This is the authors' version of an article published by Taylor & Francis in the Journal of Intelligent Transportation Systems 25(2), available at https://doi.org/10.1080/15472450.2020.1846127

A Multi-Study Investigation of Drivers and Passengers' Gesture and Voice Input Preferences for In-Vehicle Interactions

Laura-Bianca Bilius and Radu-Daniel Vatavu

MintViz Lab, MANSiD Research Center, Ştefan cel Mare University of Suceava, Suceava 720229, Romania

ARTICLE HISTORY
Compiled August 6, 2020

ABSTRACT
We conduct an examination of the preferences of drivers and passengers alike for in-vehicle interactions with a multi-study approach consisting of (1) a targeted literature survey of applications and user interfaces designed to support interactions with in-vehicle controls and systems based on gesture and voice input; (2) a large-scale survey (N=160 participants) to understand drivers and passengers' preferences for driving and traveling by car; and (3) an end-user elicitation study (N=40 drivers and passengers) to collect preferences for gesture and voice input commands for in-vehicle interaction. We analyze and discuss the gesture and voice commands proposed by our participants and describe their characteristics, such as production times for gesture input and the vocabulary size of voice commands.

KEYWORDS
Gesture input; Voice input; Connected cars; Smart cars; Study.

1. INTRODUCTION

The automotive industry has been innovating in smart technology integrated in connected cars toward improving driving safety, but also in terms of new systems, controls, and features to deliver a more comfortable and enjoyable car travel experience to drivers and passengers alike (Holtl & Trommer, 2013; Lin, 2019; Ohn-Bar & Trivedi, 2014; Varhelyi, Kaufmann, Johnsson, & Almqvist, 2020). One important aspect regarding driving safety relates to drivers' distraction while interacting with in-vehicle systems, such as the infotainment system, or with their own mobile devices, such as smartphones and smartwatches and, in the near future, possibly smartglasses (Aiordăchioae, Vatavu, & Popovici, 2019) and finger augmentation devices (Gheran & Vatavu, 2020). For instance, drivers can get easily distracted when operating the user interfaces (UIs) of such systems and applications, which more often than not means taking their eyes off the road and hands off the steering wheel. Therefore, designing interactions with smart technology inside the vehicle, including interactions based on natural input, such as gesture and voice commands, is likely to have positive consequences on driving safety in the context of connected vehicles (Shladover, 2018).

The problem of driver distraction has been repeatedly called out as an important threat to road safety. The average time during which drivers take their eyes off the

Contact Author. Email: radu.vatavu@usm.ro

road in order to focus on targets from inside the vehicle has been quantified to around 1.6 s, while *"the average time of the longest glances at a mobile phone while texting is longer than 2 seconds"* (European Road Safety Observatory, 2015) (p. 10). Moreover, according to the 2019 Road Safety Facts Report of the Association for Safe International Road Travel, *"road traffic injuries are predicted to become the fifth leading cause of death by 2030"* and *"nearly 1.25 millions people die in road crashes each year, on average 3,287 deaths a day"* (ASIRT, 2019). The number of annual road traffic deaths reported by WHO (2018) in December 2018, 1.35 million, is close to this figure. NHTSA examined the use of mobile devices while driving and reported that 3,166 people were killed because of distracted driving in 2017 alone, representing 8.5 percent of the total fatalities for that year (NHTSA, 2019). In this context, several studies have shown that the use of hands-free phones, speed dialing, and voice input inside the vehicle reduce driver distractions since there is less need for eye glances away from the road and for hand movements away from the steering wheel compared to other input modalities (Fitch, Bartholomew, Hanowski, & Perez, 2015; Jeanne Breen Consulting, 2009). For instance, experimental results reported by Fitch et al. (2015) from a sample of 204 drivers, 1,564 cell phone calls, and 844 text messages sent while driving showed that the mean values for the total amount of time that their participants' visual attention was away from the road during the span of each subtask was 23.3 s for text messaging, 8.2 s for browsing content on the phone, and 7.8 s for dialing, while hands-free tasks were executed in a substantially smaller time span of just 2.7 s. In situations of information conflict with multiple in-vehicle systems competing for the drivers' attention, Lansdown, Brook-Carter, and Kersloot (2002) found that drivers experience equal task disruption regardless of the sensory channel of the secondary task.

While cars become more complex toward smart, connected, and autonomous vehicles, but also toward integration with the Internet-of-Things (Igorevich & Min, 2015; Taherkhani, Kawaguchi, Shirmohammad, & Sato, 2016) and Augmented Reality (Langlois, That, & Mermillod, 2016; Qiu et al., 2017; Schipor & Vatavu, 2019), enabling drivers to connect their smart devices, such as smartphones and smart wearables, to the in-vehicle infotainment system, the range of safety-related concerns and potential distractions for drivers is likely to increase as the number of devices and systems that need to be operated inside the vehicle grows as well. For example, Villing (2011) studied interactions between drivers and in-vehicle systems via voice, graphical user interfaces (GUIs), and multimodal input, and reported that speech exhibited the poorest task performance while GUIs were the fastest modality but required more attention from drivers to operate accurately. Regarding gesture input, Gable, May, and Walker (2014) examined usability heuristics for gesture-based interaction inside the vehicle, and concluded that more investigations were needed to produce substantial guidelines and standards for practitioners at that time. Angelini et al. (2016) conducted a comparative analysis of gesture, voice, and touch input for in-vehicle interactions with the infotainment system, and reported that speech and touch-based input presented several advantages over gesture input performed on the steering wheel.

In this context, there is a critical practical need for designing in-vehicle interactions based on input modalities that are safe for drivers to operate effectively. Thus, safe operation becomes an important requirement for gesture, touch, and voice input inside the vehicle next to intuitiveness, efficiency, and learnability (Gable et al., 2014; May, Gable, & Walker, 2017; Villing, 2011; Wu, Wang, Liu, Qiu, & Zhang, 2019). In this regard, several studies have focused on improving gesture recognition accuracy (Kim & Choi, 2019; Mahetalia, Mehta, Varudandi, & Samdani, 2018; Tarvekar, 2018; Zobl, Geiger, Schuller, Lang, & Rigoll, 2003) or on improving speech recognition techniques (Chen, Guido, & Chen, 2007; Chung & Chen, 2001; Khan, Akmal, Ali, & Naeem, 2017), while other work has focused on understanding users' preferences for and performance with gesture and voice commands inside the vehicle (Akyol, Canzler, Bengler, & Hahn, 2000; Gable et al., 2014; Khare et al., 2009; May et al., 2017; Wu et al., 2019). Of these, a few studies have focused on passengers performing in-vehicle interactions (Lin, 2019; Ohn-Bar & Trivedi, 2014). Although this large body of scientific literature has reported useful results and created knowledge for designing gesture and voice-based interactions inside the vehicle, this knowledge is dispersed and in need of confirmation and consolidation (Bilius & Vatavu, 2020). Moreover, a comprehensive, unitary treatment of gesture and voice input, drivers and passengers, and various research methods adopted for studying in-vehicle interactions is needed.

Our approach in this work is a multi-study investigation method in order to collect a diversity of data regarding drivers and passengers in relation to their preferences and needs for in-vehicle interactions, behavior, and use of smart technology while driving or traveling by car. Our contributions are as follows:

(1) We perform a targeted literature survey (Huelin, Iheanacho, Payne, & Sandman, 2016) in which we classify prior work on in-vehicle interactions according to input modalities, type of study, type of participants (drivers and/or passengers), sample of participants, and focus on implications for driving safety.

(2) We conduct an examination of drivers' preferences, needs, and desires for controlling various car functions and in-vehicle systems, and we relate the results obtained on a representative sample of N=80 drivers to their self-reported driving experience and driving behavior. We also address car passengers in our study, a category of car travelers that has been largely overlooked by prior work. (From the total number of 64 references identified and discussed in our targeted literature review, just 6 references (9.4%) addressed passengers.) We document and report passengers' preferences for interactive gesture and voice commands to control in-vehicle systems while traveling by car. To this end, we report results from a sample of N=80 passengers (non-drivers), which we compare to the results obtained on the drivers group.

(3) We conduct an end-user elicitation study with an additional number of N=20 drivers and N=20 passengers (non-drivers) to understand what kind of gesture and voice commands they would like to use while driving or traveling by car. We discuss the specific gesture and voice commands elicited from our participants, perform agreement analysis and report agreement rates to describe the consensus reached by drivers and passengers regarding their preferred gesture and voice input commands, and analyze the characteristics of gesture and voice input, such as gesture production times and the vocabulary size of voice commands.

2. TARGETED LITERATURE REVIEW

We overview in this section prior work that examined user preferences for gesture and/or voice commands for in-vehicle interactions as well as studies that described prototypes of interactive systems and/or conducted evaluations of drivers' task performance with gesture and voice input inside the vehicle. To identify relevant references from the scientific literature regarding these topics, we ran the following query:

```
(gesture OR voice) AND (car or vehicle)
```

on the titles of papers available from the ACM Digital Library and the IEEE Xplore database. We obtained a number of 53 references from ACM DL and 79 from IEEE Xplore, from which we selected only those articles that presented (i) an actual system that users (drivers or passengers) could control via voice and/or gesture input or (ii) the results of a user study involving gesture and/or voice input performed inside the vehicle. We excluded papers that exclusively addressed other topics, such as feedback and output modalities for interactions inside the vehicle; papers that focused solely on recognition algorithms and techniques for voice and/or gesture input without any discussion of underlying interaction aspects; and papers that addressed toy or robotic vehicles. In the end, we compiled a total number of 36+16=52 papers relevant for the purpose of our investigation. Next to these, we included other 12 papers that we identified by means of references and citations of the initial set of papers. Overall, we analyzed a number of 64 papers as part of our targeted literature review (see Tables 1 and 2), which we classified according to the following criteria:

(1) *Input modality* with three categories: *gesture*, *voice*, and *multimodal* (gesture and voice) input. Some studies focused on gesture input exclusively, others on voice input, while others addressed multimodal interactions. We did not distinguish subcategories of these input modalities, *e.g.*, touch gestures *vs.* mid-air gestures *vs.* stroke gestures performed on the steering wheel, nor between voice commands and natural speech interactions, which would make our analysis more complex than needed. However, we used the last column of Tables 1 and 2 to provide such details and a brief description of each study for the interested reader.

(2) *Type of end users* with two categories: *drivers* and *passengers*. Our analysis showed that the majority of references focused on drivers, while only six papers identified by our queries addressed passengers; see Tables 1 and 2.

(3) *Type of study* with three categories: *end-user elicitation*, *system evaluation*, and *questionnaire administration*.

(4) *Sample size* represents the number of participants (either drivers or passengers) employed in the user studies reported by prior work. For all the papers that reported such studies, the sample size varied between 1 and 617 participants (M=38.4, SD=83.4), and the majority of the studies involved less than 20 participants (51.8%) and less than 30 participants (75.0%), respectively.

(5) *Focus on driving safety.* We highlighted references that specifically addressed driving safety by discussing such aspects as part of their motivation, results, implications, or conclusion sections. For example, Angelini et al. (2014) reported that 41% of their study participants were concerned about safety regarding the use of gestures on the steering wheel, and Ma and Du (2017) discussed usability of gesture input that should be "*easy to learn and remember, but also intuitive, safe, effective, comfortable and natural to use.*"

Tables 1 and 2 list the 64 references identified in our targeted literature review, arranged chronologically and classified according to the criteria above.

Gesture input for in-vehicle interactions, especially in the form of eyes-free mid-air gestures, has been examined to support safe operation of car controls and in-vehicle systems while driving. For example, Alba-Castro, González Agulla, and Loira (2014) investigated drivers' preferences for free-hand gestures and touch input and showed that touchscreens were easier to operate, but free-hand gestures presented convenient characteristics that made them safer to use. Wu, Gable, May, Choi, and Walker (2016) focused on mid-air and touch input and concluded that mid-air gestures were recommendable on the premises that future sensing technology would provide better support for such gestures and, until then, touchscreen input was recommendable. Another study conducted by Häuslschmid, Menrad, and Butz (2015) examined drivers' use of free-hand and micro gestures for in-vehicle interactions with results revealing good user experience. Yet another study, conducted by May, Gable, and Walker (2014), evaluated a mid-air gesture interface against direct touch input and found that many participants preferred touch, although mid-air gestures were also favorably perceived. Laack, Kirsch, Tuzar, and Blessing (2016) evaluated drivers' perceptions of in-vehicle gesture input and found that free-hand gestures were more preferred compared to using conventional in-vehicle controls. Several studies (Lin, 2019; May et al., 2017; Vieira et al., 2016; Wu et al., 2019) have implemented the end-user elicitation method (Wobbrock, Aung, Rothrock, & Myers, 2005; Wobbrock, Morris, & Wilson, 2009) to collect drivers' preferences for gestures they would like to use to control various car functions and in-vehicle systems. For example, May et al. (2017) proposed mid-air gestures for interacting with menus and recommended guidelines for designers to promote driving safety and ease of use for gesture input inside the vehicle. Vieira et al. (2016) reported that the participants of their gesture elicitation study preferred simple gestures resembling those implemented on smartphones with touchscreens. Lin (2019) focused on passengers and compiled a set of gestures for adjusting Highly Automated Vehicle dynamics.

