The Impact of Different Human-Machine Interface Feedback Modalities on Older Participants' User Experience of CAVs in a Simulator Environment

Iveta Eimontaite[1], Alexandra Voinescu[2], Chris Alford[1], Praminda Caleb-Solly[3], Phillip Morgan[4]

[1] University of the West of England, Faculty of Health and Applied Sciences, Bristol BS16 1QY, UK
{Iveta.Eimontaite, Chris.Alfird}@uwe.ac.uk
[2] University of Bath, Department of Psychology, Claverton Down, Bath, BA2 7AY, UK
A.Voinescu@bath.ac.uk
[3] University of the West of England, Bristol Robotics Laboratory and Institute of Bio-Sensing Technologies, Bristol BS16 1QY, UK
Praminda.Caleb-Solly@uwe.ac.uk
[4] Cardiff University, School of Psychology, Cardiff, CF10 3AT, UK
MorganPhil@cardiff.ac.uk

Abstract. Rapidly developing Autonomous Vehicle (AV) technology has potential to provide solutions to some of the aging population challenges, such as social isolation resulting from an inability to be independently mobile. However for AVs success, users' acceptance is essential. Fifteen participants (M 70 years) participated in an autonomous driving simulator trial with voice-based CAV status feedback in a decision-making scenario – whether to pick up a friend on the way. The within-subject conditions/journeys were: Audio feedback (Audio)/Pick-Up; Audio/No-Pick-Up; No-Audio/Pick-Up. Additionally, the effect of feedback during different external journey conditions was also considered, resulting in two between-subjects conditions – day and night travel. Participants physiological, cognitive and affective measures show greater situational awareness and workload ratings in the No-Audio/Pick-Up condition with increased Post-trial trust rating and overall higher positive affect. These results indicate that the greatest concentration was required in the no-sound condition, suggesting that sound/multimodal feedback improved ease of operation and journey experience.

Keywords: Connected Autonomous Vehicles · Human-Machine Interaction · Feedback Modalities · Older Participants · Hear Rate · Trust · Task Load

1 Introduction

An ageing population is a global reality, posing challenges for older adults with ageing-related impairments of how to remain an independent and active member of society

---

for as long as possible. At the same time, rapid development of Autonomous and Connected Autonomous Vehicles (AVs and CAVs) offers a potential solution to some of these problems, as it such a mobility option could provide independence and flexibility when people are no longer able to drive a conventional vehicle (and indeed for individuals who have never driven). This technology is already being trialed on roads with Uber and Google cars, as well as other car manufacturers, such as Volvo, Toyota, racing to deploy their autonomous vehicles on mass scales [1]. However, rapid progress of technology development has greatly emphasized the need for the investigation into human factors and user attitudes of AVs and CAVs so that the most optimal systems and interfaces can be designed and thoroughly tested with users before manufacture and deployment.

Older adults who no longer drive themselves, or may soon face not being able to drive, are seen as potential early adopters of CAV's, particularly those which offer Level 5/full autonomy [2]. While achieving independence is a priority to maintaining quality of later life [3], ensuring user acceptance of CAVs is predicated by building trust and confidence in the technology [4].

Driving, with its high cognitive demand and combined with rapidly changing and improving technology, such as CAVs, can be a cause of insecurity and anxiety for many people. Uncertainty and decision-making anxiety influence mental and physical illnesses [5], and it is expected that reducing uncertainty through the use of effective Human-Machine Interfaces (HMIs) can help improve the mental and physical well-being for the individual. Indeed, information communication designed efficiently can increase user empowerment and sense of control in an unfamiliar situation. For example, in a healthcare context, ergonomically designed leaflets encourage patient- doctor discussion [6] leading to feelings of being in control and greater empowerment [7]. Furthermore, feeling in control and empowered to deal with unfamiliar reduces stress and improves general well-being [8–10].

Designing effective in-vehicle human-machine interfaces (HMIs) is an important step to increasing user acceptance; especially for very high or fully AVs that are likely not going to need conventional driving controls. An experiment was designed to test user experience of such a HMI developed based on leading human factors HMI design principles and views from an older adult population group on factors such as HMI accessibility, usability, and functionality. The aim of the current CAV simulation study is to investigate participants' behavior and associated physiological response to different visibility conditions (day vs. night) as well as CAV information communication methods (sound and text notifications vs. only text notifications) and decision-making during the trip (decision to pick a friend either provided or not). It was predicted that we would observe increased physiological arousal and increased cognitive demands and more negative affective measures in the night condition, and, in particular, with No Audio and Friend Pick Up trials.

