Journal of Safety Research 89 (2024) 172–180

Contents lists available at ScienceDirect

Journal of Safety Research

journal homepage: www.elsevier.com/locate/jsr

Audiovisual presentation of variable message signs improves message processing in distracted drivers during partially automated driving

Marina Pi-Ruano [a,b,1], Javier Roca [a,b,1,*], Pilar Tejero [a,c,1]

[a] ERI-Lectura, Universitat de València (UVEG), Spain
[b] Departamento de Psicología Evolutiva y de la Educación, Universitat de València (UVEG), Spain
[c] Departamento de Psicología Básica, Universitat de València (UVEG), Spain

ARTICLE INFO

Keywords:
Audiovisual messages
Autonomous vehicle
Distraction
Eye movements
Variable message signs

ABSTRACT

Introduction: Highly automated driving is expected to reduce the accident risk occurrence by human errors, but it can also increase driver distraction. Previous evidence shows that auditory signals can help drivers take over in critical situations. However, it is still uncertain whether the potential benefit of verbal auditory signals could be generalized to driving situations where drivers are visually and auditorily distracted. *Method:* Our first objective was to compare the effectiveness of complementary audio messages (audio + visual condition) and visual only (visual condition) variable message signs (VMS) messages. The second objective was to explore the potential use of oral messages with traffic information to help highly-automated vehicle drivers identify critical situations. Eye-tracking data were also registered. Twenty-four volunteers participated in a driving simulator study, completing two tasks: (a) a TV series task, where they had to pay attention to an episode of a TV series while traveling along the route; and (b) a VMS task, where they had to recover the manual control of the car if the VMS message was a 'critical message.' *Results:* General results showed that, when the audio was available, the participants: (a) had a higher ability to discriminate the VMS messages, (b) were less conservative, (c) responded earlier, and (d) their pattern of fixations was more efficient. A complementary analysis showed that the counterbalance order was a moderating factor for the discrimination ability and the response distance measures. This evidence suggests a potential learning effect, not cancelled by counterbalancing the order of the conditions. *Conclusion:* The processing of traffic messages may improve when provided as oral and visual messages. *Practical Applications:* These results would be of special interest for engineers designing highly automated cars, considering that the design of automated systems must ensure that the driver's attention is sufficient to take over control.

1. Introduction

In past years, an increasing number of active safety systems have been integrated in vehicles, and highly automated driving is becoming a reality (Gold, Damböck, Lorenz, & Bengler, 2013). The Society of Automotive Engineers (SAE) established six levels of driving automation (SAE, 2021): from level 0, wherein no automation exists at all; to level 5, wherein the driver is not required to drive at any time. Currently, a considerable number of new vehicles sold would include a Level 2 system, in which the driver assistance system executes both steering and acceleration/deceleration functions for most of the time, and the human driver supervises the system performance while they deal with the remaining aspects of the dynamic driving task. This implies that the automatic features may be intended for use only under certain operational design domains like highways (Eriksson & Stanton, 2017). This emerging technology is expected to reduce the accident risk occurrence by human errors and has the potential to ensure safe driving (Han & Ju, 2021). However, the automated driving is able to increase driver distraction as well (Lee, McGehee, Brown, & Reyes, 2002), and thus the design of automated systems must ensure that the driver's attention is sufficient to take over control and manage the driving situation.

Driver distraction has been defined as "a diversion of attention from driving, because the driver is temporarily focusing on an object, person, task or event not related to driving, which reduces the driver's awareness, decision making ability and/or performance, leading to an increased risk of corrective actions, near-crashes, or crashes" (Regan

* Corresponding author.
E-mail addresses: marina.pi@uv.es (M. Pi-Ruano), javier.roca@uv.es (J. Roca), pilar.tejero@uv.es (P. Tejero).
[1] Avenida Blasco Ibáñez, 21, 46010 Valencia (Spain).

https://doi.org/10.1016/j.jsr.2024.01.014
Received 23 May 2023; Received in revised form 3 November 2023; Accepted 30 January 2024
Available online 10 February 2024

et al., 2008). In SAE 2 level vehicles, as drivers are temporarily freed from some basic driving operations, they may engage in hazardous, non-driving related tasks (NDRTs), such as social networking or watching a video, which can involve physical, visual, auditory and cognitive distraction (Donmez et al., 2006; Han & Ju, 2021; Papantoniou, Papadimitriou, & Yannis, 2017; SAE, 2021).

Driver distraction affects attention negatively (Stanton & Young, 2005) and different strategies have been implemented to deal with it in autonomous vehicles (Beattie, Baillie, Halvey, & McCall, 2014; Chen, Sabic, Mishler, Parker, & Yamaguchi, 2022; Geitner, Biondi, Skrypchuk, Jennings, & Birrell, 2019; Han & Ju, 2021). In particular, most of the previous research focused on comparing different types of warning signals to require the driver to take over when automated driving fails or is limited. For example, Geitner et al. (2019) compared auditory, tactile, and multimodal (auditory-tactile) warnings of unexpected events during automated driving and found that reactions times were slower for the tactile warning, while no differences were found for auditory and multimodal warnings. Han and Ju (2021) studied visual and auditory alarms for drowsy and distracted states, and found that participants recognized faster the auditory information compared to the visual one, independently of the attentional state of the participants. Consequently, these authors concluded that only giving visual warnings to drivers is not enough to make them change their attention states to perform a successful situation recognition. In addition, taking into account the driver attentional state, it was found that auditory alarms with an intense output would be appropriate for a distracted state. More specifically, Beattie et al. (2014) studied if spatialized auditory feedback (that is, the sounds of driving actions like acceleration, braking or indications were presented in different spatial locations) could be used to enhance driver awareness to the intended actions of autonomous vehicles. They found that their spatialized auditory presentation method alerted drivers much more compared to when there was no sound available. Recently, Chen et al. (2022) focused their research on auditory alarms and examined how the directionality of lateral warnings affected drivers' detection and actions to avoid collisions.

