovered that when drivers described difficult moments while performing secondary tasks, the description generally involved subsidiary sub-tasks, such as checking mirrors. For example, D14 said, *"Whether that car will be still there or moves forward. To estimate the speed of the car, I need to look in the rear mirror, as well as the side mirror, to handle the wheel, and to clarify the situation."* D15 stated, *"When passing a crosswalk, people may cut in [...] I need to see what is coming, I get confused at answering, as I see the rear mirror, bicycle may come, see the side of the vehicle, and see the front of the vehicle."* These suggest that the occurrence of subsidiary sub-tasks, such as checking mirrors, could be a useful in-situ measure for driver interruptibility classification from the driver's perspective. In addition, the presence of people (e.g., bicyclists or pedestrians) near the vehicle could also serve as a useful in-situ measure. Note that when using such vehicular sensor data, system developers should carefully consider the privacy concerns of users and bystanders [59].

6.3 Towards Flow Control of Driver-Vehicle Voice Interactions

In data communications, flow control refers to a technical concept to ensure that a sender is not overwhelming a receiver by sending packets faster than it can consume [39]. Our findings show that drivers are more likely interruptible as a secondary-task demand decreases. In addition, when a secondary-task demand is high, drivers are more likely to be interruptible for a shorter length of secondary tasks. These imply that drivers can be more interruptible by controlling secondary tasks or voice interactions. In other words, the concept of flow control can be borrowed to manage the voice interaction flows such that a system (sender) does not overwhelm a driver (receiver). Intelligent systems can broadly support the following flow control initiatives for voice interaction, depending on who has the control of an interaction (e.g., system or user, or both): system-initiative, user-initiative, and mixed-initiative interaction (as suggested in [29]).

In *system-initiative flow control*, proactive interactions are initiated and led by systems; thus, a system can be designed to automatically control the flow of interactions by considering the interruptibility of their drivers. When interruptibility is low (i.e., less opportune moments for interruption), the system can defer voice interactions until it finds opportune moments later. Furthermore, the system can dynamically adjust the interaction workload. Our results show that a driver's task performance degrades as the length of the secondary tasks increases. This implies that a shorter task is more likely to be interruptible. The literature shows a variety of automatic text summarization techniques [25], which can shorten the length, and may reduce the cognitive demand associated with the interactions. For example, when the multitasking demand is high, the systems could summarize the information (text) that is delivered to the drivers in order to improve driver interruptibility. The system could *decompose* an interaction into multiple *micro-interactions* using existing automatic text decomposition techniques, or alternatively, voice interaction designers can compose such micro-interaction tasks. The system could then delay a micro-interaction when the driver is uninterruptible and could resume later when a driver becomes interruptible.

In *user-initiative flow control*, which is the current model of in-vehicle interface usage, users have full control on the interaction flow (e.g., starting and stopping), although the current voice systems lack fine-grained, explicit flow controls unlike natural conversations (e.g., temporarily holding conversations until opportune moments appear). In the interviews, it is interesting that the drivers commonly mentioned that they could have a better engagement in the secondary tasks if they would have more control over the interactions. The drivers also explained how they control conversations with other humans when they need to pay more attention to driving. For example, D12 mentioned *"Although I could stop the (system-initiated) conversation by saying 'no,' driving changes from moment to moment. So, I said 'yes' at the beginning, but may not (engage in the conversation after saying 'yes') [...] I would pause the conversations if I were to converse with humans. But once I start the (system-initiated) conversation (when not being able to continually engage in it) I can't do anything but stop answering."* This result implies that there needs to be a more natural means of controlling the driver-vehicle interactions. It could be helpful to enable simple control commands, such as "Repeat" to easily recover after moments with low auditory-verbal performance. Furthermore, the user could explicitly control the flow by enabling pause and resume commands such as "Hold on" and "Resume."

