

To Please in a Pod: Employing an Anthropomorphic Agent-Interlocutor to Enhance Trust and User Experience in an Autonomous, Self-Driving Vehicle

David R. Large, Kyle Harrington,
Gary Burnett

Human Factors Research Group,
University of Nottingham
Nottingham, UK
{ david.r.large; kyle.harrington; gary.burnett }
@nottingham.ac.uk

Jacob Luton, Peter Thomas,
Pete Bennett

Jaguar Land Rover Research
Coventry, UK
jakeluton@hotmail.com;
{ pthoma95; pbennet7 }
@jaguarlandrover.com

ABSTRACT

Recognising that one of the aims of conversation is to build, maintain and strengthen positive relationships with others, the study explores whether passengers in an autonomous vehicle display similar behaviour during transactions with an on-board conversational agent-interface; moreover, whether related attributes (e.g. trust) transcend to the vehicle itself. Employing a counterbalanced, within-subjects design, thirty-four participants were transported in a self-driving pod using an expansive testing arena. Participants undertook three journeys with an anthropomorphic agent-interlocutor (via Wizard-of-Oz), a voice-command interface, or a traditional touch-surface; each delivered equivalent task-related information. Results show that the agent-interlocutor was the most preferred interface, attracting the highest ratings of trust, and significantly enhancing the pleasure and sense of control over the journey experience, despite the inclusion of ‘trust challenges’ as part of the design. The findings can help support the design and development of in-vehicle agent-based voice interfaces to enhance trust and user experience in autonomous cars.

Author Keywords

trust, user experience, conversational user interface, anthropomorphism, autonomous driving, pod

CSS Concepts

• **Human-centered computing~Natural language interfaces** • Human-centered computing~User studies

INTRODUCTION

Autonomous, self-driving vehicles are expected to revolutionise everyday travel with anticipated benefits of

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

AutomotiveUI '19, September 21–25, 2019, Utrecht, Netherlands

© 2019 Association for Computing Machinery.

ACM ISBN 978-1-4503-6884-1/19/09...\$15.00

<https://doi.org/10.1145/3342197.3344545>

improved road safety, efficiency, comfort and mobility. First experiences are likely to be in driverless ‘pods’ that operate in contained, ‘geo-fenced’ environments (university campuses, airports etc.) [1], with several examples already in existence. Nevertheless, major concerns have been expressed regarding the public’s willingness to adopt the technology [2], in particular, relating to issues of trust and the overall user experience [3, 4].

Trust and Driver Acceptance

Trust in technology is considered to be the extent to which people believe that technology will perform effectively and without a negative or injurious outcome [5]. Trust therefore shapes an individual’s attitudes and ultimately determines their behaviour, such as their intention to use the system [6, 7], the extent to which they rely upon the technology and their operational strategies towards its use [8, 9]. Intertwined with trust is the concept of acceptance. Acceptance has been defined in the automotive domain as “*the degree to which an individual incorporates the system in his/her driving, or if the system is not available, intends to use it*” [10] (p.18). The determinants of drivers’ trust and acceptance are thus complex and interrelated, and derive from various factors, including the individual’s understanding of the system limits and the context in which it is implemented [11].

Various theoretical models have been proposed to define and evaluate acceptance (for example, the Technology Acceptance Model, TAM) [11]. While these were originally developed in the information technology domain, they have been widely adapted for other contexts, such as driving, with most models now incorporating additional factors. In the context of autonomous driving, relevant factors include: the degree to which users can predict and understand the operation of the vehicle (system transparency) and the degree of user perception on the performance of the vehicle or technological component (technical competence) [12]. In addition, factors such as reliability and dependability can affect trust and acceptance in relation to automated vehicle technologies: reliability is defined as the ability of a device or system to perform a required function under stated conditions for a specified period, whereas dependability

refers to the frequency of automation breakdowns or errors [13]. Errors (such as false alarms) can therefore have a profound effect on users' perception of reliability and dependability. For example, a reduction in the occurrence of errors notably increases drivers' perception of the reliability of the system and positively impacts trust development [14].

Privacy and security concerns have also been highlighted as potential barriers to the trust and acceptance of autonomous vehicles. *Privacy* is linked to the handling of personal data, for example, ensuring that the user knows which data are being collected and how they will be used [15], and typically centres around two aspects: unauthorized access due to security breaches or the lack of internal controls, and the risk of secondary use, that is, the re-use of personal data for unrelated purposes without the user's consent [16]. In contrast, *security* refers to the technical guarantees that ensure measures against threat of intentional attack on systems/software, and the legal requirements and good practices with regard to privacy, are met [15].

A common ground in trust and acceptance research is that human behaviour is not determined by objective factors, but rather by the user's subjective perceptions, based on their individual attitudes, expectations and experience [7]. Thus, even a well-designed system that evidently performs effectively and without inflicting a negative or injurious outcome, may not necessarily warrant a user's trust or acceptance. As such, we look for guidance to social psychology (from which our understanding and operationalisation of trust-in-technology originate). Here, trust is defined as: "*a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behaviour of another*" [17] (p.395). In other words, trust is a belief by a person in the integrity of another, and centres around 'human' factors such as benevolence and honesty [18]. In practice, this means that humans calibrate trust in another by making attributions based on personal qualities and characteristics, often identified through speech and conversation.

Speech, Conversation and Talking Technology

Philosophical debates identify speech as one of the quintessential markers of humanness [19]. It is the primary means of social identification amongst humans, and implicates more parts of the brain than any other function [20]. Speech is peppered with salient, socially-relevant, cues, above and beyond the lexical content, that humans quickly become experts at extracting and comprehending based on vocal characteristics such as pitch, cadence, speech rate and volume: these are subsequently used to provide systematic guidance for determining gender, personality and emotion-specific actions, such as who to like and trust [21, 19].

When formed as conversation ("*any interactive spoken exchange between two or more people*"), speech serves many purposes, which can be broadly categorised as either transactional (task-based) or social (interactional) [22]. Although these may overlap within natural conversation

[23], transactional conversations pursue a practical goal, whereas the more social features of conversation aim to build, maintain and strengthen positive relationships with the other interlocutor(s) [24, 25]. Social conversation therefore includes aspects such as greetings and small talk that can help develop common ground [26], trust and rapport between interlocutors [23].

