User-Defined Voice and Mid-Air Gesture Commands for Maneuver-based Interventions in Automated Vehicles

Henrik Detjen
henrik.detjen@hs-ruhrwest.de
University of Applied Sciences Ruhr West
Bottrop, DE

Sarah Faltaous
sarah.faltaous@uni-due.de
University of Duisburg-Essen
Essen, DE

Stefan Geisler
stefan.geisler@hs-ruhrwest.de
University of Applied Sciences Ruhr West
Bottrop, DE

Stefan Schneegass
stefan.schneegass@uni-due.de
University of Duisburg-Essen
Essen, DE

Figure 1: We investigate voice and gesture control as possibilities for MBI, here: A gesture to stop the car

ABSTRACT

For highly automated vehicles (AVs), new interaction concepts need to be developed. Even in AVs, the driver might want to intervene and override the automation from time to time. To create the possibility of control, we explore vehicle control through maneuver-based interventions (MBI). Thereby, we focus on explicit, contact-less interaction, which could be beneficial in future AV designs, where the driver is not necessarily bound to classical controls. We propose a set of freehand gestures and keywords for voice control derived in a user-centered design process. Further, we discuss properties, applicability and user impressions of both interaction modalities. Voice control seems to be an efficient way to select a maneuver and free-hand gestures could be used, if voice channel is blocked, e.g., through conversation with passengers.

MuC '19, September 8–11, 2019, Hamburg, Germany

ACM ISBN 978-1-4503-7198-8/19/09…$15.00
https://doi.org/10.1145/3340764.3340798

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**.

KEYWORDS

Automotive HMI; UCD; Mid-Air Gestures; Voice Control

ACM Reference Format:
Henrik Detjen, Sarah Faltaous, Stefan Geisler, and Stefan Schneegass. 2019. User-Defined Voice and Mid-Air Gesture Commands for Maneuver-based Interventions in Automated Vehicles. In *Mensch und Computer 2019 (MuC '19), September 8–11, 2019, Hamburg, Germany*. ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3340764.3340798

1 INTRODUCTION

Automation changes the driver-vehicle interaction. The role of a driver seems to disappear. Thus, the human in the car is not entirely involved in the driving task anymore. The vision is that drivers will become passengers in the majority of time and will focus on new activities while driving [11]. Also, the interior design of automated vehicles (AVs) might be completely different from today's cars. Hence, we need to redesign traditional in-car interaction. Even if traditional control elements like haptic interfaces or touch panels might be present in AVs, the driver might not reach them comfortably at all times, e.g., because he has tilt the seat back while watching a movie on his laptop.

When it comes to AVs, people are also sceptical about handing over control to an autopilot [4, 12]. While traveling, the driver might want to intervene and override the automation, e.g., to overtake a truck that blocks the scenic view or because its more fun to control the vehicle oneself. Proving the driver possibilities to control the AV can foster trust and acceptance. But it is unclear how and to which extend drivers want to intervene in future AVs. In consequence, it might be necessary to provide different levels of control.

In this paper, we explore AV control through maneuver-based intervention (MBI) [3]. MBI is an abstraction of vehicle operations to maneuvers like parking or lane changing. Thereby, we focus on explicit, contact-less interaction, which could be beneficial in future AV designs. Here, contact-less interaction with voice or gestures reduces the time to switch between tasks, because it requires no repositioning and offers more comfort accordingly. As proposed by Desmet & Hassenzahl [1], we create a possibility to enhance future user experiences with AVs and increase users' acceptance.

Our main research questions are, from a users perspective:

RQ1. How should voice and gesture commands for MBI look like?

RQ2. To what degree is contact-less interaction feasible for MBI?

To answer the first question, we derive a set of free-hand gestures and keywords for voice control through a user-centered design process. To answer the second question, we analyze the practical feasibility of resulting designs of both interaction modalities by taking execution times and user impressions into account.