In *mixed-initiative flow control*, users and systems collaboratively control the voice interaction flow. For example, when there is high "uncertainty" in driver interruptibility, the system makes a joint-decision on whether to start interactions by asking drivers to resolve the uncertainty. Instead of employing a dialog, the system could hint forthcoming interactions (e.g., by delivering an advanced cueing/warning) that consider the users' responses by inferring the driver's intention based on their driving behavior. In this work, it was determined that the drivers reduced the vehicle speed while engaging in secondary tasks. This speed reduction pattern could be used to infer the driver's intention. The system could utilize both dialogs and hints or either one based on the driving safety constraint. The advanced cueing may increase interruptibility since they have more time to prepare for the interactions (e.g., reducing speed). For example, drivers could change lanes in advance according to the cue. Alternatively, a visual cue could be used, for example by showing the location of a forthcoming interaction in a navigation map, to inform where the drivers will engage in the voice interaction. After informing the driver with a proactive interaction by using such an auditory or visual cue, in-vehicle systems may not initiate interactions, for example, if the driver does not reduce the speed of the vehicle.

6.4 In-vehicle Multitasking and Risk Management Strategies

Driving is a safety-critical task that is performed daily by hundreds of millions of people. In the U.S. for example, there are more than 227 million drivers and 284 million vehicles in 2018 [3]. Currently, most in-vehicle voice interfaces enable drivers to initialize the interactions with the systems at any moment, regardless of driving conditions, under the assumption that drivers are rational—drivers will only engage in in-vehicle activities as long as driving conditions allow. However, previous studies warn that drivers are not that rational [28, 62]. For example, Horrey et al. demonstrated that drivers tend to initiate secondary tasks regardless of their current driving conditions, even when they are fully aware of the relative demands of the road [28]. This suggests that any failures in adaptive behaviors during a difficult section of driving could potentially lead to cognitive distractions to the drivers.

The most ideal way would be blocking any usage while the vehicle is not stationary. This approach has been extensively studied in the field of human-computer interaction [32, 33, 36]. Unfortunately, prior studies have also shown that such restrictive approaches could not be less effective in practice [16, 19]. For example, Creaser et al. showed that drivers may try to bypass the blocking application [16]. This implies that there needs to be a more neutral or comfortable means of intelligently controlling driver-vehicle interactions for safety improvements as suggested in the intelligent positive computing framework [43]; e.g., context-aware adaptive controlling [32]. To develop such means, our findings and the concept of flow control mechanisms could be useful. For example, as discussed for the system-initiative flow control mechanism, in-vehicle system could warn or defer voice interactions when the driver is uninterruptible (e.g., unsafe).

6.5 Limitations

Although this study made the first step toward understanding how driving and secondary-task demands and the adaptive behaviors are related to driver interruptibility, our results should be carefully interpreted. First, the results may not be generalizable to in-vehicle voice interactions that require additional types of input and/or output modality, given that there are still many in-vehicle interfaces that partly involve visual and/or manual operations [2]. Second, we compared driving safety across maneuver types and n-back types. We acknowledge that there could be other factors that may influence the driving safety. In our work, as one measure, driving safety was estimated based on the difference of driving performances between when concurrently executing driving and voice tasks and when only executing the driving task. For the estimation, the steering wheel reversal rate (SRR) [62] was used as a measure for the driving performance. Besides maneuver types, any vehicle control metrics, including the SRR, can be potentially influenced by other factors. These factors may include environmental factors (e.g., traffic density, vehicle speed, traffic rules), driver characteristics (e.g., driving experience, age, and road familiarity), and the existence of additional internal/external interruption (e.g., mind wandering) [41, 62].

7 CONCLUSION

We quantitatively and qualitatively examined the real-road interruptibility dataset. Our results showed that in addition to the demand of the driving task, the verbal task demand also significantly affects driver interruptibility. In addition, diverse adaptive behaviors, such as reducing the vehicle speed and adjusting the voice task engagement, significantly mediates the effect of the verbal-task demand on driver interruptibility. Based on our findings, we emphasize the importance of intelligent flow control for proactive voice interaction and elaborate on the design of different types of flow control mechanisms.

New technologies in intelligent vehicles and the popularity of auditory-verbal interfaces enable a variety of proactive or system-initiated interactions. We hope our findings and the concept of flow control for voice interactions may provide the primary step to enable safe and gratifying proactive in-vehicle voice interactions.