In general, it could be considered that auditory alarms would outperform visual warnings alone because driving is already a visually demanding task (Sabic et al., 2021). However, it is still uncertain whether the potential benefit of auditory signals could be generalized to other driving situations, for example, when drivers are both visually and auditorily distracted (e.g., engaged in an audiovisual distracting activity, such as watching a video on a cell phone). Moreover, the effects of more complex auditory messages on driver's attention, such as verbal messages providing critical traffic information, have been rarely reported in the literature. In this regard, Tejero, Pi-Ruano and Roca (2020) studied the role of auditory verbal messages, which were given as a complement for the visual content of variable message signs (VMS), while drivers with and without dyslexia were engaged in manual driving in a simulator. They found that the availability of audio versions of the messages had positive effects for all participants (with and without dyslexia), as shown by higher accuracy in a message classification task, as well as better driving performance (in particular, safer following distances to a leading car). These results suggested that the processing of the traffic messages was easier when they were received as oral messages, possibly because the car-following and the VMS reading tasks draw on the same visual resources, which were not required to process an oral message.

The objective of the current study is twofold. First, it aims at expanding the previous results reported by Tejero et al. (2020) to participants distracted by an audiovisual secondary task. Therefore, our first goal is to compare the effectiveness of complementary audio messages (audio + visual condition) and visual only (visual condition) VMS messages. According to the previous evidence provided by Tejero et al. (2020), we could expect that the audio + visual messages would help drivers discriminate critical and non-critical traffic messages. But, importantly, in the current experiment, drivers were engaged in an audiovisual distracting activity. Therefore, our results would be useful to discuss whether the previously reported advantage of the audio + visual messages is a consequence of the competition for visual resources when only visual tasks are required to the participants, or it would still be observed even when participants are engaged in an audiovisual activity.

In addition, in the current study we will explore the potential use of oral messages with traffic information to help highly-automated vehicle drivers to identify critical situations, which require the driver to take over control. When driving on a highway, one can encounter VMS providing critical information on the current state of the road and traffic. So far, a driver using a SAE 2 level car (or even a SAE 3 level at some particular situations) is responsible of monitoring such information to decide how to act accordingly. In this context, we expected that providing complementary audio versions of VMS contents would allow drivers in a SAE 2 level of automation (therefore, not required to control the vehicle's course and speed) better discriminate critical and non-critical traffic messages, even that they are engaged in an audiovisual distracting activity.

Finally, we also registered eye-tracking data to provide an estimation of the visual resources required by the visual and the audio + visual messages. To this respect, we expected that the visual demands of audio + visual messages were lower, as evidenced by differences in the eye-movement pattern (e.g., fewer and faster glances at the VMS).

2. Materials and methods

2.1. Participants

The initial sample consisted of 29 adult volunteers (16 women). All of them were native speakers of Spanish and had normal or corrected-to-normal vision. The participants were recruited through advertisements at the University of Valencia and the Polytechnic University of Valencia, and by using a recruiting online questionnaire, which was available at the website of the ERI-Lectura (University of Valencia, Spain). To confirm that potential participants were eligible, they completed the PROLEC-SE-R, a battery of tests for assessing reading skills in Spanish (Cuetos, Arribas, & Ramos, 2016). Participants had no previous history of reading difficulties and showed normal current reading performance (see Table 1 for more details). In addition, other exclusion criteria to participate were having non-corrected visual problems, other sensory impairments, cognitive or neurological disorders, and not having a driving license.

Three participants could not finish the experimental session: two of them due to some discomfort when using the driving simulator and the third one because of technical problems. In addition, we excluded one more participant who failed to properly follow the task instructions (in particular, he did not answer in any of the critical VMS, neither in the visual nor in the audio + visual condition). Finally, we also excluded the matched participant of the previous one, in order to keep the order of counterbalance. Therefore, the final sample consisted of 24 participants (12 women). See the descriptive data of the sample in Table 1. We performed power analyses with G*Power v3.1.9.6 (Erdfelder, Faul, & Buchner, 1996), considering a repeated measures ANOVA with $\alpha = 0.05$, $1 - \beta = 0.80$, and a correlation among repeated measures between 0.70

Table 1
Descriptive statistics (Min = minimum value, Max = maximum value, M = mean, SD = standard deviation, SEM = standard error of the mean) of socio-demographic and reading performance data.

	Min	Max	M	SD	SEM
Word reading score (PROLEC_SE_R_LP)	158	257	197.3	28.7	5.9
Pseudoword reading score (PROLEC_SE_R_PS)	77	188	110.3	27.8	5.7
Age	18	45	24.3	7.2	1.5
Years of formal education	12	18	14.1	1.6	0.3
Years of driving experience	0.1	20.7	4.2	4.6	0.9

and 0.90. According to these analyses, a sample of 24 participants would be of sufficient power to reliably find within-participants main effects between $\eta^2 = 0.050$ and $\eta^2 = 0.017$).

The study was conducted according to the international ethical standards and was part of a larger research project positively evaluated by the Ethics Committee of Research in Humans at the University of Valencia.

2.2. Apparatus

We used a Carnetsoft driving simulator (https://www.rijschool-simulator.nl/). The simulator was housed in a light- and sound-attenuated room of the Faculty of Psychology and Speech Therapy of the University of Valencia. It included a comfortable seat, three 24-inch monitors (Benq GL2450), an operator interface monitor, and a set of driving controls consisting of a steering wheel, a steering gear box, and three pedals (Logitech G2). The central monitor was located at 38 cm from the participant's eyes, so that the visual angle of the screen was 70° (on the horizontal axis) by 43° (on the vertical axis). The height of the seat was adjusted so that the line of sight of the participant fell on the focal point on the apparent horizon line in the driving environment displayed on this monitor. The other two simulation monitors were placed right next to the central one, one at each side. The simulator gave auditory feedback about the own vehicle engine and tires. The simulation software run on a personal computer (Intel Core i7-3770 3,4GHz with a Nvidia GTX770 2 GB graphic card). The participant's actions on the vehicle controls, as well as speed and horizontal position of the participant's vehicle on the road lane, were processed at a sample rate of 120 Hz, although continuous variables were down-sampled by the simulator and recorded at 20 Hz.

An Apple iPad 2 (9.7″ screen) connected to a speaker was used to present multimedia stimuli – a TV series – to the participants. The screen of the iPad was located 23 cm towards the left visual field from the central point of view, immediately below the monitors of the simulator. The distance from the participant's eyes to the iPad was 43 cm, so that the visual angle of the screen was 26° (on the horizontal axis) by 20° (on the vertical axis). The volume of the speaker was high enough to be listened to clearly above the sound of the simulated vehicle engine.

