CS889 Paper Final - Human, What Do I Do? A Study on Intervene Input Methods for Autonomous Vehicles

Kapil Haresh Vigneswaren
David Cheriton School of Computer Science
University of Waterloo
Waterloo, ON, Canada, N2L 3G1
khvignes@uwaterloo.ca

ABSTRACT

As autonomous vehicles get more advanced, we need to ensure that users are able to be confident in these vehicles and trust them, in order for them to be a success. This is especially hard in some autonomy levels such as the Level 3 Conditional Automation, where participants don't have to pay attention all the time but need to be available for the car if a request to intervene is invoked by the car. One way we see autonomous vehicles being able to build trust with users is to allow users to know that given a situation where there is a request to intervene, the user can input their decision quickly - especially in the case of cooperative driving environments.

Inspired by research done by Walch et. al [19], we present a study that looks into the use of touch screens, voice commands and gesture controls in the input of responses to intervene requests in autonomous vehicles. We ran a user study with 5 participants from the class, where each participant undergoes a 4 condition x 3 input method within-subject experiment. At the end of the study that allows us to see input speed, perceived and actual effectiveness as well as ease of use, we see that while voice commands are typically the fastest, touch screens are typically more accurate and effective (both perceived and based on quantitative data), a well as the easiest in most cases. Voice commands however were easier and more effective than gesture control, which was the worst in all measurable way and based on participant feedback as well. We hope to see this study repeated in the future with a larger, more diverse population of participants to get an even better idea on how these input methods perform when used by people who are older and less technologically savvy.

1. BACKGROUND OF AUTONOMOUS VEHICLES

What is an autonomous vehicle? Basically, autonomous vehicles, also known as autonomous cars or self driving cars, are motor vehicles like cars and trucks that are able to perform the various driving tasks from the time it leaves an origin location to a set destination, without needing input from a human [20]. It is guided by various sensors and cameras that would allow the car's software to understand it's surrounding and allow it to plan it's path around the various obstacles on it's own. It is often confused, however, as many cars today already have some form of autonomy, from self parking technology, to lane assist and adaptive cruise control systems. We feel that it is necessary for us to first ensure that the true meaning of an autonomous vehicle is understood before talking about our study.

The SAE International J3016 Standard has defined autonomous vehicles to have 6 varied levels of automation as below [2][4]:

- **Level 0 No Automation**
  The vehicle may include some warning functions like backup radars at this level.
- **Level 1 Driver Assistance**
  Driver is primarily in control of the vehicle and makes all decisions, however there is additional technology to assist the driver in driving duties, like adaptive cruise control.
- **Level 2 Partial Automation**
  Under certain circumstances, for example - highway conditions, the car will be able to fully control steering, braking and acceleration. The driver would still need to be paying attention and be ready to take over control from the vehicle at any time, however.
- **Level 3 Conditional Automation**
  The car is able to be running in autonomous mode in most conditions, except under some situations like construction zones or bad weather, where a human driver would have to be available for the car. The human driver would not need to pay attention at all times, but would still need to be a licensed driver.
- **Level 4 High Automation**
  The vehicle would be able to run autonomously under all conditions, and if a human doesn't take over control when the vehicle notifies a human to intervene, the vehicle would be able to still control itself. In theory, no driver input is needed.
- **Level 5 Full Automation** There is no need for a human driver at all, or any form of manual intervening. The vehicle would be able to be operated without any passengers or with passengers who are unable to drive (i.e. children and the disabled).

At this point in time, the most advanced production vehicle is only at the Level 3 autonomy level, and even that, only one model is currently on sale in certain markets in the world and it is only operational in Level 3 mode under specific conditions (i.e. below 37 miles per hour) [6]. Most production vehicles with autonomous capabilities typically fall under the Level 2 category, which requires driver attentiveness at all times. We see that most manufacturers have committed to having some offering of autonomous driving by the year 2020 [9], with many first trying to get Level 3 systems into the market before level 4 and 5 systems. Due to this, we will be looking at Level 3 systems in this report.

2. TRUST IN AUTONOMOUS VEHICLES

The main issue with autonomous vehicles is trust - in many cases, people do not trust them enough, despite data showing that autonomous vehicles are inherently safer than vehicles controlled by human drivers. Part of the reason for this is because of prior isolated incidents where a semi-autonomous vehicle had crashed in the early days of these vehicles entering the mass market, in addition to prior security issues with modern vehicles that have a lot of technology in them [14]. There has been a number of prior studies done by multiple bodies, in various countries that polled the public if they trust autonomous vehicles or otherwise. A summary of the results are as follows:

- In Canada, a study in 2017 was conducted to identify consumer perceptions on autonomous vehicles. About half of Canadian consumers surveyed say they trust autonomous vehicles to get them to their destination but only 30 per cent would replace their current vehicle with a self-driving car [15].

- A 2018 study by the Advocates for Auto and Highway Safety in the United States showed that an overwhelming 64% of people were concerned or very concerned with being on the road with autonomous vehicles around them [10] [11].

It seems to be that many don't trust autonomous vehicles as much as manufacturers and engineers want them to. Autonomous vehicles need to be able to gain the trust of the people in these vehicles, as well at the people outside of the vehicle in order for these vehicles to be a commercial success, as if people cannot trust them, no one would want to use them. Focusing on within the vehicle, we need to ensure a rider will know clearly what is going on and more importantly, they would be able to quickly be notified of an unexpected situation and be able to quickly send a response to the vehicle.

