Effects of Wording and Gendered Voices on Acceptability of Voice Assistants in Future Autonomous Vehicles

Iris Jestin
School of Computer Science
University of Nottingham
Nottingham, UK
irisjestin@gmail.com

Joel E. Fischer
School of Computer Science
University of Nottingham
Nottingham, UK
joel.fischer@nottingham.ac.uk

Maria J. Galvez Trigo
School of Computer Science
University of Lincoln
Lincoln, UK
mgalveztrigo@lincoln.ac.uk

David R. Large
Human Factors Research Group
Faculty of Engineering, University of Nottingham
Nottingham, UK
david.r.large@nottingham.ac.uk

Gary E. Burnett
Human Factors Research Group
Faculty of Engineering, University of Nottingham
Nottingham, UK
gary.burnett@nottingham.ac.uk

ABSTRACT

Voice assistants in future autonomous vehicles may play a major role in supporting the driver during periods of a transfer of control with the vehicle (handover and handback). However, little is known about the effects of different qualities of the voice assistant on its perceived acceptability, and thus its potential to support the driver's trust in the vehicle. A desktop study was carried out with 18 participants, investigating the effects of three gendered voices and different wording of prompts during handover and handback driving scenarios on measures of acceptability. Participants rated prompts by the voice assistant in nine different driving scenarios, using 5-point Likert style items in a during and post-study questionnaire as well as a short interview at the end. A commanding/formally worded prompt was rated higher on most of the desirable measures of acceptability as compared to an informally worded prompt. The 'Matthew' voice used was perceived to be less artificial and more desirable than the 'Joanna' voice and the gender-ambiguous 'Jordan' voice; however, we caution against interpreting these results as indicative of a general preference of gender, and instead discuss our results to throw light on the complex socio-phonetic nature of voices (including gender) and wording of voice assistants, and the need for careful consideration while designing the same. Results gained facilitate the drawing of insights needed to take better care when designing the voice and wording for voice assistants in future autonomous vehicles.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**.

KEYWORDS

autonomous vehicles, voice assistant, acceptability, gendered voice, wording, handover, handback

CUI '22, July 26–28, 2022, Glasgow, UK

ACM Reference Format:
Iris Jestin, Joel E. Fischer, Maria J. Galvez Trigo, David R. Large, and Gary E. Burnett. 2022. Effects of Wording and Gendered Voices on Acceptability of Voice Assistants in Future Autonomous Vehicles. In *CUI '22: Conversational User Interfaces, July 26–28 2022, Glasgow, UK*. ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3543829.3543836

1 INTRODUCTION

With a rise in commercial popularity of autonomous and semi-autonomous vehicles, people may increasingly be able to engage in other activities while the car is in control. In these vehicles, the voice assistant (VA) may play a major role in acting as the mediator between the driver and the vehicle in the future. For instance it may inform the driver of various decisions taken by the vehicle and thus, may be an integral part of building the trust of the driver in the vehicle. This is particularly crucial in the case of a transfer of control from the car to the driver (handover) and vice versa (handback) [8, 20].

Owing to the importance a voice assistant may play in an autonomous vehicle, it is vital to ensure that the prompts given by it engage the attention of the driver at the right time with the right level of information for the given context, especially in the critical case of transfer of control. In an autonomous vehicle, the gender of the voice assistant's voice has been shown to have an influence on how anthropomorphic it is perceived as [7]. Commercial voice assistants frequently have a female-sounding voice as default, and research has shown that people tend to find female voices more agreeable and pleasant [16]. However, gendered voices also embed gender stereotypes and can contribute to sexualised and gendered language being used particularly when interacting with voice assistants with a female-sounding voice [27, 30]. Given recent development of so-called "genderless" voices (e.g., "Q"[1]) we were interested to study the effects of different gendered-voices on their acceptability to support transfer of control scenarios. We were hopeful that, in the spirit of gender equality, we would not find a significant effect of gender in our study, which could be seen to provide designers with more freedom to design voices with any gender in mind.

[1] https://www.genderlessvoice.com/

Further, research has shown that the wording of prompts by voice assistants (the way in which the message is formulated) affects drivers of autonomous vehicles in important ways. For instance, the wording of the prompts given by a voice assistant has been shown to affect the perceived urgency in collision avoidance systems [3], and allow drivers to form collision avoidance plans [6]. Research also showed that whether a prompt 'style' is informative or social affects the perceived ease of use and usability of the autonomous vehicle [13]. In our study we were interested in studying the effects of wording in different driving scenarios on acceptability in conjunction with gendered voices, which to our knowledge has not been studied in this combination. Hence, this study researches the effects of gendered voice and the wording of the prompts given by a voice assistant in an autonomous vehicle.

The study presented in this paper investigates how the acceptability of prompts to support different handover-handback scenarios is affected by different gendered voices and the wording of the voice assistant. Due to the COVID-19 pandemic we carried out a desktop study, in which we explored the perceptions of three gendered voices and different levels of wording of the prompts produced by the voice assistant, by measuring it against 16 measures of acceptability. Participants were shown different driving scenario videos upon which voice prompts by the voice assistant in each scenario are played. The prompts were systematically varied in their wording style and the gender of the voice. The participants were asked to answer a questionnaire to collect their ratings of acceptability for each scenario, and took part in a short interview afterwards. With the results of this study, we hope to be able to draw insights for designing more carefully the voice assistants that help building the driver's trust and acceptability for future autonomous vehicles. This leads us to formulate the following overarching research question broken down into two parts, rq1 and rq2:

RQ: How do gendered voices and the wording of the prompts affect the acceptability of a voice assistant to support driving scenarios for future autonomous vehicles?

rq1: How is the acceptability of a voice assistant affected by different gendered voices?

rq2: How is the acceptability of a voice assistant affected by the wording of the prompts it produces?

The findings show that a commanding/formally worded prompt was rated higher on most of the desirable measures of acceptability as compared to an informally worded prompt, but also show the informally worded prompt rated as more entertaining and friendly. Further, while the 'Matthew' voice used was perceived to be less artificial and more desirable than the 'Joanna' voice and and the gender-ambiguous 'Jordan' voice, we discuss how this is not to be taken as indicative of a general preference of gender. We further discuss the findings to highlight the complex socio-phonetic nature of voices (including gender and wording) and stress that care should be taken when designing voices for voice assistants for autonomous vehicles.

This paper first reviews literature related to autonomous vehicles, the role of gendered voices in voice assistants and the role of the wording of prompts by voice assistants. After a detailed description of how the study was conducted, the findings which address the research questions are presented. Following this, the findings are discussed in relation to relevant literature, along with the study's limitations.

2 RELATED WORK

Autonomous vehicles and automated driving are thought of as helping to remove or reduce human error involved in driving, which causes the majority of road traffic accidents; thus, protecting drivers, passengers, cyclists and pedestrian [2]. Automated driving could therefore help reduce the economic and societal costs caused by motor vehicle accidents and consequent injuries, and smooth traffic flow and reduce the time and money spent for commuting per day [4].

