Driving in the future: a study on automotive interaction without a steering wheel

Dennis Sbircea
HCI-E MSc Final Project Report 2017
UCL Interaction Centre, University College London
Supervisor: Dr Duncan Brumby

ABSTRACT
The advent of autonomous cars promises to bring disruptive change to the world of driving. Already, concepts for future fully automated vehicles indicate a trend towards abandoning the steering wheel. Consequently, exploring future interaction models becomes increasingly important for the automotive domain. This dissertation aims to determine the most relevant user preferences, expectations, and concerns to consider when designing new interaction models for driving tasks in autonomous vehicles without traditional controls such as steering wheels. To do so, a mixed-methods research approach was taken. Participants completed a lane change task experiment in a simulated autonomous vehicle using self-generated commands, either unimodal (gestures or speech) or multimodal (gestures and speech). Self-reported measures of automation trust were also taken throughout the experiment by means of a 7-point Likert scale. Once finished with the experiment, participants then took part in a semi-structured interview probing their impression on the simulated experience, their preferences of modalities and steering commands, and their error recovery behaviour. Our findings indicate that most participants would be willing to try driving in a car lacking a steering wheel despite reporting limited trust and that they would prefer using gestures to direct the car. There was also strong agreement on what specific form commands should take and how errors should be handled. Finally, the implications of the results are discussed with relevance to research and automotive interaction design.

Author Keywords
Automotive user interfaces; autonomous vehicles; gesture interaction; speech interaction; automation trust; error recovery; multimodal interfaces; interaction elicitation.

ACM Classification Keywords
H.5.2 User Interfaces: interaction styles; I.2.9 Robotics: autonomous vehicles.

MSC HCI-E FINAL PROJECT REPORT
Project report submitted in part fulfilment of the requirements for the degree of Master of Science (Human-Computer Interaction with Ergonomics) in the Faculty of Brain Sciences, University College London, 2017.

NOTE BY THE UNIVERSITY
This project report is submitted as an examination paper. No responsibility can be held by London University for the accuracy or completeness of the material therein.

MSc Contribution Type
Empirical.

1. INTRODUCTION
Do cars need to have a steering wheel? Currently, it seems inconceivable that cars without steering wheels can be anything more than science fiction. However, concepts for future fully automated vehicles often do not have steering wheels. With automation, drivers could simply input a destination and let the car handle all steering functions while they work or relax. Despite being unencumbered by the task of steering, drivers may still want, for whatever reason, to change lanes, pull to the side of the road, and so on. In this context, how would they do so?

In recent years, the automotive industry has witnessed significant advances in the development of autonomous vehicles, with their long-promised advent now seeming inevitable in the following decades. The advances have translated into a wide range of automated safety and efficiency features found in cars today (e.g., cruise control, automatic headlights), meaning drivers must now negotiate with a rapidly increasing number of self-operating systems. This trend is sustained by automotive manufacturers who, to remain competitive, also use automation to expand the comfort and entertainment opportunities in their cars (Campbell, Egerstedt, How, & Murray, 2010). Furthermore, expectations of interactions with in-car technology are undergoing a profound change with the emergence of a new generation of drivers (Owens, Antin, Doerzaph, & Willis, 2015).

Driver-vehicle interactions are evolving in a fundamental way: cars are no longer seen as an extension of the driver, controlled continuously to travel from A to B. Instead, vehicles are now a distinct entity with which control is shared (Pettersson & Ju, 2017). It has been suggested that, ultimately, fully automated vehicles could make it unnecessary for drivers to focus on the transportation task at all (Kun, Boll, & Schmidt, 2016). This creates new challenges to vehicle interaction design due to how human-system relationships change the more the system gains agency (Flemisch et al., 2012). The shift in attitude is reflected by the automotive industry’s diversity of approaches to increased automation.

One notable issue regarding automated vehicles is the necessity of the steering wheel, with new concept models featuring a retractable steering wheel (e.g., BMW’s iNEXT)
or no steering wheel at all (e.g., Google’s Waymo). Indeed, it is possible that the automated vehicles of the future will not have a steering wheel due to safety concerns. In Schaefer and Hill’s study (as cited in Schaefer & Straub, 2016), the authors note people’s tendency to intervene in the autonomous control settings or fully take over control even when the vehicle is functioning at an optimal performance. Research has also shown that drivers taking a supervisory role in automated vehicles are confronted with a more complex driving task as the number of driving processes for which they are responsible is increased (Banks, Stanton, & Harvey, 2014). In this context, Schaefer and Straub (2016) have suggested that removing traditional control systems may reduce feelings of responsibility specific to vehicle control and the number of inappropriate interventions that could impact car safety.

As there is no definite consensus in the literature on the terminology used to classify the various activities performed in the car, it is necessary to clarify that this dissertation makes use of the distinction suggested by Pfleging and Schmidt (2015). Therefore, in lieu of using the traditional trisected definition (i.e., primary, secondary, and tertiary driving tasks), this dissertation will refer to all activities related to safely controlling the vehicle (e.g., lateral and longitudinal control, activating wipers) as “driving tasks”, and to all activities not related to driving (e.g., operating comfort or infotainment systems, eating, communicating with people) as “non-driving tasks”. This approach is taken here as traditional definitions become obsolete with increased vehicle automation.

Even in fully automated vehicles, users might still want the ability to control specific aspects of the driving task, such as commanding the car to change lanes, to pull to the side of the road, to change destinations, and so on (Niculescu, Dix, & Yeo, 2017). Car manufacturers wishing to give users the option to do so safely and easily, with minimal disruption to any non-driving task they might be engaged in, might therefore be searching for alternatives to the steering wheel and pedals. Multimodal interaction involving speech and gesture commands might be the answer, as previous work suggests its performance is comparable to traditional interaction in a lane change task (Pfleging, Schneegass, & Schmidt, 2012).

If an alternative to the steering wheel is to be proposed, which modality would be the most appropriate? What form should specific commands take? To what extent would users be open to this new interaction model? How would they handle the system in the event it became unresponsive? Currently, there is a gap in the literature regarding these questions and about the broader topic of multimodal interaction for driving tasks in automated vehicles. Given the vastness of the subject, an exploratory study is necessary to determine the most promising directions for future research and interaction design.

In this dissertation, an empirical approach is taken with the aim of gauging user preferences regarding multimodal interaction in a common driving task, eliciting alternative commands for this scenario and determining the most intuitive ones, and understanding error recovery behaviour. To gain a comprehensive view of the subject and why it is worth investigating, relevant literature on automotive user interfaces, driver error, automation trust, and interaction elicitation is reviewed. Next, details regarding the dissertation’s methodology are presented, including the driving simulation experiment based on a lane change task, the automation trust scale administered after each experimental condition, and the semi-structured interview probing participants on their experience. A detailed account of the quantitative and qualitative findings is then offered. Finally, the results are discussed in terms of the dissertation’s aims, relevant past work, and implications for future research and design.

2. PAST WORK

The research question addressed in this study is what are the most relevant user expectations, preferences, and challenges to consider when designing new interaction models for driving tasks in autonomous vehicles without traditional controls such as steering wheels. To understand this problem at a conceptual level, but also to ascertain the most appropriate methodological approach for it, relevant prior work will be reviewed. Four somewhat dissimilar research areas will be covered in the context of automated vehicles: automotive user interfaces (gesture, speech, and multimodal), driver error, automation trust, and interaction elicitation. The following literature review will contend they are relevant to determining the focus of this exploratory study and the specific form of its experimental design.

Gesture interfaces

Gestural interaction is regarded as a natural means of interaction with computer systems because it takes advantage of people’s innate ability to communicate via body language (Angelini et al., 2014). Indeed, in their observational study of small groups around drawing interfaces, Tang (1991) noted how people use gestures efficiently to communicate significant information such as expressing ideas, mediating group interaction, and demonstrating a sequence of actions. To further underline the innateness of this modality, users seem to have similar ideas on gesture kinematics and on mapping gestures onto commands. In a gesture elicitation study for smartphone commands focusing on guessability, Ruiz, Li, and Lank (2011) demonstrate broad agreement among their participants on gesture sets, suggesting consistent logical mappings. These mappings may be relevant to automotive interaction design if they extend to paradigms beyond mobile computing.

Systems based on gesture interaction are becoming ubiquitous because of the revolution in consumer
electronics, which has brought us phones, tablets, large displays, and even laptops capable of multi-touch recognition. Furthermore, third-party systems such as the Nintendo Wii and the Xbox Kinect have emerged, which are capable of recognising gestures made in three dimensions and of understanding additional aspects of kinematics such as depth. Regarding automotive interaction, many studies have investigated the prospect of making gestures on the steering wheel itself (Angelini et al., 2014) or on a central touchscreen (Döring et al., 2011), while others have explored opportunities to perform the gestures in mid-air (Rahman, Saboune, & El Saddik, 2011). In their paper, Ruiz, Li, and Lank (2011) draw a clear distinction between surface gestures (i.e., performed in two dimensions on a touchscreen) and motion gestures (i.e., performed in three dimensions using a controller or captured by a camera); though this distinction is made for the purpose of mobile interaction research, it is also relevant for in-car interaction research, as it is unclear which would be preferred among users. It seems, therefore, that gesture recognition technology has advanced sufficiently to be used for automotive interaction.

Gestural interaction has gained increased interest in recent years with regards to the research and design of in-vehicle interfaces. Particular attention has been paid to introducing gestures to manipulate non-driving tasks such as accessing infotainment and entertainment features (Koyama et al., 2014; Ma, Xu, & Du, 2016). This is partly because research has suggested they can reduce the driver’s cognitive and visual load (Bach, Jæger, Skov, & Thomassen, 2008; Riener, 2012). In their study, Bach et al (2008) conducted an experiment comparing tactile, touch, and gesture in-vehicle interaction techniques, focusing on the effects on drivers. Their results demonstrate that gesture interaction implies significantly fewer short-duration eye glances to perform assignments and no long-duration (above 2 seconds) glances at all. However, the above studies all assume a non-automated scenario in which the driver is continuously engaged in the driving task.

The current study aims to explore the use of gestures for driving task manipulation when the steering wheel is removed. To do so, gestural interaction will be investigated in a distinct experimental condition, independent of other modalities. Furthermore, the question of how gestures should best be performed remains. The present study aims to shed light on this by giving users the opportunity to use both surface and motion gestures and recording their choices; as no steering wheel will be provided in the experiment, surface gestures will be made directly on the simulation display.

**Voice interfaces**

Speech input is sometimes regarded as the optimal modality for information entry in a multitasking setting such as driving due to its lack of competition for visual resources (Tsimhoni, Smith, & Green, 2004). Though processing human language is a notoriously difficult and complex task for computers to perform, tremendous progress has been made in this domain in recent years.

Today, systems making use of voice interaction are becoming increasingly common in consumer electronics with the introduction of AI conversational agents such as Siri, Cortana, or Google Now. Luger and Sellen (2016) compared these three platforms in a qualitative study aiming to shed light on interactional factors affecting everyday use. At a broad level, their interview-derived findings indicate that users tend to use and regard the systems in a way that overestimates their intelligence, capabilities, and goals. Despite user expectations of conversational agents remaining far from the practical reality of use (Luger & Sellen, 2016), given a sufficiently simple interaction model, voice interfaces could be appropriate for automotive interaction.

Speech interfaces have been widely implemented in vehicles for selected features (e.g., for making a call or inputting a destination in the navigation system). Previous research has suggested that they are potentially easy and safe to use in an automotive context due to their hands-free nature (Winter, Grost, & Tsimhoni, 2010). However, it appears that only a minority of drivers use speech input regularly due to various reasons such as the increased cognitive load and effort for learning and remembering commands (Pickering, Burnham, & Richardson, 2007).

With regards to research, particular focus has been placed on investigating the extent to which using voice commands to complete a secondary, non-driving task (e.g., dictating emails, inputting destination address) can distract drivers from their primary transportation task (Lee, Caven, Haake, & Brown, 2001; Tsimhoni et al., 2004). In a study by Jamson, Westerman, Hockey, and Carsten (2004), a driving simulator experiment was conducted to explore the effects of a speech-based e-mail system on driving performance in a car-following task. In support of the above studies, their findings suggest that participants showed poorer anticipation of braking requirements and shorter time to collision when using the speech-based system compared to baseline performance (i.e., no e-mail system).

The current study aims to shed light on the use of voice commands for driving task manipulation in an automated vehicle context. To do so, speech interaction will be investigated in a distinct experimental condition, independent of other modalities.

**Multimodal interfaces**

Multimodal systems are defined as “those that process two or more combined user input modes – such as speech, touch, manual gestures, gaze [...] – in a coordinated manner with multimedia system output” (Oviatt, 2009). For the purposes of this review, only past studies discussing multimodal interfaces combining speech and gestures will be considered. The advantages of employing multimodal
interfaces lie, among others, in presenting an interaction closer to human-human communication and in increasing the robustness of interaction by using redundant or complementary information (Reeves et al., 2004). Multimodal interfaces are gaining increased importance due to how far relevant software and hardware have advanced, how the disadvantages of any single modality can be overcome by combining them appropriately, and how naturally multimodal interaction fits with now-ubiquitous mobile devices (Cutugno, Leano, Rinaldi, & Mignini, 2012; Turk, 2014). It seems, therefore, that multimodal interfaces are likely candidates for automotive interaction.

Within the broad topic of multimodal interaction, there is also the issue of how speech and gestures should be combined. Drawing a clear distinction of multimodality, Oviatt (2009) describes three styles for combining different modalities: fused modalities (i.e., multiple modalities playing a part in a single interaction step), temporally cascaded modalities (i.e., each modality supplying partial information in a specific sequence), and redundant modalities (i.e., each modality is available in each interaction step). The current study aims to address the interaction style issue by giving participants the freedom to use any style and recording their preferences.

With regards to in-vehicle interfaces, there have been attempts in research and in industry to introduce multimodal interaction combining gestures with voice commands and study their feasibility (Müller, Weinberg, & Vetro, 2011; Pfleging et al., 2012). In the study by Pfleging and colleagues (2012), participants were presented with a prototype employing a temporally cascaded style to perform various non-driving tasks, which, when compared on a lane change task against baseline interaction, was not revealed to be significantly more or less distracting and was even rated ‘good’ on the System Usability Scale. However, the notable “put-that-there” prototype by Bolt (1980) also makes a strong case for fused modalities. The above studies, however, all assume a non-automated scenario and focus on investigating multimodal interaction for non-driving tasks such as making phone calls, composing emails, and so on.