2 RELATED WORK

Maneuver-Based Intervention

Our MBI approach is based on the maneuver-based driving concept by Winner and Hakuli [16]. Maneuver-based driving involves certain maneuvers that can be picked by the driver to make decisions on a tactical level, e.g., lane change to the left. Instead of the driver fully overtaking control from the automation for an intervention, the car still handles the execution of these maneuvers on operational level (cf. Figure 2), for example choosing speed and steering angle. Prerequisite of a maneuver interface is a set of maneuvers, a so called a maneuver catalog, through which the driver makes tactical-level decisions. Schreiber [14] developed a driver-centered maneuver catalog with the goal of high expectancy compliance, short input times and few input errors: *Start*, *Turn Right*, *Turn Left*, *Lane Change Right*, *Lane Change Left*, *Straight*, *Hold at Stop-Line*, *Hold on Side-Strip* and *Parking*. This basic set of maneuvers allows a complete driving mission on country roads and highways. Therefore, we use the set for our following investigation. Complex maneuvers, "Overtaking" for instance, can be realized through combination of base maneuvers: *Lane Change Left*, speed adjustment (parameter), *Lane Change Right*.

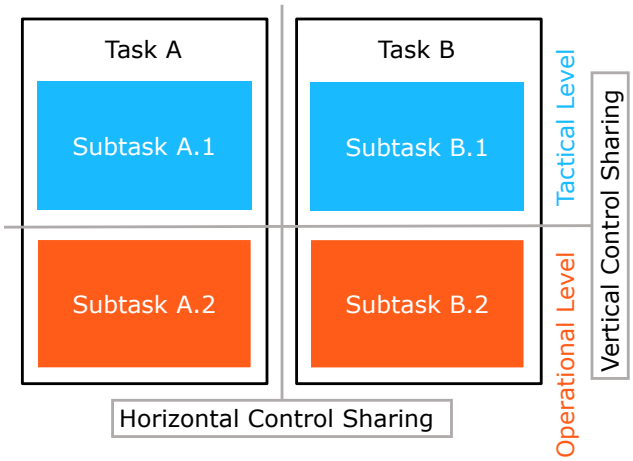

Figure 2: Vertical and Horizontal Control Sharing Between Agents. The automation of driving tasks influences the driver-vehicle interaction concept. Vehicle automation can be done in two directions: Horizontal, between tasks (inter-task sharing) and vertical, between subtasks (intra-task sharing). Inter-task sharing requires cooperation between agents (focus on shared control), while intra-task sharing leads to independent agents (focus on control transitions).

While maneuver-based driving was primary developed to keep the driver in the loop of regulation and control (SAE level 2 and 3 [13]), MBI is meant for higher automation levels (SAE 4 and 5), where the driver is not necessarily involved in the driving task at all times. There, MBI creates the possibility of control. Systems like Hotzenplotz [6] prove that users perceive it, compared to pure automation, as more positive, attractive and less boring. Also, users are more satisfied with MBI than with pure automation, because they feel competent and autonomous. In sum, MBI has the potential to unite the strength of automated driving with user needs and requirements [3]. Nevertheless, these kind of interventions

require a novel type of in-vehicle control since the classical automotive user interfaces are not designed to enter such maneuvers.

User Interfaces for MBI

For maneuver-based driving, Kauer et al. developed a touch screen, which is embedded in the steering wheel, to directly select maneuvers by tapping [8]. Another interface of Franz uses an indirect swipe mechanism on a armrest-touchpad controller, which interacts with a head-up display (display and feedback there) [5]. User interfaces for maneuver-based driving rely on touch and touch gestures to interact fast and efficient, while the driver monitors the environment.

For MBI, user interfaces have been explored little by now. Tscharn et al. [15] examined so-called "non-critical, spontaneous interventions". For their study, they used more complex maneuvers: selection of a parking lot, highway exit, obstacle evasion, and picking up a friend. They compared touch-based and voice-based user interfaces and combined it with mid-air gesture input. Voice or touch for the selection of maneuvers (e.g., parking) and pointing gestures to concrete the maneuvers location (e.g., which parking lot). They found that speech-based interaction was more natural, intuitive, and less stressful than touch-based interaction. Considering this and the pictured change of driver activities and changing novel cockpit designs in the beginning (cf. Section 1), we pick voice and mid-air gestures as interaction modalities for our following investigation. While Tscharn et al. focus on a specific set of maneuvers, the objective of our work is to generate commands for a holistic maneuver catalogue.