Finally, to analyze each participant's glance behavior, we used SMI Eye-Tracking Glasses (v1.1). This device registers eye data at a sampling rate of 30 Hz. Calibration and verification of the eye-tracker was performed using a 3-dot pattern.

2.3. Driving environment

Participants traveled along the right lane of a route in a simulated motorway environment (see Fig. 1). We used a 14.40-kilometer-long route that mostly consisted of straight road segments and some slightly curved sections with two lanes on each direction and a speed limit of 120 km/h. The route also included two U-turns, with two one-direction lanes each, in which the speed limit was 80 km/h. There were no other cars driving along the road, with the exception of a preceding car (50 m in front of the participant's) and a following car (50 m behind the participant's), which kept constant the distance and did not interact with the participant.

Fig. 1. Frontal view of the simulated scenery.

The VMS were embedded in the driving simulation, posted above the road, presented along the 120 km/h-sections of the route. The VMS were 3-D models of one of the most commonly used VMS on the Spanish roads, allowing to display messages consisting of up to three lines of text, 12 or fewer characters each line. The design features followed the specifications of the standard norm UNE-EN 12966 (Asociación Española de Normalización y Certificación, 2015). To represent the characters of text messages, we used the patterns available in CarDim (I.P.S. Vial, S.L.), a computer application aimed at creating sign models for traffic engineering projects.

2.4. Tasks and stimuli

Participants were asked to complete two tasks: a TV series task and a VMS task, as described below.

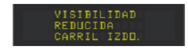
TV series task. The participants watched the first episode of a TV series while traveling along the route. The TV series was titled "*La otra mirada,*" a Spanish production filmed by BoomerangTV for the Spanish national and public channel *Televisión Española (TVE)*. None of the participants had previously watched this series. The first episode lasted 71 min and 28 s. The TV series can be freely accessed here: https://www.rtve.es/alacarta/videos/la-otra-mirada/otra-mirada-tabaco-pantalones-jazz/4581988/. Participants were asked to pay attention to the TV series while they were completing the route with the autonomous mode activated. For each condition, the TV series lasted 35 min. It began before encountering the third practice trial (five minutes after beginning to drive) in each condition. The experimenters ran the TV series when the participants arrived in a specific place of the route. In addition, to encourage attention to the TV series, they were informed that they would have to answer some questions at the end of the session: a three multiple-answer test with 10 questions involving the whole episode (see Supplementary Material 1 and 2).

VMS task. The participants were asked to recover the manual control of the car if the message presented on the VMS was a 'critical message' (i.e., a message informing on circumstances that would require the driver to take control of the car). In contrast, no response was required if the message presented was an 'informative message' (i.e., a message informing on circumstances that would not require any action from the driver). Four 'critical messages' and four 'informative messages' were used (see Table 2) based on those from Tejero et al. (2020). Particularly, the critical messages were considered as those that would require a lane change because of a sudden change in the road circumstances. On the contrary, the informative messages were considered as those that would not require a lane change.

So as to ensure the engagement of the participants with this task, care was taken not to use messages that allowed completing the task by only reading a few words at the beginning or the end of the message. In particular, as it can be seen in Table 2, at least one message in each category (informative vs. critical) began with exactly the same words in the first line (e.g., "Right lane…", "Left lane…", "Compulsory…", or "Visibility…") and ended with exactly the same words in the last line of the VMS (e.g., "…directions", "…roadworks", "…right lane", or "…left lane") of the VMS.

The participants were instructed to give the response at a distance to the VMS as far as possible, without making errors, and maintaining attention to the TV series. The responses were given by pressing one of the two levers located behind the steering wheel, the most preferred one by each participant. After some seconds, they were fostered to recover the autonomous mode by pressing one of the two central buttons of the wheel. The indication was given by the flickering of a green pictogram in

Table 2
Text VMS used in this study with their corresponding translations.

Informative messages				
Informative messages translation	Left lane occupied by roadworks	Right lane for all directions	Compulsory use of right lane	Reduced visibility on left lane
Critical messages				
Critical messages translation	Right lane occupied by roadworks	Left lane for all directions	Compulsory use of left lane	Reduced visibility on right lane

Note: The words '*izquierdo*' (the Spanish word for 'left') and '*derecho*' ('right') were represented by the abbreviations '*izdo.*' and '*dcho.*', as it is common in VMS on Spanish roads.

the dashboard and an auditory message. When the autonomous driving mode had been activated, the pictogram stopped flickering and remained steadily illuminated until the autonomous mode was deactivated again.

In one of the conditions of the experiment (see section 2.5), we also used audio files to additionally present the text displayed on the VMS as an oral message (audio + visual condition). Such audio files were prepared using a text-to-speech application available at https://fromtexttospeech.com/ (selecting the 'Spanish' language, 'Isabella' voice, and 'medium' speed), and were played through the audio system of the driving simulator. The abbreviations "*izdo.*" (*izquierdo*, left) and "*dcho.*" (*derecho,* right) were read in their complete form (see Table 2). The audio started when the participant's vehicle was just at a distance of 250 m to the VMS. According to pilot testing in the simulator, at such distance the text message on the VMS was not legible yet. As the audio message duration was about 2 s, the text message on the VMS was still not fully legible for the participant at the end of the audio message, which occurred when the participant's vehicle was at a minimum distance of 183 m to the VMS. We discarded a concurrent presentation of the audio and visual messages to prevent possible interference effects that may lead the participants to focus on only one of them, depending on their preferences (Dalton et al., 2013; Liu, 2001).

The experimental trials were defined as the periods beginning when the participant's vehicle was at a distance of 350 m to the next VMS (a distance at which the message on the VMS was not legible yet) and ending when the driver gave a response regarding the message displayed on the VMS, or, if no response occurred, when the driver reached the test VMS location. All other periods of the drive were considered as non-trial periods.

2.5. Procedure

Firstly, the participants read and signed an informed consent form. Then, their visual acuity was tested by using an E chart, to check that they were able to obtain a maximum LogMAR value of 0.1 with their usual visual correction. Next, they completed the PROLEC-SE-R (Cuetos, Arribas, & Ramos, 2016), and they filled in a questionnaire asking for some sociodemographic and educational data as well as for information on driving habits.

