2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)
Windsor Oceanico Hotel, Rio de Janeiro, Brazil, November 1-4, 2016

Touch versus Mid-Air Gesture Interfaces in Road Scenarios - Measuring Driver Performance Degradation

Thomas Kopinski[1], Jan Eberwein[2], Stefan Geisler[2] and Uwe Handmann[2]

***Abstract*— We present a study aimed at comparing the degradation of the driver's performance during touch gesture vs mid-air gesture use for infotainment system control. To this end, 17 participants were asked to perform the Lane Change Test. This requires each participant to steer a vehicle in a simulated driving environment while interacting with an infotainment system via touch and mid-air gestures. The decrease in performance is measured as the deviation from an optimal baseline. This study concludes comparable deviations from the baseline for the secondary task of infotainment interaction for both interaction variants. This is significant as all participants are experienced in touch interaction, however have had no experience at all with mid-air gesture interaction, favoring mid-air gestures for the long-term scenario.**

I. Introduction

Modern vehicles are equipped with a multitude of infotainment features for audio entertainment, sat navigation or connectivity. When designing a user interface within the frame of this scenario, one important aim - as compared to entertainment devices or apps - is to minimize driver distraction and degradation of driving performance. The German car manufacturers BMW and Volkswagen recently announced and presented the introduction of mid-air gesture control with the goal to reduce driver distraction, but until now, no quantitative research results have been published. A usability study for in-car gesture control systems is presented by Zobl et al. [1], but it does not focus on driver distraction. The integration of mid-air gestures into the human-machine interaction (HMI), in addition to touchscreen and/or turn-push-controllers, leads to multimodal user interfaces. Pfleging et al. have developed a similar approach [2]. In their work, classical controls are combined with voice control and gesture-enabled touch areas. The perceived usability and task load of their approach are similar to the base line with classical controls. Positive user feedback concerning mid-air gesture control for in-car interaction was observed by Loehmann et al. [3]. In a qualitative feedback, drivers stated that they felt less visual distraction nor any negative influence on driving performance, simultaneously rating the user experience high. In [4] a driving simulator study is presented comparing an HMI with buttons and rotaries in a low-mounted position to mid-air gestures. The results indicate less distraction and better driving performance for the gesture system. This is in line with the observations made by Kopinski et al. when developing a gesture recognition system [5]. The findings of Pickering et al. [6] show that hand gestures can indeed well be utilized for secondary tasks such as infotainment control, in order to maintain the goal of eyes-on-the-road hands-on-the-wheel. However more extensive research needs to be addressed towards the question of cognitive load and driver distraction as well as applicability of individual gestures to specific tasks. They furthermore address the issue of acceptability which seems to vary between different user groups. Doring et al. [7] present a study with 12 participants demonstrating the validity of gestural input for controlling infotainment systems while being able to keep the hands on the wheel. Using a multi-touch steering wheel they prove that gestural input can reduce distraction significantly opposed to traditional input devices. Further evaluation by conducting experiments with the Lane Change Task (LCT) are performed to corroborate the theory, however no significant insights could be gained from these experiments. However, further research is required to get quantitative data comparing touch screen to contactless gesture input. Therefore, we conducted a simulator study focusing on the Lane Change Test stating the hypothesis that touchless interaction via mid-air gestures is less distractive during a car drive than touch interaction.

Fig. 1. Participant with the interface nearby (for touch interaction).

[1]Thomas Kopinski is with the ENSTA ParisTech/UEI Lab, 858 Blvd des Maréchaux , 91762 Palaiseau, France thomas.kopinski@ensta-paristech.fr

[2]Jan Eberwein, Stefan Geisler and Uwe Handmann are with the Computer Science Institute, Hochschule Ruhr West, 46236 Bottrop, Germany firstname.lastname@hs-rw.de

Fig. 2. Participant with the interface further distant (for mid-air interaction).

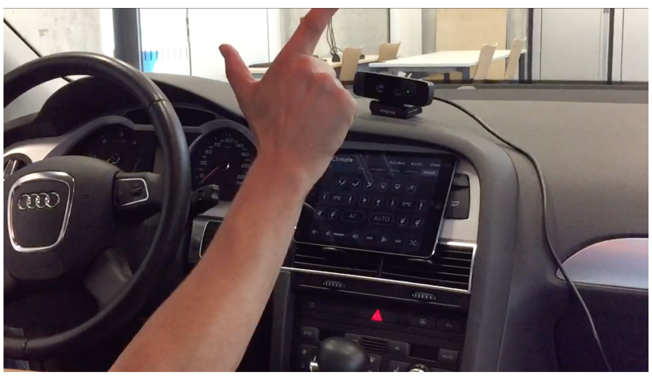

Fig. 3. Our system setup with an iPad and the Creative Gesture Camera as described in Section II.

