Gesture and Voice Commands to Interact With AR Windshield Display in Automated Vehicle: A Remote Elicitation Study

Nabil Al Nahin Ch
University of New Hampshire
Durham, New Hampshire, USA
nabilalnahin.ch@unh.edu

Diana Tosca
New York University
New York, New York, USA
dt2290@nyu.edu

Tyanna Crump
Wellesley College
Wellesley, Massachusetts, USA
tcrump@wellesley.edu

Alberta A. Ansah
University of New Hampshire
Durham, New Hampshire, USA
alberta.ansah@unh.edu

Andrew L. Kun
University of New Hampshire
Durham, New Hampshire, USA
andrew.kun@unh.edu

Orit Shaer
Wellesley College
Wellesley, Massachusetts, USA
oshaer@wellesley.edu

ABSTRACT

Augmented reality (AR) windshield display (WSD) offers promising ways to engage in non-driving tasks in automated vehicles. Previous studies explored different ways WSD can be used to present driving and other tasks-related information and how that can affect driving performance, user experience, and performance in secondary tasks. Our goal for this study was to examine how drivers expect to use gesture and voice commands for interacting with WSD for performing complex, multi-step personal and work-related tasks in an automated vehicle. In this remote unmoderated online elicitation study, 31 participants proposed 373 gestures and 373 voice commands for performing 24 tasks. We analyzed the elicited interactions, their preferred modality of interaction, and the reasons behind this preference. Lastly, we discuss our results and their implications for designing AR WSD in automated vehicles.

CCS CONCEPTS

• **Human-centered computing → Mixed / augmented reality**; **User studies**.

KEYWORDS

Windshield display, gesture, voice commands, automated driving, head-up display

ACM Reference Format:
Nabil Al Nahin Ch, Diana Tosca, Tyanna Crump, Alberta A. Ansah, Andrew L. Kun, and Orit Shaer. 2022. Gesture and Voice Commands to Interact With AR Windshield Display in Automated Vehicle: A Remote Elicitation Study. In *14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '22), September 17–20, 2022, Seoul, Republic of Korea.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3543174.3545257

AutomotiveUI '22, September 17–20, 2022, Seoul, Republic of Korea

ACM ISBN 978-1-4503-9415-4/22/09…$15.00
https://doi.org/10.1145/3543174.3545257

1 INTRODUCTION

Around the world, people spend a significant amount of time traveling for work [9, 22] and they often perform both work-related and personal tasks during their commute even when driving [55, 67]. Since driving requires visual attention and manual action, the types of non-driving tasks people can perform while driving are limited and they can negatively affect driving performance. However, introduction of conditionally automated (SAE level 3-4) and eventually fully automated (SAE level 5) [13] vehicles will allow drivers to safely engage in different non-driving tasks [32, 55]. However, the user interface (UI) of the vehicles currently on the road is not designed to support that. Automotive UI has been changing over the years as the number of functions for infotainment, entertainment and driver assistance increased [28, 64]. We are likely to see even more changes in automotive UI and interaction techniques to support non-driving activities in automated vehicles [31].

Traditionally, the center console display or dashboard instrument cluster is used to present information to the driver in a car. In recent years, heads-up displays (HUDs) have also been used to project useful driving related information like speed and navigation instructions on a small part of the windshield. Augmented reality (AR) windshield display (WSD) transforms the windshield of a car into transparent display by superimposing information over the driving scene [25, 66]. WSDs offer promising ways to display both driving related information (e.g. speed and navigation instructions) and non-driving task information, especially in automated vehicles where drivers will be able to engage safely in different non-driving related tasks for some periods of time. In addition to the obvious benefit of having extended display real estate, WSDs allow drivers to engage in non-driving tasks without having to fully take their eyes off the road. But WSD introduces a challenge in terms of ways to interact with it since the distance between the driver and the WSDs makes the traditional haptic or touch interface less convenient. Researchers have investigated the potential and challenges of using AR windshield [20, 24, 59], and different ways drivers can interact with WSDs like gesture [2, 19], speech [75] and gaze [58]. But we still do not know how people expect to interact with WSDs to engage in distinct types of complex multi-step non-driving tasks in automated vehicles. In this remote unmoderated online elicitation study, we look into this problem from the drivers' point of view, in order to explore how they expect to use gesture and voice commands to perform certain non-driving tasks in automated vehicles. Our study examines the following research questions:

RQ1: How do drivers *expect* to use gestures and voice commands to interact with windshield displays of automated vehicles to perform non-driving tasks?

RQ2: How do drivers *prefer* to interact with windshield displays of automated vehicles to perform non-driving tasks?

The rest of the paper is structured as follows. Section 2 provides an overview of research on AR WSD and how user-elicitation studies have been used to study gesture and voice commands for interacting with vehicle user interface. Section 3 describes the study in details, and we present the findings of the study in section 4. In section 5, we discuss design implications of our findings and limitations of our study. We end the paper with section 6 where we conclude our thoughts on the study and discuss future work.

2 RELATED WORK

2.1 Augmented Reality Windshield Display

Even though several automakers have added heads-up display (HUD) feature to their vehicles [51], there are still no commercially available vehicles that extend this feature to the whole windshield with AR WSD. Research shows that WSD technology has potential to enhance drivers' safety and in-car experience, specially in automated and conditionally automated vehicles [20, 24, 59]. Providing vehicle and navigation information using WSD can increase trust and acceptance in automated vehicles [17, 76] and help elder drivers to drive safely [29]. WSD can also be used to effectively communicate uncertainty information in automated driving [33]. Since looking at WSDs requires similar gaze angles as looking at the road, this technology can be used to enhance driver's spatial and situational awareness [40, 70], even in low visibility weather conditions [12, 41]. In conditionally automated vehicles, one of the topic of interest is how to assist drivers in taking over control of the vehicle [49, 57]. WSD technology can be used to positively influence takeover process [42] and shorten takeover time in case of vehicle malfunctions [17]. Additionally, WSDs can also assist drivers in performing non-driving task (NDT). Research shows that while performing NDTs using WSDs, drivers can better maintain their attention on the road [37], which leads to enhanced situational awareness [21] and improved NDT performance [54]. Drivers also experience lower cognitive workload while performing NDT [38] and find it easier to take back control of the vehicle [38, 63]. While these studies demonstrate how AR applications and WSD technology can be used as the UI of automated vehicles, it is important to gain insight into how people expect to interact with WSD. Understanding drivers' expectations and preferences has implications for, and should inform the design of future automotive UI.

Different gesture and speech based interfaces have been explored for interacting with head-up display or AR windshield display technology. The focus of research so far has been investigating how commonly used gestures and voice commands can be applied to interact with this unconventional interface. Researchers explored gestures on windshield [2] and finger gestures on steering wheel surface [36] for non-driving related tasks like media and climate control using head-up display. Some researchers took a different approach by restricting gestures to finger pointing so that drivers won't have to release their hands from steering wheel in order to interact with the windshield display [8, 19]. Similarly, research on speech interaction has been limited to exploring simple non-driving task performed on windshield display. Wang et al. proposed a steering wheel mounted button and mic to use voice commands to control a virtual assistant displayed on the windshield [73]. Other studies investigated texting and selecting from choice list using speech interaction and head-up display to evaluate the effect on driving safety and multitask performance [68, 75]. Researchers also compared gesture and voice command with other interaction modalities for head-up display. Betancur et al. and Angelini et al. found these two interaction modalities comparable in terms of user acceptance and perceived usability, but haptic method was better accepted compared to both gesture and speech [3, 6]. In this study, we focus on complex personal and work-related tasks those involve multiple steps to complete.

2.2 User-Elicitation Studies: Gestures and Voice Commands

User-elicitation studies are a particular type of participatory design approach in which end-users are shown *referents* (effects of an action) and are asked to demonstrate the corresponding *signs* (interactions that result in the specified referent) [46, 69, 78]. Research shows that end-users prefer interactions produced using user-elicitation approach compared to interactions designed solely by experts [48]. Even though user-elicitation studies are usually confined to a lab, online user-elicitation studies have also been successfully used to produce set of user-defined interactions [1].