The goal of the user interface inside the vehicle is primarily to provide a clear sense of what the car can see and what the virtual driver intends to do. This is what industry experts are currently working on, in their prototype autonomous vehicles. Research shows that a lot of the communication between a human driver and their passenger occurs non-verbally (in the case of taxi services or where someone else is behind the wheel), so the challenge for the industry has been to recreate that experience, between human users and the autonomous vehicle itself [8]. Recreating this experience will in turn help to build trust between consumers (riders) and the autonomous vehicle. However, the part where users would interact and respond with the car hasn't been really looked into. Users would only be able to fully trust an autonomous vehicle if they know that in a situation where the user needs to make a decision, they can input their decision quickly and easily. This is especially important in Level 3 Conditional Automation autonomous vehicles, where driver input may be needed in exceptional circumstances, but unlike in Level 2 vehicles, the driver does not need to pay attention all the time, until the car needs assistance. This problem only becomes more complex when cooperative driving interfaces are introduced, where a user inputs their decision into the car in some form instead of just directly taking control of the wheel. This is an interaction interface we expect to see in Level 3 vehicles.

3. PRIOR WORK & MOTIVATION

Some autonomous vehicle companies like Waymo and Lyft have been starting to perform user studies to see how various user experiences would have an impact on consumer perception and trust [18] [8]. In fact, in the case of Waymo's in person studies, results so far indicate that there has been a positive shift in perception towards autonomous vehicles, once people could see what the car was doing and what it was about to do next [8].

Some of these studies in person have introduced surprising results as well. A small scale study by Intel in the last year found surprising reactions towards the use of voice technology - that is, giving the vehicle a voice so it can verbally announce actions and information to be conveyed to the rider. Voice was seen as a trust-building technology by autonomous vehicle companies. While some did find the car talking to them was comforting, others found it annoying and repetitive over time especially once they became comfortable with the car [13].

Research and academia too has looked into the effects of user interfaces and experiences towards the riders in autonomous vehicles. One rather interesting paper by Heldin et. al looked at the impact of visualizing car uncertainty on drivers trust during an automated driving scenario through the use of a simulator study. A between-group design experiment was carried out where a continuous representation of the uncertainty of the cars ability to autonomously drive during snow conditions was displayed to one of the groups, whereas omitted for the control group, and the time drivers in the simulation spent performing other tasks and the time it took for them to take over control if needed was measured [12]. From their study, it was noted that drivers provided with the uncertainty information performed better in take-over situations and they were are also more comfortable with performing other tasks while driving, as compared to drivers without this information. This was a sign that the drivers could actually trust and felt like they could depend on the system more than the control group who did not have this uncertainty information of the vehicle. This uncertainty information allowed one to better understand the current state of the vehicle.

A second study by van der Heiden looked at the effect of priming drivers before handing over control to them. In their study they looked primarily at how pre-alerts affect the hand-over of control from a semi-autonomous car to a human driver. While they did not look at the effect of priming drivers from a trust perspective directly, this study did show the effect of priming on parameters that would affect the experience of a rider in an autonomous vehicle, such as the element of surprise (denoted by the effect on pre-alerts on heart rates of the individuals studied) [17]. In their study, there were signs that indicated that the voice based pre-alerts would allow for less stressful situations for riders, based on the typically lower heart rate for those participants who did get a pre-alert.

While we have seen a fair amount of work on allowing riders of autonomous vehicles to be informed of the current state of the vehicle and what it is about to do next, we have not seen that many studies in the field of getting input and instructions from the driver, in the case of autonomous vehicles. We have only, so far, seen a study done by Walch et. al [19], which looked at the usage of cooperative interfaces to avoid full hand-overs in situations in which the system needs the driver to make a decision - which is a method some Level 3 autonomous vehicle manufacturers may take, instead of getting a driver to directly control the steering and brakes themselves. This paper, released about 2 years ago, used a computer simulator to simulate an autonomous driving experience, and under situations of uncertainty, the simulation would present the request to the user to make a decision using a text display, and a voice based announcement of the same message [19]. Participants then had to input their choice/decision either using a touch screen or by voice (voice recognition was done through a Wizard-of-Oz). In all situations, there was an option for the driver to get the car to do some command, or the driver completely takes over control to maneuver the car around the situation or obstacle.

From this study, it was found that interaction via speech took a little bit longer than via touch. While it is understood the speech recognition was done through the Wizard of Oz method, thus the reaction time of the experimenter is also included, but potential errors of real speech recognition requiring a repetition of the command are excluded. It appeared that many drivers tend to mistrust speech recognition systems, hence the ability of these systems to support natural dialogs could improve their robustness could make people trust voice input more. However, it wasn't studied in this experiment, the effect of natural voice recognition in the case of providing input to the car when the car asks a rider for a decision.

In today's world of speech recognition, we are seeing more solutions that support natural voice recognition in cars, with manufacturers introducing natural voice recognition systems built into the core of the vehicle itself - allowing for a user to perform many actions with their voice, without having to remember a command phrase [16] [3]. More people have also used smart home devices that primarily use natural voice recognition as well as mobile phones and tables with this feature, which could have potentially changed perceptions as well.

In addition, there have been newer, novel ways of interacting with vehicles as well, namely gesture control. This technology was initially described as a proof of concept in the early 2000s, and has since been implemented in production cars [5] [1]. We would be interested as well to see how the use of newer input methods in cars could be applied for inputting decisions in autonomous vehicles, especially when quick input is required under varying conditions (i.e. rider is in conversation, using their phone, eating food, or is looking away from the road ahead). We would also be interested to see how voice input and gesture input compares against traditional input methods in cars (i.e. touch screens).

In our study we aim to test the various input methods available in the context of Level 3 autonomous vehicles that would involve cooperative driving interfaces (just like the proposed setup in the Walch et. Al paper), specifically in the situations the vehicle needs a user to make a decision on behalf of the car. We would like to study the differences for the three input methods we see in cars today, namely touch screens, voice control and gesture control, specifically in terms of ease of use according to the user, reaction speed (how long it takes for the input process to be completed, once a user has made a decision) and how effective each input method is (the number of attempts it takes to get correct input in by the user, as well as perceived effectiveness and accuracy) and perceived ease of use. This study would benefit the manufacturers of vehicles currently working on Level 3 vehicles which would have some form of co-operative driving (which is one approach manufacturers could take to implement the Level 3 autonomous features in their cars).