II. IN-CAR SETTING

Our setup targets a system within the vehicle emulating potential HMI scenarios within which the user is enabled to manipulate infotainment functions via mid-air gestures. It consists of the Creative Gesture Camera recording the VOI in the interior of the vehicle just before the front console with a lateral resolution of 320×160 of the depth sensor. The iPad is mounted to the front console and runs an application with typical infotainment scenarios (media, maps, contacts, phone, climate). Figure 4 shows four of the sixteen (sub-)screens of our infotainment system. Overall, typical functions such as media or navigation selection, navigating through submenus, browsing and turning music on/off are addressable through the mid-air gestures.

The camera is connected to a standard laptop which in turn is responsible for recording the VOI in the nearby driver zone, cropping of the recorded point cloud (with a cropping

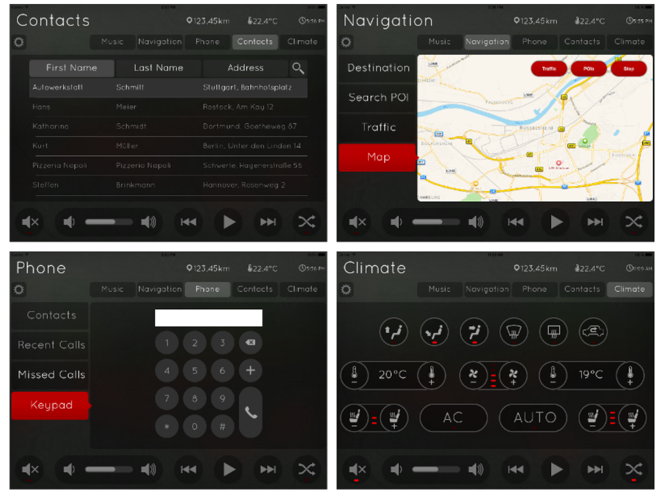

Fig. 4. Sample screens of the infotainment system running on a mobile tablet: Contacts, Navigation, Phone and Climate (top to bottom, left to right).

module to remove irrelevant arm parts) and processing the cloud to compute the features to subsequently pass them to the a convolutional neural network (CNN) for classification. Training is based on a large-scale database consisting of 600.000 hand posture samples obtained from 20 different individuals on the REHAP dataset [8]. We implemented an averaging window to record 20 snapshots in a row and produce a classification results for each. The final decision, which corresponds to the interpreted gesture sent to the iPad, is produced by max-voting. As our systems works with 35-40 fps this is more than sufficient to balance the confidence in decision-making and real-time capability. An additional delay after each gesture is sent, creates a more realistic setting for the application to be tested as the system pauses for interaction and proceeds with the received input[1].

The question remains whether mid-air gestures are more suitable for infotainment interaction than touch gestures in terms of gaze diversion, driver distraction and ease of use. We therefore simulate the scenario on the road by a realistic setting in front of a driving simulator and our infotainment system tailored to the task itself. In order to measure the influence of touch versus mid-air gestures on the performance of the driver the Lane Change Test is performed in a series of experiments and is explained in the subsequent section.

III. THE LANE CHANGE TEST

The Lane Change Test (ISO 26022 standard [9]) aims at measuring the degradation of human performance with respect to a certain primary task (PT) while conducting a secondary task (ST). The result serves as an estimate for the demand of the ST.

The PT for the participant is to drive a vehicle in a simulator on a predefined course with three lanes at a fixed velocity

[1]See the attached video for a live demonstration. Results in real-time are difficult to measure, however this short demonstration shows the robustness of our approach, its applicability and extendability as well as the work-flow.

of 60km/h. During this drive the participant is supposed to keep the vehicle as stable as possible in the middle of the lane and change the lane according to the road signs appearing at approximately equidistant points to the left and right of the driving course. Moreover, an ST is presented to the participant, creating an additional cognitive load and therefore requiring her/him to redirect attention. Within this study, the ST is to interact with an infotainment system realized on an iPad mini via touch or mid-air gestures.