2 Method

2.1 Participants

Eighteen participants (5 female) took part in the trial (*M* age = 70-years, *SD* = 4.70, range = 61-81). Data from three (one female) were removed from the analysis due to technical issues with the HMI/dashboard during at least one of the journeys. All participants had a valid driving license and reported driving between 2, 500 and 7, 400 miles per year. All had corrected vision and seven indicated that they had corrected hearing. All participants were highly functional and able to complete the entire trial. The following exclusion criteria were set to ensure comfort and safety of the participants in the simulator: the presence of any severe health conditions (i.e. epilepsy, neurological impairments, heart surgery) and severe simulator sickness after screening.

Participation was voluntary and each participant received a £20 voucher to cover transportation and associated costs. The trial was approved by the University of the West of England, Health and Applied Sciences (HAS) Research Ethics committee.

2.2 Design and Materials

Design. The study used a mixed design to investigate what effects the between-subjects variable of Visibility (day vs. night) and within-subject variables of Audio (audio and on screen notifications vs. on screen notifications only) and Pick-Up (pick up a friend vs no pick up of a friend) have on participants' experiences in the simulated autonomous vehicles. Participants experienced three journeys in the trial:

"Audio/Pick-Up", (baseline measure)

"Audio/No Pick-Up", and

"No Audio/Pick-Up".

In addition to pre-trial and post-trial measures, each journey was followed by repeated measures of trust, situation awareness, task load, and user experience questionnaires. Physiological measures (EDA, Heart Rate, and Skin Temperature) were measured throughout the three completed journeys (more detailed list of measures in *Measures*).

Simulator. The CAV simulation environment consisted of a Lutz Pathfinder pod shell (Fig. 1 and 2), three large screens on to which the images from the simulator was projected. The pod was a 2-seater adapted pod with two doors (right and left). For the purpose of our study, the steering wheel was removed and there were no pedals, as we simulated Level 5 autonomous vehicle journeys [2] with no human input (except for initial journey set-up via the HMI) as the simulator was programmed to run autonomously for the full journeys incorporating all elements of all journeys and in all environments.

The Bristol Robotics Laboratory (BRL) driving simulator environment was set-up and modelled on a Honda Civic family hatchback car with a five-speed gearbox (noting no transmission devices such as a gearstick within the simulator). The driving environment was projected at a resolution of 1280 × 1024 pixels onto three large forward projector screens giving a 210° horizontal forward field of view. A stereo sound system provided simulated engine, road, and traffic sounds. The driving simulation was generated by the SCANeR II® software (OKTAL Sydac, France). The experimental driving route consisted of a two-way road passing through a rural neighbourhood with traffic (including cars and buses) passing in the other lane in the opposite direction. The scenarios also included people/avatars walking along pavements or waiting at e.g., bus stops and animals/avatars positioned for the vehicle to slow down.

Human- Machine Interface (HMI). The design of the HMI was informed by literature reviews [11] and public engagement workshops (e.g., [12]) that synthesized best practices (mainly human factors and ergonomics) and recommendations (including participant group needs and requirements), as well as existing knowledge and experience of designing accessible interfaces for older adults. Example principles included: reduce screen clutter, use large icons, and use of icons that are highly intuitive (e.g., fit with mental models). The HMI displayed the vehicle speed, time remaining until destination, a safe stop button, and a journey map (Fig. 1b tablet on the left), as well as journey set up/change options and basic infotainment options (Fig. 1b tablet on the right).

