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

"Gap after the next two vehicles": A Spatio-temporally Situated Dialog for Human-car Cooperative Driving

Martin Ernst Heckmann, Dennis Orth, Dorothea Kolossa

2018

Preprint:

This is an accepted article published in 13. ITG Fachtagung Sprachkommunikation Oldenburg. The final authenticated version is available online at: https://doi.org/[DOI not available]

"Gap after the next two vehicles": A Spatio-temporally Situated Dialog for a Cooperative Driving Assistant

Martin Heckmann[1], Dennis Orth[1,2], Dorothea Kolossa[2]

[1]Honda Research Institute Europe, Offenbach/Main, Germany
[2]Institute of Communication Acoustics, Ruhr-Universität Bochum, Bochum, Germany

Email: martin.heckmann@honda-ri.de, {dennis.orth, dorothea.kolossa}@rub.de

Preprint version. Paper published in: 13. ITG Fachtagung Sprachkommunikation/Speech Communication. © 2018 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

"Gap after the next two vehicles": A Spatio-temporally Situated Dialog for a Cooperative Driving Assistant

Martin Heckmann[1], *Dennis Orth*[1,2], *Dorothea Kolossa*[2]

[1]Honda Research Institute Europe, Offenbach/Main, Germany
[2]Institute of Communication Acoustics, Ruhr-Universität Bochum, Bochum, Germany
Email: martin.heckmann@honda-ri.de, {dennis.orth, dorothea.kolossa}@rub.de

Abstract

We present a spatio-temporally situated dialog that is implemented in a driver assistance system. The system supports the driver in turning left at a busy urban intersection by providing verbal information on the vehicles arriving from the right. This is a highly dynamic scenario, as the location of the vehicles significantly changes while the system is referring to them. Consequently, the time the system needs for producing the utterances and the time drivers need to comprehend them and prepare their action has to be taken into account in the dialog management. We introduce a dialog concept, which predicts the future traffic participants' state and plans the utterances such that they align with the driver's expectations. We have implemented and evaluated this system in a prototype vehicle.

1 Introduction

So far, speech-based intelligent personal assistants can mainly provide users with internet-based information or control home appliances. The next step will be the development of systems that are mobile and can cooperate with users in the physical world.

1.1 Situated dialogs for cooperative activities

When humans engage in a cooperative activity, they typically have to make references to objects in the world to coordinate their joint goal. They use precise verbal descriptions, pointing gestures, head nods or eye gaze to make these references. Such situated dialogs, i.e., dialogs that are targeted at and embedded in the physical environment, have also been investigated for human-machine interaction. In these dialogs, two related problems have to be solved: identifying the object the human attention is directed to, and communicating the system's attention to the human. For the first problem, many approaches have been proposed which integrate speech with pointing gestures or eye gaze [1–11]. For the second problem, communicating the system's attention to the user, systems have been equipped with the capability to generate referring expressions. The underlying algorithms integrate knowledge on the objects present and determine which set of properties is needed to single out the target from the distractors [12–14]. These algorithms also have been extended to integrate multimodal cues [7, 8, 13, 15, 16]. The corresponding scenarios all have in common that they are at least quasi-static, i. e., the environment does not change while either the human or the system is speaking. In contrast to this, [17] address situated language understanding in a moving vehicle. While driving, users can query the system regarding specific target buildings in their surroundings. In this scenario the timing of the utterance is very important as the environment changes quickly. We also consider a dialog embedded into a dynamic traffic situation. In the next section, we will introduce our scenario based on the cooperative driving concept we have developed.