3 METHOD

To explore voice and free-hand gesture interaction from a users perspective, we utilized Wobbrock's [17] elicitation method which was originally developed for tabletop gestures. It is a user-centered design approach, where commands are derived directly from the user by showing him the effect of an action and asking him to perform its cause instead of presenting the user a set of predetermined commands devised by someone else. In our case, we show the user a specific maneuver and ask him to perform a fitting gesture (see Figure 1), respectively voice command.

Participants

We invited 20 participants ($m = 14$, $f = 6$) with an average age of 31.5 years ($SD = 13.1$, $Min = 19$, $Max = 61$). Nine participants used voice assistant systems (e.g., Google Search, Smarthome). One participant had experience with free-hand gesture interaction (MS Kinect). After an explanation of the upcoming test scenario (without explicitly mentioning maneuvers), the experiment began.

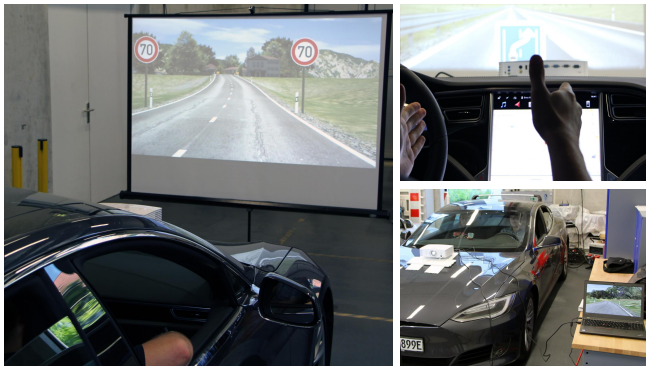

Figure 3: Experimental setup of the user study. We used a Tesla P60 and a large projection screen showing the driving maneuvers.

Setup & Procedure

To create an authentic driving atmosphere, the study was conducted in a stationary vehicle (Tesla P60). We presented a video of an simulator-generated autonomous ride on a canvas in front of the vehicle (see Figure 3). At certain points in the video course, when a maneuver began, we displayed a symbol for ten seconds (adapted symbols from Kauer et al. [8], see Figure 3) and participants reacted with their corresponding voice or gesture command. The video took about eight minutes and covered all driving maneuvers (see Section 2) at least once. The first part (ca. 4 minutes) of the track got users comfortable with the study setup. In the second part (ca. 4 minutes), we recorded the data for our following evaluation with an action cam. Participants drove the track twice in a counterbalanced within-subject design. One time only with gestural responses and one time only with voice responses. At least, after the two runs, the participants completed a custom questionnaire containing questions about acceptance, preferred input style and general feedback.

4 RESULTS

To answer our first research question (How should voice and gesture commands for MBI look like?), we cluster responses from our user study and identify a condensed set of free-hand gesture and voice commands in the following. To cluster the gestural responses systematically, we first categorize them by performance nature, form and handedness. Further, we discuss properties, applicability and user impressions of both interaction modalities to answer our second research question (To what degree is contact-less interaction feasible for MBI?).

Mid-Air Gesture Classification

To define a mid-air gesture set, we classified all gesture video samples first. As main classification dimensions, we used

