A Comparative Usability and Visual Attention Study of Android Auto and CarPlay in a Realistic Driving Setting

Chiara Melega
Department of Psychology and Cognitive Sciences
University of Trento
Rovereto, Italy
chiara.melega@studenti.unitn.it

Icaro Re Depaolini
Department of Psychology and Cognitive Sciences
University of Trento
Rovereto, Italy
icaro.redepaolini@studenti.unitn.it

Pietro Cau
Department of Psychology and Cognitive Sciences
University of Trento
Rovereto, Italy
pietro.cau@studenti.unitn.it

Tommaso Ceccherini
Department of Psychology and Cognitive Science
University of Trento
Rovereto, Italy
tommaso.ceccherini@studenti.unitn.it

Lucrezia Di Bari
Department of Psychology and Cognitive Science
University of Trento
Rovereto, Italy
lucrezia.dibari@studenti.unitn.it

1 INTRODUCTION

In-Vehicle Infotainment Systems (IVIS) allow drivers to perform a variety of tasks while driving, typically by interacting through touch or voice with a screen placed at the top of the central console of the vehicle. The tasks that can be performed with the IVIS range from playing music, to receiving and making phone calls, to handling navigation and road directions.

IVIS are born from the necessity for more connectivity inside vehicles and are seeing increased adoption, with many commercial car manufacturers beginning to offer them by default in most of their car models. Moreover with the mass adoption of smartphones worldwide, in the early 2000s, a new use-case has emerged: The use of these devices in the car [1]. To meet this user's necessity, car manufacturers started to develop new IVIS systems that allowed for connectivity with smartphones.

Another important aspect that led to the development of IVIS is the risk posed by using smartphones while driving. In the study by Oviedo-Trespalacios, King, Haque & Washington (2017) it was shown that 49% of drivers use a mobile phone for talking while at the wheel,. and 50% of them reported mobile phone use for browsing or texting [1, 8]. One of the aims of IVIS is that of preventing such behaviours by offering a convenient and safer alternative to smartphone usage that still allows drivers to enjoy a certain level of connectivity.

For the development of IVIS, car manufacturers often rely on tech companies, and today two of the most commonly used platforms are Google's "Android Auto" and Apple's "CarPlay". In a recent survey 79 to 84% of respondents are either interested or consider these platforms to be an essential feature in their vehicles [4]. These are effectively operating systems that seamlessly connect and integrate smartphone functionalities with the IVIS and offer an interface specifically designed for usage while driving.

Just as in any other in-vehicle system, meant to be used whilst operating a vehicle, considerations about usability and distraction are of vital importance. Several studies have been conducted to assess the impact of driver cell phone use on accidents. In a study conducted by Ige, Banstola & Pilkington (2016), the use of mobile phones while driving has been considered as one of the most serious forms of distraction because of the wide scale of demands on the driver's attention [2].

At the same time, 72% of drivers are aware of danger linked to mobile phone calls while driving, and 94% of them are aware of the danger linked to texting or browsing while driving [8]. If IVIS systems have the aim of tackling such problems, reducing the impact of phone usage on car accidents, considerations about their usability and distraction should be carefully taken into account. Few studies have been conducted in that direction, so it is not clear whether IVIS systems are able to fulfil one of their primary goals.

In order to investigate the impact of IVIS on driver distraction, we conducted a research aimed at measuring and comparing the usability of Android Auto and CarPlay by performing a series of usability tests, on different tasks, in a realistic setting. Secondly, we aimed at measuring and comparing the impact that these systems have on visual attention, by using gaze detection algorithms during tasks.

Therefore, we formulated the following hypotheses:

(1) Different IVIS (CarPlay vs Android Auto) will have significantly different impacts on visual attention.
(2) Different IVIS (CarPlay vs Android Auto) will have significantly different impacts performance in a simulated driving task.
(3) Different IVIS (CarPlay vs Android Auto) will have significantly different impacts on time on task.
(4) Different IVIS (CarPlay vs Android Auto) will have different usability scores.