The present study aims to explore the use of multimodal commands to manipulate driving tasks in an automated vehicle context. To do so, we will investigate gesture and voice commands used together in a distinct experimental condition. This should, therefore, shed some light on behaviour and preferences within each modality but also on whether multimodal commands are preferable to unimodal ones.

We can infer that the interface modalities discussed in the above sections (gesture, speech, multimodal) have the potential to play an important role in a new interaction model aiming to replace the steering wheel. Also, this review has underlined some of the gaps in the literature surrounding this issue which it will address. However, there are further relevant factors to be considered if a more comprehensive view of user behaviour is to be obtained. For instance, what should a user do if the autonomous vehicle is not responding as expected, or if the wrong command was issued? Thus, the following sections will review the literature relating to driver error.

**Driver error**

Driver error is considered the prime reason behind over 90% of all automobile crashes and is believed to contribute to crash occurrence or injury severity even when the main reason behind the crash is attributed to the vehicle, road, or environment (Rumar, 1990). Whatever the reason, automobile crashes are the source of immense personal tragedy and economic cost all around the world. Fortunately, the advent of autonomous vehicles has the potential to fundamentally change the transportation landscape by not only averting fatal accidents, but also increasing road capacity, lowering emissions, and providing independence to the elderly and disabled (Fagnant & Kockelman, 2015).

Despite this promising future and even more promising safety record of automated vehicles tested on the road today, the possibility of such a system committing a potentially disastrous error still exists and therefore must be considered. For example, a Tesla Model S driver was fatally injured last year after crashing into a tractor-trailer crossing the highway (Yadron & Tynan, 2016). As neither the autopilot nor the driver noticed the white side of the incoming vehicle against the brightly lit sky, the brakes were not applied. Cases such as this are especially relevant in the context discussed in the present study, in which the driver does not have access to a steering wheel.

Regardless of the modality used in our scenario, errors could occur due to a failure of the in-car sensors employed to detect and interpret the driver’s input, or due to the driver making an incorrect command because of distraction, cultural differences, or physical constraints. In such cases, time is essential to a successful error recovery, both in terms of time needed to determine the next best course of action and in terms of time needed to perform that action.

Currently, handover of control from the vehicle to the driver is considered the optimal approach to error, with extensive research being conducted to determine the best conditions for handover. For instance, Van Der Heiden, Iqbal, and Janssen (2017) conducted a driving simulator experiment in which participants performed tasks on their phone while the car is in autonomous mode. The authors compared handover performance across three conditions, manipulating the presence and type of pre-alerts; they concluded that including pre-alerts resulted in safer handover situations as it caused people to pay more attention to the road and disengage from the non-driving task earlier.
Despite efforts to make semi-autonomous vehicles safer by optimising the handover process, the question of how errors should be handled in fully automated vehicles lacking the traditional driving interface remains. To shed light on optimal error recovery procedure in this context, research must first gain insight into the most common go-to reactions participants might have when confronted with errors. The current study aims to address this gap and uncover new directions for further research by presenting participants with ‘artificial’ errors throughout the experiment and recording their responses. For experimental consistency and participant safety, this approach demands the use of a driving simulator.

Aside from the above, error is also an important factor to consider because it affects users’ trust in the automated vehicle to take them to their selected destination safely and efficiently. Automation trust is in and of itself worthy of consideration because, like error, it also affects driver behaviour, as the following section will discuss.

**Automation trust**

Following the revolution in consumer electronics, automated systems have become ubiquitous, with many being highly trusted by users. Though this trust has improved the efficiency of many processes, automation-related accidents are still common occurrences. While a part of these accidents are caused by system failures, they are also sometimes caused by humans over- or under-trusting the automation (Hoff & Bashir, 2013). Furthermore, past research has demonstrated that automation trust as a variable is not always static, but that it changes over time according to a range of factors (Yang, Unhelkar, Li, & Shah, 2017); this is especially the case for emerging technologies. It is not surprising, then, that some users may be reluctant to trust an automated vehicle given the risk involved if the system ever failed.

In their study investigating user requirements for autonomous vehicles, Niculescu, Dix, and Yeo (2017) conducted moderator-led discussions with the aid of materials (i.e., video documentaries, prototypes, and sketching) to encourage participants to share ideas. They uncovered several high-impact factors related to in-car safety and driver empowerment that, more broadly, paint a detailed picture of users’ trust in automated driving systems. For instance, users reported trusting the system more if they had good awareness of its capabilities, road obstacles, and the system’s backup plans in case of failure, indicating that availability of information affects automation trust. Notably, parents reported having particularly low trust with regards to allowing their children to use autonomous vehicles alone due to concerns about car hijacking or sabotage. Furthermore, many of their participants reported a need for the ability to manipulate the driving task if so desired (e.g., if the car is believed to be taking the wrong path or driving too slowly), indicating that trust is related to the presence of full or partial manual override.

While prior work has investigated automation trust in autonomous vehicles at a broad level, the literature has yet to investigate this topic at the more detailed level of specific interaction models. Crucially to the present study, there has been no work on which interaction modality would be most trusted by users for this purpose. The current paper aims to shed light on this gap by probing participants on their trust following every experimental condition. As several earlier studies have reliably assessed automation trust with single-item trust ratings (Brown & Galster, 2004; Hergeth, Lorenz, Vilimek, & Krems, 2016), a corresponding approach will be taken here.

**Interaction elicitation**

Aside from probing user preference and behaviour with regards to the above topics, the present study also aims to gain an understanding of what users see as the most appropriate forms for commands manipulating a given driving task. To achieve this, the most promising methodological approach seems to be interaction elicitation, a technique emerging from the domain of participatory design in which users are informed on the desired effect of an action and asked to come up with commands to reach that effect (Morris et al., 2014). Interaction elicitation techniques are particularly useful because they often produce inputs that end-users prefer to those designed by HCI professionals, possibly because professionals tend to create inputs that are more complex physically and/or conceptually (Morris, Wobbrock, & Wilson, 2010). Furthermore, interaction elicitation allows for increased focus on the end users’ desires as opposed to ease of implementation (Morris et al., 2014), meaning that applying this technique in an automotive interaction context should yield intuitive and memorable input sets.

To reduce any legacy biases participants might have and to increase the novelty of the commands they produce, an approach suggested by Morris and colleagues (2014) is adopted in the current study: specifically, instructing participants to produce multiple interaction proposals to force them to move beyond simple, legacy-inspired techniques. This is optimal as previous research has shown that forcing designers to create a large set of initial ideas leads to better final designs (Dow et al., 2012).

As recording behaviour and eliciting commands for every possible driving task goes beyond the scope of this paper, it is necessary to narrow the focus to one specific driving task. In their multimodal interaction elicitation study, Pfleeging et al (2012) argue that the task used for elicitation should be a simple one, with a clear goal and no need for extensive practice on the participant’s part. The present study will emulate their choice of task and employ a Lane Change Task for the experiment, which is a standardized method used to compare driving performance across...
different conditions, usually including a secondary task (Petzoldt, Brüggemann, & Krem, 2014).

Finally, to encourage the elicitation of the widest variety of commands and to avoid restricting users to any specific interaction style, a multimodal command recognition system, capable of quickly and reliably processing any input, is required to be introduced to the autonomous vehicle simulator. As the development of such a system is many years away from becoming a reality, an alternative is needed. Therefore, the present paper will make use of the Wizard of Oz prototyping technique, which was originally established to research speech and natural language interfaces in an age where the underlying technology was not mature enough for testing purposes (Gould, Conti, & Hovanyecz, 1983). This technique has been extensively used in the automotive research community to investigate speech- and gesture-based interfaces (Alpern & Minardo, 2003; Lathrop et al., 2005), as well as user behaviour in autonomous vehicles (Lundgren et al., 2017), and should therefore be appropriate for the purposes of the present study.

In summary, this dissertation has so far reviewed past research on several key topics related to manipulating driving tasks in automated vehicles without the use of a steering wheel. We have identified three modalities as likely candidates to replace traditional interaction models, two unimodal and one multimodal. Firstly, the possible benefit of employing gestures is a certain level of innateness originating from people’s natural understanding of body language and the ubiquity of gestures in modern electronic devices. Secondly, speech interfaces are also considered due to them becoming increasingly intelligent, being already present in certain in-vehicle features, and allowing hands-free use. Thirdly, multimodal interaction combining speech and gestures might be preferable because of its flexibility and redundancy as shown by past work. The above topics must be considered if the aim of this study is to understand which modality is preferred by participants as a replacement for the steering wheel and why.

Aside from the question of modality, this dissertation has also reviewed other topics relevant to the scenario of driving without a steering wheel. As one of the aims of the study is to explore error recovery behaviour, driver error was discussed along with its implications and research-driven solutions. Next, the necessity of gauging automation trust in our specific scenario was discussed by reviewing research underlining its relevance to driver behaviour. Finally, regarding the aim of identifying the most appropriate user-defined steering commands for our scenario, we discussed the merits of interaction elicitation as a methodological approach, along with other techniques meant to increase elicitation quality.

3. METHODS

Participants
15 participants were recruited using an opportunity sample. To be eligible for testing, participants must have normal or corrected-to-normal eyesight, have no motor impairments in the upper limbs, and be fluent English speakers. The sample consisted of students from University College London with a mean age of 25.6 (Range = 22-34, SD = 3.45); of these, 10 were female. 13 participants possessed a driving license, while one did not and one was still in driving school. Participants were compensated £10 for their time.

Design
For this simulator experiment, a within-subjects design was adopted in which participants completed a simple Lane Change Task by giving commands to the car. The Independent Variables (IVs) were the three experimental conditions under which participants had to perform the same Lane Change Task, explained below:

In the Gestures condition, participants were instructed to only control the car via gestures in the air in front of them, or to use the simulator screen as a touch screen. Anything they said during this condition had no effect on the control of the vehicle.

For the Voice Commands condition, participants were instructed to only make use of voice commands: single words, small groups of words, or short phrases. Any gesture they made during this condition had no effect on the simulation.

In the Multimodal condition, participants were instructed to use both gestures and voice commands to control the vehicle. Participants were free to use these in any combination and any order they wished, as long as they used both in response to the lane change prompts. Any gesture or voice command made in isolation of its counterpart had no effect on the simulation.

The Dependent Variables (DV) were the inputs participants issued to complete the Lane Change Task. Data (video and audio) was collected by means of a video camera placed behind and to the side of the participant. Additionally, the trust participants had in the automation they worked with was probed with a 7-point Likert scale, on which a score of 7 would represent absolute trust in the automation presented in the simulator. Here, the IVs were the experimental conditions, while the DV were the scores participants gave on the provided automation trust scale. Finally, further qualitative data was collected through a 15-minute semi-structured interview administered after the experiment, detailed in the Procedure.

Materials
An experimental setup was used in which participants performed a Lane Change Task via a simulated vehicle. Figure 1 illustrates how the task equipment was placed relative to the participant. Subjects were not provided with
a direct means of longitudinal or latitudinal control (i.e., no pedals or steering wheel) to simulate a highly automated vehicle. The vehicle was controlled by the researcher using a Speedlink Strike NX gamepad behind the participants’ visual field as part of a Wizard of Oz technique. Participants sat in a Playseat Evolution racing seat with the front frame removed, approx. 0.5 meters from the screen, while performing the task to provide a more realistic driving feel.

For data collection, a Sony NEX-3 camera was placed on a tripod behind participants, giving an ‘over-the-shoulder’ view of their gestures and the simulator display, as well as being in good range to record their voice commands clearly. This data would then be analysed to note patterns, preferences, errors, common features, and performance in the participants’ responses. For the interviews, a Sony Xperia Z3 Compact smartphone was used to collect audio data. This data would then be analysed using NVivo software package.

Procedure
For each condition, participants performed two experimental blocks, each consisting of a track with 20 prompts for lane changes. The order in which the conditions were presented was balanced across participants, while the lane-change prompts were presented in a pseudo-random fashion (i.e., random order adjusted to prevent lanes from being targeted twice in a row) to avoid learning effects and to force subjects to change lanes after every sign gantry.

Before commencing, participants were reminded of the experimental procedure and what was expected of them; they were also reminded of their rights to withdraw and handed a consent form to sign. After gathering demographic information, participants were told they would take part in a lane change task on a 5-lane highway: throughout each block, they would be prompted 20 times by overhead signs to change to a specific lane. Participants could direct the vehicle by producing appropriate inputs according to the experimental conditions. In the Gestures condition, participants could do any kind of gesture that could reasonably be interpreted as a lane change command, either in mid-air (i.e., motion gestures) or on the simulation display (i.e., surface). In the Voice Commands condition, participants could produce any input that featured basic grammar (i.e., single words or small groups of words). In the Multimodal condition, participants were required to use both gestures and voice commands, but could do so in any style, order, or combination. Participants were instructed to avoid reusing commands given in the first block during the second block of every condition; the purpose of this is to elicit more diverse input. They were also asked respond to
the prompts as accurately and quickly as possible, to make their gesture inputs in clear view of the camera, and to give their voice commands in a clear voice, one voice command at a time.

Finally, they were told that the simulator was a prototype and that there might be errors with the command interpretation software. This means that within every experimental block performed, the researcher would occasionally (i.e., 1-3 times per block) trigger the simulator to intentionally disobey the participants’ commands, either by taking no action, overshooting, or doing the opposite of the command. Participants were instructed to continue attempting the task regardless of how the vehicle behaved and to not let the errors discourage them from trying new inputs if they wished. The purpose of these interruptions was to introduce participants to situations from which they had to recover so that the commands they use to do so could be recorded. The errors were timed in such a way as to give participants enough time to reach their target if they responded quickly.

Once done with instructions, participants were invited to sit in the driving seat of the simulator, facing the monitor. Participants were required to perform at least one training block at the beginning of the experiment to practice their inputs and get a feel of vehicle responsiveness. Following this, the experimental blocks were presented. As soon as the driving task started, a multi-page instructions screen appeared (Figure 3), reminding the participant of the experimental procedure and their task. Participants could commence the block by pressing the “Start” button on the final instruction page.