Then, they completed a training on the VMS task, with no simultaneous driving, on which the messages were presented on the central monitor, one at a time, at least six times each, in different orders: 1) Text messages, sequential presentation, 2) Audio messages, sequential presentation, 3) Text messages, first random presentation, 4) Audio messages, first random presentation, 5) Audio messages, second random presentation, and 6) Text messages, second random presentation.

Next, participants got comfortable in the driver's seat and put on the SMI eye-tracking glasses. Each participant could keep his or her own optical correction. After that, they were asked to complete the simulated route, performing the VMS and the TV series tasks. Each participant completed the simulated route twice, separated by a 10-minute break to rest, using a counterbalanced order across participants. At one of the times, the messages were only available as text messages on the VMS (visual condition). At the other time, oral message versions of the text VMS were additionally presented (audio + visual condition). In each condition, 10 initial training trials with blank VMS were incorporated for participants to get used to the simulator, which were followed by 8 practice trials to allow the participants to practice the VMS task while driving. Two of the eight practice trials were completed with the manual driving mode and with no TV series, and the other six were presented with the autonomous mode activated and the TV series on. After these practice trials, 40 experimental VMS (eight messages × five repetitions) were randomly presented for each participant in each condition. Therefore, participants completed 116 trials in total. On average, participants took about 80 min to complete the two drives. Driving was not stopped except to perform a recalibration of the eye-tracker, if necessary (which was accomplished at the U-turns) or during the 10-minute break between the two experimental blocks. After that, the participants answered some questions about the TV series. In appreciation for participating in the study, they were offered a light meal (a soft beverage and a snack) at the end of the session.

2.6. Measures, design, and statistical analyses

Data from trial sections were analyzed. The differences between message conditions were tested as a within-subject design (visual vs. audio + visual).

For each participant, we obtained the following measures. Measure 1 (M1): the Signal Detection Theory (SDT) measure of sensitivity, d', to discriminate critical and non-critical messages in the VMS task. M2: the SDT measure of response bias (β). M3: the mean response distance, defined as the mean distance (in meters) from the participant's car to the VMS, when the response in the VMS task was given. This measure was computed excluding the trials on which the participant gave a wrong response in the VMS task.

Regarding the measures obtained from the eye-tracker, we analyzed in each trial the subsequent variables. M4: the mean latency of the first glance at the road (ms). M5: the number of glances at the road. M6: the mean duration of the glances at the road. M7: the sum of the time glancing at the road. All these indicators were averaged per participant and condition, and considering only the trials with a correct response. Data from two participants were excluded from the eye-tracking analyses due to problems in the register (lost connection between the eye-tracker glasses and the simulator).

Furthermore, we analyzed the answers that the participants gave to the final test about the TV series episode. M8: the sum of the questions' scores.

All measures were analyzed by separate analyses of variance (ANOVA), performed using the IBM SPSS Statistics 26 software (IBM Corp., 2019). In addition, as further explained in Section 3.2, a complementary ANOVA analysis was performed for all measures, including message condition as a within-participant factor (visual vs. audio + visual) and counterbalance order as a between-participant factor (visual first vs. audio + visual first).

3. Results

3.1. General results

Table 3 shows some descriptive statistics of the measures obtained.

3.1.1. VMS task

M1: Sensitivity (*d′*). The effect of message condition was significant (F(1, 23) = 14.607, p =.001, η_p^2 =.388). The ability to discriminate critical and non-critical VMS messages was higher in the audio + visual condition than in the visual one (see Fig. 2).

M2: Response bias (*β*). The effect of message condition was significant (F(1, 23) = 8.193, p =.009, η_p^2 =.263). The participants were more conservative in the visual condition than in the audio + visual one (see Fig. 3).

M3: Mean response distance. This measure was computed for only 20 participants because 4 participants failed to respond to any of the messages in the visual condition. The effect of message condition was significant (F(1, 18) = 542.85, p <.001, η_p^2 =.968). The participants responded correctly more quickly in the audio + visual condition than in the visual one (see Fig. 4).

3.1.2. Eye movement measures

The eye-tracker measures could only be obtained for 18 participants because 1 participant had no valid values for the audio + visual condition, 3 participants had no valid values for the visual condition, and for 2 participants it was not possible to record the eye-tracker measures.

M4: Mean latency of the first glance at the road. The effect of message condition was significant (F(1, 16) = 15.173, p =.001, η_p^2 =.487). The average latency of the first glance at the road was shorter in the visual condition than in the audio + visual one.

M5: Number of glances at the road. The effect of message condition was significant (F(1, 16) = 73.125, p <.001, η_p^2 =.820). There were more glances at the road in the visual than in the audio + visual condition.

M6: Mean duration of glances at the road. The effect of message condition was significant (F(1, 16) = 6.181, p =.024, η_p^2 =.279). The mean duration of fixations was longer in the visual condition than in the audio + visual one.

M7: Sum of the time glancing at the road. The effect of message condition was significant (F(1, 16) = 75.597, p <.001, η_p^2 =.825). The sum of the time staring at the road was longer in the visual than in the audio + visual condition.

Table 3
Percentage of correct responses in the VMS task and mean values across participants of the other measures obtained during the experimental trials (standard deviation between parentheses, 95% CI between brackets), by message condition (Visual / Audio + visual).

		Visual	Audio + visual
VMS task	Correct responses (%)	85.88 (3.48) [83.48, 88.05]	99.89 (0.32) [99.42, 99.99]
	False alarms (%)	0.13 (0.33) [−0.01, 0.26]	0.00 (0.00) [0.00, 0.00]
	d'	2.71 (1.52) [2.07, 3.35]	3.91 (0.07) [3.88, 3.94]
	β	1.76 (1.23) [1.24, 2.28]	1.04 (0.17) [0.96, 1.11]
	Mean response distance (m)	66.23 (27.24) [53.83, 78.63]	142.22 (18.34) [133.87, 150.57]
Eye movement pattern	Mean latency of the first glance at the road (ms)	1657.36 (956.02) [1181.94, 2132.77]	3141.15 (1501.21) [2394.62, 3887.69]
	Number of glances at the road	12 (2.69) [10.56, 13.24]	6 (3.55) [3.89, 7.41]
	Mean duration of glances at the road (ms)	322.85 (77.08) [284.52, 361.18]	272.53 (95.82) [224.88, 320.18]
	Sum of the time glancing at the road (ms)	3632.27 (1057.11) [3106.58, 4157.96]	1531.95 (1222.79) [923.87, 2140.03]

3.1.3. TV series task

M8: Global scores. The mean score was 9.5 out of 10 (SD = 0.8) points at the final test about the TV series episode. All participants answered correctly a minimum of eight questions. The mean response for the visual condition questions was 4.71 and the mean response for the audio + visual condition was 4.75 (F(1, 22) = 0.116, p =.737, η_p^2 =.005).