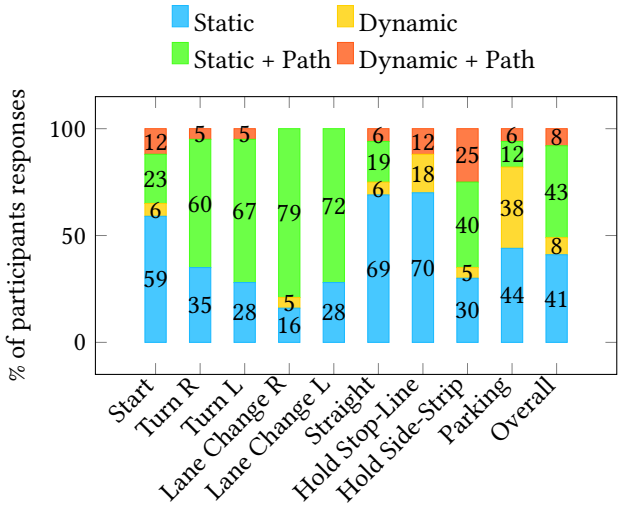

Figure 4: Mid-Air Gesture Form Classification for each Maneuver

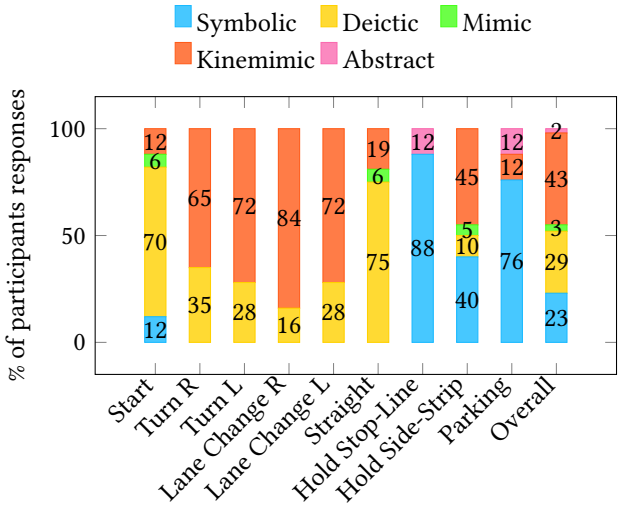

Figure 5: Mid-Air Gesture Nature Classification for each Maneuver

form, nature and handedness. The form dimension describes the performed movement, if a gesture is static or dynamic and whether it is performed along a path [17]. Figure 4 shows the overall distribution of the gestural responses' form. Gestures were performed static or static along a path for most maneuvers.

The nature describes the type of performance. We redefine the nature dimensions from Wobbrock [17] (symbolic, physical, metaphorical, abstract) for the automotive context by integrating Geiger's [7] gesture taxonomy categories. Our final nature categories are:

(1) symbolic: visual depictions
(2) deictic: pointing gestures, special case of symbolic
(3) metaphorical/mimic: gestures acting on, with or like something else
(4) kinemimic: gestures imitating a movement, special case of metaphorical/mimic
(5) abstract: not fitting in one of the categories before

Figure 5 shows the overall distribution of the performed gestural responses' form. For maneuvers *Start* and *Straight*, most users performed a deictic gesture, for maneuvers *Turn* and *Lane Change* they commonly chose a kinemimic gesture, and for maneuvers *Hold at Side-Strip*, *Hold at Stop-Line* and *Parking* they used symbolic gestures.

In terms of handedness, we made the following observations: For the dichotomous maneuvers (*Turn L/R*, *Lane Change L/R*), nearly every participant mirrored his mid-air gesture in both directions with the right hand. This is surprising, because the left shoulder joint is more flexible to the right side. When both hands were used, one hand copied the other, e.g. for stop maneuvers, to intensify the commands

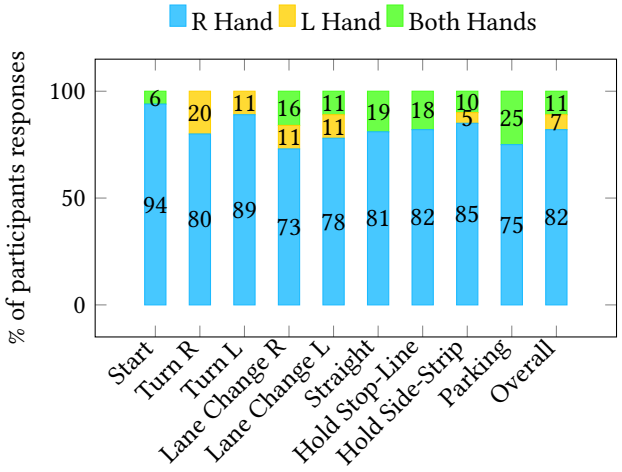

Figure 6: Mid-Air Gesture Handedness for each Maneuver

urgency. The number of fingers used, depends on the nature of the gesture. While pointing gestures were performed with one finger, swipe gestures were performed with all fingers or simplified, with 2 fingers (index and middle finger). Figure 6 shows the overall distribution of participants gestural responses' handedness. In sum, the the right hand was used in the majority of cases, while the left hand or both hands were used infrequently.