![Figure 3. Screenshot of the on-screen instructions, here showing second page.](image)

The impression of a prototype automated vehicle simulator was given by employing Wizard of Oz technique, which consisted of the researcher changing lanes for the participant without their knowledge. To prevent participants from becoming aware of the Wizard of Oz technique, the researcher sat behind them, controlling the car with a game controller under the table with one hand while pretending to be texting with the other hand. If participants asked about how the command interpretation software worked, they were shown a MiVue 538 dashboard camera pointed towards them (switched on, not recording) and told they would see the software itself after the experiment. The researcher would mostly ignore the lane change prompts (except to be aware of the participants’ target lane) on the screen and instead follow the participants’ inputs closely, including going off the road if participants were, for example, on the far-left lane and commanded the vehicle to go further left.

Once the vehicle crossed the track finish line, the simulation ended. Participants were given a few minutes’ break while the new experimental block was prepared. If subjects had just completed a condition, they were also administered the automation trust scale, which took the following form: “On a scale of 1 to 7, 1 being “not at all” and 7 being “completely”, to what extent do you trust the automation to help you complete the task successfully?”. This single 7-point Likert scale was administered after each condition for modality-specific responses.

Once all the blocks were completed, participants were invited to take part in a 15-minute semi-structured interview about their experience. The first part of the interview was designed to gather information about the impression participants got of the simulation in general (e.g., “How did it feel to have control over a car yet no pedals or steering wheel?”, “How authentic was the simulation?”), about their modality preferences (e.g., “What was your favourite modality and why?”, “Which modalities were the most and least difficult to work with?”), about error recovery (e.g., “Was error recovery more or less difficult in any condition?”, “Why do you think the errors occurred?”), and about their impression of their own performance (e.g., “How much would you say you experimented with different inputs throughout the study?”, “How satisfied are you with your choice of inputs?”).

The second part of the interview focused on asking participants to reflect on their input choices in each experimental condition, gathering information on how the choices were made (e.g., “Why did you choose the gestures that you did? Based on what criteria?”), input preference (e.g., “Which specific input fit the task best?”), and whether their experience with any existing technology affected their choices (e.g., “Were your choices inspired from existing technology?”); modality-specific questions were also presented (e.g., “Did you prefer making gestures in the air or on a surface? Why?”, “What multimodal interaction style did you prefer? Why?”). Finally, participants were invited to imagine the system they used to control the vehicle was refined and introduced as a feature in a fully automated vehicle, then were asked whether they would use it, trust it, and prefer it over changing lanes manually. Once participants completed the experiment and the interview, they were informed about the Wizard of Oz technique used for the simulation, debriefed on the study, and handed their compensation.
4. RESULTS

Automation trust ratings

Participants completed a single-item questionnaire that asked “On a scale of 1 to 7, 1 being “not at all” and 7 being “completely”, to what extent do you trust the automation to help you complete the task successfully?”. Responses were given on a 7-point Likert scale and can be taken to reflect the level of trust in the automation. Across all conditions, participants gave an average score of 4.96 out of 7 (SE = .19), representing a ‘minor’ level of trust. Pairwise comparisons, illustrated in Figure 4, revealed that participants had significantly ($p = .045$) higher trust in the automation following the Multimodal condition ($M = 5.2$, $SE = .26$) than following the Voice Commands condition ($M = 4.53$, $SE = .25$). Trust following the Gestures condition ($M = 5$, $SE = .27$) was not significantly different than after the other two conditions (all $p > .204$).

A one-way repeated measures ANOVA was run on the Likert scale data to determine whether interaction modality affected participants’ trust in the automation to help them complete their task successfully. The analysis revealed that our manipulations of modality had no significant effect on participants’ self-reported level of trust in the automation, $F(3, 42)=2.43$, $p = .079$.

![Automation trust](image)

Figure 4. Mean trust ratings for each modality condition. Error bars represent Standard Error.

Observational data

In this study, participants were recorded interacting with the simulator via gestural, speech, or multimodal (gestures and speech) commands to complete a lane change task. To analyse this data, a thematic analysis was performed in an iterative process on the video recordings of participants giving their commands. As each recording was viewed, all observed inputs were described in detail; particular attention was given to the form of the interaction (i.e., kinematics for gestures, phrasing for voice commands) and to the mental model it hinted at (i.e., degree of user involvement, hand-direction correspondence). Following this, the commands were clustered into broad groups based on their type and common traits; more specific sub-groups were generated based on common variations of the main command types. Groups and sub-groups were refined with further video analysis. There was some loss of data from three blocks in the Voice Commands condition due to recording equipment failure; however, the voice commands and the participants’ behaviour were written down as soon as the data loss was discovered.

Before going into detail on performance in each condition, a number of general observations can be made on participant behaviour across all conditions. All participants had a good grasp of what the lane change task required them to do. To foster creativity, all participants were instructed to use different inputs in the second block of every condition; when prompted to change inputs, participants did so, but a majority only made small variations of the same input (e.g., making the same gesture with a closed fist instead of an open hand). Moreover, about half of the participants experienced some degree of left-right confusion, leading to errors in Voice Commands and Multimodal conditions when “left” and “right” were used verbally. There was no apparent difference in performance or behaviour between drivers and non-drivers. Finally, while care was taken to preserve the Wizard of Oz deception, four participants asked the researcher directly whether they were taking part in a Wizard of Oz experiment (noting the simulator’s ability to reliably interpret a wide variety of spontaneous inputs). However, further inspection of the data from these four participants revealed no obvious differences from the others. Having made these general remarks, the following sections will consider the observed behaviour in each different experimental condition.

Gestural steering commands

The commands given in the Gestures condition presented the greatest amount of diversity across all experimental conditions. The degree to which participants experimented with different inputs varied, ranging from only trying one main type of gesture (which was the most common) to four types. With one exception, all participants used their hands to complete their task in this condition. From the data, the following clusters emerged, presented in order of frequency of occurrence: signalling, pointing, steering, and lane-specific gestures. In the following sections, each of these clusters will be discussed in turn.

Signalling refers to using one’s hands to indicate the direction the vehicle should move towards. To give an example, Figure 5 shows a schematic representation of a typical signalling motion. Inputs in this cluster did not target a specific lane, but simply communicated changing lanes to either left or right. This was by far the most popular input in this condition, with 11 out of 15 participants using a variation of this type of gesture at one point or another. In all cases, signalling was performed by participants as a
motion gesture (as opposed to a surface gesture). Hand shapes varied, with most participants waving a relaxed open hand to indicate left or right, and pointing with the index finger in either direction coming second. Other hand shapes in this cluster included holding up both index fingers and flexing them, waving a closed fist, snapping fingers, and pointing with the thumb.

Further variations were observed in the signalling cluster. Seven participants performed signalling gestures in a passive manner (i.e., they would make the gesture, rest, then wait for the car to obey), while four made them in an active fashion (i.e., they would make the gesture then hold their hands in the end position until the car reached the target lane). A notable feature of passive commands was repetition, where participants repeated the same gesture once for every additional lane they wanted to change in that direction. Though there were instances where participants used both hands simultaneously, most participants performed signalling with one hand at a time. In all instances where the gestures were one-handed, correspondence between the hand used and the intended direction was observed (e.g., indicating left by waving left hand). This correspondence was followed strictly by five participants, while five others occasionally switched hands and one strictly followed a reversed correspondence.

Preferred to make a surface gesture, tapping the display surface on the target lane’s position, while two preferred to make a motion gesture, simply pointing in the air.

**Figure 5. Schematic representation of a typical signaling gesture directing the car from the middle lane to the far-left one.**

**Steering** refers to gestures that directly alter the course of the car with every motion. To give an example, Figure 7 shows a schematic representation of a typical steering motion. Inputs in this cluster are similar to signalling as they both communicate direction, but crucially they are more precise in that participants expected the car to closely follow every movement made. In all cases, participants performed steering in the air in an active manner, constantly adjusting the vehicle’s course and even turning the opposite direction once the target lane has been reached as if steering a car normally.

Six participants chose to make steering gestures at some point in the study, with moderate variation in terms of handedness and form. There was an even split between use of one and both hands to perform steering; when one-handed, participants tended to switch hands because of fatigue. The most popular form of steering encountered was holding both hands up and moving them as if turning an invisible steering wheel. Other forms included holding one hand up, turning it in the desired direction, and steering as if using a joystick; one participant notably performed steering by leaning their whole body towards the left or right.

**Figure 6. Schematic representation of a typical pointing gesture directing the car from the middle lane to the far-left one.**

**Pointing** refers to using one’s hands to indicate a specific lane the vehicle should change to. To give an example, Figure 6 shows a schematic representation of a typical pointing motion. The information communicated through this gesture is more precise than in signalling and the kinematics reflected this: in all instances, participants indicated the target with the index finger of their dominant hand in a short precise motion.

While six participants made pointing gestures at one point in the simulation, the commands in this cluster only varied in two respects. Firstly, five participants performed pointing in a passive manner (i.e., they would indicate the target lane, rest, and wait for the car to respond) while only one did so in an active manner (i.e., continuously pointing towards target, adjusting position of index to match the target from their point of view). Secondly, four participants...
Lane-specific gestures refer to gestures that participants performed to command the car to change to a specific lane, similar to pointing gestures. However, pointing gestures are directional, while lane-specific gestures are symbolic, using a specific hand shape for a specific lane. This type of command was only encountered in two participants, with one developing a specific gesture for the middle lane and the other developing a system of specific gestures for each lane. The middle lane gesture (i.e., holding open hand vertically towards the centre of the display) was used exclusively as an aid to signalling gestures in cases where the participant was unsure of exactly how many lanes had to be changed; once the vehicle reached the middle lane, the participant switched back to signalling for the remaining one or two lanes. The previously mentioned system refers to a set of gestures which the participant used to encode information on two levels: raising fingers to communicate a lane’s number from one to five and raising either their left or right hand to communicate whether the count started from the left or right. For example, if the participant held their right hand up with four fingers raised, this meant “fourth lane from the right”.

Voice steering commands
Participants in the Voice Commands condition could only issue verbal instructions and could not use gestures as in the above condition. The range of verbal commands given by participants will be considered in this section. Inputs given in the Voice Commands condition presented the least amount of diversity across all experimental conditions. Though participants did experiment with several forms of phrasing, only two main clusters of voice commands emerged, presented below in order of frequency of occurrence: naming direction and naming specific lane.

Naming direction refers to voice commands which communicate the participants’ intended direction. This cluster of commands was a very popular choice as all 15 participants tried it at some point. The most common phrasing encountered in this cluster was simply saying “left” or “right”; other forms, such as saying “go to the right” and replacing left/right with one/two, were also encountered but were not different at a conceptual level. All commands were uttered in a passive manner (i.e., speaking the voice command and waiting for the car to obey) with one exception in which the participant was constantly saying “left” or “right” to adjust the vehicle’s course (i.e., similar to steering). Participants varied somewhat in cases where they wanted to change multiple lanes with one command, using either repetition (e.g., “left-left-left”) or numbers (e.g., “three lefts”) to communicate it.

Naming specific lane refers to voice commands which communicate the exact lane the car should change to. Commands in this cluster, performed by 12 out of 15 participants, were exclusively passive and varied somewhat in terms of phrasing: the most common form was assigning numbers to each lane, one to five, left to right (e.g., “second lane” for centre-left lane). Other forms included “far-left”, “left shoulder”, and “middle lane” but were less common.

Multimodal steering commands
Commands given in the Multimodal condition did not differ significantly in form from the other two experimental conditions. Despite experimental conditions being presented in counter-balanced order, no multimodal inputs contained gestures or voice commands that did not fit any of the clusters discussed above. Gestures seemed to be marginally dominant between the two modalities, with several participants sometimes forgetting to perform the speech inputs during the lane change task. “Naming direction” voice commands were notably more popular choices in the multimodal condition compared to “naming specific lane” voice commands.

Participants varied in terms of the interaction style used, with most performing the gestures and the voice commands simultaneously. In instances where commands were given simultaneously but in a cascaded style, the most common was performing the gesture first, followed closely by the voice command.

There was no obvious pattern of using a particular interaction style with a particular type of multimodal command. From the data, the following clusters emerged, presented below in order of frequency of occurrence: signalling + naming direction, steering + naming direction, pointing + naming direction, steering + naming specific lane, signalling + naming specific lane, and pointing + naming specific lane.

Signalling + naming direction refers to multimodal commands combining signalling gestures and “naming direction” voice commands. With 11 out of 15 participants using a variation of this type at some point in the experiment, it is by far the most popular choice in this experimental condition. In most cases, both the gestures and the voice commands were done in a passive manner, though there were several instances of participants combining active signalling (i.e., holding their hand in the gesture’s end position until the target lane has been reached) with passive voice commands. In all but one case, participants encoded the same information in their gestures as their voice commands (e.g., waving twice towards the left and uttering “two left”). The exception to this was one participant who communicated direction with their gesture and number of lanes to change with their speech (e.g., waving once towards the right and uttering “three”).

Steering + naming direction refers to multimodal commands combining steering gestures and “naming direction” voice commands. In all four instances of participants making a variation of this input, steering was used constantly as if controlling the car directly while the voice command tended to only be uttered once at the beginning of the command. Participants varied in terms of the form they used for gestures, with three adopting a
steering wheel style and one adopting a joystick style. Voice Commands exclusively took the form of “left” and “right”.

**Pointing + naming direction** refers to multimodal commands combining pointing gestures and “naming direction” voice commands. The four participants who tried this type of multimodal input were very uniform in their approach: both the gestures and the voice commands were performed in a passive manner, all pointing gestures were made in the air only, and all voice commands exclusively took the form of “left” and “right”.

**Steering + naming specific lane** refers to multimodal commands combining steering gestures and “naming specific lane” voice commands. The two participants who tried this type of multimodal input were very similar on several accounts: the steering gesture was performed in an active fashion as if controlling the car’s course directly while the voice commands were uttered in a passive manner, all gestures adopted the steering wheel form, and all voice commands exclusively used a number system (i.e., one to five, assigned left to right) to select a lane.

Two other multimodal commands were observed during the experiment: signalling + naming specific lane and pointing + naming specific lane. However, each combination was performed by only one participant. All gestures and voice commands were made in a passive manner (i.e., giving the command once then waiting for the result), as opposed to an active manner (i.e., constantly using the command to adjust the car’s course). In both cases, the “naming specific lane” voice commands used a number system to identify each lane (one to five, left to right). Similar to both cases, the gestures were made without touching the surface and adopted the same form as in the other conditions (i.e., waving hand, pointing).