A key interactional issue in future autonomous vehicles is the transfer of control between driver and car. A vehicle 'handover' or 'take-over-request', according to literature [28], is defined as the transfer of vehicle control from the vehicle to the driver, when the vehicle automation is doubtful about being able to handle the approaching situation and requires the driver to take over control. Similarly, a vehicle 'handback', is defined as the resuming of control by the vehicle automation or the transfer of control from the driver back to the vehicle.

According to the Society of Automotive Engineers (SAE), there are six levels of automation in a vehicle [2]. As much as there is a move towards the full automation of vehicles, mainstream production of vehicles still remains at Level 2 automation [1]. This study is relevant to level 4 automation, where the driver may have the option to take control of the vehicle but the vehicle can perform all driving tasks under certain conditions, without the help of the driver.

The presence of a voice assistant in an autonomous vehicle and the type of information conveyed by it may help to improve the driver's awareness such as during emergency manual take-overs [8]. The voice and personality of an in-vehicle navigation system has an effect on the level of confidence that the drivers place in the commands and influences issues of trust and attention [12]. A voice assistant can act as a mediator and build the trust of a driver in a vehicle [8, 12]. Owing to the potential importance a VA may play in a future autonomous vehicle, this study explores the effects of the wording of prompts and the (gendered) sound of the VA's voice on its acceptability.

2.1 Gendered voice of a voice assistant

Research has suggested that female-sounding autonomous vehicle voice assistants are in general seen to be more likeable and friendly [7]. The authors suggest that the gender of a voice assistant's voice in a fully autonomous vehicle has a higher influence on the perceived anthropomorphism, animacy and social presence scales, than the potential (visual) embodiment of an agent. Their results show that their participants thought of both voice-only and embodied-only female agents to be more competent than both the male agents. The study confirmed the stereotype where the male voice was seen to be more appropriate in a safety critical scenario and the female voice being perceived as more warm, friendly and approachable. This is in line with the classic study by Nass and Moon [20] which reported that a female-voiced computer was seen by participants to be more informed about stereotypical topics such

as love and relationships than a male-voiced computer which was rated to be more informed on a topic such as computers. The results of the study also reported that positive praise was perceived to be more compelling by participants, if it came from a male-voiced computer as opposed to a female-voiced computer. Obinali points out that people are more likely to be accepting of a voice that is perceived to be pleasant [21].

Another study shows that the design of the voice assistant influences how the autonomous vehicle is perceived and hence strongly affects the adoption of the autonomous vehicle [13]. Lee et al.'s study shows that participants rated the autonomous vehicle to have a higher perceived ease of use (PEU) and thus perceived usability (PU), when the voice assistant was consistent with the traditional gender stereotypes of informative male and social female. These results reinforce the importance of considering both agent style and gender when designing voice assistants for autonomous vehicles, for an increased PEU and PU, and hence a stronger adoption intention. However, it also suggests that designers have the power to shape social norms and stereotypes in the context of voice and gender.

In the study by Tolmeijer et al. [26], which measured trust and trait attribution of a VA when comparing variations of gender and pitch (male-low, male-high, gender-ambiguous, female-low, female-high) by the task context that the VA was deployed in, it was found that a higher pitched voice was trusted more than a lower pitched voice. Similar to previous research, trust and trait attribution stereotypes were context dependent. The gender-ambiguous voice was not seen to have a negative impact on trust as compared to the gendered voices, which encourages the future use of gender ambiguous voices for VAs without reinforcing gender stereotypes. But further research needs to be done on the effects of gender ambiguous voices, which has motivated this study in part.

According to Sutton [24], the gender of the voice assistant that is assigned to it by its users, is not determined by voice alone, but multiple elements in a Voice User Interface (VUI)'s design along with it [20, 24]. How gender in a voice is seen, varies across people and each person has their own understanding of what "male" and "female" sounds like [17]. This is influenced by people's expectations [23] which are formed by previous experience [9].

Using the terms 'genderless' or 'gender neutral' can arguably be seen as inappropriate because it suggests the absence of gender or the lack of importance of gender when designing technology [15, 24]. Therefore, the use of the term gender-ambiguous is encouraged, and is used in this study. 'Gender ambiguous' voices can be pulled into being seen as either male or female despite being carefully designed to sound non-binary [9]. This is because of the influence of other elements such as the physical appearance of the product, product branding, context, pitch of the voice used, etc. on the perception of the gender assigned to a VA by its users. Hence, it is necessary to design these elements carefully, with a sensitivity towards gender [24].

As can be seen from the literature, the gendered-sounding voices of voice assistants play an integral role in ascribing stereotypical traits to it such as friendliness (female) and technical competence (male), while adopting a gender-ambiguous voice may avoid some of the gendered social norms and stereotypes [26]. However, there is a lack of studies particularly looking at effects of gender-ambiguous voices. Therefore, the effects of gendered voices, including that of a gender-ambiguous voice, on the perceived acceptability of a VA will be part of this study.

2.2 Wording of prompts by a voice assistant

There are a range of implications of the wording of prompts produced (i.e., "spoken") by a VA. Research shows that perceived urgency and annoyance is impacted by the semantic factors of the warnings issued in the collision avoidance systems of simulated driving [3]. The signal words chosen in the warnings have an impact on the perceived urgency of the interruption. Information provided in a spoken warning is effective in helping the driver plan for collision avoidance strategies rather than the simple application of brakes [6]. In the earlier mentioned research by Lee et al. [13], which studied the perception and adoption intention of autonomous vehicles through voice agent gender and style, the authors discussed how the style of the voice commands being informative or social affected the perceived ease of use and the perceived usability of the AV, which was influenced by the gender stereotypes of informative male and sociable female.

The type of information in the prompts given in a semi-autonomous car to driver communication has a significant effect on the driver's behaviour and attitude, and is further discussed in the research by Koo et al.. The study goes on to prove how too much information, even when useful, overwhelms the driver and affects performance whereas the wrong kind of information decreases the driver's sense of responsibility and causes confusion, especially in the case of transfer of control between car to driver or vice versa. This suggests that the semantic characteristics and wording of the warnings given by a voice assistant in an autonomous vehicle impacts the driver's behaviour towards and perception of the vehicle and its surroundings [10].

Research by Wong et al. has shown that a higher assertive voice of a VA results in a faster driver reaction time and sense of urgency, as compared to a lower assertive voice in a self-driving car [29]. The study varied assertiveness of the voice by manipulating the wording and the tone of the voice, and was aimed at understanding the effects of assertiveness of the VA while the driver was immersed in a non-driving task. The results showed that the higher assertive voice was able to be perceived as urgent enough to distract the driver from the secondary task that they were immersed in, irrespective of its cognitive demand, as opposed to the lower assertive voice. It was seen that the smallest change in the wording and acoustic elements of the command being said, changed the driver's reaction time and perceived urgency. Hence the study throws light on the power of the wording and tone of the voice commands, in bringing the driver's attention from multiple secondary tasks to the primary driving task.

The wording or the content of the prompts said by a voice assistant and its para-linguistic cues are deeply related to each other. How a user experiences a voice-enabled device depends on both the voice as well as the prompt's linguistic content [18]. Users find it easier to trust the device when there is a consistency between how the voice sounds and what it is saying [19]. Speech Synthesis Markup Language offers only a limited control over the design

of speech in this development frameworks for building voice assistants, and there is a need for better, varied control to be made available [22].

