Driven to Discussion: Engaging Drivers in Conversation with a Digital Assistant as a Countermeasure to Passive Task-Related Fatigue

David R Large [1*], Gary Burnett [1], Vicki Antrobus [1], Lee Skrypchuk [2]

[1] Human Factors Research Group, Faculty of Engineering, University of Nottingham, UK
[2] Jaguar Land Rover Research, International Digital Laboratory, Coventry, UK
[*] david.r.large@nottingham.ac.uk

Abstract: Using a Wizard-of-Oz approach, we explored the effectiveness of engaging drivers in conversation with a digital assistant as an operational strategy to combat the symptoms of passive task-related fatigue. Twenty participants undertook two 30-minute drives in a medium-fidelity driving simulator between 13:00 and 16:30, when circadian and homeostatic influences naturally reduce alertness. Participants were asked to follow a lead-car travelling at a constant speed of 68mph, in a sparsely-populated UK motorway scenario. During one of the counterbalanced drives, participants were engaged in conversation by a digital assistant ('Vid'). Results show that interacting with Vid had a positive effect on driving performance and arousal, evidenced by better lane-keeping, earlier response to a potential hazard situation, larger pupil diameter, and an increased spread of attention to the road-scene (i.e. fewer fixations concentrated on the road-centre indicating a lower incidence of 'cognitive tunnelling'). Drivers also reported higher levels of alertness and lower sleepiness following the Vid drive. Subjective workload ratings suggest that drivers exerted less effort to 'stay awake' when engaged with Vid. The findings support the development and application of in-vehicle natural language interfaces, and can be used to inform the design of novel countermeasures for driver fatigue.

1. Introduction

Accident data suggest that driver fatigue is responsible for 20% of all road traffic accidents [1], although such data only account for situations in which drivers actually fall asleep at the wheel [2]. In addition, the early signs of fatigue, such as inattention, poor decision making and delayed reaction times, are thought to have an equally profound and debilitating effect on driving performance, and therefore contribute to a much higher proportion of collisions [3].

Fatigue is a state of reduced mental alertness that impairs performance. It is defined as a gradual and cumulative process associated with "a loss of efficiency, and a disinclination for any kind of effort" [4]. Fatigue in driving is commonly discussed in the context of pre-existing physiological factors, most notably sleep disorders, such as obstructive sleep apnoea [5] and primary insomnia [6]. However, fatigue at the wheel does not require drivers to be pre-disposed to such conditions. It can also occur due to other factors, including low brain activation levels, sustained attention and interruption of circadian rhythms (the 'body clock') [1]. Fatigue is therefore particularly prevalent during afternoon and night-time driving when physiological activity is diminishing [7], monotonous driving [8], and in situations where driving hours are long [3].

Fatigue is synonymous with sleepiness – and the terms are therefore often used interchangeably [3]. Strictly speaking, sleepiness (or drowsiness) is an indication of the likelihood of falling asleep and can be considered as difficulty in remaining awake. Nevertheless, the general sensation of weariness, feelings of inhibition and impaired activity – that define *fatigue* – are symptomatic of the onset of *sleep* [3]. Moreover, if uninterrupted, fatigue can naturally lead to sleep [9].

In an automotive context, fatigue-related research has traditionally focussed on understanding and mitigating the effects of *active* fatigue, which is derived from continuous and prolonged, task-related, perceptual-motor activities [10]. However, the role of a driver is nowadays largely defined by extended system monitoring or vigilance tasks, complex decision-making, and rare or even no overt physical activity. Such activities naturally demand greater cognitive than physical effort [11], and consequently expose drivers to the risks of *passive* fatigue [10].

When drivers are fatigued in this manner, vigilance and alertness deteriorate. This can result in deleterious effects on driving performance, including increased lane drift [12, 13, 14], poor speed control [14] and late corrections to lane positioning [12, 15], although fatigue may be quite severe before routine driving performance is noticeably affected [9]. Such decrements in performance are often attributed to *microsleeps* (or daytime parahypnagogia) – momentary episodes in which an individual loses and subsequently re-gains awareness after a brief lapse in consciousness. Such episodes are typically highlighted by physical indicators (e.g. droopy eyes, slow eyelid-closure, and head nodding), and may last for less than a second to a few seconds. However, decreases in physiological arousal, slowed sensorimotor functions and impaired information processing may be evident at much lower levels of fatigue. This can retard a driver's vigilance and their ability to respond to unusual and emergency events, resulting in slower reaction times to traffic controls and hazards [16].

1.1. Measuring Fatigue

Methodological approaches for assessing the onset and progression of driver fatigue are abundant and extremely varied. Common metrics are based on physiological measurements [9, 17] and/or behavioural indicators/eye activity [18, 19]. Driving performance [12, 15] and subjective assessment [20, 21] are also commonly employed. Nevertheless, there remains some debate

regarding the interpretation of such data and the reliability of some of these measures [9].

For example, physiological measures, such as heart-rate and heart-rate variability, are generally considered to be good relative indicators of workload/fatigue, but effects can be inconsistent. For example, studies show that heart-rate increases and heart-rate variability decreases during demanding mental processing, although it has also been shown that heart-rate decreases significantly during a monotonous driving task [22].

Drivers exhibit certain behavioural characteristics when fatigued, such as changes to eye activity, with elevated blink duration and frequency often associated with increasing fatigue [22]. Nevertheless, research has also shown that quiescence in eye movements is one of the earliest reliable signs of sleepiness [23]. Therefore, increases in the duration and reductions in frequency of visual fixations may also accompany the onset and escalation of fatigue. In particular, fatigued drivers are likely to focus on the road centre at the expense of other driving related objects such as signs, bicyclists, scenery etc. in the forward and peripheral road scene [19]. Percent Road Centre (PRC) is defined as the proportion of time that a driver's eyes are focussed on the *road centre* – typically, a 20° (horizontal) x 15° (vertical) rectangular area centred around the driver's mean point of fixation [19]. PRC has been shown to be sensitive to secondary task workload [19] and is thought to be equally perspicuous in situations of driver fatigue.