As mentioned before, there has not been much study in the field of providing input to autonomous vehicles when the vehicle is driving itself and needs a user to help in the decision making process. To the best of our knowledge, a study on the use of gesture control with autonomous vehicles has not been done before as well.

4. STUDY DESIGN & METHODOLOGY

Typically in a study of this nature, based on prior studies by Walch et. al and van der Heiden et. al, the best way to reenact the autonomous driving experience is to use a computer simulator, or some form of video based system, instead of actually putting people into a physical car [17] [19]. This is mainly because running it in a simulator would be much easier to run the study on participants, as well as it being much safer for participants as well. In the case of the study by both authors, a simulator was used, however this does require a lot more setup, from powerful computers to run specialised simulation software, to additional driving hardware (like a game controller steering wheel). There has been some free simulation software like CARLA, that would allow researchers to create simulated autonomous driving experiences [7]. However, this too does require significant compute power to allow the simulation software to even run in the first place, which is something we do not have on hand for this study, as this is a proof of concept study.

At the core of it, we are primarily studying how easy, effective and fast it is for an individual to input a command to the car when their concentration on something other than the road would be broken, and they now need to give a command to the car before they can return back to whatever they were doing. As a result, due to limitations on our end in terms of hardware available to run simulations, as well as the limited time available (being a course project), we argue that we are able to get promising and reasonably valid patterns with the use of pre-recorded driving videos that would give the illusion of a self driving car (since the participant would not have to do any of the driving).

Prior studies have also tested on the reaction times of participants under multiple conditions, namely when they were using their mobile phone [17] and when they were watching videos [19]. We extend this a little further and test when participants are occupied in a conversation, and when the participants are busy eating. Talking to others and eating are typically activities that already happen in cars today, so we expect that people will continue to do this when they are in an autonomous car as well. We also wanted to see the effect of participants just disconnecting from the outside world (like a dreaming mode) when they are busy watching the scenery outside (not the direction of travel), on the ease of use and effectiveness of an input method, and time it takes for them to complete the input. This is another possible situation we foresee many people to do in autonomous cars, as we already see many people doing this when they are riding in a train.

In summary, we tested the perceived and actual effectiveness, ease of use of an input method and how long it takes for people to complete the input action under 4 different circumstances, namely:

1. While they are using their mobile phone or tablet
2. While they are conversing with someone else in the car
3. While they are eating finger food/snacks (i.e. chips or TimBits)
4. While they are watching the outdoors pass by (not in the direction of travel)

In terms of input methods under the study, we looked at 3 different input methods, the use of a touch screen, voice control, and gesture control to input the chosen action to the car. We realised the recognition of all 3 input methods using a Wizard-Of-Oz methodology, as detailed per below:

- **Touch screen**
  For the touch screen, we presented the options available on a sheet of paper. To re-enact the "touch" experience, a participant needed to touch the appropriate option on the piece of paper. A researcher will be watching the touch action to confirm that the touch input was received successfully.

- **Voice recognition**
  We re-enacted a natural voice recognition system, that is a system that would be able to understand natural language and won't require a participant to say a specific catch phrase (i.e. instead of saying a number assigned to an option just like in [19], a participant can just say a summary of the instruction choice displayed on the screen). Once again, the Wizard-of-Oz style voice recognition is used, where a researcher listens to the voice input and confirms that the instruction received is clear and understood, or otherwise.

- **Gesture control**
  Participants were first trained on the gesture control system, where to input a selected option (say, option 1), the participant needs to "draw" the number 1 in the air with their index finger. The researcher acted as a Wizard-of-Oz and watched the gesture to confirm that the option selected has been understood.

As we needed to test for time taken to complete the input method, and ease of use of all 3 input methods across all 4 conditions, we needed to have all participants to do all 3 input methods sequentially. We weren't too concerned about the fact that the participant already knows their choice, since in this study we are looking at the ease of use and effectiveness of a particular input method, and the time taken to complete the input action once the decision has been made. All participants will receive training at the start of the study on how all 3 input methods would work.

In summary, the following would be the methodology of our study;

1. **Introduction of the Study**
   Participants were seated in front of a laptop computer that displayed the driving video footage. Once participants were comfortably seated, they were be briefed on the purpose of the study and how it would take place. Participants were free to ask questions at any time.

2. **Demo of full autonomous mode**
   We played a fully autonomous driving video that showed a car driving itself from the point of view of a driver inside it. This was to try and reduce the novelty effect of being an autonomous car. We demonstrated a few situations where the driver input is needed by the autonomous vehicle in order to proceed.

3. **Demo of input methods**
   Participants were then shown how all 3 input methods, that is the touch screen, voice control and gesture control worked in this study. Participants were given the chance to try all three input methods out, and we only continued the study once participants have fully understood the three different methods.

