



UNIVERSITY OF LEEDS

This is a repository copy of *Effects of adopting voice assistants countermeasures on alerting drowsy drivers across age*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/233018/>

Version: Accepted Version

Article:

Zhang, W., Carsten, O. orcid.org/0000-0003-0285-8046, Goodridge, C.M. et al. (1 more author) (2025) Effects of adopting voice assistants countermeasures on alerting drowsy drivers across age. *Applied Ergonomics*, 129. 104616. ISSN: 0003-6870

<https://doi.org/10.1016/j.apergo.2025.104616>

This is an author produced version of an article published in *Applied Ergonomics*, made available under the terms of the Creative Commons Attribution License (CC-BY), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

Effects of Adopting Voice Assistants Countermeasures on Alerting Drowsy Drivers Across Age

Wei Zhang ¹, Oliver Carsten ², Courtney M. Goodridge ², Wei Zhang ^{1*}

1 Department of Industrial Engineering, Tsinghua University, China

2 Institute for Transport Studies, University of Leeds, United Kingdom

* Corresponding author: Wei Zhang, zhangwei@mail.tsinghua.edu.cn,
Department of Industrial Engineering, Tsinghua University, China

Abstract

Drowsy driving is one of the leading contributing factors to traffic accidents. As drivers continue driving despite being aware of their drowsy states, it is necessary to adopt active countermeasures to avoid them falling asleep behind the wheel. Based on the multiple resources model, this study proposed five countermeasure types using voice assistants (*no-countermeasure* vs. *meaninglessly listening* vs. *meaningfully listening* vs. *repeating* vs. *answering*) and investigated their effects on countering drowsiness for young and middle-aged drivers. A five-week simulated manual driving experiment was conducted to evaluate the effectiveness of countermeasures based on driver performance, physiological indicators, eye movement indicators, and subjective ratings. Results indicated that *repeating* and *answering* were the two most effective countermeasures. The self-reported sleepiness of these two countermeasures (6.328 and 5.276 for *repeating* and *answering*, respectively) were significantly ($p < 0.05$) lower than that of the *no-countermeasure* (7.983). The physiological arousal state in term of skin conductance level for the two (1.798 and 1.990 for *repeating* and *answering*, respectively) were significantly ($p < 0.05$) higher than that of the *no-countermeasure* (1.022). Consistent patterns were also found in driving performance, eye movement indicators, psychomotor vigilance task accuracy, and subjective acceptance ratings. These findings suggested that countermeasures involving more resource stages can enhance drivers' alertness. The trend of higher alertness in the *answering* condition compared to the *repeating* condition highlighted the advantage of increasing cognitive load in countering drowsiness. Although young and middle-aged drowsy drivers did not differ on their average responses to different countermeasures, they showed differences in their physiological activation in reacting to the countermeasures. Drowsiness countermeasures should be tailored to fulfill individual needs, particularly for middle-aged drivers. This study not only clarifies how drowsiness countermeasures work but also offers practical guidance for vehicle designers on using in-vehicle voice assistants effectively.

Keywords: drowsiness countermeasures; voice assistants; drivers' alertness; young drivers; middle-aged drivers

1 Introduction

1.1 Background

Drowsy driving is one of the leading contributory factors to traffic accidents, resulting in approximately 20% of all motor vehicle crashes (National Sleep Foundation, 2023) and 1.8 % of fatal crashes from 2017 to 2021 in the U.S.A. (NHTSA, 2023). Drowsiness is a state of low alertness associated with lower frontal lobe functioning, greater exertion of mental effort, and an involuntary tendency for rapid skeletal muscle inhibition (*APA Dictionary of Psychology*, n.d.). The state of drowsiness may result from low arousal (e.g., long, monotonous driving), high workload (e.g., multitasking while driving), or sleep-related factors (e.g., sleep deprivation) (Ayas et al., 2023). It could reduce drivers' attention, impair their hazard perception, and slower their responses, and thus increase the risk of accidents (Kaplan & Prato, 2012; Smith et al., 2009).

Given the common causes of driver drowsiness, implementing timely countermeasures is crucial for driving safety. More than 16 methods to mitigate drowsiness have been reported in the literature (Gershon et al., 2011; Meng et al., 2015), including talking with passengers, taking a short nap, listening to music or the radio, opening the window, drinking, and etc. However, drivers continue to drive despite being aware of their drowsy states and available effective countermeasures (Jackson et al., 2011; Nordbakke & Sagberg, 2007). This tendency might stem from drivers' reliance on their past experiences, perceived social norms, journey goals, or a lack of understanding for the risks associated with drowsy driving due to reduced brain activation levels (Jackson et al., 2011; Large et al., 2017). Therefore, it is important to adopt active countermeasures for drowsiness to avoid drivers falling asleep behind the wheel.

With the widespread of in-vehicle voice assistants (RationalStat LLC, 2023), it is promising to adopt voice assistants as active drowsiness countermeasures. Moreover, given the significant rise of global median age (UN DESA, 2024) and the high

prevalence of sleepy driving across ages (Obst et al., 2011), exploring the feasibility of voice-assistant based drowsiness countermeasures for drivers of different ages has become increasingly important. By adopting voice assistants, this study innovatively proposed different countermeasures and conducted a simulated driving experiment to evaluate their effectiveness in countering drowsiness among drivers of different age groups.

1.2 Adopting voice assistants as drowsiness countermeasures

In-vehicle voice assistants provide different types of interaction such as playing the radio and starting conversations. They may be promising drowsiness countermeasures with several advantages. First, adopting such assistants to counter drowsiness is convenient. In 2023, over 20 million vehicles have equipped with automotive voice assistants, with an installation rate higher than 80 % (RationalStat LLC, 2023). These widespread voice assistants can provide convenient ways to counter drowsiness when the monitoring systems detect drivers' drowsiness (Arakawa, 2021; Lu et al., 2022). Second, engaging drowsy drivers in voice interactions is a proactive drowsiness countermeasure. Compared to passively receiving countermeasures such as feedback and warnings to be notified of their drowsy state (Aidman et al., 2015; Nishigaki & Shirakata, 2019), drivers can actively participate in the alerting process. In this process, voice assistants play the role of passengers, encouraging drivers to listen, think, or talk. Third, interacting with voice assistants have little competition in attention resources needed for driving. According to the multiple resources model, tasks interfere each other when they share the same resource stages (e.g., perception, cognition, responding), sensory modalities (auditory vs. visual), codes (spatial vs. verbal), and visual channels (Wickens, 2002). Driving primarily involves visual perception, spatial cognition, and motor skills. In contrast, alert maintenance tasks, such as memory recall tasks, auditory motor tasks, and gamified tasks (Gershon et al., 2009; Saxby et al., 2017; Takayama & Nass, 2008), often overlap with the same attention resources with the driving task. However, interacting with voice assistants primarily involves auditory perception,

cognitive processing, and verbal responding. These are largely distinct from driving demands. The drowsy drivers are, then, allowed to stay engaged in the alerting process with minimal interference.

Furthermore, interacting with voice assistants forms different types of drowsiness countermeasures, which would exert various drowsiness countering effects. As the Table 1 shows, countermeasures vary in the number of resource stages required for information processing and the extent of cognitive load involved, ranging from listening to music to engaging in conversations. On the one hand, increasing the number of resource stages in voice interactions might enhance alertness of drowsy drivers. For example, listening to music increases drivers' mental arousal through auditory perception alone. However, engaging conversations alert drivers with more resource stages, including perception, cognition, and responding. Since physical and mental arousals could produce combined arousal states (Oken et al., 2006; Yang et al., 2017), conversating with voice assistants that engage drivers both mentally and physically might exhibit more arousal states than passive listening. It was preliminary evidenced by Papachristos et al. (2024), who engaged drowsy drivers in listening to music and observed that drivers' vocal participation during the process yields a beneficial alerting effect, compared to passive listening or physical participation. On the other hand, moderately increasing drowsy drivers' cognitive load might also help to improve drivers' alertness. As underload is one of the reasons of drowsiness, increasing driver's cognitive load can alleviate the underload induced drowsiness (May & Baldwin, 2009). For instance, listening to more meaningful content (*vs.* less meaningful content) and engaging deeper conversations (*vs.* lighter conversations) could induce higher cognitive load for the drowsy drivers, helping to counteract underload-induced drowsiness by activating cognitive processes. This was supported by a non-driving-related study, which suggested that increasing workload from low to moderate levels was associated with reductions in fatigue (Grech et al., 2009).

Table 1. Involvement of resources in driving tasks and countermeasures using voice assistants

Task type	Information processing stages with perception modalities and codes			Common instances
	(Visual) Perception	(Spatial) Cognition	(Motor) Responding	
Driving Tasks	✓	✓	✓	
	(Auditory) Perception	(Verbal) Cognition	(Verbal) Responding	
	✓			Listening to music
Countermeasure types	✓	✓		Listening to the radio
using voice assistants	✓	✓	✓	Engaging in light conversation
	✓	✓✓	✓	Engaging in deep conversation

Note: ✓ means involvement of the corresponding resource; ✓✓ means the strong involvement of the corresponding resource.

How these various voice-assistant-based countermeasures varying in resource stages and cognitive load would counter drowsiness remains unknown. Previous studies suggest that countermeasures involving human auditory and verbal processes can effectively alert drowsy drivers (Atchley et al., 2014; Mahajan, 2021; Large et al., 2017; Orsini et al., 2024; Takayama & Nass, 2008; Papachristos et al., 2024). Listening to music or engaging in conversation with a voice assistant, compared to having no countermeasure, has been shown to increase driver alertness. For example, Orsini et al. (2024) found that listening to music had a positive acute effect on perceived driving fatigue compared to not listening. Other studies have also showed that verbal response tasks could enhance alertness and driving performance more effectively than listening tasks (Atchley et al., 2014; Takayama & Nass, 2008). For instance, Takayama et al. (2008) indicated that drivers preferred and drove more safely with the verbally repeating task compared to the passively listening task. However, the above-mentioned studies differed in task types, interaction materials, countermeasure durations, and driving conditions, making it difficult to compare the effectiveness of various

countermeasures among these studies. Moreover, most of those studies assessed countermeasure effectiveness using specific auditory (e.g., radio listening; Atchley et al., 2014) or verbal tasks (e.g., repeating phrases; Takayama & Nass, 2008). Few studies have explored the usage of in-vehicle voice assistants to counter drowsiness, even though these assistants are increasingly integrated into vehicles for infotainment and driving support (Lee & Jeon, 2022). Although some researchers have employed a digital voice assistant to counter passive task-related fatigue (Mahajan et al., 2021; Large et al., 2017), they did not compare different types of countermeasures in terms of resource stages and cognitive involvement. Therefore, the effects of different countermeasures using voice assistants on countering drowsiness remain unclear. Inspired by the benefit of increasing resource stages (Oken et al., 2006; Yang et al., 2017) and enhancing cognitive load (May & Baldwin, 2009; Grech et al., 2009) in enhancing drowsy drivers' alertness, the following hypotheses were proposed:

H1: By adopting voice assistants, countermeasures with more resource stages alert drowsy drivers better than those with less resource stages.

H2: Increasing cognitive load in countermeasures can enhance drowsy drivers' alertness.

1.3 Age and drowsiness countermeasures

Effective countermeasures should ideally be generalized to the general population. The number and proportion of middle-aged drivers are expected to increase with the global aging trend. This is reflected in the rise of the global median age from 22.2 years in 1950 to 30.4 years in 2023 (UN DESA, 2024). This increasing trend is particularly evident in countries such as Japan (from 21.3 in 1950 to 49.0 in 2023), China (from 22.2 in 1950 to 39.1 in 2023), and the United States (from 29.0 in 1950 to 38.0 in 2023 (UN DESA, 2024). Drowsy driving is a common issue reported by drivers of different ages (National Sleep Foundation, 2023). It is necessary to consider effects of age when evaluating the effectiveness of drowsiness countermeasures as the literature (Anstey et al., 2005) indicates that driving related capabilities decline with age.

Various countermeasures using voice assistants may benefit drowsy drivers of different age differently. As age increases, drivers are less physiologically susceptible to sleep loss and sleepiness-related driving impairment. This is supported by the middle-aged or older drivers' lower level of sleepiness-related neurophysiological indicators and lower subjective reported sleepiness under sleepy-inducing driving conditions (Lowden et al., 2009; Vaz Fragoso et al., 2015). Additionally, middle-aged drivers (30+ years) have more driving experience and hold more negative attitudes toward sleepy driving than young drivers (17–29 years) (Watling and Watling, 2015). As a result, they tend to self-regulate by avoiding challenging situations like nighttime driving (Charlton et al., 2006) and by reducing their driving distances (Braitman et al., 2011). These self-regulatory behaviors might help them compensate the impaired driving-related cognitive abilities like divided attention, which decrease with age (Anstey et al., 2005; Murman et al., 2015). On the other hand, as voice interactions while driving constitute a dual task, age-related cognitive declines may interfere drivers of middle age to perform the dual task as effectively as their younger counterpart. Specifically, middle-aged drivers show longer reaction times and higher error rates in dual tasks, regardless of resource competition between tasks (Doroudgar et al., 2017; Fraser et al., 2010; Song et al., 2017). Therefore, the middle-aged drivers might have a relatively decreased ability to allocate their attention to the driving task and the voice interaction task simultaneously than young drivers. This could threaten their driving safety and reduce the drowsiness countermeasure efficiency.