ACKNOWLEDGMENTS

This research was supported by Next-Generation Information Computing Development Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (NRF-2017M3C4A7065960) and by the 2019 KK-JRC Smart Project.

REFERENCES

[1] Piotr D. Adamczyk and Brian P. Bailey. 2004. If Not Now, when?: The Effects of Interruption at Different Moments Within Task Execution. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '04)*. ACM, New York, NY, USA, 271–278.

[2] National Highway Traffic Safety Administration. 2012. Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices. (2012).

[3] National Highway Traffic Safety Administration. 2019. Highway Statistics 2018. (2019).

[4] Christoph Anderson, Isabel Hübener, Ann-Kathrin Seipp, Sandra Ohly, Klaus David, and Veljko Pejovic. 2018. A Survey of Attention Management Systems in Ubiquitous Computing Environments. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 2, Article 58 (July 2018), 27 pages.

[5] Brian P. Bailey and Shamsi T. Iqbal. 2008. Understanding Changes in Mental Workload During Execution of Goal-directed Tasks and Its Application for Interruption Management. *ACM Trans. Comput.-Hum. Interact.* 14, 4, Article 21 (Jan. 2008), 28 pages.

[6] Kristen E Beede and Steven J Kass. 2006. Engrossed in conversation: The impact of cell phones on simulated driving performance. *Accident Analysis & Prevention* 38, 2 (2006), 415–421.

[7] Thierry Bellet, Marie-Pierre Bruyas, Héléne Tattegrain-Veste, J F Forzy, Anabela Simões, José Carvalhais, Philip Lockwood, Jérôme Boudy, Bernard Baligand, S. Damiani, and Michael Opitz. 2002. "Real Time" Analysis of the Driving Situation in order to Manage On-Board Information. In *Proceeding of the E-Safety Congress and Exhibition-IT Solutions for Safety and Security in Intelligent Transport.*

[8] Kenneth R Boff, Lloyd Kaufman, and James P Thomas. 1994. *Handbook of perception and human performance. Volume 2. Cognitive processes and performance.* Technical Report. Harry G. Armstrong Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, Ohio.

[9] Duncan P. Brumby, Dario D. Salvucci, and Andrew Howes. 2009. Focus on Driving: How Cognitive Constraints Shape the Adaptation of Strategy When Dialing While Driving. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. Association for Computing Machinery, New York, NY, USA, 1629–1638.

[10] Marie-Pierre Bruyas, Laëtitia Dumont, and Bron France. 2015. Sensitivity of Detection Response Task (DRT) to Driving Demand and Task Difficulty. *Accident Reconstruction Journal* 25 (2015), 12–15.

[11] JD Catherine Chase. 2014. US state and federal laws targeting distracted driving. *Annals of advances in automotive medicine* 58 (2014), 84.

[12] Woohyeok Choi, Sangkeun Park, Duyeon Kim, Youn-kyung Lim, and Uichin Lee. 2019. Multi-Stage Receptivity Model for Mobile Just-In-Time Health Intervention. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 2, Article Article 39 (June 2019), 26 pages.

[13] Woohyeok Choi, Sangkeun Park, Duyeon Kim, Youn-kyung Lim, and Uichin Lee. 2019. Multi-Stage Receptivity Model for Mobile Just-In-Time Health Intervention. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 2, Article Article 39 (June 2019), 26 pages. https://doi.org/10.1145/3328910

[14] Pushpa Choudhary and Nagendra R Velaga. 2019. Driver behaviour at the onset of yellow signal: a comparative study of distraction caused by use of a phone and a music player. *Transportation research part F: traffic psychology and behaviour* 62 (2019), 135–148.

[15] Fokie Cnossen, Talib Rothengatter, and Theo Meijman. 2000. Strategic changes in task performance in simulated car driving as an adaptive response to task demands. *Transportation Research Part F: Traffic Psychology and Behaviour* 3, 3 (2000), 123–140.