1.2 The Assistance on Demand concept

Advanced Driver Assistance Systems (ADAS) are designed to support the driver by avoiding safety-critical situations or by issuing warnings to the driver, e. g., blind spot warning. Occasional warnings when there is no danger or when the driver has already perceived it have the potential to annoy the driver and reduce acceptance of such systems. To overcome these limitations, we have recently proposed the so-called "Assistance on demand" (AOD) concept [18]. This concept is inspired by observations of drivers' natural behavior, e. g., while turning left at an unsignalized intersection from a subordinate road into a superordinate road. At high traffic densities this is a very challenging task for drivers [19]. When driving with a front seat passenger, they often use the opportunity to ask the passenger to support them by monitoring the right-side traffic and provide feedback on suitable time gaps to enter the intersection. This creates a situation where passenger and driver cooperate in the driving task. The cooperation is accomplished naturally, using speech as the medium of communication. In our AOD approach we transfer this idea to the interaction with the assistance system. The driver can activate it via speech and it will give feedback, also via speech, on the current traffic situation to provide support in the maneuver decision. In [20] a very similar approach for the use case of "turning left at a rural intersection with oncoming traffic" was applied in a simulator with positive results for the system. However, the AOD approach more strongly emphasizes the collaborative sharing of tasks between the driver and the system in a more demanding use case. Additionally, [20] proposed a system-initiative approach in which the system would identify a need for assistance and suggest it to the driver. In contrast, the premise of the AOD approach is to minimize the possible annoyance by the system and to leave it largely to the driver when and how to activate it. We have already shown in simulator studies that a system based on the AOD approach was well accepted by the participants [21, 22]. Following this we implemented the system in a prototype vehicle to further refine and evaluate it [23]. In this paper we illustrate the spatio-temporally situated dialog model we devised for this system.

Sec. 2 will summarize the requirements for the dialog in our scenario and outline the main ideas behind the dialog concept we developed. This will be followed by a detailed description of the implementation of the concept. After that we will give an overview of the overall driver assistance system and the integration of the dialog in Sec. 3. Finally, in Sec. 4 we will discuss open issues and their possible solutions followed by a conclusion in the last section.

2 Spatio-temporally Situated Dialog

In the following, we will first outline the prerequisites for a situated dialog in our left-turn scenario. Next we will present the main attributes of the resulting dialog model with an example and finally, we give a detailed description of the implementation of the prototype system.

2.1 Situated Dialog for Cooperative Driving

In our scenario the driver and the system cooperatively manage the driving task. There is a clear role assignment between the driver and the system based on the request formulated by the driver ("Please watch right.") While the system is monitoring the right side, the driver is expected to observe the remaining directions, in particular the left side. As a consequence, the driver should not need to spend additional visual attention to interpret the system's feedback. Hence, the system's feedback should be based on speech only. Further limiting factors are that the objects are moving quickly and that additional features of the objects, such as type (car, truck, pedestrian ...), size or color are not reliably available from the environment sensor. The latter only allows for referring to the traffic objects as "vehicles." Hence, the timing of the system behavior has to be the main structuring element in 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. Thus, 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. 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.

2.2 General Dialog Logic

We will now illustrate the dialog logic in more detail based on a few possible 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}. Consequently, the system should start announcing a fitting gap at the latest when $t_{\mathsf{trigger}} = T_{\mathsf{announce}}$. A gap just large enough for a specific driver is the so-called

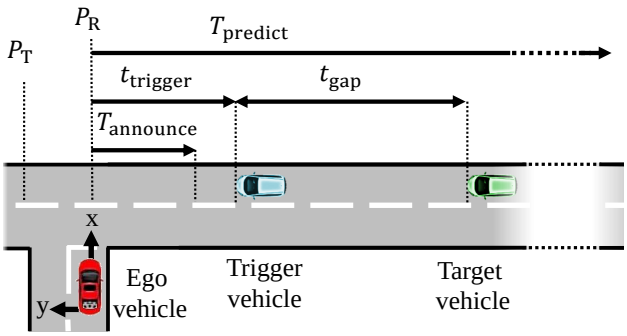


Figure 1: Top view of a T-intersection, showing the ego vehicle (red) waiting for the vehicles from the right to pass the intersection.

critical gap T_{gapcrit} [24, 25]. Hence, the system should announce a gap if $t_{\mathsf{gap}} \geq T_{\mathsf{gapcrit}}$.