It was suspected that middle-aged drivers, who are less prone to drowsiness but have impaired attention allocation capability, could respond better to lighter countermeasures that require less resource stages and cognitive effort. In contrast, young drivers, with relatively stronger attention allocation capability for dual tasks but are more prone to drowsiness, might benefit more from stronger countermeasures involving more resource stages and cognitive effort.

The research exploring how age affects the effectiveness of various voice-assistants-based countermeasures in countering drowsiness is limited. Song et al. (2017) found that young drivers were more affected in driving performance by fatigue but were benefitted more from the verbal responding task, while older drivers were relative unaffected by fatigue. Nevertheless, their study mainly focused on exploring different drivers' drowsiness development over time and investigating only the effects of verbal task as drowsiness countermeasure. Comparisons among different countermeasures for alerting drowsy drivers of different age groups remain underexplored. In addition, acceptance of a device can significantly influence the intention of drivers to use such devices (Yilmaz et al., 2022). It is important to understand how drivers of different ages accept voice assistants as drowsiness countermeasures. Even though some studies have indicated that older generations often face barriers to adopting new technologies (Tacken et al., 2005), the others have shown that middle-aged adults exhibited more positive intentions to use voice assistance than the young ones (Zhong et al., 2024). Although age-related differences in the acceptance of voice assistants have been examined (Zhong et al., 2024; Gollasch et al., 2021), most evaluations have considered voice assistants primarily as infotainment or in-vehicle control systems. Investigations on drivers' acceptance of voice assistants by serving them as drowsiness countermeasures have not been reported. How countermeasures that vary in resource stages and cognitive load would counter driver drowsiness and influence drivers' acceptance of different ages needs further investigation. Therefore, considering the trend of increasing resistance in drowsiness (Lowden et al., 2009) alongside the decline in cognitive capabilities with age (Anstey et al., 2005), this study formulated the following hypothesis:

H3: Young drivers may benefit more from countermeasures that involve more mental resource stages and higher cognitive load, while middle-aged drivers may suit better with lighter countermeasures.

1.4 Aim of the present study

In summary, the effect of voice-assistants-based countermeasures on countering drowsiness of drivers in different ages needs further investigation. Filling these gaps can provide insights into how countermeasures should be employed to drowsy drivers by using in-vehicle technologies. A simulated driving experiment in this study was conducted to explore the effects of five countermeasure types using voice assistants on countering young and middle-aged drivers' drowsiness, as measured by self-reported sleepiness, driving performance, physiological indicators, and eye movement indicators. In addition, differences in drivers' acceptance of using in-vehicle voice assistants between the two age groups were also analyzed and discussed. By adopting voice assistants, this study aimed to answer the following three research questions (RQs):

RQ1: Do countermeasures involving more resource stages (i.e., perception, cognition, responding) benefit more in alerting drowsy drivers?

RQ2: Does increased cognitive load in countermeasures alert drowsy drivers more?

RQ3: How does the effectiveness of different countermeasures vary in alerting drowsy drivers of young and middle-aged groups?

2 Method

2.1 Participants

This study focused on recruiting young (18-35 years) and middle-aged (36-64 years) male drivers, as previous research indicates that this demographic reports the highest frequency of drowsy driving incidents (Obst et al., 2011). This classification also aligned with the new age group classification considering the aging population (Mazer et al., 2021).

The minimum sample size required was 26. This was calculated using the G*Power software (3.1.9.6) using the effect size of 0.25, an α of 0.05, a power of 0.90, countermeasures of 5, a correlation of 0.5 among repeated measures, and a non-sphericity correction of 1 (G*Power, 2023).

A total of 30 participants were recruited through advertising and snowballing. One participant quitted due to personal reasons, leaving 15 participants in the young group (24-35 years, $M = 27.93$, $SD = 2.62$) and 14 in the middle-aged group (36-60 years, $M = 44.5$, $SD = 8.03$). All participants had valid driver's license and several years of driving experience (young group: $M = 5$, $SD = 2.45$; middle-aged group: $M = 20$, $SD = 5.35$).

For inclusion, participants needed to meet the following criteria: 1) physically healthy and cognitively intact; 2) own a valid driving license with over three years of driving experience; 3) regular driving of at least three times a week; 4) normal or corrected-to-normal vision and hearing; 5) regular sleep habits with no sleep-related disorders or shift work in the past month; 6) absence of cardiovascular, digestive, or psychological disorders; 7) no excessive daytime sleepiness (Epworth Sleepiness Scale score < 16) (Johns, 1991; Mahajan, 2021); 8) no plans for or recent engagement in extensive physical activities; 9) do not understand, speak, or write in French.

All participants signed an informed consent form before they joined the study. The experiment was approved by the Institutional Review Board of the Department of Industrial Engineering, Tsinghua University.

2.2 Experimental design

This study adopted a 2×5 mixed factorial design. The factors included age group and countermeasure type. The age group included young and middle-aged group. It was a between-subject variable. The countermeasure type included no countermeasure, meaninglessly listening, meaningfully listening, repeating, and answering. This variable was a within-subject variable. All participants completed driving tasks under each of the five countermeasure types.

2.3 Countermeasures and materials

The countermeasure types are listed in Table 2. This study used poems as materials for drowsy countermeasures because they are rhythmic with powerful effectiveness in eliciting peak emotional responses, as known as arousal (Wassiliwizky et al., 2017).

Two types of auditory poems with similar rhythms were selected. For the meaningless listening condition, the French poem “la Nuit de décembre” was used. Since all participants were required to not understand, speak, or write in French, this poem was meaningless to them. For the meaningful listening, repeating, and answering conditions, ten classical ancient Chinese poems were used. All the participants were familiar with these Chinese poems. They could comprehend and recite these poems easily since the primary study. Moreover, all the Chinese poems followed a five-character-quatrain structure, consisting of four lines with five characters per line. Since every Chinese character is monosyllabic, the five characters in each line corresponded exactly to five syllables. For example, “The Peasant's Lot” comprises four lines with five syllables for each line, which is easy and quick for participants to recite (see Figure 1 a)). For meaningfully listening condition, ten Chinese poems were compiled into a collection. For repeating and answering conditions, Chinese poems were displayed line by line.

All poems were recorded in a consistent tone and speed (i.e., around 95 characters/minutes) by a female speaker fluent in both French and Chinese. Each recording was processed to ensure similar rhythm across both languages, see audio track in Figure 1 b). Poems were played at an average volume of approximately 68 dB. A female experimenter manipulated the voice assistant to display all countermeasures. Each countermeasure lasted for approximately 3.5 minutes.

Table 2. Five countermeasure types

Countermeasure	Explanation
No countermeasure	Participants accepted no countermeasures. [Involve no process of auditory perception, verbal cognition, or verbal responding]
Meaninglessly listening	Participants listened to a French poem. [Involve process of auditory perception only]
Meaningfully listening	Participants listened to a series of Chinese poems. [Involve process of auditory perception and verbal cognition]
Repeating	Participants repeated one line of Chinese poems that was played by the voice assistant, imitating its tone and speed (e.g., the voice assistant played “chú hé rì dāng wǔ” and participants repeated “chú hé rì dāng wǔ”). [Involve process of auditory perception, verbal cognition, and verbal responding]
Answering	Participants answered the next line in response to the line of Chinese poems played by the voice assistant, imitating its tone and speed (e.g., the voice assistant played “chú hé rì dāng wǔ” and participants answered the next line “hàn dī hé xià tǔ”). [Involve process of auditory perception, verbal cognition (higher), and verbal responding]

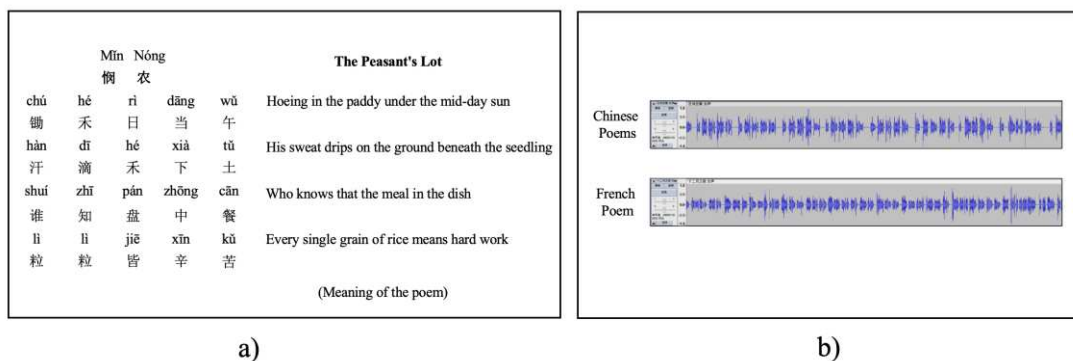


Figure 1. Illustration of countermeasure materials (a): an instance of Chinese poem, b): audio tracks of Chinese poem series and the French poem)

2.4 Measurements

The dependent variables included driving performance, physiological measures, eye movements, mental workload, and acceptance of countermeasures. The variables related to driving performance included standard deviation of lane position (SDLP), standard deviation of vehicle heading angular (SDVHA), and standard deviation of driving speed (SD-Speed). The literature has shown that a decrease of SDLP and SDVHA and an increase of SD-Speed are related to increased alertness and reduced driving drowsiness (Atchley et al., 2014; Large et al., 2017; Portouli et al., 2007).

The physiological measures included skin conductance level (SCL) (μS), the standard deviation of normal-to-normal intervals (SDNN), and breathing rate (breaths per minute). These variables were extracted from the raw data of the physiological signals of electrodermal activity, cardiac activity, and respiratory activity, respectively. The literature has shown that an increase of these variables is associated with the enhance of alertness (Bundele & Banerjee, 2010; De Naurois et al., 2019; Kiashari et al., 2018). In addition, these variables are all sensitive to age. All the SDNN (Bonnemeier et al., 2003), SCL, and the standard deviation of the SCL (Doberenz et al., 2011) tend to decrease with age.

The eye movement measures included pupil diameter (PD) (mm), proportion of visit duration on the road center (PVRC) (%), and blink rate (times per minute). PVRC was calculated as the proportion of visit duration within the road center area during the time of interest (TOI). The literature has shown that an increase of PD and PVRC and an decrease of blink rate are associated with enhanced alertness and reduced drowsiness (De Naurois et al., 2019; Large et al., 2017).

Subjective ratings for both mental workload and acceptance for countermeasures were collected. The mental workload was measured using the unweighted NASA-TLX scale (Hart & Staveland, 1988). This scale is composed of six sub-items: mental demand, physical demand, temporal demand, performance, effort, and frustration. The acceptance for countermeasure was collected using a questionnaire comprising five

items for usefulness and four items for satisfaction (Van Der Laan et al., 1997). A 7-point Likert scale was used in this questionnaire to evaluate drivers' opinions on different countermeasures.

2.5 Drowsiness inducement and detection

Drowsiness was induced by manipulating the awake time, the experimental environment, and the driving task. For the awake time, the participant was required to wake up one hour earlier than usual on the day of his visit and stayed awake for 16 hours before the experiment. The experiment visit time ranged from 21:00 PM to 01:00 AM, lasting one hour for each participant. The specific visiting time for each participant was scheduled based on their wake-up time. Moreover, the participant was instructed to refrain from caffeine, mints, teas, alcohol, or related beverages, smoking, or having a nap for the entire day of the experiment. For the environment, the experiment was conducted in a quiet and dark room. For the driving task, a monotonous night driving scenario was created by the driving simulator software. The participant was asked to follow a truck traveling at a constant speed of 90 km/h on a four-lane highway (two lanes in each direction), with a speed limit of 120 km/h.

Participants' drowsiness level was detected through self-reported Karolinska Sleepiness (KSS) ratings and performance on a psychomotor vigilance task (PVT). The KSS was used to measure the participants' drowsiness level (Åkerstedt & Gillberg, 1990). It was based on a 9-point Likert scale from 1 (extremely alert) to 9 (very sleepy, great effort to keep awake, fighting sleep). Every 5 minutes, the message "Please report your state based on the KSS scale" was emitted to request the participant to report his own drowsiness level. The participant rated his drowsiness by pressing a number button on the keypad on the steering wheel. One minute after each KSS rating, the participant performed the PVT, a sensitive drowsiness detection method requiring a button-press response to visual stimuli. A decrease of PVT performance often indicates drowsiness (Basner & Dinges, 2011; Philip et al., 2005). In this study, a simplified PVT was designed. The participant was required to press the R3 button on the steering wheel

whenever a red square icon in the gray area changed to triangle (see Figure 2). These red triangle icons appeared at random intervals (2-10 seconds) with a 1-second display duration. Responses with a reaction time of 600 ms or more were classified as lapses (Kozak et al., 2005). The PVT was conducted every eight minutes during driving and lasted for approximately two minutes. The experimenter implemented the drowsiness countermeasures only when the participants' KSS rating was equal to or larger than 7 and their PVT accuracy was lower than 70%.