Self-report techniques are also commonly employed to determine the effects of fatigue. In general, two approaches exist: determining the driver's current perception of their level of sleepiness, i.e. sleepiness as a state characteristic (e.g. Stanford Sleepiness Scale [21]), or gaining a general perception of their propensity for sleepiness as a component of daily life, i.e. sleepiness as a trait characteristic (e.g. Epworth Sleepiness Scale [20]). In an experimental situation, techniques such as Epworth Sleepiness Scale [20] are therefore beneficial to assess or screen potential participants, whereas the Stanford Sleepiness Scale [21] is of most value to measure variations in alertness associated with specific tasks or activities under investigation. Nevertheless, there are inherent difficulties in the self-assessment of sleepiness. For example, the ability to adequately express feelings and report fatigue decreases in particular with increasing sleepiness as gaps in consciousness can occur [11]. Moreover, the gradual slip in alertness associated with increasing fatigue can be difficult to perceive by the sleepy person themselves [24].

1.2. Countermeasures for Fatigue

Research suggests that many drivers continue to drive tired despite being aware of their fatigued state [1]. It is likely that some drivers fail to fully appreciate the risks associated with fatigued driving, or justify their actions based on past experience, perceived societal norms and journey goals. Alternatively, it is possible that the low arousal state and reduced brain activation levels associated with fatigue means that drivers may not be consciously aware of the extent of their fatigued state, crucially at the time that they are most at risk, and may therefore lack the wherewithal to initiate effective proactive countermeasures.

Public awareness campaigns, such as 'THINK!' in the UK [25], are generally considered to be an effective method to improve drivers' understanding of the dangers of driving while fatigued, but still require drivers to actively engage in preventive measures. Infrastructural changes, such as changes to the road layout by adding visual stimuli, curves, rumble-strips etc. have been shown to modestly reduce the onset of fatigue and improve driving performance, but do not reliably counteract the effects of sleep loss, nor do they adequately restore driving performance to the level of a well-rested driver [1].

A variety of in-vehicle devices have also been proposed [26]. However, many of these devices only respond to the later signs of fatigue, and therefore rely on physical indicators, such as elevated blink-rate, eye and eyelid activity, and mannerisms such as nodding and yawning. The concern is that by the time that fatigue is identified in this manner, a driver's performance is likely to already be significantly impaired. Such interventions may also encourage greater risk taking, shifting the responsibility for recognising fatigue from the driver to the technology – a fatigued driver may continue driving with the belief that the system will alert them if necessary, and will therefore only modify their behaviour following intervention [1].

1.3. Social interaction and Conversation

One of the most effective operational strategies to combat the effects of fatigue is social interaction and conversation [27, 28, 29]. Social interaction, particularly conversation, has been shown to maintain alertness amongst aircraft pilots even during the circadian nadir (the lowest point of natural circadian fluctuations) [29]. Moreover, research has shown that during long-haul flights, the *absence* of conversation amongst pilots can be a predictor of declining physiological alertness [30]. Nevertheless, to maintain alertness in this manner, the subject must be actively involved in the conversation and not just listening [27].

One of the advantages of using conversation to mitigate the effects of fatigue while driving is that it does not require that the driver is presenting the symptoms of fatigue before intervention occurs. Moreover, drivers can be engaged in conversation as a matter of course, without drawing conscious attention to it as an 'intervention'. Consequently, conversation may be equally as effective at preventing the onset of fatigue as 'treating' the symptoms [28]. However, conducting a conversation requires a second interlocutor, with whom to converse. This is a role that comfortably suits a co-pilot in aviation and is often naturally adopted by a front seat passenger (if present) in a driving context.

Whilst it is clearly impractical to enforce drivers to recruit a passenger to accompany them during long, monotonous or night-time journeys, advances in speech-recognition technologies, speech synthesis, natural language understanding (NLU) and dialogue management over recent years mean that modern voice-user interfaces (VUIs) are increasingly conversational. In addition, host devices are also often embodied by a digital personality. Such technology may therefore provide a promising and viable alternative to a human interlocutor.

Moreover, given the success and popularity of digital assistants offering voice-based services (e.g. Siri, Cortana and Alexa) in current personal devices such as smartphones, it is possible to envisage a future scenario in which cars are also embodied by digital assistants that interact with drivers using free-flowing, conversational dialogue akin to conversing with a loquacious passenger. Although likely to be primarily intended to 'discuss' navigational advice, vehicle status warnings, infotainment services etc., the inherent 'conversation' (humans appear to be unable to supress their natural responses to speech, even when it originates from a computer [31]) may inadvertently keep drivers alert, and help maintain driving performance.

1.4. Aims and Hypotheses

The aim of the study was to explore the effectiveness of conversing with a digital driving assistant as a countermeasure to passive task-related fatigue. Based on the literature, it was hypothesized that drivers, who were engaged in conversation with a digital driving assistant would be less likely to exhibit the effects of passive, task-related fatigue.

Fig. 1. *Medium-fidelity driving simulator showing motorway scenario used during study (top) and driving scenario showing potential hazard situation (bottom). Note: pedestrian moves to side of car but does not enter roadway.*

2. Method

2.1. Approach and Overview

To study the effect of using conversation with a digital driving assistant as a countermeasure to passive task-related fatigue, it was important create a situation in which drivers would begin to exhibit fatigue-induced symptoms. Literature suggests that a combination of time-on-task, time-of-day and driving monotony are likely to induce such symptoms amongst drivers, with effects revealed in a driving simulator during relatively short (40 minute) drives [15] and afternoon testing [12], when circadian and homeostatic influences naturally reduce participants' alertness [32]. Adopting a similar approach for the current study, all testing was scheduled to occur in the afternoon (between 14:00 and 16:30), and drivers were asked to consume a normal lunch and refrain from consuming caffeine prior to attending. Drivers were required to drive for 1 hour. Thirty minutes accompanied by a digital assistant and thirty minutes driving alone (the order of drives was counterbalanced to avoid experience effects). In common with the aforementioned studies, a range of measures was captured to detect the onset and progression of driver fatigue.