Elicitation methods have been successfully applied to generate different types of user-defined gestures for various interfaces. Elicitation studies have been used to investigate unistroke gestures [77], single-hand microgestures [11], multi-touch gestures on large and small surfaces [5, 15, 18, 34, 78], gestures using tangible interfaces [69], motion gestures for mobile computing [62], and gestures for augmented reality (AR) [35, 56] and virtual reality (VR) [80] environment. This method has also been used to study gestures for various in-vehicle user interfaces. Burnett et al. conducted an elicitation study to understand how drivers use swipe gestures on in-vehicle touch screen interface [10]. Several other studies examined how drivers use air-gestures to perform various commonly used non-driving tasks like climate control, navigation control and media control using vehicle infotainment system [16, 27, 43, 79]. Experimenting with unconventional gesture interactions like interactions via fabric-based wearable device [50] and performing gestures on steering wheel surface [4] has also used user-elicitation method. In recent years, we have started to see user elicitation being applied to study gesture interaction in automated vehicles. In their effort to study gesture interaction in highly automated vehicle, Weidner and Broll investigated how drivers interact with stereoscopic 3D display [74] and Lin studied gestures for adjusting vehicle dynamics [39].

User elicitation has also been applied to study speech interaction for various interfaces. Volkel et al. used this method to elicit dialogues for interacting with voice assistants [72]. Other researchers used user elicitation to study speech interaction as part of a multimodal human-computer interaction interface [44, 60, 61]. In their study, Hoffmann et al. elicited voice commands along with surface gestures and mid-air gestures for interacting with smart home and they found that people preferred speech interaction compared to mid-air gestures [26]. Other studies explored uesr-defined voice commands and gestures for using a web browser on TV [46] and controlling an unmanned aerial vehicle (UAV) [53]. Researchers applied elicitation method to study voice commands in vehicles for maneuver-based interventions [14] and various non-driving tasks like phone call, controlling climate, navigation, media etc [7].

These studies demonstrate that user elicitation can be applied to explore gesture and speech interaction in different contexts. Even though elicitation studies have been conducted to investigate gesture and voice command for interacting with different in-vehicle user interfaces, to our knowledge, this is the first study focusing on AR windshield display as the user interface of automated vehicles. In contrast to most automotive UI elicitation studies which were conducted in a lab, our unmoderated online elicitation study aimed to reach a relatively large and diverse group of participants. The referents we chose were also more complex and of open nature compared to other studies because our goal was not to generate a set of user-defined interactions, but to examine how drivers expect to interact with this new and unconventional user interface. Contribution of this study is the findings which demonstrate user expectation and preference for gesture and voice based interaction with WSD in automated vehicle.

3 METHOD

We conducted a within subjects remote unmoderated user elicitation study to examine how drivers expect to use gestures and voice commands to interact with AR windshield display of automated vehicles. Participants were presented with various referents accompanied by a pair of images depicting the changes in vehicle UI for those referents and they were asked to demonstrate (using video recordings) which voice or gesture interaction they would use for each referent. We developed a web app for conducting remote video-based elicitation studies. After looking at the images and reading description of each referent, participants could easily record their responses on the same web app without using any external application. Figure 1 shows an example of a referent presented on the web app and the interface for participants to record their video responses for gesture and voice commands.

3.1 Referents and experimental task

From previous studies, we know that drivers perform both work related and personal activities while commuting [67]. So in our study, we presented participants with scenarios related to both work and personal tasks. Our study presented 24 referents (effect of performing a task) divided into four scenarios. Each scenario had five to eight referents (see table 1). Each participant was presented with referents from two randomly selected scenarios; one personal task scenario from *Audiobook* and *Karaoke*, and one work related scenario from *Podcast* and *Presentation.* So each participant demonstrated interactions for either 11 or 13 referents depending on which scenarios they were presented with. When presenting a referent to the participant, the task that was being performed for that referent was described in text and the referent (effect of that task being performed) was shown using two images of AR windshield display. For example, in Figure 1, we see the images presented to the participants for the referent *open karaoke application* from the *karaoke* scenario. For this referent, the task described to the participant was "How would you open the karaoke application? What input command would result in the following user interface?". The images show the effect of completing this task. The first image (top) of Figure 1 shows the AR windshield display of an automated vehicle where the car is in automated mode. The second image (bottom) shows the karaoke application on the AR windshield display, which is the referent. For each referent presented to the participants, they were asked four questions.

(1) Using a voice command, how would you complete this task?
(2) Using a gesture, how would you complete this task?
(3) What is your preferred interaction method for this task?
(4) Why do you prefer this interaction method for this task?

For the first two questions, participants recorded two videos demonstrating a gesture and a voice command they would use to complete the task. For the third question, the participants chose either gesture or voice command as their preferred interaction method for that task. For the last question, participants wrote the reasoning behind their choice of preferred interaction method.

3.2 Interacting with AR windshield

It may be difficult to conceptualize the idea of AR WSD used as the user interface of automated vehicles since not many people are familiar with these emerging technologies. So, we developed images depicting how completing different tasks might look like on the AR windshield display. The design of AR windshield display we used in the study (see Figure 1) was speculative but it was based on research conducted in this domain and commercially available products on the market that use similar technology. Most of the vehicles with heads up displays [51] show driving related information on a small part of the windshield directly in front of the driver. Similarly, the AR windshield display design we developed presented speed limit, vehicle speed, status of vehicle automation, and the amount of time left before the driver will need to take over the control of the car on the bottom part of the windshield in front of the driver. The non-driving task related information was presented on the passenger side of the windshield so that the driver can monitor the road ahead while engaging in other tasks. All driving and non-driving related information was projected on top of a driving scene, making the windshield a large transparent display.

3.3 Procedure

At the beginning of this online elicitation study, the first web page briefly introduced the study to the participants. The introduction page described the aim of the study and the participants were informed that they will have to record video responses to complete the study. After signing the consent form, the participants were introduced to the AR windshield display and automated vehicle technology. They watched a one minute video of how interacting with AR windshield display in an automated vehicle may look like. Then the participants received a detailed description of the tasks they were expected to perform in the study. After that the participants watched a tutorial of the interface they will be using to record

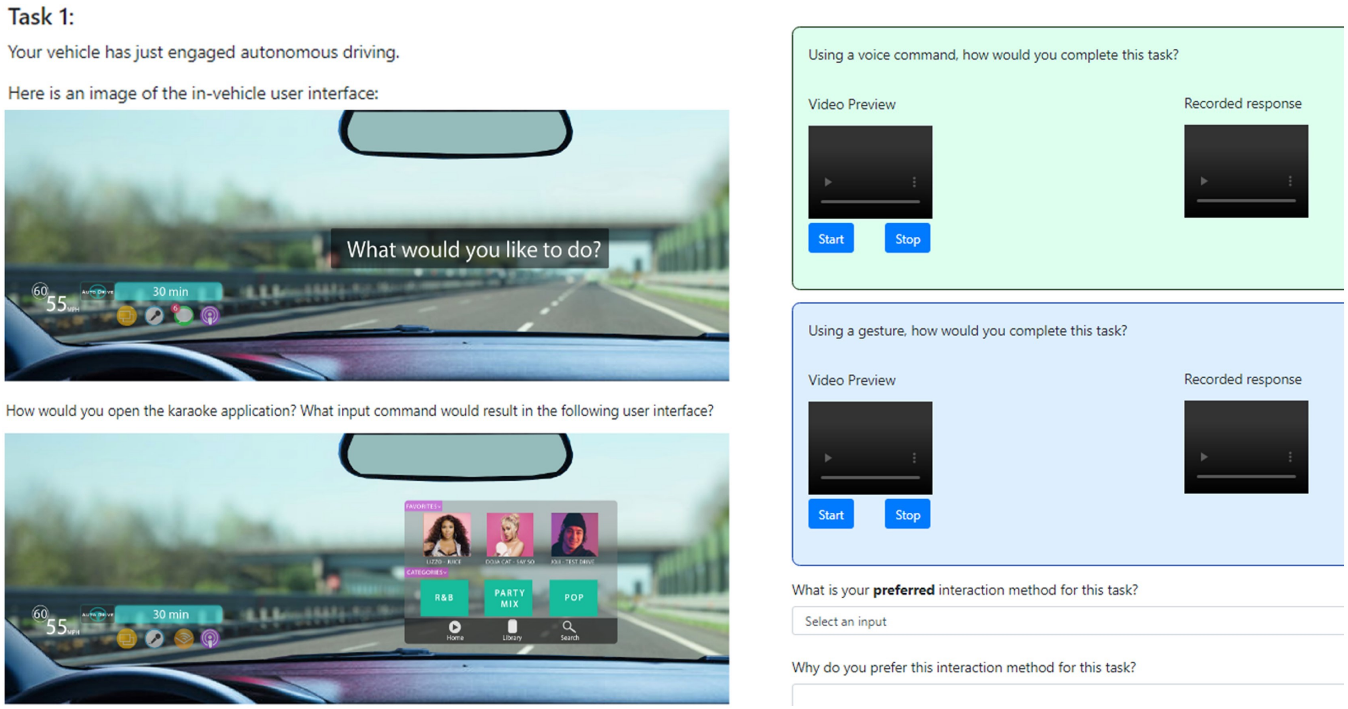

Figure 1: Web app interface for participants to see the referents (left) and record their responses (right). Driving information and non-driving task projected on augmented reality windshield display to present the referent *Open karaoke application*.

their video responses. They also got a chance to practice recording video responses in order to get familiar with the process. After that the participants were presented with the first referent of a randomly selected personal task-related scenario (*audiobook* and *karaoke*). At this time, they also received a brief description of the scenario and the related tasks. For example, in the *audiobook* scenario, the referents included opening the audiobook application, playing a specific audiobook, bookmarking a section of the audiobook, rewinding 30 seconds, and exiting the application to go to the main menu. All the referents of a scenario were presented in sequence. After finishing responding to referents of the first scenario, the process repeated for a randomly selected work-related scenario (*podcast* and *presentation*). The participants could only proceed to the next referent after recording two video responses for gesture and speech interaction, choosing their preferred interaction for that referent, and describing why they preferred that interaction method. They completed the study by answering some demographic questions.

