CORA, a Prototype for a Cooperative Speech-Based On-Demand Intersection Assistant

Martin Heckmann
Honda Research Institute
Europe GmbH
D-63073 Offenbach, Germany
martin.heckmann@honda-ri.de

Nico Steinhardt
Honda Research Institute
Europe GmbH
D-63073 Offenbach, Germany
nico.steinhardt@honda-ri.de

Dennis Orth
Ruhr-University Bochum
Faculty of Electrical Engineering
and Information Technology
D-44780 Bochum, Germany
dennis.orth@rub.de

Bram Bolder
Honda Research Institute
Europe GmbH
D-63073 Offenbach, Germany
bram.bolder@honda-ri.de

Mark Dunn
Honda Research Institute
Europe GmbH
D-63073 Offenbach, Germany
mark.dunn@honda-ri.de

Dorothea Kolossa
Ruhr-University Bochum
Faculty of Electrical Engineering
and Information Technology
D-44780 Bochum, Germany
dorothea.kolossa@rub.de

AutomotiveUI '19 Adjunct , September 21–25, 2019, Utrecht, Netherlands

ACM ISBN 978-1-4503-6920-6/19/09.
https://doi.org/10.1145/3349263.3349599

Abstract
We present the first speech-based advanced driver assistance prototype. It is based on our previously proposed on-demand communication concept for the interaction between the driver and his or her vehicle. Using this concept, drivers can flexibly activate the system via speech whenever they want to receive assistance. We could show via driver simulator studies that an instantiation of this concept as an intersection assistant, supporting the driver in turning left, was well received by drivers and preferred to an alternative, vision-based system. In this paper, we present the prototype implementation and give details on how we adapted it to the intricacy of urban traffic as well as to the shortcomings of current sensor technology in establishing an adequate environment perception. The accompanying video gives an impression of the interaction between the driver and the system when cooperatively turning left from a subordinate road into crossing traffic.

Author Keywords
Advanced driver assistance system (ADAS), cooperative driving, assistance on demand, speech control, prototype.

CCS Concepts
•Human-centered computing → Natural language interfaces;

Introduction

Motivated by the desire to overcome acceptance barriers of Advanced Driver Assistance Systems (ADAS), e.g., caused by occasional false alarms [1], we previously proposed the Assistance On Demand (AOD) concept and implemented it in a speech-based left-turn assistant [16, 17]. This concept allows the driver to flexibly engage in an interaction with the ADAS. When the driver encounters a challenging situation and wishes for assistance, she can communicate this to the system via speech. We derived the interaction concept for the left-turn assistant from observations of drivers' natural behaviour: When turning left at an unsignalized intersection, drivers often use the opportunity to ask a present front seat passenger for support, e.g. by transferring to them the task of monitoring the traffic from the right, and requesting feedback about suitable gaps in traffic to enter the intersection. The AOD concept transfers this human-human cooperation to a human-system-cooperation with the system acting as a co-passenger which is giving support when requested (i.e. only "on demand") and which is communicating with the driver via speech.

Such a cooperative interaction approach had previously been proposed for the less challenging task of turning left in oncoming traffic [1] and, more recently, for the hand-over from autonomous driving to manual control [6]. The scenario we propose, turning left from a subordinate road, entails a clearer role assignment and hence emphasizes the cooperative aspect more prominently. Additionally, according to subjective reports and to objective workload measures, turning left at an unsignalized urban intersection from a subordinate road into crossing traffic with high traffic density is one of the most challenging tasks for drivers [5, 18].

Our evaluation of the AOD concept in a driving simulator showed that drivers highly appreciate the concept and prefer it to an alternative, vision-based interaction. In a subsequent evaluation we personalized the system to the drivers. For each driver we determined her personally preferred minimal gap in traffic which allows her to merge into the crossing traffic [10]. Using these for personalizing the system recommendations significantly further improved the acceptance of the system [11]. Subsequently, we implemented the AOD concept in a prototype vehicle [9] and evaluated it in urban traffic [12].

In this paper, we present how we adapted the interaction concept from the simulator to its application in real traffic. We discuss the modifications we made, owing to the uncertainties in the perception and the complex behavior of the traffic participants. The accompanying video shows the system in action while supporting a driver in turning left from a subordinate road into crossing traffic

Interaction Concept

In the following, we will first outline the general concept of the interaction. Next we will present the main attributes of the resulting situated dialog model.