2.2. Participants

Twenty-three people took part in the study, comprising 18 male and 5 female (median age range, 31-40 years old). The primary concern in selecting participants was that they were experienced and active drivers (mean time with licence, 11 years 10 months; current annual mileage, 7046), and did not suffer from existing sleep-related disorders. Thus, anybody presenting with a known sleep disorder, or indicating a pre-inclination towards excessive daytime sleepiness (identified using the Epworth Sleepiness Scale [20]), were excluded from testing. Driving experience and exposure were not considered as independent variables. Participants received £20 (GBP) of shopping vouchers as compensation for their time.

2.3. Apparatus, Design and Procedure

The study took place in the University of Nottingham Human Factors driving simulator. This medium-fidelity, fixed-based simulator (Figure 1) comprises an Audi TT car located within a curved screen, affording approximately 270 degrees forward and side image of the driving scene via three overhead HD projectors. A Thrustmaster force-feedback steering wheel and pedal set are faithfully integrated with the original Audi steering wheel and pedals. STISIM Drive (version 3) was used to create a standard three-lane UK motorway scenario, with a leading car travelling at a constant speed of 68 mph in lane one. Participants were asked to follow the lead car during each of the two drives, which were counterbalanced.

Table 1 Examples of Vid's opening gambits

Category	Statement
Calendar	*"I looked at your calendar and you have a meeting upcoming today at 3 o'clock. Would you like me to set a reminder for your meeting?"*
Shopping	*"It looks like you've got a few things to do on your way home this evening. You need to buy milk. Would you like me to set a reminder for you to buy milk?"*
Journey	*"There is congestion ahead. This may delay you by 5 minutes would you like me to direct you around the congestion? ... OK. I am calculating a reroute."*

During one of the drives, participants were provided with a digital driving assistant, described as *"a fully operational, prototype natural language system currently under development by a major car manufacturer"*. The digital assistant was affectionately called 'David', but this was abbreviated to 'Vid' for the purpose of the study. Vid introduced itself to participants before they began driving, and initiated conversation using the same opening gambits for each participant (see examples in Table 1). These were selected to represent likely interactions with a digital driving assistant, based on discussions with industry experts, and have been used in several other similar studies (see: [33, 34]). All interjections were designed to invite a response from participants, but did not necessarily demand a detailed reply.

Participants were told that they could use natural, 'conversational' language when interacting with Vid, and were specifically advised that they could initiate conversation. They were invited to interact with the system as they imagine they might should they own and use such a system in their own vehicle. The intercourse therefore developed differently based on participants' responses and interjections, but maintained a common thread.

2.2.1 Wizard-of-Oz:
For the purpose of conducting the study, the digital assistant was created using Wizard-of-Oz simulation, with a professional actor (Pablo) embodying Vid. Pablo was situated out of sight and used Google Hangouts to communicate with the participant. A second person accompanied Pablo and was responsible for retrieving facts and figures in real-time from the Internet and other sources, as dictated by the course of the conversation (e.g. locating requested music tracks, providing further details about news items etc.). Utterances were generated in real time, guided by a comprehensive script with appropriate language and phrasing, although Pablo was able to respond freely as necessary. Pablo delivered utterances using a subtle computer inflexion, honed through extensive training and practice sessions to conform with our understanding of drivers' expectations of future talking technology.

2.4. Measures

A range of objective and subjective performance measures were captured to assess driver alertness and workload. All measures were subsequently compared across the two drives, identified as 'Control' and 'Vid' (with drivers interacting with the digital assistant during the latter). Unless otherwise stated, mean values were calculated across each drive and compared using paired-samples t-tests. The following measures were collected:

2.3.1 Driving Performance:
Fatigue is likely to impact on both lateral and longitudinal driving behaviour, with increased variability associated with fatigued drivers [12, 13, 14, 15]. Consequently, standard deviation of lane position (SDLP) and standard deviation of speed and headway were obtained from the driving simulator.

2.3.2 Response to Hazard:
A stationary vehicle was situated on the hard-shoulder of the motorway approximately 30 miles into each drive. As participants approached the vehicle during their second drive, a pedestrian walked from behind the vehicle in the direction of the roadway, as if attending to their car (Figure 1). The headway to the parked vehicle at which drivers responded to this potential hazard (as indicated by them lifting their foot from the accelerator pedal) was recorded. A simulated 'lost signal', prior to the hazard event ensured that drivers were not engaged with Vid at the time of the hazard, thereby avoiding any confounding distraction effects.

2.3.3 Physiological Measures:
Heart-rate variability was captured using an Empatica E4 wristband. Heart-rate (HR) is defined as the number of heartbeats within a fixed period of time (usually a minute). Heartbeats have variable time durations with different oscillation patterns. Heart-rate variability (HRV) thus measures the variability of the interbeat interval (IBI) in a specified time period. HRV can be influenced by the amount and type of effort exerted or demanded by task execution (see: [35]).

2.3.4 Visual Behaviour:
Blink duration and frequency, number and duration of fixations (across the entire drive), spread of visual attention (PRC), and pupil diameter were captured using SMI eye-tracking glasses (ETG).

2.3.5 Subjective Sleepiness:
The Stanford Sleepiness Scale (SSS) [21] was used to capture self-reported sleepiness before and after each drive, with additional written comments provided by participants, immediately after submitting ratings.