3.4 Participants

We recruited 43 participants using the online subject recruiting platform Prolific [52]. Since our goal was to examine how drivers expect to interact, we recruited participants who reported that they had a valid driver license. In this paper we present results based on the responses of 31 participants who proposed gestures and voice commands for all of the assigned referents. Of these 31 participants 18 identified as men, 13 as women. They were aged between 19 and 63 ($M = 29, SD = 11$). They were located in Africa, Europe, and North America. Since each participant was presented with referents from two randomly selected scenarios (discussed in section 3.1), among the 31 participants, 16 participants proposed interactions for referents from *audiobook* and *presentation* scenarios and 15 participants proposed interactions for referents from *karaoke* and *podcast* scenarios. All 43 participants were compensated with $7 for their time.

4 RESULTS

4.1 User-defined gestures and voice commands (RQ1)

Thirty one participants proposed a total of 373 gestures and 373 voice commands for 24 referents (11 referents x 15 participants + 13 referents x 16 participants = 373). Out of those, 278 gestures and 284 voice commands were distinct. We considered distinctness on a per-referent basis, meaning a gesture or voice command proposed for two different referents were considered two distinct interactions. Multiple participants proposing the same gesture or voice command for the same referent was not considered distinct. Table 1 presents the popular gestures and voice commands for each referent that were proposed by at least two participants. In cases where the only selected interaction was proposed for multiple referents or none of the participants agreed on any interaction, we selected one gesture or voice command from all proposals. Gestures from Table 1 are shown in Figure 2.

4.1.1 Agreement rate. To compute the degree of consensus among participants, we calculated agreement rate A_r using the formula proposed by Vatavu and Wobbrock [71].

Table 1: Most commonly suggested gestures and voice commands per referent and the number of participants who suggested them.

No.	Referent	Gesture	Voice command
Audiobook (16 participants)			
1	Open audiobook application	Tap(3), Swipe left then tap(2)	Open audiobook(7)
2	Play audiobook <audiobook name>	Tap(6), Hold one finger upright(2)	Play <audiobook name>(4), Resume <audiobook name>(2)
3	Bookmark audiobook section	Tap(3), Two fingers crossed(1)	Add bookmark(1)
4	Rewind 30 seconds	Swirl counter clockwise(3), Tap(3), Wave left(2)	Rewind 30 seconds(4), Rewind(3), Go back 30 seconds(2)
5	Exit audiobook application	Make X with both hands(2), Close palm(2), Wave right(2)	Exit application(2), Return to main menu(2)
Karaoke (15 participants)			
6	Open karaoke application	Tap(4), Mimic holding mic(2), Mimic holding mic and move left and right(2)	Open karaoke(5), Karaoke(2)
7	Select category <category name>	Tap(6), Hold index and little finger upright(1)	Open <category name>(3), <category name>(2)
8	Play song <song name>	Tap(5), Hold seven fingers upright(2)	Play <song name>(8)
9	Play vocals in background	Tap(5), Hold two fingers upright(1)	Play vocals(3), Enable vocals(2)
10	Exit karaoke application	Wave left(4), Wave right(3)	Exit karaoke application(1)
Podcast (15 participants)			
11	Open podcast application	Tap(3), Using two fingers make a circle and tap in the center(1)	Open podcast(5), Open podcast application(3), Open the podcast app(2)
12	Play podcast <podcast name>	Hold three fingers upright(3), Swipe up then tap(3), Tap(2)	Play <podcast name>(6)
13	Bookmark podcast section	Tap(2), Make a plus sign with two fingers(1)	Add a bookmark here(1)
14	Skip to section <section name>	Tap(4), Point left then make a circle with palm(1)	Skip to summary(2)
15	Text <colleague name>a link of podcast	Tap(5), Make a T with two fingers(1)	Send podcast to <colleague name>(1)
16	Exit podcast application	Wave left(3), Make X with both hands(2)	Exit application(2), Close podcast open main menu(1)
Presentation (16 participants)			
17	Open presentation application	Tap(6), Show palm then thumbs up(1)	Open presentation(8), Open presentation application(2), Open my presentation(2)
18	Open presentation <presentation name>	Tap(4), Swipe left then tap(2)	Open <presentation name>(8), Open <presentation name> presentation(2)
19	Start the timer	Tap(4), Make the ok gesture(2)	Start timer(3), Start the timer(2)
20	Go to next slide	Wave left(5), Move hand pointing right(1)	Next slide(8), Next(2)
21	Pause the timer	Tap(4), Show palm(3), Move palm forward(2)	Stop timer(4), Pause the timer(3), Pause(3), Pause timer(2)
22	Display all slides	Tap(4), Close all fingers from open position(2), Open fingers from closed position(2)	Display all the slides(2)
23	Get feedback on presentation	Tap(5), Make two fists(1)	Feedback(2), Give me feedback on my presentation(2)
24	Exit presentation application	Wave right(3), Wave downward(2)	Close the presentation application(1)

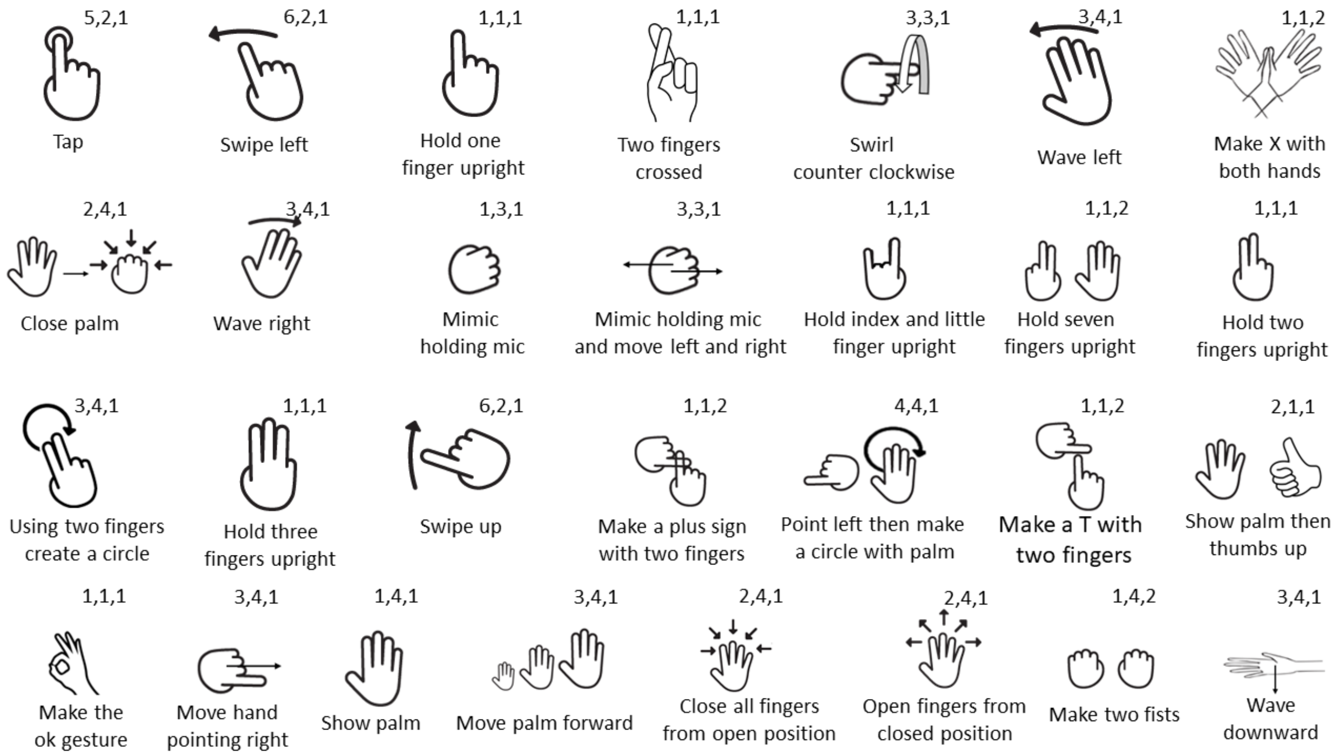

Figure 2: Popular gestures proposed by the participants. Gestures are presented as they appear in Table 1. The number on top of the gesture represents in which category along 3 dimensions (Table 3) they fall into. For example, the numbers 5,2,1 next to *tap* gesture means it falls into 5^{th} and 2^{nd} category in dimension *form* and *nature* and the gesture requires 1 hand to perform.