A common proposition in HCI is that humans appear to lack the wherewithal to overcome instinctive behaviours, and interact with a talking computer in a similar manner to talking to another human, demonstrating humanlike behaviours and making similar attributions [19]. For example, different digital 'personalities', created by varying the vocal characteristics and language content of spoken language interfaces, have been shown to influence trust, performance, learning and even consumers' buying habits during research studies [19]. Similar effects have been noted in the automotive domain, with participants recognising unique 'personalities' associated with different voices employed to deliver navigational instructions, even though the content remained the same: this influenced their attitudes towards the navigational device, including how much they liked it, their preferences for use, and the level of trust that they associated with it [27]. In social robotics, conversational interaction has been highlighted as an important factor in ensuring long-term rapport building and use [28]. The accommodation of social conversations therefore appears to be a critical factor in developing trust in a social agent [29].

Anthropomorphism

Attributing human motivations, characteristics or behaviour to inanimate objects, such as talking technology (as observed in the aforementioned examples) – and building expectations on the basis of this – is evidence of *anthropomorphism* [30]. Although the term has been used pejoratively in science when novelty features (such as a human face) have been added to non-human entities, anthropomorphism is not simply the titivation of artefacts with superficial human features and characteristics, but rather the process of inductive inference whereby people are inspired to believe (at some level) that the artefact has capacity for rational thought (agency) and conscious feeling (experience) [31]. This is typically inspired by peoples' experience of these features and characteristics (voice, social behaviour etc.) in humans.

Indeed, passengers of an autonomous vehicle that was *anthropomorphised* (given a human name, gender and voice) rated their vehicle as having more humanlike cognitive capacities than those who occupied a vehicle with the same autonomous features but without the associated anthropomorphic cues [32]. Participants in the study also reported trusting their vehicle more, were more relaxed in an accident, and blamed their vehicle and related entities less for an accident caused by another driver [32]. In addition, increases in the level of trust in automation were noted following the addition of anthropomorphic features (voice

and gender) to an existing audio-visual Human-Machine Interface (HMI) during conditionally-automated driving [33]. In this simulator-based study, the speech interface explained what the vehicle was intending to do during handovers. The value of using anthropomorphism to improve system transparency (that is, what the vehicle is doing and why) was also highlighted by Miglani, Diels, and Terken [34] who predicted associated increases in the level of trust.

Other authors, notably Eriksson and Stanton [35] have postulated the importance of employing a conversational agent HMI (their so-called “chatty co-pilot”) – drawing upon the aviation literature – as an effective and natural means of providing calibrated trust to human users of a conditionally-automated vehicle. The suggestion here is that the chatty co-pilot encourages *appropriate* levels of trust relevant to the system capability and reliability, thereby ensuring the system is used (appropriately) and complacency effects are minimised.

Study Aims and Scope

Evidently, there is considerable research exploring the use of human characteristics – especially conversational speech – within automotive HMIs to leverage human qualities and capabilities and embody these within the technology. This has been inspired to a large extent by the recent proliferation and popularity of personal devices like Amazon Echo and Google Home. While the research has generally revealed positive effects associated with talking technology during simulated driving (to enhance the subjective evaluation of the vehicle, and improve users’ behaviour and performance), this study is the first to investigate how an anthropomorphised agent-interlocutor employing natural, conversational language could influence trust and the user experience in a genuinely autonomous ‘pod’ vehicle. It builds on the work of Antrobus et al. [36], who explored the use of a natural language interface to improve the trust and acceptance of an autonomous taxi in a driving simulator. It is worth highlighting at the outset that this is not an evaluation of technology that has been deliberately designed to hold emotional intelligence or influence users’ emotions. Instead, it is a controlled experiment motivated by the psychological construct of anthropomorphism, which is subsequently used as a theoretical basis to interpret the observed behaviour. The study therefore aims to expose natural, instinctive behaviours and opinions that were motivated by the presence of a voice, and the interaction style it affords, and it is therefore also highly relevant to current trends and research interest in conversational user interfaces.

METHOD

Use-Case and Script Development

Prior to undertaking the driving study itself, a series of focus groups were conducted to determine the content and delivery of the ‘anthropomorphic agent’ dialogue, and to explore any concerns that people may have when using such an interface in an autonomous vehicle. In light of the findings from the focus groups, and in consideration of the limitations of the

experimental setup, we decided to focus on three use-cases: entertainment (news, sporting headlines, music etc.), office/scheduling (managing and creating to-do lists, calendars and emails) and system notifications (requests for vehicle diagnostics etc.). For each of these use-cases, we devised a variety of predetermined opening gambits and responses in order to form the basis of the anthropomorphic agent’s conversational exchanges, and the style with which to deliver these.

Participants

A representative sample of experienced drivers were recruited to take part (n=34), comprising 17 male and 17 female participants. Ages ranged from 21–58 years with a mean age of 40, driving experience from 3.5–40 years with a mean of 20.6 years, and self-reported annual mileage from 6,000 to 60,000 with a mean of 15,691. All participants were employees of Jaguar Land Rover (primarily from non-technical roles, i.e. administrative/support).

Apparatus

The study was conducted in the Urban Development Lab (UDL) indoor testing environment in Coventry using a driverless pod supplied by the RDM Group (Figure 1). The pod operated fully-autonomously throughout the study. The area was presented as an urban scenario, with shop fronts etc. projected onto timber constructions and perimeter screens to emulate commercial premises etc. The layout enabled multiple routes to be followed. During the study, participants were recorded using a GoPro camera for subsequent analysis, as well as a second camera for immediate streaming and observation. A small booth on the edge of the testing area