4. **Begin study on 4 different conditions the participant will be under**
   We tested the input speed, ease of use and effectiveness of different input methods under various conditions, as per the list below. It is important to note that for all 4 different conditions, 4 different videos were used - this is to prevent the participant from ideally knowing when to expect the request to intervene to arrive from the car. In all cases, we presented the request to intervene through an in car display (represented by a sheet of paper with the numbered options), and over voice (represented by a researcher announcing the request to intervene and the options available for the car). Despite the options being numbered, in the case of the voice recognition, the participant can say their selected option in natural language, instead of saying the option number. The conditions to be tested are as follows:

- **Using mobile phone/tablet**
  Participants were told to use the provided tablet to scroll through a news feed of content (i.e. Reddit's home page) when they were be interrupted with a request to intervene.
- **Conversation with someone**
  Under this situation, the participant was in conversation with the researcher, as we did not have any extra people to help with simulating a conversation with someone. When needed, the *researcher will play a pre-recorded voice request to intervene and administer the paper version of the request to intervene*, but will remain in conversation with the participant throughout the request to intervene and after the request to intervene.
- **Eating finger food**
  Participants were supplied with finger food like chips and Timbits, and were interrupted while they were eating with a request to intervene.
- **Watching outdoors (not in direction of travel)**
  Participants were requested to gaze away from the main screen showing the driving feed of the simulation. A secondary screen was setup to show a video of the outdoors in passing (i.e. view out of a train window). Participants were interrupted with the request to intervene at a certain point.

In general, when needed, participants were interrupted once during each condition with a request to intervene. Once the researcher had shown the options available and announced them over voice, the participant demonstrated all 3 input methods with their choice on what the car should do next. In all three input methods, the time it took for them to complete the input method would be measured. In the case the input is unclear, the participant was asked to repeat until it is understood by the researcher (who acts as the Wizard-of-Oz here). The number of attempts needed to get correct input in was counted. Each condition required no more than 3-4 minutes to test, leading to a 12-16 minute study per participant. No rewards were provided to the participants, aside from the free food in the eating finger food study.

At the end of all 4 situations, a semi structured interview was run in order to get more qualitative data that would give us some ideas on the ease of use of each input method from a participant's perspective. We ran this study on a very small sample of 5 people, due to time limitations in this study period, being a class project. As a result, we aimed more towards obtaining qualitative data from this study, with some quantitative data as well.

To summarise, each participant underwent a 4 (conditions) x 3 (input method) within-subject driving simulator study, similar to what was done in Walch et. al's paper[19].

5. DATA COLLECTION

Fundamentally in this study, we are testing 3 different constructs, namely the time taken for an input operation to be completed, the ease of use of the input operation and how effective the input method was. We measured the different constructs, across the 4 different situations the participant will be under, using the following methods:

- **Time taken to complete an input method**
  To collect data for this, we measured the time, in seconds, it took for an individual participant to complete the input action for each input method under a given situation the participant will be under.
- **Ease of use of an input method**
  We conducted a semi-structured interview at the end to understand which input method was deemed easy to use and which wasn't from the participants.
- **Effectiveness of an input method**
  This was measured through the number of attempts it took the participant to issue an input using a particular input type, that the researcher (who acted as the sensors/intelligence of the car in this study) to understand. We also got some feedback from the participants through the semi structured interview as to which input methods felt more effective and accurate, than others from their perspective.

For all the quantitative data collection and measurements being done in this study, it was be separated by the situation the participant is under. That is, we did not total up the number of attempts to issue a gesture command while the user is eating with the number of attempts to issue a gesture command when the user is using a mobile phone. The reason for this was because we expected that the situation an individual is under will have an effect on the time taken to complete, effectiveness and ease of use of a particular input method, and it is important to see if a certain input method would perform better under certain conditions over others.

From there, we went ahead to analyse the results to see the correlation between situations and the effectiveness (both perceived vs actual number of attempts), ease of use and time taken to complete a particular input method.

6. RESULTS

At the end of the study and interview, we collected and tabulated our results up. The following shows our result by each construct. Our participants have been anonymised as Participant 1 - Participant 5.

Time Taken to Complete an Input Method

The following tables show the raw times we obtained for each condition the participant was under. On the grand scheme of things, the big surprise is how voice commands seemed to be faster than touch screens in most situations for most people - which is quite an opposite from the Walch et. al paper [19].

A. Using a mobile device/tablet

Participant	**Time Taken (Seconds)**		
	Touch Screen	**Voice Command**	**Gesture Control**
Participant 1	4.05	**1.82**	5.05
Participant 2	1.67	**1.43**	1.78
Participant 3	2.21	**0.70**	3.29
Participant 4	**1.00**	1.73	2.00
Participant 5	1.13	**0.95**	2.47

Table 1. Times taken, in seconds, by each participant for each input method, when using a mobile device/tablet to read when a request to intervene was issued

We note that in this table, for 4 out of the 5 participants, the use of the voice command to input their choice was the fastest. In fact, in some participant cases (i.e. participant 1 and 3), we actually see voice command being more that 2 times faster than touch. One participant did have touch screen being the faster input method, but gesture control was consistently the slowest option across all our participants. Voice command was likely the fastest for most mainly because of the fact that under this condition, participants had to hold the tablet that was provided to them to use for reading and browsing a news feed - which would have been a little cumbersome to hold a tablet of iPad size in just one hand and stretch to touch the screen - or to balance the tablet in one hand while doing the necessary gesture. Instead, the voice command does not even require the participant to move at all and they can continue to hold the tablet and use it while issuing the command.

It is expected that we would see participant times for a particular input method under a certain condition to vary based on participant (i.e. participant 3 having double the time for touch screen compared to participant 4) - this is because different participants may have different internal response speed, and it is difficult to measure how fast are nervous responses happening in a participants nervous system. However we expect that given the same participant, they would have the same internal response speed across all three input methods. Nevertheless, comparing participants based on the order of which input method was the fastest still gives us a reasonable idea of input speed in general for a particular input method.

B. Conversation with Someone

Participant	**Time Taken (Seconds)**		
	Touch Screen	**Voice Command**	**Gesture Control**
Participant 1	**2.05**	3.50	2.23
Participant 2	3.13	2.78	**2.42**
Participant 3	1.62	**1.52**	1.98
Participant 4	1.90	**1.16**	2.93
Participant 5	**0.95**	1.60	1.62

Table 2. Times taken, in seconds, by each participant for each input method, when in a conversation with someone when a request to intervene was issued

In this condition, we note that 2 participants had voice being the fastest, while 2 participants had touch screen being the fastest input method. Only one participant (Participant 2) had gesture control being the fastest - and even that, it was only slightly faster than the voice command. This drop in the number of people who had touch screen being the fastest input method is mainly because participants had to end their conversation and context switch what they were about to say to the car. These context switches can be expensive and slow - especially if the conversation was relatively intense.