$$A_r = \frac{|P_r|}{|P_r|-1} \sum_{P_i \subseteq P_r} \left(\frac{|P_i|}{|P_r|}\right)^2 - \frac{1}{|P_r|-1}$$

In this equation, A_r is the agreement rate for referent r, P_r is the set of proposed interactions for referent r, and P_i is the subset of identical proposed interactions for that referent. The lower bound of the range for A_r is 0 (if all participants propose unique gestures/voice commands for the referent) and the upper bound is 1 (if all participants propose the same gesture/voice command). For each referent, we calculated one agreement rate for the gestures and another for the voice commands proposed by the participants.

Let's consider agreement rate for the referent *Rewind 30 seconds* as an example. Sixteen participants proposed voice commands and gestures for this referent. Among those, there were 11 distinct gestures and 10 distinct voice commands. Groups of three, three, and two participants proposed three unique gestures and each of the other eight unique gestures were proposed by one participant. For voice commands, groups of four, three, and two participants proposed three unique ones and each of the other seven unique voice commands were proposed by one participant. Agreement rate of gestures and voice commands for this referent was calculated using the following equations.

$$A_{rewind-30-seconds(gesture)} = \frac{16}{15}\left(8\left(\frac{1}{16}\right)^2 + \left(\frac{3}{16}\right)^2 + \left(\frac{3}{16}\right)^2 + \left(\frac{2}{16}\right)^2\right) - \frac{1}{15} = 0.06$$

$$A_{rewind-30-seconds(voice-command)} = \frac{16}{15}\left(7\left(\frac{1}{16}\right)^2 + \left(\frac{4}{16}\right)^2 + \left(\frac{3}{16}\right)^2 + \left(\frac{2}{16}\right)^2\right) - \frac{1}{15} = 0.08$$

Agreement rates for all referents are presented in Table 2. Agreement rate of gestures for different referents ranged between 0.01 and 0.14 ($M = 0.07, SD = 0.04$) and agreement rate of voice commands ranged between 0 and 0.27 ($M = 0.08, SD = 0.09$). We did not find any significant difference in agreement rate between gesture and voice command using a two sample t-test ($t(30) = -0.62, p = 0.54$).

4.1.2 Gesture classification. In order to better understand the gestures proposed by the participants, we manually classified them

Table 2: Gesture and voice command agreement rate and preferred interaction for each referent.

Scenario	No.	Referent	Agreement rate		Preferred interaction	
			Gesture	Voice command	Gesture	Voice command
Audiobook (16 participants)	1	Open audiobook application	0.03	0.18	19%	81%
	2	Play audiobook <audiobook name>	0.13	0.06	31%	69%
	3	Bookmark audiobook section	0.03	0.00	12%	88%
	4	Rewind 30 seconds	0.06	0.08	44%	56%
	5	Exit audiobook application	0.03	0.02	31%	69%
Karaoke (15 participants)	6	Open karaoke application	0.08	0.10	7%	93%
	7	Select category <category name>	0.14	0.04	7%	93%
	8	Play song <song name>	0.10	0.27	0%	100%
	9	Play vocals in background	0.10	0.04	27%	73%
	10	Exit karaoke application	0.09	0.00	33%	67%
Podcast (15 participants)	11	Open podcast application	0.03	0.13	0%	100%
	12	Play podcast <podcast name>	0.07	0.14	13%	87%
	13	Bookmark podcast section	0.01	0.00	27%	73%
	14	Skip to section <section name>	0.06	0.01	20%	80%
	15	Text <colleague name>a link of podcast	0.10	0.00	7%	93%
	16	Exit podcast application	0.04	0.01	47%	53%
Presentation (16 participants)	17	Open presentation application	0.13	0.25	25%	75%
	18	Open presentation <presentation name>	0.06	0.24	31%	69%
	19	Start the timer	0.05	0.03	19%	81%
	20	Go to next slide	0.08	0.24	69%	31%
	21	Pause the timer	0.08	0.11	50%	50%
	22	Display all slides	0.07	0.01	50%	50%
	23	Get feedback on presentation	0.08	0.02	12%	88%
	24	Exit presentation application	0.03	0.00	37%	63%

along three dimensions (Table 3). Based on the taxonomy of gestures performed on surface [78] and in AR environment [56], we proposed characterization dimensions that are relevant for interacting with AR WSD. Table 3 presents our proposed taxonomy of gestures for interacting with AR WSD. The *form* dimension consists of six categories that describe whether the hand pose changed or remained the same and whether the hand moved or remained in the same location while performing the gesture. It is applied separately to right and left hand in a two-handed gesture. *Tap* and *swipe* are special cases of *static pose and path*. We distinguished them because of their similarity to using touch surface. The *nature* dimension consists of four categories; *symbolic*, *touch*, *metaphorical*, and *abstract*. Gestures visually depicting a symbol were categorized as *symbolic* gestures. For example, the gesture *hold two fingers upright* like the letter 'V' for the referent *play vocals in background* can be considered a *symbolic* gesture. Gestures pretending to act on a touch surface by pointing finger towards the windshield were classified as *touch* gestures. All the *tap* and *swipe* gestures fall into this category. Gestures expressed through a metaphor were categorized as *metaphorical* gestures. For example, the gesture *mimic holding mic* for the referent *open karaoke application* can be considered a *metaphorical* gesture. Rest of the gestures were arbitrary in nature and were considered as *abstract* gestures. Some complex gestures can fall into multiple categories in the same dimension. For example, the gesture *tap and move both hands like opening a book* for the referent *open audiobook application* falls into both *touch* and *metaphorical* categories in the *nature* dimension. Based on which hand the participant used for the gesture, the *handedness* dimension was divided into three categories; *right*, *left*, and *both*.

Proportion of gestures in each category along three dimensions are presented in Figure 3. In the *form* dimension for right hand, large proportion of gestures were *tap* (37%) and *static pose and path* (32%). For left hand, *static pose and path* (43%) and *static pose* (32%) were most common. In the *nature* dimension, gestures were dominantly *touch* (50%) and *abstract* (39%). For *handedness* dimension, gestures were primarily *right* (72%) handed and gestures using *both* (17%) hands were more common than *left* (11%) handed gestures.

4.1.3 Voice command classification. In order to better understand the voice commands proposed by the participants, we manually classified them along four dimensions (Table 4) based on the taxonomy proposed by Hoffmann et al. [26]. Table 4 presents the taxonomy of voice commands for interacting with AR WSD. The *form* dimension consists of four categories that describe the number of words used for a voice command and whether the command uses sentence structure. We considered any name in the voice command like name of an audiobook, podcast, presentation, or song as one word. The *nature* dimension consists of two categories; *action* and *state*. *Action* voice commands utter the action to execute and *state* voice commands express the desired condition. For example, the voice command *karaoke* for the referent *open karaoke application* falls into *state* category while the voice command *open*

Table 3: Taxonomy of gestures for interacting with AR windshield display.

Dimension	Category	Description
Form	Static pose	Hand pose is held in one location.
	Dynamic pose	Hand pose changes in one location.
	Static pose and path	Hand pose is held and hand relocates.
	Dynamic pose and path	Hand pose changes and hand relocates.
	Tap	Static pointing pose moving toward display
	Swipe	Static pointing pose moving across display
Nature	Symbolic	Gesture visually depicts a symbol.
	Touch	Gesture pretends to act on touch surface.
	Metaphorical	Gesture is metaphorical.
	Abstract	Gesture mapping is arbitrary.
Handedness	Right	Gesture performed using right hand.
	Left	Gesture performed using left hand.
	Both	Gesture performed using both hands.