Based on literature, it is probable that the wording of the prompts said by a voice assistant in autonomous vehicles has a significant effect on its perception by the driver. Hence, the effects of wording on the perceived acceptability of a VA in a self-driving car will be investigated in this study.

3 THE STUDY

3.1 Participants

18 participants (9 male and 9 female, 19-55 years of age), with a valid driving license, were recruited to take part after the study was approved by the University's ethics committee. The participants were recruited from UK (6), UAE (7) and India (5), and hence had a range of mixed backgrounds. They were recruited by opportunistic sampling via social media and through direct or indirect friends and family. All participants were given information regarding the study before the experiment, gave informed consent and were debriefed after the study.

3.2 Study design

A 3x3 (**gendered voices** ['Matthew', male | 'Joanna', female | 'Jordan', gender-ambiguous] **x wording** [commanding | formal | informal]) factorial within-subjects study was carried out, which explored six driving scenarios of handover and handback. The study design was counterbalanced to counteract order effects; each participant experienced nine scenarios, experiencing every combination of gender by wording once, and every driving scenario at least once. Thus, participants were presented with a different gender-wording combination for each scenario. A manipulation check to examine whether the different levels of wording and gendered voices were being perceived as intended, was carried out for each scenario.

For a given scenario, three different levels of wording of the prompts said by the voice assistant were created by modifying the tone of voice (i.e., the wording) into commanding, formal, or informal wording while keeping the semantic meaning the same. An example of a prompt can be seen in Figure 1. The dependent variable comprised 16 measures of acceptability; The 16 items were: "Appropriate", "Artificial", "Acceptable", "Annoying", "Assertive", "Effective", "Entertaining", "Friendly", "Human-Like", "Raising Alertness", "Undesirable", "Helpful", "Trustworthy", "Distracting", "Useful" and "Understandable", adapted from the voice rating sheet developed by Large and Burnett [11].. The independent variables were the gendered voices and the wording of the voice assistant.

3.3 Scenario design

Drawing from the literature [14, 28], the handover and handback driving scenarios (see Table 1) were created based on a combination of three factors: 1. whether the transfer of control was initiated by the driver or the car; 2. whether the direction of transfer was from car to driver (handover) or vice versa (handback); 3. whether the transfer of control was safety-critical or optional. Thus, the scenarios were; i) car-initiated handback (optional) ii) car-initiated handback (safety-critical) iii) car-initiated handover (safety-critical) iv) driver-initiated handback (optional) v) driver-initiated handback but car rejects (safety-critical) vi) driver-initiated handover (safety-critical).

The examples for car-initiated handover and driver-initiated handover i.e; the 'fog' scenario and the 'sharp turn' scenario, are aligned with the already existing taxonomy and literature for handovers [14]. 'Handback', or the transfer of control from driver to the car has not been thoroughly investigated in studies and a clear and precise taxonomy for the same is yet to be formed [28]. Hence, the four handback scenarios are not aligned with literature but inspired by the already existing taxonomy for handovers. The 'approaching a bridge' scenario is a car-initiated handback and the 'car swerving' scenario is a car-initiated handback safety-critical situation. The 'entering a highway' is a driver-initiated handback whereas the 'road marking' scenario is a driver-initiated handback.

For each of these scenarios, a prompt in one of the three intended wordings was scripted, which was played to the participants in one of the three gendered voices.

3.4 Apparatus

The desktop study was conducted via video call on Zoom, Microsoft Teams or Google Meet according to the participants' convenience, where they were shown first-person driving videos via a shared screen for simulation purposes and played the voice assistant prompts as audio clips. The video clips shown were selected from YouTube according to the driving scenario under consideration. The audio clips of the voice assistant prompts were played using the voice user experience prototyping tool 'Voiceflow'. The online questionnaires were created using Microsoft Forms and the interview was voice recorded on a mobile phone's voice recorder. The setup can be seen in Figure 1.

3.5 Materials

The prompts were scripted and played using the voice prototyping tool, 'Voiceflow', which converts the text of the prompts entered into the speak block into voice output. The voice used for both the male and female voices were the Amazon Alexa's US English voices called 'Matthew' and 'Joanna'. Due to a lack of gender-ambiguous voices in Voiceflow we created the 'Jordan' voice using 'Speech Synthesis Markup Language' (SSML). 'Jordan' is based on the female voice 'Joanna' and was created by manipulating the SSML tags for prosody 'pitch', 'rate' and 'volume'.

The wording of the prompts was created according to literature [10], which suggests that both 'how' and 'why' information is needed in prompts given by a VA of a semi-autonomous vehicle, for safety-critical scenarios to lead to the highest driving performance. Hence, the prompts used in the driving scenarios had both 'how' and 'why' messages in them for each of the wordings, commanding, formal and informal.

The During Study questionnaire, had two five-point Likert scale questions (from strongly disagree to strongly agree), that were repeated for each of the nine scenarios shown to the participants; thus, higher scores of the participants are consistent with a higher level of agreement with the statements. The first question served the purpose of a manipulation check, to see whether the gender and wording of the voice assistant was being perceived as intended (participants were not told the name of the voices so as not to

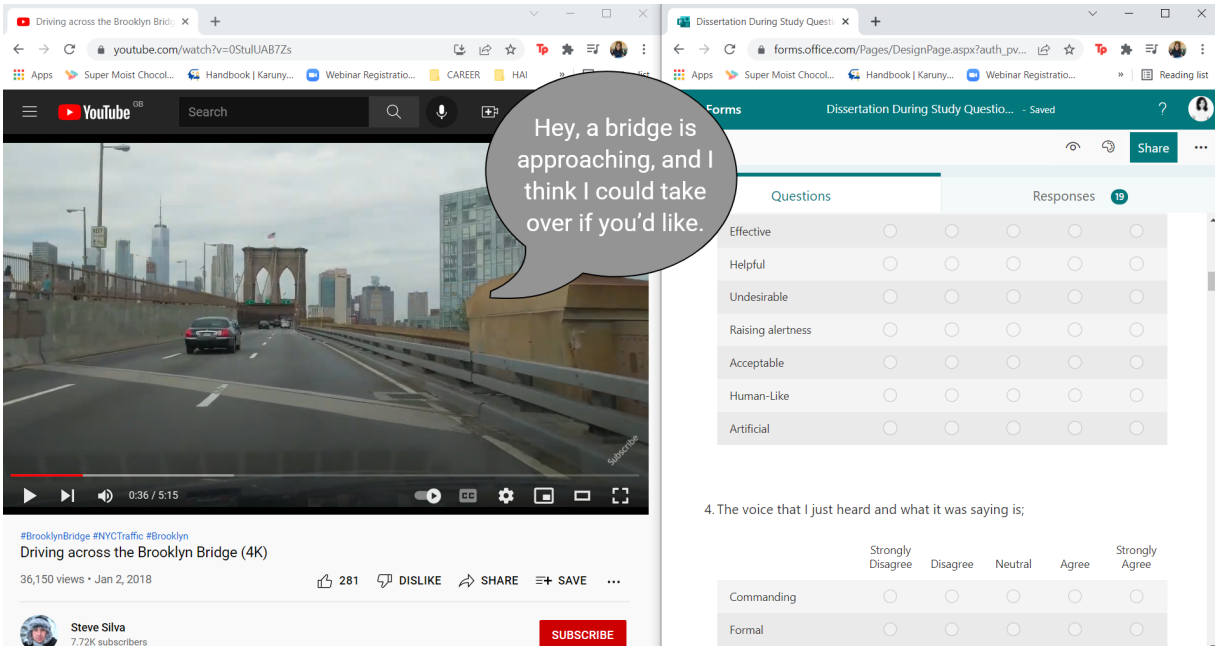

Figure 1: The desktop study setup showing a YouTube video for the 'Approaching a bridge' driving scenario and the prompt by the voice assistant (informal wording), along with the During Study questionnaire on Microsoft Forms.