Table 1: User Agreement Score for each Maneuver

Maneuver	A_{Speech}	$A_{Gesture}$
Start	.19	.31
Turn R	.81	.48
Turn L	.81	.46
Lane Change R	.22	.45
Lane Change L	.17	.41
Straight	.65	.43
Hold Stop-Line	.68	.61
Hold Side-Strip	.15	.17
Parking	1	.262
Overall	.51	.40

User Defined Voice and Mid-Air Gesture Command Set

After the systematic classification of gestures, we clustered similar voice and gesture responses for each maneuver. Afterwards, we assigned these clusters to our proposed command sets.

User Agreement on Commands. To measure the degree of consensus between users, we calculated the clustering agreement score [17] (see Table 1). Larger clusters lead to higher agreement scores. Our participants showed high consensus for voice and gesture commands on most maneuvers (Voice: $M = .51$, $Min = .15$, $Max = 1$; Gestures: $M = .4$, $Min = .17$, $Max = .61$).

Mapping of Clusters to a Command Set. To generate a voice and mid-air gesture command set, we mapped the found clusters of user responses to our command set for each maneuver. We only used clusters with $n >= 3$ for our mapping and thereby excluded single responses and small clusters, which might have been coincidence. Despite of this threshold, the coverage of our final mapping in relation to all responses is high. The mapped voice command set includes 90% of all samples and the gesture command set covers 86% of all samples. Figure 7 shows the suggested mapping for gesture commands, Table 2 shows the suggested mapping for voice commands.

Execution Times

The execution time is important, because it determines, if the interaction is fast enough for real applications. An example: When we drive 100 km/h (27.8 m/s) on a highway and decide to leave the highway 100m before an exit, the available time period for a commands performance and system-side recognition, processing and execution is 3.6 seconds combined. To estimate how long an interaction takes, we measured the execution time for each participants response. For the voice