In the case where touch screens were faster than voice command, for those participants we noticed that the touch screen was almost up to twice as fast compared to the voice command. Again, we expect this to be because of the part where for these participants, it may have been faster for them to touch a button than prepare themselves to say the command and actually say the it out itself after being in a conversation.

C. Eating Finger Food

Participant	**Time Taken (Seconds)**		
	Touch Screen	**Voice Command**	**Gesture Control**
Participant 1	3.35	**1.68**	4.38
Participant 2	**1.25**	1.82	1.31
Participant 3	1.26	**1.03**	2.46
Participant 4	**0.89**	1.17	1.65
Participant 5	0.96	**0.95**	4.34

Table 3. Times taken, in seconds, by each participant for each input method, eating finger food, when a request to intervene was issued

We note that under the eating condition, 3 out of our 5 participants had voice command being the fastest input method, followed by touch screen. We also see that for 4 out of our 5 participants (except Participant 2), gesture control was the slowest input method, with some participants (i.e. Participant 1 and 5) having gesture control almost 3-4 times longer than their fastest input time (which for both Participant 1 and 5, was voice command).

One of the likely reasons for this was because when participants were eating, they needed to also have one hand holding the food container while the other hand was holding the finger food before the participant actually consumed the food. This was seen to be quite awkward and cumbersome for the participants as they had to make sure the food container did not fall while doing the gestures.

We also did not see that many participants having voice command being the fastest input method (compared to using the mobile device/tablet condition) as participants tend to also have some food remaining in their mouth making it hard for them to issue the voice command. Participants did not want to mess the touchscreen up as well, leading to quite a few trying to clean their hands up once they freed their hand from holding the food before using the touchscreen. This condition was arguably one of the worst case conditions as not only were the hands of participants occupied with holding the food/food container, their mouths also were full with food as well. However this is a condition we expect to see to be often the case in autonomous cars in the future, considering we already see people eating finger food like fruits while driving, on roads today. The availability of autonomous technologies in cars is likely to lead more people to eat in the car, while being driven to their destination.

D. Watching the Outdoors

Participant	Time Taken (Seconds)		
	Touch Screen	Voice Command	Gesture Control
Participant 1	**1.38**	2.76	2.00
Participant 2	**1.32**	1.74	2.03
Participant 3	2.22	**0.86**	1.56
Participant 4	1.18	**0.95**	1.95
Participant 5	1.22	**1.03**	1.55

Table 4. Times taken, in seconds, by each participant for each input method, when ”watching the outdoors” when a request to intervene was issued

Just like the finger food test, we see 3 of our participants having voice commands as the fastest input method, while the remaining 2 had touch screen as their fastest input method. We see gesture control being the slowest for all participants except Participant 1 and 3, where gesture control was still faster than their slowest input method, which were voice command and touch screen respectively.

The watching the outdoors condition presents a condition where participants are most distracted, as unlike all other conditions tested, the participant's main vision will not be the direction of travel. In other words, the participants cannot see what is about to happen ahead, or even glance/peek often enough, when they were presented this condition. As a result, participants can't really prime themselves based on what is happening in front of the car. The results presented here are consistent with what we expected, as participants were able to just say what they wanted the car to do, without having to move themselves and align towards the touch screen. The generally longer time taken to draw the gesture was amplified when one had to also shift themselves back in the seat to face the road, before the gesture was drawn, as one had to align the area where the gesture will be drawn to be within the gesture ”sensor” area. The gesture ”sensor” area is actually how gesture control is typically setup in normal cars that have it today.

Number of Attempts to Issue Input (Effectiveness of Input Method)

The following section details our results on the number of attempts a participant took to input a particular type of command, under a certain condition they were placed in. The number of attempts it took gives an idea on how effective an input method is for a certain condition - ideally an input method that required the least amount of attempts is one that is robust and allows a user to easily perform the input command accurately. When the Wizard of Oz was unable to understand the participant input, the participant would need to attempt to input their choice again using the same input method (be it touch screen, voice command or gesture control).

On the grand scheme of things, it appears that the touch screen consistently was the only input method that needed just one input attempt. Voice command had one instance where a participant needed to do the voice command input twice, while for gesture control, 2 different participants needed to do their gestures twice. This result is an indication of how robust and effective touch screens were in the context of in-car interactions, and voice actually did not do too bad. Gesture control's results were not ideal - especially considering this experiment was done with a wizard of Oz (we expect sensors to be more strict with detecting and understanding input.

A. Using a mobile device/tablet

Participants	Number of Attempts		
	Touch Screen	Voice Command	Gesture Control
Participant 1	1	1	**2**
Participant 2	1	1	1
Participant 3	1	1	1
Participant 4	1	1	1
Participant 5	1	1	1

Table 5. Number of attempts by each participant for each input method, when using a mobile device/tablet to read when a request to intervene was issued

In nearly all cases we see that participants only needed one attempt to input their choice - with the exception of Participant 1 for gesture control. We noticed that the participant had tried to repeatedly draw their choice multiple times consecutively (i.e. given the choice number 1, the participant kept drawing the number 1 repeatedly in the air, leading the wizard of Oz to think the participant was inputting the choice 111 instead of 1). We did ask the participant about this later on in the interview, and the participant was hoping that by repeatedly drawing, it would mean the car would be able to react faster. This is a tendency we see in people over the years with conventional input methods like mice and buttons, where people try to click the same thing or press the same button repeatedly in the hope it will make the system register the input faster (but this is never the case in reality).