3.2. Results by counterbalanced order

The results described in the previous section were obtained by averaging the conditions across the two possible counterbalanced orders (visual condition first and audio + visual condition first). However, we also performed a complementary analysis to check if the counterbalanced order was a relevant factor moderating the reported results, and we observed significant results. As this evidence suggests a potential learning effect that is not cancelled by counterbalancing the order of the conditions, we decided to report this complementary analysis. Table 4 shows the descriptive statistics of the measures that showed a different pattern, as compared to the general results.

3.2.1. VMS task

Sensitivity (*d′*). The interaction between message condition and the order of counterbalance was significant (F(1,22) = 16.219, p =.001, η_p^2 =.424), as well as the two main effects of message condition (F(1,22) = 24.272, p <.001, η_p^2 =.525) and order of counterbalance (F(1,22) = 15.246, p =.001, η_p^2 =.409). Taking into account the interaction (see Table 4), when the participants began the experiment with the visual condition, their ability to discriminate the VMS messages was clearly lower (F(1,22) = 40.086, p <.001, η_p^2 =.646) in the visual condition (1.747) as compared to the audio + visual condition (3.919). In contrast, when participants began the experiment with the audio + visual condition, a similar sensitivity was observed (F(1,22) = 0.404, p =.531, η_p^2 =.018) in both the visual (3.673) and the audio + visual (3.892) conditions (see Fig. 5).

Mean response distance. The interaction between message condition and the order of counterbalance was significant (F(1,18) = 12.318, p =.003, η_p^2 =.406), as well as the main effect of message condition (F(1,18) = 542.845, p <.001, η_p^2 =.968). Taking into account the interaction (see Table 4), when the participants began the experiment with the visual condition, they gave their correct responses in the VMS task at a clearly shorter distance to the VMS (F(1,18) = 299.461, p <.001, η_p^2 =.943) in the visual condition (47.414) as compared to the audio + visual condition (138.425). The same was true when the participants began the experiment with the audio + visual condition: they gave their correct responses at a shorter distance to the VMS (F(1,18) = 244.762, p <.001, η_p^2 =.931) in the visual condition (77.227) as compared to the audio + visual condition (144.409). However, considering the reported significant interaction, the difference between both message conditions was smaller in the latter (67.182) as compared to the former (91.011) order of counterbalance (see Fig. 6).

No other message condition × order of counterbalance significant interaction was found on the rest of the measures registered in the VMS task (all p >.229).

3.2.2. Eye movement pattern

No significant condition × order of counterbalance interaction was found on the eye-tracking measures reported (all p >.161).

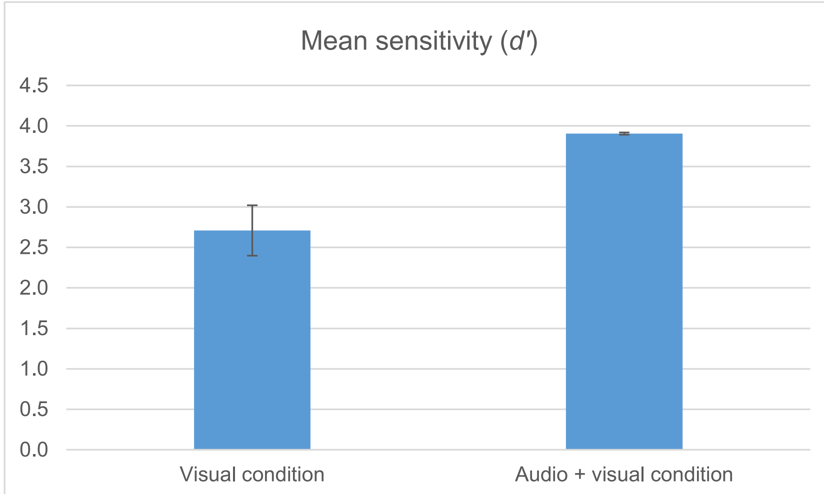


Fig. 2. Mean sensitivity for the experimental conditions (visual vs. audio + visual).

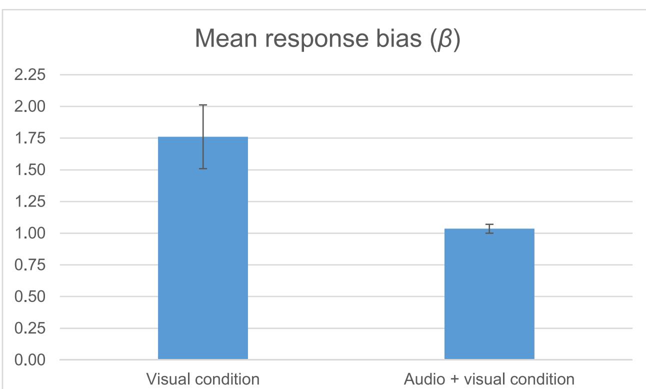


Fig. 3. Mean response bias for the experimental conditions (visual vs. audio + visual).

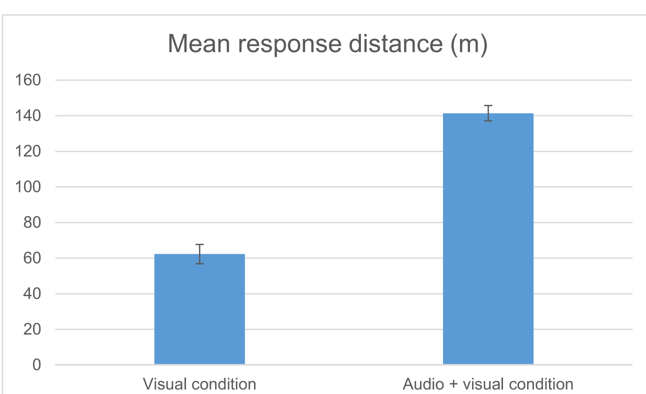


Fig. 4. Mean distance of responses for the experimental conditions (visual vs. audio + visual).