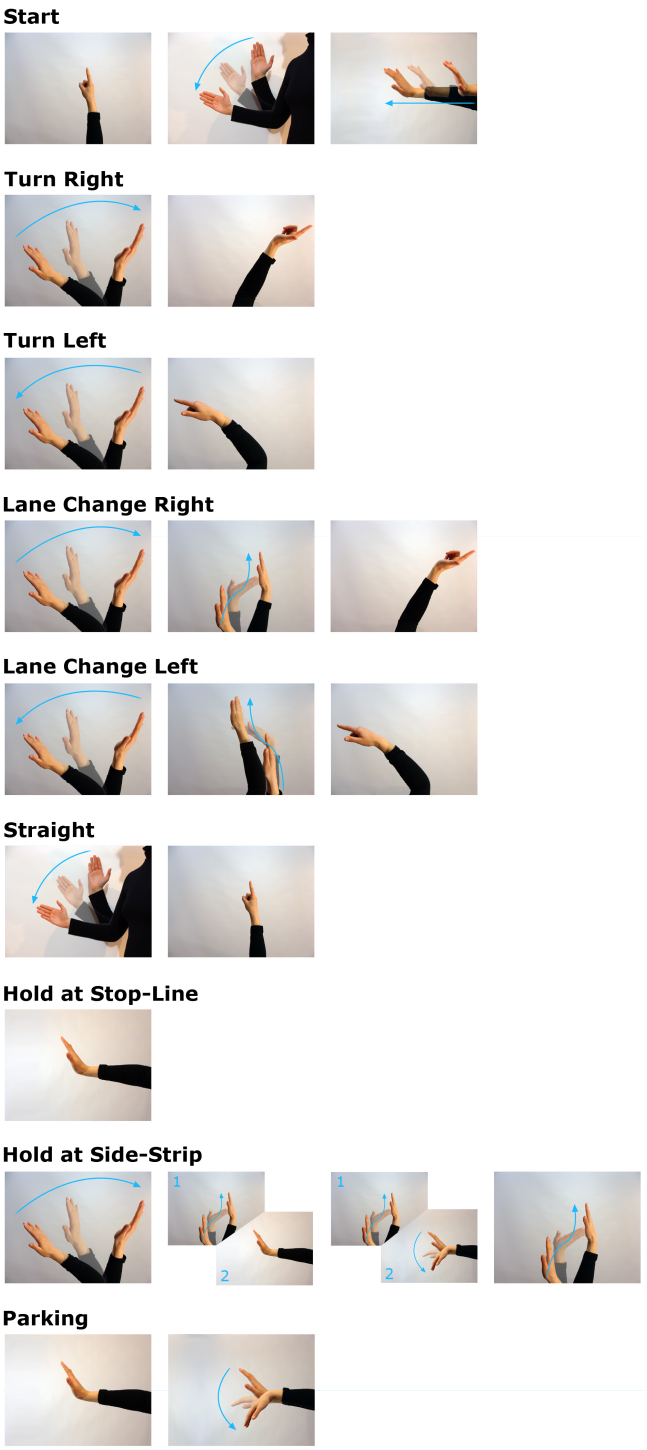

Figure 7: User-Defined Mid-air Gesture Commands Generated in the Study, from bird perspective or from side-perspective for better view of up-down movement and hands

Table 2: User-defined voice command set generated in the study

Maneuver	Keyphrases (in EBNF)
Start	set off, start [(drive | straight ahead)], begin, straight, go [straight]
Turn R	turn right
Turn L	turn left
Lane Change R	[(move to | select)] right (lane | track), lane change [to the] right, [keep] right [ahead]
Lane Change L	veer [to the] left [and continue], [on] left lane, lane change [to the] left, (pass | overtake | pull ahead) left, [drive] left
Straight	[(go | follow the road)] straight, continue [driving] straight [on]
Hold Stop-Line	stop [at the line], hold [independently]
Hold Side-Strip	stop [right], (pull | right) over, hold [right [on the edge of the roadway]]
Parking	park