Next to gesture commands, voice input represents a common modality to support the design and engineering of automotive UIs that implement natural interactions inside the vehicle, enabling drivers to control various car functions without taking their eyes off the road. Barón and Green (2006) studied usability and safety aspects of speech-based interfaces for vehicles and concluded that drivers generally perform better and spend less time looking off the road when using voice input control. Gellatly and Dingus (1998) investigated the effects of automatic speech recognition inside the vehicle and reported that voice input can improve the safety and usability of in-vehicle interactions. In another study (Sokol, Chen, & Donmez, 2017), drivers' responses to a voice-controlled system in the presence of background noise were analyzed to evaluate technical performance. Alvarez et al. (2010) proposed a prototype for a conversational voice interface to reduce driver distraction and increase overall driving satisfaction. Prior work has also proposed voice-based systems and applications for drivers to control various car functions (Fissore, Laface, & Ruscitti, 1992; Jeong & Liu, 2018; Khan et al., 2017). A shortcoming of voice recognition is represented by the capacity of the system to understand input in the presence of background noise, various accents, and speed of speech (Petkar, 2016). The effect of voice input on driving was studied by Roider, Reisig, and Gross (2018), Reimer et al. (2014), Zheng, McDonald, and Pickering (2008), and Kujala and Grahn (2017) that reported few glances away from the road, intuitiveness of voice input commands, little distraction for drivers, low overall task completion times, and better user experience compared to other input modalities. Kim, Jeong, and Lee (2019) investigated the effects of error recovery strategies for voice input systems and recommended that users should repeat the input command and systems should provide input examples to users.

Table 1. An overview of previous work identified in our targeted literature survey for using gesture and voice input inside the vehicle (continues on the next page).

Reference	Input modality		Study†	End users		N‡	Safety*	Description
	Gesture	Voice		Drivers	Pass.			
Akyol et al. (2000)	✔		♦	✔		n/a		Audio message playback controlled with gestures.
Ohn-Bar and Trivedi (2014)	✔		♣	✔	✔	8		Gesture control for smartphone operation, music, photos, and navigation.
Sokol et al. (2017)		✔	♥	✔		20		Voice input for music control, navigation, phone contacts.
Khare et al. (2009)	✔	✔	♦	✔		n/a	✔	Gesture and voice input control of various in-vehicle systems.
Pfleging, Schneegass, and Schmidt (2012)	✔	✔	♣♥	✔		12		Gesture and voice input control of car comfort settings, assistance, and infotainment.
Alvarez et al. (2010)		✔	♦	✔		n/a		Voice input for driver-car dialogue systems.
Wu et al. (2019)	✔		♣♥	✔		16	✔	User-defined gesture input for in-vehicle information systems.
Wu et al. (2016)	✔		♥♦	✔		14		Four gesture-based interaction techniques for menu navigation.
Vieira et al. (2016)	✔		♥♣	✔		13		User-defined gestures for touch input inside the vehicle.
Angelini et al. (2016)	✔	✔	♥♣	✔		60	✔	Gesture, speech, and touch input for in-vehicle infotainment systems.
Graichen, Graichen, and Krems (2019)	✔		♥♦	✔		109	✔	Touch and gesture input with focus on user experience and reducing driver's distraction.
May et al. (2017)	✔		♣	✔		161	✔	User-defined free-hand gestures and online survey of preferences.
Gable et al. (2014)	✔			✔		n/a	✔	Usability heuristics applied to mid-air gesture input.
May et al. (2014)	✔		♦	✔		26		Mid-air gestures for menu navigation.
Lin (2019)	✔		♥♣		✔	12		User-defined gestures for automated vehicle dynamics addressing passengers.
Alba-Castro et al. (2014)	✔		♦♥	✔	✔	23		Hand gestures for controlling the infotainment system.
Duffield and Krupenia (2015)	✔	✔	♦♥	✔		15		Driver interaction preferences for a multi-interactive, multi-display truck.
Gellatly and Dingus (1998)		✔	♣	✔		12	✔	Effects of speech recognition, input modality, and driver age on driving performance.
Häuslschmid et al. (2015)	✔		♦♥	✔		24		Free-hand and micro gestures study to investigate impact on driving performance.
Laack et al. (2016)	✔		♥	✔		45		System for gesture recognition inside the vehicle.
Sauras, Gil, Gill, Pisu, and Taiber (2017)	✔	✔	♦	✔		n/a		Human-vehicle interaction system based on voice input and deictic gestures.
Villing (2011)		✔	♦♥	✔		10		In-vehicle dialogue with voice, gesture, and multimodal interaction.
Riener et al. (2013)	✔		♣♥	✔		12	✔	Spatial extent of gesture input, time performance of gestures articulated inside the vehicle.
Detjen et al. (2019)	✔	✔	♣♥	✔		20		Mid-air gesture and voice control for automated vehicles.
Sterkenburg, Landry, and Jeon (2017)	✔		♦♥	✔		24	✔	Mid-air gesture input.
Rahman, Saboune, and El Saddik (2011)	✔		♦	✔		8		Hand gesture control for intuitive interactions inside the vehicle.
Weidner and Broll (2019)	✔		♣♥	✔		23		User-defined gestures for the in-vehicle environment.
Ecker, Broy, Hertzschuch, and Butz (2010)	✔		♦♥	✔		28		Touch gestures and pie menus.
Lee, Yoon, and Shin (2015)	✔		♦♥	✔		15		Gestures performed on the steering wheel.
Sterkenburg et al. (2016)	✔		♦	✔		n/a		Mid-air gesture input.
Ohn-Bar, Tran, and Trivedi (2012)	✔		♦	✔	✔	8		Hand gestures for the in-vehicle infotainment system.
Roider et al. (2018)		✔	♦♥	✔		25	✔	Study on the efficiency of voice input for in-vehicle environments.
Roider, Rümelin, Pfleging, and Gross (2017)	✔	✔	♦♥	✔		29		Study on the performance of natural input in various conditions.
Fariman et al. (2016)	✔		♣	✔		22		User-defined gesture vocabulary for automotive climate control.
May et al. (2016)	✔		♦	✔		25		Comparison between three mid-air gestures for menu selection tasks.
Angelini et al. (2014)	✔		♣	✔		40	✔	User-defined gestures performed on the steering wheel.

†The type of study can be participatory design in the form of end-user elicitation of gesture or voice commands (♣), pre-designed commands (♦), or questionnaire administration (♥). ‡Number of participants. *Safety was explicitly indicated to the study participants.

Table 2. An overview of previous work identified in our targeted literature survey for using gesture and voice input inside the vehicle (continued from the previous page).

Reference	Input modality		Study[†]	End users		N[‡]	Safety[*]	Description
	Gesture	Voice		Drivers	Pass.			
Brand et al. (2016)	✔		♣♥	✔		10		Mid-air gestures to control the content displayed by a Head-Up Display.
Roider and Gross (2018)	✔		♣	✔		91		Deictic gestures while driving.
Angelini et al. (2013)	✔		♦	✔		9		Micro gestures performed on the steering wheel.
Jeong and Liu (2018)		✔	♦	✔		35		Model for task completion time and workload of remote-manual and voice controls.
Hessam et al. (2017)	✔		♣	✔		79		Gesture vocabulary for in-vehicle interactions.
Mahr, Endres, Müller, and Schneeberger (2011)	✔		♣♥	✔		24		User preference for in-vehicle micro-gestures.
Reimer et al. (2014)		✔	♦	✔		64	✔	The effects of in-vehicle voice input.
Burnett et al. (2013)	✔		♦	✔		20		Swipe gestures on touchscreens.
Ecker, Broy, Butz, and De Luca (2009)	✔		♦	✔		16		Touch gestures for in-vehicle interactions.
Roider and Raab (2018)	✔		♦	✔		21		Visual feedback (light) for mid-air gesture input.
Neßelrath, Moniri, and Feld (2016)	✔	✔	♦	✔		n/a		A prototype that allows a multimodal (speech, gaze and gestures) interaction with cars.
Zobl et al. (2003)	✔		♦	✔		1		Hand gesture recognition for in-vehicle interactions.
Sauras-Perez, Taiber, and Smith (2014)	✔		♣♥	✔		15		Gestures to control functions.
Ma and Du (2017)	✔		♣	✔		34	✔	Usability of gesture input, proposal of a gesture set.
Smith, Csech, Murdoch, and Shaker (2018)	✔		♦	✔	✔	5		Gesture recognition technique using a mm-wave radar sensor.
Khan et al. (2017)		✔	♦	✔		n/a		In-vehicle car control, application in MATLAB.
Fissore et al. (1992)		✔	♦	✔		617		Voice dialing for the in-vehicle environment.
Manawadu, Kamezaki, Ishikawa, Kawano, and Sugano (2016)	✔		♦♥	✔		20	✔	Hand gesture interface to control autonomous vehicles.
Zheng et al. (2008)		✔	♦	✔		12	✔	Study on the impact of voice interfaces for the in-vehicle environment.
Sachara, Kopinski, Gepperth, and Handmann (2017)	✔		♦	✔		20		Free-hand gesture recognition for infotainment systems.
Verma and Choudhary (2018)	✔		♦	✔		5		Framework for dynamic hand gesture recognition to control the in-vehicle tasks.
Raja, Ghaderi, and Sigg (2018)	✔		♦♥	✔	✔	40		Behavior recognition and gesture input for personal assistants.
Alpern and Minardo (2003)	✔		♦	✔		8		Study on using gestures for secondary tasks while driving.
Kim et al. (2019)		✔	♦♥	✔		47		Effect of recognition errors when interacting with voice input.
Ahmad et al. (2018)	✔		♦♥	✔		24		Mid-air pointing selection for automotive displays.
Kujala and Grahn (2017)		✔	♦♥	✔		41	✔	Comparison between touchscreen keyboard, handwriting, and voice recognition.
Shakeri, Williamson, and Brewster (2018)	✔		♦♥	✔		17	✔	Unimodal and multimodal ultrasound feedback for mid-air gesture input.
Shakeri, Williamson, and Brewster (2017)	✔		♦♥	✔		17	✔	Effects of various feedback modalities on mid-air gesture input.
This study	✔	✔	♣♥	✔	✔	200	✔	Fills a gap in the scientific literature by reporting on drivers' and passengers' preferences alike for both gesture and voice input inside the car and by using both participatory design and questionnaire administration research methods.

[†]The type of study can be participatory design in the form of end-user elicitation of gesture or voice commands (♣), pre-designed commands (♦), or questionnaire administration (♥). [‡]Number of participants. [*]Safety was explicitly indicated to the study participants.

3. BEHAVIOR AND PREFERENCES FOR DRIVING, TRAVELING BY CAR, AND CONTROLLING CAR FUNCTIONS

We conducted a study to collect and understand the general preferences of drivers and passengers regarding their use of smart devices, including devices regularly used in the car, their perceptions of and preferences for gesture and voice input to control various car functions, and their driving behavior and car traveling experience. Our research method consisted of online questionnaire administration.

3.1. Participants

A total number of N=160 volunteers (N_d=80 drivers[1] and N_p=80 non-drivers,[2] referred to in the following as "passengers") participated in our study and answered general demographic questions regarding their age, gender, current occupation, use of technology, including smart devices, driving history and driving behavior, perceived utility of gesture and voice input, and their preferences regarding car functions they would like to control using gesture and voice input commands. We recruited drivers and passengers via email lists, Facebook and professional social networks, and from the student, research, and staff body of our university. We divided the sample of 160 drivers and passengers into four age groups of equal size; see Tables 3 to 6.

3.1.1. Drivers

Tables 3 and 4 present demographic details of the drivers group (N_d=80 participants). Their occupations included, in alphabetical order: beauty professional, business owner, caretaker, economist, doctor, lawyer, mechanical engineer, military personnel, professor, secretary, software engineer, student, and one participant was unemployed (retired). Overall, the wide range of age and diversity of driving frequency, context, and experience of the 80 participants from our study reflect a representative sample of drivers.

Table 3. Demographic details of the sample of drivers (N_d=80 participants) and their self-reported driving behavior; continues with Table 4.

	Age group	N	Mean age (SD) [yrs.]	Gender (F/M)	Driving frequency				Who else is in the car?	
					Every day	Almost every day	Once a week	Occa-sionally	Driver alone	Family, friends, co-workers
1	<20 yrs.	20	18.9 (0.3)	25%, 75%	15%	30%	30%	25%	75%	60%, 50%, 5%
2	[20,35)	20	24.9 (4.2)	35%, 65%	35%	25%	20%	20%	75%	50%, 40%, 0%
3	[35,50)	20	40.3 (3.6)	40%, 60%	65%	30%	5%	0%	65%	100%, 25%, 15%
4	≥50 yrs.	20	53.2 (2.9)	35%, 65%	60%	15%	20%	5%	90%	50%, 20%, 20%
	Overall	**80**	**34.3 yrs.**	**34%, 66%**	**43.8%**	**25%**	**18.8%**	**12.5%**	**76.3%**	**65%, 34%, 10%**

[1]The subscript letter "d" from the notation N_d refers to drivers.
[2]The subscript letter "p" from the notation N_p refers to passengers.

Table 4. Demographic details of the sample of drivers (N_d=80 participants), their self-reported driving behavior and experience; continued from Table 3.