[16] Janet I. Creaser, Christopher J. Edwards, Nichole L. Morris, and Max Donath. 2015. Are cellular phone blocking applications effective for novice teen drivers? *Journal of Safety Research* 54 (2015), 75.e29 – 78.

[17] Dick De Waard. 1996. *The measurement of drivers' mental workload.* Groningen University, Traffic Research Center Netherlands.

[18] David M DeJoy. 1989. The optimism bias and traffic accident risk perception. *Accident Analysis & Prevention* 21, 4 (1989), 333–340.

[19] M Kit Delgado, Kathryn J Wanner, and Catherine McDonald. 2016. Adolescent cellphone use while driving: An overview of the literature and promising future directions for prevention. *Media and communication* 4, 3 (2016), 79.

[20] Frank A Drews, Monisha Pasupathi, and David L Strayer. 2008. Passenger and cell phone conversations in simulated driving. *Journal of Experimental Psychology: Applied* 14, 4 (2008), 392.

[21] Wolfgang Fastenmeier and Herbert Gstalter. 2007. Driving task analysis as a tool in traffic safety research and practice. *Safety Science* 45, 9 (2007), 952–979.

[22] Vérane Faure, Régis Lobjois, and Nicolas Benguigui. 2016. The effects of driving environment complexity and dual tasking on drivers' mental workload and eye blink behavior. *Transportation Research Part F: Traffic Psychology and Behaviour* 40 (2016), 78–90.