SCENARIOS		Type		
		safety-critical		optional
Initiated by	car	**Fog** (handover). "The visibility is low due to the fog. Please take over the car now." (formal)	**Car swerving** (handback). "Taking over control now! You're not responding. The car is swerving." (commanding)	**Approaching bridge** (handback). "Hey a bridge is approaching, and I think I could take over if you'd like." (informal)
	driver	**Sharp turn** (handover). "Yes, that's a good idea. A sharp turn is approaching." (formal)	**Road marking** (handback). "Can't take over control! Road markings aren't clear." (commanding)	**Entering highway** (handback). "Cool, yea I'll take over. Was just going to suggest it. We're entering a highway." (informal)

Table 1: Handover and handback scenarios with sample voice assistant prompts (wording level) by type and whether initiated by car or driver as designed by drawing on the taxonomy for handovers [14].

bias their gender perception). The second question, inspired by the voice rating sheet developed by Large and Burnett, asked, "The wording of voice in this context is..." followed by the sixteen items to measure acceptability (see section 3.2).

The post study questionnaire had three questions to measure the overall gender preference for the VA. This was followed by a question for each driving scenario to see which wording of the prompt said by the VA the participants thought would have been appropriate in that scenario. Again these were five-point Likert scale questions, from 1 (strongly disagree) to 5 (strongly agree).

The recorded interview at the end was semi-structured with questions such as; "From the different scenarios that you've just experienced, tell me a time when you felt the voice was appropriate/inappropriate for the context and why?", "Could you tell the difference when the voices were changing between commanding,

formal and informal?". We will make use of the interview statements to contextualise the quantitative results presented in the next section.

3.6 Procedure

The participants were briefed on the aim of the study and on what they would be asked to do as part of it. The participants were given an information sheet with all the necessary details regarding the study, following which they were asked to fill in a consent form. As part of the study, the screen was shared and a YouTube video pertaining to the driving scenario under consideration was played for them to see. Along with this, a detailed description of what happens in the particular scenario was read out. Following this, an audio clip of the voice assistant prompt was played by triggering the clip in Voiceflow. The participant was then asked to answer the related questions in the 'During Study questionnaire'. The above was repeated for each of the 9 driving scenarios. At the end, the participants were asked to fill in the 'Post Study Questionnaire' following which a short voice recorded interview was conducted. The interview was semi-structured and aimed at understanding their outlook on the scenarios, voices and the study in general. All the participants were debriefed and thanked at the end of the study.

4 FINDINGS

First, in order to raise confidence that the levels of the independent variables (IVs) *wording* and *gendered voices* were indeed perceived as significantly different from one another, statistical analyses were performed on the manipulation check questions. One-way independent measures ANOVAs were conducted on the manipulation check responses, with *wording* (commanding, formal, informal) of the prompts as the IV and the rating of how participants perceived the wording to be commanding, formal or informal as the dependent variables (DVs). The pairwise comparisons showed that the commanding and formally worded prompts were not perceived as significantly different from each other. As a result, we have collapsed the two levels *commanding* and *formal* of the IV wording into one level, and henceforth treat this in our analysis as a joint level 'commanding/formal'.

Furthermore, to rule out that the *type* of scenario (safety-critical vs. optional) had an effect on how the wording was perceived (and/or an interaction effect with the IV *wording*), we ran a 3x2 (wording x safety-criticality of driving scenario) factorial independent measures ANOVA against the rating to which the wording was perceived to be commanding, formal or informal. The results demonstrate that there is no statistically significant main effect of *type* of scenario on perceived wording, and that there is a statistically significant main effect of actual wording of the prompt on how the wording is perceived.

A final manipulation check was carried out to ensure that the different *gendered voices* were perceived as intended (i.e., recognisably female, male, and gender-ambiguous). One-way independent measures ANOVAs were conducted on the manipulation check responses, with *gendered voices* ('Joanna', 'Matthew', 'Jordan') of the prompts as the IV and the rating of how much the participants agreed that voice sounded female, male or gender-ambiguous as the DVs. The results showed that *gendered voices* does have a significant effect on the perceived gender at $p < 0.001$. The pairwise comparisons showed that all 3 voices are rated as significantly different from each other at $p < 0.001$, both when the DV is female or male sounding voice. For the 'Jordan' voice, participants agreed significantly more that it sounded gender-ambiguous than the 'Joanna' or 'Matthew' voice, which also did not differ significantly from each other. The results confirm the distinction in perception of the 'Joanna voice' sounding female, the 'Matthew' voice sounding male, and the 'Jordan' voice sounding gender-ambiguous, as was intended.

Following Wong et al.'s [29] approach to analysing Likert-scale data using parametric tests, we conducted a 3x2 (gender x wording) factorial repeated measures ANOVA on the 16 measures of acceptability, to analyse the effects of wording and gendered voices on acceptability. We present this 'main analysis' in the following two subsections. Effects with a $p - value < 0.05$ were considered to be significant.

4.1 Main analysis 1: effects of gendered voices

From the results of statistical analyses shown in Table 2, it can be seen that *gendered voices* had a statistically significant main effect on 2 out of the 16 measures of acceptability, at $p < 0.05$. The pairwise comparison seen in Figure 2 shows that the male 'Matthew' voice is rated as the least artificial and least undesirable out of the three, whereas the gender-ambiguous 'Jordan' voice is rated the most artificial and most undesirable. The 'Jordan' voice is rated significantly more undesirable than the 'Matthew' voice. However at this point it is important to state that we do not claim, nor intend to show that a male voice is *generally* preferred to a gender-ambiguous voice; it is important that, as we will discuss, we interpret the results with regard to *just these voices* that we compared, namely, 'Matthew', 'Jordan', and 'Joanna'.

The post-study interviews provided some more context to understand the findings regarding the artificiality and undesirability of the 'Jordan' voice, where one of the participants reported, "I wasn't a big fan... It wasn't very human-like". Another participant was of the opinion that, "The artificial voice was quite unsettling", thus clearly characterising the 'Jordan' voice as 'artificial', a sentiment which was echoed by other participants.

Thus, rq1 can be answered in the following way: while no significant effect of gendered voices on 14 of the 16 acceptability measures was found, a statistically significant effect for gendered voices was found on artificiality and undesirability. The male voice 'Matthew' was rated significantly less undesirable (i.e. more desirable) than the gender-ambiguous voice 'Jordan'.