	Age group	N	Driving context							Driving experience	
			Work duties	Commuting	Driving others	Private trips	Shopping	Visiting family	Holiday travel	Mileage [thousands km] (Mean, SD)	Years (Mean, SD)
1	<20 yrs.	20	30%	70%	45%	55%	80%	80%	55%	3.8 (4.4)	0.95 (0.15)
2	[20,35)	20	25%	60%	25%	65%	70%	85%	80%	42.2 (35.2)	5.35 (3.58)
3	[35,50)	20	45%	90%	50%	60%	90%	70%	80%	223.7 (226.5)	16.39 (5.52)
4	≥50 yrs.	20	50%	70%	30%	60%	90%	80%	70%	272.6 (252.8)	18.64 (8.72)
	Overall	**80**	**37.5%**	**72.5%**	**37.5%**	**60%**	**82.5%**	**78.8%**	**71.3%**	**135.5 (203.3)**	**10.33 (9.17)**

3.1.2. Passengers

Tables 5 and 6 present demographic details of the non-drivers group (N_p=80 participants) who did not own a driver's license at the moment of the study. Professional occupations of the passengers included, in alphabetical order: bank clerk, cashier, engineer, flight attendant, librarian, mechanical engineer, medical assistant, professor, secretary, software engineer, student, and one participant was unemployed (retired).

Table 5. Demographic details of the sample of passengers (N_p=80 participants) and details regarding their regular traveling by car; continues with Table 6.

	Age group	N	Mean age (SD) [yrs.]	Gender (F/M)	Context of travel by car				Occupied seats	
					Every day	Almost every day	Once a week	Occasionally	Front seat	Back seat
1	<20 yrs.	20	18.8 (0.4)	60%, 40%	5%	20%	45%	30%	65%	45%
2	[20,35)	20	23.8 (3.4)	35%, 65%	20%	40%	15%	25%	80%	45%
3	[35,50)	20	40.0 (4.2)	85%, 15%	35%	45%	10%	10%	95%	10%
4	≥50 yrs.	20	54.1 (4.8)	50%, 50%	40%	15%	15%	30%	65%	50%
	Overall	**80**	**34.2**	**58%, 42%**	**25%**	**30%**	**21.3%**	**23.7%**	**76.3%**	**37.5%**

Table 6. Demographic details of the sample of passengers (N_p=80 participants) and details regarding their regular traveling by car; continued from Table 5.

	Age group	N	Main purpose for traveling by car					
			Work duties	Commuting	Private travel	Shopping	Visiting family	Holiday travel
1	<20 yrs.	20	5%	55%	50%	65%	75%	65%
2	[20,35)	20	5%	55%	40%	70%	65%	65%
3	[35,50)	20	40%	75%	45%	85%	75%	70%
4	≥50 yrs.	20	5%	45%	15%	65%	75%	45%
	Overall	**80**	**13.75%**	**57.5%**	**37.5%**	**71.25%**	**72.5%**	**61.25%**

3.2. Task

Participants filled in a Google Forms questionnaire, which they received by email or via Facebook; in a few cases, the questionnaires were filled in by the experimenter during phone calls (with five of the participants). We designed two forms of the questionnaire, one for drivers and one for passengers, containing specific questions for each category of car travelers. The two questionnaires overlapped on parts regarding demographic information, *e.g.*, age, current occupation, etc. Specific questions regarded driving habits and experience (for drivers), reasons for traveling by car (for passengers), and preferences for using gesture and voice input in the car (for both drivers and passengers).

3.2.1. Questionnaire

The following information was collected from drivers:

(a) *Demographic information*: age, gender, and professional occupation or education in case of unemployment at the time of the study.

(b) *Driving experience and habits:* the number of years since they were driving, the approximate number of kilometers driven, main purpose for using the car, and information regarding the regular occupants of the car (other passengers).

(c) *Use of smart devices*. We asked drivers to indicate devices they were using on a regular basis by choosing from the following list: *laptop*, *tablet*, *smartphone*, *smartwatch*, *smartglasses*, *smart TV*, *smart ring*, *head-up display*, *GPS device*, *portable radio/MP3 player*, *wireless headphones*, *video game consoles*, and *other*; in case the latter option was chosen, participants were asked to specify the type of device they were using on a regular basis. The questionnaire also asked which, if any, of these devices were controlled using either gesture or voice input.

(d) *Use of smart devices in the car:* drivers indicated which devices, from the above list, were also used in the car on a regular basis.

(e) *Perceived utility of gesture and voice input in the car:* drivers specified the degree of perceived utility for gesture and voice input to control each of the following ten car functions: (1) make phone calls using the in-vehicle integrated Bluetooth interface, (2) adjust car mirrors, (3) control the air conditioner (AC), (4) control the aerators mouth, (5) control the integrated multimedia system/radio/CD, (6) GPS navigation system, (7) head-down display (HDD), (8) head-up display (HUD), (9) control the car windows, and (10) adjust the chairs. Participants' answers were collected using two 5-point Likert scales, one for gesture and one for voice input, with the following options: 1 – "not useful at all," 2 – "little useful," 3 – "moderately useful," 4 – "useful," and 5 – "very useful."

A similar form of the questionnaire (except for the driving experience and habits section) was employed for passengers.

3.3. Results

Tables 3 to 6 show a wide variety of driving behavior, including frequency and context of driving, and a variety of reasons for passengers to choose the car as a means of transportation. In the following, we focus on our participants' perceptions of gesture and voice input in the car and on their preferences for controlling various car functions.

3.3.1. Self-Reported Use of Smart Devices

Figure 1 summarizes device use, in percentages, reported by our sample of drivers and passengers and Table 7 highlights which of those devices were controlled at least once using either voice or gesture input. Results showed that most participants employed laptops (82.5% of the drivers and 83.7% of the passengers, $\chi^2_{(1)}$=0, p=1, *n.s.*), smartphones (95% and 90%, $\chi^2_{(1)}$=0.811, p=.368, *n.s.*), GPS navigation (87.5% and 73.7%, $\chi^2_{(1)}$=4.001, p=.045), and radio devices (88.7% and 85%, respectively,

$\chi^2_{(1)}$=0.219, p=.640, $n.s.$). Few participants reported using smartwatches (21.2% of the drivers and 16.2% of the passengers, $\chi^2_{(1)}$=0.370, p=.543, $n.s.$), and very few reported having used Augmented Reality smartglasses or smart rings (less than 10%, $\chi^2_{(1)}$=0, p=1, $n.s.$). Participants reported using gestures to operate smartphones, video game consoles, and the in-vehicle infotainment system, while voice input was used to interact with personal assistants, such as Alexa or Amazon Echo, and to operate smartphones and GPS navigation, among others; see Table 7.

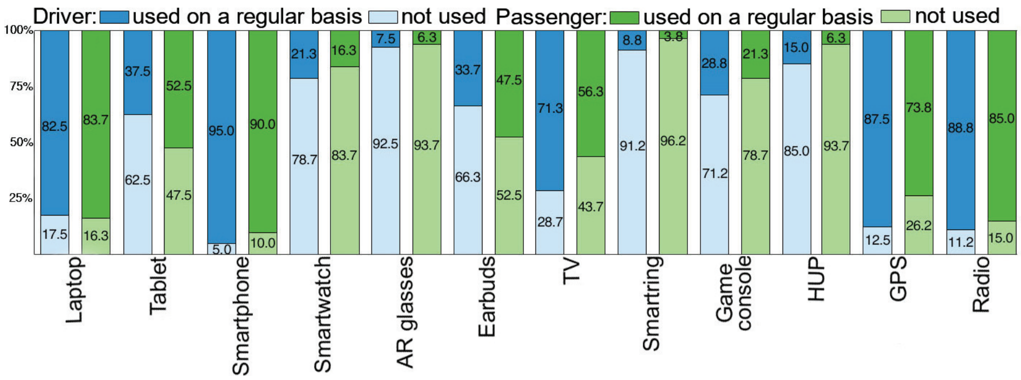

Figure 1. Overview of devices used by drivers and passengers on a regular basis.

Table 7. Details regarding devices used on a regular basis (continuation of Figure 1).

	Age group	Devices used by drivers			Devices used by passengers		
		N	Voice input	Gesture input	N	Voice input	Gesture input
1	<20 yrs.	20	Smartphone, robots, Siri, laptop, GPS, smart TV	Smartphone, smart TV, tablet, in-car LCD display	20	Smartphone, tablet, laptop	Smartphone, gates, lights
2	[20,35)	20	Smartphone, Alexa, GPS, in-car infotainment system, lights and heating	Smartphone, Kinect, video games	20	Smartphone, GPS, Amazon Echo, laptop, Alexa, Google voice assistant, Siri, smart door, smart TV	Smartphone, XBox, Kinect, Oculus Rift, Nintendo Wii, tablet board, car electric windows
3	[35,50)	20	Smartphone, Siri, XBox	Smartphone, XBox	20	Smartphone, GPS, Siri	Smartphone, XBox
4	≥50 yrs.	20	Smartphone	Myo, Leap Motion	20	Smartphone, smart TV	Smartphone

3.3.2. Perceived Utility of Gesture and Voice Input

We asked participants to rate their perceived utility of gesture and voice input; see Figures 2 and 3 that show the preferences of drivers and passengers, respectively, reported for each age group. For the drivers group, we observed more preference for voice input compared to gesture commands for all age groups, *e.g.*, drivers less than 20 years old considered voice input "useful" (rating of 4) and "very useful" (rating of 5) in 90% of cases, while they thought gesture commands were useful and very useful, respectively, according to 70% of their answers; see Figure 2. Preferences for voice input increased with age, *e.g.*, drivers over 50 years considered voice input "very useful" in 65% of their answers and "useful" in other 25% of their responses, while gesture commands were "useful" in 50% of the cases and "very useful" in other 25%. For passengers, we noticed a similar trend where voice input was preferred to gestures overall, but there were more ratings toward the lower end of the Likert scale showing that gesture commands were perceived less useful in general, *e.g.*, passengers between 20 and 35 years old considered gesture commands “not useful at all” and “little useful,” according to 40% of their answers; see Figure 3.

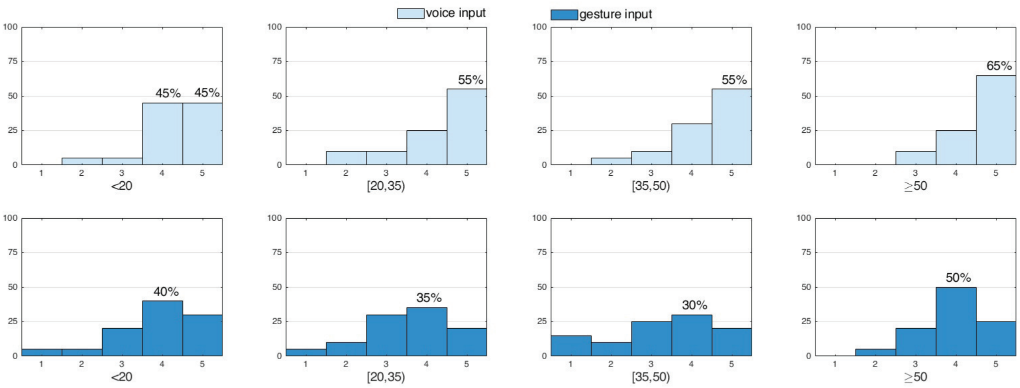

Figure 2. Drivers’ preferences for gesture and voice input collected using a 5-point Likert scale. *Notes:* 1 indicates that gesture/voice is “not useful at all” and 5 “very useful.” The mode is highlighted in each chart, *e.g.*, 55% of the drivers with ages between 35 and 50 years considered voice input very useful.

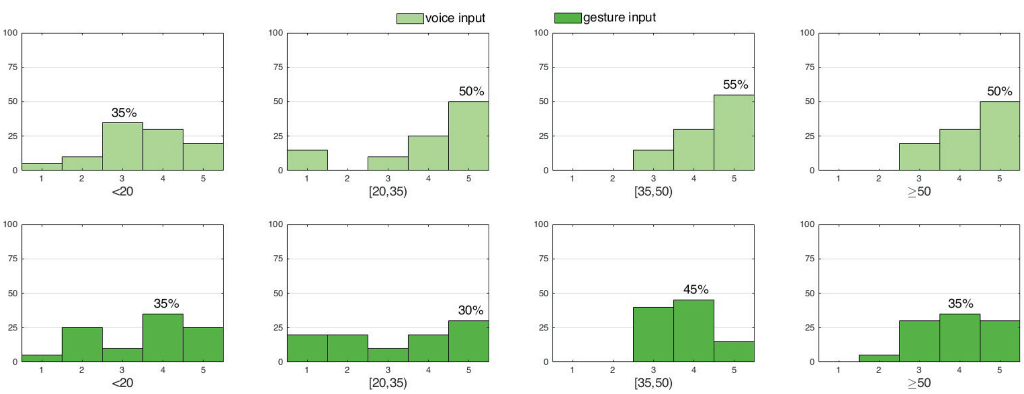

Figure 3. Passengers’ preferences for using gesture and voice input collected using 5-point Likert scales. *Notes:* 1 indicates that gesture/voice is “not useful at all” and 5 “very useful.” The mode is highlighted in each chart, *e.g.*, 45% of the passengers between 35 and 50 years old considered gesture input “useful.”