Assuming the trigger vehicle is still sufficiently far away to make an announcement, i. e., $t_{\mathsf{trigger}} \geq T_{\mathsf{announce}}$, and the gap behind this vehicle is large enough for the driver to make the turn, i. e., $t_{\mathsf{gap}} \geq T_{\mathsf{gapcrit}}$, 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. The system's prediction horizon is limited mainly by the system's perception range and the accumulating uncertainties in the future environment state with an increasing prediction horizon. We assume that the system can look approximately D_{predict} meters to the right due to sensor range limitations and that all vehicles are moving at a constant speed. Assuming a maximum speed of the traffic participants $v_{\max}$ defined by the current speed limit, the minimal prediction time T_{predict} (Fig. 1) can be calculated as $T_{\mathsf{predict}} = D_{\mathsf{predict}}/v_{\max}$. Due to this limited and situation-dependent perception horizon T_{predict}, it is advisable to make the announcement of the fitting gap as late as possible, i.e., when $t_{\mathsf{trigger}} = T_{\mathsf{announce}}$. This maximizes the likelihood that all vehicles relevant for the planned announcement have already been detected. In case of $t_{\mathsf{gap}} < T_{\mathsf{gapcrit}}$ 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 after the trigger vehicle has passed the trigger point P_T, leaving the target vehicle as the closest vehicle from the right. The trigger point is $l_{\mathsf{vehicledefault}}$ to the left of the reference point. The value $l_{\mathsf{vehicledefault}}$ represents a default vehicle length, which we use as the real vehicle length is difficult to obtain reliably from our sensors. The next announcement will only be possible after the trigger vehicle has passed the trigger point. Then it is removed from the environment representation and the vehicle which is closest at that time is chosen as the new trigger vehicle.

In dense traffic, very often t_{gap} is smaller than T_{announce}, the time needed for an announcement. Hence there would not be enough time to make the following announcement once the trigger vehicle has passed. This is particularly problematic when the system wants to announce a fitting gap. We propose to deal with those cases by conceptually combining the current trigger vehicle and the target vehicle to a trigger vehicle group. We call the vehicle following this group the *target follower*. If the gap between the last vehicle of this trigger vehicle group (the target vehicle) and the target follower has been measured

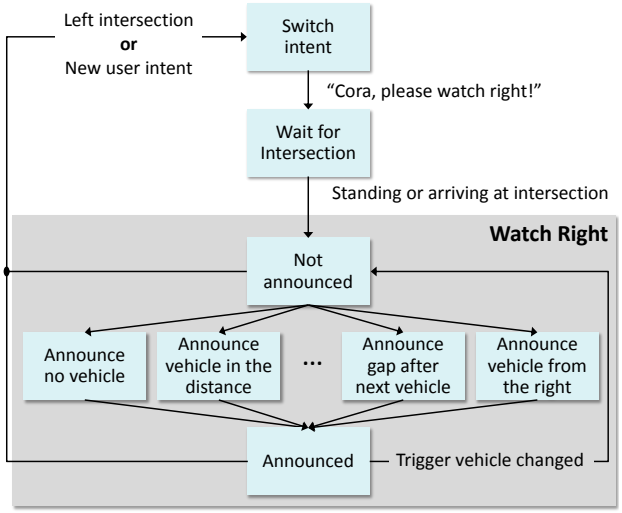


Figure 2: Overview of the dialog manager state machine.

to be larger than T_{gapcrit} the system outputs "Gap after the next two vehicles." However, this requires a rather long prediction, which might be compromised by T_{predict}.

So far we have only given an overview of the general logic of the dialog management. In the next section we will introduce the hierarchical dialog state machine we have developed to make the wide range of situations encountered in real traffic manageable.

2.3 Dialog State Machine

The dialog state machine is organized in two hierarchical levels. On the top level, the interpretation of the user intent and the ego vehicle state are handled. The feedback on the traffic situation is orchestrated in the *Watch Right* state and its subordinate states.