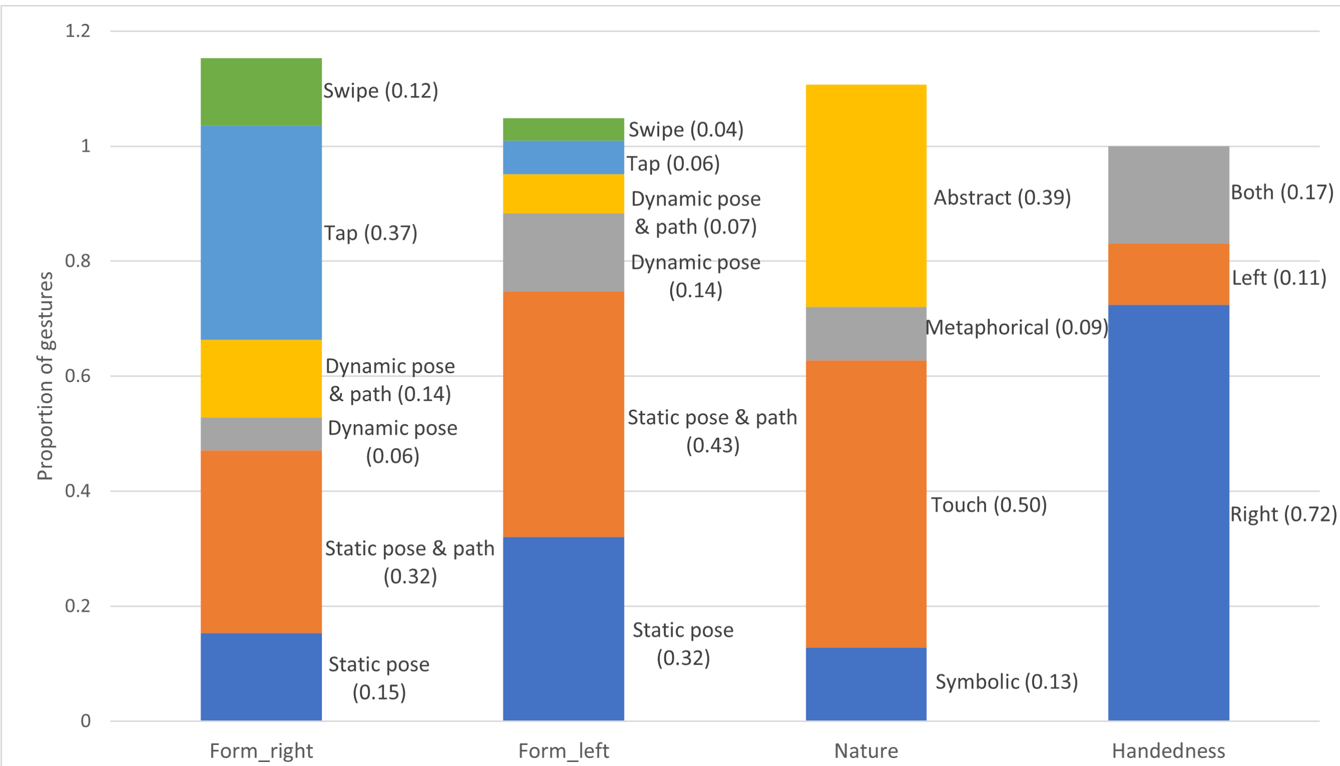

Figure 3: Percentage of gestures in each category along the three dimensions described in Table 3. The *form* dimension has been calculated for each hand separately. Summation of all gestures in a dimension can be larger than 1 because some gestures fall into multiple categories in the same dimension.

karaoke falls into *action* category. The *context* dimension describes whether specific context is needed for executing the command. For example, the voice command *return to main menu* for the referent *exit audiobook application* falls into *no-context* category while the voice command *exit application* falls into *in-context* category. Lastly, voice commands were categorized along the *complexity* dimension based on whether they consist of a single command or multiple commands. For example, the voice command *exit karaoke return to main menu* for the referent *exit karaoke application* falls into *compound* category while the voice command *turn off* falls into *simple* category.

Proportion of voice commands along four dimensions are presented in Figure 4. In *form* dimension, most voice commands fall into *sentence* (52%) and *two words* (33%) categories. For *nature*, *context*, and *complexity* dimensions, most of the voice commands stated an *action* (89%), needed to be considered *in-context* (64%), and consisted of a *simple* (86%) command.

4.2 Interaction preference (RQ2)

For 21 out of the 24 referents, participants preferred voice commands over gesture interaction. Table 2 presents the percentage of participants who preferred gesture or voice commands for performing each task. On average, participants preferred voice commands over gestures in 74% (276/373) of the instances for interacting with augmented reality windshield display. We analyzed the number of instances participants used specific words to describe why they preferred an interaction modality and found that the ease of use was by far the most common theme. Among the 276 instances where participants preferred voice commands, they often used words like *easy/easier/simple/simpler* (124) to describe interaction using voice commands and the word *hard/difficult* (22) to describe gesture interactions. For example, one participant wrote *"Easier to say, harder to show karaoke."*. Participants also thought voice commands were more *precise/accurate* (29), easier to *understand* (12), and *quicker/faster/fast* (22) to perform. One example of such opinion: *"Easier, quicker, more precise than using a gesture for such a task."*. Not requiring to use *hand/hands* (16) for interaction was also appreciated by the participants; *"I think it would be safer, considering I have to put my hands on the wheel, even if I assume the car would alert me in time."*. Some of these themes were also common among the 97 instances where participants preferred gesture interaction over voice commands. Some participants found gesture interaction *easy/easier/simple/simpler* (41) for performing certain tasks; *"Since it's a long name I find it easier to point and click on the podcast I want to play."*. Other words participants used to explain why they preferred gesture interactions were *fast/faster* (23) and *convenient* (5). One participant wrote *"I think it's faster and more convenient, considering the fact I don't have to put my hands on the wheel."*.

5 DISCUSSION

In this elicitation study, participants proposed 373 gestures and 373 voice commands for performing 24 tasks using AR WSD. Our results show no significant difference between the agreement rates of gestures and voice commands. The consensus among participants for both gestures and voice commands were low compared to previous studies [7, 26, 56]. This could be because some of the referents in this study were complex and specific in nature. So the participants had to think of a complex gesture or come up with a voice command that conveyed some specific information. Since the participants never used an AR WSD in real life, the elicited interactions were less likely to be shaped by legacy bias [47], which in turn may have resulted in diverse gestures and voice commands.

We found that a large proportion of gestures for both right and left hand falls into *static pose and path* category. Proportion of *tap* gestures are comparatively higher for right hand and proportion of *static pose* gestures are higher for left hand. This suggests in a two-handed gesture, people often used their right hand for movement while keeping left hand pose static. This is consistent with existing research on two-handed interaction [65], which shows that in case of asymmetric bi-manual actions, dominant hand often operates in spatial reference to the action of non-dominant hand [23].

For interacting with WSD, participant mostly used *simple* voice commands that stated an *action* to perform. A large proportion of voice commands needed to be considered *in-context*. For example, to execute the voice command *rewind 30 seconds*, the system needs to know which application the participant is referring to. We've also observed participants use similar gestures and voice commands to perform similar tasks in different scenarios (across different applications). This emphasizes the importance of contextual awareness in designing such interaction interfaces.

Results of the study clearly show that participants preferred voice commands over gesture interaction. This is in accordance with previous studies that found similar preference for smart-home [26] and in-vehicle [7] interactions. Whether the participants preferred gesture or voice commands, the main factor behind participants' preference was ease of use. For some participants it meant performing a task fast and for others it meant ease of referring to a specific name or action directly through a voice command or gesture. This suggests that even though more participants may prefer voice interactions for most of the tasks, preference still depends on the individual and the task they are performing. It is worth noting that there are scenarios, which we didn't study here, in which gesture might be preferred to voice commands. For example, when the driver is participating in a meeting while in the car or engaging in a conversation with passengers, they might not want to interrupt the conversation by using voice commands. Also, in the case of simple tasks, gestures might be much quicker to perform, and thus more convenient, than spoken instructions (cf. [45]).

Even though the participants never used AR WSD in an automated vehicle in real life, we still see some legacy bias affecting the elicited interactions. A large proportion of gestures proposed were similar to touch surface gestures. Legacy bias can limit uncovering novel interactions [47], but it can also help users to adopt new interfaces [30]. Designing air-gestures based on surface gestures or using touch surface for interacting with AR WSD may make it easier for drivers to adopt this new and unconventional interface.