4.2 Main analysis 2: effects of wording

To rule out whether the *transfer of control* in the scenario (*handover vs. handback*) had an unintended main or interaction effect, a 2x2 (handover-handback/wording) factorial between subjects ANOVA was conducted to test the effect of handover-handback and wording on the 16 acceptability measures. Whether the transfer of control was a handover or a handback did not have a statistically significant main effect on any of the 16 acceptability measures, and there was no interaction effect between wording of the prompts and the

Acceptability Measures	F(2,32)	p-value	Mean *('Joanna', female)*	Mean *('Matthew', male)*	Mean *('Jordan', gender-ambiguous)*
Artificial	3.462	0.044 *	3.337	2.806	3.513
Undesirable	3.870	0.031 *	2.39	2.014	2.691

Table 2: F, p and mean values of acceptability measures with a significant main effect of gendered voices ('artificial' and 'undesirable'). Higher scores indicate higher levels of agreement. * denotes significance at the 0.05 level.

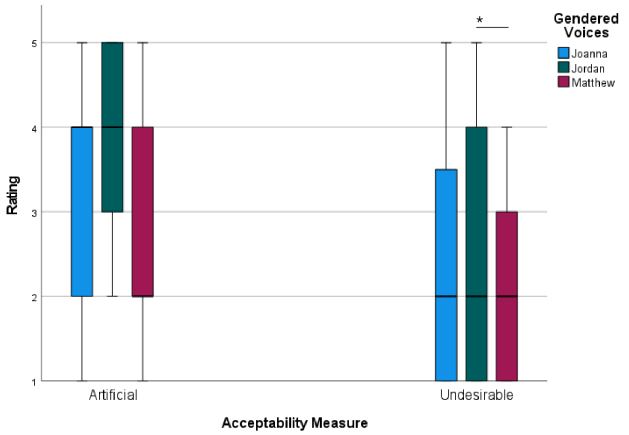

Figure 2: Boxplots showing participants' average ratings from 1 (strongly disagree) to 5 (strongly agree) for the two acceptability measures with a significant effect of *gendered voice* by level (type of voice) of the voice assistant. * denotes a significant difference at 0.05 level between the 'Jordan' and the 'Matthew' voice for how undesirable participants rated the voice. The lower edge of the box denotes the first quartile, and the upper edge the third quartile of the range of the ratings. The lower whisker and the upper whisker of the box plot denote the minimum and maximum ratings respectively. The thick line across the box denotes the median of the values.

handover-handback driving scenarios; thus, transfer of control was ruled out as a confounding factor.

Regarding the effect of *wording* of the prompts on acceptability, Table 3 shows a statistically significant main effect of *wording* at $p < 0.05$ on 10 out of the 16 measures of acceptability (listed in Table 3). The wording of the prompt did not have a statistically significant effect on the remaining 6 out of the 16 measures of acceptability (artificial, undesirable, distracting, human-like, helpful and trustworthy). The pairwise comparisons in Figure 3 show that the *commanding/formal* wording of prompts is considered as significantly more (at the 0.05 level) acceptable, effective, raising alertness, understandable and useful, and strongly more (at the $p < 0.01$ level) appropriate and assertive than the *informal* wording.

Conversely, the *informal* wording is rated as significantly more (at the 0.05 level) annoying, and strongly more (at the 0.01 level) entertaining and friendly than the *commanding/formal* wording.

This can be contextualised with interview statements in which many participants expressed that the commanding or formal voices seemed to better catch their attention. Quoting one of the participants who said, "To put trust in the machine to drive for you, I would like a more formal voice". A preference for a commanding/formal rather than informal prompts by voice assistants in future autonomous vehicles appears to be suggested by these findings; however, for certain qualities the informal wording was preferred. This is implied by interview statements such as, "I prefer a less commanding or formal, more laid back kind of a voice". Another participant went on to say, "I guess I liked the informal voice more than the formal one". This suggests that, as we will return to in the discussion, there are cases where designers will want to consider a more informal wording.

In summary, rq2 can be answered as follows: A statistically significant effect of wording for 10 out of the 16 acceptability measures was found, showing that a *commanding/formal* wording was rated significantly more acceptable, effective, raising alertness, understandable, useful, and strongly more appropriate and assertive than the *informal* wording; while the *informal* wording was rated as significantly more annoying, and strongly more entertaining and friendly than the *commanding/formal* wording.

4.3 Follow-up analysis: Does the driving scenario impact perceived acceptability of the wording?

The results from the interview suggested that the way the participants perceived the acceptability of the wording of a prompt could potentially have been influenced by the driving scenario under consideration. Quoting one of the participants, they said, "A lot of my decisions were based on the context" [of the scenario]. Another participant went on to say, "As the context was more critical, the voices became more commanding", suggesting that the driving scenario may have been a confounding factor. To analyse whether the driving scenario and wording of the prompts had an effect on the perceived acceptability of the VA, we ran a 2x6 (wording x scenarios) factorial between subjects ANOVA on the measures of acceptability (there were six scenarios, see Table 1). The results confirmed that there was a statistically significant effect for two of the acceptability measures, for how entertaining and friendly the prompt was rated. A statistically significant main effect of the driving scenario on the acceptability measure 'friendly' was found, with $F(5, 149) = 5.048, p < 0.001$; and a significant main effect was found on the measure 'entertaining', with $F(5, 150) = 3.139, p = 0.01$. An interaction effect between the wording and the driving scenario was found on the acceptability measure 'assertive'.

The results of this follow-up analysis show that the different wording levels of the prompts can interact with the scenario (i.e., for 'assertive', and that the scenario has a significant effect on

Acceptability Measures	F(1,16)	p-value	Mean *commanding/formal*	Mean *informal*
Acceptable	7.120	0.017 *	3.784	3.276
Annoying	5.082	0.039 *	2.04	2.45
Appropriate	10.028	0.006 **	3.91	3.29
Assertive	23.719	<.001 **	3.715	2.786
Effective	8.166	0.011 *	3.936	3.41
Entertaining	9.787	0.006 **	1.974	2.46
Friendly	29.038	<.001 **	2.808	3.78
Raising Alertness	8.203	0.011 *	3.75	3.276
Understandable	5.291	0.035 *	4.142	3.743
Useful	6.046	0.026 *	3.902	3.39

Table 3: F, p and mean values of acceptability measures with a significant main effect of wording. Asterisks indicate significance at the * 0.05 level and the ** 0.01 level.