We also asked drivers and passengers about the car functions they would like to control using gesture and voice input; see Figure 4 for their preferences. Overall, drivers preferred voice over gesture input when compared to passengers: 50% of the drivers favored voice input and 23.9% favored gesture input compared to 45.4% and 27.5% of the passengers, respectively, but we did not find a statistical significant result ($\chi^2_{(3)}$=1.173, p=.760, *n.s.*). For example, 78.8% of the drivers preferred voice commands to control GPS navigation compared to 63.7% of the passengers ($\chi^2_{(1)}$=3.692, p=.05, *n.s.*) and 30.0% *vs.* 25.0% for controlling car seats ($\chi^2_{(1)}$=0.282, p=.595, *n.s.*). Exceptions were also observed, *e.g.*, 57.5% of the drivers preferred gestures to control the in-vehicle AC compared to 46.2% of the passengers ($\chi^2_{(1)}$=1.602, p=.206, *n.s.*). Overall, passengers reported more preference for using gesture commands compared to drivers, but results were again not statistically significant, *e.g.*, 28.8% *vs.* 27.5% for controlling the AC with gestures ($\chi^2_{(1)}$=0, p=.1, *n.s.*). For some functions, some participants preferred not to use either gesture or voice input, such as 17.5% of the drivers and 15% of the passengers for car windows control ($\chi^2_{(1)}$=0.046, p=.830, *n.s.*) and 40% of the drivers and 22.5% of the passengers for controlling car mirrors ($\chi^2_{(1)}$=4.916, p = .027<.05).

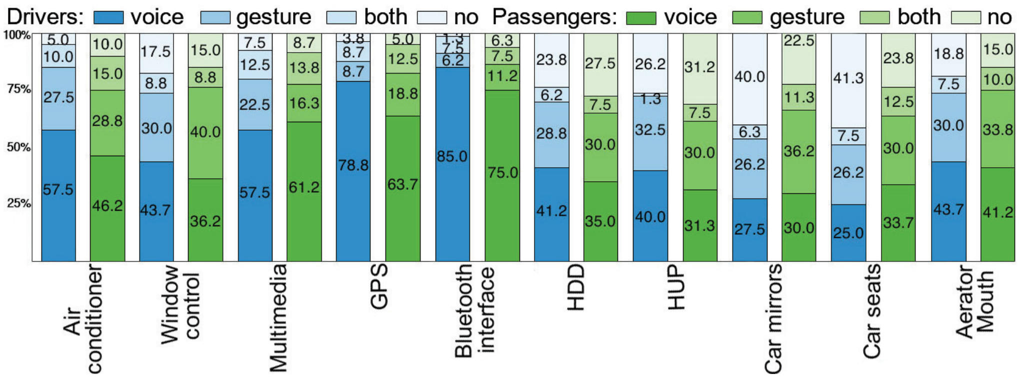

Figure 4. Preferences for controlling specific car functions using voice and gesture input.

4. ELICITATION OF PREFERRED GESTURE AND VOICE INPUT COMMANDS FOR CONTROLLING CAR FUNCTIONS

Informed by the results regarding drivers' and passengers' preferences for controlling car functions, we applied the end-user elicitation method (Vatavu & Wobbrock, 2015, 2016; Villarreal-Narvaez, Vanderdonckt, Vatavu, & Wobbrock, 2020; Wobbrock et al., 2005, 2009) to understand proposals of actual gesture and voice commands that drivers and passengers would like to use to control various car functions.

4.1. Participants

Twenty drivers (N_d=20) and twenty non-drivers or passengers (N_p=20), not having participated in the first study, volunteered for the end-user gesture and voice elicitation experiment. We selected volunteers to form the same age groups as in the previous study, *i.e.*, participants younger than 20 years old, between 20 and 35 years old, between 35 and 50, and older than 50 years; see Table 8. All participants were right handed.

Table 8. Demographic details about the drivers (N_d=20) and passengers (N_p=20) that volunteered to take part in the end-user gesture and voice elicitation study.

	Age group	Drivers			Passengers		
		N	Mean age (SD) [yrs]	Gender (F/M)	N	Mean age (SD) [yrs]	Gender (F/M)
1	<20 yrs.	5	18.8 (0.44)	40%/60%	5	18.8 (0.44)	20%/80%
2	[20,35)	5	32.0 (2.91)	40%/60%	5	27.4 (2.88)	40%/60%
3	[35,50)	5	41.6 (6.42)	60%/40%	5	42.8 (3.34)	60%/40%
4	≥50 yrs.	5	56.0 (4.30)	60%/40%	5	56.6 (5.68)	60%/40%
	Overall	**20**	**37.1 (3.51)**	**50%/50%**	**20**	**36.4 (3.08)**	**45%/55%**

4.2. Task

Following the end-user elicitation method (Wobbrock et al., 2005, 2009), we presented participants with a list of 20 car functions, explored in the previous study,[3] denoted in the following as "referents" according to the terminology from Wobbrock et al. (2009); see Table 9. Then, we asked participants to think about and propose *gesture and voice commands that they would like to use to control those functions in the car.* Each participant proposed 20 gestures and 20 voice commands, one for each referent. The only instructions that we provided to participants were that their commands should be executed *safely* while in the car (either when driving − for drivers, or when occupying a passengers' seat − for passengers), be *intuitive*, *easy to remember,* and *easy to execute*. Except for these general guidelines, we did not constrain participants' choice of commands to any particular gesture or voice input technology (*e.g.*, free-hand or touch input) to avoid legacy bias (Morris et al., 2014), and we told them that any gesture and voice commands they would propose were recognizable by our system. Consequently, no actual gesture or voice input system was present in the car during the study in order to encourage participants to focus on commands and not on the particular characteristics of some gesture/voice input and recognition system. Moreover, there were no constraints imposed on the gesture size, speed, or location of the gestures articulated inside the vehicle or on the vocabulary permitted for voice input commands. During the experiment, participants were seated in the car, either in the driver's seat for drivers or in one of the other seats for passengers. The order of referents was randomized per participant. To avoid potentially hazardous situations, the car used in the study was parked during the entire duration of the elicitation experiment. The commands proposed by participants were video recorded for analysis purposes.

4.3. Dimensions of Analysis and Measures

We employed the following criteria to evaluate and characterize drivers' and passengers' preferences for gestures commands which they would like to use in the car:

(1) **Gesture structure** characterizes gestures with respect to the information value contained by the hand poses and overall movement of the fingers, hands, and arms performed during the articulation of the gesture. We considered three categories for this criteria: *posture*, *motion*, and *mixed* gestures, inspired by Gheran, Vanderdonckt, and Vatavu (2018) and Vatavu and Pentiuc (2008). For example, we considered that the "thumb-up" gesture falls in the *posture* category since there is no movement that is relevant for the meaning of the gesture; a circle-shaped gesture performed in mid-air would fall into the *motion* category, since the pose of the hand during the movement is less important than the movement itself and does not bring key meaning to the gesture; while a compound gesture composed of a "thumb-up" hand pose followed by a circle shape drawn in mid-air would fall into the *mixed* category.

(2) **Handedness** with three categories according to the hand(s) used to articulate the gesture. We considered gestures performed with the *dominant* hand to be different from gestures performed with the *nondominant* hand or with *both hands*, even if the same hand poses and motion were adopted during both articulations.

[3]Note that in the previous study we had 10 functions, but they included both "on" and "off" commands, *e.g.*, participants from the first study were asked about their preferences for controlling the air conditioner, where in this second study we asked for their specific commands to turn the air conditioner on and off.

Table 9. Referents used for the end-user gesture and voice input elicitation study.

	Referent	Preference for voice and gesture input†	Used in other elicitation studies
1	Answer phone call	96.2%	Ohn-Bar and Trivedi (2014); Sokol et al. (2017); Vieira et al. (2016); Wu et al. (2019)
2	Close phone call	96.2%	Ohn-Bar and Trivedi (2014); Sokol et al. (2017); Vieira et al. (2016); Wu et al. (2019)
3	Turn on Air Conditioner	92.5%	Wu et al. (2019)
4	Turn off Air Conditioner	92.5%	Wu et al. (2019)
5	Turn on multimedia	91.9%	Akyol et al. (2000); Graichen et al. (2019); Ohn-Bar and Trivedi (2014); Sokol et al. (2017); Wu et al. (2019)
6	Turn off multimedia	91.9%	Akyol et al. (2000); Graichen et al. (2019); Ohn-Bar and Trivedi (2014); Sokol et al. (2017); Wu et al. (2019)
7	Next song	91.9%	Akyol et al. (2000); Angelini et al. (2016); Ohn-Bar and Trivedi (2014); Sokol et al. (2017); Vieira et al. (2016); Wu et al. (2019)
8	Previous song	91.9%	Akyol et al. (2000); Angelini et al. (2016); Ohn-Bar and Trivedi (2014); Sokol et al. (2017); Vieira et al. (2016); Wu et al. (2019)
9	Roll down window	83.75%	Pfleging et al. (2012)
10	Roll up window	83.75%	Pfleging et al. (2012)
11	Turn on HUD	71.3%	Pfleging et al. (2012)
12	Turn off HUD	71.3%	Pfleging et al. (2012)
13	Turn on HDD	74.35%	Graichen et al. (2019); May et al. (2014); May et al. (2017); Vieira et al. (2016); Wu et al. (2016)
14	Turn off HDD	74.35%	Graichen et al. (2019); May et al. (2014); May et al. (2017); Vieira et al. (2016); Wu et al. (2016)
15	Turn on GPS	95.6%	Graichen et al. (2019); Ohn-Bar and Trivedi (2014); Sokol et al. (2017); Wu et al. (2019)
16	Turn off GPS	95.6%	Graichen et al. (2019); Ohn-Bar and Trivedi (2014); Sokol et al. (2017); Wu et al. (2019)
17	Temperature increase	92.5%	Pfleging et al. (2012); Wu et al. (2019)
18	Temperature decrease	92.5%	Pfleging et al. (2012); Wu et al. (2019)
19	Elevate aerator mouth	83.1%	Wu et al. (2019)
20	Descend aerator mouth	83.1%	Wu et al. (2019)

†Average preference for both drivers and passengers, resulted from the informative study; see Section 3.

(3) **Scale** characterizes the amplitude of the articulated gestures with three categories: *small*, *medium*, and *large*, corresponding to movements performed at finger, hand and wrist, and arm level; see Gheran et al. (2018) that employed this criterion in their gesture elicitation study on smart ring gestures. For example, we considered a small circle performed with one finger to be different from the same circle-shaped gesture performed with the entire arm.

(4) **Complexity** divides gestures into *simple* and *compound*. For example, we considered a circle shape performed in mid-air as a *simple* gesture, but a circle followed by the "thumbs-up" hand pose as a *compound* gesture.

(5) **Locale.** Gestures performed in various spatial locations inside the vehicle were considered to be different, even if they reduced to the same hand poses and motion trajectory. For this criterion, we considered three categories: gestures performed *in the air*, *on other surfaces*, and on the *steering wheel*.

(6) **Body-referenced locale.** Gestures performed in different spatial regions around the participant's body were considered different as well. For example, the same circular motion performed to the *right*, in *front*, *back*, or to the *left* of the body was considered to represent different gestures by consequence of the specific location, relative to the user's body, where hands movement was performed.

We also defined the **Production-Time** dependent variable to measure the time, in

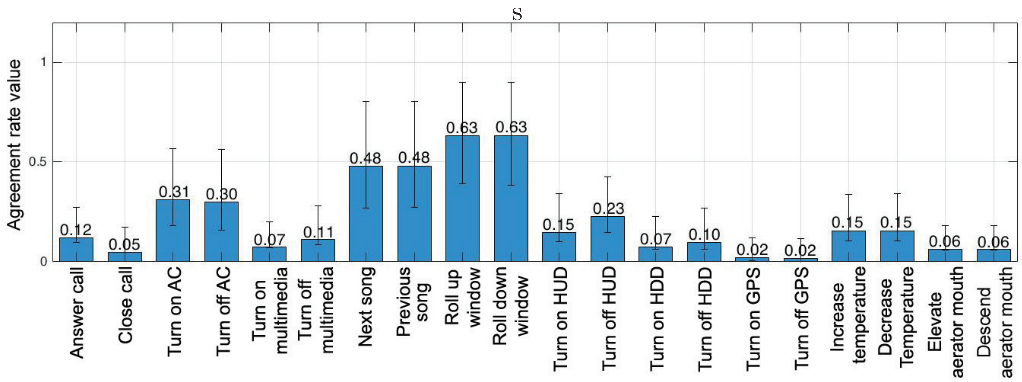

Figure 5. Agreement rates obtained for gestures elicited from drivers (mean AR=.210).

seconds, needed by our participants to execute the gesture for each car function. For gesture input, production times represent the equivalent of task times for general tasks studied in Human-Computer Interaction; see Leiva, Martín-Albo, Plamondon, and Vatavu (2018) and Leiva, Martín-Albo, and Vatavu (2018) for a discussion.