5.1 Limitations

One limitation of this study is that it was conducted online. Interactions and preference might be different within a vehicle because of the driving context. We did not consider scenarios where speech modality is engaged, like the driver is talking on the phone or conversing with passengers. Finally, in this study, we had participants from different countries in Europe, North America, and Africa, but we did not consider how cultural differences might affect the way drivers interact or their preference for modality.

6 CONCLUSION & FUTURE WORK

In this study, we investigated drivers' interaction for 24 personal and work-related tasks. Future research that intends to generate user-defined interaction sets can benefit from our findings, which provide insights on how the drivers imagine they will be using gestures and voice commands to interact with AR WSD of automated vehicles. In future work, additional scenarios need to be examined where either of the interaction modalities are engaged. Simulator studies should

Table 4: Taxonomy of voice commands for interacting with AR windshield display.

Dimension	Category	Description
Form	Single word	Voice command consists out of a single word
	Two words	Voice command consists out of two words
	More words	Voice command consists out of more words without sentence structure
	Sentence	Voice command uses sentence structure
Nature	Action	Voice command states the action to perform
	State	Voice command describes the desired condition
Context	In-context	Voice command requires specific context
	No-context	Voice command does not require specific context
Complexity	Simple	Voice command consists of a single voice command
	Compound	Voice command can be decomposed into simple voice commands

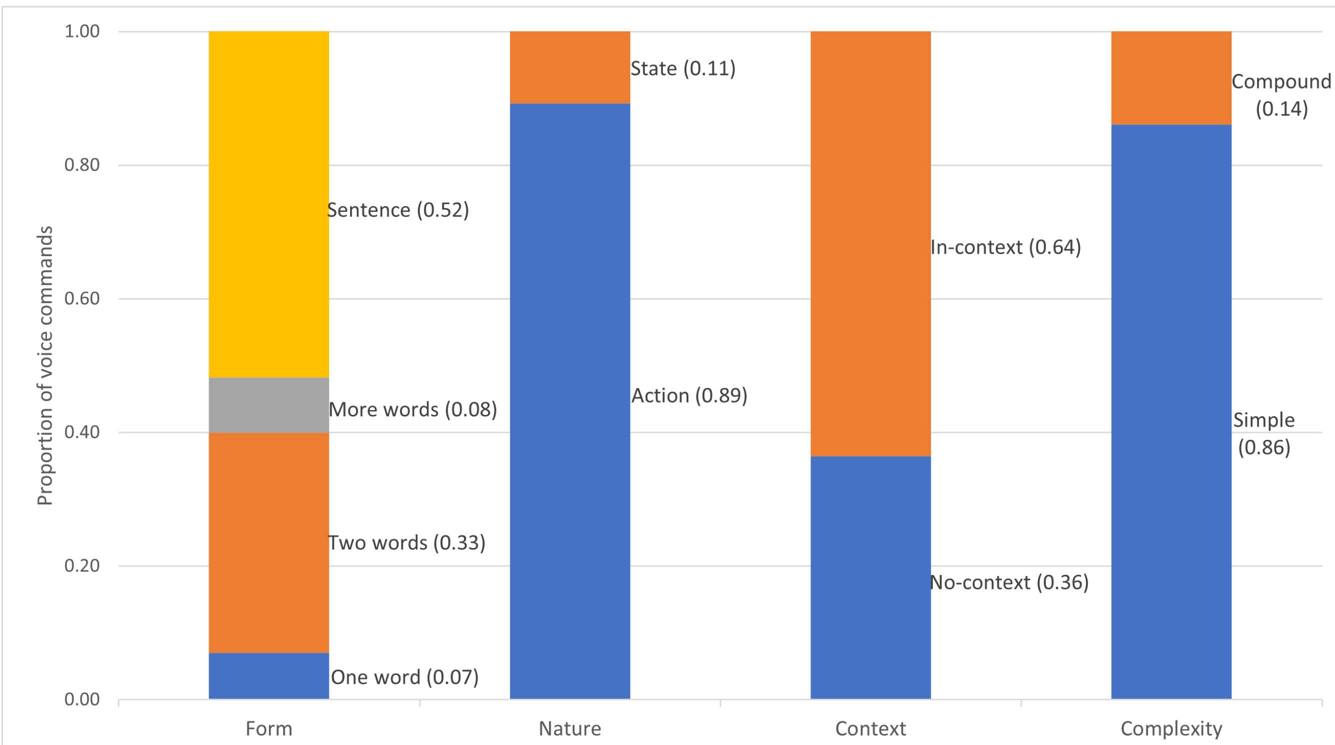

Figure 4: Percentage of voice commands in each category along the four dimensions described in Table 4.

also be used to assess transition to and from personal or work-related tasks to driving and how different interaction modalities will influence this transition.

ACKNOWLEDGMENTS

This work was in part supported by NSF grants CMMI- (University of New Hampshire: 1840085; Wellesley College: 1840031). Any opinions, findings, and conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NSF.

REFERENCES

[1] Abdullah X Ali, Meredith Ringel Morris, and Jacob O Wobbrock. 2019. Crowdlicit: A system for conducting distributed end-user elicitation and identification studies. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–12.

[2] Micah Alpern and Katie Minardo. 2003. Developing a car gesture interface for use as a secondary task. In *CHI'03 extended abstracts on Human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 932–933.

[3] Leonardo Angelini, Jürgen Baumgartner, Francesco Carrino, Stefano Carrino, Maurizio Caon, Omar Khaled, Jürgen Sauer, Denis Lalanne, Elena Mugellini, and Andreas Sonderegger. 2016. Comparing gesture, speech and touch interaction modalities for in-vehicle infotainment systems. In *Actes de la 28ième conférence francophone sur l'Interaction Homme-Machine*. HAL, Lyon, France, 188–196.

[4] Leonardo Angelini, Francesco Carrino, Stefano Carrino, Maurizio Caon, Omar Abou Khaled, Jürgen Baumgartner, Andreas Sonderegger, Denis Lalanne, and Elena Mugellini. 2014. Gesturing on the steering wheel: A user-elicited taxonomy. In *Proceedings of the 6th international conference on automotive user*