2.3.6 Mood Assessment:
The UWIST Mood Adjective Check-List (UMACL) [36] was delivered before and after each drive. Ratings were aggregated and interpreted as specified in Matthews et al. [36] to provide ratings of hedonic tone ('happiness'), tense arousal ('mood') and energetic arousal ('alertness').

2.3.7 Workload Ratings:
The NASA Task Load Index (TLX) rating scale [37] was used to capture workload ratings before and after each drive. Ratings were aggregated

to provide Total Workload ('Raw-TLX'), in line with common practice [37].

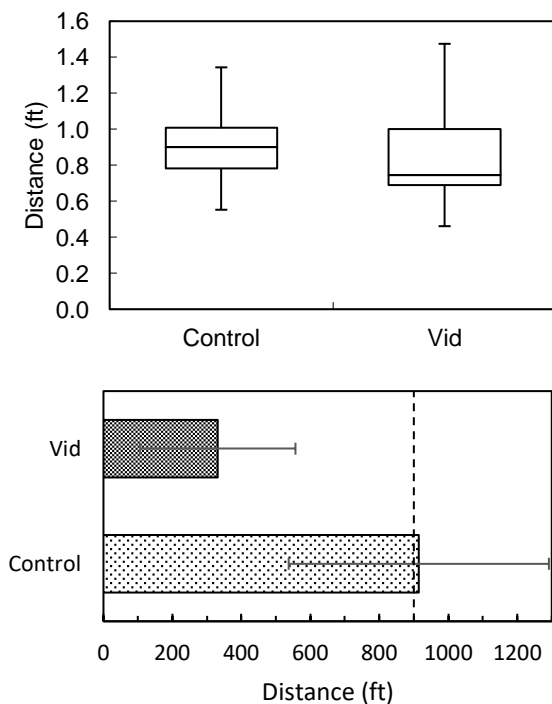*2.3.8 Post-Study Feedback:* General feedback was captured upon completion of both drives.

3. Results

Paired-samples t-tests were conducted to compare mean values between conditions ('Control' and 'Vid'), with the null hypothesis that there were no differences between conditions (i.e. the intervention had no effect on driver fatigue).

3.1. Driving Performance

3.1.1 Lateral Control (SDLP): There was a significant difference between conditions ($t(18)=3.24$, $p = .005$), with drivers demonstrating significantly more variability in lane keeping during the Control drive (M=0.93ft, SD=.28), compared to the Vid drive (M=.814ft, SD=.259) (Figure 2).

3.1.2 Longitudinal Control (Standard Deviation of Vehicle Speed and Headway): There were no significant differences in longitudinal vehicle control, with standard deviation of speed similar between conditions ($p = .881$). Differences in the standard deviation of headway were approaching significance ($p = .101$) with longer headway evident during Vid drives (M=267.8,ft SD=563.0) compared to Control drives (M=226.6ft, SD=497.0).

3.1.3 Response to Hazard: Figure 2 shows mean values (with standard error bars) for the distance at which drivers responded to a potential hazard situation (i.e. lifted their foot from the accelerator pedal in anticipation of braking). Given that this only occurred during drive 2, and that drives were counterbalanced, results are effectively 'within-subjects' (i.e. with n=11). As such, any statistical analysis in severely compromised, and therefore raw data only are reported, to provide an indication of behaviour. These suggest that drivers responded much sooner (559 feet or 170m) before the hazard during the Vid drive. In contrast, participants had already passed the pedestrian and parked car before lifting their foot from the accelerator pedal during the Control drive (Figure 2).

3.2. Physiological Measures

3.2.1 Heart-Rate Variability: There were no significant differences in HRV between the Control (M=0.837, SD=.101 and M=0.068, SD=.020, respectively) and Vid (M=0.843, SD=.107 and M=0.067, SD=.021, respectively) drives ($p = .518$ and .631, respectively).

3.3. Visual Behaviour

3.3.1 Blinks: There was no significant difference in the number of blinks during each drive between Control (M=1097.9, SD=556.4) and Vid (M=1107.3, SD=424.6) drives ($p = .918$). In addition, there was no significant difference in mean duration of blinks between Control (M=397.4, SD=61.5) and Vid (M=383.5, SD=42.4) drives ($p = .124$).