Speech as Sole Interaction Modality

We decided to solely rely on speech as the interaction modality. This allows the driver to fully focus her visual attention on the actual driving task. In addition to speech [14, 15, 4], in particular auditory [15] and tactile information [13, 3, 7, 8] have already been investigated quite thoroughly as alternatives to visual communication in the automotive context. We decided for speech-based communication as we expected it to be perceived as very intuitive by the driver and because it allows to also communicate more complex concepts. Additionally, this purely speech-based interaction mimicks the human-human interaction mentioned before. To increase the naturalness of the interaction, in our im-

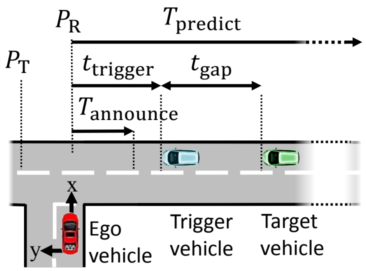

Figure 1: Top view of a T-intersection, showing the ego vehicle (red) waiting for the vehicles from the right to pass the intersection.

plementation the driver does not need to push a button but can directly address the system with a wake-up word (in our case "CORA" as an acronym for *COpeartive Assistant*) followed by her request (e.g., "CORA, please watch right.").

Situated Dialog for Cooperative Driving

When adapting our interaction concept from the simulator to the prototype vehicle, we quickly realized that this required a much more complex dialog model. The situations we encountered (e.g., vehicles stopping behind a bus at a bus stop or vehicles stopping in the center of the road to prepare a left-turn into the road where we were standing) were far richer than those we had designed in the simulator. Additionally, specific features of the objects, such as type (car, truck, pedestrian ...), size or color are not reliably available from the environment sensors. This only allows for referring to the traffic objects as "vehicles". Hence, timing of the system's announcements is key in structuring the dialog. As a consequence of the high temporal dynamics of the environment, the traffic situation changes while the system makes an announcement. This means that the announcements of the system have to be directed towards a predicted future state of the environment. Furthermore, the prediction not only needs to take into account the time needed to output the utterance, but also the time the driver needs to understand it, verify the traffic situation and take a decision. Based on these observations, we introduce the concepts of a *trigger vehicle*, which triggers the evaluation of the situation by the system, and a *target vehicle*, to which the announcements of the system are targeted. We assume that the driver, if no additional information is presented, will associate the system's reference to the vehicle with the shortest time of arrival to the ego vehicle. This is what we call the trigger vehicle (compare Fig. 1). To avoid confusion of the driver and prevent an overly verbose dialog, each trigger vehicle can typically only trigger one verbal reference (an announcement) of the system. When the trigger vehicle has passed to the left side, we search for a new trigger vehicle.

General Dialog Logic

We will now illustrate the dialog logic in more detail based on two situations. During the approach of the trigger vehicle, several conditions are tested, which can trigger an announcement of the system. The reference point P_R for the calculation of the time of arrival is in the center of the ego vehicle (compare Fig. 1). We call this time of arrival t_{trigger} for the trigger vehicle. Consequently, we denote the time gap between the trigger and the target vehicle as t_{gap} and measure it from the front of the trigger vehicle to the front of the target vehicle. We accumulate the time the system needs to output the utterance, and the time the driver needs to understand it, verify the traffic situation and take a decision in T_{announce} (for the prototype we set $T_{\text{announce}} = 2.5\,\text{s}$). Consequently, the system should start announcing a fitting gap at the latest when $t_{\text{trigger}} = T_{\text{announce}}$. Making the announcement earlier increases the uncertainties in the prediction. A gap just large enough for a specific driver is the so-called critical gap $T_{\text{gap crit}}$ [10, 2]. Hence, the system should announce a gap if $t_{\text{gap}} \geq T_{\text{gap crit}}$.

Assuming the trigger vehicle is still sufficiently far away to make an announcement, i.e., $t_{\text{trigger}} \geq T_{\text{announce}}$, and the gap behind this vehicle is large enough for the driver to make the turn, i.e., $t_{\text{gap}} \geq T_{\text{gap crit}}$, the system should announce “gap after next vehicle.” In this utterance, "next vehicle" refers to the trigger vehicle which triggered the announcement and the "gap" is related to the target vehicle, the distance of which determines if the driver can make the turn. In case of $t_{\text{gap}} < T_{\text{gap crit}}$ the system will announce "vehicle from the right". To guarantee that the driver will relate this announcement to the target vehicle, the announce-

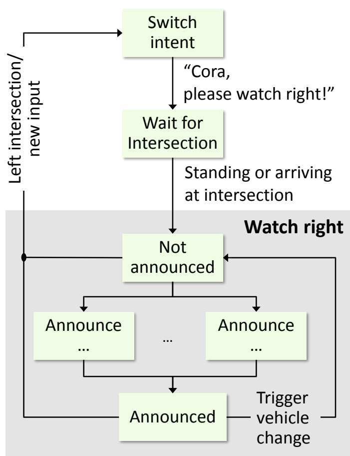

Figure 2: Overview of the dialog manager state machine.