The following measures were employed to evaluate and characterize drivers' and passengers' preferences for voice input commands they would like to use in the car:

(1) **Number of words** of the command, *e.g.*, "turn radio on" has three words.
(2) **Number of distinct words** characterizes the size of the vocabulary employed by our participants to propose voice commands for the entire set of 20 referents.
(3) **Language preference**. Participants were free to propose commands either in their native language (anonymous) or in English, whichever seemed more fit.

We employed the **Agreement-Rate (AR)** measure of Vatavu and Wobbrock (2015) to compute the degree of agreement between participants' proposals for gesture and voice commands they would like to use for in-vehicle interactions, as follows:

$$AR(r) = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}\delta_{i,j}}{N \cdot (N-1)} \quad (1)$$

where r denotes the referent (*e.g.*, "Next song"); N is the number of proposals (or participants, since one proposal was elicited from each participant); and $\delta_{i,j}$ computes to 1 if the commands proposed by participants i and j are equivalent and to 0 otherwise. An agreement value closer to 1 denotes more consensus among the proposals elicited from participants, while an agreement value of 0 represents no consensus or absolute disagreement. We computed AR for each referent, each input modality (gesture and voice), and each category of users (drivers and passengers), which resulted in 20 $\times$ 2 $\times$ 2 = 80 agreement rates. To evaluate gesture similarity and equivalence, we used the structure, handedness, scale, complexity, locale, and body-referenced locale categories. For voice input, we considered that two commands were equivalent if they had the same words, not necessarily in the same order; *e.g.*, we considered "volume up" and "up volume" equivalent. Also, we considered two voice commands equivalent if they contained the same words, even if the commands were proposed in two different languages (native and English, respectively).

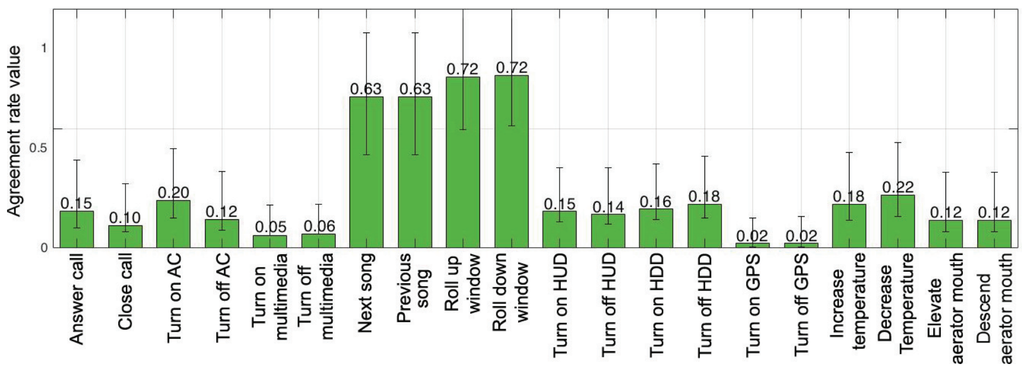

Figure 6. Agreement rates obtained for gestures elicited from passengers (mean AR=.235).

4.4. Results: Gesture Commands

We collected a total number of 800 proposals for gesture commands from our sample consisting of 20 drivers, 20 passengers, and 20 referents representing various functions to control inside the vehicle. In the following, we report agreement analysis results, discuss some of the most frequently occurring gesture commands among participants' proposals, and characterize gestures using specific measures of gesture articulation performance, such as production time, scale, handedness, and locale.

4.4.1. Agreement over Gesture Commands

Figures 5 and 6 present agreement rates for each referent considered in our study. Overall, agreement among drivers varied between .02 and .63 (M=.21, SD=.20, Mdn=.14) and among passengers between .02 and .72 (M=.23, SD=.23, Mdn=.15). A Wilcoxon rank sum test (with continuity correction) indicated no statistically significant difference between drivers and passengers (W=179.5, p = .588>.05, *n.s.*). Some referents, such as "Next song," "Previous song," "'Roll up window," and "Roll down window" received high consensus for drivers and passengers alike (average .55 for drivers and .68 for passengers, respectively). This result is probably due to implicit directionality semantics embedded in those referents, *e.g.*, left and right for advancing to the next and previous song, and up and down for sliding the car window. An interesting observation is that pairs of dichotomous referents, *e.g.*, "Turn on AC" and "Turn off AC," "Next song" and "Previous song," etc., received similar levels of agreement. The average absolute difference in agreement rates for dichotomous referents was .02 (SD=.03) for drivers and passengers alike. Some referents resulted in low agreement rates, *e.g.*, .05 and .10 for "Close call" and .07 and .05 for "Turn on multimedia." According to the recommendations of Vatavu and Wobbrock (2015) for interpreting the magnitudes of agreement rates, 75% of the referents received "low" (*i.e.*, less than .10) and "medium" (less than .30) agreement in the drivers group and 80% in the passengers group. The rest (25% and 20%, respectively) reached "high" (*i.e.*, greater than .30, but less than .50) and "very high" (greater than .50) agreement rates. Next, we look at the individual gestures proposed by our participants and discuss frequent gesture proposals.

4.4.2. Frequent Gesture Proposals

We employed the six criteria from the previous section (gesture structure, handedness, scale, complexity, locale, and body-referenced locale) to describe each gesture proposal in a space with six dimensions and a total number of $3\times3\times3\times2\times3\times4=648$ possible characterizations. For example, a swipe movement to effect"Answer phone call" can be described by the tuple ("motion", "dominant hand", "medium scale", "simple", "in the air", "to the right of the body"). Figures 7 and 8 show a selection of our participants' most frequent gesture commands proposed for each referent. For example, five drivers suggested the "call me" emblematic hand gesture for answering a phone call (Figure 7a,b), and three drivers proposed "thumbs-up" followed by swipe down to end the call. Most passengers used the "call me" hand pose as well for answering and ending the call (Figure 8a,b). Frequent gestures for turning on and off the AC were circular motions performed in mid-air, for which the directions were clockwise for turning the AC on and counterclockwise for turning it off (Figure 7c,d and Figure 8c,d). To motivate their choice for this specific gesture, one participant mentioned:

"This gesture reminds me of propeller movements."

Swipe up and swipe down gestures performed with the open hand were employed by our participants to roll up and down the car windows (see Figure 7i,j and Figure 8i,j), for which we identified the following motivation:

"This gesture is intuitive for rolling up the window, safe, and easy to perform."

Four of the drivers chose to point their hands to the center of the console in order to turn on the multimedia system, just like they would push a physical button, while six of the drivers proposed a specific hand pose to turn the music off (Figure 7e,f):

"Pushing an imaginary button in mid-air reminds me of using a remote control or one of the buttons from the car."

"I think that this hand pose [Figure 7f] *is emblematic for ending a certain action."*

Passengers also prefered to draw a waveform shape in mid air for turning on and off the multimedia system; see Figure 8e,f. The gestures proposed by passengers to change the song (Figure 8g,h) were influenced by their experience using smartphones:

"Browsing through the songs reminds me of my smartphone, so it feels natural to me to use the same gestures."

Three of the drivers and seven of the passengers controlled the aerator mouth using gestures that manipulated imaginary objects in mid-air; see Figure 7s,t and Figure 8s,t. Regarding the GPS referents, three of the passengers suggested a gesture to fold out an imaginary map (Figure 8o,p) and two of the drivers preferred to draw in mid-air a curve shape going from left to right (Figure 7o,p):

"Navigation by GPS reminds me of using paper maps."

The gestures thumbs-up, thumbs-down, swipe up, and swipe down to increase and decrease the temperature inside the vehicle were associated by passengers with gestures they would use to communicate such requests to other people; see Figure 8q,r:

"When I want to communicate this to another person, I make this gesture."

Gestures proposed by passengers to turn on and off the Head-Up and Head-Down displays were performed by picking up invisible objects and placing them near the displays, among other proposals (Figure 8k,l,m,n):

"It is like moving information around, but more like taking invisible information and make it visible on the display."

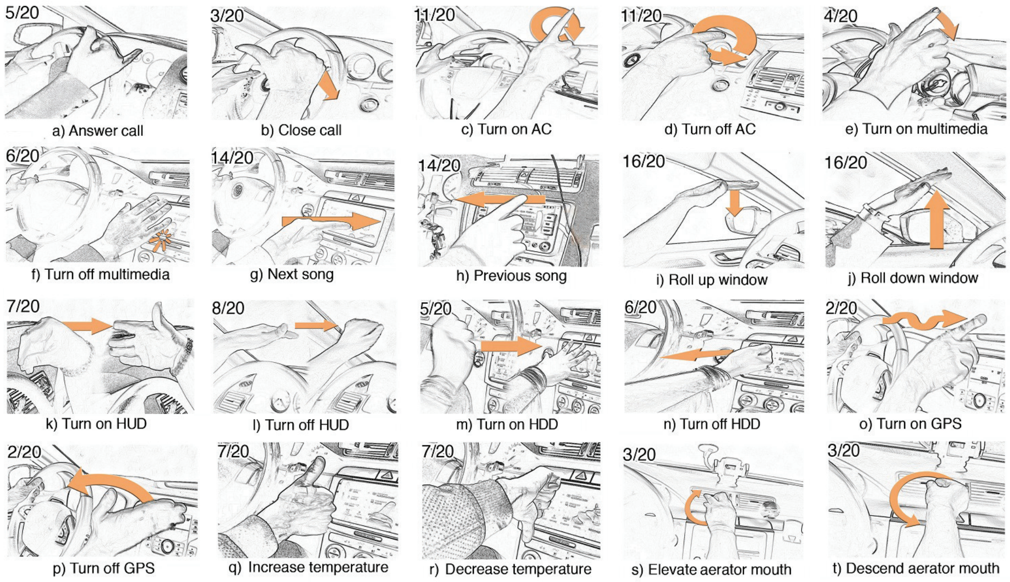

Figure 7. A selection of gestures frequently observed among drivers' proposals for commands they would like to use in the car. *Notes:* the numbers from the top-left of each figure (*e.g.*, 5/20 for the "Answer call") show the numbers of drivers, out of 20, proposing the gesture illustrated in that figure. The gestures are referenced and discussed in the text.

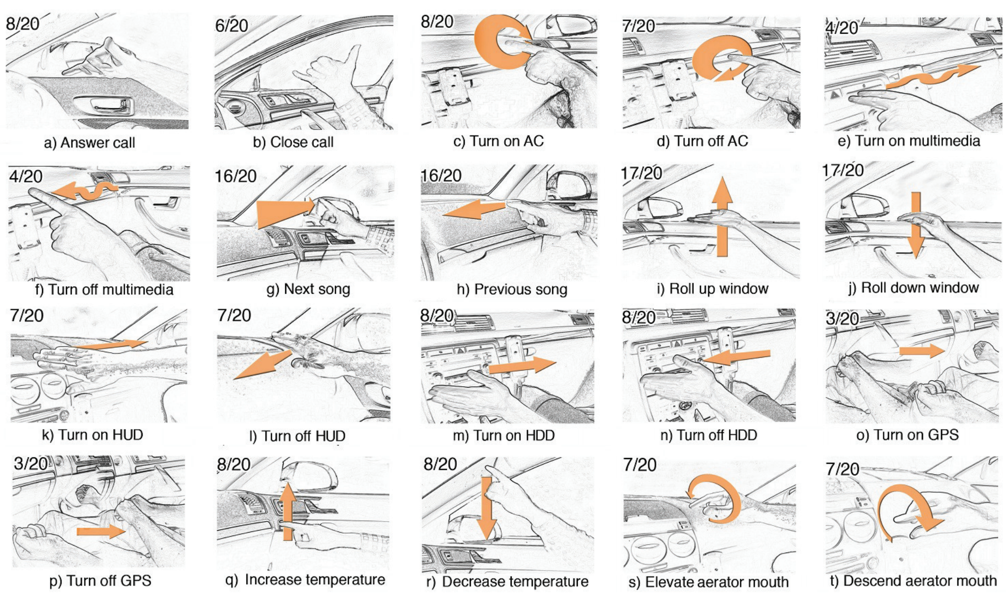

Figure 8. A selection of gestures frequently observed among our passengers' proposals for commands they would like to use in the car. *Notes:* the numbers from the top-left of each figure (*e.g.*, 8/20 for the "Answer call") show the numbers of passengers, out of 20, proposing the gesture illustrated in that figure. The gestures are referenced and discussed in the text.