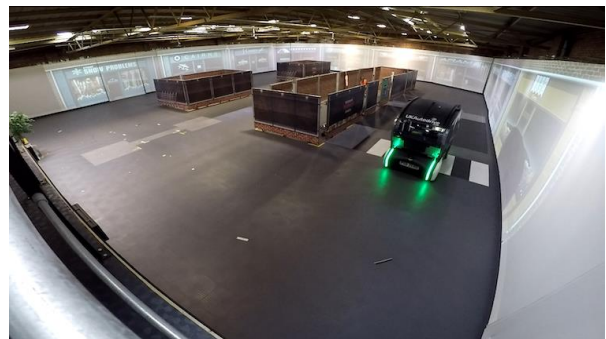


Figure 1. Arena view of pod and testing environment

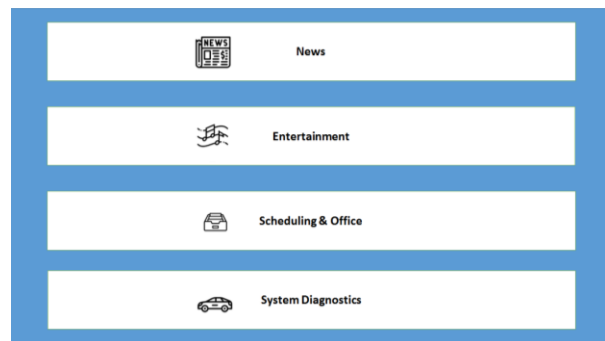


Figure 2. Touchscreen main menu

housed a professional actor, who delivered the anthropomorphic agent dialogue in real time, for example in response to conversation initiated by participants, using a Wizard-of-Oz approach [37]. Participants were told that they were interacting with a prototype highly-capable, conversational interface; they were not aware that they were conversing with another human. In addition, a Microsoft Surface touch-screen tablet with a bespoke PowerPoint presentation was installed in the vehicle, although this was not used with the conversational user interface.

Experimental Design

To ensure a thorough investigation, three interfaces were evaluated in a counterbalanced, within-subjects design: touchscreen, voice command and the anthropomorphic agent interlocutor. The voice command condition was included to explore whether any observed differences were as a result of the communication modality (voice versus touch) or due to the ‘anthropomorphic’ nature of the speech interface. Each interface provided equivalent task-related information relevant to the use-case under examination at set intervals throughout each drive, although the scope to engage in further interactions naturally differed between interfaces.

Touchscreen

A touchscreen interface was installed in the centre of the pod. Participants were able to interact with this using a bespoke PowerPoint presentation navigable via embedded hyperlinks, to appear as a fully-functioning, interactive HMI.

Voice Command (*‘AutoCab’*)

Participants were required to interact using specific voice commands relevant to each use-case. System responses were delivered in real-time by the actor using a Wizard-of-Oz approach. However, the actor was restricted in the range of responses available, and was instructed only to respond to correctly formatted commands (as would be expected in commercially-available voice-command technology). To aid participants, the range of usable commands was provided as a static, visual prompt on the touchscreen, which also remained present during this drive.

Anthropomorphic Agent-Interlocutor (*‘UltraCab’*)

Participants were able to interact with the anthropomorphic agent using free-flowing, conversational language. System responses were conversational in nature, based on the script informed by the focus groups, and also included more personable qualities (use of the first-person pronoun, participant’s name, politeness etc.) in line with findings from both the focus groups and previous work [38, 32]. Responses were composed and delivered in real-time by the actor using the Wizard-of-Oz approach.

Procedure

Following an initial briefing and the collection of demographic and background data (including anthropomorphic tendency [39]), participants completed three drives, with each drive relating exclusively to a different HMI. Each drive lasted approximately eight minutes and followed one of three different routes (HMIs

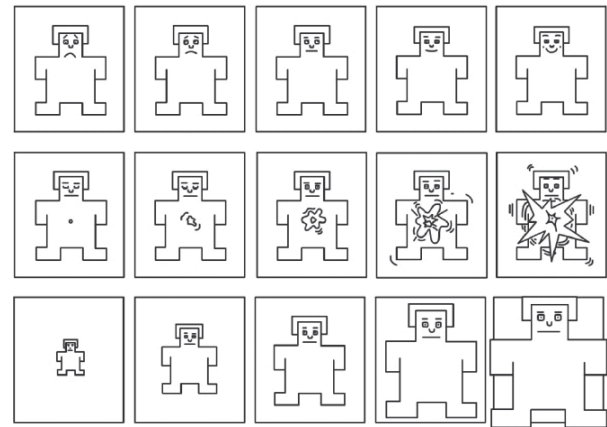


Figure 3. Self-Assessment Manikin [43], showing Pleasure (top), Arousal and Dominance scales (bottom)

and route-selection were counterbalanced between participants). To explore issues of reliability, participants were presented with a ‘trust challenge’ towards the end of each journey, inspired by the methodology employed by Antrobus, et al. [36]. For example, the pod abruptly stopped, claiming to have detected a pedestrian in the road ahead, and informed the participant that the journey would resume once the pedestrian had moved; in practice, the roadway ahead was clear. Security and privacy ‘challenges’ were also apparent throughout each journey in the form of requests to access participants’ personal email accounts and diaries etc.

Following each drive, participants completed the trust-in-automation rating scale [5]. To measure the user (affective) experience, they completed the Self-Assessment Manikin (SAM) [40]. The SAM is a non-verbal, pictorial assessment technique that operationalises user experience through the constructs of pleasure, arousal and dominance (Figure 3). Finally, participants rated the perceived anthropomorphism of each interface [41].

After experiencing journeys with all three HMIs, participants were invited to undertake a fourth drive using the HMI of their choice (later interpreted as an objective indication of preference): no ‘trust challenges’ were present during this ‘free-choice’ drive. In addition, participants ranked the three HMIs in terms of their overall preference. Finally, a structured, post-study interview was conducted to elucidate participants’ ratings. Each trial lasted approximately 1½ hours, including briefing and debriefing.

RESULTS

A repeated-measures ANOVA with post-hoc pairwise comparisons was conducted for each measure, unless otherwise stated. Post-study interviews were transcribed and analysed in a systematic qualitative interpretation using inductive thematic analysis [42]. In addition, conversational exchanges with UltraCab were transcribed and analysed, although this is reported elsewhere (see: [43] for details).

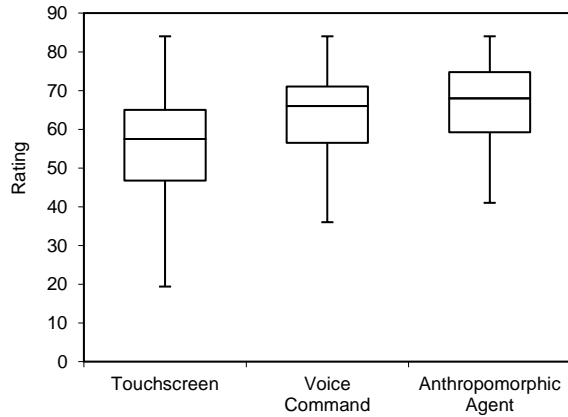


Figure 4. Trust in Technology ratings [7]

Trust-in-Automation

Responses to Jian et al.'s [5] trust in automation scale indicated significant differences in the level of Trust that participants placed in the different HMIs (Figure 4) ($F(2,66) = 11.80, p < .001, \eta_p^2 = .26$). Pairwise comparisons revealed that the trust placed in the anthropomorphic agent (mean: 67.1) was significantly higher than trust placed in the voice command (mean: 63.2, $p = .041$) and touchscreen interfaces (mean: 54.9, $p < .001$). Significant differences were also observed between the voice command and touchscreen ($p = .007$).