“Stay” commands

“Stay” commands refer to a type of command that was observed in every experimental condition. Participants used them as an aid to other inputs to ensure the vehicle would stay on the target lane once it was reached. They were performed by four participants in the Gestures condition, two in the Voice Commands condition, and four in the Multimodal condition. Notably, these were not always the same participants, with a total of seven participants making “stay” commands at any point during the simulation. In the Gestures condition, they exclusively took the form of a short wave towards the centre of the display with a vertically-held open hand. In the Voice Commands condition, participants mostly phrased the command as “stay” or “keep straight”. In the Multimodal condition, participants simply performed the above gestures and voice commands simultaneously.

“Stay” commands were used as a separate gesture after passive inputs and as a final part of active inputs. Overall, while there seems to be no obvious pattern of making “stay” commands more often with certain clusters of input, their use does seem to be related to error recovery and error prevention. The frequency of “stay” commands seemed to increase following errors; they were also observed when participants were experimenting with a new input.

**Error recovery**

All participants experienced errors in all three conditions, between one and three per block; these took three possible forms: ignoring the participant’s command, going to the opposite direction as the command, and going past the target lane. All participants displayed similar error recovery behaviour regardless of modality condition or how the errors manifested themselves: participants would always repeat the given command several times in a more pronounced manner. Participants’ gestures would become more deliberate and obvious, while the voice commands would become louder and clearer. If the vehicle overshot the target lane, participants would mirror their input to match the vehicle’s new position relative to the target. Except for “stay” commands, no other error-related commands were observed.

**Interview Data**

After completing the lane change task in all experimental conditions, participants were interviewed about their experience with the simulation and their preferences for steering commands. A thematic analysis was performed on the participants’ interview data. The analysis focused on understanding several aspects of the participants’ experience: how they felt about the idea of using alternative interaction models to change lanes, which modality and specific command they were most satisfied with, the context of their choices, and error recovery behaviour.

Participants’ impressions of the experience of controlling a car using gestures and/or voice commands were generally mixed, with six participants reporting a positive impression, five reporting a negative one, and the rest being undecided. Participants in the first group often cited the ease of interaction as their main reason:

“It’s much easier and can make me more relaxed. Maybe it’s not real driving, but I don’t have to use many parts of my body to drive.” [P12]

Participants who reported a negative impression often mentioned not feeling sufficiently in control:

“It’s awkward. I think even though the system is quite reliable, it’s still…I’m not in control. Of course I can command the system, but I have had a driving license for more than 10 years now; I feel more in control when I physically touch something and steer.” [P13]

When asked whether they would trust the system they used if it was refined and introduced as a feature in a fully automated vehicle, the majority of participants agreed that they would trust it enough to try it, but that they would need to be reassured of its safety and reliability first. When asked
whether they would prefer the system they used in the simulation to changing lanes manually, participants were split, with many adding they prefer both so that they could choose according to road conditions:

“... they felt most natural, because I'm a driver. The one where I did the steering wheel made me feel like I was in control and felt a bit safer. [...] I definitely trusted myself more with the gestures.” [P6]

Preference of interaction modality
With regards to participants' preference of modality, the gestures were the most popular by far. The most common reasons given for this included a lack of ambiguity and an increased feeling of control:

“... they felt most natural, because I'm a driver. The one where I did the steering wheel made me feel like I was in control and felt a bit safer. [...] I definitely trusted myself more with the gestures.” [P6]

When participants were critical of using gestures for lane changing, participants often mentioned long reaction times:

“I found that gestures took a long time to do unless your hand was near the screen.” [P1]

Three participants reported speech as their preferred modality, often citing hands-free control and increased speed:

“I preferred voice commands, because they're easier. I think using my hands all the time would be too tiring, doing this again and again - I'd rather sit like this, relax, and say what I want.” [P4]

Criticisms of employing voice commands for directing a vehicle included fatigue, confusing left and right, and having to interrupt music. Notably, some non-native English speakers felt disadvantaged when using voice commands:

“I didn't like using voice commands because English is not my native language and I have to think too much.” [P9]

Multimodal commands were the least popular, with only two participants reporting them as their favourite. They explained that combining gestures and voice commands offered increased flexibility and reliability:

“I think I liked the multimodal, because you could sort of combine your commands. For example, when I used the hand I could say "two lanes" and it would know because my hand was pointing to the left, then it would go two lanes to the left. Otherwise I would need to say a long sentence which includes the word "left".” [P7]

For many participants, however, combining gestures and voice commands seemed to increase fatigue and the likelihood of errors, either through giving conflicting commands or forgetting one part of the command:

“If I have to use both at the same time, I worry I will forget to do or say something. I think I prefer only one modality.” [P5]

In terms of which modality was considered the easiest to work with, most participants reported the gestures, motivating that they felt more intuitive and unambiguous:

“Gestures were the easiest because they were very intuitive and I didn’t have to think.” [P3]

In terms of which modality was considered the most difficult to work with, most participants reported the voice commands, citing difficulties counting the exact number of lanes to change or confusing left and right:

“I think voice commands were quite hard to work with. Sometimes you can't really express in a second how many lanes or what direction, sometimes your brain gets a bit confused.” [P7]

Preference of steering commands
Throughout the interviews, participants were asked which specific command was their favourite for the lane change task they had just completed. Responses did not differ from how frequent they occurred across participants. With regards to gestures, most participants reported signalling gestures, with steering being the second most popular. In terms of speech, more than half of participants preferred “naming direction” commands. Finally, most participants favoured multimodal commands that combined signalling gestures with “naming direction” voice commands, with pointing and “naming specific lane” multimodal commands ranked second.

Choice criteria: simplicity
Participants were also probed on what criteria they had in mind when creating the new commands. Across all three conditions, the top cited criterion was “simplicity”, as in making an input easy for the user to perform and easy for the system to understand. With regards to gestures, this often meant low physical effort and few parts to a motion:

“I think I chose them based on the effort. I always try to use the one which implies the least physical effort.” [P7]

In terms of voice commands, simplicity often meant being brief and using unambiguous phrasing:
“They must be very short, very simple. Also, it can be recognized by the machine easily, it’s not confusing. I think saying “three left” is the most direct way to control the car.” [P12]

With regards to multimodal commands, participants were mindful of whether the simplicity of the components was preserved once combined:

“They are the easiest on their own, and the easiest in combination” [P3]

Choice criteria: intuitiveness
The second most cited criterion for input elicitation was, across all conditions, how intuitive the commands felt. This is independent of simplicity because “intuitive” is a description of how the commands are understood while “simple” is a description of how they are presented. Despite this, there was no consensus on what exactly made commands feel natural or intuitive, with most participants agreeing on one point only:

“I don’t know, it just seemed natural.” [P2]

Choice criteria: modality-specific
The third most cited criterion for input choice differed according to the modality in question. For gestures, this was precision:

“I wanted to be correct. I want my car to be there and do so in the simplest and most accurate way.” [P10]

With regards to voice commands, participants valued speed after simplicity and intuitiveness:

“...they’re short and quick, especially saying “left/right”. For longer distances I just said “right-right-right-right-right.” [P11]

In the Multimodal condition, participants reported attempting to choose gestures and voice commands that complemented each other so that both could ‘confirm’ their intention:

“I think I just really wanted to drive home the point that I wanted to go in this direction, to try to add a bit more specificity.” [P1]

Legacy biases and inspiration
When asked whether their choices of commands were inspired from their experience with any existing technology, several participants suggested their gestures were inspired from video games (e.g., Kinect, Wii, racing games) and their voice commands from voice assistants (e.g., Siri, Google Now). Most participants, however, reported no particular source of inspiration.

Driver error
Participants were also probed on their impression of the errors experienced in the simulator. When asked how well they thought they recovered from errors, most reported having a good overall performance:

“I think it was pretty easy, unless I was really close to the gantry. So when I noticed that the car was going straight and I wanted it to go a bit further right or left, I just repeat my gesture. It wasn’t hard.” [P3]

Despite the above, all participants expressed concern over the system’s reliability at some point during the study, noting the errors as the cause. Some participants reported feeling disengaged from the simulation as a result:

“Probably the system is not 100% reliable, and understandably so. This is why I feel so disconnected; I expect the system to make mistakes.” [P13]

When asked about what caused the errors, most participants suggested it was a combination of themselves not knowing what the system best responded to and the system failing to recognize their commands correctly:

“I don’t know what kind of input library the system uses and I think the errors come from me testing the system.” [P10]

When probed on whether error recovery difficulty was the same across all experimental conditions, most participants reported recovery was easiest when using only gestures:

“It was easier with gestures, because when I wanted to turn left I just made a movement and then when I wanted to stop I made another movement.” [P15]

Conversely, they reported recovery was the most difficult when using only voice commands:

“It was easier with the gestures because it’s easier to use directional movements than thinking about which is left and which is right. I was at my worst performance when I was using the voice commands.” [P14]

5. DISCUSSION
In this interaction elicitation study, participants completed a lane change task experiment in a simulated autonomous vehicle using commands alternative to the steering wheel and pedals, either unimodal (gestures or voice commands) or multimodal (gestures and voice commands). After completing each experimental condition, participants reported their trust in the automation on a 7-point Likert scale. Having finalized the above, participants then took part in a semi-structured interview probing their impression on the experience, their preferences of modalities and steering commands, and error recovery behaviour.

In this discussion, the findings of the present study are related to past research and to the design of in-vehicle interactions to manipulate driving tasks. Furthermore, suggestions as to how this could provide direction for future research and benefit automotive interaction design are considered. Finally, the study’s limitations are examined.

There is a recent trend in the automotive industry to either ‘hide’ or remove the steering wheel from driving interfaces in new concept models. One of the aims of this dissertation
was to explore users’ feelings on controlling a car with gestures and/or voice commands instead of a steering wheel. Our interview findings revealed a variety of impressions, ranging from delighted, to neutral, to displeased. When prompted for details, most participants focused on the idea of empowerment: those who had a positive impression reported a more relaxed experience unencumbered by steering, while those who had a negative impression reported losing out on the experience due to relinquishing control. As such, it is possible that this variety of reactions to the presented scenario is a reflection of each participant’s eagerness to allow the car to perform the driving task with minimal driver input.

Our empowerment-related findings echo the ones of Niculescu, Dix, and Yeo (2017), whose study describes how lack of control in automated vehicles was a frequent issue among enthusiastic drivers. This user group often requested a more flexible car allowing passengers to manipulate lateral and longitudinal control, which matches what several of our own participants reported. The implication to automotive interaction design is that there are users who are happy to give up control and users who are not, though the proportions of these groups might change as new generations of drivers emerge (Owens, Antin, Doerzaph, & Willis, 2015). As such, if the industry’s goal is to expand the market penetration of fully automated vehicles by appealing to enthusiastic drivers, then introducing an optional steering wheel might be the optimal choice for now.

A further goal of this study was to investigate users’ preferred interaction model for manipulating driving tasks in autonomous vehicles lacking a steering wheel. In the above literature review, three likely candidates were suggested: gestures because of their innateness, speech due to hands-free interaction, and multimodal due to flexibility. During the interview phase of the study, participants were asked about their favourite interaction modality. Our findings reveal that gestural interaction was by far the most popular choice. Participants often reported that using gestures felt intuitive to them, as if using body language to communicate with a person. Gestures were regarded by participants as the easiest to work with and often helped them overcome spoken language barriers (i.e., left/right confusion, impulse to speak native language) encountered in the multimodal condition. Also, most participants reported error recovery was easiest when performed using gestures. Disadvantages were also mentioned by several participants, specifically that gestures sometimes implied longer reaction times compared to the other modalities.

Our findings regarding participants’ preference for gestures supports and reflects past research describing gestural human–computer interaction as natural (Angelini et al., 2014) as well as appropriate for in-vehicle interaction (Alpern & Minardo, 2003; Ma et al., 2016; Pfleging et al., 2012). The implication to automotive interaction design is that a gesture-based interaction model is the optimal alternative to using a steering wheel for driving task manipulations. However, in a Wizard of Oz study by Mignot, Valot, and Carbonell (1993) investigating multimodal (speech and gestures) interaction when using work-related applications, the authors concluded that gestures were mainly used for simple and direct commands, while speech was more often used for abstract commands. It is possible, therefore, that our findings are the result of the simple and direct nature of the lane change task favouring the use of gestures. As such, the preferred modality for a straightforward task like changing lanes might not be necessarily the same for a more abstract task such as changing the automated vehicle’s destination.

Speech and multimodal interaction were not our participants’ preferred choice for directing the car to change lanes without a steering wheel. Our findings indicate that the most common reasons behind this were fatigue due to excessive speaking, occasionally giving conflicting commands (especially left/right confusion), and, in the case of the multimodal condition, concerns about forgetting part of the command. As these issues are not exclusive to the lane change task, it is likely they will be relevant in design and future research exploring the use of speech and multimodal interfaces. Another factor to consider when including voice interaction in driving interfaces is language, as many of our participants who were non-native English speakers reported difficulties producing the commands despite being fluent. They cited requiring too much time to translate their desired input into English, especially when performing under the pressure and time constraints of errors.

Another goal of the present study was to shed light on what users see as the most appropriate forms of steering commands within each of the three modalities. To do so, two methodological approaches were employed: firstly, participants were recorded while eliciting commands for a lane change task as part of a driving simulator experiment; secondly, participants were probed on their input preferences and thinking during elicitation via a semi-structured interview. A diverse range of commands was produced this way. The results of the thematic analyses performed on both the observational data and interview data indicate strong agreement among participants with regards to input preferences. In the Gestures condition, participants favoured commands falling into the signalling cluster (i.e., using one’s hands to indicate the direction the vehicle should move towards). Also, there was a clear preference for making gestures in mid-air (i.e., motion gestures). In the Voice Commands condition, participants favoured inputs in the ‘naming direction’ cluster (i.e., communicating one’s intended direction). In the Multimodal condition, participants strongly preferred combining signalling gestures with ‘naming direction’ voice commands.
The findings concerning user preferences for specific steering commands within each modality suggest strong agreement among participants despite the variety of inputs observed. This is consistent with the results of the gesture elicitation study by Ruiz, Li, & Lank (2011), who also noted agreement among their own participants. Our findings support their suggestion that participants develop a consistent logical mapping when eliciting commands for a desired effect. For instance, our participants often took advantage of hand-direction correspondence (e.g., signalling left with left hand) when performing gestures, with a few even claiming in the interviews to have done so unconsciously.