4.4.3. Gesture Articulation Performance

In the following, we characterize the gestures proposed by our participants in terms of their performance and properties. To avoid pseudoreplication problems affecting Chi-Square tests for count data under repeated measurements (Lazic, 2010), we computed aggregated normalized values, between 0 and 1, for each referent and each gesture category, *e.g.*, if 7 of the 20 drivers performed gestures with a *mixed* structure for the "Next song" referent, then we computed the ratio 7/20=.35 as the measured value for that referent. This approach (Lazic, 2010) enabled us to use the t-test under normality and homoscedasticity, Welch's variant of the t-test under normality and heteroscedasticity, and the Wilcoxon-Mann-Whitney test for non-normal data. Figure 9 shows the distribution of gestures elicited from drivers and passengers according to their structure, handedness, scale, complexity, locale, and body-referenced locale. We found that drivers proposed and articulated *simple* gestures in 88.75% of the cases and passengers in 82.00% (W=268.5, p=.063, *n.s.*). Also, drivers performed *motion* gestures in 74.5% of their articulations and passengers in 69.25% ($t_{(38)}$=1.052, p=.299, *n.s.*). The scale of gestures was *medium* for 51.25% of the drivers' and for 57% of the passengers' gesture articulations ($t_{(38)} = -0.833$, p=.410, *n.s*), and drivers tended to proposed less *large* gestures compared to passengers, 12.25% *vs.* 22.0%, although the result was not statistically signifcant (W=161.5, p=.297, *n.s.*). The majority of the gestures were articulated using the *dominant* hand, 89.75% and 74.5%, respectively (W=330, p<.001) and just a few passengers (2.5%) proposed bimanual gestures. Also, gestures were performed vastly in mid-air by 91.75% of the drivers and 99.5% of the passengers, respectively (W=42, p<.001). We also found that, on average, drivers were significantly faster than passengers, and performed gestures in 1.49 s (SD=.29) compared to 1.67 s (SD=.23) for passengers ($t_{(38)}$=2.138, p=.039<.05, Cohen's d=.67).

Drivers preferred to execute gesture commands to the right of their body (80.75% *vs.* 23.0%, W=356, p<.001), while passengers exploited the space around their body to a greater extent and, consequently, they performed gestures in front (42%) and to the left (35%) as well. Even if drivers had the steering wheel to use as a support for touch input, very few (4.75%) actually proposed such gestures. The majority of mid-air gestures that was elicited in our study suggests that gesture acquisition technology and solutions based on finger augmentation devices, such as smart rings (Gheran & Vatavu, 2020), might work well inside the vehicle for both drivers and passengers to complement or replace the use of steering wheel controls for drivers and other in-vehicle controls for passengers, and they should be examined more closely in future work.

4.5. Results: Voice Commands

We collected a total number of 800 proposals for voice commands from our sample of 20 drivers, 20 passengers, and 20 referents representing functions to control in the car. In the following, we discuss some of the most frequently proposed voice commands among our participants' proposals, and we characterize the elicited commands using specific measures, such as the average number of words and the vocabulary size.

4.5.1. Frequent Proposals for Voice Input Commands

Table 10 lists the most frequent voice input commands proposed by both drivers and passengers for our set of 20 referents. Except the "Turn on multimedia" referent for which two different commands were proposed by the two groups ("Music" and "Start

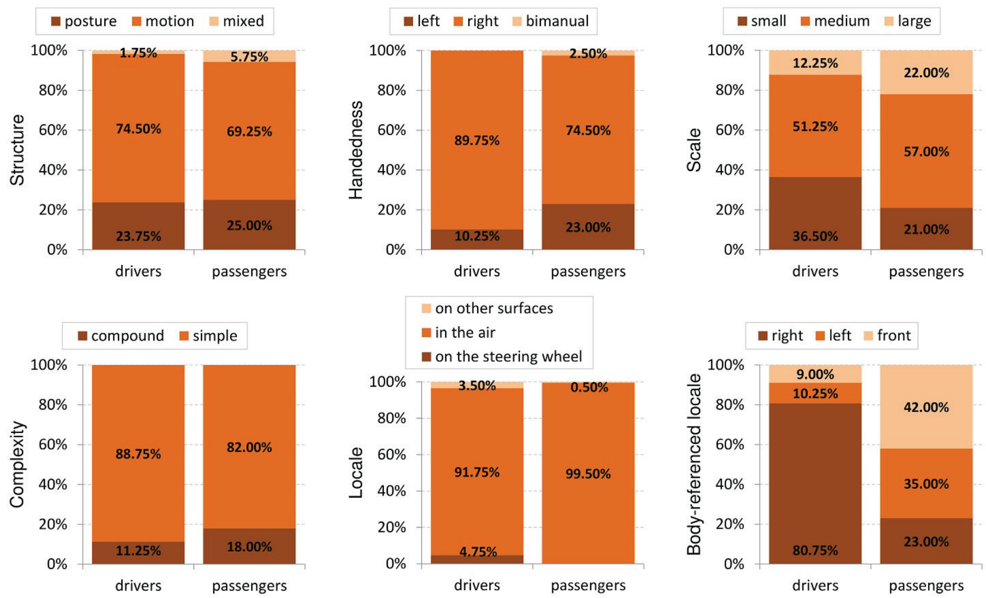

Figure 9. Drivers and passengers' gestures characterized along the six dimensions of our analysis.

music," respectively), all the referents resulted in the same voice input commands occurring most frequently among our participants' proposals and with similar frequencies. For example, eight of the drivers (40%) and nine of the passengers (45%) proposed the voice commands "Stop air" to turn off the air conditioner ($\chi^2_{(1)}$=0, p=1, $n.s.$). Except for the temperature control functions that use three words, all the other most frequent voice commands are composed of two words only. Table 11 presents frequency results of the top-10 most frequently used words by both drivers and passengers. The words "Start," "Stop," "Close," "Open," "Up," and "Down" appear frequently as they were reused by our participants for specifying the "on" and "off" states of multiple functions in the car, such as the Head-Down Display, Head-Up Display, GPS, multimedia system, and the AC. The top-10 most frequently used words represent 44.4% of all the words used by drivers and 37.01% of the words used by passengers to propose voice input commands. A percent of 40.38% of the vocabulary used by our participants is represented by just 10 words.

Table 10. A selection of voice input commands frequently observed among our participants' proposals for commands they would like to use in the car. *Notes:* the numbers following each command (*e.g.*, 9/20 for "answer phone call") show the number of participants proposing the gesture illustrated in that figure out of the sample of 20 drivers (second column) or 20 passengers (third column).

	Function	Drivers	Passengers
1	Answer phone call	Answer (9/20)	Answer (10/20)
2	Close phone call	Close (9/20)	Close (7/20)
3	Turn on Air Conditioner	Start air (6/20)	Start air (6/20)
4	Turn off Air Conditioner	Stop air (8/20)	Stop air (9/20)
5	Turn on multimedia	Music (9/20)	Start music (7/20)
6	Turn off multimedia	Stop music (8/20)	Stop music (10/20)
7	Next song	Next song (11/20)	Next song (11/20)
8	Previous song	Previous song (11/20)	Previous song (11/20)
9	Roll down window	Window down (8/20)	Window down (11/20)
10	Roll up window	Window up (9/20)	Window up (11/20)
11	Turn on HUD	Open HUD (8/20)	Open HUD (10/20)
12	Turn off HUD	Stop HUD (8/20)	Stop HUD (10/20)
13	Turn on HDD	Open HDD (8/20)	Open HDD (9/20)
14	Turn off HDD	Stop HDD (8/20)	Stop HDD (10/20)
15	Turn on GPS	Open GPS (8/20)	Open GPS (9/20)
16	Turn off GPS	Close GPS (10/20)	Close GPS (9/20)
17	Temperature increase	Rise the temperature (9/20)	Rise the temperature (7/20)
18	Temperature decrease	Drop the temperature (9/20)	Drop the temperature (7/20)
19	Elevate the aerator mouth	Air up (8/20)	Air up (8/20)
20	Descend the aerator mouth	Air down (8/20)	Air down (8/20)

Table 11. The top-10 most frequently used words by drivers (N=798 words) and passengers (N=904 words) in their proposed voice commands to control the various car functions. *Note:* all commands were translated into English.

	Drivers		Passengers		Drivers and Passengers	
	Word	**Frequency**	**Word**	**Frequency**	**Word**	**Frequency**
1	Stop	7.26%	Stop	5.64%	Stop	6.40%
2	Start	5.63%	Air	4.42%	Air	4.93%
3	Air	5.51%	Music	3.98%	Start	4.70%
4	Close	4.63%	Window	3.98%	Window	4.05%
5	Window	4.13%	Start	3.87%	Close	3.99%
6	Music	3.88%	Close	3.42%	Music	3.93%
7	Up	3.75%	Down	3.09%	Up	3.40%
8	GPS	3.50%	Up	3.09%	GPS	3.29%
9	Down	3.38%	GPS	3.09%	Down	3.23%
10	Open	2.63%	Song	2.43%	Open	2.46%
	Total	**44.3%**		**37.01%**		**40.38%**

4.5.2. Voice Input Articulation Characteristics

Figure 10 shows measures of performance describing our participants' use of voice input commands. We found that drivers used a smaller number of words per command (2.02) compared to passengers (2.27), representing a statistically significant difference ($t_{(24.816)} = -2.635$, p=.014) as well as a smaller vocabulary (a mean number of 21.0 distinct words per participant) compared to the vocabulary of the passengers (22.4 distinct words per participant, W=151, p=.185, *n.s.*). The majority of the participants (70.5% of the drivers and 72.0% of the passengers) proposed voice input commands in their native language (anonymous for review), while approximatively 30% of our participants preferred to propose commands in English.

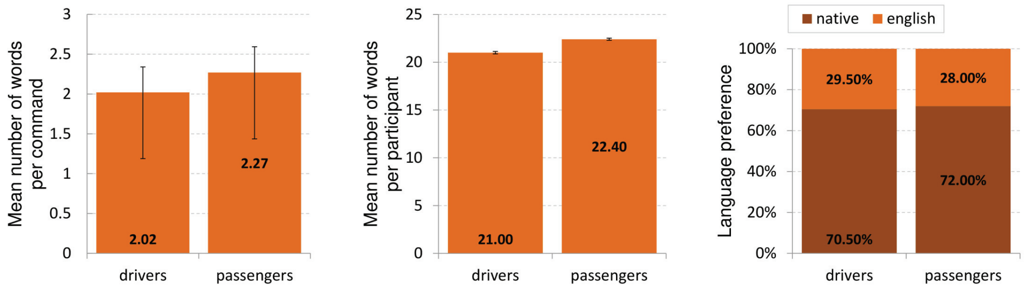

Figure 10. Articulation characteristics of the voice commands elicited from drivers and passengers.

5. CONCLUSION AND FUTURE WORK

We examined in this work drivers' and passengers' preferences for gesture and voice input for in-vehicle interactions and controlling various car functions. We presented results from a targeted literature survey, but also from two studies involving a large body of participants (160 and 40, respectively, representing 200 drivers and non-drivers). Our results showed many preferences for gesture and voice input in the car in terms of gesture structure, scale, and body-referenced gestures as well as differences between the preferences of drivers and passengers with regards to voice commands, *e.g.*, a smaller number of words employed by drivers per command. These results come to complement the existing literature on gesture and voice input and interactions inside the vehicle by providing the perspectives of both drivers and passengers. There are several limitations to our work, including the fact that gestures were performed in a stationary car, we did not record numerical representations of the elicited gestures (Vatavu, 2019), and did not perform an evaluation of user performance with preferred gesture and voice input commands compared to designer's defined commands. Future work will look at implementing a selection of gesture and voice commands resulted from our study and on evaluating usability and user experience in connection with safe operation and articulation of those commands inside the car. Another interesting direction of work is represented by the opportunity to include pedestrians in the interaction process with connected vehicles (Hamdani, Benamar, & Younis, 2019) toward extending the range of possible users and interactions in the context of Vehicle-to-Everything communications.

We hope that our results on the preferences for in-vehicle interactions from the perspective of drivers and passengers alike will be informative for researchers and practitioners toward designing and engineering automotive user interfaces based on natural input modalities inside the vehicle.

Funding

This work was supported by a grant of the Romanian Ministry of Research and Innovation, CCCDI-UEFISCDI, project number PN-III-P1-1.2-PCCDI-2017-0917, contract no. 21PCCDI/2018, within PNCDI III. Research was conducted in the Machine Intelligence and Information Visualization Lab (MintViz) of the MANSiD Research Center. The infrastructure was provided by the University of Suceava and was partially supported from the project *"Integrated center for research, development and innovation in Advanced Materials, Nanotechnologies, and Distributed Systems for fabrication* *and control"*, No. 671/09.04.2015, Sectoral Operational Program for Increase of the Economic Competitiveness, co-funded from the European Regional Development Fund.

References

Ahmad, B. I., Hare, C., Singh, H., Shabani, A., Lindsay, B., Skrypchuk, L., ... Godsill, S. (2018). Selection facilitation schemes for predictive touch with mid-air pointing gestures in automotive displays. In *Proc. of the 10th int. conf. on automotive user interfaces and interactive vehicular applications* (pp. 21–32). New York, NY, USA: ACM. Retrieved from `https://doi.org/10.1145/3239060.3239067`

Aiordăchioae, A., Vatavu, R.-D., & Popovici, D.-M. (2019). A design space for vehicular lifelogging to support creation of digital content in connected cars. In *Proc. of the acm sigchi symposium on engineering interactive computing systems.* ACM. Retrieved from `https://doi.org/10.1145/3319499.3328234`

Akyol, S., Canzler, U., Bengler, K., & Hahn, W. (2000). Gesture control for use in automobiles. In *Proceedings of the IAPR Conference on Machine Vision Applications* (p. 349-352).