B. Conversation with Someone

Participants	Number of Attempts		
	Touch Screen	Voice Command	Gesture Control
Participant 1	1	**2**	1
Participant 2	1	1	1
Participant 3	1	1	1
Participant 4	1	1	1
Participant 5	1	1	1

Table 6. Number of attempts by each participant for each input method, when in a conversation with someone when a request to intervene was issued

Once again, we see Participant 1 being the only participant who needed to repeat an input action, during the test of the voice command input method. This was the only input method in this condition that needed a participant to repeat. We note the reason why this participant had to repeat was because when the participant issued the voice command, the vagueness in their command made it not possible for the wizard of Oz to understand what they were trying to say. We have to remember that while the voice command setup in this study did support natural language, it does not mean a vague command can be issued by the user and the system will be able to figure out what the participant was trying to say.

C. Eating Finger Food

Participants	Number of Attempts		
	Touch Screen	Voice Command	Gesture Control
Participant 1	1	1	1
Participant 2	1	1	1
Participant 3	1	1	1
Participant 4	1	1	1
Participant 5	1	1	**2**

Table 7. Number of attempts by each participant for each input method, when eating finger food when a request to intervene was issued

We note that just like in the case of where the participant was using a mobile phone/tablet to read, one participant (Participant 5) had to repeat the gesture control input. All other input methods did not need repetition and all participants except Participant 5 needed only 1 attempt for each input method.

In the case of Participant 5, we asked them at the end of the study what caused them to do an input that wasn't understood by the wizard of Oz - we were told that the balancing act of not messing up the area with food bits, as well as making sure the food which they were holding wasn't going to fall wasn't easy - leading them to fumble while trying to do the gesture.

D. Watching the Outdoors

Participants	Number of Attempts		
	Touch Screen	Voice Command	Gesture Control
Participant 1	1	1	1
Participant 2	1	1	1
Participant 3	1	1	1
Participant 4	1	1	1
Participant 5	1	1	1

Table 8. Number of attempts by each participant for each input method, when eating finger food when a request to intervene was issued

In this condition, we note that no participant needed to repeat any of their input attempts. This was the only condition that had this observation. While participants were not looking at the road ahead, they still had their hands free and could easily say what they wanted the car to do, with no obstructions (unlike situations like the eating finger food situation).

Interview - Perceived Effectiveness and Accuracy of Input Method

We performed an interview at the end of the measured user study where we recorded the time taken a user took to complete the input method and the number of attempts a user needed to complete the input method successfully.

One of the questions we asked participants was how accurate and effective they felt a particular input method was. The idea behind this was in order to gain trust, we need to also understand what input methods are perceived accurate and effective in the various contexts users are placed under, and was there a difference in the perceived accuracy and effectiveness when compared with quantitative data (i.e. did participants find a particular input method more accurate than it actually was). It is important to note accuracy may not typically equate to ease of use (which we cover later in this paper), as an accurate and effective input method could possibly be difficult and cumbersome to use in reality.

From the perspective of touch screens, in general our participants found the touch screen to be very accurate and reassuring to use, making them like the touch screen a lot. Participants felt that the touch screen input was definitely recognized - this is likely because the touch screen involves a rather concrete action that can be seen, that is actually touching an icon on the screen, even in this paper prototype form. It also is likely because touch screens have been around for a while and have become ubiquitous in modern society, from mobile devices to in-car entertainment systems. This familiarity is likely what lead most people to feel like it would always work and was accurate. Some participants did mention that the touch screen allowed them to read their options faster than the voice over announcement which helped them react better. Comparing to our quantitative results, we note that touch screens were the only input method that consistently needed just one attempt - regardless of context. This is consistent with our findings from the participants as mentioned above.

In the case of voice control, our participants indicated in the interview that in general the voice control was alright but it wasn't as confidence inspiring as the touchscreen from the perspective of accuracy and effectiveness. They didn't feel that the voice commands would reliably get the job done, unlike the touchscreen. While one participant said that voice command did feel faster to respond, others brought up concerns such as accent limitations and concerns that the car won't understand the command or interpret the command wrongly, putting them in a potentially dangerous or unwanted situation. These are all valid concerns that have been addressed a lot (especially in the case of accents) in later generation natural language voice command systems (like Google Assistant and Amazon Alexa), however these systems aren't perfect too. Interestingly, we can say that participants raise these concerns likely from their past experience with using voice assistants that allow natural language input like Amazon Alexa or Google Assistant especially on earlier generations of these systems.

When compared with the quantitative results in terms of accuracy and speed, we see that one participant did need to repeat their attempt to input through the voice command when they were in conversation with someone. However, we also note that typically voice command did appear to be the fastest input method in all conditions except when in conversation with someone (tied with touch screen). Especially in the using mobile phone condition, voice command was clearly seen as the fastest option for 4 out of 5 participants. Hence it is interesting to see that raw input speed is not necessarily guaranteeing user satisfaction and allows users to trust and use a particular input system with confidence.

Gesture control was disliked by every participant unanimously. Our participants indicated that for one, it felt awkward and unnatural, leading them to be unsure if the gesture was successfully identified by the car (or rather in this case, the wizard of Oz). Since there was only feedback when there was an incorrect input method, participants were not entirely sure if the gesture detected was correct or otherwise. Interestingly enough, we did not give feedback to the participants when there was a success case in the voice command study, however, it is probably because participants typically spent longer in the case of gesture control to input, that they probably felt that if they needed to redo the gesture, it may take too long and put them in danger in a car - unlike the voice command where since it was fast enough to input, participants likely felt they still had enough time to repeat if needed. These views were not helped by the fact that 2 participants did have to repeat their inputs in the gesture control study, which likely had an effect on their perceptions on the accuracy and effectiveness of the input method. Arguably, one participant did say that it is plausible that gesture control is more effective and accurate for users with accents, where a voice control system may not work well.