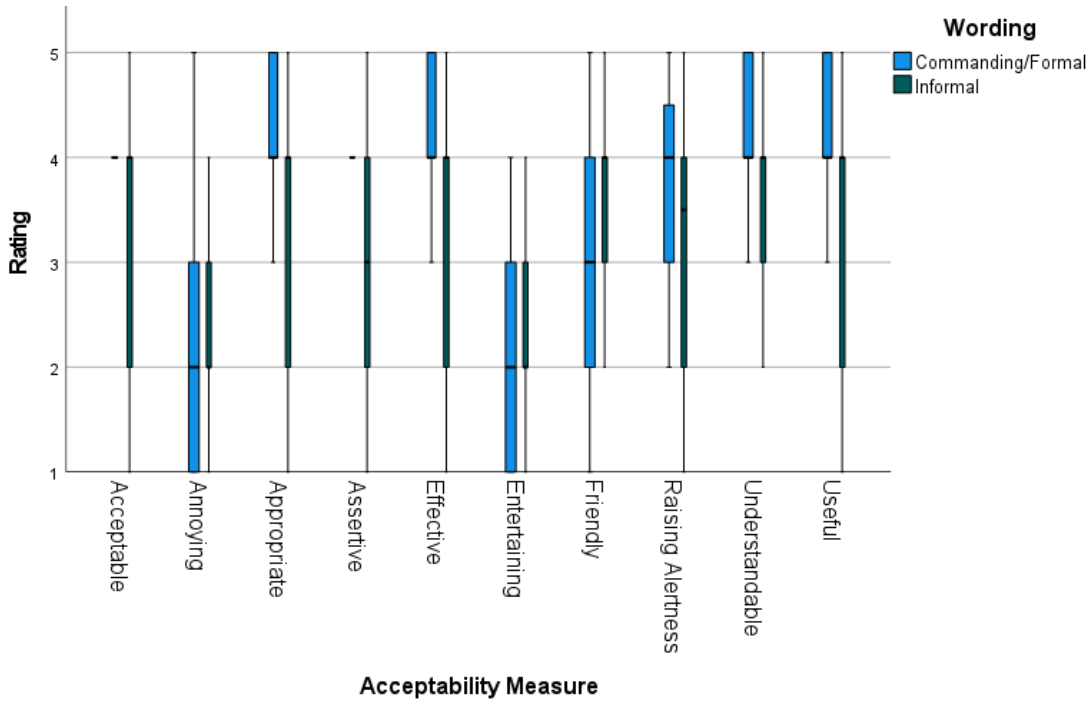

Figure 3: Boxplots showing participants' average ratings from 1 (strongly disagree) to 5 (strongly agree) for the 10 acceptability measures with a significant effect of *wording* by level of the prompt. The lower edge of the box denotes the first quartile, and the upper edge the third quartile of the range of the ratings. The lower whisker and the upper whisker of the box plot denote the minimum and maximum ratings respectively. The thick line across the box denotes the median of the values.

those acceptability measures (friendly and entertaining) for which the *informal* wording was also rated significantly higher. These findings do support, at least partially, our participants' interview statements emphasising that the context of the scenario matters for the acceptability of the prompt.

5 DISCUSSION

This study aimed to analyse the effects of *wording* and *gendered voices* of a voice assistant on its acceptability to support handover and handback scenarios in future autonomous vehicles. Herein we discuss our main findings from the study in relation to the literature.

5.1 Effects of gendered voices

Our main finding regarding gendered voices was that while no significant effect of gendered voices on 14 of the 16 acceptability measures was found, a statistically significant effect for gendered voices was found on artificiality and undesirability. The male voice

'Matthew' was rated significantly less undesirable than the gender-ambiguous voice 'Jordan'.

It is important to interpret and discuss these results carefully, so as not to jump to conclusions; these results must not be interpreted to suggest a *general* preference of gender-sounding voices for voice assistants. The first important caveat is that our results relate to the voices 'Matthew', 'Jordan', and 'Joanna'. It is important to stress that our results are about *just these* voices, not male or female or gender-ambiguous voices in general.

First, let's discuss some further caveats when it comes to 'designing' a gender-ambiguous voice. Due to the lack of gender-ambiguous voices in the Voiceflow tool we used in our study, we created the 'Jordan' voice by manipulating the SSML tags for prosody 'pitch', 'rate' and 'volume', of the 'Joanna' voice. Hence, it is not surprising that many participants reported in the interviews that they did not come to recognise a gender-ambiguous voice, with one participant referring to it as the "artificial female" voice. This echoes existing research which points out, based on previous experience with gendered voices, people tend to pull voices that are designed to be gender-ambiguous into a male or female category [9]. This characterisation is also consistent with the findings of the inferential statistics that showed that the 'Jordan' voice was seen as the most artificial. Even though Cambre and Kulkarni state the need for deliberately designing voices of VAs to sound distinctly non-human [5], the interview results of this study suggest otherwise. Participants expressed their disfavour for the 'Jordan' voice by referring to it as "artificial" and "freaky". Few of the participants went on to convey their preference for a "more human-like" voice. It makes sense then that the 'Jordan' voice was rated as significantly more artificial and undesirable, given that simply tweaking the attributes of a female voice, such as pitch, rate and volume, doesn't necessarily result in a 'gender-ambiguous' voice, nor a voice that sounds pleasant. We strongly suggest that what our findings show is that not any voice can simply be manipulated in order to sound 'gender-ambiguous', while also maintaining (or gaining) a sound associated with positive qualities. For designers of voice experiences this further demonstrates the need for better control over the auditory qualities of speech to be made available in the current development frameworks for voice assistants and conversational interfaces [5, 22].

Second, regarding findings on the effects of (binary) gender of conversational and voice assistants in the literature, some of our participants expressed preference for the 'Joanna' voice over the 'Matthew' voice, describing it as 'soothing' and 'trusting' in the interviews, and thus echoing popular research which suggests that a female voice is preferred to a male voice in a voice assistant, including those in autonomous vehicles [7, 20, 21]. However, this preference was not borne out by the results from the inferential statistical analysis, which suggests that even though the 'Matthew' voice was rated more desirable and less artificial compared to the 'Joanna' voice, the differences were not significant. We were pleased that our study did not replicate the findings by Lee et al., who found that participants perceived a higher ease of use and usability when the autonomous vehicle voice assistant abides by traditional gender stereotypes of informative male and friendly female [13]. Again, we do not want to make statements about gender preference in general, but we think it can be read as an encouraging sign for gender-equality that there were no significant differences for the male 'Matthew' voice and the female 'Joanna' voice.

Finally, it is important to connect our research to important work in HCI and CUI drawing on socio-phonetics, the study of the social factors influencing the production and perception of speech that shapes socio-cultural identities. Aligning with the study by Sutton et al. [25], our findings reinforce the idea that the qualities associated with human-sounding voices (including those human-sounding voices of a voice assistant) by people, are influenced by factors such as geography, sex and gender, age, sexuality, social class, accent, pitch, and dialect. Thus, in our study, in manipulating the pitch, rate and volume for the voice to sound 'gender-ambiguous' we may have inadvertently changed other, subtle characteristics of the voice that influence how people have perceived its acceptability. Our lesson for the CUI community from this is that, due to the sensitive and complex socio-phonetic nature of voices, careful consideration of the subtle ways in which different auditory elements are at play must be taken when designing voices for voice assistants and interfaces.

Due to the complex socio-phonetic nature of (gendered) voices our results do not conclusively point at what the voice in a future autonomous vehicle should sound like, and instead we suggest more research is needed and care should be taken when designing voices. If anything, we would feel most confident to suggest our findings support the research findings by Cambre and Kulkarni [5], which encourages a move away from the one-voice-fits-all approach.