Alba-Castro, J. L., González Agulla, E., & Loira, F. (2014). Hand Gestures to Control Infotainment Equipment in Cars. In *Proc. of IEEE Intelligent Vehicles Symposium* (pp. 1–6). Retrieved from `http://dx.doi.org/10.1109/IVS.2014.6856614`

Alpern, M., & Minardo, K. (2003). Developing a car gesture interface for use as a secondary task. In *CHI '03 Extended Abstracts on Human Factors in Computing Systems* (pp. 932–933). New York, NY, USA: ACM. Retrieved from `https://doi.org/10.1145/765891.766078`

Alvarez, I., Martin Hammond, A., Dunbar, J., Taiber, J., Wilson, D.-M., & Gilbert, J. (2010). Voice interfaced vehicle user help. In *Proc. of the 2nd int. conf. on automotive user interfaces and interactive vehicular applications* (p. 42-49).

Angelini, L., Baumgartner, J., Carrino, F., Carrino, S., Caon, M., Khaled, O. A., ... Sonderegger, A. (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* (p. 188-196). Fribourg, Switzerland. Retrieved from `https://hal.archives-ouvertes.fr/hal-01384007`

Angelini, L., Carrino, F., Carrino, S., Caon, M., Khaled, O. A., Baumgartner, J., ... Mugellini, E. (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* (pp. 1–8). New York, NY, USA: ACM. Retrieved from `https://doi.org/10.1145/2667317.2667414`

Angelini, L., Carrino, F., Carrino, S., Caon, M., Lalanne, D., Khaled, O. A., & Mugellini, E. (2013). Opportunistic synergy: A classifier fusion engine for micro-gesture recognition. In *Proceedings of the 5th international conference on automotive user interfaces and interactive vehicular applications* (pp. 30–37). New York, NY, USA: ACM. Retrieved from `https://doi.org/10.1145/2516540.2516563`

ASIRT. (2019). *Road safety facts by the association for safe international road travel.* Retrieved from `https://www.asirt.org/safe-travel/road-safety-facts/`

Barón, A., & Green, P. (2006). *Safety and usability of speech interfaces for in-vehicle tasks while driving: A brief literature review.* Ann Arbor, Mich.: University of Michigan, Transportation Research Institute, Human Factors Division. Retrieved from `https://books.google.ro/books?id=XDBANgAACAAJ`

Bilius, L.-B., & Vatavu, R.-D. (2020). A Synopsis of Input Modalities for In-Vehicle Infotainment and Consumption of Interactive Media. In *Proc. of the ACM International Conference on Interactive Media Experiences* (pp. 195–199). New York, NY, USA: ACM. Retrieved from `https://doi.org/10.1145/3391614.3399400`

Chen, S., Guido, R. C., & Chen, S. (2007, Dec). Voice activity detection in car environment using support vector machine and wavelet transform. In *Proc. of the 9th IEEE International Symposium on Multimedia Workshops* (p. 252-255). Retrieved from `http://dx.doi.org/10.1109/ISM.Workshops.2007.50`

Chung, C.-J., & Chen, W.-Y. (2001). Low complexity voice recognition in car environment. In *Proc. of the IEEE 3rd Workshop on Signal Processing Advances in Wireless Communications* (p. 380-383).

Duffield, T. J., & Krupenia, S. (2015). Drivers' interaction preferences in autonomous vehicle multimodal interactive systems. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *59*(1), 1302-1306. Retrieved from `https://doi.org/10.1177/1541931215591213`

Ecker, R., Broy, V., Butz, A., & De Luca, A. (2009). PieTouch: A Direct Touch Gesture Interface for Interacting with in-Vehicle Information Systems. In *Proceedings of the 11th international conference on human-computer interaction with mobile devices and services.* New York, NY, USA: ACM. Retrieved from `https://doi.org/10.1145/1613858.1613887`

Ecker, R., Broy, V., Hertzschuch, K., & Butz, A. (2010). Visual cues supporting direct touch gesture interaction with in-vehicle information systems. In *Proceedings of the 2nd international conference on automotive user interfaces and interactive vehicular applications* (pp. 80–87). New York, NY, USA: ACM. Retrieved from `https://doi.org/10.1145/1969773.1969788`

European Road Safety Observatory. (2015). *Cell phone use while driving.* Retrieved from `https://ec.europa.eu/transport/road_safety/sites/roadsafety/files/erso-synthesis-2015-cellphone-detail_en.pdf`

Fissore, L., Laface, P., & Ruscitti, P. (1992, March). HMM modeling for speaker independent voice dialing in car environment. In *Proc. of the IEEE Int. Conference on Acoustics, Speech, and Signal Processing* (Vol. 1, p. 249-252). Retrieved from `http://dx.doi.org/10.1109/ICASSP.1992.225925`

Fitch, G. M., Bartholomew, P. R., Hanowski, R. J., & Perez, M. A. (2015). Drivers' visual behavior when using handheld and hands-free cell phones. *Journal of Safety Research*, *54*, 105.e29–108. Retrieved from `http://www.sciencedirect.com/science/article/pii/S0022437515000407`

Gable, T. M., May, K. R., & Walker, B. N. (2014). Applying popular usability heuristics to gesture interaction in the vehicle. In *Adjunct Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 1–7). New York, NY, USA: ACM. Retrieved from `http://doi.acm.org/10.1145/2667239.2667298`

Gellatly, A., & Dingus, T. (1998). Speech recognition and automotive applications: Using speech to perform in-vehicle tasks. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *42*, 1247-1251. Retrieved from `http://dx.doi.org/10.1177/154193129804201715`

Gheran, B.-F., Vanderdonckt, J., & Vatavu, R.-D. (2018). Gestures for smart rings: Empirical results, insights, and design implications. In *Proceedings of the 2018 designing interactive systems conference* (pp. 623–635). New York, NY, USA: ACM. Retrieved from `https://doi.org/10.1145/3196709.3196741`

Gheran, B.-F., & Vatavu, R.-D. (2020). From Controls on the Steering Wheel to Controls on the Finger: Using Smart Rings for In-Vehicle Interactions. In *Companion Publication of the 2020 ACM Designing Interactive Systems Conference* (pp. 299–304). New York, NY, USA: ACM. Retrieved from https://doi.org/10.1145/3393914.3395851

Graichen, L., Graichen, M., & Krems, J. (2019). Evaluation of gesture-based in-vehicle interaction: User experience and the potential to reduce driver distraction. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *61*, 774-792. Retrieved from http://dx.doi/org/10.1177/0018720818824253

Hamdani, S. E., Benamar, N., & Younis, M. (2019). A protocol for pedestrian crossing and increased vehicular flow in smart cities. *Journal of Intelligent Transportation Systems*, 1–20. Retrieved from https://doi.org/10.1080/15472450.2019.1683451

Holtl, A., & Trommer, S. (2013). Driver assistance systems for transport system efficiency: Influencing factors on user acceptance. *Journal of Intelligent Transportation Systems*, *17*(3), 245–254. Retrieved from https://doi.org/10.1080/15472450.2012.716646

Huelin, R., Iheanacho, I., Payne, K., & Sandman, K. (2016). *What's in a Name? Systematic and Non-Systematic Literature Reviews, and Why the Distinction Matters.* Retrieved from https://www.evidera.com/wp-content/uploads/2015/06/Whats-in-a-Name-Systematic-and-Non-Systematic-Literature-Reviews-and-Why-the-Distinction-Matters.pdf

Häuslschmid, R., Menrad, B., & Butz, A. (2015). Freehand vs. micro gestures in the car: Driving performance and user experience. In *Proc. of the IEEE Symposium on 3D User Interfaces (3DUI)* (pp. 159–160). Retrieved from http://dx.doi.org/10.1109/3DUI.2015.7131749

Igorevich, R. R., & Min, D. (2015). Smart Car Use Case: Dynamic Reconfigurable IoT Convergence with BigData. In *Proceedings of the 2015 International Conference on Big Data Applications and Services* (pp. 254–258). New York, NY, USA: ACM. Retrieved from https://doi.org/10.1145/2837060.2837108

Jeanne Breen Consulting. (2009). *Car telephone use and road safety, final report.* Retrieved from https://ec.europa.eu/transport/road_safety/sites/roadsafety/files/pdf/car_telephone_use_and_road_safety.pdf

Jeong, H., & Liu, Y. (2018). Cognitive Modeling of Remote-Manual and Voice Controls for In-Vehicle Human-Automation Systems. In *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 135–136). New York, NY, USA: ACM. Retrieved from https://doi.org/10.1145/3173386.3176958

Khan, S., Akmal, H., Ali, I., & Naeem, N. (2017, Nov). Efficient and unique learning of in-car voice control for engineering education. In *Proc. of the International Multi-topic Conference* (p. 1-6).

Khare, A., Ghosh, H., Wattamwar, S. S., Sinha, A., Bhowmick, B., Kumar, K. C., & Kopparapu, S. K. (2009). Multimodal interaction in modern automobiles. In *Proceedings of the 1st International Workshop on Multimodal Interfaces for Automotive Applications (MIAA)* (Vol. 3, pp. 17–20).

Kim, J., Jeong, M., & Lee, S. C. (2019). Why Did This Voice Agent Not Understand Me?: Error Recovery Strategy for in-Vehicle Voice User Interface. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications: Adjunct Proceedings* (pp. 146–150). New York, NY, USA: ACM. Retrieved from https://doi.org/10.1145/3349263.3351513

Kim, K., & Choi, J. (2019, Feb). Passengers' gesture recognition model in self-driving vehicles: Gesture recognition model of the passengers' obstruction of the vision of the driver. In *Proc. of the 2019 ieee 4th international conference on computer and communication systems (icccs)* (p. 239-242). Retrieved from http://dx.doi.org/10.1109/CCOMS.2019.8821649

Kujala, T., & Grahn, H. (2017). Visual distraction effects of in-car text entry methods: Comparing keyboard, handwriting and voice recognition. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 1–10). New York, NY, USA: ACM. Retrieved from https://doi.org/10.1145/3122986.3122987

Laack, A. v., Kirsch, O., Tuzar, G.-D., & Blessing, J. (2016). Controlling vehicle functions with natural body language. In B. Weyers & A. Dittmar (Eds.), *Mensch und Computer 2016 – Workshopband.* Aachen: Gesellschaft für Informatik e.V.

Langlois, S., That, T. N., & Mermillod, P. (2016). Virtual head-up displays for augmented reality in cars: A user testing to validate the congruence. In *Proceedings of the european conference on cognitive ergonomics.* New York, NY, USA: ACM. Retrieved from https://doi.org/10.1145/2970930.2970946

Lansdown, T. C., Brook-Carter, N., & Kersloot, T. (2002). Primary task disruption from multiple in-vehicle systems. *Journal of Intelligent Transportation Systems*, *7*(2), 151–168. Retrieved from https://doi.org/10.1080/10248070190048736

Lazic, S. E. (2010). The problem of pseudoreplication in neuroscientific studies: is it affecting your analysis? *BMC Neuroscience*, *11*(5). Retrieved from https://dx.doi.org/10.1186%2F1471-2202-11-5

Lee, S. H., Yoon, S.-O., & Shin, J. H. (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* (p. 94–99). New York, NY, USA: ACM. Retrieved from https://doi.org/10.1145/2809730.2809739

Leiva, L. A., Martín-Albo, D., Plamondon, R., & Vatavu, R.-D. (2018). KeyTime: Super-Accurate Prediction of Stroke Gesture Production Times. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.* New York, NY, USA: ACM. Retrieved from https://doi.org/10.1145/3173574.3173813

Leiva, L. A., Martín-Albo, D., & Vatavu, R.-D. (2018). GATO: Predicting Human Performance with Multistroke and Multitouch Gesture Input. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services.* New York, NY, USA: ACM. Retrieved from https://doi.org/10.1145/3229434.3229478

Lin, H. (2019). *Using passenger elicitation for developing gesture design guidelines for adjusting highly automated vehicle dynamics.* A Thesis Presented to The Academic Faculty, Georgia Tech Library.