Announcements
Okay, I'm watching.
Okay, I will watch.
No vehicle from the right.
Still no vehicle from the right.
Vehicle in the distance.
Vehicle from the right.
Still vehicle from the right.
No vehicle after the next vehicle.
No vehicle after the next two vehicles.
Gap after the next vehicle.
Gap after the next two vehicles.

Table 1: Set of announcements of the system.

ment is only made once the trigger vehicle has reached the trigger point P_{T}, leaving the target vehicle as the closest vehicle from the right (P_{T} is left of the reference point as we track the front of the vehicles but want the rear to have passed the intersection). No furhter announcements will be made for this trigger vehicle. After it has passed the trigger point, it is removed from the environment representation and the vehicle which is closest at that time is chosen as the new trigger vehicle.

Dialog State Machine
We will now give a short overview on the dialog state machine and the announcements the system makes. More details can be found in [9]. The dialog state machine is organized in two hierarchical levels (compare Fig. 2). On the top level, the system continuously waits for the wake-up word. Once the wake-up word is detected, the system interprets the user intent. If the user wants the system to watch the right side, the system transits to *Wait for intersection*. It remains there until the arrival at the intersection is detected. After that or if the ego vehicle was already standing at the intersection, it transits to the *Watch right* state. This state is subdivided into several sub-states. The system remains in this state until it detects the departure from the intersection or the driver canceled the interaction via speech.

In the *Not announced* state of the *Watch right* state the system determines the current trigger vehicle and continuously checks if the conditions to make an announcement are fulfilled. If this is the case, it transits in the corresponding announcement state, makes the announcement and continues to *Announced*. There it waits until until the current trigger vehicle crosses the trigger point. Then a switch to the new trigger vehicle takes place and the system transits back to *Not announced* and the cycle restarts. Table 1 shows a list of all announcements the system makes.

System Overview
So far we only presented the dialog model and its implementation in the dialog manager. In this section we give a short overview of the remaining building blocks of the prototype system. More details can be found in [9].

System Hardware and Software
We used a modified 2012 model-year Honda CR-V with automatic transmission to implement the prototype system. In addition to the standard equipment, it features 6 Ibeo Automotive Systems LUX LIDAR sensors, which have an overall coverage of 360°. The trunk contains computing hardware to store and process the sensor data. The vehicle CAN bus provides access to the ego vehicle's current velocity, acceleration and the state of the accelerator as well as the brake pedal at a frame rate of $f_{\mathsf{CAN}} = 50\,\mathrm{Hz}$. We acquire the speech commands from the driver via an XCore 7-channel microphone array. For speech feedback we use the standard audio equipment of the vehicle. Speech enhancement, recognition and synthesis are accomplished via the SSE™, VoCon™ and Vocalizer™ software respectively, all from Nuance.

Scene understanding
The LIDAR sensors provide estimates of the position and velocity of all traffic participants in the vicinity of the ego vehicle. However, the vehicles passing from the left side frequently occlude the vehicles approaching from the right side. To mitigate this, we extrapolate the track of a vehicle whose track was lost (see [9] for more details). Based on these extrapolated vehicle tracks, we estimate the gaps between vehicles from the right side. The AOD system should only give feedback on the current traffic situation to the driver while the ego vehicle is standing at an intersection. Therefore, in [9] we implemented methods to detect the arrival and departure of the ego vehicle from the intersection.

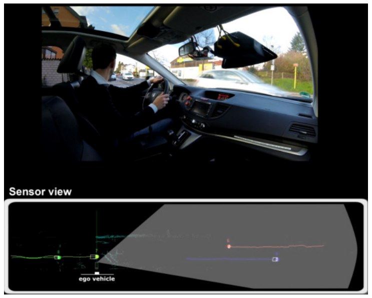

Figure 3: Screenshot of accompanying video.