[23] Gregory M Fitch, Kevin Grove, Richard J Hanowski, and Miguel A Perez. 2014. Compensatory behavior of drivers when conversing on a cell phone: Investigation with naturalistic driving data. *Transportation research record* 2434, 1 (2014), 1–8.
[24] R Fuller. 1991. Behavior analysis and unsafe driving: Warning-learning trap ahead! *Journal of Applied Behavior Analysis* 24, 1 (1991), 73–75.
[25] Mahak Gambhir and Vishal Gupta. 2017. Recent automatic text summarization techniques: a survey. *Artificial Intelligence Review* 47, 1 (2017), 1–66.
[26] Salvador García, Alberto Fernández, Julián Luengo, and Francisco Herrera. 2010. Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power. *Information Sciences* 180, 10 (2010), 2044–2064.
[27] Lisbeth Harms. 1991. Variation in drivers' cognitive load. Effects of driving through village areas and rural junctions. *Ergonomics* 34, 2 (1991), 151–160.
[28] William J Horrey and Mary F Lesch. 2009. Driver-initiated distractions: Examining strategic adaptation for in-vehicle task initiation. *Accident Analysis & Prevention* 41, 1 (2009), 115–122.
[29] Eric Horvitz. 1999. Principles of Mixed-initiative User Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '99)*. ACM, New York, NY, USA, 159–166.
[30] David B Kaber, Yulan Liang, Yu Zhang, Meghan L Rogers, and Shruti Gangakhedkar. 2012. Driver performance effects of simultaneous visual and cognitive distraction and adaptation behavior. *Transportation research part F: traffic psychology and behaviour* 15, 5 (2012), 491–501.
[31] Auk Kim, Woohyeok Choi, Jungmi Park, Kyeyoon Kim, and Uichin Lee. 2018. Interrupting Drivers for Interactions: Predicting Opportune Moments for In-vehicle Proactive Auditory-verbal Tasks. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 4, Article 175 (Dec. 2018), 28 pages.
[32] Inyeop Kim, Gyuwon Jung, Hayoung Jung, Minsam Ko, and Uichin Lee. 2017. Let's FOCUS: Mitigating Mobile Phone Use in College Classrooms. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 63.
[33] Jaejeung Kim, Chiwoo Cho, and Uichin Lee. 2017. Technology Supported Behavior Restriction for Mitigating Self-Interruptions in Multi-device Environments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 64.
[34] Sheila G Klauer, Feng Guo, Jeremy Sudweeks, and Thomas A Dingus. 2010. *An analysis of driver inattention using a case-crossover approach on 100-car data.* Technical Report.
[35] Rex B Kline. 2015. *Principles and practice of structural equation modeling.* Guilford publications.
[36] Minsam Ko, Seungwoo Choi, Koji Yatani, and Uichin Lee. 2016. Lock n' LoL: group-based limiting assistance app to mitigate smartphone distractions in group activities. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 998–1010.
[37] Joe W. Kotrlik and Heather A. Williams. 2003. The incorporation of effect size in information technology, learning, information technology, learning, and performance research and performance research. *Information Technology, Learning, and Performance Journal* 21, 1 (2003), 1.
[38] Georgios K. Kountouriotis, Panagiotis Spyridakos, Oliver M.J. Carsten, and Natasha Merat. 2016. Identifying cognitive distraction using steering wheel reversal rates. *Accident Analysis & Prevention* 96 (2016), 39 – 45.
[39] James Kurose and Keith Ross. 2016. *Computer Networking: A Top-Down Approach (7th Edition)*. Pearson.
[40] Eun-Kyu Lee, Mario Gerla, Giovanni Pau, Uichin Lee, and Jae-Han Lim. 2016. Internet of Vehicles: From intelligent grid to autonomous cars and vehicular fogs. *International Journal of Distributed Sensor Networks* 12, 9 (2016), 1550147716665500.
[41] John D Lee. 2014. Dynamics of driver distraction: The process of engaging and disengaging. *Annals of advances in automotive medicine* 58 (2014), 24–32.
[42] John D. Lee and David L. Strayer. 2004. Preface to the Special Section on Driver Distraction. *Human Factors* 46, 4 (2004), 583–586.
[43] Uichin Lee, Kyungsik Han, Hyunsung Cho, Kyong-Mee Chung, Hwajung Hong, Sung-Ju Lee, Youngtae Noh, Sooyoung Park, and John M. Carroll. 2019. Intelligent positive computing with mobile, wearable, and IoT devices: Literature review and research directions. *Ad Hoc Networks* 83 (2019), 8 – 24. https://doi.org/10.1016/j.adhoc.2018.08.021
[44] Jonathan Levy and Harold Pashler. 2008. Task prioritisation in multitasking during driving: opportunity to abort a concurrent task does not insulate braking responses from dual-task slowing. *Applied Cognitive Psychology* 22, 4 (2008), 507–525.
[45] Geoffrey R Loftus. 1978. On interpretation of interactions. *Memory & Cognition* 6, 3 (1978), 312–319.
[46] Sean MacLain. 2017. Toyota's Talking Car Wants to Be Your Clingy BFF. (30 October 2017).
[47] Philipp Mayring. 2004. Qualitative content analysis. *A companion to qualitative research* 1 (2004), 159–176.
[48] Bruce Mehler, Bryan Reimer, and Jeffery Dusek. 2011. MIT AgeLab Delayed Digit Recall Task (n-back). (2011).
[49] Barbara Metz, Andreas Landau, and Volker Hargutt. 2015. Frequency and impact of hands-free telephoning while driving–Results from naturalistic driving data. *Transportation research part F: traffic psychology and behaviour* 29 (2015), 1–13.
[50] Christian Meurisch, Maria-Dorina Ionescu, Benedikt Schmidt, and Max Mühlhäuser. 2017. Reference Model of Next-generation Digital Personal Assistant: Integrating Proactive Behavior. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '17)*. ACM, New York, NY, USA, 149–152.

[51] Jeff Miller and Moritz Durst. 2014. "Just do it when you get a chance": the effects of a background task on primary task performance. *Attention, Perception, & Psychophysics* 76, 8 (01 Nov 2014), 2560–2574.

[52] R Montanari, G Wenzel, S Mattes, F Kuhn, F BELLOTTI, and D Morreale. 2002. COMUNICAR: Integrated on-vehicle human machine interface designed to avoid driver information overload. In *9th World Congress on Intelligent Transport SystemsITS America, ITS Japan, ERTICO (Intelligent Transport Systems and Services-Europe).*

[53] Carol M Musil, Camille B Warner, Piyanee Klainin Yobas, and Susan L Jones. 2002. A comparison of imputation techniques for handling missing data. *Western Journal of Nursing Research* 24, 7 (2002), 815–829.