2.6 Apparatus

2.6.1 Driving simulator

The driving simulator in this study was composed of a driving-simulated software (STISIM DRIVE Build 3.22.05, System Technology Inc., Hawthorne, CA, USA), an operational system, a Dell workstation (Dell Precision 3640 Tower, Dell Inc., Round Rock, TX, USA), an audio system, and a 65-inch monitor with 4k resolution (see Figure 2). The operation system included a steering wheel with force feedback, a brake pedal, an acceleration pedal, and a clutch pedal. Driving data was recorded at a frequency of 120 Hz.

2.6.2 Physiological equipment

An eye tracker (Tobii X3 Pro, Tobii Inc., Danderyd, Stockholm, Sweden) was used to track participants' eye movement behaviors with the pupillary responses at a sampling rate of 250 Hz. A portable wireless physiological system, ErgoLAB (Kingfar Co., Beijing, China), with modules of electrodermal activity (EDA), photoplethysmography (PPG), and respiration (RESP), was used to record the physiological signals at 64 Hz. As Figure 2 shows, the EDA was measured using two electrode patches attached to the index and ring fingers. The PPG was captured by an infrared sensor on the left earlobe, and RESP was collected via a strap on the abdomen.

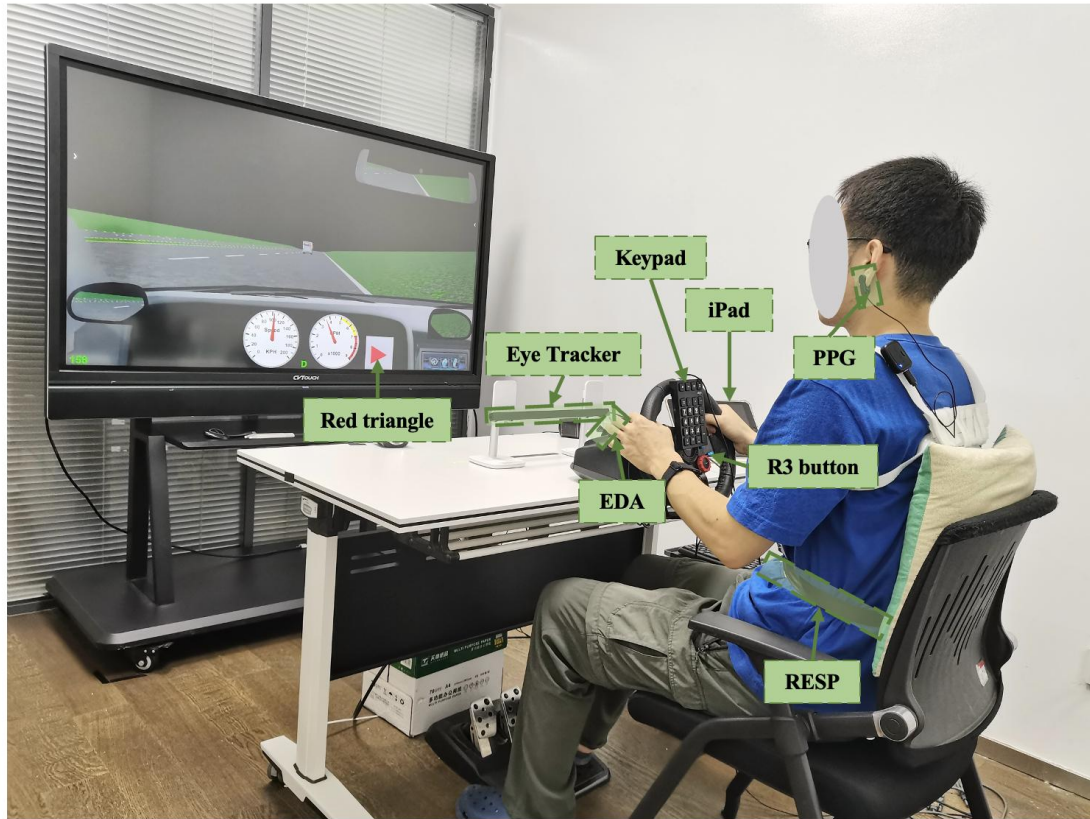


Figure 2. Experimental setup

2.6.3 Drowsiness detection and countermeasure equipment

The drowsiness detection and countermeasure equipment included the Eprime 3.0 software (Psychology Software Tools, Inc. Sharpsburg, PA, USA), a ThinkCentre workstation (ThinkCentre M80, Lenovo Ltd., Beijing, China), an iPad (iPad 9th generation, Apple Inc., Cupertino, CA, USA), a numeric keypad, and an audio system. The iPad positioned at the top right corner of the desk was used to display the KSS scale, while the keypad was installed on the steering wheel and was used to collect KSS responses. The ThinkCentre workstation was linked to the Dell system that ran the driving scenario, enabling real-time PVT responses collection. A customized voice interaction program in Eprime 3.0 enabled sequential or line-by-line playback of poems to simulate an interactive voice assistant. Interfaces of real-time PVT response results, KSS reporting outcomes, and voice interaction programs were positioned out of participant's view.

2.7 Procedures

Each participant completed five visits, scheduled one week apart. At the first visit, the experimenter introduced the objectives of the study and the noninvasive physiological system to the participant. The participant signed an informed consent form and completed a demographic questionnaire about his age, driving experience, and health state.

The experiment process for each visit was similar. At first, the experimenter helped the participants to wear all physiological devices, completed eye movement calibration, and had the participant rested for 5 minutes. Then, the participant completed two practice sessions: a three-minute session to familiarize him with the driving simulator and a five-minute session to practice KSS reporting and PVT while driving. The participant was instructed to keep his thumbs near the R3 button on the steering wheel to complete PVT. During the formal driving test, the participant drove continuously and completed KSS reporting and the PVT for approximately every 5 minutes (see Figure 3 a)). For drowsiness detection and countermeasure implementation, this study employed the widely used “Wizard of Oz” (WoZ) method (Dahlbäck et al., 1993). The participant was told that he was interacting with a real digital voice assistant. In fact, the interaction was operated by a female experimenter, who monitored the experiment process without the awareness of the participant. She implemented countermeasures when each participant’s drowsiness level met the countermeasure criterion (i.e., $KSS \geq 7$ & $PVT \text{ accuracy} \leq 70\%$). To minimize the potential alerting effect of PVT, all countermeasures were conducted one minute after the last PVT. Once the countermeasure criterion was met, the experimenter either played a continuous poem for the meaninglessly or meaningfully listening conditions, controlled the line-by-line playback of poems based on participants’ responses for the repeating and answering conditions, or played nothing for the no countermeasure condition. This was accomplished by using the customized voice interaction program. Following each countermeasure, the participant reported his KSS rating immediately and completed a PVT again. After the experiment, the participant filled out questionnaires on mental

workload during the countermeasure and the subjective ratings for countermeasure acceptance. The experiment in each week lasted for 50-70 minutes, depending on when the participant's drowsiness level met the countering criteria. The sequence of the five countermeasure types was balanced by using a Latin Square design. After five visits, each participant received a cash compensation of 1,000 RMB (i.e., approximately 140 USD) for his participation in the experiment.

2.8 Data analysis

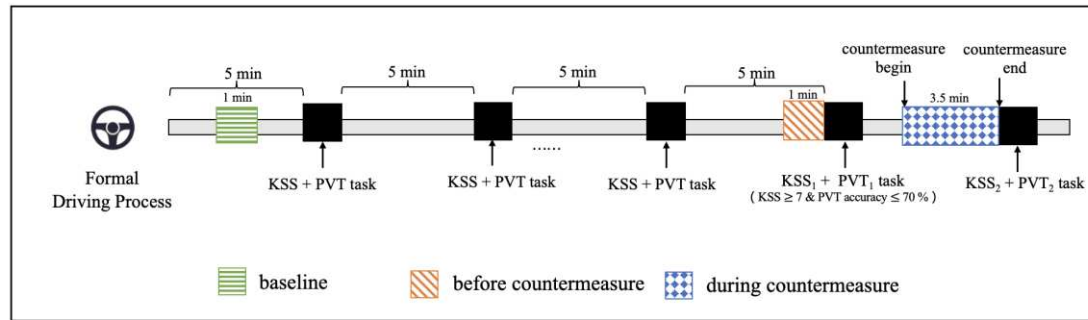
2.8.1 Data preprocessing

To evaluate the effectiveness of the countermeasures, different timings regarding the implementation of countermeasures were extracted for indicators (see Figure 3 b)). For KSS ratings and PVT accuracy, the corresponding values (i.e., KSS_1 and PVT_1), which exactly met the countermeasure criteria, were treated as the “before countermeasure” indicators. The KSS ratings and PVT accuracy (i.e., KSS_2 and PVT_2) after the countermeasure were treated as the “after countermeasure” indicators.

Driving performance data were exported from the STISIM software and averaged according to different time intervals. Figure 3 b) shows that the period of 1 minute right before the instructional speech for KSS_1 rating served as the time interval of the “before countermeasure” (i.e., orange square) and the exact 3.5-minute duration of the countermeasure served as the time interval of “during countermeasure” (i.e., blue square).

Physiological and eye movement signals were first extracted and averaged using the ErgoLAB software. The time intervals of “before countermeasure” and “during countermeasure” for both indicators were set same as driving performance data. Then, a multistage normalization method was applied to normalize data across different age groups, baseline states, and weekly calibration variances (Yang et al., 2021). Data were normalized by dividing the mean values from the “before countermeasure” or “during countermeasure” intervals by the mean “baseline” value. The baseline interval was

defined as a one-minute period from the start of the formal driving session in each week (i.e., green square) (see Figure 3 b)).



a)

Indicators	Before countermeasure	During countermeasure	After countermeasure
KSS, PVT accuracy	KSS_1, PVT_1		KSS_2, PVT_2
Driving behaviors	Average in	Average in	
Physiological indicators (normalized)	Average in Average in	Average in Average in	
Eye movement indicators (normalized)	Average in Average in	Average in Average in	

b)

Figure 3. a) formal driving process in each week and b) data preprocess for different indicators in each week

2.8.2 Statistical analysis

RStudio (2023.10.31) was used for statistical analysis. For ordinal data, the non-parametric Friedman test was conducted (Friedman, 1937). If the Friedman test showed statistically significant differences, a Wilcoxon signed-rank test was employed for the *post-hoc* test (Wilcoxon, 1992). For continuous data, the linear mixed model (LMM) accounting for both fixed and random effects, was adopted (Baayen et al., 2008; Wan & Sarter, 2022). Residual normality was checked using Q-Q plots and the one-sample Kolmogorov-Smirnov test (Venables & Ripley, 2002). The homogeneity of variance was assessed using the Levene's test (Levene, 1960). If the model assumption was violated, the data was transformed into a logarithm function. In the LMM data analysis, age group, countermeasure type, countermeasure timing, and their interaction effects

were considered as fixed-effect variables. The individual variance was considered as a random variable. The paired *t*-test was used for the *post hoc* analysis if the main effect was significant and for the simple effect analysis if the interaction effect was significant. The significant level of all analyses was set at 0.05 and corrected by a Bonferroni adjustment.

To check the success of manipulation and the countermeasure outcomes, this study examined the effects of countermeasure type on KSS ratings and PVT accuracy by separating age group and countermeasure timing. To compare the effectiveness of different countermeasures on alerting drowsy drivers, the effects of age group, countermeasure type, countermeasure timing, and their interaction effects on driving performance, physiological indicators, and eye movement indicators were analyzed. In addition, the heterogeneity of participant in the effect of overall countermeasures on the above indicators was analyzed by adopting causal processing methods. Finally, the mental workload and acceptance in different countermeasures between two age groups were analyzed. Data and analysis code can be found in the following link (<https://github.com/zhangweibetty/Drowsy-countermeasures>).

3 Results

3.1 Manipulation checks and results

3.1.1 Karolinska Sleepiness ratings

For the young group, before the countermeasure, there was no main effect of countermeasure type on KSS ratings ($\chi^2(4) = 3.591, p = 0.464$). After the countermeasure, the main effect of countermeasure type on the KSS rating was significant ($\chi^2(4) = 29.813, p < 0.001$) (see Figure 4). *Post hoc* analysis suggested that the answering condition ($M = 4.767, SE = 0.518$) led to the lowest KSS rating, which was significantly lower than those of the meaningfully listening ($M = 7.100, SE = 0.391$) ($p < 0.01$), the meaninglessly listening ($M = 7.567, SE = 0.271$) ($p < 0.05$), and the no countermeasure ($M = 8.033, SE = 0.172$) ($p < 0.05$) conditions. The repeating condition

($M = 6.067$, $SE = 0.396$) achieved the second lowest KSS ratings, which was significantly lower than that of the no countermeasure condition ($p < 0.05$). No significant differences in KSS ratings were observed between the answering and repeating conditions.

For the middle-aged group, before the countermeasure, there was a main effect of countermeasure type on KSS ratings ($\chi^2(4) = 10.488$, $p < 0.05$), but *post hoc* analysis revealed no significant differences among the five countermeasure types. After the countermeasure, the main effect of countermeasure type on KSS ratings was significant ($\chi^2(4) = 21.780$, $p < 0.001$) (see Figure 4). The answering condition ($M = 5.821$, $SE = 0.468$) produced the lowest KSS rating, which was significantly lower than both the meaningless listening ($M = 7.536$, $SE = 0.350$) and no-countermeasure ($M = 7.929$, $SE = 0.245$) conditions ($p < 0.05$ for each comparison).