Interview - Ease of Use of Input Methods

As part of our interview, we also asked our participants their views on the ease of use of the three different input methods. While it is possible to get an idea on the ease of use based on the time it took to complete the input for a particular input method (where a faster input method is most likely easier), it is still no match to actually talking to users and understanding their perceptions and experiences as well.

On the whole, our participants felt that the touchscreen was easiest to use. In most situations participants felt that they would end up using this input method. This is likely because of the familiarity of touch screens, considering touch screens have become present in many devices today and people have experienced using one. Having said that, our participants did mention that the one time that touch screens felt difficult to use was when they were eating finger food. While some were OK with touching the screen with dirty fingers, others were not. More so, participants did mention that anytime they had to hold something in their other hand before reaching to touch the screen (i.e. when holding a tablet/mobile phone, or even holding a food container), it felt difficult and cumbersome.

In terms of voice command, participants felt that it was easy enough to use - though not as easy as the touch screen. It was not easy to use when participants were in conversation with another party, however. Our participants mentioned that the voice command was very disruptive to a conversation, and at times like that, they did prefer alternative input methods like the touch screen and gesture control, since they could still continue on with the conversation. This was expected, however we did not foresee issues with voice commands when eating until our participants mentioned issues when trying to input through the voice command service when they had food in their mouth. One of the biggest benefits of having voice commands was when watching the outdoors, and when participants were eating and did not have food in their mouth - in situations like this, our participants told us that it allowed them to continue watching the outdoors, and not get the touch screen dirty with fingers covered in food grime.

We expected that gesture control will be universally disliked from the perspective of ease of use, based on the results obtained in the previous sections of this paper. This was indeed true when we did ask our participants as well. In general participants unanimously agreed that gestures were the worst input method, and it felt very awkward using them. While in some cases like in a conversation (where they could continue talking and still issue a gesture), or when they had food in their mouth and their hands were covered in food grime, it was nice to have the gesture control, however these two situations were not enough to make up for the fact that gestures were awkward to do. Participants also said that they intentionally drew the gestures slowly in the fear that the car won't be able to recognise the gesture. This is a valid concern, especially in a critical situation in the car, which also explains why gesture control was generally the slowest input method, in the earlier part of the results section of this paper. In summary, given all three options, our participants agreed that gesture control would be their last resort input method.

DISCUSSION

We definitely have obtained some interesting results as shown in the previous section of this paper. Based on these results, we would like to present our further analysis and reasoning, once we had successfully conducted the study.

As we reference Walch et. al [19] as one of the inspirations for this paper, it would only be fair for us to take a look to see how did our results compare with Walch et. al's results in general. It is understood that in their paper, speech was generally found to be slower than touch screens. Many drivers also didn't quite trust the voice system in that study as well.

By comparison, in our paper, we see that voice commands typically had a faster input time compared to touch screens in almost all conditions. We have run our study on a greater variety of conditions compared to Walsh et. al, showing us that voice worked reasonably well in all situations, and in some of these situations, it worked exceptionally well from a speed perspective. From the perspective of user trust, we looked into this with the use of the effectiveness and accuracy measure and interview questions in our study. From this, we see that voice commands came up to be not bad, but not as outstanding as touch screens. Having said that, it turned out that voice commands were way better from an effectiveness and accuracy, speed, and ease of use of gesture control.

On the topic of gesture control - we did not expect to see the results for this input method to be this negative. While there was expectation that this input method would likely need a little bit more time than others, we did not expect the difference to be as large as per indicated in our timed results here. More so, we expected users to like gesture control in more situations, such as when they were watching the outdoors, but as it turned out, people did prefer to use touch screens or voice typically and only would want to use gestures if there was no choice (i.e. hands are full/dirty and mouth is full with food).

We also acknowledge that in the case that a user took more than one attempt to input, the time recorded to complete the input would be the time taken to complete all attempts for that particular input method. We chose to measure it this way, as it mimics a real world situation if the car doesn't understand the input - since in that case, time is still being used up in the failed attempts as well. We know based on the results there were 3 instances of participants having to repeat an input method. Based on the results however, we do not see this methodology of recording time having a significant effect on our results - if at all, it helps to yield more realistic results in our study.

THREATS TO VALIDITY

While care has been taken to minimise threats to the validity of this study, we still would like to document the following potential threats to validity:

- **Participant diversity wasn't great**

  As we were limited to only selecting students in the class to participate in this study, we were not able to get an extremely diverse pool of participants. For one, we did not have many participants (being a class project. In addition, our participants typically were between the ages of 22-30, and were all computer science graduate students here at the University of Waterloo. This itself is a concern as it means our study was only made up of people who had a relatively strong tech background and were most likely to be accepting of new technology. Despite this group of people who would be most likely to enjoy new technology and be the most optimistic group of participants towards new technology, we did see the general dislike for gesture control in the context of autonomous cars - which is a relatively new idea. Our study presents preliminary results that gives an idea as to how people may accept these new input methods. Nevertheless, it would be interesting to see how this study would play out in a group of older, less technologically savvy participants.

- **No physically challenged participants**

  Unfortunately as part of our study, we were not able to test our ideas on physically challenged participants. Part of the appeal of autonomous vehicles, including ones that involve co-operative driving like in this study is that they would allow for the physically challenged to enjoy the freedom of mobility since there isn't a need to operate a traditional pedal and steering wheel setup. These people are going to be very important in the development of autonomous vehicles and hence need to be involved in studies like this.

- **The potential presence of the novelty effect**

  It is possible for us to say that some participants may not have had the novelty effect with gestures, while others may have had this novelty effect. At the same time, it is possible for us to say that participants are more likely to be very used to touch screens and voice commands since many other consumer electronic devices like mobile phones and tablets have these input methods.

