Honda Research Institute EU
Honda Research Institute Europe GmbH
https://www.honda-ri.de/

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

Martin Ernst Heckmann, Dennis Orth, Nico Andreas Steinhardt, Bram Bolder, Mark Dunn, Dorothea Kolossa

2019

Preprint:

This is an accepted article published in AutomotiveUI ’19 Adjunct . The final authenticated version is available online at:
https://doi.org/10.1145/3349263.3349599

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

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.
AutomotiveUI '19 Adjunct , September 21–25, 2019, Utrecht, Netherlands
©2019 Copyright is held by the owner/author(s).
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 a 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;

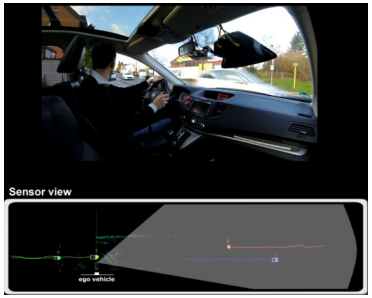

Figure 1: Screenshot of accompanying video.

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 [17, 18]. The AOD concept allows the driver to flexibly engage in an interaction with the ADAS via speech whenever she encounters a challenging situation and wishes for assistance. We derived the interaction concept for the left-turn assistant from observations of drivers' natural behaviour: When turning left at an unsignalized intersection, they 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 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, such a scenario is one of the most challenging tasks for drivers [5, 19]. 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 [17, 18]. 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 [11, 10]. Using these for personalizing the system recommendations significantly further improved the acceptance of the system [12]. Subsequently, we implemented the AOD concept in a prototype vehicle [9] and evaluated it in urban traffic [13].

In this paper we outline the necessary adaptations of the interaction concept from the simulator towards real traffic and the main building blocks of a prototype system. The accompanying video shows the interaction between the driver and the prototype system while cooperatively turning left at an urban intersection[1] (compare Fig. 1).

Interaction Concept

In the following, we will outline the general interaction concept and the resulting situated dialog model.

Speech as Sole Interaction Modality

As the system only communicates via speech, the driver can fully focus her visual attention on the actual driving task. In addition to speech [15, 16, 4], in particular auditory [16] and tactile information [14, 3, 7, 8] have already been investigated quite thoroughly as alternatives to visual communication in the automotive context. We decided for a 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, it mimicks the human-human interaction mentioned before. In our implementation, the driver does not need to push a button but can directly address the system with a wake-up word ("CORA" as an acronym for *COpeartive Assistant*) followed by her request (e.g., "CORA, please watch right.").

[1] The video can be accessed at https://www.ruhr-uni-bochum.de/ika/forschung/forschungsbereich_kolossa/forsch_kolossa_engl.htm

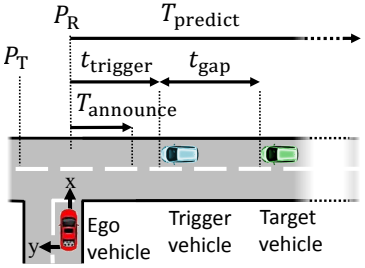

Figure 2: Top view of a T-intersection, showing the ego vehicle (red) waiting for the vehicles from the right to pass the intersection.

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 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, the timing of the system's announcements (its verbal references) is key in structuring the dialog, in particular as often the traffic situation changes while the system makes an announcement. Consequently, the announcements 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 (compare Fig. 2). 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. To avoid confusion of the driver and prevent an overly verbose dialog, each trigger vehicle can typically only trigger one 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. 2). 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 $T_{\text{announce}} = 2.5\,\text{s}$). For this reason, the system starts announcing a fitting gap when $t_{\text{trigger}} = T_{\text{announce}}$. Earlier announcements would increase the uncertainties in the prediction. A gap just large enough for a specific driver is the so-called critical gap $T_{\text{gap crit}}$ [11, 2]. Hence, the system announces 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 announces "Gap after next vehicle." In this utterance, "next vehicle" refers to the trigger vehicle, the closest vehicle to the right and which triggered the announcement, and "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 announcement 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 this time is chosen as the new trigger vehicle.

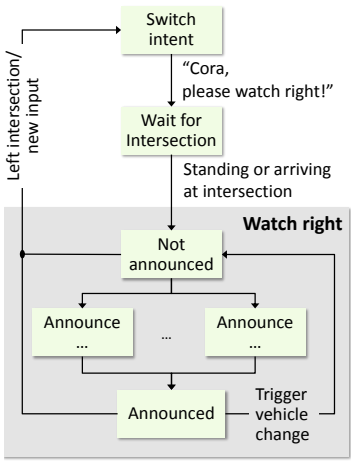

Figure 3: Overview of the dialog manager state machine.

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.

Dialog State Machine

The dialog state machine is organized in two hierarchical levels (compare Fig. 3 and see [9] for more details). 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 and transits to *Wait for intersection* if the user wants the system to watch the right side. From there it transits to the *Watch right* state, either after arriving at the intersection or direclty if the ego vehicle is already standing at the intersection. 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* sub-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 it switches to a new trigger vehicle and transits back to *Not announced* and the cycle restarts. Table 1 shows a list of all announcements the system makes.

System Overview

In this section we give a short overview of the remaining building blocks of the system (see [9] for more details).

System Hardware and Software

We implement the prototype system in a modified 2012 model-year Honda CR-V with automatic transmission. 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 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 50 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

We use the position and velocity estimates for the traffic participants provided by the LIDAR sensor and apply a tracking algorithm which can handle occlusions of the vehicles approaching from the right by the vehicles passing from the left [9]. 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.

Conclusion

In this paper and with the accompanying video, 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. Recordings form real urban traffic 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. The formal evaluations of the system in the driving simulator [17, 12] and the informal evaluations based on the subjective impressions of a few people driving with the prototype system indicate that the benefits of the system transfer from the simulator to the real world.

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, and Martin Heckmann. 2017a. Predicting driver left-turn behavior from few training samples using a maximum a posteriori method. In *IEEE Conf. Intell. Transp. Syst. Proceedings, ITSC*, Vol. 2018-March. IEEE, 1–6. DOI: http://dx.doi.org/10.1109/ITSC.2017.8317721
11. Dennis Orth, Dorothea Kolossa, Milton Sarria Paja, Kersten Schaller, Andreas Pech, and Martin Heckmann. 2017b. 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
12. Dennis Orth, Nadja Schömig, Christian Mark, Monika Jagiellowicz-Kaufmann, Dorothea Kolossa, and Martin

Heckmann. 2017c. 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

13. 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

14. 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

15. 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

16. 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

17. 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

18. 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

19. 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