Fig. 2. *Driving performance measures: Boxplot showing standard deviation of lateral lane position (top), and mean distance from hazard (@900ft) that drivers responded (i.e. lifted their foot from the accelerator pedal), with standard error bars (bottom).*
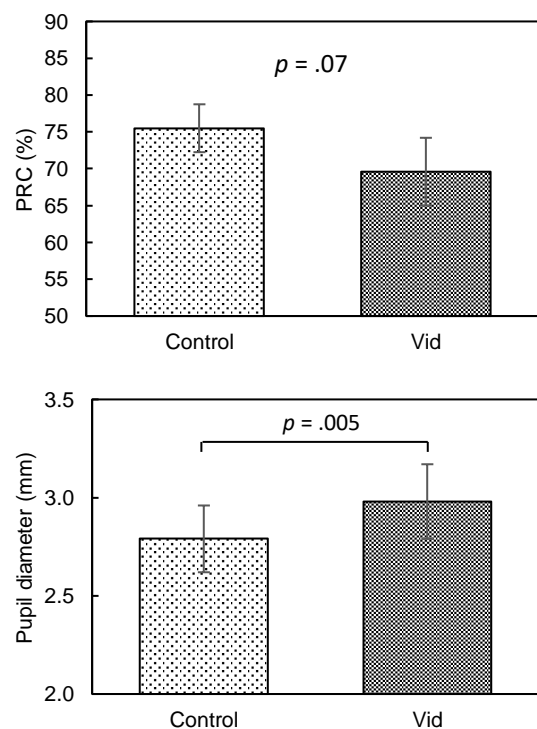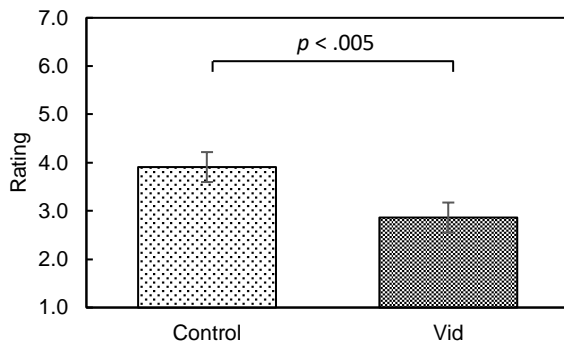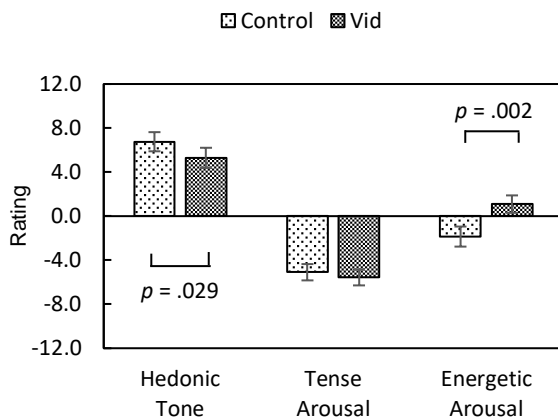
Fig. 3. *Visual behaviour: Mean percent road centre (top), and mean pupil diameter (bottom), with standard error bars.*

5

Fig. 4. *Mean sleepiness ratings, using SSS [21] with standard error bars. Note: scale ranges from 1 to 7, where 1.0 = 'feeling active, alert or wide awake' and 7.0 = 'no longer fighting sleep'.*

Fig. 6. *Workload Ratings: Mean NASA-TLX 'total workload' ratings showing constituent subscales, with standard error bars. Note: maximum possible value is 240.*

Fig. 5. *Mood assessment: Mean UWIST mood adjective assessment ratings [33], with standard error bars. Note: possible values ranges from -12 to +12.*
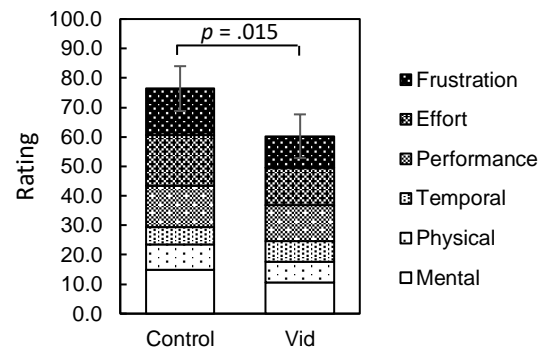
compared to Vid (M=2.86, SD=1.46) (Figure 4). Mean values suggest that drivers were approaching "somewhat foggy" (4 rating) after the Control drive, but "awake, but relaxed" (3 rating) after driving with Vid [21].

3.4.2 Comments: Written responses support the ratings, with participants stating reasons, such as: *"fatigue from driving, time of day, monotonous journey"*, *"tiredness from driving"*, *"consistent and monotonous driving"* and *"lack of stimulus and low task load"* cited after the Control drive. In contrast, lower ratings of sleepiness after the Vid drive were supported by comments, such as: *"interaction with Vid prevented boredom"*, *"interaction with [Vid] ... made me feel more active"*, *"Vid helped keep me from losing concentration through tiredness/boredom"* and *"Vid kept me fully alert during [the] drive"*.

3.5. Mood Assessment

There were significant differences in 'hedonic tone' (happiness) and 'energetic arousal' (alertness) between the Control (M=6.76, SD=3.97 and M=-1.857, SD=4.15, respectively) and Vid drives (M=5.29, SD=4.23 and M=1.095, SD=3.62, respectively) (t(20)=2.353, p = .029 and t(20)=3.623, p = .002, respectively). Drivers indicated higher levels of 'alertness', but lower levels of 'happiness' associated with the Vid drive. There was no significant difference in 'mood' (tense arousal) between driving conditions (p = .557) (Figure 5).

3.6. Driver Workload

Total Workload was determined by amalgamating 'raw' (un-weighted) ratings for subscales. There was a significant difference in Total Workload between drives (t(22)=2.645, p = .015), with participants associating lower workload with the Vid drive (M=60.22, SD=36.4), compared to the Control drive (M = 76.39, SD=37.3). Significant differences were also evident in Mental (t(22)=2.855, p = .009) and Effort (t(22)=1.885, p = .073) subscales (Figure 6), with participants indicating higher Mental workload and Effort during the Control drive.

3.3.2 Fixations: There was no significant difference in number of fixations during each drive between Control (M=4020.8, SD=694.8) and Vid (M=4175.0, SD=839.6) drives (p = .434), and also no significant difference in the mean duration of fixations between Control (M=255.9, SD=71.5) and Vid (M=249.7, SD=69.0) drives (p = .656).

3.3.3 Spread of Visual Attention (PRC): There was a near significant difference in PRC between Control (M=75.5%, SD=12.2) and Vid (M=69.6%, SD=21.6) drives (t(21)=1.91, p = .07), with participants spending more time focussed on the 'road centre' during the Control drive, suggesting 'cognitive tunnelling' (Figure 3).

3.3.4 Pupil Diameter: There was a significant difference in pupil size between Control (M=2.79mm, SD=.89) and Vid (M=2.98mm, SD=.96) drives (t(25)=3.046, p = .005). Pupil diameter was larger during the Vid drive (Figure 3).