3.1.2 Psychomotor vigilance task accuracy

For the young group, the main effect of countermeasure type on PVT accuracy was not significant before the countermeasure ($F(4, 56) = 0.822$, $p = 0.517$) nor after the countermeasure ($F(4, 56) = 1.114$, $p = 0.359$). For the middle-aged group, the main effect of countermeasure type on PVT accuracy was not significant before the countermeasure ($F(4, 65) = 0.926$, $p = 0.455$). However, after the countermeasure, the main effect of countermeasure type on PVT accuracy was significant ($F(4, 65) = 2.533$, $p < 0.05$) (see Figure 4). *Post hoc* analysis suggested that both the answering ($M = 0.651$, $SE = 0.040$) and meaningfully listening ($M = 0.657$, $SE = 0.057$) conditions resulted in significant higher PVT accuracy than the no countermeasure condition ($M = 0.425$, $SE = 0.074$) ($p < 0.05$ for each comparison).

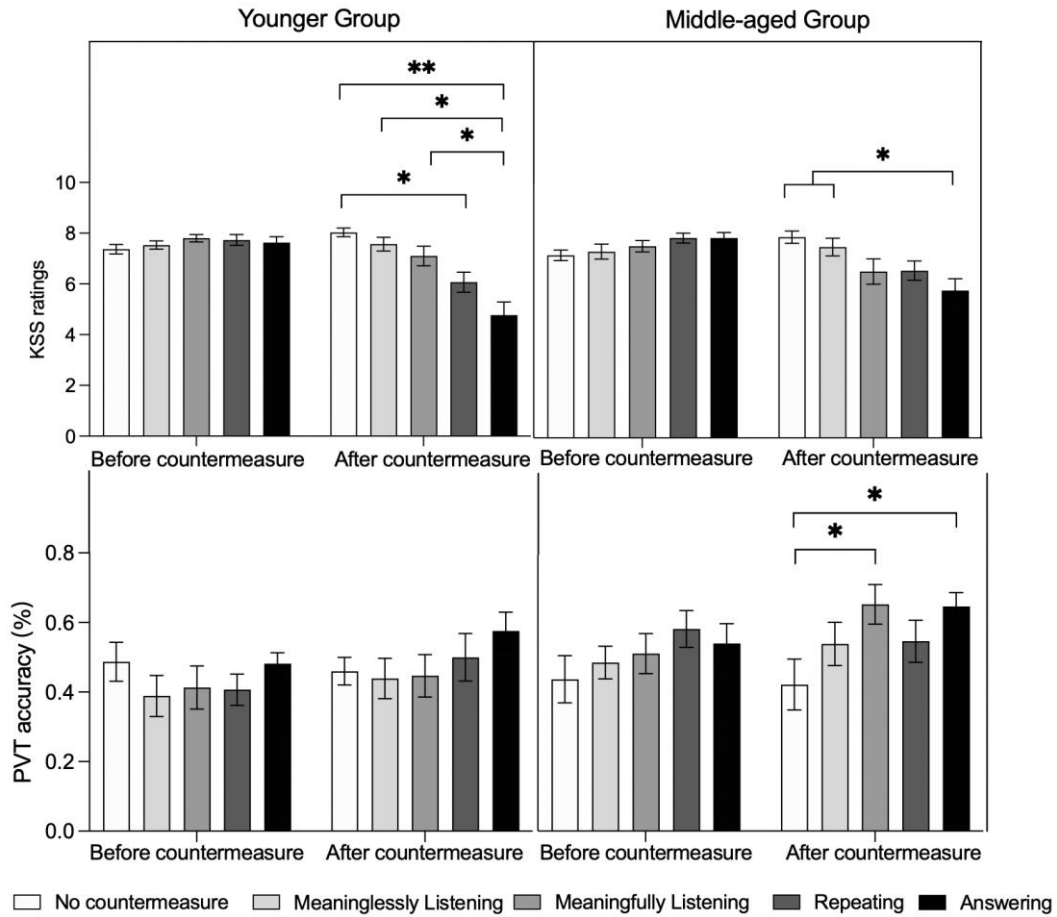


Figure 4. Means and standard errors of KSS ratings and PVT accuracy for different countermeasure types before and after countermeasures (* $p < 0.05$; ** $p < 0.01$)

3.2 Driving performance

3.2.1 Standard deviation of line position

Although no main effects of countermeasure type or countermeasure timing on SDLP were found, there were significant interaction effects between countermeasure type and timing (see Table 3). Figure 5 shows that no significant differences were observed on SDLP among the five countermeasure types before the countermeasure. However, during the countermeasure, participants had smaller SDLP when repeating ($M = 0.796$, $SE = 0.041$) than when meaningfully listening ($M = 1.018$, $SE = 0.069$) ($p < 0.01$), meaninglessly listening ($M = 1.087$, $SE = 0.059$) ($p < 0.001$), and receiving no countermeasure ($M = 1.067$, $SE = 0.087$) ($p < 0.05$). Additionally, SDLP was also smaller in the answering condition ($M = 0.832$, $SE = 0.064$) than in the meaninglessly listening condition ($p < 0.01$).

Table 3. Effects of age group, countermeasure type, and countermeasure timing on dependent variables

<i>Driving behaviors</i>						
	SDLP		SDVHA		SD-speed	
	F(p)	η_p^2	F(p)	η_p^2	F(p)	η_p^2
Age group	0.009	< 0.001	0.451	0.016	0.370	0.102
Countermeasure type	1.200	0.019	1.060	0.017	1.190	0.014
Countermeasure timing	0.226	0.001	2.661	0.011	27.620 ***	0.019
Age group $\times$ Countermeasure type	0.001	0.008	1.422	0.023	4.308 **	0.066
Age group $\times$ Countermeasure timing	0.503	< 0.001	0.780	0.003	0.254	0.001
Countermeasure timing $\times$ Countermeasure type	6.917 ***	0.102	8.052 ***	0.117	1.442	0.023
Age group $\times$ Countermeasure timing $\times$ Countermeasure type	0.385	0.006	0.423	0.007	0.607	0.010
<i>Physiological measurements</i>						
	SCL		SDNN		Breathing rate	
	F(p)	η_p^2	F(p)	η_p^2	F(p)	η_p^2
Age group	0.007	< 0.001	0.267	0.010	0.047	0.002
Countermeasure type	9.084 ***	0.130	2.265 +	0.036	2.069	0.033
Countermeasure timing			17.266			
	30.998 ***	0.113	***	0.066	0.776	0.003
Age group $\times$ Countermeasure type	1.607	0.026	0.397	0.006	1.181	0.019
Age group $\times$ Countermeasure timing	0.031	< 0.001	0.057	< 0.001	0.660	0.003
Countermeasure timing $\times$ Countermeasure type	5.913 ***	0.089	6.626 ***	0.098	0.648	0.011
Age group $\times$ Countermeasure timing $\times$ Countermeasure type	0.159	0.003	0.392	0.006	0.668	0.011
<i>Eye movement indicators</i>						
	Pupil diameter		PVRC		Blink rate	
	F(p)	η_p^2	F(p)	η_p^2	F(p)	η_p^2
Age group	0.008	< 0.001	2.123	0.073	0.493	0.018
Countermeasure type	9.571 ***	0.136	0.613	0.010	1.669	0.027
Countermeasure timing	29.240 ***	0.107	4.115 *	0.017	0.756	0.003
Age group $\times$ Countermeasure type	0.974	0.016	0.499	0.008	0.965	0.016
Age group $\times$ Countermeasure timing	0.695	0.003	0.762	0.003	0.011	< 0.001
Countermeasure timing $\times$ Countermeasure type	12.349 ***	0.169	2.901 *	0.046	0.262	0.004
Age group $\times$ Countermeasure timing $\times$ Countermeasure type	0.354	0.006	0.406	0.007	0.364	0.005

Note: F(p) indicates the F values and significance signs, suggesting * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; η_p^2 indicates the partial eta square, namely, determines the effect size; SDLP- standard deviation of lane position; SDVHA- standard deviation of vehicle heading angular; SDSpeed- standard deviation of driving speed; SCL- skin conductance level; SDNN- standard deviation of normal-to-normal intervals; PVRC- proportion of visit duration on the road center.

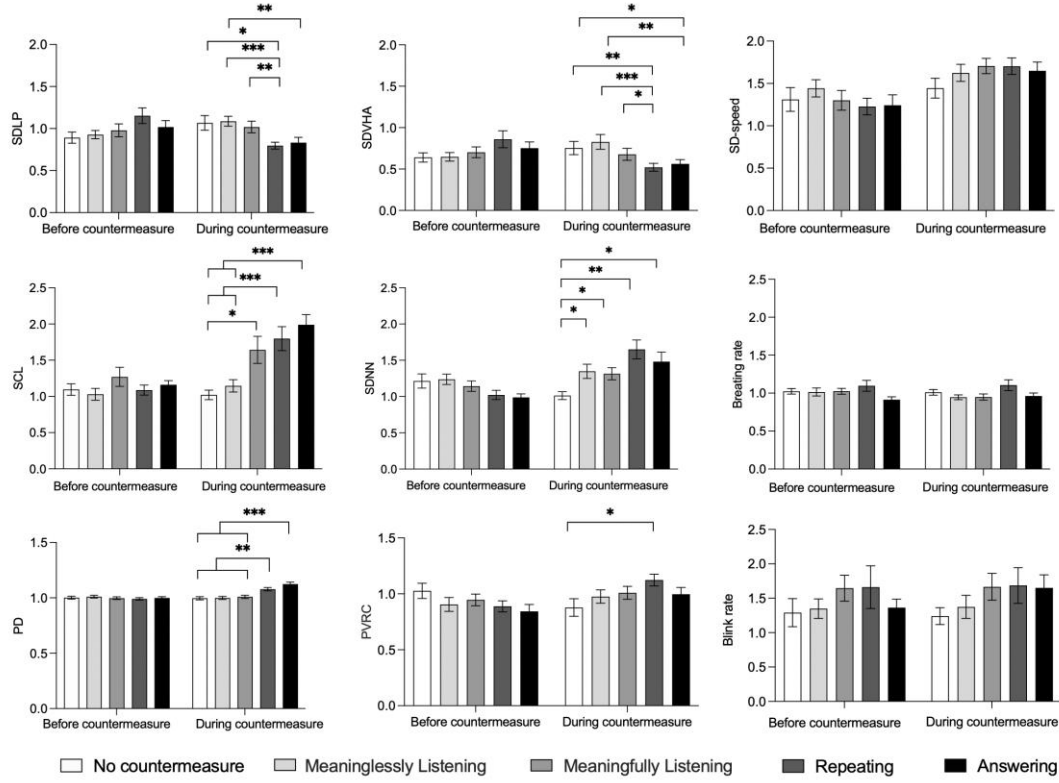


Figure 5. Means and standard errors of dependent variables for different countermeasure types before and during countermeasures (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$)

3.2.2 Standard deviation of vehicle heading angular

Table 3 shows that there were significant interaction effects between countermeasure type and countermeasure timing, although the main effects of countermeasure type and countermeasure timing were not significant. Simple effect analysis suggests that before the countermeasure, no significant differences were observed on SDVHA ($F(4, 112) = 1.620, p = 0.174$) (see Figure 5). However, during the countermeasure, the five countermeasure types produced significant differences ($F(4, 112) = 10.74, p < 0.001$). Participants' SDVHA when answering ($M = 0.563, SE = 0.052$) was significantly lower than accepting no countermeasure ($M = 0.755, SE = 0.080$) ($p < 0.05$) and meaninglessly listening ($M = 0.828, SE = 0.089$) ($p < 0.01$). Moreover, SDVHA in the repeating condition ($M = 0.522, SE = 0.047$) was also lower

than that in the no countermeasure ($p < 0.01$), meaninglessly listening ($p < 0.001$), and meaningfully listening ($M = 0.678$, $SE = 0.073$) ($p < 0.05$) conditions.

3.2.3 Standard deviation of driving speed

There was a main effect of countermeasure timing on SD-speed (see Table 3). Participants' SD-speed during the countermeasure ($M = 4.897$, $SE = 0.632$) was significantly larger than that before the countermeasure ($M = 3.528$, $SE = .643$) ($p < 0.001$). Although the interaction effects between age group and countermeasure type were significant, comparisons of different countermeasure types within each age group were not reported when countermeasure timing was not considered.

3.3 Physiological measurements

3.3.1 Skin conductance level

There was a significant main effect of countermeasure type on SCL (see Table 3). *Post hoc* analysis results suggested that the answering condition ($M = 1.575$, $SE = 0.131$) resulted in higher SCL than the meaninglessly listening ($M = 1.088$, $SE = 0.084$) and no countermeasure ($M = 1.059$, $SE = 0.073$) conditions ($p < 0.001$ for each comparison). The SCL in the repeating condition ($M = 1.443$, $SE = 0.142$) was also higher than in the meaninglessly listening ($p < 0.05$) and no countermeasure ($p < 0.01$) conditions. The meaningfully listening also showed increased SCL when compared with the no countermeasure condition ($p < 0.05$).