2 METHODS

2.1 Realistic Setting

When studying interaction with IVIS it's important to faithfully recreate the setting in which they will be used, i.e. while driving a vehicle.

When using IVIS on the road we expect users to simultaneously perform and pay attention to two tasks: using the system and driving, which can be itself subdivided in a number of sub-tasks. Users will do this by "multitasking" and by switching between tasks at appropriate times [7]. The attentional and cognitive demands posed by this dual task "setting" have a strong impact on the modality of interaction with the system.

For example, users may not be able to continuously look at the interface, they will likely have to pay attention to the road and choose appropriate times to look at the screen with quick glances, and then interact with it using only one hand, maybe even while not looking.

Therefore placing participants in front of a screen and asking them to perform tasks is not an appropriate way of studying these systems, since it would not reflect the interaction pattern of a real use situation.

In order to create a realistic setting we believed that it was important, as a starting point, to carry out the tests in a realistic use environment. Therefore we conducted the study inside a real (stationary) car, with a standard 8-inch IVIS screen equipped from the factory, positioned above the centre console. This meant the participants were subject to realistic constraints in terms of screen size, position, distance and orientation.

As we discussed before, interacting with a touch screen while driving creates a particular pattern of interaction born from the necessity to perform two tasks at the same time: driving and interacting with the interface [7].

In order to elicit a similar pattern of interaction from the participants we choose to conduct the study following a dual task paradigm. In such an experimental paradigm participants are asked to perform two tasks simultaneously: in our case an IVIS task (using the IVIS interface) and a simulated driving task.

2.2 Driving Task

For "driving task" we intend a task that is meant to induce a similar attentional requirement as driving. In the choice of the task it was important for us that the task was something that required subjects to look at it, but that they could still briefly look away from, from time to time.

Furthermore, since we are interested in the impact on visual attention, we required a task from which we could extract an objective continuous measure of visual attention. As a result, we excluded tasks like Reaction Time tasks or GO/NO-GO tasks, often found in the existing literature on driver distraction, but only capture performance at discrete intervals of time.

The task we settled on is a pursuit-tracking task, specifically, we adapted the one used by [5]. The task consisted in a screen placed in front of the participant, above the steering wheel, in which a ball and a triangle are displayed, the ball moves from side to side, in a manner that appears to be random, but follows a combination of sine waves, as described in [5]. The triangle can also be moved side to side and is controlled by the participant through a wireless one-handed joystick (JoyCon), used with the left hand. The right hand, instead, will be used to perform the IVIS tasks. The participant is instructed to move the triangle following as close as possible the movement of the circle. This, as described in [5], is meant to simulate the visual pattern of someone driving on a winding road.

As output of the task, a score is calculated as the mean absolute error, using the average distance between circle and triangle during the whole trial.

Figure 1: Driving task setup

2.3 IVIS Task

While subjects performed the driving task, they were also asked to perform a task on the IVIS screen. For the purposes of studying visual attention exclusively, a very standardised set of tasks might have been more appropriate, but since we were also interested in usability, we didn't give specific directions on "how" the tasks should be performed. Users were given the prompt of the task without any other suggestions. For example, some participants used touch only, while others voice controls. While this introduces an added level of variability in the data, it also replicates a more realistic usage pattern.

The three tasks the participants were asked to perform on the IVIS were the following:

- *Task 1:* Use the IVIS system to get road directions for a specific place, avoiding tolls.
- *Task 2:* Use the IVIS system to call Bill Gates.
- *Task 3:* Use the IVIS system to play the first song in the "Liked Songs" playlist.

It has to be noted that for both platforms Google Maps, Spotify and the platform's default phone application were used to perform the tasks.

2.4 Gaze Detection

In order to have a direct measure of visual attention a gaze detection algorithm was used to analyse the video of participants performing the two tasks (driving and IVIS) simultaneously, this allowed us to know exactly how much participants "looked at the road" (the driving task) and how much they instead looked at the IVIS screen.