3.4. Subjective Sleepiness

3.4.1 SSS: There was a significant difference in sleepiness ratings following the Control and Vid drives (t(21)=5.161, p < .005), with participants indicating higher levels of sleepiness after the Control drive (M=3.91, SD=1.48)

6

3.7. Post-Study Feedback

After both drives had been completed, participants were asked if the Vid system helped them remain alert while driving. Twenty participants responded positively, indicating that the system provided *"something (someone) else to communicate with"* and *"a sense of company on the journey"*. In addition, advocates of the system said that it *"alleviated boredom"*, *"broke monotony"* and kept them occupied *"in the same way that having a passenger converse with you keeps your mind occupied"*. Conversation was highlighted in particular as a positive attribute of the system: *"talking heightened my alertness"*, *"by talking about news etc., it helped to remain alert to driving"*. In contrast, the three participants who said that Vid did not help them remain alert, stated that *"the constant interruptions are annoying"*, it was no more effective *"than a radio would have [been]"*, and *"[it] didn't affect me much"*. It was noteworthy that when asked what they would do on a normal journey to remain alert, conversation (e.g. *"talk to passengers"*, *"call someone for company"*) was highlighted as the most popular operational strategy amongst participants.

4. Discussion

There is substantial evidence to support the use of conversation as a potential strategy to keep drivers alert and minimise the effects of fatigue [27, 28, 29]. A novel aspect of the current investigation was to utilise a digital assistant ('Vid') as the second interlocutor. Results obtained from a broad range of measures suggest that conducting a conversation with the digital assistant had a positive effect on driver arousal during an extended, monotonous journey, compared to the same journey undertaken without accompaniment ('Control'). In particular, driving with Vid encouraged higher self-reported levels of alertness ('energetic arousal') [36] and lower levels of subjective sleepiness [21]. In addition, participants associated lower workload [37] with the Vid drive – this can be interpreted as less perceived effort required to 'keep awake', alert and focussed on driving: a conclusion that is also supported by comments received during the post-study feedback. In practice, one might suspect that engaging in conversation could elevate workload (compared to a situation where no dialogue took place). The aim of the approach explored here is to ensure that an 'appropriate' level of workload is achieved (through conversation with the digital assistant), to enable drivers to remain 'engaged' with the driving task, but avoid situations of 'underload' [38], in particular.

Objectively, driving performance was superior throughout the drive in which participants interacted with Vid, evidenced by better lane-keeping. In contrast, this metric suggested that driving performance was 'impaired' during the Control drive [39]. In addition, engagement with Vid enabled drivers to respond more effectively (i.e. sooner) to a potential hazard situation, whereas participants had notably passed the hazard (on average) before effecting the same response during the Control drive. Physical indicators captured during the Vid drive, such as increased pupil diameter, and a trend towards an increased spread of attention to the road-scene (i.e. fewer fixations were concentrated on the road-centre indicating a lower incidence

of 'cognitive tunnelling'), also indicate that drivers were more alert when they were engaged in conversation with the digital assistant.

Despite this evidence, some of the measures collected during the study revealed no differences between conditions. For example, heart-rate variability (HRV) was comparable between drives, and the number and mean duration of blinks and fixations were inconclusive. Such measures might normally be expected to differentiate between different levels of fatigue. Failure to elicit differences in these measures, during the current study, may indicate that the experimental conditions were insufficient to elicit the full symptoms of fatigue. Alternatively, given that the effects of fatigue are cumulative, the analysis approach (i.e. taking mean values across the entire drive) may have masked differences for these measures. It is also feasible that factors inherent to the driving simulator may have influenced the efficacy of these measures. For example, high luminance levels (associated with the projected scene) may have effected pupil dilation [40], and lower levels of perceived risk associated with the simulation experience, may have influenced HRV across both drives.

Overall, results appear to suggest that engaging in conversation with a digital assistant increased driver arousal. Nevertheless, it is possible that drivers' behaviour may have been influenced by the novelty of the situation, and this may have encouraged them to 'test' the system. Indeed, one participant commented during the post-study feedback: *"the anticipation of hearing what the Vid would do made me alert. [I was] almost excited to 'play' with its capabilities"*. It is also interesting to note that participants indicated lower levels of 'happiness' (hedonic tone) [36] associated with the Vid drive. This suggests that although conversing with Vid may have had a positive impact on alertness, further work is required to refine the nature of the conversation to ensure that it is also a 'pleasurable' experience.

A further word of caution concerns the nature of fatigue under investigation. Given the low demands and monotony of the driving scenario, drivers were expected to exhibit symptoms of *passive* task-related fatigue. While the evidence suggests that engaging in conversation can increase driver alertness in this situation (i.e. when drivers were experiencing cognitive underload), it is feasible that extensive or complex interactions with a digital assistant could actually increase workload for drivers, particularly if they are already highly-loaded (e.g. when negotiating high traffic density, complex road networks etc.). This is likely to *actively* fatigue drivers and could impair performance as a result. Consequently, interventions, such as engaging drivers in conversation, must be bound by the current state of the driver. Driver-state monitoring systems (expected to adorn next generation vehicles) could provide a viable method to determine driver workload prior to initiating conversation (as indeed a human passenger would be likely to do).