Self-Assessment Manikin: Pleasure

There were statistically significant differences between all three conditions for ratings of Pleasure ($F(3,66) = 33.3, p < .001, \eta_p^2 = .50$) (Figure 5). Pairwise comparisons revealed that mean ratings for Pleasure were highest (most positive affect) in response to the anthropomorphic agent (mean: 4.5), compared to the voice command (mean: 3.8, $p = .009$) and touchscreen interfaces (mean: 2.5, $p < .001$). Significant differences were also observed between the voice command and touchscreen ($p < .001$).

Self-Assessment Manikin: Arousal

There were also significant differences associated with the level of Arousal ($F(2,66) = 9.96, p < .001, \eta_p^2 = .23$). The touchscreen notably attracted the highest ratings (mean: 3.6), which were significantly higher than the voice command (mean: 3.0, $p = .005$) and the anthropomorphic agent (mean: 2.6, $p = .001$). Differences between the voice command and anthropomorphic agent were not significant ($p = .07$) (Figure 6).

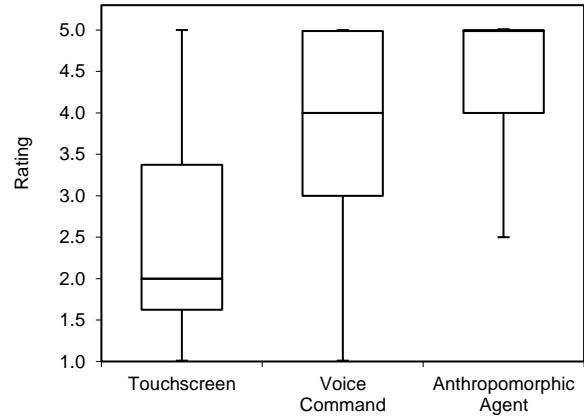


Figure 5. Ratings of Pleasure [43]

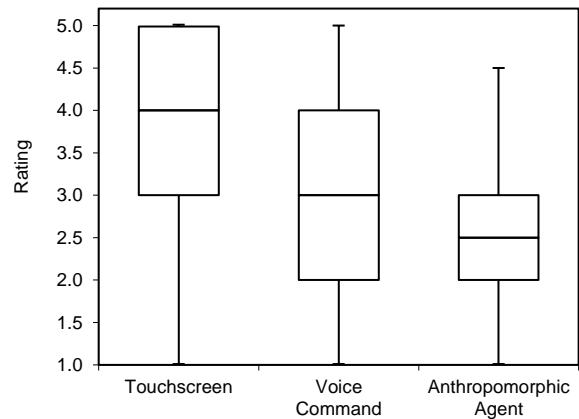


Figure 6. Ratings of Arousal [43]

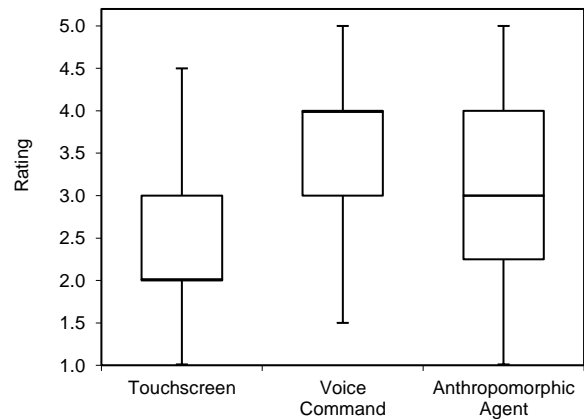


Figure 7. Ratings of Dominance [43]

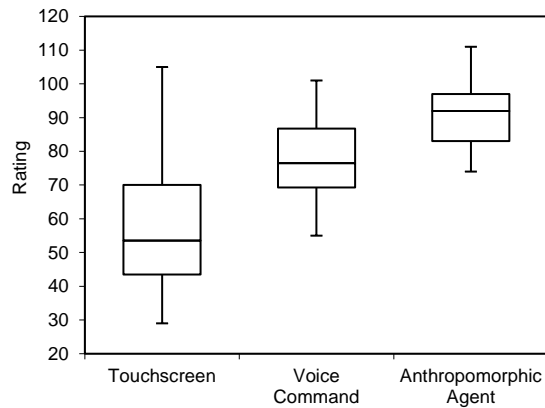


Figure 8. Ratings of Perceived Anthropomorphism [45]

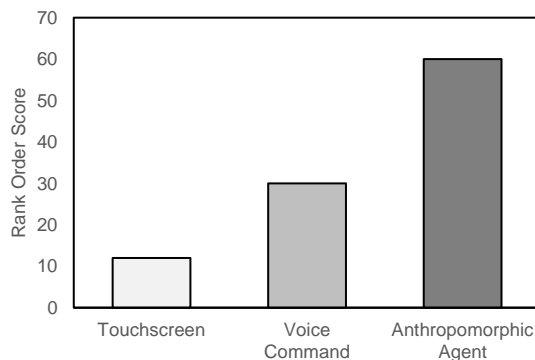


Figure 9. Pairwise Rank Scores

Self-Assessment Manikin: Dominance

Statistically significant differences were observed for ratings of Dominance ($F(2,66) = 12.7, p < .001, \eta_p^2 = .28$). While the voice command and anthropomorphic agent were comparable (with mean ratings of 3.4 and 3.3, respectively), these both attracted significantly higher ratings of Dominance than for the touchscreen (mean: 2.4, $p < .001$ and $p = .001$, respectively) (Figure 7).

Perceived Anthropomorphism of the Interfaces

Statistically significant differences were also noted for the Perceived Anthropomorphism of the interfaces ($F(3,66) = 56.2, p < .001, \eta_p^2 = .63$), with the anthropomorphic agent rated significantly higher than both the touchscreen and voice command interfaces (mean ratings: 91.4, 77.9 and 57.6, respectively, all $p < .001$) (Figure 8).