The video was recorded through a webcam placed just above the screen displaying the driving task, for the analysis we used a pre-trained machine learning model able to perform eye contact detection [6], which allowed us to estimate with good accuracy when the participant was looking at the driving task, and when he wasn't.

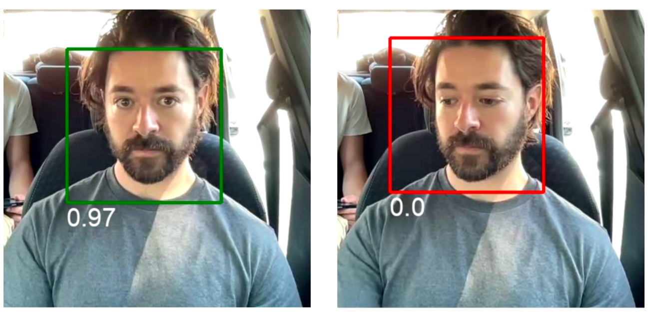

Figure 2: Eye Contact Detection - Looking at the driving task vs looking at the IVIS

3 EXPERIMENT DESIGN

The experiment was conducted following a within subject approach, meaning that all participants performed all tasks for each platform, in a randomised order. In total, 27 Participants were recruited following a snowball sampling approach.

Before beginning the session each participant was introduced to the experiment, asked to sign an informed consent form and to fill a questionnaire. The questionnaire contained demographic questions as well as questions regarding familiarity with the platforms used in the experiment (Android Auto and CarPlay), their respective operating systems (Android and iOS) and general driving experience (years of driver licence ownership).

The participants were then sat in the driver's seat of the stationary vehicle, while an experimenter, tasked with instructing them, sat beside them in the passenger seat. Two other experimenters sat in the back, one handling post-task questionnaires and keeping track of tasks, and one as observer and note-taker (Figure 3).

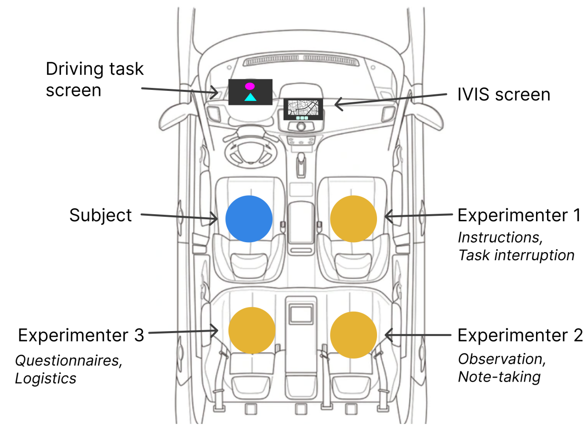

Figure 3: Vehicle Setup

In order to familiarise participants with the driving task, the experiment began with two trials of the driving task, which lasted 45 seconds each. This helped us reduce a possible carryover effect caused by subjects becoming more familiar with the task during the experiment.

The familiarisation trials were followed by a 45 second control trial, in which participants were again asked to perform the driving task only. This time the score (mean absolute error) was recorded in order to have a performance baseline on the driving task for each participant.

After the control trial the experimental trials began. In these trials the participants were asked to perform the driving task while also performing a task on the IVIS (Figure 4). In total we had each participant perform six such trials: one for each of the three IVIS tasks (Music, Calling and Navigation) on both platforms (AndroidAuto and CarPlay). The order of the platforms, and the order of the tasks within them, was randomised amongst participants in order to reduce the bias caused by a possible order effect between analogous tasks on different platforms.

Figure 4: Experimental Trial - The participant performs the driving task while completing a task on the IVIS

During these tasks, one of the experimenters had a second Joy-Con through which he was able to interrupt the trial once the task was completed. This also allowed us to precisely calculate time on task.

After each task, subjects were asked to answer a SEQ questionnaire [10] to rate how difficult they found performing that task. After completing the tasks for each platform they were also asked to fill a SUS questionnaire [3] to rate the usability of the interfaces.

After the experiment, each participant took part in a post-experiment interview, aimed at collecting subjective impressions on the difficulty of the tasks, opinions and personal preferences regarding the platforms.