It is also worth reflecting on the Wizard-of-Oz methodology employed during the study. This highly novel approach provided an excellent platform to observe and comment upon the behaviour of drivers engaged with a digital assistant offering highly-capable, natural-language abilities. The challenge of using this approach in the current context was to ensure that the system conformed with participants' expectations. This meant offering advanced natural language capabilities that notably exceeded current

technology, yet still conformed with participants' expectations of how future talking technology might perform. In addition, it was important to avoid potential 'uncanny valley' effects (the sense of eeriness and feelings of revulsion that can accompany technology that appears too human) [41]. Consequently, Vid's utterances were delivered in a slightly stilted ('computerised') manner, but conformed with formal English language grammar and syntax, and all exchanges assumed mutual understanding between interactants (as you would expect during human-human conversation). Consequently, common linguistic techniques, such as back-channelling, fillers, vague language, hedged requests etc. [34] could be utilised and understood (by both parties). While the responses and capability of Vid may therefore not be truly representative of future talking technology, or indeed, designers' aspirations for such, it was a necessary approach to create a 'plausible' digital assistant. Utilising a Wizard-of-Oz method ensured that it was possible to successfully create and maintain this illusion – it is worth noting that all participants expressed genuine surprise when they were informed that they had been interacting with another human-being rather than the technology as described, following completion of the study. Naturally, a short-coming of this approach (as indeed, with any 'controlled' conversation) is that it was difficult to maintain the same balance of talking/listening between participants, so the overall experience may have differed somewhat.

As a final cautionary note, it is worth highlighting that operational countermeasures (such as conversation), do not necessarily address the underlying physiological causes of fatigue, but rather aim to enhance alertness and performance temporarily so that operational safety and performance are maintained [27]. Moreover, individual differences – such as age, individual susceptibility to sleep loss, and lifestyle factors such as new parenthood, personality and mood – can also play a major part in susceptibility to fatigue [1]. While the current evidence suggests that conversation with a digital assistant may act as an effective countermeasure to the onset and effects of passive task-related fatigue in some situations, it is important for drivers to understand the need to individually assess their own fatigue, both before and during driving.

5. Conclusion

Using a Wizard-of-Oz approach, we explored the effect of engaging drivers in conversation with a digital assistant ('Vid') on driver fatigue by comparing performance with a Control drive, in which no digital assistant was present. Results suggest that communicating with Vid had a positive effect on driver arousal and performance, evidenced by better lane-keeping (lower SDLP) and quicker responses to a potential hazard situation. Participants also reported higher levels of alertness, and lower sleepiness and workload, associated with the Vid drive. These ratings were supported by physical indicators, such as increased pupil diameter, and a trend towards a lower percentage of visual attention (PRC) directed towards the road centre, suggesting that drivers were more alert when they were also engaged in conversation with the digital assistant. Based on these findings, we therefore conclude that engaging in conversation with a digital assistant may provide a viable method to combat the onset and effects of passive task-related fatigue, and hence, improve road-safety. Nevertheless, some metrics (particularly physiological measures) were inconclusive, and it is also worth noting that participants indicated lower levels of 'happiness' associated with the Vid drive. Further work is therefore required to refine the nature of the conversation to ensure that it is also a 'pleasurable' experience. Future investigations could also consider longer episodes of driving, and potentially sleep-deprive participants prior to testing, to ensure that they are truly fatigued. In addition, the effect of conversation on active fatigue (i.e. considering drivers who are negotiating more demanding driving situations) could be explored.

7. References

[1] P. Jackson, C. Hilditch, A. Holmes, N. Reed, N. Merat and L. Smith, "Fatigue and road safety: a critical analysis of recent evidence," London: Department for Transport, 2011.

[2] K. Dobbie, "Fatigue-related Crashes: An Analysis of Fatigue-related Crashes on Australian Roads using an Operational Definition of Fatigue," 2002.

[3] D. Dinges, "An overview of sleepiness and accidents," *Journal of Sleep Research,* vol. 4, no. 2, pp. 4-14, 1995.

[4] E. Grandjean, "Fatigue in industry.," *British Journal of Industrial Medicine,* vol. 36, pp. 175-186, 1979.

[5] J. R. Stradling and R. J. O. Davies, "Sleep 1: Obstructive sleep apnoea/hypopnoea syndrome: definitions, epidemiology and natural history," *Thorax,* vol. 59, pp. 73-78, 2004.

[6] J. A. Shekleton, N. L. Rogers and S. M. Rajaratnam, "Searching for the daytime impairments of primary insomnia," *Sleep Medicine Reviews,* vol. 14, no. 1, pp. 47-60, 2010.

[7] L. Torsvall and T. Åkerstedt, "Sleepiness on the job: continuously measured EEG changes in train drivers," *Electroencephalography and clinical Neurophysiology,* vol. 66, no. 6, pp. 502-511, 1987.

[8] J. A. Horne and L. A. Reyner, "Sleep related vehicle accidents," *British Medical Journal,* vol. 310, pp. 565-567, 1995.

[9] S. K. Lal and A. Craig, "A critical review of the psychophysiology of driver fatigue," *Biological psychology,* vol. 55, no. 3, pp. 173-194, 2001.

[10] P. Desmond and P. Hancock, "Active and passive fatigue states.," in *Stress, workload, and fatigue*, Mahwah, NJ, Erlbaum, 2001, pp. 455-465.

[11] I. D. Brown, "Driver Fatigue," *Human Factors: The*

Journal of the Human Factors and Ergonomics Society, vol. 36, no. 2, pp. 298-314, 1994.

[12] S. Otmani, T. Pebayle, J. Roge and A. Muzet, "Effect of driving duration and partial sleep deprivation on subsequent alertness and performance of car drivers," *Physiology and Behavior,* vol. 84, no. 5, pp. 715-724, 2005.

[13] S. N. Biggs, A. Smith, J. Dorrian, K. Reid, D. Dawson, v. d. H. C. and S. Baulk, "Perception of simulated driving performance after sleep restriction and caffeine," *Journal of Psychosomatic Research,* vol. 63, no. 6, pp. 573-577, 2007.

[14] S. D. Baulk, S. N. Biggs, K. J. Reid, C. J. van den Heuvel and D. Dawson, "Chasing the silver bullet: measuring driver fatigue using simple and complex tasks," *Accident Analysis and Prevention,* vol. 40, no. 1, pp. 396-402, 2008.