The issue of logical mappings is not so unambiguous, however. Since in the present dissertation participants elicited inputs for a single driving task (i.e., changing lanes) whereas in the study by Ruiz and colleagues (2011) commands were elicited for a variety of mobile functions, the extent to which the logical mappings in both studies match is unclear. It is also unclear how much speech interaction also benefits from logical mappings, as past work has suggested they are better suited for abstract commands (Mignot et al., 1993). Further research in this area is therefore required. The implication for automotive interaction design is that there is potential for consistent logical mappings to be exploited to develop taxonomies of inputs. As noted by Ruiz et al (2011), the task of interaction design would then become one of shaping established taxonomies of commands while preserving the logical mappings of inputs to results.

The present dissertation aimed to gain insight into the behaviour drivers might exhibit when confronted with errors in an autonomous vehicle lacking a steering wheel. To do so, participants were exposed to ‘artificial’ errors throughout the experiment, meaning that the researcher would occasionally (i.e., 1-3 times per block) trigger the simulator to intentionally disobey the participants’ commands, either by taking no action, overshooting, or doing the opposite of the command. Our findings, derived from observational and interview data, indicate that all participants displayed similar error recovery behaviour across all modalities: repeating the given command several times in a more pronounced manner. Also, while some participants made use of “stay” commands to aid the recovery process, no other error-specific commands were observed (e.g., undo). It is possible this behaviour was the result of a lack of feedback from the system; participants were never sure of whether the system understood their command and, even if it did, how well it would execute it. However, it is also a possible result of a feeling of disengagement which some participants reported due to errors occurring regardless of modality or performance.

While most participants reported having an easy time recovering from errors, all expressed concern over the system’s reliability at some point during the study because of them. Many reported that if some of the errors in the simulation occurred in real life, the consequences would be dire, and that they would like to have a steering wheel available to recover from errors. Indeed, crashes like the one involving a Tesla Model S last year, in which the driver could have taken over control and avoided the trailer-truck manually when the system failed, are good arguments for keeping the steering wheel at least as an optional feature in the car. In the more distant future, however, it might be safer to remove the steering wheel, as Schaefer & Straub (2016) note in their study that people tend to intervene in the autonomous control settings even when the vehicle is functioning at optimal performance, and that removing the steering wheel entirely may reduce the number of inappropriate interventions.

Another goal of the current study was to gauge participant trust in the automation they worked with and to determine whether trust levels varied depending on interaction modality. Our results related to the automation trust ratings revealed that, overall, participants only had a minor amount of trust in the system to help them complete the lane change task successfully. Participants trusted the automation the most when using multimodal commands, followed by gestural, then speech interaction. No significant effect of modality on participants’ self-reported level of trust was observed. The findings are interesting as they suggest participants generally trusted their least preferred modality the most.

Multimodal interaction receiving the highest trust ratings is consistent with previous research suggesting that employing multimodal interfaces increases interaction robustness by using redundant or complementary information (Reeves et al., 2004). However, this does not help explain why our participants nevertheless preferred other modalities, and further research is required to shed light on this. In the interview, many subjects reported conditional trust in a hypothetical real-life version of the system: if it was sufficiently developed, if it followed strict safety standards, and – crucially – if the participants were aware of its safety records. This type of conditional trust is consistent with participant comments on autonomous vehicle safety in the study by Niculescu et al. (2017). The implication for automotive interaction design is, therefore, that users need to have explicit awareness of the vehicle’s safety capabilities before feeling reassured inside it.

Our findings contribute to a growing body of work on in-vehicle interactions in autonomous vehicles. The current dissertation is distinct from previous lines of research, however, in that we focus on issues surrounding driving task manipulation without a steering wheel. Our results show that most participants would be willing to try driving in such a scenario despite limited trust, that they would prefer using gestures to direct the car, that there is strong agreement on what form commands would take and how errors should be handled. In the following section, we
consider several important limitations of the current dissertation and some directions that future research in this area might examine.

Limitations and future research
Our results suggest that there might be value in using gestures to manipulate driving tasks in autonomous vehicles without a steering wheel. It seems that not only do participants prefer this modality, but that they also consider it less difficult and ambiguous. Future studies may therefore conduct a more in-depth investigation of the benefits and challenges of employing gestures for driving tasks.

As our findings indicate, there is a group of enthusiastic drivers for which empowerment in the driving task is strongly related to how enjoyable the driving experience is. Future work might investigate whether their needs can be reconciled with the surrender of control that autonomous vehicles imply.

Another relevant point relates to the experimental task. This dissertation employed a basic lane change task to elicit user-defined commands to direct a car without a steering wheel. While this may be sufficient for an exploratory study, further research in this area may employ tasks that are more representative of real driving situations. For instance, the simulation could prompt users to perform a range of operations (e.g., direct the car to park, to avoid a particular street, or to change destinations) while driving around a city. These operations should ideally vary in terms of complexity to explore a range of situations future drivers might find themselves in. Furthermore, employing a simulator that resembles a car more closely would foster participant engagement in the task.

One limitation of the present study concerns the automation trust scales. While the sample used may have been of appropriate size for the thematic analyses performed on the observational and interview data (as saturation was reached after 15 participants), it might have been too small to reliably analyse quantitative data. We performed a parametric one-way repeated measures ANOVA on the ordinal data from the Likert scales, which is not problematic assuming a normal sampling distribution (i.e., n > 30). While our data seems to be normally distributed upon inspection, the results of our quantitative data analysis may be skewed because of our sample size.

Our finding that multimodal commands are the least preferred choice among users for in-vehicle interaction is inconsistent with previous research demonstrating their relative success, such as the work by Pfleging et al. (2012). However, while this study shows that multimodal interaction is not significantly more distracting than baseline interaction and that it rates ‘good’ on the System Usability Scale, it does not take user preference into account. Furthermore, Pfleging et al. (2012) investigate multimodal interaction only for non-driving tasks and employ a researcher-defined interaction model. As such, it is unclear whether users were less pleased with multimodal commands in the present dissertation due to using them for a driving task, due to employing a user-defined interaction model, or a different reason entirely. Further work in this area is required to determine the feasibility of multimodal interaction for in-vehicle interfaces.

The introduction of errors in the driving simulator experiment, while helping achieve a better understanding of error recovery behaviour, has likely produced the unintended consequence of affecting participant choice during the interaction elicitation process. As implied by some comments made during the interviews, some participants might have been discouraged from trying an input again after encountering error the first time. Also, others might have tried to read too much into how and why errors occurred and adjusted their commands to fit their mental model of how the system worked. We suggest that future interaction elicitation studies in this area avoid a similar introduction of errors.

Finally, participants in this study were exposed to the driving simulator for approximately 25 minutes on average across all conditions and did not consider long-term adaptation to a driving interface lacking a steering wheel. During the interviews, about half of participants commented on how they felt they would have elicited more appropriate commands if given more time with the simulator. As users gain experience with a system, the way they regard it and interact with it may change. Further research is therefore required to shed light on long-term user relations with autonomous vehicles lacking a steering wheel.

6. CONCLUSION
The advent of autonomous vehicles is inevitable as technology advances. As this revolution takes place on the streets, the driving interface is unlikely to remain unchanged for long. Consequently, investigating automotive interaction design in this area becomes increasingly important. This dissertation set out to explore how drivers might still be in control in the possible future in which fully automated vehicles will present no traditional driving interfaces such as a steering wheel or pedals. A driving simulator experiment was conducted to better understand which interaction modality would be the most appropriate replacement to the steering wheel, how specific steering commands might look like, and how users might regard other relevant issues like error and automation trust.
Participants took part in an interaction elicitation study involving a lane change task, completed a brief questionnaire, and provided in-depth answers via a semi-structured interview. Our findings provide detailed insight into user preferences, expectations, and concerns, as well as a taxonomy of gestural, speech, and multimodal commands. Furthermore, our results are put into context and discussed with relevance to future research directions and automotive interaction design. The contribution of this dissertation is
the first exploratory study on the use of multimodal interaction to manipulate lateral control of a fully autonomous vehicle.

ACKNOWLEDGEMENTS
This project would not have been possible without the knowledge and guidance of Dr Duncan Brumby, who has proven to be an excellent supervisor. I wholeheartedly enjoyed researching this topic and would like to continue to do so in the future.

I would also like to extend special thanks to Florian Weber and the team at BMW Group, whose invaluable contribution has also made this project a reality. Danke schön!

My deep appreciation also goes to my parents and Coralie Perrin for their unending support and encouragement when I most needed it. It means the world to me.

Lastly, thank you to all my participants and colleagues who provided essential data and feedback for this project. It has been an absolute pleasure to be part of this community.

REFERENCES


Oviatt, S. (2009). Human-computer interaction. Design issues, solutions, and applications. CRC Press. Retrieved from https://books.google.fr/books?hl=en&lr=&id=tRVRK8UhuacC&oi=fnd&pg=PA85&ots=DBaQne1VCS&sig=guQ5A5FmXy-vDDMQEa-gD84YGh4#v=onepage&q&f=false


Multimodal Interaction in the Car - Combining Speech and Gestures on the Steering Wheel, (c), 155–162.


APPENDIX 1: INTERVIEW GUIDE

Thank you for taking part in our experiment. If you are ready, it is now time for the interview part of the study.

- How did it feel to have control over a car yet no pedals or steering wheel?
- How authentic was the simulation? Did it feel like a driverless car?
- What was your favourite modality? Did you prefer using gestures, voice commands, or both?
  - Please motivate.
- Which modalities were the most and least difficult to work with?
- To which modalities do you think the simulation responded best?
- How well do you think you were able to recover from errors?
  - Was error recovery more or less difficult in any condition?
  - Why do you think the errors occurred?
- How much would you say you experimented with different inputs throughout the experiment?
- How satisfied are you with your choice of inputs?
  - i.e. would you have required more testing before you settled on a set of inputs?
- Let’s talk about the gestures you used…
  - Did you prefer making gestures in the air or on a surface?
  - Why did you choose the gestures that you did?
    - Based on what criteria?
  - Which ones do you think fit the task best?
  - What do you think your best gestures all have in common?
  - Were your choices inspired from existing technology?
    - Which?
- Let’s talk about the voice commands you used…
  - Why did you choose the voice commands that you did?
    - Based on what criteria?
  - Which ones do you think fit the task best?
  - What do you think your best voice commands all have in common?
  - Were your choices inspired from existing technology?
    - Which?
- Regarding the multimodal condition…
  - What multimodal interaction style did you prefer?
    - Cascading? Simultaneous?
    - Please motivate.
  - Why did you choose the gestures and voice commands that you did?
    - Based on what criteria?
  - Which input do you think fit your task best?
  - Were your choices inspired from existing technology?
    - Which?
- If a developed and refined version of what you did today became a feature for driverless cars, would you use it? Would you trust it? Would you prefer it over changing lanes manually?
- Is there anything you would like to add?
APPENDIX 2: INTERVIEW TRANSCRIPTS

NB: interviewer text is underlined.

Participant 1
Thank you for taking part in our experiment, it is now time for the interview part of the study. So, how did it feel to be in control of a car yet no pedals and no steering wheel?

A bit worrying…umm, like at the start it didn’t seem to…it got a bit easier but then it was still unpredictable at times

How authentic was the simulation? To what extent does it feel like an autonomous vehicle?

It seemed kind of autonomous…I was worried about the accuracy of it; I wasn’t sure how it learned and how accurate it could be…

What was your favourite modality? Did you prefer gestures, voice commands, both?

I didn’t mind either, but I found that gestures took a long time to do unless your hand was near the screen and then voice seemed to be less predictable so it took more time to do voice. Gestures were my favourite.

Which modalities were the least and most difficult to work with?

Voice commands were more difficult because they were less predictable. It was easier to use gestures in some but not all cases. If I held my hand up and sort of tried to move the screen in the direction of my target, then that was a pain. If I just tapped my target lane, then that was fine.

How well do you think you were able to recover from errors?

In a simulation it’s OK, but if I was actually driving and like trying to desperately swerve, I would have probably been dead.

Was recovery more or less difficult in any specific condition?

Yeah. I think sometimes the voice was more difficult because it’s faster to say “right, right, right, right”…you’re not sure which ones of those are going to be recognized. Like with the gestures it’s pretty straight-forward, especially if you tap.

Why do you think the errors occurred?

I don’t know…I think it just wasn’t familiar enough. There wasn’t really good feedback…

You would have liked some sort of feedback then?

Yeah definitely. If it said, you know, “this is the lane you’re going to”, then I could change it, as opposed to waiting until the end when I overshot the target lane

What kind of feedback would you prefer? Audio? Visual?

Anything, anything. As long as it happens as soon as I issue the command so I can react quickly if I want to change.

How much would you say you experimented with different inputs throughout the study?

A bit, but I found it kind of…uh, started drawing blanks at some point. I think the tap one seemed like a good one.

How satisfied are you with your choice of inputs?

If there were only two modalities, like voice and gesture, I think I would feel more comfortable if they gave me a modality which the system recognizes really well, then I would just use that. I would be happy to adapt to whatever choice they make just as long as it works really well.

Let’s talk about the gestures you used…

Did you prefer making gestures in the air in front of the simulation or on the actual surface?

I really value precision over anything else; physically tapping where I want to go seemed the most precise so I would prefer the surface.

Why did you choose the gestures that you did? Based on what criteria?

I was trying to be different, trying to find different ways of doing it. Experimenting.

Which ones do you think fit the task best?

Tapping the target lane.

What do you think your best gestures all have in common?

They all have precision in common

Where your choices inspired from any existing technology?

Not really…can’t think of anything.

Let’s talk about the voice commands you used…

Why did you choose the voice commands that you did?

Just because…I chose the “right, right” and the “two rights” and commands just for precision and trying to be really quick so that if there is a problem I can correct it quickly.

Which ones do you think fit the task best?

In terms of voice…I was trying to think of a name for each lane; I thought of “left shoulder, right shoulder” and that sometimes didn’t work. I think the best thing for precision would be if there was a specific name for each lane.

What do you think your best voice commands all have in common?

Fast, fast, and precise, I think.

Where they inspired from any existing technology?

My mom backseat driving (laughs)

Let’s talk about the multimodal condition…
What multimodal interaction style did you prefer? Doing the gestures and the voice commands at the same time or one after the other?

I’m not really sure, I didn’t think too much about which one I should do first. I think I normally did the gesture first and if that didn’t work I would be more specific with the voice command, trying to add accuracy.

Why did you choose the combinations of gestures and voice commands that you did?