Here is an overview of the measures we collected for each participant:

Questionnaire data

- Demographics
- Platform and OS familiarity
- Post-task questionnaire (SEQ)
- Post-platform questionnaire (SUS)

Experimental data

- Driving task performance
- Time on task
- Visual attention data

Qualitative data

- Qualitative observations during each task
- Post-experiment Interviews

4 RESULTS

4.1 Quantitative

To address the hypotheses established in the previous section (see Objective and Hypothesis), we conducted a series of both descriptive and inferential statistical tests based on the collected data. Initially, we cleaned the dataset of any null values and plotted graphs related to key analysis items, specifically focusing on data distributions by platform and task. The graphs indicated no significant scale differences in the data that would warrant initial concern.

Therefore, to test the first three hypotheses, we performed an inferential statistical analysis using as a main statistical tool, a two-way ANOVA with repeated measures, yielding the following results.

As shown in Tables 3, 1, there was no significant main effect of the platform on both *time_on_task* and *average_game_error* ($p > 0.05$).

However, we found a significant main effect for the type of task in both metrics ($p_1 < 0.05$, $p_2 < 0.05$). Furthermore, the ANOVA results for *percentage_of_distraction* revealed significant differences for both the main effect of the platform and the type of task ($p < 0.05$).

Given the observed variance differences across tasks, we performed three ad-hoc two sample t-tests on the different tasks for the different reference metrics, adjusting for Bonferroni's correction.

The post-hoc analyses of the task pairs indicated significant differences in *time_on_task*, *average_game_error*, and *percentage_of_distraction* for Task 3. These findings, supported by the plot visualisations, suggest that the Android platform overall performed better, particularly in terms of *average_game_error* and *time_on_task*.

The results of these statistical analyses demonstrate that the platform did not significantly impact time on task and driving task performance. However, it was observed a main effect of the platform on visual attention: more specifically, by looking at the plots (Fig. 5), Android Auto appears to demand less visual attention than CarPlay.

Regarding hypothesis number four (see Objective and Hypothesis), we decided to perform a paired t-test on both the individual SUS results and the SEQ ones. As we can see from the data in the table (Table 4), we found no significant differences for either SUS or SEQ, which means that the two platforms are perceived to be more or less the same in terms of usability.

In summary, the analysis demonstrated that while the platform mainly influenced distraction levels, the type of task played a significant role in all performance metrics. Furthermore from a quantitative perspective, usability perceptions did not differ significantly between the platforms.

4.2 Qualitative

In addition to quantitative measures, qualitative data was collected through interviews and observations for each user to support and contextualise the findings. This data served two key and integrated purposes:

(1) Guide the analysis of performance data.
(2) Ground the quantitative results in participants' perceptions and real-time experiences.

Interviews and observations also provided valuable insights for future research and potential improvements to the IVIS. Interviews focused on understanding participants' experiences with the system, including the challenges they faced during tasks and the strategies they adopted. Observations served to validate these reported experiences by directly measuring task performance. We start from analysing the results obtained from the interviews conducted at the