Content of the Video

The accompanying video shows a prototypical interaction between the driver and the system (the video can be accessed here https://www.ruhr-uni-bochum.de/ika/forschung/forschungsbereich_kolossa/forsch_kolossa_engl.htm). In the video the interior of the vehicle, including the driver, and large parts of the outside traffic scene, left of and in front of the vehicle, can be seen (compare Fig. 3). The lower part of the video, depicted as *Sensor view*, shows a visualization of the traffic objects and their traces as perceived by the LIDAR scanner (including the additional post-processing mentioned above). In this view the position of the ego vehicle is marked and its predicted trajectory, based on the angle of the steering wheel, is shown via a line. The driver activates the system once she has stopped at the intersection. The activation is visualized by highlighting the region of the traffic scene attended to by the system. The system gives feedback on the traffic scene on the right side which allows the driver to concentrate her attention on the left side. Once the system announces a potential occasion to make the turn, the driver verifies the system's announcement via a glance to the right and then starts to turn. Based on the accelerator pedal position the system detects that the driver starts the turn and consequently stops providing information to the driver.

Conclusion

In this paper, we presented a prototype implementation of our Assistance On Demand concept which allows the driver to flexibly request assistance from the ADAS via speech. We implemented the concept as a left-turn assistant hosted on a production Honda CR-V equipped with additional sensing and computing hardware. During the prototype development we made numerous recordings in urban traffic. These recordings guided the development of the situated dialog model which implements the interaction between the driver and the ADAS. Owing to the multitude of different situations we encountered during the recordings, each calling for different information to be transmitted to the driver in order to be of maximum use to her, this dialog model features a comprehensive set of different states. To allow the driver to integrate the information in her decision process, the system needs to inform the driver on predicted future traffic states. We previously evaluated the adequateness of the predictions of the system using a dataset of 115 vehicles passing from the right [12]. Despite the quite challenging scenario with respect to the perception of the vehicles and the establishment of an environment representation, the system made correct predictions in more than 90 % of the cases. The development of the prototype sets the foundation for a formal evaluation of the benefits of the system to the driver in real traffic in a user study. Our current observations are based on formal evaluations of the system in a driving simulator [16, 11] and informal evaluations based on the subjective impressions of a few people driving with the system (all associates of Honda Research Institute). These subjective impressions indicate that the benefits of the system do not only transfer from the simulator to the real world but might even be more pronounced as the cognitive demand in real traffic is much higher for these situations than in the simulator. The simulator studies already showed that the usefulness of the system was rated the higher the more complex the situation was [16].

Acknowledgments

We want to thank Heiko Wersing, Nadja Schömig and Martina Hasenjäger for fruitful discussions and support in our experiments.

REFERENCES

1. Lars Biester. 2009. *Cooperative automation in automobiles.* Ph.D. Dissertation. Humboldt-Universität zu Berlin. DOI: `http://dx.doi.org/10.18452/15935`