2.3.1 Top-level State Machine

Fig. 2 shows the hierarchical state machine of the proposed AOD dialog manager. On the top level, the state machine iterates through the following states:

- *Switch intent*: The system is waiting for an input from the driver. If the natural language understanding (NLU) component recognizes that the driver's intent is to make the system watch the right-hand side, it will transition to *Wait for Intersection*. Currently there are no other intents implemented.
- *Wait for Intersection*: The system analyzes the state of the ego vehicle. If the arrival at the intersection has already been detected, the system outputs "Okay. I am watching" and immediately transits to *Watch Right* otherwise it will output "Okay. I will watch," waits until the arrival at the intersection is detected, and then also transits to *Watch Right*.
- *Watch Right*: In this state the traffic environment is analyzed and feedback is given to the driver. This is implemented in a subordinate state machine, which is explained in more detail in the next section.

2.4 Watch-Right State Machine

The Watch-Right State Machine consists of three different types of states:

- *Not Announced*: The system reaches this state if no announcement has yet been made for the current trigger vehicle. In this state, the conditions for transition into one of the *Announcement* states are checked.
- *Announcement*: In this set of states, the system gives feedback to the driver on the current traffic situation. Once the system has started to give the feedback for a specific trigger vehicle, it transits immediately to the *Announced* state.
- *Announced*: The system waits in this state 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*.

From the *Not Announced* state, a transition to one of the following announcement states takes place if the corresponding condition is met:

- *Announce No Vehicle*: If there is no vehicle approaching the intersection from the right, the system outputs "No vehicle from the right". An internal timer assures that this announcement is performed at most every $T_{\text{novehicle}}$ seconds.
- *Announce Vehicle in the Distance*: If there was previously no trigger vehicle and $t_{\text{trigger}} \geq T_{\text{announce}} + T_{\text{gapcrit}}$ the system outputs "Vehicle in the distance." This condition ensures that the distance to the approaching vehicle is still large enough for the driver after the announcement has finished.
- *Announce No Vehicle after Next*: If there is no target vehicle and $t_{\text{trigger}} \leq T_{\text{announce}}$ the system outputs "No vehicle after the next vehicle."
- *Announce Gap*: If $t_{\text{gap}} \geq T_{\text{gapcrit}}$ and $t_{\text{trigger}} \leq T_{\text{announce}}$ the system outputs "Gap after next vehicle."
- *Announce Vehicle*: If $t_{\text{gap}} < T_{\text{gapcrit}}$ the system will wait until the trigger vehicle reaches the trigger point and then outputs "Vehicle from the right." This state also has an internal timer (T_{vehicle}) which prevents that this message is repeated too frequently in dense traffic situations.
- *Announce Gap after Next Two Vehicles/No Vehicle After Next Two Vehicles*: If $t_{\text{gap}} < T_{\text{announce}}$ the trigger vehicle and target vehicle are conceptually combined. We call the vehicle behind the target vehicle target follower and the gap to it $T_{\text{gapfollower}}$. If $T_{\text{gapfollower}} \geq T_{\text{gapcrit}}$ and $t_{\text{trigger}} \leq T_{\text{announce}}$ the system announces "Gap after the next two vehicles." If there is no target follower the system announces "No vehicle after the next two vehicles" instead.

In every announcement state, the text of the corresponding utterance is sent to the TTS module and the system immediately transits to the *Announced* state. As outlined above, the system remains in this state until the trigger vehicle changes. The system also leaves the state *Announced* if there is no trigger vehicle present at all or if we need to make an additional announcement for the same trigger vehicle. The latter occurs if, e.g., the system has announced "Vehicle in the distance" and needs to make another announcement when $t_{\text{trigger}} = T_{\text{announce}}$ or the trigger vehicle reaches P_T.