Table 1: ANOVA Time on Task

Factor	p-value
Platform	0.28
Task Type	3.9×10^{-10}

Table 2: ANOVA Average Game Error

Factor	p-value
Platform	0.25
Task Type	1.2×10^{-8}

Table 3: ANOVA percentage distraction

Factor	p-value
Platform	0.038
Task Type	3.5×10^{-8}

Table 4: Paired t-test SUS and SEQ

Metric	p-value
SUS	0.55
SEQ	0.60

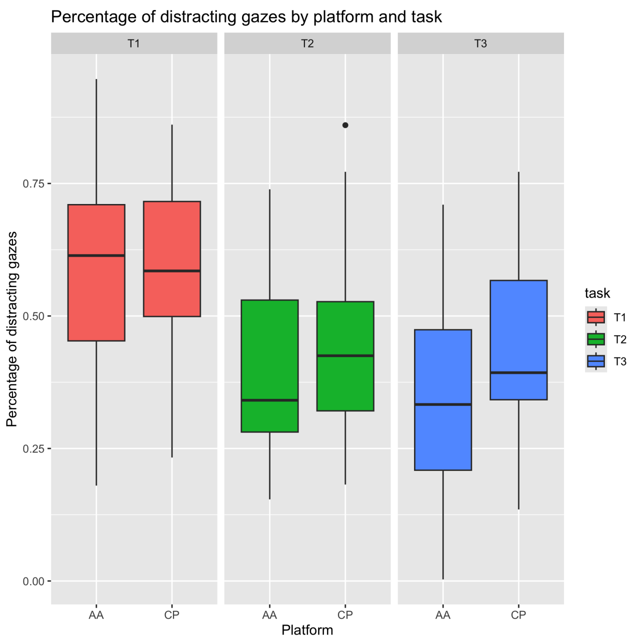

Figure 5: Percentage of distracting gaze (visual attention metric) by platform and task

end of the tests. User feedback from interviews aligned with quantitative data on task difficulty, by revealing navigation as the most challenging task. As a matter of fact, the additional request of avoiding tolls significantly increased task complexity and completion time, contributing to the perceived difficulty. Next, user preferences diverged regarding aesthetics and usability. In terms of aesthetics, a clear preference emerged for Apple CarPlay (66.6However, when it comes to usability, Android Auto was perceived as more user-friendly and intuitive. Reasons included the more prominent voice assistant button, which was available on every screen, compared to Apple CarPlay's Siri (only visible in the calls app). Additionally, the multi-step navigation task on Android Auto mirrored the familiar Google Maps application, which is the default for Android, contributing to its perceived ease of use. It's important to note that familiarity plays a significant role in both aesthetic preference and perceived usability. In fact, many participants reported favouring the platform that matched their phone's operating system. According to our observations, the voice assistant usage doesn't differ between interfaces but among tasks: task 3 (call) appears to be completed with the assistance of the voice command more times than task 2 (music). These differences may be affected by the requirements to complete each task: task 3 and 1 required the use of the keyboard, while task 2 relied on searching through icons in order to be completed. Moreover, some participants attempted to use the voice assistant, but it failed and eventually completed the task manually by increasing the time on task. Failure with the voice assistant mainly consisted in: misunderstanding the command by the voice assistant or confusion among the voice assistants specific for each app. For what concerns the difficulty of the tasks, it emerged that task 1 is harder to perform due to the "avoid tolls" option to find and select. As a matter of fact, we observed that the majority of the time on task 1 was dedicated to this requirement. The reason might be that the users' expectations about the steps to perform the task were different from the platforms' ones (gulf of execution and evaluation). In particular, it emerged that task 1: needed more steps than the others, offered multiple possible paths and unmatched interaction possibilities with the users' search expectations.

Table 5: Percentage of Voice Assistant Use through Tasks and Interface

	Task 1	Task 2	Task 3
CP	29%	3%	51%
AA	25%	3%	62%

5 LIMITATIONS AND FUTURE WORK

It's important to note that despite our efforts this experiment has some limitations. Firstly, although suitable to elicit the pattern of interaction with the IVIS that we were looking for, our experimental setting is not necessarily fully representative of true usage conditions. We noticed that a small number of participants, when struggling with the IVIS task, would completely disregard the driving task, sometimes even for more than 10 seconds at a time. Such a situation is not realistic, as, if replicated in a real driving setting, would surely result in a crash. We attribute this to a failure of the driving task to transmit a sufficient sense of danger. In order to mitigate this problem, we recommend future researchers using this method to implement some feedback in the driving task, maybe haptic or auditory, once a certain threshold of error is surpassed in order to bring the attention of the participant back to the driving task. Moreover as an additional experimental research, it would have been useful to compare the performance of distraction using the phones (as a baseline) instead of IVIS systems, in order to inquire about the role of "decreasing the distractions" by which IVIS systems have initially been developed. Finally, while detecting the gaze, we measured only overt visual attention: locations of gaze do not necessarily reflect directly what human observers are paying attention to. While overt visual attention is the act of physically directing the eyes to a stimulus, covert visual attention is related to a mental shift of attention without physical movement [9].