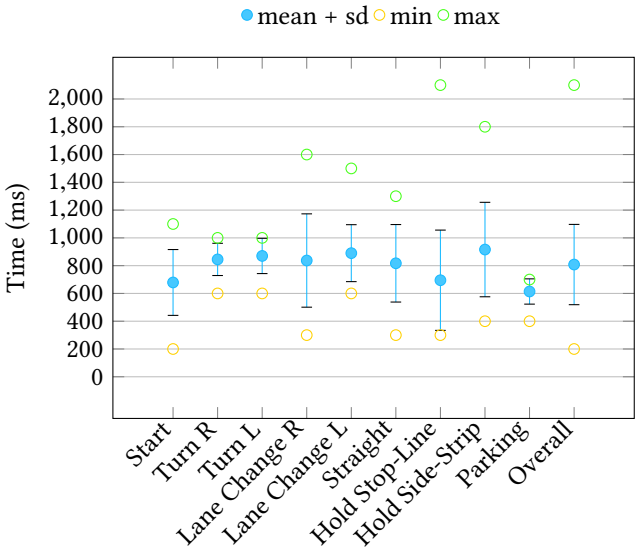

Figure 8: Execution Times for Voice Control

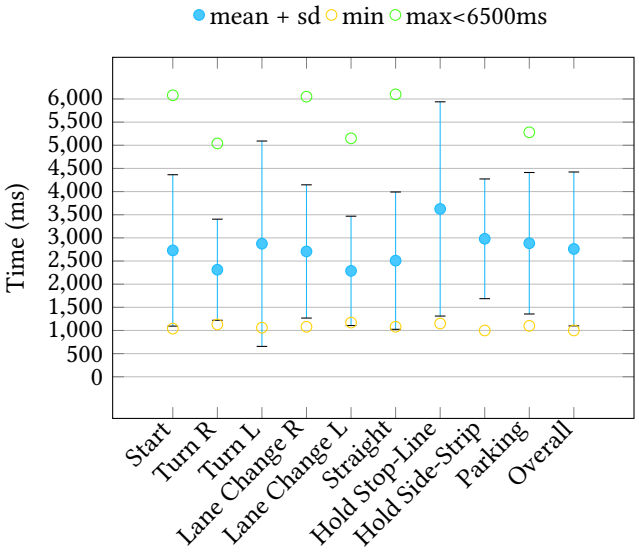

Figure 9: Execution Times for Mid-Air Gesture Control

commands, we translated the commands from German to English. Next, we used the Google Cloud Text-to-Speech API to create audio files (voice: "en-US-Wavenet-D") for the commands. Afterwards, we sent the audio files to the Google Cloud Speech-to-Text API to get the duration of each command. For mid-air gestures, the execution time is defined from start of a motion (muscle contraction) till end of this motion (muscle relaxation). The observed length of the command varies depending on the maneuver, as shown in Figure 8 and Figure 9. Some static gestures were performed for the whole maneuver length (~10s). For instance, for *Hold at Stop-Line*, participants did not release their gesture, until the car actually stopped. Overall, voice commands ($M = 0.88s$, $SD = 0.29s$, $Min = 0.2s$, $Max = 2.1s$) can be performed significantly ($t(153) = 1,98, p < .01$) faster than mid-air gestures ($M = 2.76s$, $SD = 1.66s$, $Min = 1.0s$, $Max = 10.1s$). Summarized for our interaction scenario above, voice control is sufficiently fast, but gesture control could be too slow in some cases.

Acceptance & Preferences

The driver's acceptance of speech and gesture interfaces is a critical factor. Previous work [2] shows that people could imagine using voice control for maneuver-based driving ("You can also tell a person the way"), at least to a limited extent ("outside noise" and "conversations"), while the possible use of gesture control was more ambiguous ("unclear gestures", "acceptance", "freedom of movement"). In our video study, this judgments have been confirmed. On a 6-point likert-scale ($Min = 1$, $Max = 6$), the voice control was rated good ($M = 4.65$, $SD = 1.11$) and free-hand gesture control rather bad ($M = 3.3$, $SD = 1.15$). Consequently, 14 out of

20 participants (70%) preferred voice control over gesture control for MBI.

5 DISCUSSION AND FUTURE WORK

Method: The user-centered design process we used is fundamentally different from methods like Nielsen's expert-based design for gestures [9] and has been topic of debates (activity-centered design vs user-centered design, c.f. [10]). Though the final system needs to be defined by experts, the results of this study will support this process by giving valuable insights about user preferences and mental models.

A limitation of this study is that derived command sets are bound to participants' culture, language and car properties, especially space and technology. Future studies should use a more representative sample, and discuss the need for user configuration.

Mapping & Mental Models of Users: Participants abbreviated the maneuvers *Lane Change* and *Turn* by uttering only the direction ("right", "left") and performed similar gestures for *Turn L/R-Lane Change L/R*, *Start-Straight*, and *Stop-Parking*. Hence, this commands overlap. At first glance, the overlapping of commands seems problematic. A command should be unambiguous and not result in different maneuvers. In practice, however, none of the maneuvers with overlapping commands can occur simultaneously. By including the context, a system can distinguish between them at any time, for example "right [turn]" and "right [lane change]", assumed that the vehicle is crossing one line at once. With this extension, the mapping is conflict free. Further, this simplification of the provided maneuvers made by the users, indicates that their mental model of maneuvering is not fitting to the maneuver catalogue. We recommend further research on users mental models for MBI, because the user interface might be drastically reduced to a few commands on input side: **Straight** (*Straight*, *Start*), **Stop** (*Hold at Side-Strip*, *Hold at Stop-Line*), **Left** (*Turn L*, *Lane Change L*), **Right** (*Turn R*, *Lane Change R*), and **Parking** (*Parking*).

6 CONCLUSION

We examined, how drivers would interact with AVs based on MBI using mid-air gesture and voice control as primary selection mechanism. The study was conducted in a real car to improve the participants' impression of the available space and sound characteristics. Both, voice and gesture interaction are suitable for MBI in different ways. Voice control is faster, has higher acceptance and is preferred by most users. Thus, our results show that voice might be used as a primary input for maneuvers selection. From participants' recorded responses, we proposed a set of voice and mid-air gesture commands for MBI. This command sets can be used as a foundation for future MBI system design.

REFERENCES

[1] Pieter Desmet and Marc Hassenzahl. 2012. Towards happiness: Possibility-driven design. In *Human-computer interaction: The agency perspective*. Springer, 3–27. https://doi.org/10.1007/978-3-642-25691-2_1

[2] Henrik Detjen, Stefan Geisler, Maurizio Salini, Martin Wozniak, and Colja Borgmann. 2018. Teilautomatisiertes Fahren via Sprachsteuerung: Erwartungen und Anforderungen. In *Mensch und Computer 2018 - Workshopband*, Raimund Dachselt and Gerhard Weber (Eds.). Gesellschaft für Informatik e.V., Bonn. https://doi.org/10.18420/muc2018-ws15-0455

[3] Henrik Detjen, Stefan Schneegass, and Stefan Geisler. 2019. Maneuver-based Driving for Intervention in Autonomous Cars. In *CHI'19 Workshop on "Looking into the Future: Weaving the Threads of Vehicle Automation"*. ACM.