In addition to the logic outlined so far we have implemented a few shortcuts and additional states. These are mainly needed to handle the arrival at the intersection. However, due to space constraints they are not further detailed here.

In the *Not Announced* and the *Announced* states, there are regular checks if a new driver utterance has been received or the ego vehicle has left the intersection. In these cases the system interprets this as a signal to abort the cur-

rent dialog and will transit back to the *Switch Intent* state.

3 System Overview

We have implemented the dialog model in a prototype driver assistance system. This prototype system runs in real time on a modified 2012 model-year Honda CR-V and can be used to assess the system's behavior in real traffic. The system comprises the following components [23]:

- *Sensor data acquisition*: In addition to the standard equipment, the prototype vehicle features 6 Ibeo Automotive Systems LUX LIDAR sensors. They sample the environment with $f_{\mathsf{LIDAR}} = 25\,\mathrm{Hz}$ and have an overall coverage of 360°. We use them to estimate the positions and velocities of all traffic participants in the vicinity of the ego vehicle. 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 ($f_{\mathsf{CAN}} = 50\,\mathrm{Hz}$). We acquire the speech commands from the driver via a XCore 7-channel microphone array.
- *Scene understanding*: To be able to handle occlusions of the vehicles approaching from the right by the vehicles passing from the left side we extended the tracking algorithms provided by the LIDAR sensor [23]. Based on these corrected 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, we use methods to precisely detect the arrival and departure of the ego vehicle from the intersection [23]. In particular if the departure detection is delayed, the system will continue to give feedback on the traffic whilst the vehicle has already left the intersection, which is perceived as very annoying.
  To facilitate the interaction with the AOD system, the automatic speech recognition (ASR) continuously listens for the wake-up word. We have chosen 'CORA' as the wake-up word as an acronym for *Cooperative Assistant*. Once the wake-up word is detected, the recognition of the driver's utterance is triggered. Speech enhancement, recognition and synthesis are accomplished via the SSE™, VoCon™ and Vocalizer™ software respectively, all from Nuance. A simple NLU component analyzes the utterance and forwards the result to the dialog manager, i. e., it filters the driver's utterance for the word indicating the direction ("front", "left" or "right").
- *Dialog Manager*: The dialog manager is activated when the NLU has detected the intent of the driver to make the system watch right. Then it continuously ($f_{\mathsf{dialog}} = 10\,\mathrm{Hz}$) receives the data from the gap estimation and ego vehicle state estimation module, analyzes it and shapes the dialog as outlined in Sec. 2.3. For the dialog manager, $T_{\mathsf{announce}} = 2.5\,\mathrm{s}$ is used as an estimate of the cumulative time needed for the system to output an utterance and for the driver to comprehend it. Given the approximate range of the LIDAR of 150 m, the maximum time for which we can predict the traffic situation on the right side (T_{predict}) is 11.0 s when vehicles are assumed to move at a maximum velocity of $50\,\mathrm{km\,h^{-1}}$, the speed limit in the area where we have tested the system. As mentioned previously, the actual detection distance unfortunately depends on the current situation.
- *System Output*: The system interacts with the environment only via speech. For this purpose it uses a text-to-speech (TTS) module and outputs the utterance via the standard audio equipment of the vehicle.

4 Limitations and Future Work

The sensor capabilities of the prototype system constrain the dialog to a large extent.

One direction we are currently investigating to overcome these restrictions is to infer the traffic participant type based on an analysis of the traffic participants' current and past movement behavior. This will allow us to adapt the dialog better to the current situation.

So far, in our predictions we assume that the velocity of the traffic vehicles is constant. In an analysis of the behavior of the system in real traffic, we could show that despite this rather restrictive assumption, we can achieve a 90% prediction accuracy and that a more sophisticated second order model for the traffic participants does not yield to further improvements [26].

Another limiting factor is the rather short sensor range. Humans can easily look three to four times further into the intersection. This limited sensor range and the resulting limited prediction range also confines the reasoning we can perform on the development of the traffic situation. The grouping of two closely following vehicles into one trigger vehicle already pushes at the boundaries of the prediction range in our scenario. However, these limitations can only be tackled with alternative or additional sensors.