Preferences

Given the ability to choose, thirty participants (88%) chose the anthropomorphic agent to accompany them during their final, ‘free-choice’ drive. In addition, rankings indicated a strong preference for voice interfaces in general, and the anthropomorphic agent in particular. A pairwise rank analysis was conducted. This confirmed a clear overall preference for the anthropomorphic agent (Figure 9).

Post-Study Interviews: Thematic Analysis

The following six themes emerged during the analysis of the post-study interview transcripts. These are presented and discussed below.

1. System Transparency and Building Mental Models
2. Challenging Trust
3. Privacy and Security
4. On the Edge of the Uncanny Valley
5. Natural and Easy to Use
6. The Personable Touch

System Transparency and Building Mental Models. Many participants stated that they felt more in control of the interaction with the anthropomorphic agent (“*Much preferred [UltraCab], it made you feel in control*” P23), though it was also noted that the feelings of control did not necessarily transcend to the vehicle itself – one participant mentioned that they were unable to stop the vehicle by asking UltraCab, for example (this was in fact true for all interfaces). Those participants who felt more in control with the touchscreen or voice command interface commented that they understood how these operated (i.e. were able to quickly form an accurate mental model).

“I think once I’d figured out the touch screen a bit more I felt a bit more in control of that.” P20

Conversely, difficulty in forming a coherent mental model for the anthropomorphic agent (for example, how it arrived at each decision or formulated each response) was seen as a limiting factor by those participants who were more critical of this particular HMI. This is likely to be influenced by factors such as system transparency, which was also mentioned by some, and already highlighted as a potential barrier to trust and technology acceptance [11].

Challenging Trust. Trust was frequently mentioned during the post-study interviews, particularly in relation to the trust challenges. Although several participants recognised the need to develop trust over time (“*I think trust is gained by experience, so the more it happens the more trust is built up*” P27), the anthropomorphic agent appeared to attract inherently higher levels of initial trust, from the outset, even despite the erroneous declaration:

“I trust it, I trusted [UltraCab] from the first go round to be honest, it told me why it stopped” P21

As part of the experimental design, all HMIs provided the same explanation following the stoppages. It is therefore possible that participants were more reassured by the agent’s explanation because they felt that they could question it to gain clarification (given the dubious nature of the stoppage) – though there was little evidence within the conversational exchanges (also transcribed – see: [43]) that any participants actually sought further explanation or clarification. An alternate explanation is that an element of trust that had been established through the social aspects of the conversation [22]. Indeed, some participants felt that the agent exclusively warranted *blame* (in that it had contravened their shared

trust) – even though all HMIs were equally culpable: similar accusations were notably not directed at the other HMIs:

“...that made me not trust it [UltraCab], because I was getting an error message that didn't relate to the conditions, and it made it feel less reliable... It felt like this isn't working properly.” P30

“[UltraCab] was the most frustrating probably because your eyes were telling you something and the system was telling you another.” P20

Nevertheless, some participants were more forgiving of the technology generally ('better safe than sorry'), irrespective of the manner in which information was delivered:

“[other automotive systems] tend to pick up things that are not there, so sometimes you have to take them with not 100% accuracy...I'd rather it do that than run over a cat or a dog or something I can't see” P21

Privacy and Security. Similar to the findings of the earlier focus group, many participants expressed concerns about the privacy and security of their personal information. These included concerns about financial information, targeted advertising, knowledge of their home address and potential security breaches made by the autonomous vehicle: interestingly, such concerns were targeted mostly at the anthropomorphic agent, suggesting perhaps that participants felt it *knew them better*, given the conversational exchanges that had taken place. Participants also indicated that they would be reticent about sharing their personal information (through conversation) had they been travelling in a communal vehicle, compared to had it been their own personal transport:

“I have one concern and that's for sure, and that's the security aspects of it.” P19

“What's not clear to me is when you go in, what's it got access to.” P30

Even so, privacy and security of information were not necessarily seen as barriers, although participants indicated that they would have remained in control of which data they shared (and how they shared it):

“I would have programmed in things, like my emails...I would have programmed in my reminders and it will remind me of those things, but I'll still be very much in control of them.” P27

Participants therefore tended to concur that they viewed the role of UltraCab was as an *assistant*, meaning it could conduct tasks on their behalf. This suggests that participants wanted to maintain agency and responsibility over decision-making (and the sanctity of their data), but delegate routine task execution to the agent.

On the Edge of the Uncanny Valley. Several participants expressed concerns associated with the anthropomorphic agent attempting to *deceive* them into believing it was a

human, when it was not; moreover, this had a negative effect on users' perception of the development of trust:

“I think anything it did deliberately to build my trust would have the opposite effect. An attempt to pretend to be anything other than what it is, i.e. trying to be more human, is not necessarily a good thing in my view.” P32

This raises potential concerns associated with the agent being viewed as too human – the so-called *uncanny valley* effect. This can result in a sense of eeriness and suspicion due to a perceptual tension arising from mismatched stimuli and causing incongruence between users' expectations of a system and its actual capabilities [44]. This has already been recognised as an area of concern associated with intelligent conversational agents, with researchers proposing potential language-based solutions [45, 46].

Natural and Easy to Use. Several participants recognised the benefits of using natural language as an interaction mechanism, commenting that it was quick and intuitive, and that they did not need to learn a new HMI or technique:

“The conversation was better, I think because often you can't remember how you're supposed to say a command...I could interact just by talking, it felt a lot more natural seemed to work well.” P26

The benefits of using verbal interactions more generally (i.e. including the command-based interface) were also recognised:

“I thought the voice command was very good, it's much easier to interact and find information that you'd need whilst staying aware of what's going on around you.” P19

Nevertheless, some participants felt that conversational interactions with UltraCab required more effort. This may reflect the fact that using spoken conversation inspired participants to actively engage in the co-creation of common ground, trust and rapport (as they would with another human conversational partner) – rather than just the transactional (functional) aspects required by the interaction. Conversely, a single button press on the touchscreen, for example, was quick and easy and required no emotional attachment or commitment. Even so, participants generally agreed that the anthropomorphic agent provided the most 'natural' interaction. In particular, being able to express the same information in different ways, and the ability to build and maintain common ground and mutual understanding within the interaction (so that this could be referenced in later interactions) were actually seen as most useful; again, this reflects common practices employed during human-human conversation [26].