6 CONCLUSION AND REFLECTIONS

By merging the quantitative and qualitative results, we conclude the following statements. The two platforms were similar in terms of time required to complete analogous tasks and in terms of their impact on performance in the simulated driving task. However, Android Auto generally had a significantly smaller detrimental impact on visual attention compared to CarPlay. Additionally, the study indicates the inherently different complexity of the tasks imposes varying demands on visual attention. Furthermore, familiarity emerges as a crucial variable influencing users' overall perceptions of their interactions with these systems. Moreover, it was also noticed that CarPlay is considered to be more aesthetically pleasant, while, despite no significant difference in usability scores (SUS, SEQ), in the interviews Android Auto was perceived as easier to use. Finally, there was a significant difference between the two platforms in task 3 (making a phone call). In this task, Android Auto performed significantly better in all experimental measures (time on task, average game error, and visual attention). We attribute these results to differences in the interfaces and in steps required to perform a call, although this is just an interpretation.

A ADDITIONAL RESOURCES

- Google Drive folder
- GiHhub repository
- Qualitative thematic analysis (of the interviews)

REFERENCES

[1] [n. d.]. Smartphone Use While Driving Grows Beyond Texting to Social Media, Web Surfing, Selfies, Video Chatting. https://www.prnewswire.com/news-releases/smartphone-use-while-driving-grows-beyond-texting--to-social-media-web-surfing-selfies-video-chatting-300085207.html.

[2] 2015. Mobile phone use while driving: Underestimation of a global threat. *Journal of Transport Health* (2015). https://www.sciencedirect.com/science/article/abs/pii/S221414051500700

[3] John Brooke. 1995. SUS: A quick and dirty usability scale. *Usability Eval. Ind.* 189 (11 1995).

[4] Marc Carlier. 2022. Consumer interest in Android Auto and Apple CarPlay - U.S. survey 2017. https://www.statista.com/statistics/858655/survey-among-us-vehicle-owners-on-android-auto-and-apple-carplay/.

[5] Stefano C. Castro, David L. Strayer, Dora Matzke, and Andrew Heathcote. 2019. Cognitive workload measurement and modeling under divided attention. *Journal of Experimental Psychology: Human Perception and Performance* 45, 6 (2019), 826–839. https://doi.org/10.1037/xhp0000638 Epub 2019 Apr 18.

[6] Eunji Chong, Elysha Clark-Whitney, Audrey Southerland, Elizabeth Stubbs, Chanel Miller, Eliana L. Ajodan, Melanie R. Silverman, Catherine Lord, Agata Rozga, Rebecca M. Jones, and James M. Rehg. 2020. Detection of eye contact with deep neural networks is as accurate as human experts. *Nature communications* 11, 1 (12 2020). https://doi.org/10.1038/s41467-020-19712-x

[7] Manuel Held, Jan W. Rieger, and Jelmer P. Borst. 2024. Multitasking While Driving: Central Bottleneck or Problem State Interference? *Human Factors* 66, 5 (2024), 1564–1582. https://doi.org/10.1177/00187208221143857 Epub 2022 Dec 6.

[8] Oscar Oviedo-Trespalacios, Mark King, Md Mazharul Haque, and Simon Washington. 2017. Risk factors of mobile phone use while driving in Queensland: Prevalence, attitudes, crash risk perception, and task-management strategies. *PloS one* 12, 9 (2017), e0183361. https://doi.org/10.1371/journal.pone.0183361

[9] Yashas Rai and Patrick Le Callet. 2018. Visual attention, visual salience, and perceived interest in multimedia applications. In *Academic Press Library in Signal Processing*. Vol. 6. Academic Press.

[10] Donna Tedesco and Thomas Tullis. 2006. A Comparison of Methods for Eliciting Post-Task Subjective Ratings in Usability Testing ABSTRACT. *Usability Professionals Association (UPA)*.