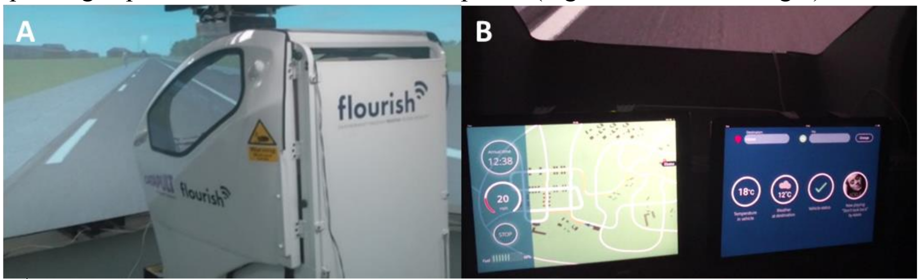

Fig. 1. (a) Exterior of the Lutz pod; (b) Interior of the Lutz pod and HMI set up

The HMI was implemented on two 12.9 inch iPad Pro with ED backlit display with iPS technology; retina display; 2732 x 2048 resolution at 264 pixels per inch, and fingerprint resistant coating.

Scenarios. Participants completed three CAV journeys. For each journey they were provided with a scenario that specified that they would be going from the Town Hall back to their home. Two of the variables we were testing were *Audio* (CAV speaks) and *Pick-Up* (user able to intervene with pre-planned journey). Therefore, two further conditions were added to manipulate our variables and compare to the baseline: *Audio/Pick-Up* vs *No Audio/Pick-Up* and *Audio/Pick-Up* vs *Audio/No Pick-Up* (Table 1).

Table 1. Independent variable specification within the experimental scenarios

Scenario	Notification modality	Decision option
Audio/Pick-Up	Audio and text	Friend pick up option
No Audio/Pick-up	Text	Friend pick up option
Audio/No Pick-Up	Audio and text	No Friend Pick up option

2.2 Scales

The trial involved the use of a combination of pen and paper questionnaires distributed for the participants to fill in before, after and during the main simulator components of the trial (Table 2. In addition to these, physiological measures of heart rate, electrodermal properties of the skin, and peripheral skin temperature were taken using an Empatica E4 wristband (Empatica, US) to measure levels of arousal and stress.

Table 2. Questionnaires and scales used during the main component of current experiment

Measure	Authors	Details	Conducted
Motion Sickness Assessment Questionnaire (MSAQ)	Gable, Walker, & Gable, 2013	The scale contains 17 items (e.g. I feel nauseated) that measure symptoms that usually occur if/when one experience simulator sickness. Participants have to rate on a 10-point Likert scale ranging from 0 "not at all" to 10 "severely" any of the symptoms described.	Pre-trial Post-trial
Positive and Negative Affect Schedule (PANAS)	Watson, Anna, & Tellegen, 1988	20 items that describe positive and negative emotions (e.g. excited, guilty). Participants have to rate the extent to which they currently felt emotions (e.g. distressed, alert) on a 5 point Likert-type scale (1 = not at all, 5 = extremely).	Pre-trial After each journey
Raw NASA Task Load index	Hart & Staveland, 1988	6 subscales and 6 items that measure subjective workload on the following dimensions: Mental demand, Physical demand, Temporal demand, Effort, Performance and Frustration Participants have to rate on a Vas-type scale ranging from low to high how much each scale dimension/subscale contributes to task load.	Pre-trial After each journey
Pitsburg Sleep Quality Index Short Form (PSQI)	Buysse et al. , 1989	The scale consisted of 4 items that refer to sleep habits/subjective sleep quality (e.g. "During the past month, when have you usually gone to bed at night").	Pre-trial
Situation Awareness rating technique (SART)	Taylor & Selcon, 1990	Situation awareness refers to the perception of environmental cues, their position and meaning, with a focus on the near past, current, and future situations. Nine items (e.g. Is the situation highly unstable and likely to change suddenly (High) or is it very stable and straightforward (Low)?) Participants have to respond on a scale from 1 "high" to 7 "low".	After each journey
Trust in Automation Questionnaire (TiA)	Gold et al., 2015	The scale contains 19 items on a Likert-type rating scale from 1 "strongly disagree" to 5 "strongly agree" that measure participants' trust in automation (e.g. The system state was always clear to me). The questionnaire is structured into five subscales (Reliability/Competence, Familiarity, Trust, Understanding, and Intention of Developers) which each contain between two and four items.	After each journey Post-trial

2.3 Procedure

Written consent from participants was obtained prior to the experiment. More specifically, an induction 'In-Vehicle Participant Workshop' was held a few months before the actual simulator trial. The purpose of the workshop was to inform participants about the aim of the project, timescale and to gather information about participants' expectations of autonomous vehicles. It also served as an element of the iterative process to design the HMI, as well as a way to keep all participants' expectations about autonomous (including connected) vehicles/driverless car at a similar level as they can likely be biased through different personal beliefs and media coverage of the topic. The experiment aims, research questions, variables, or predictions were not disclosed during this workshop.