During the interaction phase with touch gestures, the iPad is fixed in a position approximating typical setups in real vehicles (see Fig. 1). In this case it is within the driver's reach zone and positioned in such a way as to provide enough visual feedback while at the same time requiring her/him to redirect gaze during the interaction phase. Conversely, during interaction with freehand gestures, the iPad is placed further to the back, slightly elevated and out of reach of the driver, comparable to highly mounted displays in cars (Fig. 2). This is a crucial feature as we aim at creating a realistic scenario, copying the installation of each technology in modern cars. The time-of-flight camera, suggesting the recording of mid-air hand gesture interaction, is mounted to the far right of the user. As the study presumes a perfect gesture recognition, which is not available with current systems, the Wizard-of-Oz method is used, meaning that performed gestures are not recognized automatically, but by a hidden operator who controls the system.

The task for the participant is to change lane quickly and efficiently as soon as it is clearly recognizable. As driving skills and performance during the LCT can vary significantly between subjects a 'within-subject' design is facilitated where multiple tests are conducted for one participant. This compensates for fluctuations and serves as a control for the subject itself. The LCT is divided into four separate tracks, an initial baseline, to determine driver's performance before the test, followed by a drive with touch and one with touchless interaction (randomized order between participants). It concludes with a second baseline to measure learning effects. Each track lasts approx. 3 minutes and contains 18 lane change signs. The available gestures were explained to the participants followed by a 5-minute interaction phase before the test drive until each participant tested every feature at least twice and felt comfortable with the means of interaction.

IV. Infotainment system and interaction

To simulate typical interaction scenarios, an infotainment system was designed and implemented on an iPad mini. The system contains four sections - maps, radio, CD drive and telephone from left to right - with content accessible via either touch or mid-air gesture interaction. All areas are touch sensitive, the top row is reserved for navigating through the main sections. Each section can in turn be addressed by a static hand gesture, i.e. showing up to four fingers.

When in the navigation section, the user is able to interact with the maps application, showing a cropped segment of the surrounding area, by using the one-finger swipe gesture. Conversely, the participant can use the mid-air counterpart by pointing in the direction towards which the maps app should be scrolled, i.e. point left, right, up or down. While the user has the possibility to exactly determine how far the map should be scrolled, pointing in mid-air scrolls the map in the respective direction in fixed steps.

The second section contains five different radio channels visualized via the well-known cover flow component. The individual channels can be reached step-wise by clicking the left or right buttons. As soon as the radio section is opened, a station-specific audio file starts playing to provide further feedback for the user. Audio can be paused/resumed via the pause/play button at the bottom of the screen. Again, two hand gestures have been attached to the same function, namely holding the hand out flat and pointing towards the screen. Swapping of radio channels is realized via pointing to the left or right of the screen which iteratively scrolls through the channels depending on the side pointed to. Pressing a button is always highlighted as known from touch devices, helping participants recognize whether an interaction step was recognized correctly as haptic feedback is missing in the use case of mid-air gesture interaction.

The third section contains the private music collection visualized by a number of CD covers. The interaction and feedback concept corresponds with the one for the radio section.

Lastly, the fourth section contains a list of phone contacts which can be scrolled through via swiping up/down in the list or pointing via a hand gesture to the respective ends. Moreover, we realized a custom window sliding in, as soon as the 'call event' is triggered by the wizard. It displays an image of the person calling, the contact's name and details as well as buttons for accepting or declining the call. If the call is accepted, a 'call running' window slides in as long as the conversation is held. Ending the call is again possible by a simple button press. Conversely, hanging up or declining the call in the first place can be achieved via the hold gesture (flat hand). Accepting the call in the first instance is realized via pointing towards the the screen. During an incoming/running call all audio files are paused analoguous to regular system behavior.

V. Procedure of LCT

The procedure for the LCT was the same for all participants except for two factors. First of all, out of the ten available tracks 4 were randomly chosen as suggested by the ISO standard for the experiments per participant. 9 of 17 participants started with touch gestures as ST A and then interacted via mid-air gestures as ST B, for the remaining 8 this procedure was switched.