[4] Sabrina C. Eimler and Stefan Geisler. 2015. Zur Akzeptanz Autonomen Fahrens – Eine A-Priori Studie. In *Mensch und Computer 2015 – Workshopband*, Anette Weisbecker, Michael Burmester, and Albrecht Schmidt (Eds.). De Gruyter Oldenbourg, Berlin, 533–540.

[5] Benjamin Franz. 2014. *Entwicklung und Evaluation eines Interaktionskonzepts zur manöverbasierten Führung von Fahrzeugen.* Ph.D. Dissertation. Technische Universität Darmstadt. https://doi.org/10.13140/RG.2.1.1918.5364

[6] Anna-Katharina Frison, Philipp Wintersberger, Andreas Riener, and Clemens Schartmüller. 2017. Driving Hotzenplotz: A Hybrid Interface for Vehicle Control Aiming to Maximize Pleasure in Highway Driving. In *Proceedings of the 9th international conference on automotive user interfaces and interactive vehicular applications*. ACM, 236–244. https://doi.org/10.1145/3122986.3123016

[7] Michael Geiger. 2003. *Berührungslose Bedienung von Infotainment-Systemen im Fahrzeug.* Ph.D. Dissertation. Technische Universität München.

[8] Michaela Kauer, Michael Schreiber, and Ralph Bruder. 2010. How to conduct a car? A design example for maneuver based driver-vehicle interaction. In *Intelligent Vehicles Symposium (IV), 2010 IEEE*. IEEE, 1214–1221. https://doi.org/10.1109/IVS.2010.5548099

[9] Michael Nielsen, Moritz Störring, Thomas B Moeslund, and Erik Granum. 2003. A procedure for developing intuitive and ergonomic gesture interfaces for HCI. In *International gesture workshop*. Springer, 409–420. https://doi.org/10.1007/978-3-540-24598-8_38

[10] Donald A Norman. 2005. Human-centered design considered harmful. *interactions* 12, 4 (2005), 14–19. https://doi.org/10.1145/1070960.1070976

[11] Bastian Pfleging, Maurice Rang, and Nora Broy. 2016. Investigating user needs for non-driving-related activities during automated driving. In *Proc. of the 15th international conference on mobile and ubiquitous multimedia*. ACM, 91–99. https://doi.org/10.1145/3012709.3012735

[12] Christina Rödel, Susanne Stadler, Alexander Meschtscherjakov, and Manfred Tscheligi. 2014. Towards autonomous cars: the effect of autonomy levels on acceptance and user experience. In *Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM, 1–8. https://doi.org/10.1007/978-3-319-91806-8_19

[13] SAE Standard J3016 2018. *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles.* Standard. SAE, USA.

[14] Michael Schreiber. 2012. *Konzeptionierung und Evaluierung eines Ansatzes zu einer manöverbasierten Fahrzeugführung im Nutzungskontext Autobahnfahrten.* Ph.D. Dissertation. Technische Universität Darmstadt.

[15] Robert Tscharn, Marc Erich Latoschik, Diana Löffler, and Jörn Hurtienne. 2017. 'Stop over there': natural gesture and speech interaction for non-critical spontaneous intervention in autonomous driving. In *Proc. of the 19th ACM International Conference on Multimodal Interaction*. ACM, 91–100. https://doi.org/10.1145/3136755.3136787

[16] Hermann Winner and Stephan Hakuli. 2006. Conduct-by-wire–following a new paradigm for driving into the future. In *Proc. of FISITA world automotive congress*, Vol. 22. Citeseer, 27.

[17] Jacob O Wobbrock, Meredith Ringel Morris, and Andrew D Wilson. 2009. User-defined gestures for surface computing. In *Proc. of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1083–1092. https://doi.org/10.1145/1518701.1518866