4. Discussion

The main aim of this study was to test if an audio + visual presentation of the text displayed on a VMS would help drivers discriminate critical and non-critical traffic messages during an autonomous driving situation, while the driver is engaged in an audiovisual distractive activity (i.e., watching a TV series on a tablet). Therefore, our results could be useful to discuss whether the previously reported advantage of the audio + visual messages would still be observed even when participants are engaged in an audiovisual activity. In addition, we would explore the potential use of oral messages with traffic information to help drivers operating a highly-automated vehicle to identify critical situations where they would be required to take over control.

First, according to our behavioral results, participants performed better the VMS task in the audio + visual condition as compared to the visual condition, considering the sensitivity (d'), as well as the correct response distance. Therefore, the drivers were able to discriminate more correctly the critical and non-critical VMS messages at longer (i.e., safer) distances in the audio + visual condition as compared with the visual one. These results are consistent with Tejero et al. (2020) and suggest that the processing of traffic messages may improve when they were also provided as oral messages. Importantly, the advantage of the audio + visual condition was also reported in the current experiment where drivers were engaged in an audiovisual distracting activity, while in Tejero et al.'s study the tasks required to the participants involved only a competition for visual resources. Therefore, our results imply that the benefit of audio + visual messages would not be cancelled when the tasks required to the participants (i.e., the TV series task and the VMS task) involve both auditory and visual resources.

A potential explanation for the reported advantage of the audio + visual messages could be that auditory messages would work as warning signals making drivers shifting attention from the ongoing TV series to the traffic message so as to decide whether or not they have to recover the manual control. In contrast, when drivers only rely on visual messages, they have no particular signal to warn them of an incoming VMS. The effects of warning signals are well known in the attention literature. For example, according to Petersen and Posner (2012), warning signals prior to a target event would increase phasic alertness, replacing the participant's prior attentional state with a new state that involves preparation for detecting and responding to the target event. Therefore, in the current experiment, the auditory messages would have produced, first of all, an alerting effect on the participants, inducing an attentional shift from the TV series, and helping them process the traffic message and respond to the VMS task. In fact, Geitner et al. (2019) found that the auditory-tactile warnings fostered faster reactions than tactile warnings during automated driving, which would be in line with our finding of the advantage of audio + visual warnings as compared to single modality (visual) warnings alone. In addition, Han and Ju (2021) found that their participants recognized faster the auditory than the visual alarms of their experiment and Beattie et al. (2014) found that giving a spatialized auditory alarm warned drivers much more than when no sound was available.

Table 4
Mean values across participants of the sensitivity (d′) and response distance measurements obtained during the experimental trials in the VMS task (standard deviation between parentheses, 95% CI between brackets), by order of counterbalance (Visual first / Audio + visual first) and message condition (Visual / Audio + visual).

		Visual first		Audio + visual first	
		Visual	Audio + visual	Visual	Audio + visual
VMS task	d′	1.75 (1.66) [1.04, 2.46]	3.92 (0.00) [3.88, 3.96]	3.67 (0.29) [2.96, 4.38]	3.89 (0.10) [3.85, 3.93]
	Mean response distance (m)	47.41 (27.48) [29.72, 65.11]	138.43 (21.74) [124.27, 152.58]	77.23 (21.17) [62.78, 91.68]	144.41 (17.14) [132.85, 155.97]

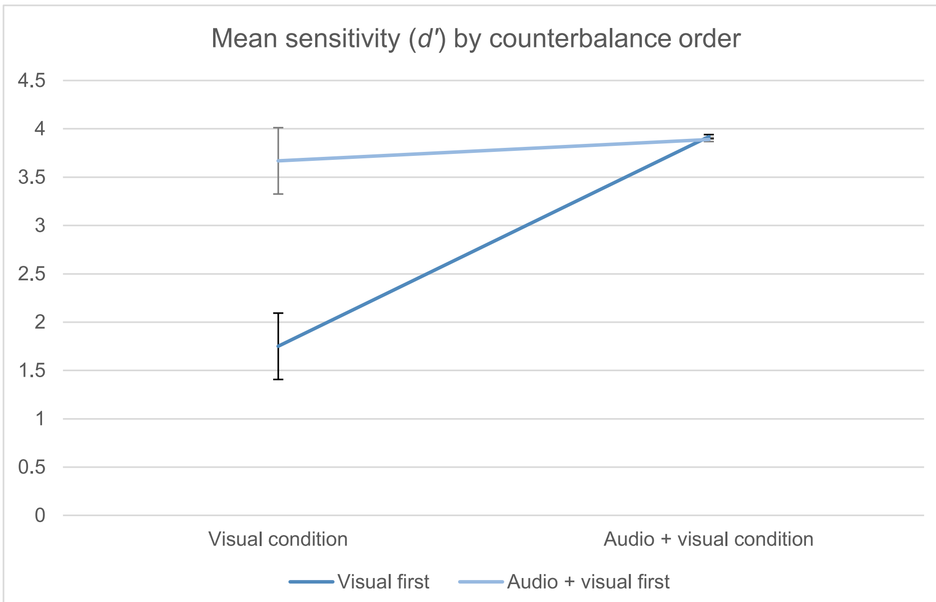


Fig. 5. Interaction between the experimental conditions (visual vs. audio + visual) and the order of counterbalance for the mean sensitivity.

Regarding the response bias (β), our participants were more conservative in the visual condition, that is to say, they were more prone to omit a response to the stimuli in this condition. This result would be in line with our previous interpretation, as it would indicate that giving an audio + visual message would foster a higher tendency to respond to the stimuli.

Interestingly, the analysis of eye tracking measures while the participants were approaching the VMS showed that the availability of an oral message in the audio + visual condition, in comparison with the visual condition, resulted in a more efficient pattern of fixations. In particular, in the audio + visual condition, participants were able to better discriminate critical and non-critical messages despite: (a) delaying their first glance at the road, (b) dedicating fewer and shorter glances to the road, and (c) investing less total glancing time to inspect the road. Taken together, these results suggest that the drivers had to check the road less often (thus less visual demands were required) when having audio + visual VMS messages and, consistently with the previously proposed warning effect of these messages, they could focus their attention on the TV series and only shift their visual attention to the road when really necessary. To this respect, it should be noted that the oral versions of the messages in the audio + visual condition began when the messages displayed on the VMS were not legible yet. Consequently, the eye-tracking results, as well as the longer response distances previously reported in the audio + visual condition, would imply that participants relied more on the oral version of the messages to give their responses, instead of waiting for the text to be fully legible. Finally, no differences were reported in the questions on the TV series. Therefore, the differences in the cognitive demands required by the visual or the audio + visual conditions had no noticeable effect on the TV series task.