  This may be a good thing in this case, to some extent as it would allow us to get ”raw” feedback from participants especially in the gesture control input method. While it is possible that some had the novelty effect when using this input method, the lack of prior experience using this input method is likely more representative of the general population anyway, which is something we want as autonomous cars are being built for everyone, not just the technologically savvy.

We hope that if this study was to be repeated and extended, ideally with a bigger, more diverse participant base, we would be able to suppress some of these threats to validity.

CONCLUSION

In conclusion, we see that based on our study, in most conditions, voice commands were the fastest, followed by touch screens and trailed by gesture control, in the case of inputting commands during a request to intervene by a Level 3 autonomous vehicle. At the same time, while voice commands are fast, they don't seem as accurate as touch screens, which only needed one attempt to input in every situation we tested, for every participant. Voice was the second most accurate, with gesture control trailing in third place. This observation was backed up by our participants who indicated that they too perceived touch screens to be the most effective and accurate. From an ease of use perspective, once again, the touch screen was seen as the easiest to use as well in most cases - though in some cases, the voice command was preferred, according to our participants. Participants also unanimously said that the gesture control was clunky and difficult to use, even in extreme cases where it wasn't easy to use the touch screen or voice command. There is still room for improvement, being an exploratory paper for a class. We hope to see this topic revisited with a more diverse participant pool to better understand how new input methods like gesture control and voice commands perform for different people.

REFERENCES

1. “Continental automotive gesture control system.” [Online]. Available: https://www.continental-automotive.com/en-gl/Passenger-Cars/Interior/Comfort-Security/Driver-Status/Gesture-Control
2. “J3016: Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems,” 2018. [Online]. Available: https://www.sae.org/standards/content/j3016_201401
3. “Mercedes-benz user experience: Revolution in the cockpit.” Jan 2018. [Online]. Available: https://www.mercedes-benz.com/en/mercedes-benz/innovation/mbux-mercedes-benz-user-experience-revolution-in-the-cockpit/
4. N. H. T. S. Administration, “Automated vehicles for safety,” Feb 2018. [Online]. Available: https://www.nhtsa.gov/technology-innovation/automated-vehicles-safety
5. S. Akyol, U. Canzler, and K. Bengler, “Gesture control for use in automobiles.”
6. R. Baldwin, “Audi takes its self-driving car where others dare not go,” Jul 2017. [Online]. Available: https://www.engadget.com/2017/07/11/audi-takes-its-self-driving-car-where-others-dare-not-go/
7. A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, “CARLA: An open urban driving simulator,” in *Proceedings of the 1st Annual Conference on Robot Learning*, 2017, pp. 1–16.
8. D. Etherington, “Waymo focuses on user experience, considers next steps,” Nov 2017. [Online]. Available: https://beta.techcrunch.com/2017/10/31/waymo-self-driving-ux/
9. D. Fagella, “Self-driving car timeline for 11 top automakers,” Jun 2017. [Online]. Available: https://venturebeat.com/2017/06/04/self-driving-car-timeline-for-11-top-automakers/

10. A. for Highway and A. Safety, *CARAVAN Public Opinion Poll: Driverless Cars*, Jan 2018. [Online]. Available: http://saferoads.org/wp-content/uploads/2018/01/AV-Poll-Report-January-2018-FINAL.pdf

11. A. J. Hawkins, “Americans still deeply skeptical about driverless cars: poll,” Jan 2018. [Online]. Available: https://www.theverge.com/2018/1/12/16883510/self-driving-car-poll-congress-bill-safety

12. T. Helldin, G. Falkman, M. Riveiro, and S. Davidsson, “Presenting system uncertainty in automotive uis for supporting trust calibration in autonomous driving,” in *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, ser. AutomotiveUI ’13. New York, NY, USA: ACM, 2013, pp. 210–217. [Online]. Available: http://doi.acm.org.proxy.lib.uwaterloo.ca/10.1145/2516540.2516554

13. M. Hutson, “People don’t trust driverless cars. researchers are trying to change that,” Dec 2017. [Online]. Available: http://www.sciencemag.org/news/2017/12/people-don-t-trust-driverless-cars-researchers-are-trying-change

14. T. Leggett, “Will we ever be able to trust self-driving cars?” Jan 2018. [Online]. Available: http://www.bbc.com/news/business-42710215

15. T. C. Press, “Half of canadians trust self-driving cars, survey says,” Sep 2017. [Online]. Available: https://www.thestar.com/business/tech_news/2017/09/14/half-of-canadians-trust-self-driving-cars-survey-says.html

16. D. Shapiro, “Nvidia powers mercedes-benz mbux, its next-gen ai cockpit,” Jan 2018. [Online]. Available: https://blogs.nvidia.com/blog/2018/01/09/mercedes-ces-2018/

17. R. M. van der Heiden, S. T. Iqbal, and C. P. Janssen, “Priming drivers before handover in semi-autonomous cars,” in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’17. New York, NY, USA: ACM, 2017, pp. 392–404. [Online]. Available: http://doi.acm.org/10.1145/3025453.3025507

18. T. Verge, Jan 2018. [Online]. Available: https://www.youtube.com/watch?v=eRISZlIFtBA

19. M. Walch, T. Sieber, P. Hock, M. Baumann, and M. Weber, “Towards cooperative driving: Involving the driver in an autonomous vehicle’s decision making,” in *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, ser. Automotive’UI 16. New York, NY, USA: ACM, 2016, pp. 261–268. [Online]. Available: http://doi.acm.org/10.1145/3003715.3005458

20. N. Zon and S. Ditta, *Robot, Take the Wheel*. Mowat Centre for Policy Innovation, University of Toronto, 2016.