[54] David Navon and Daniel Gopher. 1979. On the economy of the human-processing system. *Psychological review* 86, 3 (1979), 214.

[55] Donald A Norman and Daniel G Bobrow. 1975. On data-limited and resource-limited processes. *Cognitive psychology* 7, 1 (1975), 44–64.

[56] Ann A O'Connell and D Betsy McCoach. 2008. *Multilevel modeling of educational data.* IAP.

[57] Joakim Östlund, Björn Peters, Birgitta Thorslund, Johan Engström, Gustav Markkula, Andreas Keinath, Dorit Horst, Susann Juch, Stefan Mattes, and Uli Foehl. 2005. *Driving performance assessment - methods and metrics.* Technical Report. Information Society Technologies (IST) Programme.

[58] Oscar Oviedo-Trespalacios, Md Mazharul Haque, Mark King, and Simon Washington. 2017. Effects of road infrastructure and traffic complexity in speed adaptation behaviour of distracted drivers. *Accident Analysis & Prevention* 101 (2017), 67–77.

[59] Sangkeun Park, Joohyun Kim, Rabeb Mizouni, and Uichin Lee. 2016. Motives and Concerns of Dashcam Video Sharing. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16).* ACM, New York, NY, USA, 4758–4769.

[60] Rahul Rajan, Ted Selker, and Ian Lane. 2016. Task Load Estimation and Mediation Using Psycho-physiological Measures. In *Proceedings of the 21st International Conference on Intelligent User Interfaces (IUI '16).* ACM, New York, NY, USA, 48–59.

[61] Thomas Ranney, Scott Baldwin, Larry Smith, Elizabeth Mazzae, and Russell Pierce. 2014. Detection Response Task (DRT) Evaluation for Driver Distraction Measurement Application. (2014).

[62] Michael A Regan, John D Lee, and Kristie Young. 2008. *Driver distraction: Theory, effects, and mitigation.* CRC Press.

[63] John TE Richardson. 2011. Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review* 6, 2 (2011), 135–147.

[64] Christina Rudin-Brown and Samantha Jamson. 2013. *Behavioural adaptation and road safety: Theory, evidence and action.* CRC Press.

[65] Dario D Salvucci and Andrew Liu. 2002. The time course of a lane change: Driver control and eye-movement behavior. *Transportation research part F: traffic psychology and behaviour* 5, 2 (2002), 123–132.

[66] Dario D. Salvucci, Niels A. Taatgen, and Jelmer P. Borst. 2009. Toward a Unified Theory of the Multitasking Continuum: From Concurrent Performance to Task Switching, Interruption, and Resumption. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09).* Association for Computing Machinery, New York, NY, USA, 1819–1828. https://doi.org/10.1145/1518701.1518981

[67] Nadine Sarter. 2013. Multimodal support for interruption management: Models, empirical findings, and design recommendations. *Proc. IEEE* 101, 9 (2013), 2105–2112.

[68] Rob Semmens, Nikolas Martelaro, Pushyami Kaveti, Simon Stent, and Wendy Ju. 2019. Is Now A Good Time?: An Empirical Study of Vehicle-Driver Communication Timing. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19).* ACM, New York, NY, USA, Article 637, 12 pages.

[69] Victor Shcherb, Alexey Pelykh, Hardy Mueller, and Pavol Zibrita. 2017. OsmAnd - OSM Automated Navigation Directions - navigation software based on OpenStreetMap.

[70] Mathew P White, J Richard Eiser, and Peter R Harris. 2004. Risk perceptions of mobile phone use while driving. *Risk Analysis: An International Journal* 24, 2 (2004), 323–334.

[71] Christopher D. Wickens. 2008. Multiple Resources and Mental Workload. *Human Factors* 50, 3 (2008), 449–455.

[72] Christopher D. Wickens. 2010. Multiple Resources And Performance Prediction. *Theoretical Issues in Ergonomics Science* 3, 2 (2010), 159–177.