interfaces and interactive vehicular applications. Association for Computing Machinery, New York, NY, USA, 1–8.
[5] Lisa Anthony, Quincy Brown, Jaye Nias, Berthel Tate, and Shreya Mohan. 2012. Interaction and recognition challenges in interpreting children's touch and gesture input on mobile devices. In *Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces*. Association for Computing Machinery, New York, NY, USA, 225–234.
[6] J Alejandro Betancur, Nicolás Gómez, Mario Castro, Frederic Merienne, and Daniel Suárez. 2018. User experience comparison among touchless, haptic and voice Head-Up Displays interfaces in automobiles. *International Journal on Interactive Design and Manufacturing (IJIDeM)* 12, 4 (2018), 1469–1479.
[7] Laura-Bianca Bilius and Radu-Daniel Vatavu. 2020. A multistudy investigation of drivers and passengers' gesture and voice input preferences for in-vehicle interactions. *Journal of Intelligent Transportation Systems* 25, 2 (2020), 197–220.
[8] Daniel Brand, Kevin Büchele, and Alexander Meschtscherjakov. 2016. Pointing at the HUD: Gesture interaction using a leap motion. In *Adjunct Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. Association for Computing Machinery, New York, NY, USA, 167–172.
[9] Charlynn Burd, Michael Burrows, and Brian McKenzie. 2021. Travel time to work in the united states: 2019. *American Community Survey Reports, United States Census Bureau* 2 (2021), 2021.
[10] Gary Burnett, Elizabeth Crundall, David Large, Glyn Lawson, and Lee Skrypchuk. 2013. A study of unidirectional swipe gestures on in-vehicle touch screens. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. Association for Computing Machinery, New York, NY, USA, 22–29.
[11] Edwin Chan, Teddy Seyed, Wolfgang Stuerzlinger, Xing-Dong Yang, and Frank Maurer. 2016. User elicitation on single-hand microgestures. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 3403–3414.
[12] Vassilis Charissis and Stylianos Papanastasiou. 2010. Human–machine collaboration through vehicle head up display interface. *Cognition, Technology & Work* 12, 1 (2010), 41–50.
[13] On-Road Automated Driving (ORAD) committee. 2021. Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. https://saemobilus.sae.org/content/j3016_202104. Accessed: 2022-01-5.
[14] Henrik Detjen, Sarah Faltaous, Stefan Geisler, and Stefan Schneegass. 2019. User-defined voice and mid-air gesture commands for maneuver-based interventions in automated vehicles. In *Proceedings of Mensch und Computer 2019*. Association for Computing Machinery, New York, NY, USA, 341–348.
[15] Julien Epps, Serge Lichman, and Mike Wu. 2006. A study of hand shape use in tabletop gesture interaction. In *CHI'06 extended abstracts on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 748–753.
[16] Hessam Jahani Fariman, Hasan J Alyamani, Manolya Kavakli, and Len Hamey. 2016. Designing a user-defined gesture vocabulary for an in-vehicle climate control system. In *Proceedings of the 28th Australian Conference on Computer-Human Interaction*. Association for Computing Machinery, New York, NY, USA, 391–395.
[17] Alexander Feierle, Fabian Schlichtherle, and Klaus Bengler. 2021. Augmented Reality Head-Up Display: A Visual Support During Malfunctions in Partially Automated Driving? *IEEE Transactions on Intelligent Transportation Systems* 23, 5 (2021), 4853–4865.
[18] Leah Findlater, Ben Lee, and Jacob Wobbrock. 2012. Beyond QWERTY: augmenting touch screen keyboards with multi-touch gestures for non-alphanumeric input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 2679–2682.
[19] Kikuo Fujimura, Lijie Xu, Cuong Tran, Rishabh Bhandari, and Victor Ng-Thow-Hing. 2013. Driver queries using wheel-constrained finger pointing and 3-D head-up display visual feedback. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. Association for Computing Machinery, New York, NY, USA, 56–62.
[20] Joseph L Gabbard, Gregory M Fitch, and Hyungil Kim. 2014. Behind the glass: Driver challenges and opportunities for AR automotive applications. *Proc. IEEE* 102, 2 (2014), 124–136.
[21] Michael A Gerber, Ronald Schroeter, Li Xiaomeng, and Mohammed Elhenawy. 2020. Self-interruptions of non-driving related tasks in automated vehicles: Mobile vs head-up display. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–9.
[22] José Ignacio Giménez-Nadal, José Alberto Molina, and Jorge Velilla. 2022. Trends in commuting time of European workers: A cross-country analysis. *Transport Policy* 116 (2022), 327–342.
[23] Yves Guiard. 1987. Asymmetric division of labor in human skilled bimanual action: The kinematic chain as a model. *Journal of motor behavior* 19, 4 (1987), 486–517.
[24] Renate Haeuslschmid, Bastian Pfleging, and Florian Alt. 2016. A design space to support the development of windshield applications for the car. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 5076–5091.
[25] Renate Häuslschmid, Sven Osterwald, Marcus Lang, and Andreas Butz. 2015. Augmenting the driver's view with peripheral information on a windshield display. In *Proceedings of the 20th International Conference on Intelligent User Interfaces*. Association for Computing Machinery, New York, NY, USA, 311–321.
[26] Fabian Hoffmann, Miriam-Ida Tyroller, Felix Wende, and Niels Henze. 2019. User-defined interaction for smart homes: voice, touch, or mid-air gestures?. In *Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia*. Association for Computing Machinery, New York, NY, USA, 1–7.
[27] Hessam Jahani, Hasan J Alyamani, Manolya Kavakli, Arindam Dey, and Mark Billinghurst. 2017. User evaluation of hand gestures for designing an intelligent in-vehicle interface. In *International Conference on Design Science Research in Information System and Technology*. Springer, Springer, Berlin/Heidelberg, Germany, 104–121.
[28] Dagmar Kern and Albrecht Schmidt. 2009. Design space for driver-based automotive user interfaces. In *Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. Association for Computing Machinery, New York, NY, USA, 3–10.
[29] SeungJun Kim and Anind K Dey. 2009. Simulated augmented reality windshield display as a cognitive mapping aid for elder driver navigation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 133–142.
[30] Anne Köpsel and Nikola Bubalo. 2015. Benefiting from legacy bias. *interactions* 22, 5 (2015), 44–47.
[31] Andrew L Kun. 2018. Human-machine interaction for vehicles: Review and outlook. *Foundations and Trends® in Human–Computer Interaction* 11, 4 (2018), 201–293.
[32] Andrew L Kun, Orit Shaer, Andreas Riener, Stephen Brewster, and Clemens Schartmüller. 2019. AutoWork 2019: workshop on the future of work and well-being in automated vehicles. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications: Adjunct Proceedings*. Association for Computing Machinery, New York, NY, USA, 56–62.
[33] Alexander Kunze, Stephen J Summerskill, Russell Marshall, and Ashleigh J Filtness. 2018. Augmented reality displays for communicating uncertainty information in automated driving. In *Proceedings of the 10th international conference on automotive user interfaces and interactive vehicular applications*. Association for Computing Machinery, New York, NY, USA, 164–175.
[34] Byron Lahey, Audrey Girouard, Winslow Burleson, and Roel Vertegaal. 2011. PaperPhone: understanding the use of bend gestures in mobile devices with flexible electronic paper displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1303–1312.
[35] Jae Yeol Lee, Gue Won Rhee, and Dong Woo Seo. 2010. Hand gesture-based tangible interactions for manipulating virtual objects in a mixed reality environment. *The International Journal of Advanced Manufacturing Technology* 51, 9 (2010), 1069–1082.
[36] Sang Hun Lee, Se-One Yoon, and Jae Hoon Shin. 2015. On-wheel finger gesture control for in-vehicle systems on central consoles. In *Adjunct Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. Association for Computing Machinery, New York, NY, USA, 94–99.
[37] Xiaomeng Li, Ronald Schroeter, Andry Rakotonirainy, Jonny Kuo, and Michael G Lenné. 2020. Effects of different non-driving-related-task display modes on drivers' eye-movement patterns during take-over in an automated vehicle. *Transportation research part F: traffic psychology and behaviour* 70 (2020), 135–148.
[38] Xiaomeng Li, Ronald Schroeter, Andry Rakotonirainy, Jonny Kuo, and Michael G Lenné. 2021. Get Ready for Take-Overs: Using Head-Up Display for Drivers to Engage in Non–Driving-Related Tasks in Automated Vehicles. *Human factors* 0, 0 (2021), 00187208211056200.
[39] Hongnan Lin. 2019. Using passenger elicitation for developing gesture design guidelines for adjusting highly automated vehicle dynamics. In *Companion Publication of the 2019 on Designing Interactive Systems Conference 2019 Companion*. Association for Computing Machinery, New York, NY, USA, 97–100.
[40] Patrick Lindemann, Tae-Young Lee, and Gerhard Rigoll. 2018. Catch my drift: Elevating situation awareness for highly automated driving with an explanatory windshield display user interface. *Multimodal Technologies and Interaction* 2, 4 (2018), 71.
[41] Patrick Lindemann, Tae-Young Lee, and Gerhard Rigoll. 2018. Supporting driver situation awareness for autonomous urban driving with an augmented-reality windshield display. In *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE, IEEE, Piscataway, NJ, 358–363.
[42] Lutz Lorenz, Philipp Kerschbaum, and Josef Schumann. 2014. Designing take over scenarios for automated driving: How does augmented reality support the driver to get back into the loop?. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 58. Sage Publications Sage CA: Los Angeles, CA,

Sage Publications, Thousand Oaks, CA, 1681–1685.

[43] Keenan R May, Thomas M Gable, and Bruce N Walker. 2017. Designing an in-vehicle air gesture set using elicitation methods. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. Association for Computing Machinery, New York, NY, USA, 74–83.

[44] Christophe Mignot, Claude Valot, and Noelle Carbonell. 1993. An experimental study of future "natural" multimodal human-computer interaction. In *INTERACT'93 and CHI'93 Conference Companion on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 67–68.

[45] W Thomas Miller and Andrew L Kun. 2013. Using speech, GUIs and buttons in police vehicles: field data on user preferences for the Project54 system. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. Association for Computing Machinery, New York, NY, USA, 108–113.

[46] Meredith Ringel Morris. 2012. Web on the wall: insights from a multimodal interaction elicitation study. In *Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces*. Association for Computing Machinery, New York, NY, USA, 95–104.

[47] Meredith Ringel Morris, Andreea Danielescu, Steven Drucker, Danyel Fisher, Bongshin Lee, MC Schraefel, and Jacob O Wobbrock. 2014. Reducing legacy bias in gesture elicitation studies. *interactions* 21, 3 (2014), 40–45.