The main effect of countermeasure timing on SCL was significant. Participants' SCL during the countermeasure ($M = 1.520$, $SE = 0.151$) was higher than that before the countermeasure ($M = 1.129$, $SE = 0.088$) ($p < 0.001$). The interaction effect between countermeasure type and countermeasure timing was also significant. Figure 5 shows there were no significant differences among countermeasure types in SCL ($F(4,112) = 0.913$, $p = 0.459$) before the countermeasure was presented. However, the five countermeasure types differed on SCL ($F(4,112) = 12.035$, $p < 0.001$) during the countermeasure. Specifically, SCL was higher in the answering condition ($M = 1.990$, $SE = 0.140$) compared to the meaninglessly listening ($M = 1.147$, $SE = 0.085$) and no

countermeasure ($M = 1.022$, $SE = 0.066$) conditions ($p < 0.001$ for each comparison). The repeating condition ($M = 1.798$, $SE = 0.165$) also produced higher SCL than the meaningfully listening ($M = 1.643$, $SE = 0.187$) and no countermeasure conditions. The meaningfully listening condition resulted in higher SCL than the no countermeasure condition ($p < 0.05$).

3.3.2 Standard deviation of normal-to-normal intervals

The main effect of countermeasure timing on SDNN was significant (see Table 3). *Post hoc* analysis suggested that participants had higher SDNN during the countermeasure ($M = 1.361$, $SE = 0.110$) than before the countermeasure ($M = 1.121$, $SE = 0.074$) ($p < 0.001$). Significant interaction effects between countermeasure type and timing on SDNN were observed.

Figure 5 shows that there were no significant differences in SDNN among the five countermeasure types before the countermeasure was implemented. However, during the countermeasure, SDNN was significantly smaller in the no countermeasure condition ($M = 1.013$, $SE = 0.054$) compared to the meaninglessly listening ($M = 1.348$, $SE = 0.098$) ($p < 0.05$), meaningfully listening ($M = 1.314$, $SE = 0.084$) ($p < 0.05$), repeating ($M = 1.650$, $SE = 0.130$) ($p < 0.01$), and answering ($M = 1.481$, $SE = 0.132$) ($p < 0.05$) conditions.

3.3.3 Breathing rate

No significant main effects nor interaction effects of age group, countermeasure type, and countermeasure timing on breathing rate were found (see Table 3).

3.4 Eye movement variables

3.4.1 Pupil diameter

Table 3 shows that there was a significant main effect of countermeasure type on participants' PD. Participants had a significant larger PD in the answering condition ($M = 1.062$, $SE = 0.019$) than in the meaningfully listening ($M = 1.003$, $SE = 0.013$), meaninglessly listening ($M = 1.005$, $SE = 0.013$), and no countermeasure ($M = 0.999$, $SE = 0.013$) conditions ($p < 0.01$ for each comparison).

The main effect of countermeasure timing on PD was also significant. Participants generally had larger PD during the countermeasure ($M = 1.042$, $SE = 0.018$) than before the countermeasure ($M = 1.00$, $SE = 0.012$) ($p < 0.001$). The interaction effects between countermeasure type and countermeasure timing were also significant. Figure 5 shows that there were no significant differences in the impact of the five countermeasure types on PD before the countermeasure was presented. During the countermeasure, however, the answering condition ($M = 1.124$, $SE = 0.018$) resulted in significantly larger PD than the meaningfully listening ($M = 1.009$, $SE = 0.015$), meaninglessly listening ($M = 0.999$, $SE = 0.014$), and no countermeasure conditions ($M = 0.997$, $SE = 0.014$) ($p < 0.001$ for each comparison). The repeating condition ($M = 1.078$, $SE = 0.015$) also resulted in a larger PD than the other three conditions, except for the answering condition ($p < 0.01$ for each comparison).

3.4.2 Proportion of visit duration on the road center

There was a main effect of countermeasure timing on PVRC (see Table 3). Participants had higher PVRC during the countermeasure ($M = 0.997$, $SE = 0.063$) than before the countermeasure was presented ($M = .922$, $SE = 0.060$) ($p < 0.01$).

The interaction effects between countermeasure type and countermeasure timing on PVRC were also significant. Figure 5 shows that the five countermeasure conditions produced no significant differences on the PVRC before the countermeasure was presented. During the countermeasure, the repeating condition ($M = 1.124$, $SE = 0.053$) led to a higher PVRC than the no countermeasure condition ($M = 0.878$, $SE = 0.079$) ($p < 0.05$).

3.4.3 Blink rate

No main effects or interaction effects of age group, countermeasure type, and countermeasure timing were found on blink rate (see Table 3).

3.5 Heterogeneity

This study further explored heterogeneity of participant in the effect of overall countermeasures by adopting causal processing methods (Bolger et al., 2019;

Goodridge et al., 2024). A rule of thumb indicates that causal effect heterogeneity is noteworthy if its standard deviation is 0.25 (i.e., relative size) or greater relative to the fixed effect (Bolger et al., 2019). The overall countermeasures (see Table 4), therefore, demonstrated notable causal heterogeneity effect of participant on SDLP, SCL, SDNN, and the breathing rate.

Table 4. Heterogeneity of participant on the effects of overall countermeasures.

	Parameter estimates		Relative size
	Fixed effect	SD	
<i>Driving performance</i>			
SDLP	0.125	0.057	0.457
SDVH	0.149	0.030	0.201
SD-speed	0.946	0.154	0.163
<i>Physiological indicators</i>			
SCL	0.844	0.241	0.285
SDNN	0.356	0.103	0.289
Breathing rate	0.105	0.053	0.501
<i>Eye movement indicators</i>			
PVRC	0.150	0.009	0.058
Pupil Diameter	0.126	0.019	0.149
Blink rate	0.180	0.002	0.010

Note: SD: standard deviation; relative size = SD/ Fixed effect.

Furthermore, given that age was more variable in the middle-aged group (SD = 8.026) than in the young group (SD = 2.615), this study explored whether age could predict between-participant heterogeneity in SDLP, SCL, SDNN, and the breathing rate. By adding the effects of age and its interaction effects with countermeasures in LMM models, the results suggested that age accounts for 78.27%, 25.05%, 17.67%, and 35.97% of the between-participants heterogeneity in the causal effects of countermeasures on SDLP, SCL, SDNN, and breathing rate, respectively. As shown in Figure 6, what is particularly interesting is the relationship between age and the total implied heterogeneity of the countermeasures on the physiological indicators. Namely, the

effects of the countermeasures on all measures of physiological activation were more variable for the middle-aged drivers than the young ones. Hence whilst there were no interaction effects in physiological activation as a function of age and countermeasures, it seems that middle-aged drivers tended to be more variable in physiological activation.

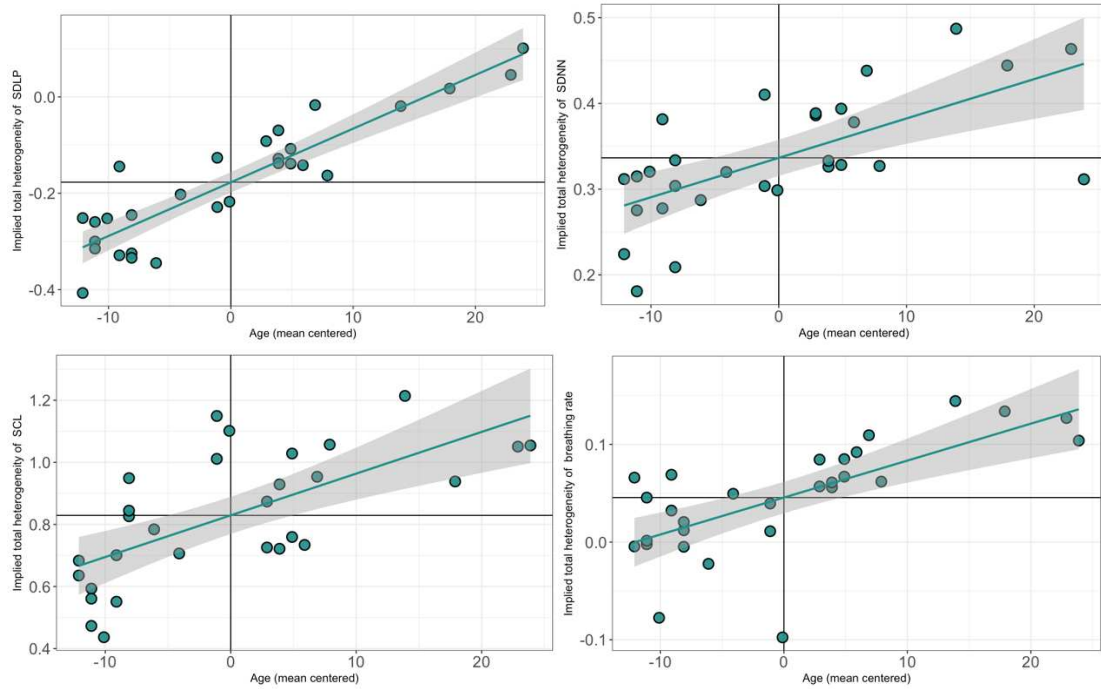


Figure 6. Relationship between the implied total heterogeneity of dependent variables and mean centered age. The vertical line on the x axis denotes mean age. The horizontal line on the y axis denotes the average effect of overall countermeasures on SDLP, SDNN, SCL, and breathing rate, respectively.

3.6 Subjective ratings

3.6.1 Mental workload

The main effect of countermeasure type was significant on the NASA-TLX score. *Post hoc* test results suggested that the repeating condition ($M = 37.334$, $SE = 4.841$) resulted in a significant lower workload score than the no countermeasure condition ($M = 44.828$, $SE = 4.710$) ($p < 0.05$).

There was a marginal significant three-way interaction effects among age group, countermeasure type, and workload dimension. For the young group, the interaction effects between countermeasure type and dimension were significant ($F(20, 406) =$

1.891, $p < 0.05$). Figure 7 shows, that the answering condition ($M = 58.667$, $SE = 7.784$) produced a higher score of mental demand than the meaninglessly listening condition ($M = 29.333$, $SE = 5.873$) ($p < 0.05$). However, the answering condition ($M = 18.333$, $SE = 3.984$) resulted in a lower workload score on frustration dimension compared to the meaninglessly listening condition ($M = 46.667$, $SE = 8.247$) ($p < 0.05$). For the middle-aged group, the interaction effects between countermeasure type and workload dimension were not significant ($F(20, 377) = 0.745$, $p = 0.779$).

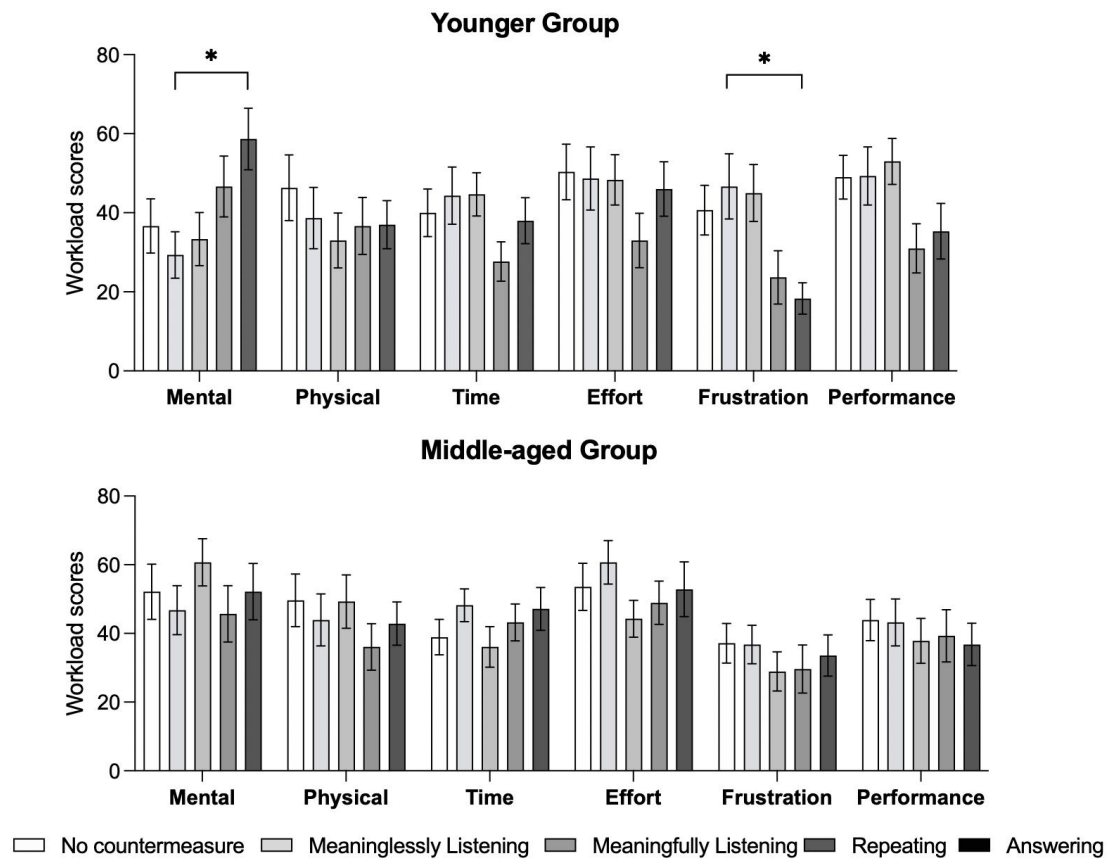


Figure 7. Means and standard errors of subjective rating of mental workload for countermeasure types and age groups (* $p < 0.05$)

3.6.2 Acceptance

For the young group, the Friedman test indicated a significant main effect of countermeasure type on usefulness ratings ($\chi^2(3) = 24.147$, $p < 0.001$) (see Figure 8 a)). *Post-hoc* analysis suggested that answering ($M = 1.587$, $SE = 0.111$) was perceived as the most useful countermeasure, with score significantly higher than those for

meaningfully listening ($M = 0.160$, $SE = 0.331$) ($p < 0.05$) and meaninglessly listening ($M = -0.733$, $SE = 0.267$) ($p < 0.01$). Repeating ($M = 1.120$, $SE = 0.258$) was rated as the second most useful method. It was also significantly more useful than meaninglessly listening ($p < 0.05$). The main effect of countermeasure type on satisfaction ratings was also significant ($\chi^2(3) = 19.468$, $p < 0.001$) (see Figure 8 c)). Compared to the meaninglessly listening condition ($M = -0.683$, $SE = 0.208$), the answering ($M = 1.083$, $SE = 0.190$) ($p < 0.01$) and repeating ($M = 0.750$, $SE = 0.234$) ($p < 0.05$) conditions were rated as more satisfying.