In its current version, the system does not check if an announcement that it has started is still valid after it is finished. To minimize potential confusions of the driver due to changes in the traffic situation after the start of the announcement, in future work, we will implement a process which constantly verifies that the current traffic situation is still in line with the last announcement and upon a deviation reacts appropriately.

5 Conclusion

In this work we have presented a spatio-temporally situated dialog system, which makes references to quickly moving objects. Our system supports the driver in turning left. The driving task already entails a high load on the visual attention of the driver. Therefore, our system only uses verbal feedback to reference the surrounding vehicles. We have shown how this can be achieved via a predictive dialog planning, which uses precise timings of the utterances. We have implemented a prototype system in a vehicle. The system can be activated by speech commands and is able to recognize when the driver is standing at the intersection. It then tracks the vehicles approaching from the right using LIDAR sensors, estimates the gaps between the vehicles and gives the driver recommendations on suitable gaps until it detects the driver starting the left turn. The ongoing evaluations of the system in real traffic were key in the development of the dialog model. They have helped us to identify many relevant aspects of the possible traffic conditions and options for handling them.

Future work will be aimed at addressing the remaining issues outlined above, in order to arrive at a dialog model that is fully adaptive to the environment, including all possible dynamic changes of the traffic conditions.

References

[1] R. A. Bolt, ""Put-that-there": Voice and Gesture at the Graphics Interface," *Proc. 7th Annu. Conf. Comput. Graph. Interact. Tech. - SIGGRAPH '80*, pp. 262–270, 1980.

[2] L. Heck, D. Hakkani-Tür, M. Chinthakunta, G. Tur, R. Iyer, P. Parthasarathy, L. Stifelman, E. Shriberg, and A. Fidler, "Multi-Modal Conversational Search and Browse," in *First Work. Speech, Lang. Audio Multimed.*, 2013.

[3] R. Stiefelhagen, C. Fügen, P. Gieselmann, H. Holzapfel, K. Nickel, and A. Waibel, "Natural human-robot interaction using speech, head pose and gestures," in *2004 IEEE/RSJ Int. Conf. Intell. Robot. Syst. (IEEE Cat. No.04CH37566)*, vol. 3, pp. 2422–2427, 2004.

[4] B. Schauerte and G. A. Fink, "Focusing computational visual attention in multi-modal human-robot interaction," in *Int. Conf. Multimodal Interfaces Work. Mach. Learn. Multimodal Interact. - ICMI-MLMI '10*, p. 1, 2010.

[5] C. Matuszek, L. Bo, L. S. Zettlemoyer, and D. Fox, "Learning from Unscripted Deictic Gesture and Language for Human-Robot Interactions," in *Proc. AAAI 2014*, pp. 2556–2563, 2014.

[6] G. Canal, S. Escalera, and C. Angulo, "A real-time Human-Robot Interaction system based on gestures for assistive scenarios," *Comput. Vis. Image Underst.*, vol. 149, pp. 65–77, 2016.

[7] J. Rickel and W. L. Johnson, "Animated Agents for Procedural Traning in Virtual Reality: Perception, Cognition and Motor Control," *Appl. Artif. Intell.*, vol. 13, no. 4-5, pp. 343–382, 1999.

[8] Y. I. Nakano, G. Reinstein, T. Stocky, and J. Cassell, "Towards a model of face-to-face grounding," in *Proc. 41st Annu. Meet. Assoc. Comput. Linguist. - ACL '03*, vol. 1, pp. 553–561, 2003.

[9] R. Iida, M. Yasuhara, and T. Tokunaga, "Multi-modal Reference Resolution in Situated Dialogue by Integrating Linguistic and Extra-Linguistic Clues," in *Proc. 5th Int. Jt. Conf. Nat. Lang. Process. (IJCNLP 2011)*, no. 2003, pp. 84–92, 2011.