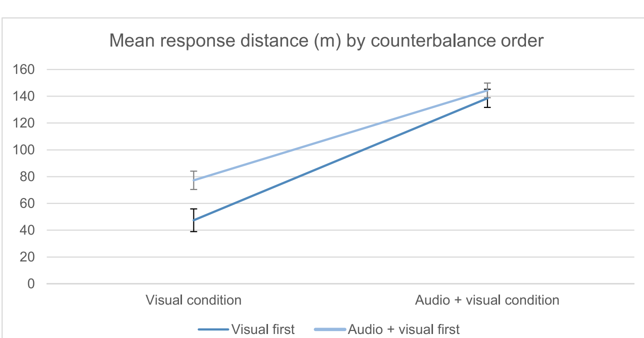


Fig. 6. Interaction between the experimental conditions (visual vs. audio + visual) and the order of counterbalance for the mean distance of responses.

Broadbent et al. (2023) studied gaze behavior in manual driving and compared a single-task condition (only driving) to a dual-task condition (driving and performing a modified 2-back task). They found that, when distracted by a cognitive demanding secondary task, participants reduced the number of fixations but it increased their mean duration. In addition, they spent a greater percentage of time fixating on the road compared to the mirrors, especially the right wing mirror. In contrast, in our experiment, adding an audio message (audio + visual condition), resulted in both fewer and faster fixations, as well as in spending less time glancing at the road. Therefore, the pattern of eye movements in the current experiment suggests that the audio + visual condition would not be working as an extra cognitive demanding secondary task but, instead, it would be actually promoting a more efficient pattern of fixations.

Xu et al., (2023) compared gaze behavior in different levels of driving automation (SAE 2 level and SAE 3 level) and proposed an integrated gaze control model during take-over driving. According to these authors, the control of gaze is the result of the interaction between bottom-up and top-down flows of information. Applying this model to our findings, we can infer that our participants would rely more on the bottom-up influence of the audio message as they showed fewer and later glances in the audio + visual condition as compared to the visual condition. In fact, in the latter condition, they checked the road more frequently and for longer periods, which could reflect their internal state of top-down expectation about the incoming VMS.

Finally, as reported in section 3.2., we also performed a complementary analysis suggesting a potential learning effect that is not cancelled by counterbalancing the order of the conditions. According to this complementary analysis, when the participants began the experiment with the visual condition, their ability to discriminate the VMS messages was clearly lower and responded at a shorter distance as compared to the (posterior) audio + visual condition. In contrast, when they began the experiment with the audio + visual condition, their sensitivity was similarly high in the audio + visual and in the only visual conditions. This finding suggests that the audio + visual condition resulted in a positive effect on the VMS task performance (higher d'), which was not cancelled when participants completed the next block of trials without the audio messages. Therefore, this unexpected result suggests that the audio messages could have induced a learning effect beyond their exposure, allowing the participants to better discriminate between the critical and non-critical messages when they had to rely only on the visual messages later. To this respect, it should be noted that message presentation was randomized for each participant and block of trials (therefore, it was not possible to learn the location of the critical and non-critical messages between blocks) and that the correct responses in the visual condition after the audio + visual condition were produced also at a shorter distance (therefore, participants still had to wait for the text in the VMS to be readable).

4.1. Limitations

The current study has some relevant limitations that must be considered in the interpretation of the results. First, we report a single experiment with 24 participants. Therefore, further research will be useful to confirm the effects reported using different experimental conditions (e.g., other SAE levels of automation), diverse groups of drivers (e.g., with and without difficulties to read the text messages, such as drivers with dyslexia or any driver in adverse conditions) and to verify that no other significant effects would also be found with larger samples. Second, our data were gathered using a driving simulator, and thus, real-driving studies would be required to expand our results to real-life settings. Finally, as previously discussed, our results suggest a potential warning effect of the oral messages, but no physiological measures of alertness were taken in the current study.

5. Conclusion

According to the results reported in the current study, the processing of traffic messages may improve when they are provided as both oral and visual messages. In particular, drivers were able to better discriminate critical and non-critical messages when an audio + visual message were used, even that they were performing a distractive secondary task requiring audiovisual cognitive resources (i.e., watching a TV series episode). We think that these results would be of special interest for engineers designing and developing highly automated cars, considering that the design of automated systems must ensure that the driver's attention is sufficient to take over control and manage the driving situation.

Funding

This article is part of the R+D+i project PID2019-106562GB-I00, funded by MCIN/AEI/10.13039/501100011033, and the project AICO/2021/290, funded by the Conselleria de Educación, Universidades y Empleo (Generalitat Valenciana). In addition, Marina Pi-Ruano received a PhD grant (ACIF/2019/160), funded by the Conselleria d'Innovació, Universitats, Ciència i Societat Digital & European Social Fund.

CRediT authorship contribution statement

Marina Pi-Ruano: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Software, Visualization, Writing – original draft. **Javier Roca:** Conceptualization, Data curation, Funding acquisition, Methodology, Project administration, Resources, Software, Supervision, Writing – review & editing. **Pilar Tejero:** Conceptualization, Data curation, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jsr.2024.01.014.