During the simulator journey phase, participants were seated in the simulator and experienced the virtual journeys while interacting with the HMI. At the beginning of the first journey, the participant received the journey familiarization instruction sheet that contained an overview of the journeys tasks, including familiarization with the Lutz pod and simulation environment, instructions for the first journey and what to do if experiencing motion sickness symptoms. Half of the participants were in Day driving condition (clear visibility, 12:00 noon) with the others in the Night condition (limited visibility, 00:00 midnight). Audio/No Pick-Up were always included as a second journey, and the order of No Audio/Pick-Up and Audio/Pick-Up was counterbalanced between participants. Each journey lasted 6-minutes. After each journey, participants completed measures (Table 1; "After each journey" as well as usability and user experience scales which are not discussed here) and were provided with a scenario description for the next journey.

The testing session lasted for approximately 165-minutes, with increased variability depending inter-subject individual differences (ranging from 140 – 240 minutes).Noting that breaks were offered and taken throughout the experiment at the request of the participant and researcher. After the journey phase, participant performance was measured using a Standardized Assessment Framework, developed for the current project, which includes a combination of cognitive tasks and questionnaires (not detailed here).

2.4 Analysis

Z-scores were calculated for raw physiological data recorded through the Empatica E4 wristband (as recommended by Ben-Shakhar (1985)). Time stamps for when journeys started and ended were taken and z-scores for each physiological measure for these time durations were compared between the three different scenarios ("Audio/No Pick-Up ", "Audio/Pick-Up ", and "No Audio/Pick-Up ") and two conditions (*Day* and *Night*).

Questionnaires, scales and measures were pre-processed following original questionnaire coding instructions. Quantitative analysis was performed with *IBM SPSS Statistics 23*. Mixed analysis of variance (ANOVA) tests were conducted with pre-planned comparisons with the baseline condition. Bonferroni corrections were applied for the post hoc comparisons.

3 Results

3.1 Controlling for group differences

To control for any between subject differences, participants' age, average number of miles they drive in a year, PSQI, MSAQ, PANAS, and NASA TLX scores were compared between day and night group participants. Independent *t*-tests did not reveal any significant differences in these scores (all $ts(13) \leq 1.15$, $ps \geq .27$)[2], indicating that both participant groups were homogenous.

Further analysis were performed by comparing measures taken to assess participants' physiological states, affective, and cognitive scores depending on the three different scenarios (*Audio/Pick-Up*, *Audio/No Pick-Up*, *No Audio/Pick-Up*) and group (*Day/ Night)*. The results are presented in the following sections.

3.2 Physiological arousal as a function of scenario

EDA. A mixed ANOVA with a between-subject measure of Group (Day/ Night) and within-subject measure of Scenario with EDA as the dependent variable showed a non-significant main effect of Scenario, although there was a trend, $F(2, 24) = 2.73$, $p = .09$, $\eta_p^2 = .19$. There was a non-significant main effect of Group, and a non-significant Scenario x Group interaction (all $Fs(2, 24) \leq 0.67$, all $ps \geq .52$) (Fig. 2a).

Heart Rate.An equivalent mixed ANOVA compared participants' heart rate over the three Scenario types indicated a significant main effect of Scenario, $F(2, 24) = 4.11$, $p = .03$, $\eta_p^2 = .26$) a non-significant main effect of Group, and a non-significant interaction between Scenario x Group (all $Fs(2, 24) \leq 0.46$, all $ps \geq .64$).

Exploring the main effect of Scenario, Bonferroni pairwise comparisons indicated that *No-Audio/Pick-Up* Scenario resulted in the highest heart rate compared to *Audio/Pick-Up* ($p = .05$) and *Audio/No Pick-Up* ($p = .046$). There was no difference between *Audio/Pick-Up* and *Audio/No Pick* scenarios ($p > .99$) (Fig. 2a).

Skin Temperature. A mixed ANOVA revealed a significant main effect of Scenario ($F(2, 24) = 7.03$, $p = .004$, $\eta_p^2 = .369$), but a non-significant main effect of Group, and a non-significant Scenario x Group interaction (all $Fs(2, 24) \leq 0.92$, all $ps \geq .41$).