The Personable Touch. Participants frequently commented on the 'personable' aspects of the anthropomorphic agent (“I think it brought together a slightly personal touch, without sounding too automated.” P30), and recognised that these

were inspired by the use of conversational dialogue, specifically mentioning the use of social etiquette (politeness, apology, etc.). Comments suggest that participants believed this enhanced the perception of the agent as being ‘friendly’, ‘helpful’ and ‘intelligent’. Moreover, these qualities were seen to reinforce trust, comfort and even companionship:

“I trusted it more...when it prompted questions and was a bit friendlier in its answers...I instantly felt more comfortable.” P34

“you could ask questions, also felt less lonely, because you’re in there by yourself, you could relax” P23

“I think he is helpful and makes the journey pleasant, so it doesn’t feel so, like you are in a machine. So it feels more intelligent, you can trust it.” P31

Similar observations have also been made during studies in social robotics, in which people spoke openly about personal matters with a conversational robot, using it as an emotional outlet that reduced feelings of loneliness [47].

DISCUSSION

The study explored the efficacy of using an anthropomorphic agent employing conversational language to engender trust and enhance the user experience in an autonomous, self-driving ‘pod’ vehicle. The approach was motivated by our understanding of the role of conversation in human-human interactions, which is used in part to build trust and maintain and strengthen a positive relationship with the other interlocutor(s) [24, 25]. Overall, results show that the anthropomorphic agent-interlocutor was the most preferred interface. It invited the highest ratings of trust, and significantly increased the pleasure and sense of dominance (or control) over the journey experience. Ratings associated with ‘arousal’ were contrary to initial expectations, suggesting (on face value at least) that both the touchscreen and voice command interface were more ‘exciting’ and ‘engaging’ than the anthropomorphic agent. Nevertheless, while the pleasure and dominance scales have distinct positive and negative valences – ‘happy’ versus ‘sad’, and ‘under control’ versus ‘in control’, the semantic anchors utilised by the SAM [40] arousal scale lack unique associations. For example, the first picture (rating 1) shows an individual who is very calm (Figure 3). As such, this could be interpreted as ‘relaxed’, but ‘bored’ or even ‘lazy’ may be equally applied. In contrast, the last picture (rating 6) shows an individual who is literally ‘bursting’ with arousal. Thus, interpretations associated with a 6-rating could include extreme emotional states of ‘excitation’ and ‘euphoria’, but this could equally be interpreted as severe rage, agitation or anger. It is therefore feasible that the low ratings associated with the agent suggest a positive result in that participants were more relaxed when interacting using conversation, and conversely, that participants were highly agitated when interacting with the touchscreen. This is important in a

driving-related context, particular if drivers may be required to take control of the vehicle at some point.

As expected, participants associated higher levels of anthropomorphism with the agent. Indeed, the ‘personable’ aspects were a strong feature of the post-study interview, with participants positively reinforcing elements such as the use of social etiquette (politeness, apology, etc.). Participants also commented that they found the free-flowing conversational nature of the agent (as opposed to the strict voice-command approach) to be more ‘natural’, enabling them to simply say what they wanted, rather than having to interact in a predefined or command-based manner. It was also commented that using conversational exchanges made the interaction friendlier and more pleasant, and that the anthropomorphic agent could therefore potentially provide companionship on longer journeys which may otherwise feel isolating. Nevertheless, some participants notably commented that interacting with the natural language interface required *more* effort – this is thought to be due to the perceived additional effort to engage with elaborative, contextual social talk (to build common ground, trust and rapport) that is seen in human conversation [22, 23].

The significantly higher reported levels of trust associated with the anthropomorphic agent are likely to be a factor of the perceived humanness, or anthropomorphism, as also observed by Waytz et al. [32]. It is suggested that this helped to overcome potential reliability issues, such as the ‘trust challenges’, explored as part of the experiment. Nevertheless, it is also worth noting that these challenges attracted concerns about ‘deception’ more generally, and this was relevant for all interfaces. While this is perhaps unsurprising, given the nature of the perturbation, it is interesting to note that this generally had less of an impact on trust when using the anthropomorphic agent, suggesting that people are more likely to accept such fallibility from an entity that they perceive to be more humanlike – there was even evidence of participants *blaming* the agent.

Deception was also mentioned in the context of the anthropomorphic agent *pretending* to be human or appearing too humanlike – this was implicit in the design of the HMI and its loquacious style of interaction, and is an interesting irony, given the Wizard-of-Oz methodology employed here, whereby a human was actually impersonating the technology. However, it is also a word of warning to designers of conversational user interfaces that care must be taken to ensure that these do not descend into the ‘uncanny valley’ [44, 45], as this can result in deleterious effects on human perception and performance, and encourage inappropriate assertions of trust and reliance.

A potential criticism associated with the anthropomorphic agent (raised by several participants) was a lack of system transparency – for example, understanding how it arrived at each decision, calculated risk or formulated responses. This difficulty in establishing a coherent mental model for this HMI may have contributed to feelings of being out of control

by these participants (although, the agent notably attracted the highest ratings of control/dominance overall). Many participants also expressed concerns about the privacy and security of their personal information. These included concerns about financial information, targeted advertising, knowledge of their home address, and potential security breaches (or ‘hacking’) of the autonomous vehicle. Participants also indicated that they would be less willing to share information with the vehicle when undertaking a journey shared with other passengers. Even so, participants expressed a strong desire that the system learned their preferences, for example their tastes in music, and performed the functions of a personal organiser, thereby improving the overall user experience. Overall, participants were therefore not entirely unwilling to share information with the autonomous vehicle (to achieve this goal), although they did indicate that they would feel much more comfortable doing so if they had greater knowledge of how their data was being collected and used. This has clear implications for trust.

Limitations

Although the study revealed positive benefits associated with using an anthropomorphic agent-interlocutor to support passengers in an autonomous pod (compared to a touchscreen and voice-command interface), care should be taken when generalising from the findings. For example, while the pod did indeed operate autonomously, the experience was rather restrained, in that the pod travelled very slowly (at a fast walking pace), braking could be abrupt, and the physical design of the vehicle restricted participants’ vision. In addition, the overall journey experience was limited (there were no other vehicles present, for example) and participants could exert no control over the pod – they could not modify the route or stop the vehicle: this could potentially have influenced their subjective ratings of the experience – indeed, some participants indicated that their ratings of trust may have been different had the pod travelled more quickly and in the presence of other traffic.