I think I just really wanted to drive home the point that I wanted to go in this direction, to try to add a bit more specificity.

Were your choices inspired from any technology?

Not really, no.

If a developed and refined version of what you did today became a feature for driverless cars, would you use it?

Maybe not. I think there are some lane changing systems out there apparently where you just click a button on the steering wheel right or left, and then the car changes lanes. I think that seems quite precise, you know you clicked it once. But if it’s just me pointing in the air, I’m afraid it might misread it and I’d just die.

Assuming you’re in an automated vehicle and, for some reason, you wanted to change lanes, would you prefer using any of the methods you used today compared to taking over control and changing lanes manually?

I’d trust it enough to try it. As long as it’s precise. I can imagine I’d rather prefer to have one of those systems, but at the moment pedal and wheel seem more reliable. If there was a button saying click which lane you want to change to, I would definitely prefer that as it would be much easier.

I get the feeling that you want a bit more structure in this kind of choice that you’re making. Am I right?

Yeah, just having like boxes in front of me, one for each lane, then you click a box and the car changes to that lane, that would be cool.

Participant 2

How did it feel to have control over the car yet have no wheels or pedals?

Boring…I like the ability to control the car manually.

How authentic was the simulation? Did it make you feel like in an autonomous car?

No, it seemed fake.

What was your favourite modality? Did you prefer gestures, voice commands, multimodal?

I preferred only gestures or only voice commands, I don’t like a combination of them. It makes me less focused.

Which modalities were the most and least difficult to work with?

Voice command was easy to work with, hardest was multimodal.

How well do you think you were able to recover from errors?

I think the simulation didn’t listen to me and I had to shout many times.

Was it easier or harder to recover from errors in any modality? Did it make a difference?

I think maybe voice was difficult to recover from.

Why do you think the errors happened?

Errors happened because I didn’t give the commands accurately.

How much would you say you experimented with different inputs throughout the study?

I think I tried mostly the same commands, just a few different ones.

Would you say the commands you used today are final?

Yeah, I’m satisfied.

Now let’s talk about the gestures...

Did you prefer making gestures in the air or on the screen?

On the surface, for sure!

Why did you choose the gestures that you did?

I think pointing to the exact lane is more accurate, instead of just pointing to the left or right. Accuracy is important.

Which gesture do you think fit the task best?

Pointing to the target lane, and then using my fingers for steering, where the fingers’ direction is the same as the lane’s direction.

What do you think your best gestures all have in common?

I don’t think they have anything in common, they were quite different. Maybe just the directions they used.

Where your choices inspired from any existing technology?

Yes, when I swap my fingers it’s like my iPad, though the pointing wasn’t inspired from anything. I also used a steering wheel gesture, like in a video game.

Let’s talk about voice commands...

Why did you choose the voice commands that you did?

First one I chose was “left” and “right” and it’s very common, just go in that direction. I would also say “left,
left” for changing two lanes to the left. The second one I used was labelling each lane with a number and then naming that number.

Which ones did you prefer the most?
The numbers.

Where you inspired from any existing technology?
No…

Let’s talk about the multimodal condition…

What kind of style did you use? Simultaneous, cascading?
I gave both commands at the same time. Because what you think is what you act…giving them at different times just cause delays…

Why did you choose these combinations?
I don’t know, it just seemed natural.

Which one do you think was the best multimodal command?
I preferred pointing and naming the lane. My hands got tired when I was doing the steering wheel.

Did you get any inspiration from any existing technology?
Siri for voice commands.

If a refined and advanced version of what you did today became a feature in real driverless cars…

How could I trust that? I would never use it. I cannot control it, I can only give it commands…

Assuming you are in an autonomous vehicle on your way somewhere and you wanted to change lanes, would you prefer to use one of the techniques you used today or would you prefer using the wheel and pedals?
Definitely wheels and pedals!

Participant 3
So how did it feel to have control over a car yet no wheel or no pedals?

Maybe because I don’t have a driving license but it felt pretty natural. Because I used to play racing games where you just press buttons such as left or right – pretty similar to what I did today so it felt rather natural.

How authentic do you think the simulation was? How much did it feel like an autonomous vehicle?
I have never used self-driving cars but it felt like a regular car because it was kind of following the laws of physics. It wasn’t just linear, it was more on curve…I’m not sure how to say.

What was your favourite modality? Did you prefer gestures, voice commands, or both?

Definitely the gestures, because when I had both I had trouble remembering what was left and what was right, so I made mistakes – gestures are pretty unambiguous on where I want to go and I didn’t have to think so much. When I had both, sometimes I said something that contradicted my gesture, and it was always the case that my gesture was right and my voice was wrong, not the other way around. Gestures required less thinking on my part.

Which modality was the most difficult to work with and which one was the easiest to work with?

Gestures were the easiest because they were very intuitive and I didn’t have to think. The most difficult was both, because I made many mistakes where my gestures said different things than my voice.

How well do you think you were able to recover from errors?
I think it was pretty easy, unless I was really close to the gantry. So basically when I noticed that the car was going straight and I wanted it to go a bit further right or left, I just repeat my gesture. It wasn’t hard.

Was error recovery more difficult in one condition or another? Was it easier?
I guess gestures take less time as well; I was a bit stressed when the car made a mistake and then I had to think: what I need to say, what is my command…with gestures, it was easier for me. I just repeated what I said or did, because I assumed that this was something in the code, not on my part; or I just made a very small gesture and it didn’t register it.

Did you consider the possibility of creating an undo command?
No.

Why do you think the errors occurred?
I think some of them might have been because my gesture was too unnoticeable. So in the voice condition I might have not spoken clearly enough, and in the gesture condition I might have made too small of a gesture to be registered by the system.

Did you try many different types of gestures, voice commands, or did you mostly use the same inputs?

I think I used more by the end because I knew I could experiment within the condition. It was pretty cool. I found that the best time to experiment was at the beginning when I had no commands. I tried a few things, but when I tried one thing for example in the voice command, it didn’t work and then I didn’t try it again. I can’t remember specifically which one but it was in the second condition. Overall, moderate experimentation, nothing crazy.

Are you satisfied with your choice of inputs?
I think I would need more testing before I could say I found the perfect inputs. Especially since I wasn’t quite sure how it interprets my gestures, so I just did gestures that felt intuitive to me, but sometimes the car went over the lane I wanted to go to, so then I tried to adjust this gesture a little bit and check, basically experiment, within a gesture. I changed it by making it shorter, snappier, or by making it longer. And in the condition where I used a simulated car wheel, I thought I could do the same thing: for example, I should go two lanes to the right, so then I did the right gesture two times, but then it didn’t do anything, it just went back to its place, so then I made the gesture longer.

Let’s talk about the gestures you used…

Did you prefer making gestures in the air or on the surface? I was pointing, touching the screen at some point. I preferred doing gestures in the air.

Why did you choose the gestures that you did? I guess I didn’t think too much about. I just went with my instinct – if my hand goes to the left, then the car would go to the left. Then I simulated a car wheel. Then I pointed to the screen to my target.

Which one of these do you think was the most successful? I think the waving in the air, with an open hand.

What do you think your best gestures all have in common? I think they are low-effort. When I was doing the wheel, my hands got tired and with waving one hand: (1) it requires using one hand, not two, and (2) you can put your hands down in between, when you don’t need to change lanes.

Where your choices inspired from any existing technology? I don’t know, because I don’t use the Kinect. Maybe the Wii controller, because sometimes you just move it like that, like waving, but not consciously.

Let’s talk about the voice commands you used…

Why did you choose the commands that you did? I did “left-right”, or “left-left”. Basically, I felt like I needed to say the direction, because when I tried to say “one” once, it didn’t respond, so I guessed it needed directional controls. I tried “left-left” and it worked so I stuck with it. Then I thought it was too much to say “left” four times, because it was too slow. So then I tried “four left”, like multiplying, which worked fine as well.

Which one do you think was the most successful? I think the multiplication one…”one left”, “three rights”…it seemed to be understood by the system pretty well.

What do you think your best voice commands all have in common? They are short and easy to say.

Were your choices inspired from any existing technology? No.

Let’s talk about the multimodal condition…

What interaction style did you prefer? Cascading, simultaneous? I think I started talking first, and then as I was finishing my voice command I did the gesture, almost at the same time really.

I have a question about a specific input you made in this condition: at one point, you were saying “left-left” and you were doing a continuous, steering gesture. Could you walk me through your rationale? I didn’t think about it; It was horrible. It was so confusing, because you have this numerical system and then this continuous system of the wheel, and then I wasn’t sure what was prioritized if I made a mistake.

Why did you choose the combinations that you used? Basically experimentation. I was trying to be the most efficient, trying different things. Mostly I tried to do a combination of things I already did and knew they worked. Towards the end I tried something else.

Which one do you think was the most successful in this condition? I think the waving gesture and the voice commands “two left”. They are the easiest on their own, and also the easiest in combination. When they were more lanes between the car and the target lane, it was easier to use the system of saying “first from the right” or “middle lane”, because sometimes I wasn’t sure how many lanes I had to change but I knew which lane I had to go to.

If a developed and refined version of what you did today became a feature for a driverless car, would you use it? Would you trust it? I think I would use the gestures. I would use the system, see how it behaves, then decide to trust it or not.

Would you prefer this system over changing lanes manually? No.

Participant 4 So how did it feel to have control over a car yet no wheel or pedals? I think it was easier than driving the car. I think it’s just me, because I’m learning how to drive at the moment, and I keep getting confused with things like the clutch and gas...if everything was automatic like here I think it would be so much easier.
What was your favourite modality?
I preferred voice commands, because they're easier. I think using my hands all the time would be too tiring, doing this again and again - I'd rather sit like this, relax, and say what I want.

Which condition was the most difficult to work with?
Gestures. Multimodal was in the middle.

In which condition did the simulation respond best?
I think voice commands.

How well do you think you were able to recover from errors?
Quite well, actually. I think it was quite good.

Was error recovery easier or harder in certain conditions?
Maybe it was harder when I got closer to the sign which made me confused about left or right. It was also easier to recover with voice commands.

Why do you think the errors happened?
I thought I was the one who made the errors. I sometimes used the wrong gestures or commands. I think it worked quite well.

How much would you say you experimented with different inputs?
I think I just kept getting back to the same things, so not a lot at all!

How satisfied are you with your choice of inputs?
I think I'm pretty satisfied!

Let's talk about the gestures that you used...
Did you prefer making gestures in the air or on the surface?
I trusted interaction more when touching the screen compared to gestures. I think in a real situation I would use those, but here I did more gestures in the air just because it was fun.

Why did you choose the gestures that you used?
It was more natural when it came to left or right. When I had to go left, I used my left hand, and vice versa. For two lefts I would go like "one, two" (gesturing twice with her hand)

Which gesture do you think was the most successful?
Waving left or right hand in a semi-circle motion, with one motion added for each extra lane.

What do you think your best gestures all have in common?
I think they are all natural, intuitive. I also did a motion when the car got to the lane I wanted to tell it to stay there.

Do you think using the "stay there" gesture improved performance?
Yeah, I think it helped. I felt like if I kept gesturing, it would keep going to the right, so I wanted to stop it from turning.

Were your choices inspired from any technology?
I think swiping came from the phone. Even games, like the Nintendo Wii.

Let's talk about the voice commands that you used...
Why did you choose the voice commands that you used?
It was just natural to say right and left.

Which ones were the most successful?
The go left/right ones. I think "go two rights" was quite confusing, so I used "go double right"

Were you inspired from any tech?
Not really, I used it like any smartphone voice interface.

Let's talk about the multimodal condition...
Describe me the style of interaction that you had in your mind?
I think I used the voice command before the gesture; I trusted the voice commands more than the gestures. With the gestures I was confused, like what do I do now?

Why did you choose these combinations?
I just thought of gestures to go along with voice commands, really.

Which ones were the most successful?
The directional circular motions, counting the lanes.

Was this inspired from any technology?
I thought of an Xbox game I used to play with the Kinect

If a developed and refined version of what you used today became a feature in a real driverless car, would you use it?
I think I'd give it a try.

Would you trust it?
Half and half

Would you prefer using it instead of the wheel and pedals?
If I knew it worked, yes, but I'd still want the wheel just in case.

Participant 5
So how did it feel to have control over a car yet no wheel or pedals?
It felt just like a game. I'm a bit anxious about the gesture, because, for example, in voice you command the car to move exactly one lane - with the gesture, you don't know when your car will stop. You have to hold your hands in that position...I think from the experience, I will choose the voice command...do you know when you play, there are joysticks and arrows? I feel like the voice commands are the arrows and the gestures are the joysticks. There's no specific haptic feedback so it's harder...

How authentic did the simulation feel?
I would say I feel like in a driverless car, but I don't feel like I'm interacting with a vehicle. I feel like I'm telling the machine to do what it has to do, but because it gives no feedback on whether it's working or not.

What was your favourite modality?
If I have to use both at the same time, I worry I will forget to do or say something. I think I prefer only one modality, probably voice.

Which condition was the most and least difficult to work with?
Most difficult - multimodal. Least difficult - voice

How well do you think you were able to recover from errors?
When I'm using voice, I'm less likely to panic. When I use gestures, I panic and then my hand movements are affected and then I panic more.

Was error recovery easier or harder in certain conditions?
Multimodal was easier than voice

Why do you think the errors happened?
Probably because I don't know which input the machine understands.

Let's talk about the gestures that you used...
Did you prefer making gestures in the air or on the surface?
Didn't touch the surface.

Why did you choose the gestures that you used?
Because it's easy. The first gesture was easy but not comfortable, that's why I went with something else. That's why I snapped my fingers either with my left or right hand, because I could do one for one, because it gave some feedback.

Which gesture do you think was the most successful?
Snapping my fingers. I think using both hands to command the car to move to the middle lane was very convenient. I just put my hands in the middle right in front of me.

Were your choices inspired from any technology?
No.

Let's talk about the voice commands that you used...
Why did you choose the voice commands that you used?
Probably because it's already part of life. I feel like it's more reliable if I use voice, also if I had to change many lanes I could first command "middle lane" and then change again instead of counting how many lanes I had to change or saying "right" four times.

Which ones were the most successful?
Middle lane, and just being specific.

Were you inspired from any tech?
Not really.

Let's talk about the multimodal condition...
Describe me the style of interaction that you had in your mind?
I think it must be complimentary, because in driving you are not only moving left or right, you are using the horn, the lights, and all the other controls. I could see myself using voice commands for small things like that and signalling, turning on the parking camera...