For the middle-aged group, the Friedman test results suggested a significant main effect of countermeasure type on usefulness ratings ($\chi^2(3) = 14.614$, $p < 0.01$). Figure 8 b) shows that both answering ($M = 1.329$, $SE = 0.262$) and repeating ($M = 1.143$, $SE = 0.243$) were perceived more useful than meaninglessly listening ($M = -0.057$, $SE = 0.315$) and meaningfully listening ($M = 0.686$, $SE = 0.276$) ($p < 0.05$ for each comparison). Although the main effect of countermeasure type on satisfaction ratings was not significant ($\chi^2(3) = 4.357$, $p = 0.225$). The paired- t tests indicated that participants rated answering ($M = 0.750$, $SE = 0.277$) as more satisfying than meaninglessly listening ($M = -0.036$, $SE = 0.250$) ($p < 0.05$) (see Figure 8 d)).

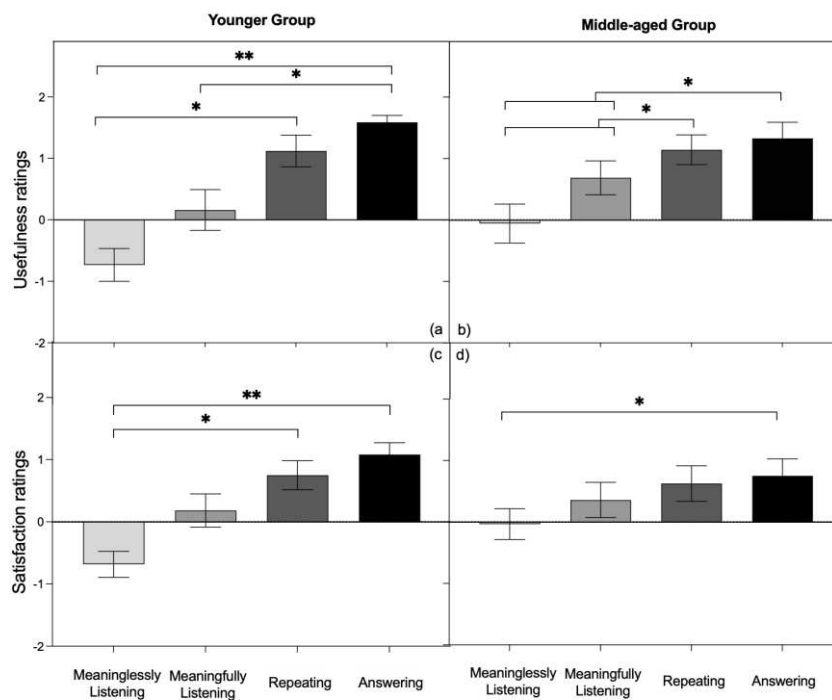


Figure 8. Means and standard errors of usefulness and satisfaction ratings for countermeasure types and age groups (* $p < 0.05$; ** $p < 0.01$).

4 Discussion

This study investigated the effects of different countermeasure type on alerting young and middle-aged drivers. Drivers' drowsiness was induced by monotonous driving at night and was detected by drivers' KSS ratings and PVT responses. The results of KSS ratings and PVT accuracy before the countermeasure suggested no significant differences among the five countermeasure types for both driver groups. It indicated that the drowsy state of drivers was consistent before the countermeasure was implemented.

4.1 Effects of different countermeasures by adopting voice assistants

It was found that the countermeasures involving more stages of resources provided greater alertness for drowsy drivers. The first hypothesis (H1) of this study was supported. Specifically, compared to the no countermeasure and listening conditions, the repeating and answering conditions were the two most effective countermeasures. It was reflected by drivers' more stable lane keeping performance, higher physiological arousal levels, greater attentions to the road, lower self-reported sleepiness, and higher PVT accuracy in repeating and answering conditions.

Driving is a cognitively demanding task, requiring drivers' perceptual abilities, attention abilities, memory, executive functions, etc. (Anstey et al., 2005). In drowsy states, drivers' perception ability, cognitive processing, and responses to driving-related stimuli are inhibited due to reduced physiological states and brain activation levels (Khan & Hong, 2015; Wörle et al., 2019). The listening tasks normally engage drivers' auditory perception only. Repeating and answering tasks, on the other hand, can induce a combination alert effect by animating drivers' auditory perception, working memory, and verbal responses through their active involvement. As a result, the combination of alerting effects leads to a more alert physical and mental state. This makes drivers to focus their attention better on the road and maintain more stable vehicle control. In

addition, higher usefulness and satisfaction scores for the repeating and answering conditions were found than the two listening conditions. This implies that participants' subjective acceptance ratings aligned with the objective findings. The present results were consistent with previous findings of the advantages of verbal processes (Atchley et al., 2014; Large et al., 2017; Mahajan, 2021; Orsini et al., 2024; Takayama & Nass, 2008) and listening tasks (Atchley et al., 2014; Takayama & Nass, 2008) in enhancing alertness of drowsy drivers. The current study further emphasized the underlying mechanisms for the superiority of verbal interactions based on resource stages, rather than simply comparing different countermeasure methods. This is an advancement beyond the previous studies.

A comparison between the answering and repeating conditions showed that increasing drivers' cognitive load led to the potential effect of achieving a more alert state of drivers. This supports our second hypothesis (H2). It is noticeable that among the five countermeasure conditions, the answering condition led to the highest mental workload young drivers. The increased mental workload, however, reduced the frustration levels for this group. Additionally, the answering condition produced a more alert state than the repeating condition. This was verified by the lower KSS ratings, higher SCL and PD, and smaller SDLP and SDVHA in answering condition, although the differences were not statistically significant. These findings were inconsistent with those of Saxby et al. (2017). Their study claimed that the increasing drivers' cognitive load by recalling of past events and their personal impacts failed to counter the subjective automation-induced fatigue and the objective loss of performance. This discrepancy may be due to the difference of the cognitive load between the two studies. The cognitive load in Saxby et al. (2017), induced by a close-call incident recall task, was believed to be higher than the poem recitation in our study.

Listening conditions, no matter the listening content was meaningless or meaningful, had limited effects on alerting drowsy drivers. The meaninglessly listening and meaningfully listening conditions did not significantly counter drowsiness beyond

minor increases in SCL and SDNN compared to the no countermeasure condition. This finding was inconsistent with those of Orsino et al. (2024)'s and Amirah and Puspasari (2019), who found listening countermeasures led to a significant reduction in driving drowsiness. This inconsistency could be attributed to facts that the participants in Orsino et al. (2024) were allowed to select the music they were familiar. Amirah and Puspasari (2019), on the other hand, adopted pop music. The familiarity and melody of the music may provide additional benefits in physiological (Gabrielsson & Wik, 2003) and motor responses (Vuong et al., 2023). Although the rhythmic poems adopted in the current study could also elicit emotional arousal (Wassiliwizky et al., 2017), they may not be as enjoyable as listening to familiar music. Nevertheless, it is noteworthy that listening to poems did not adversely affect driving performance nor aggregate drowsiness, which is consistent with the findings from a previous research (Ünal et al., 2013).

4.2 Effects of different countermeasures for age groups

The third hypothesis (H3) of this study claims that young drivers may benefit more from stronger countermeasures that involve more resource stages and higher cognitive load, while middle-aged drivers may suit better with lighter countermeasures. However, the young and middle-aged participants were not significantly different in response to different countermeasures. Despite varying levels of drowsiness susceptibility and cognitive abilities, participants in the two groups exhibited similar responses to countermeasures in terms of driving behavior, physiological measures, and eye movement patterns. Our hypothesis was not supported. This finding was inconsistent with that in Song et al. (2017), who claimed that young drivers benefited more from a verbal alert maintain task while older drivers' performance was less affected by fatigue. One possible explanation for this discrepancy is that, although drivers' resistance to drowsiness increases with age, this resistance might result from their greater mental efforts to avoid falling asleep (Lowden et al., 2009). The current study adopted countermeasures based on the same high drowsiness level for both age groups, the additional mental effort of middle-aged drivers might not be enough to compensate

their drowsiness. This resulted in similar effects of countermeasures for the two age groups. Nevertheless, the countermeasures in Song et al. (2017) were applied based on the same driving time period. In this condition, the drowsiness level between young and older drivers at the time of the countermeasure might be different. In addition, all the middle-aged participants in the current study were experienced drivers. They were less likely to increase their high-risk attention to secondary tasks over time during driving compared to less experienced drivers (Klauer et al., 2014). Therefore, although middle-aged drivers may have inferior attention allocation abilities than young drivers, their extensive experiences with dual-task management could compensate for this deficit. As a result, increasing resource stages and cognitive load did not diminish the countermeasure effects for middle-aged drivers.

Whilst there were no average differences of different countermeasures between age groups. The results of the heterogeneity analysis revealed that physiological activation attributed to the countermeasures was more variable for middle-aged drivers than young drivers. It has long been established that the whilst the central tendency of datasets can be the same, the spread of the data may differ (Anscombe, 1973; Friedman, 1937). A similar phenomenon may appear here. The heterogeneity results could imply that the countermeasures may not have the same physiological effect for different driving groups. The variability could manifest itself as middle-aged drivers having much higher physiological activation (i.e., SCL, SDNN, and breathing rate) to countermeasures than other demographics in certain instances. This pattern might be attributed to the greater physiological shifts of middle-aged drivers when transitioning from the sleepy state to countermeasure state. On the one hand, during drowsiness, age-related declines in autonomic nervous system function and associated physical states tended to result in lower physiological activation (Doberenz et al., 2011; Luebberding et al., 2014). For example, epidermal hydration decreases with age (Cho et al., 2019), reducing skin conductance level (Fowles & Venables, 1970) and skin elasticity (Lee et al., 2022). On the other hand, when countermeasures were applied, the middle-aged drivers might pay

more efforts than young drivers in responding to countermeasures while driving, to compensate their decreased cognitive abilities. This process could result in a relatively greater physiological arise from their drowsy baseline. These findings align with previous studies that older adults exhibit higher levels of cardiovascular reactivity (Hess & Ennis, 2012) and prefrontal cortex activation (Ohsugi et al., 2013) in response to cognitive engagement compared to young adults.

4.3 Practical implications

Findings of this experiment have practical implications for the design of in-vehicle voice assistants as real-world drowsiness countermeasures. First, to maintain the effectiveness of drowsiness countermeasures while minimizing potential adverse effects, voice-assistant-based countermeasures should be tailored to drivers' cognitive characteristics and physiological limitations. For young drivers, it is recommended to implement voice assistants that encourage them to speak and to think. These are effective in countering drowsiness by involving more resource stages and cognitive engagement. Nevertheless, countermeasures with excessive cognitive load should be avoided as they may lead to detrimental driving performance if they are manifested as distraction (D'Addario & Donmez, 2019). For middle-aged drivers, although they were also benefited from speaking and thinking, the design and implementation of voice-assistant-based countermeasures should be more cautious due to their higher variability in physiological activation to countermeasures. While the increased physiological activation is generally associated with improved alertness (Oken et al., 2006; Yang et al., 2017), it may also exacerbate declines in working memory (Riediger et al., 2014) and increase susceptibility to distraction (Öztürk et al., 2023), potentially threatening driving safety. Moreover, sudden increases in physiological arousal should be carefully managed for some middle-aged adults, given their higher susceptibility to cardiovascular diseases (Shih et al., 2011).

Second, it is recommended to personalize the features of voice assistants to encourage drivers to adopt them as drowsiness countermeasures. In practice, drivers

may avoid voice assistants due to low trust and satisfaction, influenced by design factors like voice characteristics, familiarity, anthropomorphism, and visual appeal (Lee & Jeon, 2022; Liu et al., 2023; Park et al., 2024). Furthermore, middle-aged adults differ significantly from young adults in their smart voice assistant usage requirements, preferences, and acceptance (Zhong et al., 2024; Gollasch et al., 2021). Although both driver groups in the current study showed high acceptance for overall countermeasures, they may have been partly driven by the novelty of the experimental context and their cooperation to “test” the system (Large et al., 2018). Therefore, personalizing voice assistant features to meet drivers’ individual needs and preferences is essential to refine drowsiness countermeasures, ensuring positive user experiences and improving effectiveness in countering drowsiness.