In addition, although all three interfaces delivered the same task-related information for each of the use-cases, the anthropomorphic agent arguably offered greater interactivity, in that participants were able to engage in conversational dialogue to seek further clarification, request music tracks etc. While this is in fact a particular benefit of employing this type of HMI over the more constrained and task-oriented experiences offered by traditional touchscreen or voice command systems, the number of interactions and information exchanged during the study may therefore have differed between conditions, and this could have affected comparative ratings. Nevertheless, the decision and propensity to engage further with the agent was at the behest of each participant. As a consequence, individual differences (age, gender, personality etc.) and cultural differences may have influenced their propensity to engage; these factors could be explored in future works. In addition, recruiting a broader range of participants (i.e. not limited to employees of Jaguar Land Rover) would be beneficial in future work,

although this was a restriction imposed during the current study.

Finally, although the agent’s responses were guided by a script that incorporated appropriate language and phrasing (informed by the focus groups and previous studies), no restrictions were placed in terms of how much elaboration was possible. Moreover, our ‘wizard’ was instructed to respond to all enquiries, avoiding clinical, out-of-domain responses, such as, “*Sorry. I don’t understand*”. As such, the interface exceeded current state-of-the-art agent-interlocutors (such as Alexa, Siri and Google Home), which promise much through their implied humanness, but fall short of the reflexive and adaptive interactivity that occurs in most human-human conversation [48]. However, this was an important part of the experimental design to ensure that the results support the long-term development of human-agent interaction, rather than simply commenting on current limitations of the technology; even so, it may have discombobulated some participants.

CONCLUSION

While other scholars have explored the effects of using a speech-based interface during simulated driving, this is the first study to consider the impact of using an anthropomorphic agent (employing conversational speech) on the development of trust and the overall user experience in a fully-autonomous pod vehicle. Based on the results, we can conclude that using an anthropomorphic agent with two-way conversational interactions increases users’ perceived trust and pleasure; moreover, passengers felt more in control of the journey experience when accompanied by the agent. It was also evident that using anthropomorphism in the design of the agent created a more ‘forgiving’ experience (compared to other, more traditional interfaces), in which passengers were apparently more willing to accept reliability and dependability indiscretions, such as those revealed by the trust challenges. Nevertheless, issues of security and privacy remained, particularly where the agent appeared to hold or have access to personal, and even intimate knowledge about the passenger. Although the approach is inspired by aspects of human-human interpersonal relationships – and results support the fact that participants imbued the agent with similar, humanlike qualities and capabilities – the study also reveals potential challenges facing designers of ‘intelligent’ conversational user interfaces, such as system transparency and the development of an appropriate mental model, particularly where users perceived the agent’s humanlike qualities to be approaching perfection. This suggests that future work should focus on tailoring the experience to engender *appropriate* human-likeness, and thereby avoid the perils of the uncanny valley [45]. Results of the study can also inform the design, and ultimate uptake and acceptance of autonomous, self-driving ‘pod’ vehicles more generally.

ACKNOWLEDGEMENTS

The research was conducted in collaboration with Jaguar Land Rover, and we would like to thank them for their advice

and support. We are also indebted to our wizard, Pablo (as ever), without whom the study would not have been possible.

REFERENCES

1. C. Miralles-Guasch and E. Domene, "Sustainable transport challenges in a suburban university: The case of the Autonomous University of Barcelona," *Transport policy*, vol. 17, no. 6, pp. 454-463, 2010.
2. M. Kyriakidis, R. Happee and J. de Winter, "Public opinion on automated driving: Results of an international questionnaire among 5000 respondents.," *Transportation research part F: traffic psychology and behaviour*, vol. 32, pp. 127-140, 2015.
3. D. Fagnant and K. Kockelman, "Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations," *Transportation Research Part A: Policy and Practice*, vol. 77, pp. 167-181, 2015.
4. S. Merritt, "Affective processes in human-automation interactions," *Human Factors*, vol. 53, no. 4, pp. 356-370, 2011.
5. J.-Y. Jian, A. M. Bisantz and C. G. Drury, "Foundations for an Empirically Determined Scale of Trust in Automated Systems," *International Journal of Cognitive Ergonomics*, vol. 4, no. 1, pp. 53-71, 2000.
6. J. Choi and Y. Ji, "Investigating the importance of trust on adopting an autonomous vehicle," *International Journal of Human-Computer Interaction*, vol. 31, no. 10, pp. 692-702, 2015.
7. M. Ghazizadeh, J. Lee and L. Boyle, "Extending the Technology Acceptance Model to assess automation," *Cognition, Technology & Work*, vol. 14, no. 1, pp. 39-49, 2012.
8. J. D. Lee and K. A. See, "Trust in automation: Designing for appropriate reliance," *Human Factors*, vol. 46, pp. 50-80, 2004.
9. J. Lee and N. Moray, "Trust, self-confidence, and operators' adaptation to automation," *International journal of human-computer studies*, vol. 40, no. 1, pp. 153-184, 1994.
10. E. Adell, A. Varhelyi and L. Nilsson, "The definition of acceptance and acceptability.," in *Driver Acceptance of New Technology. Theory, Measurement and Optimisation*, 2014, pp. 11-21.
11. F. D. Davis, "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology," *MIS Quarterly*, vol. 13, no. 3, pp. 319-340, 1989.
12. J. Choi and M. Kim, "Anthropomorphic Design: Projecting Human Characteristics to Products," in *International Association of Societies of Design Research Conference (IASDR2009)*, Seoul, Korea, 2009.
13. S. Merritt and D. Ilgen, "Not all trust is created equal: Dispositional and history-based trust in human-automation interactions," *Human Factors*, vol. 50, no. 2, pp. 194-210, 2008.
14. P. de Vries, C. Midden and D. Bouwhuis, "The effects of errors on system trust, self-confidence, and the allocation of control in route planning," *International Journal of Human-Computer Studies*, vol. 58, no. 6, pp. 719-735, 2003.
15. C. Flavián, M. Guinalú and R. Gurrea, "The role played by perceived usability, satisfaction and consumer trust on website loyalty," *Information & management*, vol. 43, no. 1, pp. 1-14, 2006.
16. M. Culnan and P. Armstrong, "Information privacy concerns, procedural fairness, and impersonal trust: An empirical investigation.," *Organization science*, vol. 10, no. 1, pp. 104-115, 1999.
17. D. M. Rousseau, S. B. Sitkin, R. S. Burt and C. Camerer, "Not so different after all: a cross-discipline view of trust," *Academy of Management Review*, vol. 23, no. 3, pp. 393-404, 1998.
18. R. Larzelere and T. Huston, "The dyadic trust scale: Toward understanding interpersonal trust in close relationships," *Journal of Marriage and the Family*, pp. .595-604, 1980.
19. C. Nass and S. Brave, *Wired for speech: How voice activates and advances the human-computer relationship*, MIT Press, 2005.
20. D. Massaro and M. Cohen, "Perceiving talking faces," *Current Directions in Psychological Science*, vol. 4, no. 4, pp. 104-109.
21. S. Pinker, *The language instinct: How the mind creates language*, Penguin, 2003.
22. G. Brown, B. Gillian and G. Yule, *Discourse analysis*, Cambridge university press, 1983.
23. C. Cheepen, *The predictability of informal conversation.*, Pinter, 1988.
24. R. Dunbar and R. Dunbar, *Grooming, gossip, and the evolution of language*, Harvard University Press, 1998.
25. H. Fang, H. Cheng, E. Clark, A. Holtzman, M. Sap, M. Ostendorf, Y. Choi and N. Smith, "Sounding board-university of washington's alexa prize submission," in *Alexa Prize Proceedings*, 2017.