Which ones were the most successful?
Complementing - moving the right hand to the right, then saying "right" at the same time, like confirming the gesture.

Was this inspired from any technology?
No.

If a developed and refined version of what you used today became a feature in a real driverless car, would you use it?
Depends on the traffic...maybe

Would you trust it?
Not yet. I may trust it for other activities but not for driving.

Would you prefer using it instead of the wheel and pedals?
On a transition (lane change?), it would be good to have control, but if I could tell it to park itself in the nearest rest area for example, I would.

Participant 6
So how did it feel to have control over a car yet no wheel or pedals?
Really did not like it. The time I felt most secure was when I was making the weird steering wheel motion. With the voice control, didn't that like that because I felt at some point I was drifting off, and then I kept getting right and left mixed up. It's not the technology that's the problem, it's me. And saying the 1, 2, 3, 4, 5 lanes - I guess that wasn't bad,
but when I got to the end of the track I found myself nodding off a bit, but that didn't happen with the gestures.

How authentic did the simulation feel?
I didn't feel like in a driverless car. It did feel authentic in the way it was moving.

What was your favourite modality?
Definitely gestures; it felt most natural, because I'm a driver. The one where I did the steering wheel made me feel like I was in control and felt a bit safer. Both was good because sometimes it felt like another way to give feedback. I definitely trusted myself more with the gestures.

Which condition was the most and least difficult to work with?
Most difficult - gesture, steering wheel. Most difficult - voice, especially with the left/right

To which modality did the system respond best?
I think voice, actually. Maybe because most of the time it was one word of information, while it had more information to process with gestures.

How much would you say you experimented with different gestures throughout the study?
I think with the multimodal that was when I experimented most and found the two I was most comfortable with. I tried pointing, which worked well.

How satisfied are you with your choice of inputs?
I think I'd need more testing. Even though I liked the [steering wheel motion] I don't know how comfortable that would be for an entire journey.

How well do you think you were able to recover from errors?
Alright. I just corrected it as soon as it happened. I got frantic when the pointing gestures didn't work though.

Was error recovery easier or harder in certain conditions?
Easier in gesture.

Why do you think the errors happened?
Don't know...

Let's talk about the gestures that you used...

Did you prefer making gestures in the air or on the surface?
Didn't touch the surface.

Why did you choose the gestures that you used?
Steering wheel felt the most natural because it felt like it had the most control and that I could control the angle of it, like I wasn't making it veer off too far to the right or the left.

Which gesture do you think was the most successful?
Steering wheel.

What do your best gestures all have in common?
I think they're similar to a real car and I feel like I can successfully predict the angle that the car is going to move.

Were your choices inspired from any technology?
No.

Let's talk about the voice commands that you used...

Why did you choose the voice commands that you used?
I chose the numbers just because I thought that would be easy for my brain to handle, I guess. The second one I kind of chose left/right, because I couldn't think of anything else; I wouldn't use it in a real scenario.

Were you inspired from any tech?
Not really.

Let's talk about the multimodal condition...

Describe me the style of interaction that you had in your mind?
I think I was mostly doing it at the same time, but probably the gestures came first and voice commands just after, to kind of solidify where I definitely want to go.

Based on what criteria did you choose these combinations?
I just went with what came to my mind at the time.

Which ones were the most successful?
I think it was the pointing and saying "1, 2, 3" because I kept getting left and right wrong even though I was compensating with the gesture.

Was this inspired from any technology?
No.

If a developed and refined version of what you used today became a feature in a real driverless car, would you use it?
Maybe, but I'd have to get used to it a lot first. I think it would take me a long time to get comfortable with it. Maybe future generations will be fine with it but not me.

Would you prefer using it instead of the wheel and pedals?
It would depend on how many other cars and/or pedestrians are about. If it was very busy I'd do it myself, but if it was quite empty I would comfortable enough to use the gestures.

Participant 7
So how did it feel to have control over a car yet no wheel or pedals?
Very good, I liked it. It felt kind of natural, but only after a while.

How authentic did the simulation feel?
The controls felt very natural but it didn't feel like I was in a car at all.

What was your favourite modality?
I think I liked the multimodal, because you could sort of combine your commands. For example, when I used the hand I could say "two lanes" and it would know because my hand was pointing to the left, then it would go two lanes to the left. Otherwise I would need to say a long sentence which includes the word "left". This is more natural, more like when we speak to people, when we "Talk with our hands". I would like to use both gestures and voice commands when interacting with a car.

Which condition was the most and least difficult to work with?
I think voice command was quite hard to work with. Sometimes you can't really express in a second how many lanes or what direction, sometimes your brain gets a bit confused. Pointing seemed a lot more quicker, maybe because it's more visual. Least difficult was multimodal.

How well do you think you were able to recover from errors?
It was hard to recover from errors when I spoke the commands - "go all the way to the left, no! go all the way to the right!" - with the hands you could react much faster and naturally.

Was error recovery more or less difficult in any condition?
It was easier in any condition where I could use my hands.

Why do you think the errors occurred?
I think most of them occurred because of my input. I felt like the system was reacting correctly most of the time.

Did you experiment much throughout the study?
I tried a few different things, but once I found a command that was the least effort while it still understood what I wanted, I stuck with that.

Are you satisfied with your choice of gestures?
Yeah I think so. Coming up with my own commands was harder than I thought. Voice commands were easier, but in the end I came up with this one-handed steering motion - like a joystick - which was great.

Let's talk about the gestures that you used.
Did you prefer making gestures in the air or on the surface?
I think it depends on the situation...if the lane I want to go to is really far, it was easier to just point on the lane and select it, but if I just wanted to go left or right once, I only did a gesture in the air.

Why did you choose the gestures that you used?
I think I chose them based on the effort. I always try to use the one which implies the least physical effort.

Which gesture do you think was the most successful?
I liked the joystick gesture with one hand. It felt natural, but maybe because I was sitting in front of the screen and it felt a bit like a computer game.

What do your best gestures all have in common?
Least amount of effort

Were your choices inspired from any technology?
Yes, especially the joystick thing. I don't know about the rest, maybe it was partly inspired by technology and partly by how we communicate naturally.

Let's talk about the voice commands that you used...
Why did you choose the voice commands that you used?
Effort. I tried making longer sentences, then refined my commands to be as short as possible.

Which one was the most successful?
"Two lanes left, two lanes right" - exactly like that. Or "go all the way to the left". I feel like there is no specific voice command that I like most because it just depends on what comes to my mind.

Were you inspired from any tech?
No.

Let's talk about the multimodal condition...
Describe me the style of interaction that you had in your mind?
I preferred using both at the same time, because it's kind of natural to express your intention in words and at the same time express your intention with your hands.

Based on what criteria did you choose these combinations?
Effort

Which ones were the most successful?
I think it was the combination of pointing to the left and to the right and saying how many lanes I want in that direction.

Was this inspired from any technology?
No.

If a developed and refined version of what you used today became a feature in a real driverless car, would you use it?
I think so.

Would you prefer using it instead of the wheel and pedals?
I think I actually prefer this gesture thing. I think taking control over the car feels like you are interrupting the driverless functionalities.
Participant 8
So how did it feel to have control over a car yet no wheel or pedals?
It's not that trust, because you don't know whether the car will follow your instruction and whether you know what you're doing.

How authentic did the simulation feel?
It didn't feel like a real car, but the road is similar to when I drive cars in real life - the speed is quite fast.

What was your favourite modality?
I preferred using voice, because for gestures, I think the car might slow down my gesture. If I just say "first lane", then the car will go there, but with gestures have to push it to go left.

Which condition was the most and least difficult to work with?
I think the gestures-only is the most difficult, because it's hard to control. Easiest would be multimodal because it's like "double security" - you're using both to confirm the command.

How well do you think you were able to recover from errors?
When the car didn't do anything, I might think the car didn't recognise the gesture or voice, so I'm a bit nervous to push the car to go to the right lane. If it went to the wrong position, I'm also nervous to try to move the car again, but this time I'm probably too late.

Was error recovery more or less difficult in any condition?
Once in the multimodal condition, the car went to the wrong direction, and I was very close to the sign, and I tried to recover from the error and it made me rush...

Why do you think the errors occurred?
Because my gestures weren't the preferred ones, or my voice was misunderstood.

Are you satisfied with your choice of gestures?
I think I could find better commands. I think if you gave me the instructions, that would be great. If I use Chinese to control my car, I might behave differently than in English. Maybe because my vocabulary is limited, I couldn't come up with much.

Let's talk about the gestures that you used...
Did you prefer making gestures in the air or on the surface?
Didn't touch the screen.

Why did you choose the gestures that you used?
Based on the way I drive the car, like I have a wheel. And the other gesture, I just went for something easy: pointing.

Which gesture do you think was the most successful?
Pointing, because it's easy to let the car know which lane I want to go to.

Were your choices inspired from any technology?
No.

Let's talk about the voice commands that you used...
Why did you choose the voice commands that you used?
Just based on ease of control.

Which one was the most successful?
Numbers! It's very accurate, it lets me trust the car because the numbers represent the lane.

Were you inspired from any tech?
Maybe navigation in the car.

Let's talk about the multimodal condition...
Describe me the style of interaction that you had in your mind?
First I used the gesture because my motion goes first, then I have to think "is it left or right?", so voice might be slower. I used them as confirmation.

Based on what criteria did you choose these combinations?
I don't know.

Which ones were the most successful?
The last one, in which I was pointing and saying number.

Was this inspired from any technology?
No.

If a developed and refined version of what you used today became a feature in a real driverless car, would you use it?
Maybe, maybe not. I would like to try; if the car responds accurately, I might use it more.

Would you prefer using it instead of the wheel and pedals?
I prefer the steering wheel.

Participant 9
So how did it feel to have control over a car yet no wheel or pedals?
It was quite unusual for me, but after I practiced for a while I think I can handle it. Sometimes I drive the real car and I will lean my body as I did in this experiment.

How authentic did the simulation feel?
It didn't feel like a real driverless car because there were no other cars on the road.

What was your favourite modality?
For me, I like to use gestures more than voice commands, which distracted me. I didn't like using voice commands because English is not my native language and I have to think.

Which condition was the most and least difficult to work with?
The hardest was the last one. I think I did "1, 2, 3, 4, 5" which is not very natural for me. The easiest one was just using my hands and body.

How well do you think you were able to recover from errors?
I think I did quite well.

Was error recovery more or less difficult in any condition? It was harder when using my body to steer because I couldn't really control the car anymore if I went too far on one side. Voice and multimodal were harder - multimodal distracted me; I didn't want to use both but the condition forced me to do that - voice was less natural because we use voice interfaces less in real life.

Do you feel you experimented a lot throughout the study? Yes I think so, but there is not a lot of variety available with voice.

Are you satisfied with your choice of gestures?
I don't know, maybe I could find better gestures but not sure.

Let's talk about the gestures that you used...
Did you prefer making gestures in the air or on the surface? In the air.

Why did you choose the gestures that you used?
I don't know...

Which gesture do you think was the most successful?
My favourite was moving my body.

What about the motion you made at the end of a gesture? [setting the car straight]
Because of the way we drive the car - you turn to the right and then back to the left to go straight. When I do this steering gesture, I feel I can go right a bit or a lot, so this is my way of making sure the car didn't go too far.

Were your choices inspired from any technology?
Cars, maybe.

Let's talk about the voice commands that you used...
Why did you choose the voice commands that you used?
Because I think the system will know the language I use so it will understand left and right.

Which one was the most successful?
Left and right.

Were you inspired from any tech?
No.

Let's talk about the multimodal condition...
Describe me the style of interaction that you had in your mind?
I think I would use gesture first, then use voice commands to modify if I feel like the gesture doesn't work well. But I prefer not to use voice at all; if the gesture registered then I won't use it.

Based on what criteria did you choose these combinations?
I don't know.

Which ones were the most successful?
Not sure.

Was this inspired from any technology?
No.

If a developed and refined version of what you used today became a feature in a real driverless car, would you use it? Maybe, if it works well.

Would you prefer using it instead of the wheel and pedals?
If we talk only about changing the lane, maybe I would prefer to tell it directly which lane I'd like to go to, because it's easier. For other things like parking, if it means having to command every movement, then I would prefer to do it manually; if it means I give the command and the car does it, that would be perfect.

Participant 10
So how did it feel to have control over a car yet no wheel or pedals?
At first it's quite tricky. My concern is: I don't know what gestures it's using. What exactly should I do to control it? After I tried it a bit, it has a decent level of reliability.

How authentic did the simulation feel?
Not very. It's like a training session. There are no other cars and it's quite simple.

What was your favourite modality?
I preferred gesture-only, because I found it takes longer to react in voice-only condition. In the multimodal, I might make mistakes where the gesture is towards the right and I said "left", then I get worried.

Which condition was the most and least difficult to work with?
Multimodal was hardest. Gesture-only was easiest and the simulation responded best to it.
Let's talk about the multimodal condition...
Describe me the style of interaction that you had in your mind?
The gestures and voice commands came at the same time; gesture is more natural. If I want to give a command, I naturally use my hands.

Based on what criteria did you choose these combinations?
Correspondence. If I say "left-left", then I would move my hand twice to match it.

Which ones were the most successful?
The one where I'm kind of pointing twice, saying "right-right".

Was this inspired from any technology?
No.

If a developed and refined version of what you used today became a feature in a real driverless car, would you use it?
If it's the only choice, I would. I think I probably would anyway. I'm just reluctant because I don't trust it. Once I see it's very reliable, I'll use it.

Would you prefer using it instead of the wheel and pedals?
Some people like the feeling of manipulation when they drive a car. I kind of enjoy my current driving experience but I kind of like this one as well. I'd want both in a car.

**Participant 11**

So how did it feel to have control over a car yet no wheel or pedals?
(laughs) I feel like I've lost more energy just talking than...

How authentic did the simulation feel?
It made me feel like I was in a driverless car.

What was your favourite modality?
Gesture-only, because I get tired from talking. I would say "turn left" to a driver if I was in the passenger seat, but I wouldn't say it to a car.

Which condition was the most and least difficult to work with?
The easiest was actually the voice condition; the simulator even responded better to that. The hardest was both. The most comfortable was the gesture.

How well do you think you were able to recover from errors?
Quickly.

Was error recovery more or less difficult in any condition?
It was easy for the speech condition because even if I'm driving in real life and I make a mistake, I would actually shout out "to the right!".