Exploring the main effect of Scenario, pairwise comparisons indicated that *No-Audio/Pick-Up* Scenario resulted in highest skin temperature, compatible with greater emotional arousal, compared to *Audio/Pick-Up* ($p = .039$) and *Audio/No Pick-Up* ($p = .030$) There was no significant difference between *Audio/Pick-Up* and *Audio/No Pick-Up* scenarios ($p > .99$) (Fig. 2a).

---

[2] Non-significant results are provided in a summarized format, i.e. all $ts(13) \leq 1.15$, all $ps \geq .27$ would mean that all other *t* values are lower than 1.15, and *p* values are higher than .27.

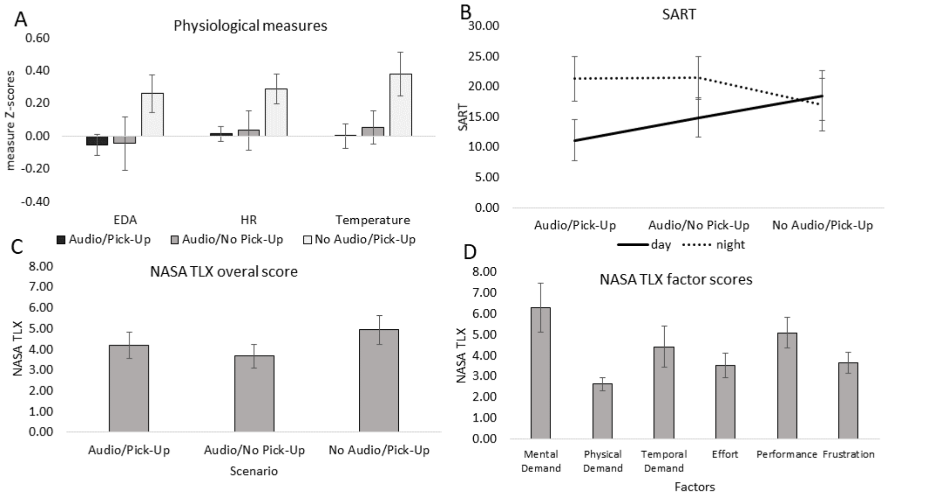

Fig. 2. (a) Physiological measures z-scores in three experimental conditions; (b) SART scores in three experimental conditions as a function of visibility (day and night), (c) NASA-TLX overall scores in the experimental conditions and (d) for each factor separately. All figures show error bars +/- SEM.

3.3 Affective states as a function of scenario

Trust. A mixed ANOVA with a between-subject measure of Group (Day/ Night) and within-subject measure of Scenario with dependent variable of trust scores revealed a non-significant main effect of Scenario, but there was a trend ($F(1.25, 16.30) = 3.72$, $p = .06$, $\eta_p^2 = .22$. The interaction between Scenario x Group was non-significant ($F(1.25, 16.30) = 2.57$, $p = .123$, $\eta_p^2 = .165$), as was the main effect of Group ($F(1, 13) = 0.53$, $p = .481$).

Exploring the marginally non-significant main effect of Scenario, pairwise comparisons indicated that the *No-Audio/Pick-Up* Scenario did result in significantly higher trust scores compared to *Audio/Pick-Up* ($p = .001$) but not compared to *No Pick-Up/Audio* ($p = .210$). The difference between *Audio/Pick-Up* and *No Pick-Up/Audio* scenarios was not significant ($p > .99$) (Table 3).

Interestingly, comparing post-test trust scores with all three scenario trust scores, revealed a significant main effect of Scenario ($F(1.37, 17.85) = 3.86$, $p = .016$, $\eta_p^2 = .23$), a non-significant main effect of Group, and a non-significant Scenario x Group interaction (all $Fs(1.37, 17.85) \leq 2.48$, all $ps \geq .126$). Pairwise comparisons indicated that trust post-test scores significantly increased compared to *Audio/Pick-Up* ($p = .032$) and *No-Audio/Pick-Up* Scenario scores ($p = .022$), but there was no significant difference with *Audio/No Pick-Up* Scenario Trust scores ($p < .999$) (Table 3).