[10] D. Hakkani-Tür, M. Slaney, A. Celikyilmaz, and L. Heck, "Eye Gaze for Spoken Language Understanding in Multi-modal Conversational Interactions," in *Proc. 16th Int. Conf. Multimodal Interact. - ICMI '14*, pp. 263–266, 2014.

[11] C. Kennington and D. Schlangen, "A simple generative model of incremental reference resolution for situated dialogue," *Comput. Speech Lang.*, vol. 41, pp. 43–67, 2017.

[12] R. Dale and E. Reiter, "Computational interpretations of the Gricean maxims in the generation of referring expressions," *Cogn. Sci.*, vol. 19, no. 2, pp. 233–263, 1995.

[13] I. Van Der Sluis and E. Krahmer, "Generating multimodal references," *Discourse Process.*, vol. 44, no. 3, pp. 145–174, 2007.

[14] E. Krahmer and K. van Deemter, "Computational Generation of Referring Expressions: A Survey," *Comput. Linguist.*, vol. 38, pp. 173–218, mar 2012.

[15] A. Kranstedt, A. Lücking, T. Pfeiffer, H. Rieser, and I. Wachsmuth, "Deixis: How to determine demonstrated objects using a pointing cone," in *Lect. Notes Comput. Sci.*, vol. 3881 LNAI, pp. 300–311, 2006.

[16] G. Skantze, A. Hjalmarsson, and C. Oertel, "Turn-taking, feedback and joint attention in situated human-robot interaction," *Speech Commun.*, vol. 65, pp. 50–66, 2014.

[17] T. Misu, A. Raux, R. Gupta, and I. Lane, "Situated language understanding for a spoken dialog system within vehicles," *Comput. Speech Lang.*, vol. 34, no. 1, pp. 186–200, 2015.

[18] N. Schömig, M. Heckmann, H. Wersing, C. Maag, and A. Neukum, "Assistance-on-demand: A speech-based assistance system for urban intersections," AutomotiveUI '16 Adjunct, pp. 51–56, ACM, 2016.

[19] A. Stinchcombe and S. Gagnon, "Estimating workload demands of turning left at intersections of varying complexity," vol. 5 of *Driving Assessment 2009*, pp. 440–446, University of Iowa Public Policy Center, 2009.

[20] L. Biester, *Cooperative automation in automobiles*. PhD thesis, Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät II, 2009.

[21] N. Schoemig, M. Heckmann, H. Wersing, C. Maag, and A. Neukum, ""Please watch right" - Evaluation of a speech-based on-demand assistance system for urban intersections," *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 54, pp. 196–210, 2018.

[22] D. Orth, N. Schömig, C. Mark, M. Jagiellowicz-Kaufmann, D. Kolossa, and M. Heckmann, "Benefits of personalization in the context of a speech-based left-turn assistant," in *AutomotiveUI '17*.

[23] D. Orth, B. Bolder, N. Steinhardt, M. Dunn, D. Kolossa, and M. Heckmann, "A Speech-based On-demand Intersection Assistant Prototype," in *To appear in Proc. IEEE Intell. Veh. Symp.*, 2018.

[24] D. Orth, D. Kolossa, M. S. Paja, K. Schaller, A. Pech, and M. Heckmann, "A maximum likelihood method for driver-specific critical-gap estimation," in *IEEE Intell. Veh. Symp. Proc.*, pp. 553–558, IEEE, jun 2017.

[25] W. Brilon, R. Koenig, and R. J. Troutbeck, "Useful estimation procedures for critical gaps," *Transp. Res. Part A*, vol. 33, pp. 161–186, apr 1999.

[26] D. Orth, N. Steinhardt, B. Bolder, M. Dunn, D. Kolossa, and M. Heckmann, "Analysis of a speech-based intersection assistant in real urban traffic," in *2018 IEEE 21th Int. Conf. on Intell. Transp. Syst. (ITSC)*.