2. Werner Brilon, Ralph Koenig, and Rod J. Troutbeck. 1999. Useful estimation procedures for critical gaps. *Elsevier Transp. Res. Part A Policy Pract.* 33 (1999), 161–186. DOI: http://dx.doi.org/10.1016/S0965-8564(98)00048-2
3. Gregory M. Fitch, Jonathan M. Hankey, Brian M. Kleiner, and Thomas A. Dingus. 2011. Driver comprehension of multiple haptic seat alerts intended for use in an integrated collision avoidance system. *Transp. Res. Part F Traffic Psychol. Behav.* 14, 4 (2011), 278–290. DOI: http://dx.doi.org/10.1016/j.trf.2011.02.001
4. Christiane Glatz, Stas Krupenia, Heinrich H Bülthoff, and Lewis L Chuang. 2018. Use the Right Sound for the Right Job: Verbal Commands and Auditory Icons for a Task-Management System Favor Different Information Processes in the Brain. In *Proc. 2018 CHI Conf. Hum. Factors Comput. Syst.* 472:1–13. DOI: http://dx.doi.org/10.1145/3173574.3174046
5. P A Hancock, G Wulf, D Thom, and P Fassnacht. 1990. Driver workload during differeing driving maneouvres. *Accid. Anal. Prev.* 22, 3 (1990), 281–290. DOI: http://dx.doi.org/10.1016/0001-4575(90)90019-H
6. B Karakaya, L Kalb, and K Bengler. 2018. Cooperative Approach to Overcome Automation Effects During the Transition Phase of Conditional Automated Vehicles. In *12. Work. Fahrerassistenzsysteme und Autom. Fahr. FAS 2018*.
7. Matti Krüger, Heiko Wersing, and Christiane B. Wiebel-Herboth. 2018. Approach for Enhancing the Perception and Prediction of Traffic Dynamics with a Tactile Interface. In *Adjun. Proc. 10th Int. Conf. Automot. User Interfaces Interact. Veh. Appl.* 164–169. DOI: http://dx.doi.org/10.1145/3239092.3265961
8. Christian Maag, Norbert Schneider, Thomas Lübbeke, Thomas H. Weisswange, and Christian Goerick. 2015. Car Gestures - Advisory warning using additional steering wheel angles. *Accid. Anal. Prev.* 83 (oct 2015), 143–153. DOI: http://dx.doi.org/10.1016/J.AAP.2015.07.020
9. Dennis Orth, Bram Bolder, Nico Steinhardt, Mark Dunn, Dorothea Kolossa, and Martin Heckmann. 2018. A Speech-based On-demand Intersection Assistant Prototype. In *IEEE Intell. Veh. Symp. Proc.* DOI:http://dx.doi.org/10.1109/IVS.2018.8500402
10. Dennis Orth, Dorothea Kolossa, Milton Sarria Paja, Kersten Schaller, Andreas Pech, and Martin Heckmann. 2017a. A maximum likelihood method for driver-specific critical-gap estimation. In *IEEE Intell. Veh. Symp. Proc.* IEEE, 553–558. DOI: http://dx.doi.org/10.1109/IVS.2017.7995776
11. Dennis Orth, Nadja Schömig, Christian Mark, Monika Jagiellowicz-Kaufmann, Dorothea Kolossa, and Martin Heckmann. 2017b. Benefits of personalization in the context of a speech-based left-turn assistant. In *AutomotiveUI 2017 - 9th Int. ACM Conf. Automot. User Interfaces Interact. Veh. Appl. Proc.* DOI: http://dx.doi.org/10.1145/3122986.3123004
12. Dennis Orth, Nico Steinhardt, Bram Bolder, Mark Dunn, Dorothea Kolossa, and Martin Heckmann. 2018. Analysis of a Speech-Based Intersection Assistant in Real Urban Traffic. In *IEEE Conf. Intell. Transp. Syst. Proceedings, ITSC*, Vol. 2018-Novem. 1273–1278. DOI: http://dx.doi.org/10.1109/ITSC.2018.8569821
13. Sebastiaan M. Petermeijer, Joost C.F. De Winter, and Klaus J. Bengler. 2016. Vibrotactile Displays: A Survey with a View on Highly Automated Driving. (2016). DOI: http://dx.doi.org/10.1109/TITS.2015.2494873
14. Bastian Pfleging, Stefan Schneegass, and Albrecht Schmidt. 2012. Multimodal interaction in the car - combining speech and gestures on the steering wheel. In *Proc. 4th Int. Conf. Automot. User Interfaces Interact. Veh. Appl.* 155–162. DOI: http://dx.doi.org/10.1145/2390256.2390282
15. Andreas Riener, Myounghoon Jeon, Ignacio Alvarez, and Anna K. Frison. 2017. Driver in the Loop: Best Practices in Automotive Sensing and Feedback Mechanisms. In *Springer*. 295–323. DOI: http://dx.doi.org/10.1007/978-3-319-49448-7_11
16. Nadja Schoemig, Martin Heckmann, Heiko Wersing, Christian Maag, and Alexandra Neukum. 2018. "Please watch right" - Evaluation of a speech-based on-demand assistance system for urban intersections. *Transp. Res. Part F Traffic Psychol. Behav.* 54 (2018), 196–210. DOI: http://dx.doi.org/10.1016/j.trf.2018.01.018
17. Nadja Schömig, Martin Heckmann, Heiko Wersing, Christian Maag, and Alexandra Neukum. 2016. Assistance-on-demand: A speech-based assistance system for urban intersections. In *Adjun. Proc. 8th Int. Conf. Automot. User Interfaces Interact. Veh. Appl.* DOI: http://dx.doi.org/10.1145/3004323.3004328
18. Arne Stinchcombe and Sylvain Gagnon. 2009. Estimating Workload Demands of Turning Left at Intersections of Varying Complexity. In *Proc. 5th Int. Driv. Symp. Hum. Factors Driv. Assess. Des.* 440–446. DOI:http://dx.doi.org/10.17077/drivingassessment.1355