26. H. H. Clark, *Using Language*, Cambridge: Cambridge University Press, 1996.
27. D. R. Large and G. E. Burnett, "The effect of different navigation voices on trust and attention while using in-vehicle navigation systems," *Journal of Safety Research*, vol. 49, pp. 69-75, 2014.
28. A. Tapus, M. Mataric and B. Scassellati, "Socially assistive robotics grand challenges of robotics.," *IEEE Robotics & Automation Magazine*, vol. 14, no. 1, pp. 35-42, 2007.
29. R. Looije, M. Neerinx and F. Cnossen, "Persuasive robotic assistant for health self-management of older adults: Design and evaluation of social behaviors," *International Journal of Human-Computer Studies*, vol. 68, no. 6, pp. 386-397, 2010.
30. Farlex Inc., "The Free Dictionary," Online. . Available: <https://www.thefreedictionary.com/anthropomorphism> . Accessed 1 March 2019. .
31. H. Gray, K. Gray and D. Wegner, "Dimensions of mind perception," *Science*, vol. 315, no. 5812, pp. 619-619, 2007.
32. A. Waytz, J. Heafner and N. Epley, "The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle," *Journal of Experimental Social Psychology*, vol. 52, pp. 113-117, 2014.
33. Y. Forster, F. Naujoks and A. Neukum, "Increasing anthropomorphism and trust in automated driving functions by adding speech output," in *IEEE Intelligent Vehicles Symposium (IV)*, 2017.
34. A. Miglani, C. Diels and J. Terken, "Compatibility between trust and non-driving related tasks in UI design for highly and fully automated driving," 2016.
35. A. Eriksson and N. A. Stanton, "The Chatty Co-Driver: A Linguistics Approach Applying Lessons Learnt from Aviation Incidents," *Safety Science*, vol. 99, no. A, pp. 94-101, 2017.
36. V. Antrobus, G. Burnett and D. Large, "'Trust me—I'm AutoCAB': Using natural language interfaces to improve the trust and acceptance of level 4/5 autonomous vehicles," in *The 6th HUMANIST Conference*, The Hague, 2018.
37. N. Dahlbäck, A. Jönsson and L. Ahrenberg, "Wizard of Oz studies—why and how," *Knowledge-based systems*, vol. 6, no. 4, pp. 258-266, 1993.
38. D. Large, L. Clark, A. Quandt, G. Burnett and L. Skrypchuk, "Steering the conversation: A linguistic exploration of natural language interactions with a digital assistant during simulated driving," *Applied ergonomics*, vol. 63, pp. 53-61, 2017.
39. M. Chin, R. Yordon, B. Clark, T. Ballion, M. Dolezal, R. Shumaker and N. Finkelstein, "Developing and anthropomorphic tendencies scale," in *Human Factors and Ergonomics Society Annual Meeting*, Los Angeles, CA, 2005.
40. M. Bradley and P. Lang, "Measuring emotion: the self-assessment manikin and the semantic differential," *Journal of behavior therapy and experimental psychiatry*, vol. 25, no. 1, pp. 49-59, 1994.
41. C. Bartneck, D. Kulić, E. Croft and S. Zoghbi, "Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots," *International journal of social robotics*, vol. 1, no. 1, pp. 71-81, 2009.
42. V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative Research in Psychology*, vol. 3, no. 2, 2006.
43. D. R. Large, L. Clark, G. Burnett, K. Harrington, J. Luton, P. Thomas and P. Bennett, "'It's Small Talk, Jim, But Not as We Know It.' Engendering Trust through Human-Agent Conversation in an Autonomous, Self-Driving Car," in *1st International Conference on Conversational User Interfaces (CUI2019)*, Dublin, Ireland, 2019.
44. M. Mori, "The Uncanny Valley," *Energy*, vol. 7, no. 4, pp. 33-35, 1970.
45. L. Clark, A. Ofemile and B. Cowan, "Exploring verbally uncanny valley effects with vague language in computer speech," in *Voice Attractiveness: Studies on Sexy, Likable, and Charismatic Speakers*, Springer, 2019.
46. D. R. Large, G. Burnett and L. Clark, "Lessons from Oz: Design Guidelines for Automotive Conversational User Interfaces," in *Automotive User Interfaces and Interactive Vehicular Applications (AutoUI2019) (under review)*, Utrecht, Netherlands, 2019.
47. A. Sabelli, T. Kanda and N. Hagita, "A conversational robot in an elderly care center: an ethnographic study," in *6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2011.
48. M. Porcheron, J. Fischer, S. Reeves and S. Sharples, "Voice interfaces in everyday life," in *In proceedings of the 2018 CHI conference on human factors in computing systems*, 2018.