Table 3. Mean PANAS and trust scores after each of three Scenarios and post-test scores overall and as a function of Group (Day/Night)

		Audio/ Pick-Up	*Audio /No Pick-Up*	*No Audio/ Pick-Up*	*Post test*
Day	PANAS positive affect	34.0 (6.8)	36.0 (9.0)	36.6 (8.5)	N/A
	PANAS negative affect	11.3 (2.5)	10.1 (0.4)	10.4 (1.1)	N/A
	Trust	3.3 (0.8)	2.9 (1.4)	3.8 (0.5)	3.4 (0.6)
Night	PANAS positive affect	36.3 (11.5)	10.1 (0.4)	39.7 (9.1)	N/A
	PANAS negative affect	7.6 (4.5)	7.4 (5.1)	8.1 (3.9)	N/A
	Trust	3.0 (0.5)	3.7 (0.7)	3.9 (0.5)	3.9 (0.5)

PANAS scores. A mixed ANOVA with a between-subject measure of Group (day/ night) and within-subject measures of Scenario and PANAS Factor (positive affect/ negative affect) revealed a significant main effect of PANAS Factor ($F(1, 13) = 108.83$ $p \leq .001$, $\eta_p{}^2 = .893$). The main effect of Scenario as well as was as main effects of Group, interactions Scenario x Group, Scenario x Factor, Factor x Group, Scenario x Factor x Group were non-significant (all $Fs(2, 26) \leq 2.54$, all $ps \geq .099$)

An investigation of the main effect of PANAS Factor revealed participants overall felt significantly more positive affect compared to negative affect ($t(14) = 10.41$, $p \leq .001$; positive affect mean = 36.60, $SD = 9.01$, negative affect mean = 9.24, $SD = 3.35$). Other comparisons were not significant (all $ts(14) \leq 1.45$, all $ps \geq .169$, Table3).

3.4 Cognitive demand as a function of scenario

Situation Awareness. A mixed ANOVA with a between-subject measure of Group (Day/ Night) and within-subject measure of Scenario and dependent variable of SART score revealed a non-significant main effect of Scenario and Group ($F(1, 13) = 1.25$, $p = .25$) and a marginally non-significant Scenario x Group interaction with a clear trend ($F(2, 26) = 3.24$, $p = .056$, $\eta_p{}^2 = .20$)(Fig. 2b).

A further investigation of marginally non-significant interaction of Scenario x Group revealed that the day participant Group scored higher than night Group participants in the *Audio/Pick-Up* Scenario, although it was a non-significant trend. Other *t*-test comparisons were also non-significant (all $ts(13) \leq 1.37$, all $ps \geq .194$).

Task Load. A mixed ANOVA with a between-subject measure of Group (day/ night) and within-subject measures of Scenario and Factor (*mental demand, physical demand, temporal demand, effort, performance, frustration*) revealed significant main effects of Scenario ($F(2, 26) = 5.46$, $p = .011$, $\eta_p{}^2 = .296$) and Factor ($F(2.19, 28.42) = 5.58$, $p \leq .001$, $\eta_p{}^2 = .300$). Other main effects and interactions were not significant (all $Fs(2.19, 28.42) \leq 1.19$, all $ps \geq .325$).

Exploration of the main effect of Scenario showed that overall cognitive load was significantly higher in *No-Audio/Pick-Up* scenario compared to *Audio/Pick-Up* scenario ($p = .011$). The mean differences between *Audio/Pick-Up* vs no pick *Audio* as well as *Audio/Pick-Up* vs. *No-Audio/Pick-Up* were not significant, $p = .34$ and $p = .41$, respectively (Fig. 2c)

Furthermore, looking at the six component factors of this measure, there were significant differences between mental demand vs. physical demand and physical demand vs. performance (*ps* = .041 and *p* = .012, respectively). Other mean differences were not significant (all *ps* ≥ .126; Fig. 2d).

4 Discussion

The current study investigated the effect of different feedback modalities (audio/no-audio/with text) within a CAV human-machine interface (HMI) on user experience of older participants who underwent a series of level 5 autonomy simulated journeys. Subjective ratings of affect included trust and positive and negative mood state. Situational awareness and workload scales provided measures of task load and engagement, and physiological measures including heart rate provided an index of physiological arousal for comparison with the subjective measures.