Ma, J., & Du, Y. (2017, Aug). Study on the evaluation method of in-vehicle gesture control. In *Proc. of the 3rd IEEE International Conference on Control Science and Systems Engineering* (p. 145-148). Retrieved from http://dx.doi.org/10.1109/CCSSE.2017.8087912

Mahetalia, J., Mehta, R., Varudandi, S., & Samdani, K. (2018). A comparative study on gesture recognition systems for car infotainment systems. In *Proc. of the IEEE Int. Conf. on System, Computation, Automation and Networking* (p. 1-7). Retrieved from http://dx.doi.org/10.1109/ICSCAN.2018.8541236

Mahr, A., Endres, C., Müller, C., & Schneeberger, T. (2011). Determining human-centered parameters of ergonomic micro-gesture interaction for drivers using the theater approach. In *Proc. of the 3rd Int. Conf. on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 151–158). New York, NY, USA: ACM. Retrieved from https://doi.org/10.1145/2381416.2381441

Manawadu, U. E., Kamezaki, M., Ishikawa, M., Kawano, T., & Sugano, S. (2016, Oct). A hand gesture based driver-vehicle interface to control lateral and longitudinal motions of an autonomous vehicle. In *Proc. of the IEEE International Conference on Systems, Man, and Cybernetics.* Retrieved from http://dx.doi.org/10.1109/SMC.2016.7844497

May, K. R., Gable, T. M., & Walker, B. N. (2014). A multimodal air gesture interface for in vehicle menu navigation. In *Adjunct Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 1–6). New York, NY, USA: ACM. Retrieved from http://doi.acm.org/10.1145/2667239.2667280

May, K. R., Gable, T. M., & Walker, B. N. (2017). Designing an in-vehicle air gesture set using elicitation methods. In *Proceedings of the 9th Int. Conf. on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 74–83). New York, NY, USA: ACM. Retrieved from http://doi.acm.org/10.1145/3122986.3123015

Morris, M. R., Danielescu, A., Drucker, S., Fisher, D., Lee, B., schraefel, m., & Wobbrock, J. O. (2014). Reducing legacy bias in gesture elicitation studies. *interactions*, *21*(3), 40–45. Retrieved from http://doi.acm.org/10.1145/2591689

Neßelrath, R., Moniri, M. M., & Feld, M. (2016, Sep.). Combining speech, gaze, and micro-gestures for the multimodal control of in-car functions. In *Proc. of the 12th International Conference on Intelligent Environments (IE)* (p. 190-193). Retrieved from http://dx.doi.org/10.1109/IE.2016.42

NHTSA. (2019). *Distracted driving.* Traffic Safety Facts, U.S. Departament of Transportation, National Highway Traffic Safety Administration. Retrieved from https://www.nhtsa.gov/risky-driving/distracted-driving

Ohn-Bar, E., Tran, C., & Trivedi, M. (2012). Hand gesture-based visual user interface for infotainment. In *Proc. of the 4th Int. Conf. on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 111–115). New York, NY, USA: ACM. Retrieved from https://doi.org/10.1145/2390256.2390274

Ohn-Bar, E., & Trivedi, M. (2014). Hand gesture recognition in real time for automotive interfaces: A multimodal vision-based approach and evaluations. *IEEE Transactions on Intelligent Transportation Systems*, *15*, 2368-2377. Retrieved from http://dx.doi.org/10.1109/TITS.2014.2337331

Petkar, H. (2016, 10). A review of challenges in automatic speech recognition. *International Journal of Computer Applications*, *151*, 23-26. Retrieved from http://dx.doi.org/10.5120/ijca2016911706

Pfleging, B., Schneegass, S., & Schmidt, A. (2012). Multimodal interaction in the car: Combining speech and gestures on the steering wheel. In *Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 155–162). New York, NY, USA: ACM. Retrieved from http://doi.acm.org/10.1145/2390256.2390282

Qiu, H., Ahmad, F., Govindan, R., Gruteser, M., Bai, F., & Kar, G. (2017). Augmented vehicular reality: Enabling extended vision for future vehicles. In *Proceedings of the 18th International Workshop on Mobile Computing Systems and Applications* (pp. 67–72). New York, NY, USA: ACM. Retrieved from https://doi.org/10.1145/3032970.3032976

Rahman, A. M., Saboune, J., & El Saddik, A. (2011). Motion-path based in car gesture control of the multimedia devices. In *Proceedings of the First ACM International Symposium on Design and Analysis of Intelligent Vehicular Networks and Applications* (pp. 69–76). New York, NY, USA: ACM. Retrieved from https://doi.org/10.1145/2069000.2069013

Raja, M., Ghaderi, V., & Sigg, S. (2018, July). WiBot! In-Vehicle Behaviour and Gesture Recognition Using Wireless Network Edge. In *Proc. of the 2018 ieee 38th international conference on distributed computing systems (icdcs)* (p. 376-387). Retrieved from https://doi.org/10.1109/ICDCS.2018.00045

Reimer, B., Mehler, B., Dobres, J., Mcanulty, H., Mehler, A., Munger, D., & Rumpold, A. (2014, 09). Effects of an 'expert mode' voice command system on task performance, glance behavior and driver physiology. In *Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 1–9). New York, NY, USA: ACM. Retrieved from https://doi.org/10.1145/2667317.2667320

Riener, A., Ferscha, A., Bachmair, F., Hagmüller, P., Lemme, A., Muttenthaler, D., ... Weger, F. (2013). Standardization of the in-car gesture interaction space. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 14–21). New York, NY, USA: ACM. Retrieved from https://doi.org/10.1145/2516540.2516544

Roider, F., & Gross, T. (2018). I see your point: Integrating gaze to enhance pointing gesture accuracy while driving. In *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 351–358). New York, NY, USA: ACM. Retrieved from https://doi.org/10.1145/3239060.3239084

Roider, F., & Raab, K. (2018, June). Implementation and evaluation of peripheral light feedback for mid-air gesture interaction in the car. In *Proc. of the 2018 14th International Conference on Intelligent Environments (IE)* (p. 87-90). Retrieved from http://dx.doi.org/10.1109/IE.2018.00021

Roider, F., Reisig, L., & Gross, T. (2018). Just look: The benefits of gaze-activated voice input in the car. In *Adjunct Proc. of the 10th Int. Conf. on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 210–214). Retrieved from https://doi.org/10.1145/3239092.3265968

Roider, F., Rümelin, S., Pfleging, B., & Gross, T. (2017). The effects of situational demands on gaze, speech and gesture input in the vehicle. In *Proc. of the 9th Int. Conf. on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 94–102). Retrieved from https://doi.org/10.1145/3122986.3122999

Sachara, F., Kopinski, T., Gepperth, A., & Handmann, U. (2017). Free-hand gesture recognition with 3D-CNNs for in-car infotainment control in real-time. In *Proc. of the IEEE 20th Int. Conf. on Intelligent Transportation Systems* (p. 959-964). Retrieved from https://doi.org/10.1109/ITSC.2017.8317684

Sauras, P., Gil, A., Gill, J., Pisu, P., & Taiber, J. (2017). VoGe: A Voice and Gesture System for Interacting with Autonomous Cars. SAE Technical Paper. Retrieved from http://dx.doi.org/10.4271/2017-01-0068

Sauras-Perez, P., Taiber, J., & Smith, J. (2014, Nov). Variability analysis of in-car gesture interaction. In *Proc. of the 2014 International Conference on Connected Vehicles and Expo (ICCVE)* (p. 777-780). Retrieved from http://dx.doi.org/10.1109/ICCVE.2014.7297657

Schipor, O. A., & Vatavu, R.-D. (2019). Towards Interactions with Augmented Reality Systems in Hyper-Connected Cars. In *EICS Workshops* (pp. 76–82). Retrieved from http://ceur-ws.org/Vol-2503/paper1_12.pdf

Shakeri, G., Williamson, J. H., & Brewster, S. (2017). Novel multimodal feedback techniques for in-car mid-air gesture interaction. In *Proc. of the 9th int. conf. on automotive user interfaces and interactive vehicular applications* (pp. 84–93). Retrieved from https://doi.org/10.1145/3122986.3123011

Shakeri, G., Williamson, J. H., & Brewster, S. (2018). May the force be with you: Ultrasound haptic feedback for mid-air gesture interaction in cars. In *Proc. of the 10th Int. Conf. on Automotive User Interfaces and Interactive Vehicular Applications.* Retrieved from https://doi.org/10.1145/3239060.3239081

Shladover, S. E. (2018). Connected and automated vehicle systems: Introduction and overview. *Journal of Intelligent Transportation Systems*, *22*(3), 190–200. Retrieved from https://doi.org/10.1080/15472450.2017.1336053

Smith, K. A., Csech, C., Murdoch, D., & Shaker, G. (2018). Gesture Recognition Using mm-Wave Sensor for Human-Car Interface. *IEEE Sensors Letters*, *2*(2).

Sokol, N., Chen, W., & Donmez, B. (2017). Voice-controlled in-vehicle systems: Effects of voice-recognition accuracy in the presence of background noise. In *Proc. of the 2017 Driving Assessment Conference* (p. 158-164). Retrieved from http://dx.doi.org/10.17077/drivingassessment.1629

Sterkenburg, J., Landry, S., & Jeon, M. (2017). Eyes-free in-vehicle gesture controls: Auditory-only displays reduced visual distraction and workload. In *Proc. of the 9th Int. Conf. on Automotive User Interfaces and Interactive Vehicular Applications Adj.* (pp. 195–200). Retrieved from https://doi.org/10.1145/3131726.3131747

Taherkhani, M. A., Kawaguchi, R., Shirmohammad, N., & Sato, M. (2016). BlueParking: An IoT Based Parking Reservation Service for Smart Cities. In *Proceedings of the Second International Conference on IoT in Urban Space* (pp. 86–88). New York, NY, USA: ACM. Retrieved from https://doi.org/10.1145/2962735.2962739

Tarvekar, M. (2018, June). Hand gesture recognition system for touch-less car interface using multiclass support vector machine. In *Proc. of the 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)* (p. 1929-1932). Retrieved from http://dx.doi.org/10.1109/ICCONS.2018.8663003

Varhelyi, A., Kaufmann, C., Johnsson, C., & Almqvist, S. (2020). Driving with and without automation on the motorway – an observational study. *Journal of Intelligent Transportation Systems*, 1–22. Retrieved from https://doi.org/10.1080/15472450.2020.1738230

Vatavu, R.-D. (2019). The dissimilarity-consensus approach to agreement analysis in gesture elicitation studies. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–13). New York, NY, USA: ACM. Retrieved from https://doi.org/10.1145/3290605.3300454

Vatavu, R.-D., & Pentiuc, S. G. (2008). Multi-level representation of gesture as command for human-computer interaction. *Computing and Informatics*, *27*(6), 837–851. Retrieved from http://dx.doi.org/ojs/index.php/cai/article/viewArticle/16

Vatavu, R.-D., & Wobbrock, J. O. (2015). Formalizing agreement analysis for elicitation studies: New measures, significance test, and toolkit. In *Proc. of the 33rd ACM Conference on Human Factors in Computing Systems* (pp. 1325–1334). Retrieved from http://doi.acm.org/10.1145/2702123.2702223

Vatavu, R.-D., & Wobbrock, J. O. (2016). Between-subjects elicitation studies: Formalization and tool support. In *Proceedings of the 2016 CHI Conference on*

Human Factors in Computing Systems (pp. 3390–3402). New York, NY, USA: ACM. Retrieved from http://doi.acm.org/10.1145/2858036.2858228

Verma, B., & Choudhary, A. (2018). Framework for dynamic hand gesture recognition using grassmann manifold for intelligent vehicles. *IET Intelligent Transport Systems*, *12*(7), 721-729. Retrieved from http://dx.doi.org/10.1049/iet-its.2017.0331

Vieira, J., Sampaio, R., Nascimento, R., Ferreira, J. P., Machado, S., Ribeiro, N., ... Mouta, S. (2016). Gestures while driving: A guessability approach for a surface gestures taxonomy for in-vehicle indirect interaction. In *Proc. of the Human Factors and Ergonomics Society Europe Chapter Annual Conference.* Retrieved from https://www.hfes-europe.org/wp-content/uploads/2015/12/Vieira2016.pdf

Villarreal-Narvaez, S., Vanderdonckt, J., Vatavu, R.-D., & Wobbrock, J. O. (2020). A systematic review of gesture elicitation studies: What can we learn from 216 studies? In *Proc. of the 2020 ACM Designing Interactive Systems Conference* (pp. 855–872). Retrieved from https://doi.org/10.1145/3357236.3395511

Villing, J. (2011). Speech, buttons or both? a comparative study of an in-car dialogue system. In *Adjunct proceedings of the 3rd international conference on automotive user interfaces and interactive vehicular applications salzburg.* Retrieved from https://www.auto-ui.org/11/docs/AUI2011_adjunctproceedings.pdf

Weidner, F., & Broll, W. (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.* New York, NY, USA: ACM. Retrieved from https://doi.org/10.1145/3365610.3365625

WHO. (2018). *Global Status Report on Road Safety.* Retrieved from https://www.who.int/violence_injury_prevention/road_safety_status/2018/en/

Wobbrock, J. O., Aung, H. H., Rothrock, B., & Myers, B. A. (2005). Maximizing the guessability of symbolic input. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems* (pp. 1869–1872). New York, NY, USA: ACM. Retrieved from http://doi.acm.org/10.1145/1056808.1057043

Wobbrock, J. O., Morris, M. R., & Wilson, A. D. (2009). User-defined gestures for surface computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1083–1092). New York, NY, USA: ACM. Retrieved from http://doi.acm.org/10.1145/1518701.1518866

Wu, H., Wang, Y., Liu, J., Qiu, J., & Zhang, X. (2019). User-defined gesture interaction for in-vehicle information systems. *Multimedia Tools and Applications*, *79*, 263–288. Retrieved from http://dx.doi.org/10.1007/s11042-019-08075-1

Wu, S., Gable, T., May, K., Choi, Y., & Walker, B. (2016). Comparison of surface gestures and air gestures for in-vehicle menu navigation. *Archives of Design Research*, *29*, 65. Retrieved from http://dx.doi.org/10.15187/adr.2016.11.29.4.65

Zheng, P., McDonald, M., & Pickering, C. (2008, Oct). Effects of intuitive voice interfaces on driving and in-vehicle task performance. In *Proc. of the 11th Int. IEEE Conference on Intelligent Transportation Systems* (p. 610-615).

Zobl, M., Geiger, M., Schuller, B., Lang, M., & Rigoll, G. (2003, July). A real-time system for hand gesture controlled operation of in-car devices. In *Proc. of the 2003 International Conference on Multimedia and Expo* (Vol. 3). Retrieved from http://dx.doi.org/10.1109/ICME.2003.1221368