[48] Meredith Ringel Morris, Jacob O Wobbrock, and Andrew D Wilson. 2010. Understanding users' preferences for surface gestures. In *Proceedings of graphics interface 2010*. Canadian Information Processing Society, 403 King Street West, Suite 205 Toronto, Ont. M5U 1LSCanada, 261–268.

[49] Divyabharathi Nagaraju, Alberta Ansah, Nabil Al Nahin Ch, Caitlin Mills, Christian P Janssen, Orit Shaer, and Andrew L Kun. 2021. How Will Drivers Take Back Control in Automated Vehicles? A Driving Simulator Test of an Interleaving Framework. In *13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. Association for Computing Machinery, New York, NY, USA, 20–27.

[50] Vijayakumar Nanjappan, Rongkai Shi, Hai-Ning Liang, Kim King-Tong Lau, Yong Yue, and Katie Atkinson. 2019. Towards a taxonomy for in-vehicle interactions using wearable smart textiles: insights from a user-elicitation study. *Multimodal Technologies and Interaction* 3, 2 (2019), 33.

[51] Brian Normile and Jane Ulitskaya. 2021. *Which Cars Have Head-Up Displays?* cars. https://www.cars.com/articles/which-cars-have-head-up-displays-434824/

[52] Stefan Palan and Christian Schitter. 2018. Prolific. ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance* 17 (2018), 22–27.

[53] Ekaterina Peshkova, Martin Hitz, and David Ahlström. 2016. Exploring user-defined gestures and voice commands to control an unmanned aerial vehicle. In *International Conference on Intelligent Technologies for Interactive Entertainment*. Springer, Springer, Berlin/Heidelberg, Germany, 47–62.

[54] L Petersen, L Robert, X Yang, and D Tilbury. 2019. Situational Awareness, Driver's Trust in Automated Driving Systems and Secondary Task Performance. *SAE Int. J. of CAV* 2, 2 (2019), 0.

[55] Bastian Pfleging, Maurice Rang, and Nora Broy. 2016. Investigating user needs for non-driving-related activities during automated driving. In *Proceedings of the 15th international conference on mobile and ubiquitous multimedia*. Association for Computing Machinery, New York, NY, USA, 91–99.

[56] Thammathip Piumsomboon, Adrian Clark, Mark Billinghurst, and Andy Cockburn. 2013. User-defined gestures for augmented reality. In *IFIP Conference on Human-Computer Interaction*. Springer, Springer, Berlin/Heidelberg, Germany, 282–299.

[57] Jonas Radlmayr, Christian Gold, Lutz Lorenz, Mehdi Farid, and Klaus Bengler. 2014. How traffic situations and non-driving related tasks affect the take-over quality in highly automated driving. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 58. Sage Publications Sage CA: Los Angeles, CA, Sage Publications, Thousand Oaks, CA, 2063–2067.

[58] Andreas Riegler, Bilal Aksoy, Andreas Riener, and Clemens Holzmann. 2020. Gaze-based Interaction with Windshield Displays for Automated Driving: Impact of Dwell Time and Feedback Design on Task Performance and Subjective Workload. In *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. Association for Computing Machinery, New York, NY, USA, 151–160.

[59] Andreas Riegler, Philipp Wintersberger, Andreas Riener, and Clemens Holzmann. 2019. Augmented Reality Windshield Displays and Their Potential to Enhance User Experience in Automated Driving. *i-com* 18, 2 (2019), 127–149.

[60] Sandrine Robbe. 1998. An empirical study of speech and gesture interaction: Toward the definition of ergonomic design guidelines. In *CHI 98 Conference Summary on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 349–350.

[61] Sandrine Robbe-Reiter, Noëlle Carbonell, and Pierre Dauchy. 2000. Expression constraints in multimodal human-computer interaction. In *Proceedings of the 5th international conference on Intelligent user interfaces*. Association for Computing Machinery, New York, NY, USA, 225–228.

[62] Jaime Ruiz, Yang Li, and Edward Lank. 2011. User-defined motion gestures for mobile interaction. In *Proceedings of the SIGCHI conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 197–206.

[63] Clemens Schartmüller, Andreas Riener, Philipp Wintersberger, and Anna-Katharina Frison. 2018. Workaholistic: on balancing typing-and handover-performance in automated driving. In *Proceedings of the 20th international conference on human-computer interaction with mobile devices and services*. Association for Computing Machinery, New York, NY, USA, 1–12.

[64] Albrecht Schmidt, Anind K Dey, Andrew L Kun, and Wolfgang Spiessl. 2010. Automotive user interfaces: human computer interaction in the car. In *CHI'10 Extended Abstracts on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 3177–3180.

[65] Orit Shaer and Eva Hornecker. 2010. *Tangible user interfaces: past, present, and future directions*. Now Publishers Inc, Hanover, MA 02339, USA.

[66] Yasuhiro Takaki, Yohei Urano, Shinji Kashiwada, Hiroshi Ando, and Koji Nakamura. 2011. Super multi-view windshield display for long-distance image information presentation. *Optics express* 19, 2 (2011), 704–716.

[67] Thomaz Teodorovicz, Andrew L Kun, Raffaella Sadun, and Orit Shaer. 2022. Multitasking while driving: A time use study of commuting knowledge workers to assess current and future uses. *International Journal of Human-Computer Studies* 162 (2022), 102789.

[68] Kathryn G Tippey, Elayaraj Sivaraj, and Thomas K Ferris. 2017. Driving while interacting with Google Glass: Investigating the combined effect of head-up display and hands-free input on driving safety and multitask performance. *Human factors* 59, 4 (2017), 671–688.

[69] Consuelo Valdes, Diana Eastman, Casey Grote, Shantanu Thatte, Orit Shaer, Ali Mazalek, Brygg Ullmer, and Miriam K Konkel. 2014. Exploring the design space of gestural interaction with active tokens through user-defined gestures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 4107–4116.

[70] Emma van Amersfoorth, Lotte Roefs, Quinta Bonekamp, Laurent Schuermans, and Bastian Pfleging. 2019. Increasing driver awareness through translucency on windshield displays. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications: Adjunct Proceedings*. Association for Computing Machinery, New York, NY, USA, 156–160.

[71] Radu-Daniel Vatavu and Jacob O Wobbrock. 2015. Formalizing agreement analysis for elicitation studies: new measures, significance test, and toolkit. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1325–1334.

[72] Sarah Theres Völkel, Daniel Buschek, Malin Eiband, Benjamin R Cowan, and Heinrich Hussmann. 2021. Eliciting and Analysing Users' Envisioned Dialogues with Perfect Voice Assistants. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–15.

[73] Ying Wang, Shengfan He, Zuerhumuer Mohedan, Yueyan Zhu, Lijun Jiang, and Zhelin Li. 2014. Design and evaluation of a steering wheel-mount speech interface for drivers' mobile use in car. In *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*. IEEE, IEEE, Piscataway, NJ, 673–678.

[74] Florian Weidner and Wolfgang Broll. 2019. Interact with your car: a user-elicited gesture set to inform future in-car user interfaces. In *Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia*. Association for Computing Machinery, New York, NY, USA, 1–12.

[75] Garrett Weinberg, Bret Harsham, and Zeljko Medenica. 2011. Evaluating the usability of a head-up display for selection from choice lists in cars. In *Proceedings of the 3rd International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. Association for Computing Machinery, New York, NY, USA, 39–46.

[76] Philipp Wintersberger, Tamara von Sawitzky, Anna-Katharina Frison, and Andreas Riener. 2017. Traffic augmentation as a means to increase trust in automated driving systems. In *Proceedings of the 12th biannual conference on italian sigchi chapter*. Association for Computing Machinery, New York, NY, USA, 1–7.

[77] Jacob O Wobbrock, Htet Htet Aung, Brandon Rothrock, and Brad A Myers. 2005. Maximizing the guessability of symbolic input. In *CHI'05 extended abstracts on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1869–1872.

[78] Jacob O Wobbrock, Meredith Ringel Morris, and Andrew D Wilson. 2009. User-defined gestures for surface computing. In *Proceedings of the SIGCHI conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 1083–1092.

[79] Huiyue Wu, Yu Wang, Jiayi Liu, Jiali Qiu, and Xiaolong Luke Zhang. 2020. User-defined gesture interaction for in-vehicle information systems. *Multimedia Tools and Applications* 79, 1 (2020), 263–288.

[80] Huiyue Wu, Yu Wang, Jiali Qiu, Jiayi Liu, and Xiaolong Zhang. 2019. User-defined gesture interaction for immersive VR shopping applications. *Behaviour & Information Technology* 38, 7 (2019), 726–741.