Skin temperature, heart rate and EDA indicated increased response, reflecting greater arousal, in the No-Audio/Pick-Up Scenario. Conversely, the lowest response was in the Audio/Pick-Up Scenario. The highest arousal scores were experienced in the No-Audio/Pick-Up condition, and this is consistent with literature showing that not knowing what is happening is related to the feeling of being out of control, while on the other hand, a feeling of being in control can decreased experienced stress and anxiety [8–10]. Furthermore, higher scores on EDA, heart rate and skin temperature measures were associated with higher experienced negative emotions and anxiety [20, 21].

Affect measures indicated overall higher positive affect scores than negative affect scores. This is perhaps not surprising as the participants were a self-selecting group. In terms of trust, the highest ratings were in the No-Audio/Pick-Up scenario. Although at first sight this result seems counterintuitive, it could be explained by considering the performance of the CAV in the simulation. In this no Audio condition participants did not know what actions the CAV simulator was going to take (No Audio), therefore successful completion of the journey could have led towards increased trust compared to scenarios where participants were informed of CAV behaviour. Furthermore, and consistent with the past literature, trust and attitudes towards automation increases with greater exposure to it [22, 23] and depends on the automation behavioural characteristics – such as working without errors [24]. Therefore, it is not surprising that participants' post-scenario measures of trust show higher trust scores compared to Audio/No Pick-Up and Audio/Pick-Up scenarios.

Furthermore, cognitive measures indicated that scenario type affected participants' cognitive response to the CAV. With situation awareness, the interaction between scenario type, and whether participants experienced the journey in a day or night environment, resulted in a trending difference in situation awareness scores. Participants in the Night condition showed increased situation awareness in the Audio/No Pick-Up and Audio/Pick-Up scenarios compared to participants in the Day, yet in the Audio/Pick-Up scenario, both groups of participants had very similar scores. Decreased situation awareness suggests that individuals relax and trust the technology, in this case the CAV. On the other hand, high situation awareness indicates that individuals are in fight or flight readiness and suggests feelings of tension towards the environment[25].

Results reveal situational awareness and workload ratings in the no-Audio, pick up a friend condition, compared to the only condition. This was supported by the greatest increase in heart rate suggesting higher levels of both physiological and subjective arousal in the No Audio condition with similar levels in the two Audio conditions. Trust scores increased significantly post-test after the pick-up conditions, with positive affect higher than negative affect throughout.

Taken together, the findings indicate that older adult participants found the simulated CAV journey a positive experience with increasing trust, based on their HMI interaction and journey experience. However, the greatest concentration was required in the no-Audio notifications condition, suggesting that sound/multimodal feedback improved ease of operation and journey experience.

The findings are important to inform future CAV HMI design guidelines for this user group. In particular, clear communication of vehicle behaviour to increase trust and user experience. Furthermore, information communication is recommended in both modalities (text and audio feedback) to help the user feel more relaxed during the journeys and trusting that the vehicle will cope with user specified decisions. Future work will explore aspects such as how the user is affected by different levels of explainability of vehicle behaviour and together with user control of the level and modality of the feedback from the vehicle.

References

1. Jee, C., Mercer, C.: Driverless car news: The great driverless car race: Where will the UK place?, https://www.techworld.com/apps-wearables/great-driverless-car-race-where-will-uk-place-3598209/, (2017).
2. SAE International: U.S. Department of transportation's new policy on automated vehicles adopts SAE International's levels of automation for defining driving automation in on-road motor vehicles. (2016).
3. Musselwhite, C., Haddad, H.: Mobility, accessibility and quality of later life. Qual. Ageing Older Adults. 11, 25–37 (2010).
4. Abraham, H., Lee, C., Brady, S., Mehler, B., Reimer, B., Coughlin, J..: Autonomous Vehicles and Alternatives to Driving: Trust, Preferences, and Effects of Age. Presented at the Transportation Research Board 96th Annual Meeting , Washington DC, United States (2017).
5. Moreno-Jiménez, B., Rodríguez-Carvajal, R., Garrosa Hernández, E., Morante Benadero, M.A., others: Terminal versus non-terminal care in physician burnout: the role of decision-making processes and attitudes to death. Salud Ment. 31, 93–101 (2008).
6. Mills, M.E., Sullivan, K.: The importance of information giving for patients newly diagnosed with cancer: a review of the literature. J. Clin. Nurs. 8, 631–642 (1999).
7. Ussher, J., Kirsten, L., Butow, P., Sandoval, M.: What do cancer support groups provide which other supportive relationships do not? The experience of peer support groups for people with cancer. Soc. Sci. Med. 62, 2565–2576 (2006).
8. Lautizi, M., Laschinger, H.K.S., Ravazzolo, S.: Workplace empowerment, job satisfaction and job stress among Italian mental health nurses: an exploratory study. J. Nurs. Manag. 17, 446–452 (2009).
9. Ozer, E.M., Bandura, A.: Mechanisms governing empowerment effects: a self-efficacy analysis. J. Pers. Soc. Psychol. 58, 472 (1990).
10. Pearson, L.C., Moomaw, W.: The relationship between teacher autonomy and stress, work satisfaction, empowerment, and professionalism. Educ. Res. Q. 29, 37 (2005).