5.2 Effects of wording

Our main finding regarding wording was that a statistically significant effect for 10 out of the 16 acceptability measures was found, showing that a *commanding/formal* wording was rated significantly more acceptable, effective, raising alertness, understandable, useful, and strongly more appropriate and assertive than the *informal* wording; while the *informal* wording was rated as significantly more annoying, and strongly more entertaining and friendly than the *commanding/formal* wording.

In line with the findings by Wong et al. [29], our findings support that the wording of the prompts said by a voice assistant has a significant effect on people's acceptability of the VA. Our current study points out that the commanding/formally worded prompts of the VA were rated higher in many of the desirable measures of acceptability as opposed to the informally worded prompts. Similar to Wong et al. 's study, which concludes that a higher assertive voice of a voice assistant in a self-driving car is seen to be more urgent than a lower assertive one, this study manipulates the wording of the prompts said by the VA, to measure acceptability. The commanding/formally worded prompt was seen to be more assertive and was also associated to a higher degree of being perceived as acceptable, appropriate, effective, understandable and useful. The higher rating of the commanding/formally worded prompt for the measure 'raising alertness' also aligns with Wong et al.'s finding that a higher assertive voice leads to faster reaction time and a higher sense of urgency [29]. While our findings also align with the findings by Wong et al. [29], in that the more 'assertive' commanding/formal wording of the prompt did not have a statistically significant effect on the perceived trustworthiness of the prompt,

conversely Large and Burnett [12] found that a higher assertive voice correlates with higher trust. Further research would need to be carried out to address this discrepancy.

A consistency of the wording of a prompt said by a VA, with how the voice sounds, helps to build trust with the users [19]. However, this study manipulated the wording of the prompt while keeping the tone consistent (other than changing the voices). Consistent with Nass and Lee's study [19], the interview results of this study had participants report, "The words were changing, but the tone remained formal, even when trying to be informal", referring to the commanding/formal prompt. This reinforces the idea that the wording and tone of the prompts by a VA need to be consistent for it to be acceptable by the users.

Regarding the informally worded prompt we found that this was perceived to be more annoying, but also strongly more entertaining and friendly than the commanding/formal wording. While how an informally worded prompt is perceived as more entertaining and friendly might be obvious, further research is needed to understand more about the context in which it is also perceived to be more annoying. Some reflections on the role of context follow.

Our findings also suggests a relationship between the wording and the context of use (e.g., the driving scenario), echoing related work that has pointed out that how a user is affected by the voice of a VA is likely dependent on the context of its use [25]. Participants reported in the interviews that they perceived the prompts to be commanding/formal or informal based on the context of the driving scenario under consideration. While whether the scenario involved, handover or handback did not show a statistically significant effect on the acceptability of the VA, the follow-up analysis showed that the scenario did have a statistically significant effect on how entertaining and friendly the prompt was rated, and the scenario had a significant interaction effect with wording on how assertive the prompt was rated. This furthermore suggests that the nature of the scenario should be considered when scripting the wording of voice assistant prompts, particularly when aiming for a more entertaining or friendly wording to delight the driver. Further research could be conducted in order to draw more definitive conclusions regarding the relationship between context/scenario and wording.

6 LIMITATIONS

Due to the social distancing rules in the pandemic we were forced to adopt a desktop study approach, which likely diminishes the ecological validity of the study. The probably limited suspension of disbelief of the participants helping them believe that they are driving on the road, may have affected their actual behaviour towards and the perception of the VA in the various driving scenarios. This was correctly put by one of the participants who said, "the amount of time there is for a reaction can only be judged in a real-life situation". Hence, the question of whether the results of the study would translate to a driver's actual behaviour on the road in real-life would need to be investigated in future work.

Further, while we find our sample size commensurate with that of similar studies in the literature [29], there may nevertheless be questions regarding the generalisability of the study findings due to the modest sample size. How the voice and the wording of the prompts said by the voice assistant is perceived is subject to subtle characteristics and varies from person to person, even though there might have been a general consensus in opinion. This may lead to participants perceiving a prompt differently to what it was designed to be perceived as, which could further affect the results.

7 CONCLUSION

This study investigated people's acceptability of voice assistants to support handover and handback driving scenarios in future autonomous vehicles, by exploring the effect of different gendered voices and the wording of the prompts said by the VA. Regarding the effects of gendered voices, we found an effect on only 2 out of 16 acceptability measures, which we interpret as an encouraging signal. Which gender a voice sounds like has little impact; however, we found the 'Jordan' voice which was our attempt to create a 'gender-ambiguous' sounding voice, was rated significantly more undesirable and artificial than the 'Matthew' voice. We stress that the result should be interpreted in the context of *just these* voices we compared, not as a general gender preference. We discuss the subtle and many characteristics by which people assign qualities to voices in light of related research on socio-phonetics and suggest that our findings support moving away from a one-voice-fits-all approach in designing voices.

Furthermore, our findings show how a commanding/formally worded prompt by a voice assistant is perceived to be significantly more acceptable for 10 out of 16 acceptability measures than an informally worded one by the participants, regardless of whether the driving scenario entailed handover or handback. Our findings also showed that an informal wording, while also judged as significantly more annoying, was also rated as significantly more entertaining and friendly. The findings on wording overall suggest that, probably due to the safety-critical nature of driving, commanding/formal prompts are preferred most of the time for handover and handback scenarios, but some of the time an informal wording can have a delightful effect on drivers. While less conclusive, parts of our analysis suggest that the type of scenario is likely to play a role in determining what kind of wording should be used. The study thus successfully addresses the research questions of how the acceptability of a voice assistant is affected by the gendered voices and wording of the prompts said by the voice assistant.

Overall, our findings demonstrate the complex nature of characteristics to be considered for the design of voice assistants in future autonomous vehicles and the impact they have on its perceived acceptability. Our findings demonstrate how careful thought must be given to designing the wording and the voice of VAs given the characteristics that our study has shown to impact its perception and acceptability.

ACKNOWLEDGMENTS

This work was supported by the Engineering and Physical Sciences Research Council [grant numbers EP/V00784X/1, EP/T022493/1]. We are grateful to our participants for taking part in this study.

REFERENCES

[1] 2022. The 6 levels of vehicle autonomy explained. https://www.synopsys.com/automotive/autonomous-driving-levels.html Retrieved February 25, 2022.
[2] National Highway Traffic Safety Administration. 2022. Automated vehicles for safety. https://www.nhtsa.gov/technology-innovation/automated-vehicles-