[15] P. Thiffault and J. Bergeron, "Monotony of road environment and driver fatigue: a simulator study," *Accident Analysis and Prevention,* vol. 35, no. 3, pp. 381-391, 2003.

[16] N. B. Powell, K. B. Schechtman, R. W. Riley, K. Li, R. Troell and C. Guilleminault, "The road to danger: the comparative risks of driving while sleepy," *Laryngoscope,* vol. 111, no. 5, pp. 887-893, 2001.

[17] K. Brookhuis and D. de Waard, "Monitoring drivers' mental workload in driving simulators using physiological measures," *Accident Analysis & Prevention,* vol. 42, no. 3, pp. 898-903, 2010.

[18] W. Wierwille and L. Ellsworth, "Evaluation of driver drowsiness by trained raters," *Accident Analysis & Prevention,* vol. 26, no. 5, pp. 571-581, 1994.

[19] T. W. Victor, J. L. Harbluk and J. A. Engström, "Sensitivity of eye-movement measures to in-vehicle task difficulty," *Transportation Research Part F: Traffic Psychology and Behaviour,* vol. 8, no. 2, pp. 167-190, 2005.

[20] M. W. Johns, "A new method for measuring daytime sleepiness: the Epworth sleepiness scale," *Sleep,* vol. 14, no. 6, pp. 540-545, 1991.

[21] E. Hoddes, V. Zarcone, H. Smythe, R. Phillips and W. C. Dement, "Quantification of sleepiness: a new approach," *Psychophysiology,* vol. 10, no. 4, pp. 431-436, 1973.

[22] G. Larue, A. Rakotonirainy and A. Pettitt, "Driving performance impairments due to hypovigilance on monotonous roads," *Accident Analysis & Prevention,* vol. 43, no. 6, pp. 2037-2046, 2011.

[23] J. Santamaria and K. Chiappa, "The EEG of drowsiness in normal adults," *Journal of Clinical Neurophysiology,* vol. 4, no. 4, pp. 327-382, 1987.

[24] N. Kleitman, Sleep and wakefulness, University of Chicago Press, 1963.

[25] Department for Transport, "Wake Up to the Signs of Tiredness. THINK! leaflet," 2005. [Online]. Available: http://webarchive.nationalarchives.gov.uk/200611010 04827/http://thinkroadsafety.gov.uk/advice/pdf/992.pd f. [Accessed 22 August 2016].

[26] N. Wright, B. Stone, T. Horberry and N. Reed, "A review of in-vehicle sleepiness detection devices," 2007.

[27] M. R. Rosekind, P. H. Gander, K. B. Gregory, R. M. Smith, D. L. Miller, R. Oyung, L. L. Webbon and J. M. Johnson, "Managing fatigue in operational settings 1: Physiological considerations and countermeasures," *Behavioral Medicine,* vol. 21, no. 4, pp. 157-165, 1996.

[28] J. A. Caldwell, M. M. Mallis, J. L. Caldwell, M. A. Paul, J. C. Miller and D. F. Neri, "Fatigue countermeasures in aviation," *Aviation, space, and environmental medicine,* vol. 80, no. 1, pp. 29-59, 2009.

[29] M. Dijkman, N. Sachs, E. Levine, M. Mallis, M. Carlin, K. Gillen, J. Powell, S. Samuel, J. Mullington, M. Rosekind and D. Dinges, "Effects of reduced stimulation on neurobehavioral alertness depend on circadian phase during human sleep deprivation," *Sleep Res,* vol. 26, p. 265, 1997.

[30] P. Cabon, J. Fouillot, A. Coblentz, R. Mollard and J. Speyer, "Etude du niveau d'éveil des équipages au cours de vols long-courriers," *Medicine Aéronautique et Spatiale,* vol. 30, pp. 321-338, 1991.

[31] C. Nass and S. Brave, Wired for speech: How voice activates and advances the human-computer relationship, Cambridge: MIT press, 2005.

[32] P. H. Ting, J. R. Hwang, J. L. Doong and M. C. Jeng, "Driver fatigue and highway driving: A simulator study," *Physiology & Behavior,* vol. 94, no. 3, pp. 448-453, 2008.

[33] D. Large, G. Burnett, B. Anyasodo and L. Skrypchuk, "Assessing Cognitive Demand during Natural Language Interactions with a Digital Driving Assistant," in *The 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutoUI2016)*, Ann Arbor, Michigan, 2016.

[34] D. Large, L. Clark, A. Quandt, G. Burnett and L. Skrypchuk, "Steering the conversation: A linguistic exploration of natural language interactions with an autonomous digital driving assistant," *Applied Ergonomics,* vol. 63, pp. 53-61, 2017.

[35] L. Mulder, D. De Waard and K. Brookhuis, "Estimating mental effort using heart rate and heart rate," in *Handbook of Ergonomics and Human Factors Methods*, London, Taylor &, 2004, pp. 201-208.

[36] G. Matthews, D. Jones and A. Chamberlain, "Refining the measurement of mood: The UWIST mood adjective checklist," *British journal of psychology,* vol. 81, no. 1, pp. 17-42, 1990.

[37] S. Hart and L. Staveland, "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research," *Advances in psychology,* vol. 52, pp. 139-183, 1988.

[38] M. Young and N. Stanton, "Malleable attentional resources theory: a new explanation for the effects of mental underload on performance," *Human factors,*

vol. 44, no. 3, pp. 365-75, 2002.

[39] K. Brookhuis, D. Waard and S. Fairclough, "Criteria for driver impairment," *Ergonomics,* vol. 46, no. 5, pp. 433-445, 2003.

[40] V. Clark and J. Kruse, "Clinical methods: the history, physical, and laboratory examinations," *Jama,* vol. 264, no. 21, pp. 2808-2809, 1990.

[41] M. Mori, "The Uncanny Valley," *Energy,* vol. 7, no. 4, pp. 33-35, 1970.