4.4 Limitations and future works

This study has several limitations. First, this study utilized the “Wizard of Oz” (WoZ) method to simulate the voice interaction by asking the participants to repeat and to answer rhythmic poems instead of interacting with a real voice assistant. Although these poems were familiar enough for participants to engage listening and speaking processes, the simulated interaction may not fully reflect the dynamics of a natural real-world conversation. Considering the benefits of verbal engagement in counteracting drowsiness, future research should explore more naturalistic dialogues to enhance drivers’ willingness to engage with voice assistants.

Second, this study compared the effectiveness of countermeasures only between young (18–35 years) and middle-aged drivers (36–64 years). The older drivers aged 65 and above were not included. The results of these countermeasures may not be generalizable to the older drivers due to the age-related decrease in physiological states (Luebberding et al., 2013) and cognitive capabilities (Anstey et al., 2005; Murman et al., 2015). Considering that many individuals aged 65 and above continue to drive despite moderate cognitive decline (Shimada et al., 2016), future research should investigate the effectiveness of these countermeasures for older drivers.

Third, due to a device limitation, certain sensitive eye movement indicators, such as PERCLOS (i.e., the percentage of time that the eyes are more than 80% closed), were not recorded. As PERCLOS is a highly validated drowsiness indicator (Abe, 2023), future studies should incorporate this parameter alongside other measures for a more robust drowsiness detection and countermeasure framework.

Fourth, countermeasure effects might be transient. Orsini et al. (2024) found that the effects of listening to self-selected music in countering drowsiness lasted for 12-25 mins. However, the current study focused only on the immediate effect of different countermeasures. How long alertness might be maintained by using our countermeasures remains unclear. As the superiority of repeating and answering countermeasures was confirmed in our findings, the duration of alertness associated with these countermeasures deserves further exploration.

5 Conclusion

This study conducted a simulated driving experiment to investigate the effects of five different countermeasures using voice assistants on countering drowsiness for young and middle-aged drivers. Results suggested that countermeasures involving more resource stages were more effective in alerting drowsy drivers. This finding was evidenced by lower self-reported sleepiness, more stable lane keeping, higher physiological arousal level, greater attention to the road, and higher PVT accuracy. In addition, drivers also had higher acceptance for the repeating and answering conditions compared to the no countermeasure and listening conditions. Moreover, the increased cognitive load of the answering condition may achieve a higher level of alertness than that of the repeating condition. Listening tasks had limited effects on alerting drowsy driver but did not adversely impact driving performance or drowsiness. In terms of age, young and middle-aged drivers did not differ on average in response to different countermeasures. Nevertheless, considering individual differences in drivers' physiological activation in reacting to countermeasures, drowsiness countermeasures

should be tailored to the individual needs, particularly for middle-aged drivers. Our findings not only clarify drowsiness countermeasure mechanisms but also offer practical guidance for vehicle designers to use in-voice assistants as anti-drowsiness tools.

Acknowledgment

We appreciate the support from National Natural Science Foundation of China with the Main Research Project on Machine Behavior and Human-Machine Collaborated Decision Making Methodology (Project Number: 72192820), and the Second Research Project on Human Behavior in Human-Machine Collaboration (Project Number: 72192824). We would like to express our special thanks to Dr. Jingyue Zheng for her support in the conceptualization of this study.

References

- Abe, T. (2023). PERCLOS-based technologies for detecting drowsiness: Current evidence and future directions. *SLEEP Advances*, 4(1), zpad006. <https://doi.org/10.1093/sleepadvances/zpad006>
- Aidman, E., Chadunow, C., Johnson, K., & Reece, J. (2015). Real-time driver drowsiness feedback improves driver alertness and self-reported driving performance. *Accident Analysis & Prevention*, 81, 8–13. <https://doi.org/10.1016/j.aap.2015.03.041>
- Åkerstedt, T., & Gillberg, M. (1990). Subjective and Objective Sleepiness in the Active Individual. *International Journal of Neuroscience*, 52(1–2), 29–37. <https://doi.org/10.3109/00207459008994241>
- Amirah, A. Y., & Puspasari, M. A. (2019). Music as Countermeasure for Driving Fatigue Using Brain Signal Indicator. *Proceedings of the 5th International Conference on Industrial and Business Engineering*, 169–172. <https://doi.org/10.1145/3364335.3364365>
- Anscombe, F. J. (1973). Graphs in Statistical Analysis. *The American Statistician*. <https://www.tandfonline.com/doi/abs/10.1080/00031305.1973.10478966>
- Anstey, K., Wood, J., Lord, S., & Walker, J. (2005). Cognitive, sensory and physical factors enabling driving safety in older adults. *Clinical Psychology Review*, 25(1), 45–65. <https://doi.org/10.1016/j.cpr.2004.07.008>
- APA Dictionary of Psychology. (n.d.). <https://dictionary.apa.org/drowsiness>
- Arakawa, T. (2021). A Review of Heartbeat Detection Systems for Automotive Applications. *Sensors*, 21(18), 6112. <https://doi.org/10.3390/s21186112>
- Atchley, P., Chan, M., & Gregersen, S. (2014). A Strategically Timed Verbal Task Improves Performance and Neurophysiological Alertness During Fatiguing Drives. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 56(3), 453–462. <https://doi.org/10.1177/0018720813500305>

- Ayas, S., Donmez, B., & Tang, X. (2023). Drowsiness Mitigation Through Driver State Monitoring Systems: A Scoping Review. *Human Factors*, 00187208231208523. <https://doi.org/10.1177/00187208231208523>
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412. <https://doi.org/10.1016/j.jml.2007.12.005>
- Basner, M., & Dinges, D. F. (2011). Maximizing Sensitivity of the Psychomotor Vigilance Test (PVT) to Sleep Loss. *Sleep*, 34(5), 581–591. <https://doi.org/10.1093/sleep/34.5.581>
- Bolger, N., Zee, K. S., Rossignac-Milon, M., & Hassin, R. R. (2019). Causal processes in psychology are heterogeneous. *Journal of Experimental Psychology: General*, 148(4), 601–618. <https://doi.org/10.1037/xge0000558>
- Bonnemeier, H., Wiegand, U. K., Brandes, A., Kluge, N., Katus, H. A., Richardt, G., & Potratz, J. (2003). Circadian profile of cardiac autonomic nervous modulation in healthy subjects: differing effects of aging and gender on heart rate variability. *Journal of cardiovascular electrophysiology*, 14(8), 791-799.
- Braitman, K. A., & Williams, A. F. (2011). Changes in self-regulatory driving among older drivers over time. *Traffic injury prevention*, 12(6), 568-575.
- Bundele, M. M., & Banerjee, R. (2010). ROC analysis of a fatigue classifier for vehicular drivers. *2010 5th IEEE International Conference Intelligent Systems*, 296–301. <https://doi.org/10.1109/IS.2010.5548362>
- Charlton, J. L., Oxley, J., Fildes, B., Oxley, P., Newstead, S., Koppel, S., & O'Hare, M. (2006). Characteristics of older drivers who adopt self-regulatory driving behaviours. *Transportation Research Part F: Traffic Psychology and Behaviour*, 9(5), 363-373.
- Cho, C., Cho, E., Kim, N., Shin, J., Woo, S., Lee, E., ... & Ha, J. (2019). Age - related biophysical changes of the epidermal and dermal skin in Korean women. *Skin Research and Technology*, 25(4), 504-511.

- D'Addario, P., & Donmez, B. (2019). The effect of cognitive distraction on perception-response time to unexpected abrupt and gradually onset roadway hazards. *Accident Analysis & Prevention*, 127, 177–185. <https://doi.org/10.1016/j.aap.2019.03.003>
- Dahlbäck, N., Jönsson, A., & Ahrenberg, L. (1993, February). Wizard of Oz studies: why and how. In Proceedings of the 1st international conference on Intelligent user interfaces (pp. 193-200).
- De Naurois, C. J., Bourdin, C., Stratulat, A., Diaz, E., & Vercher, J.-L. (2019). Detection and prediction of driver drowsiness using artificial neural network models. *Accident Analysis & Prevention*, 126, 95–104. <https://doi.org/10.1016/j.aap.2017.11.038>
- Doberenz, S., Roth, W. T., Wollburg, E., Maslowski, N. I., & Kim, S. (2011). Methodological considerations in ambulatory skin conductance monitoring. *International Journal of Psychophysiology*, 80(2), 87-95.
- Doroudgar, S., Chuang, H. M., Perry, P. J., Thomas, K., Bohnert, K., & Canedo, J. (2017). Driving performance comparing older versus younger drivers. *Traffic Injury Prevention*, 18(1), 41–46. <https://doi.org/10.1080/15389588.2016.1194980>
- Fraser, S. A., Li, K. Z. H., & Penhune, V. B. (2010). Dual-task performance reveals increased involvement of executive control in fine motor sequencing in healthy aging. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, 65(5), 526–535. <https://doi.org/10.1093/geronb/gbq036>
- Friedman, M. (1937). The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance. *Journal of the American Statistical Association*. <https://www.tandfonline.com/doi/abs/10.1080/01621459.1937.10503522>
- Fowles, D. C., & Venables, P. H. (1970). The effects of epidermal hydration and sodium reabsorption on palmar skin potential. *Psychological Bulletin*, 73(5), 363.
- Gabrielsson, A., & Wik, S. L. (2003). Strong Experiences Related to Music: A descriptive System. *Musicae Scientiae*, 7(2), 157–217. <https://doi.org/10.1177/102986490300700201>

- Gershon, P., Ronen, A., Oron-Gilad, T., & Shinar, D. (2009). The effects of an interactive cognitive task (ICT) in suppressing fatigue symptoms in driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, 12(1), 21–28. <https://doi.org/10.1016/j.trf.2008.06.004>
- Gershon, P., Shinar, D., Oron-Gilad, T., Parmet, Y., & Ronen, A. (2011). Usage and perceived effectiveness of fatigue countermeasures for professional and nonprofessional drivers. *Accident Analysis & Prevention*, 43(3), 797–803. <https://doi.org/10.1016/j.aap.2010.10.027>
- Gollasch, D., & Weber, G. (2021). Age-related differences in preferences for using voice assistants. In *Proceedings of Mensch und Computer 2021* (pp. 156-167).
- Goodridge, C. M., Gonçalves, R. C., Arabian, A., Horrobin, A., Solernou, A., Lee, Y. T., Lee, Y. M., Madigan, R., & Merat, N. (2024). Gaze entropy metrics for mental workload estimation are heterogenous during hands-off level 2 automation. *Accident Analysis & Prevention*, 202, 107560. <https://doi.org/10.1016/j.aap.2024.107560>
- G*Power. (2023). *G * Power 3.1 manual*.
- Grech, M. R., Neal, A., Yeo, G., Humphreys, M., & Smith, S. (2009). An examination of the relationship between workload and fatigue within and across consecutive days of work: Is the relationship static or dynamic? *Journal of Occupational Health Psychology*, 14(3), 231–242. <https://doi.org/10.1037/a0014952>
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In P. A. Hancock & N. Meshkati (Eds.), *Advances in Psychology* (Vol. 52, pp. 139–183). North-Holland. [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9)
- Hess, T. M., & Ennis, G. E. (2012). Age differences in the effort and costs associated with cognitive activity. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 67(4), 447-455.

- Jackson, P., Hilditch, C., Holmes, A., Reed, N., Merat, N., & Smith, L. (2011). Fatigue and Road Safety: A Critical Analysis of Recent Evidence. *Road Safety Web Publication*, 21(44).
- Johns, M. W. (1991). A New Method for Measuring Daytime Sleepiness: The Epworth Sleepiness Scale. *Sleep*, 14(6), 540–545. <https://doi.org/10.1093/sleep/14.6.540>
- Kaplan, S., & Prato, C. G. (2012). Associating Crash Avoidance Maneuvers With Driver Attributes and Accident Characteristics: A Mixed Logit Model Approach. *Traffic Injury Prevention*, 13(3), 315–326. <https://doi.org/10.1080/15389588.2011.654015>
- Khan, M. J., & Hong, K.-S. (2015). Passive BCI based on drowsiness detection: An fNIRS study. *Biomedical Optics Express*, 6(10), 4063. <https://doi.org/10.1364/BOE.6.004063>
- Kiashari, S. E. H., Nahvi, A., Homayounfard, A., & Bakhoda, H. (2018). *Monitoring the Variation in Driver Respiration Rate from Wakefulness to Drowsiness: A Non-Intrusive Method for Drowsiness Detection Using Thermal Imaging*. 3(1).
- Klauer, S. G., Guo, F., Simons-Morton, B. G., Ouimet, M. C., Lee, S. E., & Dingus, T. A. (2014). Distracted Driving and Risk of Road Crashes among Novice and Experienced Drivers. *New England Journal of Medicine*, 370(1), 54–59. <https://doi.org/10.1056/NEJMsa1204142>
- Kozak, K., Curry, R., Greenberg, J., Artz, B., Blommer, M., & Cathey, L. (2005). Leading Indicators of Drowsiness in Simulated Driving. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 49(22), 1917–1921. <https://doi.org/10.1177/154193120504902207>
- Large, D. R., Burnett, G., Antrobus, V., & Skrypchuk, L. (2017). *Stimulating Conversation: Engaging Drivers in Natural Language Interactions with an Autonomous Digital Driving Assistant to Counteract Passive Task-Related Fatigue*. In International conference on driver distraction and inattention.