safety Retrieved February 25, 2022.
[3] Carryl L. Baldwin. 2011. Verbal collision avoidance messages during simulated driving: perceived urgency, alerting effectiveness and annoyance. *Ergonomics* 54, 4 (2011), 328–337. https://doi.org/10.1080/00140139.2011.558634 arXiv:https://doi.org/10.1080/00140139.2011.558634 PMID: 21491275.
[4] Michele Bertoncello and Dominik Wee. 2021. Ten ways autonomous driving could redefine the Automotive World. https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/ten-ways-autonomous-driving-could-redefine-the-automotive-world
[5] Julia Cambre and Chinmay Kulkarni. 2019. One Voice Fits All? Social Implications and Research Challenges of Designing Voices for Smart Devices. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 223 (nov 2019), 19 pages. https://doi.org/10.1145/3359325
[6] Shun-Hui Chang, Chih-Yung Lin, Chin-Ping Fung, Jiun-Ren Hwang, and Ji-Liang Doong. 2008. Driving performance assessment: Effects of traffic accident location and alarm content. *Accident Analysis & Prevention* 40, 5 (2008), 1637–1643. https://doi.org/10.1016/j.aap.2008.05.003
[7] Jiayuan Dong, Emily Lawson, Jack Olsen, and Myounghoon Jeon. 2020. Female Voice Agents in Fully Autonomous Vehicles Are Not Only More Likeable and Comfortable, But Also More Competent. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 64, 1 (2020), 1033–1037. https://doi.org/10.1177/1071181320641248 arXiv:https://doi.org/10.1177/1071181320641248
[8] Michelle Hester, Kevin Lee, and Brian P. Dyre. 2017. "Driver Take Over": A Preliminary Exploration of Driver Trust and Performance in Autonomous Vehicles. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 61, 1 (2017), 1969–1973. https://doi.org/10.1177/1541931213601971 arXiv:https://doi.org/10.1177/1541931213601971
[9] Keith Johnson. 2006. Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics* 34, 4 (2006), 485–499. https://doi.org/10.1016/j.wocn.2005.08.004 Modelling Sociophonetic Variation.
[10] Jeamin Koo, Jungsuk Kwac, Martin Steinert, Larry Leifer, and Clifford Nass. 2014. Why did my car just do that? Explaining semi-autonomous driving actions to improve driver understanding, trust, and performance. *International Journal on Interactive Design and Manufacturing (IJIDeM)* 9 (01 2014). https://doi.org/10.1007/s12008-014-0227-2
[11] David Large and Gary Burnett. 2013. Drivers' preferences and emotional responses to satellite navigation voices. *Int. J. of Vehicle Noise and Vibration* 9 (01 2013), 28 – 46. https://doi.org/10.1504/IJVNV.2013.053815
[12] David R. Large and Gary E. Burnett. 2014. The effect of different navigation voices on trust and attention while using in-vehicle navigation systems. *Journal of Safety Research* 49 (2014), 69.e1–75. https://doi.org/10.1016/j.jsr.2014.02.009 Proceedings of the International Conference on Road Safety (RSS2013).
[13] Sanguk Lee, Rabindra Ratan, and Taiwoo Park. 2019. The Voice Makes the Car: Enhancing Autonomous Vehicle Perceptions and Adoption Intention through Voice Agent Gender and Style. *Multimodal Technologies and Interaction* 3 (03 2019), 20. https://doi.org/10.3390/mti3010020
[14] Rod Mccall, Fintan Mcgee, Alexander Meschtscherjakov, Nicolas Louveton, and Thomas Engel. 2016. Towards A Taxonomy of Autonomous Vehicle Handover Situations. https://doi.org/10.1145/3003715.3005456
[15] Jenny Medeiros. 2022. This is Q: The first genderless voice for ai. https://www.voicesummit.ai/blog-old/genderless-voices-are-finally-coming-to-ai
[16] Wade J Mitchell, Chin-Chang Ho, Himalaya Patel, and Karl F MacDorman. 2011. Does social desirability bias favor humans? Explicit–implicit evaluations of synthesized speech support a new HCI model of impression management. *Computers in Human Behavior* 27, 1 (2011), 402–412.
[17] John W Mullennix, Keith A Johnson, Meral Topcu-Durgun, and Lynn M Farnsworth. 1995. The perceptual representation of voice gender. *The Journal of the Acoustical Society of America* 98, 6 (1995), 3080–3095.
[18] Clifford Nass and Scott Brave. 2005. *Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship.* The MIT Press.
[19] Clifford Nass and Kwan Lee. 2001. Does Computer-Synthesized Speech Manifest Personality? Experimental Tests of Recognition, Similarity-Attraction, and Consistency-Attraction. *Journal of experimental psychology. Applied* 7 (10 2001), 171–81. https://doi.org/10.1037/1076-898X.7.3.171
[20] Clifford Nass and Youngme Moon. 2000. Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues* 56 (03 2000), 81–103. https://doi.org/10.1111/0022-4537.00153
[21] Chidera Obinali. 2019. The perception of gender in voice assistants. *Perception* 3 (2019), 22–2019.
[22] Sarah Perez. 2018. Alexa developers get 8 free voices to use in skills, courtesy of Amazon polly. https://techcrunch.com/2018/05/16/alexa-developers-get-8-free-voices-to-use-in-skills-courtesy-of-amazon-polly/#:~:text=Amazon%20Polly%20%7C%20TechCrunch-,Alexa%20developers%20get%208%20free%20voices%20to,skills%2C%20courtesy%20of%20Amazon%20Polly&text=Now%20Alexa's%20voice%20apps%20don,of%20the%20Amazon%20Polly%20service.
[23] Elizabeth Ann Strand. 2000. *Gender stereotype effects on speech processing.* The Ohio State University.
[24] Selina Jeanne Sutton. 2020. Gender Ambiguous, Not Genderless: Designing Gender in Voice User Interfaces (VUIs) with Sensitivity. In *Proceedings of the 2nd Conference on Conversational User Interfaces* (Bilbao, Spain) *(CUI '20)*. Association for Computing Machinery, New York, NY, USA, Article 11, 8 pages. https://doi.org/10.1145/3405755.3406123
[25] Selina Jeanne Sutton, Paul Foulkes, David Kirk, and Shaun Lawson. 2019. Voice as a Design Material: Sociophonetic Inspired Design Strategies in Human-Computer Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300833
[26] Suzanne Tolmeijer, Naim Zierau, Andreas Janson, Jalil Wahdatehagh, Jan Marco Leimeister, and Abraham Bernstein. 2021. Female by Default? – Exploring the Effect of Voice Assistant Gender and Pitch on Trait and Trust Attribution. 1–7. https://doi.org/10.1145/3411763.3451623
[27] Mark West, Rebecca Kraut, and Han Ei Chew. 2019. I'd blush if I could: closing gender divides in digital skills through education. (2019). https://unesdoc.unesco.org/ark:/48223/pf0000367416.page=1
[28] Philipp Wintersberger, Paul A. Green, and Andreas Riener. 2017. Am I Driving or Are You or Are We Both? A Taxonomy for Handover and Handback in Automated Driving.
[29] Priscilla N. Y. Wong, Duncan P. Brumby, Harsha Vardhan Ramesh Babu, and Kota Kobayashi. 2019. Voices in Self-Driving Cars Should Be Assertive to More Quickly Grab a Distracted Driver's Attention. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Utrecht, Netherlands) *(AutomotiveUI '19)*. Association for Computing Machinery, New York, NY, USA, 165–176. https://doi.org/10.1145/3342197.3344535
[30] Heather Suzanne Woods. 2018. Asking more of Siri and Alexa: feminine persona in service of surveillance capitalism. *Critical Studies in Media Communication* 35, 4 (2018), 334–349. https://doi.org/10.1080/15295036.2018.1488082 arXiv:https://doi.org/10.1080/15295036.2018.1488082