Why do you think the errors happened?
You know how I'm on this lane and I see that arrow over there on the opposite side - I don't know the term but I couldn't tell if it was on the 4th or 5th lane.

Do you feel you experimented a lot throughout the study?
Moderate; I didn't try that many.

Are you satisfied with your choice of gestures?
I'm satisfied.

Let's talk about the gestures that you used...
Did you prefer making gestures in the air or on the surface?
I preferred making gestures in the air, because it feels more natural.

Why did you choose the gestures that you used? Based on what criteria?
Based on my experience of being a back-seat driver (laughs)

Which gesture do you think was the most successful?
This one, like waving to the side. I tried pointing to the lane, but it didn't respond quickly.

What do your best gestures have in common?
I actually did more gesture with my right hand, even if I had to go to the left. I just shifted my right arm to the left. I think it's about pointing to the right lane.

Were your choices inspired from any technology?
No.

Let's talk about the voice commands that you used...
Why did you choose the voice commands that you used?
Because they're short and quick, especially saying left/right. For longer distances I just said "right-right-right-right".

Which one was the most successful?
"Middle lane" worked, also "go to the centre", "all the way to the right" worked

Were you inspired from any technology?
Maybe games, like racing games.

Let's talk about the multimodal condition...
Describe me the style of interaction that you had in your mind?
I think they came at the same time - no particular reason.

Based on what criteria did you choose these combinations?
I don't know...

Which ones were the most successful?
I think when I just did like, pointing my finger to the screen and saying "to the right".

Was this inspired from any technology?
No.

If a developed and refined version of what you used today became a feature in a real driverless car, would you use it?
What if there are other cars? I'm not sure.

Would you prefer using it instead of the wheel and pedals?
I might use the gestures, not the speech.

Participant 12
So how did it feel to have control over a car yet no wheel or pedals?
It's much easier and can make me more relaxed. Maybe it's not real driving, but I don't have to use many parts of my body to drive. I feel like I have better control this way.

How authentic did the simulation feel? Did it make you feel like in a driverless car?
No, it didn't feel like a real car at all.

What was your favourite modality?
It depends...When I only used voice, my voice gradually became weaker than at the beginning. If I only use the gesture, my hand and arm will be tired. Also, I felt at some moment that I couldn't focus, maybe because the task was too boring. At the beginning, when I used both voice command and gesture command, I felt distracted.

Which condition was the most and least difficult to work with?
Easiest was gesture, hardest was multimodal.

How well do you think you were able to recover from errors?
I expect if the car drives away from my target, the steering wheel is still there so I can use it.

Was error recovery more or less difficult in any condition?
No difference.

Why do you think the errors happened?
For the gesture, I think there was some mistake of the gesture recognition system. Maybe the wrong gesture was detected.

Do you feel you experimented a lot throughout the study?
So-so

Are you satisfied with your choice of inputs?
I gradually adjusted my commands. At the beginning, when I used my hands, I tried to keep my hands moving to control the car, because I hope the car can detect the line on the road so that I just need to do the gesture once for the car to move one lane, and twice for the car to move two lanes... I think I could find better input with more testing.

Let's talk about the gestures that you used...
Did you prefer making gestures in the air or on the surface? Didn't touch the surface.

Why did you choose the gestures that you used? Based on what criteria?
I think I was comparing - do I use my finger or my whole hand? I think the finger is much easier. But because it's easier I might get bored faster.

Which gesture do you think was the most successful?
I would say...I do the gesture once, and the car move once, I do the gesture three times, and the car moves three times.

Were your choices inspired from any technology?
The cameras you see on autonomous vehicles that can detect the lines on the road.

Let's talk about the voice commands that you used...
Why did you choose the voice commands that you used?
They must be very short, very simple. Also, it can be recognized by the machine easily, it's not confusing. I think saying "three left" is the most direct way to control the car. Also I used number 1 to mean left and number 2 to mean right. I also used "1, 2, 3, 4, 5".

Which one was the most successful?
With "left-left" I have to think how many times to say "left", but I still think it's my favourite.

Were you inspired from any technology?
No.

Let's talk about the multimodal condition...
Describe me the style of interaction that you had in your mind?
To be honest, I didn't figure out a pattern for myself...

Based on what criteria did you choose these combinations?
How easy it was for me.

Which ones were the most successful?
Not sure...I don't know...

Was this inspired from any technology?
Maybe wearables.

If a developed and refined version of what you used today became a feature in a real driverless car, would you use it?
If it has a very high level of safety, yes. I'd probably just sit on the backseat.

Would you prefer using it instead of the wheel and pedals? I want to have both.

**Participant 13**
How did it feel to control a car yet have no wheel or pedals?
It's awkward. I think even though the system is quite reliable, it's still...I'm not in control. Of course I can command the system, but I have had a driving license for more than 10 years now; I feel more in control when I physically touch something and steer.

How authentic was the simulation? Did you feel like in a driverless car at all?
I think it was quite realistic in the sense that I’m driving, but there were no other cars on the road, which is what will make it very difficult in reality. When you drive a car in a normal way, you need to look at so many things, and here it's basically one straight strip. I don't feel like it's real driving because there are no other cars, no flashing lights, no distractions.

What was your favourite modality? Why?
I think I liked the gestures, specifically just pointing to the target lane, because it felt the most physical; I was still physically interacting with the system.

What modalities were the most and least difficult to work with?
Least difficult was the pointing one, most difficult was just saying “left” and “right”, because I needed to repeat myself several times to change multiple lanes. It feels like a lot of effort.

How well do you think you were able to recover from errors?
Quite well. I think the system did one or two errors per condition; recovery was quite fast.

Was error recovery more difficult in one condition or another? Was it easier in one condition or another?
I think it was easy in every condition.

Why do you think the errors happened?
Probably the system is not 100% reliable, and understandably so. This is why I feel so disconnected; I expect the system to make mistakes.

Did you experiment much throughout the simulation?
Not really. I just chose a few and then combined them. I couldn't think of something that I wouldn't do in reality, like making crazy gestures. For me it needs to work in the simplest way.

Are you satisfied with your choice of gestures?
Yes, I would say so.

Let's talk about the gestures you used...
Why did you choose the gestures that you did?
I think how easy they are to execute. I didn’t try to make crazy gestures. It’s straightforward to me. I can’t think of anything else other than pointing or swiping. I think I am too heavily influenced by technology.

Were your choices inspired from any existing technology?
Yes, ever since Apple introduced the iPhone 10 years ago. We evolved from there, not just with Apple but with every product. I think swiping has become ubiquitous. As for pointing, I think we use that a lot in AR.

Let's talk about the voice commands you used...
Why did you choose the commands that you did?
I tried to keep it as simple as possible, because I don’t really trust autonomous systems. I think I tried to keep it as simple as possible because of my own experiences with Siri. It’s garbage. You can use it for the simplest commands, but...that’s it. I think I under-trust them so that’s why I don’t use them.

Let's talk about the multimodal condition...
What interaction style did you prefer? Simultaneous? Cascading?
I think simultaneous makes the most sense. I didn’t want to confuse the system, so by doing them all at once I thought I would make my intentions clear.

Were your choices in this condition inspired from any technology?
Not really.

If a developed and refined version of what you did today became a feature for a driverless car, would you use it?
Yes, I think so. What I missed here the most was feedback; it would have given me certainty and shown me how far I can trust the system. If I point at a lane and say “go to the first lane” and the system displayed me that it got the command, with vibration, with audio, with visual, anything, to confirm that it actually got what I said, it would be way easier. I think eventually I would use it.

Would you prefer it over changing lanes manually?
It really depends on the system. When I know that 99% of the time I can fully trust it that it does everything, I would probably use it.

Participant 14
So how did it feel to have control over a car yet no wheel or pedals?
It was novel, it was novel...it was another way to control a car, because I'm used to driving a car the traditional way. It was different.

How authentic did the simulation feel? Did it make you feel like in a driverless car?
Yeah it was. You can tell it’s a simulation but it's quite realistic: you have your lanes, this big highway, and you can change lanes, and you have markers...

What was your favourite modality?
I prefer using gestures because I think Jesus can interfere less with my driving habits. I like to listen to music a lot and talk in the car, but I need to listen to myself if I'm giving voice commands.

Which condition was the most and least difficult to work with?
Multimodal was the most difficult, because you can have a set of instructions and you only have to learn the instructions, but in the multimodal way you don't know whether it's actually going to work or not. For example, if you say "change to the left", it will change to the left, but in the multimodal there are many options, it's less clear to me what the commands are. I could forget them more easily. The gestures were the easiest.

How well do you think you were able to recover from errors?
It was fine because the car just does what you tell it to. It happened to me a couple of times that in my mind I wanted to say "move to the right" but somehow I got left and right mixed.

Was error recovery more or less difficult in any condition?
It was easier with the gestures because it's easier to use directional movements than thinking about which is left and which is right. I was at my worst performance when I was using the voice commands.

Do you feel you experimented a lot throughout the study?
I was asked to change the commands...I think I was more experimental with the voice commands.

Are you satisfied with your choice of inputs?
Yes I guess. I never thought about what kind of commands I'd give to a car. With time, I think I could come up with better, more comfortable inputs. I think in a real situation I wouldn't change two lanes at once. I would change one lane and once I know it's safe, I would repeat the command.
Let's talk about the gestures that you used...
Did you prefer making gestures in the air or on the surface?
Didn't touch the surface, forgot about it.

Why did you choose the gestures that you used? Based on what criteria?
The ones that I felt were natural to me. My hands are right before my body, so it just felt natural.

Which gesture do you think was the most successful?
The fingers. When I used the fingers it was very easy. I guess you can tire if you move your hands a lot, so I think the fingers are better.

What do you think all your best gestures have in common?
I guess I was mapping...I used my left hand to gesture to change to the left and the right hand for changing to the right.

Were your choices inspired from any technology?
Maybe Kinect? From Microsoft? Or the Wii... I guess with the Kinect you don't have a controller, so you have your hands free.

Let's talk about the voice commands that you used...
Why did you choose the voice commands that you used?
I tried to vary the commands a bit. For example I'm going to say "left" if I want to move left and then say "again" to repeat the command. I also tried to give the instructions from the beginning, like "go right three times".

Which one was the most successful?
Saying one command at a time. You have to do an extra calculation to tell how many lanes you have to change. It's because of this that I made mistakes, because I was counting the lanes, checking the sign, and then give the command. With commands like "right" that you just repeat after every lane, you don't have that problem at all.

Were you inspired from any technology?
No, because I'm not a big fan of voice commands...Siri...I never use Siri on my phone.

Let's talk about the multimodal condition...
Describe me the style of interaction that you had in your mind?
Something that was clever enough that...something that you can communicate in natural language...I understand that distance is difficult to implement, so what I would ask is simple commands...just keep it simple. I expect the car to know what I mean with my commands.

Based on what criteria did you choose these combinations?
Simplicity.

Which ones were the most successful?
Repeating the instructions: left-left, with matching gestures.

Was this inspired from any technology?
No.

If a developed and refined version of what you used today became a feature in a real driverless car, would you use it?
Yes, definitely. We've been using a steering wheel since the car was invented, 150 years ago. It doesn't mean that you have to reinvent the wheel, sorry for that...I wouldn't remove the wheel completely.

Would you prefer using this system to change lanes or do it manually?
Manually, until I know I can use the system safely.

Participant 15
So how did it feel to have control over a car yet no wheel or pedals?
I feel amazing. I don't know how it works, I still think you were controlling it. I don't know in a real scenario whether the car will be smart enough to recognize everything, and especially for the voice, if I didn't pronounce very clearly, when I drive I can't think much about my pronunciation or voice.

How authentic did the simulation feel? Did it make you feel like in a driverless car?
It felt authentic, it was just strange without the wheel.

What was your favourite modality?
I prefer both if I can input in two ways in case I have my hands busy. I think multimodal is quite flexible.

Which condition was the most and least difficult to work with?
The gestures were the easiest; they were the most natural. I think the voice was the most difficult.

In which condition was the simulator most responsive?
Can't tell the difference, really...

How well do you think you were able to recover from errors?
It was very hard to recover when I was approaching the arrow and it's just a short distance away. I couldn't always tell which way I was going and what the system understood. I would have liked some sort of feedback from the simulation to tell me where the car is going.

Was error recovery more or less difficult in any condition?
It was easier with gestures, because when I wanted to turn left I just made a movement and then when I wanted to stop I made another movement.

Why do you think the errors happened?
I don't know...

Do you feel you experimented a lot throughout the study? It was hard to find different inputs. In a real scenario I would use the same set every time, because it would be easier to remember and easier for the machine to recognize too.

Are you satisfied with your choice of inputs?
I think so, yes.

Let's talk about the gestures that you used...

Did you prefer making gestures in the air or on the surface? Didn't touch the surface, because in a real car I cannot touch the glass, so why would I do this here?

Why did you choose the gestures that you used? Based on what criteria?
In real life when we communicate with people, we use gestures like I used today.

Which gesture do you think was the most successful?
Making a waving motion in the direction and then stopping when I'm satisfied, then it will stay in that lane.

What do you think all your best gestures have in common?
The "stay" gesture which I used to tell the car to stay in its current lane.

Were your choices inspired from any technology?
Maybe from PCP module... there are some distance sensors. If there is a sensor on the side, I wave my hand at it to enhance recognition.

Let's talk about the voice commands that you used...

Why did you choose the voice commands that you used?
I just used left/right. I don't think there was a particular reason. It just felt natural. The only reason I used the others was because you forced me to find different ones.

Which one was the most successful?
Turn left, turn right...

Were you inspired from any technology?
No.

Let's talk about the multimodal condition...

Describe me the style of interaction that you had in your mind?

I think the voice and the gesture commands had the same kind of meaning, and I just repeated them here. I think it would be useful to be able to use two modalities, but maybe not together at the same time like I did.

Based on what criteria did you choose these combinations?
I don't know...I just went for what felt natural, I didn't think much about it.

Which ones were the most successful?
Same...the action signal with the stop signal.

Was this inspired from any technology?
No.

If a developed and refined version of what you used today became a feature in a real driverless car, would you use it? I don't know. I feel that if I can use the wheel as the main input and use gestures or voice only sometimes for other controls, like an assistant. I want the wheel to be there in case I need to correct it.

Would you prefer using this system to change lanes or do it manually?
I prefer the traditional way.