References

Asociación Española de Normalización y Certificación, (2015). Señalización vertical en carretera. Paneles de mensaje variable. Norma EN 12966:2014. AENOR, Madrid.
Beattie, D., Baillie, L., Halvey, M., & McCall, R. (2014). What's around the corner? Enhancing driver awareness in autonomous vehicles via in-vehicle spatial auditory displays. In *In: NordiCHI '14 Proceedings of the 8th Nordic Conference on Human-Computer Interaction*. https://doi.org/10.1145/2639189.2641206
Broadbent, D. P., D'Innocenzo, G., Ellmers, T. J., Parsler, J., Szameitat, A. J., & Bishop, D. T. (2023). Cognitive load, working memory capacity and driving performance: A preliminary fNIRS and eye tracking study. *Transportation Research: Part F, 92*, 121–132. https://doi.org/10.1016/j.trf.2022.11.013
Chen, J., Šabić, E., Mishler, S., Parker, C., & Yamaguchi, M. (2022). Effectiveness of lateral auditory collision warnings: Should warnings be toward danger or toward safety? *Human factors, 64*(2), 418–435. https://doi.org/10.1177/0018720820941618
Cuetos, F., Arribas, D., & Ramos, J. L. (2016). *Batería de evaluación de los procesos lectores en secundaria y bachillerato, revisada (PROLEC-SE-R)*. Madrid: TEA.
Dalton, P., Agarwal, P., Fraenkel, N., Baichoo, J., & Masry, A. (2013). Driving with navigational instructions: Investigating user behaviour and performance. *Accident Analysis & Prevention, 50*, 298–303.
Donmez, B., Boyle, L. N., Lee, J. D., & McGehee, D. V. (2006). Drivers' attitudes toward imperfect distraction mitigation strategies. *Transportation Research Part F: Traffic Psychology and Behaviour, 9*(6), 387–398. https://doi.org/10.1016/J.TRF.2006.02.001
Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments & Computers, 28*(1), 1–11. https://doi.org/10.3758/BF03203630
Eriksson, A., & Stanton, N. A. (2017). Takeover time in highly automated vehicles: Noncritical transitions to and from manual control. *Human Factors, 59*(4), 689–705. https://doi.org/10.1177/0018720816685832
Geitner, C., Biondi, F., Skrypchuk, L., Jennings, P., & Birrell, S. (2019). The comparison of auditory, tactile, and multimodal warnings for the effective communication of unexpected events during an automated driving scenario. *Transportation Research Part F: Traffic Psychology and Behaviour, 65*, 23–33. https://doi.org/10.1016/j.trf.2019.06.011
Gold, C., Damböck, D., Lorenz, L., & Bengler, K. (2013). "Take over!" How long does it take to get the driver back into the loop? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 57*(1), 1938–1942. https://doi.org/10.1177/1541931213571433

Han, J.-H., & Ju, D.-Y. (2021). Advanced alarm method based on driver's state in autonomous vehicles. *Electronics, 10*(22), 2796. https://doi.org/10.3390/electronics10222796

IBM Corp.. (2019). IBM SPSS Statistics for Windows (Version 26.0). [Computer software]. *IBM Corp*. https://www.ibm.com/spss.

Lee, J. D., McGehee, D. V., Brown, T. L., & Reyes, M. L. (2002). Collision warning timing, driver distraction, and driver response to imminent rear-end collisions in a high-fidelity driving simulator. *Human Factors, 44*(2), 314–334. https://doi.org/10.1518/0018720024497844

Liu, Y. C. (2001). Comparative study of the effects of auditory, visual and multimodality displays on drivers' performance in advanced traveler information Systems. *Ergonomics, 44*(4), 425–442.

Papantoniou, P., Papadimitriou, E., & Yannis, G. (2017). Review of driving performance parameters critical for distracted driving research. *Transportation Research Procedia, 25*, 1796–1805. https://doi.org/10.1016/j.trpro.2017.05.148

Petersen, S. E., & Posner, M. I. (2012). The attention system of the human brain: 20 years after. *Annual Review of Neuroscience, 35*(1), 73–89. https://doi.org/10.1146/annurev-neuro-062111-150525

Regan, M. A., Lee, J. D., & Young, K. (2008). *Driver distraction: Theory, effects, and mitigation*. CRC Press.

Šabić, E., Chen, J., & MacDonald, J. A. (2021). Toward a better understanding of in-vehicle auditory warnings and background noise. *Human Factors, 63*(2), 312–335. https://doi.org/10.1177/0018720819879311

Society of Automotive Engineers, (2021). SAE levels of driving automation refined for clarity and international audience. Retrieved from: https://www.sae.org/blog/sae-j3016-update.

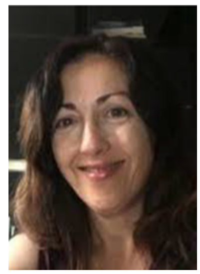
Stanton, N. A., & Young, M. S. (2005). Driver behaviour with adaptive cruise control. *Ergonomics, 48*(10), 1294–1313. https://doi.org/10.1080/00140130500252990

Tejero, P., Pi-Ruano, M., & Roca, J. (2020). Better read it to me: Benefits of audio versions of variable message signs in drivers with dyslexia. *Annals of dyslexia, 70*(3), 295–312. https://doi.org/10.1007/s11881-020-00199-6

Xu, C., Louw, T. L., Merat, N., Li, P., Hu, M., & Li, Y. (2023). Drivers' gaze patterns when resuming control with a head-up-display: Effects of automation level and time budget. *Accident; analysis and prevention, 180*, Article 106905. https://doi.org/10.1016/j.aap.2022.106905

Marina Pi-Ruano is a PhD student of the Doctoral Programme in Reading and Comprehension at the University of Valencia (Spain). She obtained a Degree in Psychology (University of Valencia) in 2016, and a Master's Degree in Dependence: Research, Treatment and Drug Pathologies (University of Valencia) in 2018. Her thesis focuses on the study of integration of multimodal information during vehicle driving.

Javier Roca is associate professor at the University of Valencia (Spain). He is also researcher at the ERI-Lectura (Interdisciplinary Research Structure on Reading) and Secretary of the Spanish Society for Experimental Psychology (SEPEX). He obtained a Degree in Psychology (University of Valencia), and a PhD in Experimental Psychology (University of Granada). His research focuses on the study of cognitive processes and its application in daily living activities, such as reading or driving a vehicle. Since 2012, he has been leading a funded research line on the difficulties of adults with and without dyslexia to acquire information from traffic signs.



Pilar Tejero Gimeno (University of Valencia, Spain) is an Associate Professor of Psychology, and also a researcher in the ERI-Lectura (university Interdisciplinary Research Structure on Reading). She obtained her PhD in Psychology from the University of Valencia in 1995. Her research has mainly focused on the processing of visual and auditory information while driving a vehicle, both in optimal conditions and under the influence of factors that affect driver performance, such as sleep deprivation, prolonged driving, or dyslexia.