- Large, D. R., Burnett, G., Antrobus, V., & Skrypchuk, L. (2018). Driven to discussion: engaging drivers in conversation with a digital assistant as a countermeasure to passive task - related fatigue. *IET Intelligent Transport Systems*, 12(6), 420-426.
- Lee, S. C., & Jeon, M. (2022). A systematic review of functions and design features of in-vehicle agents. *International Journal of Human-Computer Studies*, 165, 102864.
- Lee, H. J., Park, S. R., Kwon, D. I., Park, M. S., & Lim, D. H. (2022). Depth profiling of epidermal hydration inducing improvement of skin roughness and elasticity: In vivo study by confocal Raman spectroscopy. *Journal of Cosmetic Dermatology*, 21(10), 4810-4817.
- Levene, H. (1960). *Robust Tests for Equality of Variances*. Stanford University Press.
- Liu, J., Wan, F., Zou, J., & Zhang, J. (2023). Exploring factors affecting people's willingness to use a voice-based in-car assistant in electric cars: An empirical study. *World Electric Vehicle Journal*, 14(3), 73.
- Lowden, A., Anund, A., Kecklund, G., Peters, B., & Åkerstedt, T. (2009). Wakefulness in young and elderly subjects driving at night in a car simulator. *Accident Analysis & Prevention*, 41(5), 1001–1007. <https://doi.org/10.1016/j.aap.2009.05.014>
- Lu, K., Sjörs Dahlman, A., Karlsson, J., & Candefjord, S. (2022). Detecting driver fatigue using heart rate variability: A systematic review. *Accident Analysis & Prevention*, 178, 106830. <https://doi.org/10.1016/j.aap.2022.106830>
- Luebberding, S., Krueger, N., & Kerscher, M. (2014). Mechanical properties of human skin in vivo: a comparative evaluation in 300 men and women. *Skin Research and Technology*, 20(2), 127-135.
- Mahajan, K., Large, D. R., Burnett, G., & Velaga, N. R. (2021). Exploring the benefits of conversing with a digital voice assistant during automated driving: A parametric duration model of takeover time. *Transportation research part F: traffic psychology and behaviour*, 80, 104-126.
- May, J. F., & Baldwin, C. L. (2009). Driver fatigue: The importance of identifying causal factors of fatigue when considering detection and countermeasure

- technologies. *Transportation Research Part F: Traffic Psychology and Behaviour*, 12(3), 218–224. <https://doi.org/10.1016/j.trf.2008.11.005>
- Mazer, B., Chen, Y.-T., Vrkljan, B., Marshall, S. C., Charlton, J. L., Koppel, S., & G  linas, I. (2021). Comparison of older and middle-aged drivers' driving performance in a naturalistic setting. *Accident Analysis & Prevention*, 161, 106343. <https://doi.org/10.1016/j.aap.2021.106343>
- Meng, F., Li, S., Cao, L., Li, M., Peng, Q., Wang, C., & Zhang, W. (2015). Driving Fatigue in Professional Drivers: A Survey of Truck and Taxi Drivers. *Traffic Injury Prevention*, 16(5), 474–483. <https://doi.org/10.1080/15389588.2014.973945>
- Murman, D. L. (2015, August). The impact of age on cognition. In *Seminars in hearing* (Vol. 36, No. 03, pp. 111-121). Thieme Medical Publishers.
- National Sleep Foundation. (2023). *National Sleep Foundation's Drowsy Driving Survey*.
- NHTSA. (2023). *Fatality and Injury Reporting System Tool (FIRST) Version 5.6*. <https://cdan.dot.gov/query>
- Nishigaki, M., & Shirakata, T. (2019). Driver attention level estimation using driver model identification. *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, 3520–3525. <https://doi.org/10.1109/ITSC.2019.8917380>
- Nordbakke, S., & Sagberg, F. (2007). Sleepy at the wheel: Knowledge, symptoms and behaviour among car drivers. *Transportation Research Part F: Traffic Psychology and Behaviour*, 10(1), 1–10. <https://doi.org/10.1016/j.trf.2006.03.003>
- Obst, P., Armstrong, K., Smith, S., & Banks, T. (2011). Age and gender comparisons of driving while sleepy: Behaviours and risk perceptions. *Transportation Research Part F: Traffic Psychology and Behaviour*, 14(6), 539–542. <https://doi.org/10.1016/j.trf.2011.06.005>

- Ohsugi, H., Ohgi, S., Shigemori, K., & Schneider, E. B. (2013). Differences in dual-task performance and prefrontal cortex activation between younger and older adults. *BMC neuroscience*, 14, 1-9.
- Oken, B. S., Salinsky, M. C., & Elsas, S. M. (2006). Vigilance, alertness, or sustained attention: Physiological basis and measurement. *Clinical Neurophysiology : Official Journal of the International Federation of Clinical Neurophysiology*, 117(9), 1885–1901. <https://doi.org/10.1016/j.clinph.2006.01.017>
- Orsini, F., Baldassa, A., Grassi, M., Cellini, N., & Rossi, R. (2024). Music as a countermeasure to fatigue: A driving simulator study. *Transportation Research Part F: Traffic Psychology and Behaviour*, 103, 290–305. <https://doi.org/10.1016/j.trf.2024.04.016>
- Öztürk, İ., Merat, N., Rowe, R., & Fotios, S. (2023). The effect of cognitive load on Detection-Response Task (DRT) performance during day- and night-time driving: A driving simulator study with young and older drivers. *Transportation Research Part F: Traffic Psychology and Behaviour*, 97, 155–169. <https://doi.org/10.1016/j.trf.2023.07.002>
- Papachristos, E., Merritt, T., Schneiders, E., Jahanshiri, D., Pir, A., & Ciobanu, A. (2024, September). Sound Strategies for Safe Driving: Exploring Auditory Interventions to Counteract Passive Driver Fatigue. In *Proceedings of the Second International Symposium on Trustworthy Autonomous Systems* (pp. 1-6).
- Park, D., Lee, Y., & Kim, Y. M. (2024). Effects of autonomous driving context and anthropomorphism of in-vehicle voice agents on intimacy, trust, and intention to use. *International Journal of Human–Computer Interaction*, 40(22), 7179-7192.
- Philip, P., Sagaspe, P., Moore, N., Taillard, J., Charles, A., Guilleminault, C., & Bioulac, B. (2005). Fatigue, sleep restriction and driving performance. *Accident Analysis & Prevention*, 37(3), 473–478. <https://doi.org/10.1016/j.aap.2004.07.007>
- Portouli, E., Bekiaris, E., Papakostopoulos, V., & Maglaveras, N. (2007). On-road experiment for collecting driving behavioural data of sleepy drivers. *Somnologie*

- *Schlafforschung Und Schlafmedizin*, 11(4), 259–267.
<https://doi.org/10.1007/s11818-007-0319-3>
- RationalStat LLC. (2023). *Automotive Voice Industry Report, 2023-2024*.
<http://www.researchinchina.com/Htmls/Report/2023/73945.html>
- Riediger, M., Wrzus, C., Klipker, K., Müller, V., Schmiedek, F., & Wagner, G. G. (2014). Outside of the laboratory: Associations of working-memory performance with psychological and physiological arousal vary with age. *Psychology and aging*, 29(1), 103.
- Saxby, D. J., Matthews, G., & Neubauer, C. (2017). The relationship between cell phone use and management of driver fatigue: It's complicated. *Journal of Safety Research*, 61, 129–140. <https://doi.org/10.1016/j.jsr.2017.02.016>
- Shih, H., Lee, B., Lee, R. J., & Boyle, A. J. (2011). The aging heart and post-infarction left ventricular remodeling. *Journal of the American College of Cardiology*, 57(1), 9-17.
- Shimada, H., Tsutsumimoto, K., Lee, S., Doi, T., Makizako, H., Lee, S., ... & Suzuki, T. (2016). Driving continuity in cognitively impaired older drivers. *Geriatrics & gerontology international*, 16(4), 508-514.
- Smith, S. S., Horswill, M. S., Chambers, B., & Wetton, M. (2009). Hazard perception in novice and experienced drivers: The effects of sleepiness. *Accident Analysis & Prevention*, 41(4), 729–733. <https://doi.org/10.1016/j.aap.2009.03.016>
- Song, W., Woon, F. L., Doong, A., Persad, C., Tijerina, L., Pandit, P., Cline, C., & Giordani, B. (2017). Fatigue in Younger and Older Drivers: Effectiveness of an Alertness-Maintaining Task. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 59(6), 995–1008.
<https://doi.org/10.1177/0018720817706811>
- Tacken, M., Marcellini, F., Mollenkopf, H., Ruoppila, I., & Szeman, Z. (2005). Use and acceptance of new technology by older people. Findings of the international

- MOBILATE survey: 'Enhancing mobility in later life'. *Gerontechnology*, 3(3), 126-137.
- Takayama, L., & Nass, C. (2008). Assessing the Effectiveness of Interactive Media in Improving Drowsy Driver Safety. *Human Factors*, 50(5), 772–781. <https://doi.org/10.1518/001872008X312341>
- Ünal, A. B., de Waard, D., Epstude, K., & Steg, L. (2013). Driving with music: Effects on arousal and performance. *Transportation Research Part F: Traffic Psychology and Behaviour*, 21, 52–65. <https://doi.org/10.1016/j.trf.2013.09.004>
- UN DESA (United Nations, Department of Economic and Social Affairs, Population Division) (2024). World Population Prospects 2024, Online Edition.
- Van Der Laan, J. D., Heino, A., & De Waard, D. (1997). A simple procedure for the assessment of acceptance of advanced transport telematics. *Transportation Research Part C: Emerging Technologies*, 5(1), 1-10.
- Vaz Fragoso, C. A., Van Ness, P. H., Araujo, K. L., Iannone, L. P., & Klar Yaggi, H. (2015). Age - Related Differences in Sleep - Wake Symptoms of Adults Undergoing Polysomnography. *Journal of the American Geriatrics Society*, 63(9), 1845-1851.
- Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S*. Springer New York. <https://doi.org/10.1007/978-0-387-21706-2>
- Vuong, V., Hewan, P., Perron, M., Thaut, M. H., & Alain, C. (2023). The neural bases of familiar music listening in healthy individuals: An activation likelihood estimation meta-analysis. *Neuroscience & Biobehavioral Reviews*, 154, 105423. <https://doi.org/10.1016/j.neubiorev.2023.105423>
- Wan, Y., & Sarter, N. (2022). Attention Limitations in the Detection and Identification of Alarms in Close Temporal Proximity. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 001872082110639. <https://doi.org/10.1177/00187208211063991>

- Wassiliwizky, E., Koelsch, S., Wagner, V., Jacobsen, T., & Menninghaus, W. (2017). The emotional power of poetry: Neural circuitry, psychophysiology and compositional principles. *Social Cognitive and Affective Neuroscience*, 12(8), 1229–1240. <https://doi.org/10.1093/scan/nsx069>
- Watling, C., & Watling, H. (2015). Sleepy driving and drink driving: attitudes, behaviours, and perceived legitimacy of enforcement of younger and older drivers. In Proceedings of the 2015 Australasian Road Safety Conference (ARSC2015) (pp. 1-10). Australasian College of Road Safety (ACRS).
- Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, 3(2), Article 2.
- Wilcoxon, F. (1992). Individual Comparisons by Ranking Methods. In S. Kotz & N. L. Johnson (Eds.), *Breakthroughs in Statistics: Methodology and Distribution* (pp. 196–202). Springer. https://doi.org/10.1007/978-1-4612-4380-9_16
- Wörle, J., Metz, B., Thiele, C., & Weller, G. (2019). Detecting sleep in drivers during highly automated driving: The potential of physiological parameters. *IET Intelligent Transport Systems*, 13(8), 1241–1248. <https://doi.org/10.1049/iet-its.2018.5529>
- Yang, S., Kuo, J., Lenné, M. G., Fitzharris, M., Horberry, T., Blay, K., Wood, D., Mulvihill, C., & Truche, C. (2021). The Impacts of Temporal Variation and Individual Differences in Driver Cognitive Workload on ECG-Based Detection. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 63(5), 772–787. <https://doi.org/10.1177/0018720821990484>
- Yang, Z., Jia, W., Liu, G., & Sun, M. (2017). Quantifying mental arousal levels in daily living using additional heart rate. *Biomedical Signal Processing and Control*, 33, 368–378. <https://doi.org/10.1016/j.bspc.2016.11.003>
- Yılmaz, M. B., & Rızvanoğlu, K. (2022). Understanding users' behavioral intention to use voice assistants on smartphones through the integrated model of user

satisfaction and technology acceptance: a survey approach. *Journal of Engineering, Design and Technology*, 20(6), 1738-1764.

Zhong, R., Ma, M., Zhou, Y., Lin, Q., Li, L., & Zhang, N. (2024). User acceptance of smart home voice assistant: a comparison among younger, middle-aged, and older adults. *Universal Access in the Information Society*, 23(1), 275-292.