11. Morgan, P., Caleb-Solly, P., Voinescu, A., Williams, C.: Literature review: human-machine interface. Proj. Rep. UWE Bristol Bristol. (2016).
12. Morgan, P.L., Voinescu, A., Williams, C., Caleb-Solly, P., Alford, C., Shergold, I., Parkhurst, G., Pipe, A.: An Emerging Framework to Inform Effective Design of Human-Machine Interfaces for Older Adults Using Connected Autonomous Vehicles. In: Stanton, N.A. (ed.) Advances in Human Aspects of Transportation. pp. 325–334. Springer International Publishing (2018).
13. Gable, T.M., Walker, B.N., Gable, T.: Georgia Tech Simulator Sickness Screening Protocol. 16.
14. Watson, D., Anna, L., Tellegen, A.: Development and Validation of Brief Measures of Positive and Negative Affect: The PANAS Scales. 8.
15. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In: Advances in Psychology. pp. 139–183. Elsevier (1988).
16. Buysse, D.J., Reynolds III, C.F., Monk, T., Berman, S.R., Kupfer, D.J.: The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res. 28, 193–213 (1989).
17. Taylor, R.M., Selcon, S.J.: Cognitive Quality and Situational Awareness with Advanced Aircraft Attitude Displays. Proc. Hum. Factors Soc. Annu. Meet. 34, 26–30 (1990).
18. Gold, C., Körber, M., Hohenberger, C., Lechner, D., Bengler, K.: Trust in Automation – Before and After the Experience of Take-over Scenarios in a Highly Automated Vehicle. Procedia Manuf. 3, 3025–3032 (2015).
19. Ben-Shakhar, G.: Standardization Within Individuals: A Simple Method to Neutralize Individual Differences in Skin Conductance. Psychophysiology. 22, 292–299 (1985).
20. Bechara, A., Damasio, H., Tranel, D., Damasio, A.R.: The Iowa Gambling Task and the somatic marker hypothesis: some questions and answers. Trends Cogn. Sci. 9, 159–162 (2005).
21. Ben-Shakhar, G., Bornstein, G., Hopfensitz, A., Van Winden, F.: Reciprocity and emotions: arousal, self-reports, and expectations. CESifo Work. Pap. 1298, 1–16 (2004).
22. Bartneck, C., Suzuki, T., Kanda, T., Nomura, T.: The influence of people's culture and prior experiences with Aibo on their attitude towards robots. AI Soc. 21, 217–230 (2006).
23. Stafford, R.Q., Broadbent, E., Jayawardena, C., Unger, U., Kuo, I.H., Igic, A., Wong, R., Kerse, N., Watson, C., MacDonald, B.A.: Improved robot attitudes and emotions at a retirement home after meeting a robot. In: RO-MAN, 2010 IEEE. pp. 82–87. IEEE (2010).
24. Nomura, T., Shintani, T., Fujii, K., Hokabe, K.: Experimental investigation of relationships between anxiety, negative attitudes, and allowable distance of robots. In: Proceedings of the 2nd IASTED international conference on human computer interaction, Chamonix, France. ACTA Press. pp. 13–18 (2007).
25. McLucas, A.C.: Decision Making: Risk Management, Systems Thinking and Situation Awareness. Argos Press P/L (2003).