For each test it is emphasized that there is no measuring of the quality of the driving but rather that the focus is on finding cues as to which kind of interaction is more distracting during a simulated drive. The participant is seated and adjusts the seat to a comfortable position. Initially, the most important features of the test are mentioned and explained (sign/symbol meaning, behavior etc.) in accordance with the ISO recommendation. The participant is instructed to change

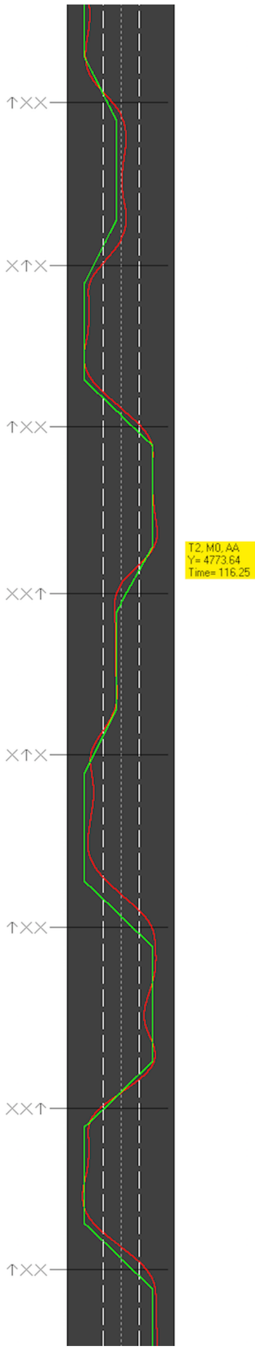

Fig. 5. Visualization of the driver's performance (red line) versus the reference drive (green line).

the lane correctly and efficiently. Focus should be put on the PT while trying to perform the ST as good as possible. She/he drives a test drive which serves as a phase of getting used to the system, hence no data is recorded here. When comfortable to perform the test, the participant is asked to drive the first baseline. The next step depends on whether the distraction is measured during mid-air or touch interaction but either way, the functionality of the infotainment system is explained along with the interaction principle. Moreover every participant is asked to test every functionality of the system once. This means that every participant interacts with the iPad via touching, swiping if the ST is touch interaction or via the corresponding mid-air gestures. Afterwards she/he is asked to drive the first dual-task run during which a supervisor gives instructions to the participant in the form of infotainment interaction tasks. This procedure begins with the START sign of the track and ends with a task after the last lane change sign. During the drive the participant is given tasks on a constant basis and in a random order, i.e. once a task is finished the next one is presented right away. This way, the participant is constantly occupied with an ST. As some participants interacted with the system more slowly than others the number of finished task varied. However all were able to perform each task at least twice, some participants even up to three times. The number of different possible task sums up to 21 (19 unique functions and the two tasks 'skip two channels left' and 'skip two channels right'). The last ST is explained and, if understood and each possible way of interaction was tested at least once (or as often until she/he felt comfortable), the participant is asked to drive the ST run, analogous to the first one. Data is recorded if the track is run successfully and the participant is asked to conclude the test with a PT run only, i.e. the second baseline drive. If a task was not understood, it was repeated until the participant performed the demanded task. If a mistake was made, e.g. swapping channel left instead of right, the mistake was briefly mentioned and the participant asked to correct it, i.e. 'go two channels right'.

VI. Results

The study was conducted with N=17 participants (5 female), all being licensed drivers aged from 23 to 44 years. All of them have had regular experience with touch gestures due to frequent smartphone and tablet use, however only little experience with in-vehicle touch screen interfaces and no experience at all with interaction via mid-air hand gestures. Of the 17 participants, 9 conducted the LCT by first interacting via touch gestures (group T), while 8 participants began with mid-air gestures (group M) as displayed in Fig. 6 and 7. A few things can be noted when comparing these findings. The average median deviation (mdev) for baseline 1 (B1) is 0.33 with an improvement to 0.31 for baseline 2 (B2) averaged over all participants. This means that on average the participants performed better on the PT towards the end of the LCT. Averaged over all participants, performance on the PT degrades, as both ST A and ST B are at about 0.56 and 0.53 - an increase of more than 0.02. The highest mdev is measurable for P3 from group T (approx. 1.05) for the ST of mid-air interaction. The second and third highest mdev is measured for P2 and P6 both of group M, namely 0.89 and 0.79 during mid-air and touch interaction respectively. In all three cases, each participant missed a sign, naturally resulting in increased mdev values. Overall, 4 of 17 participants (P1, P3, P5, P6) from group T each missed a sign completely due to distraction, resulting in no lane change or in one case a change to a wrong lane and therefore in a large mdev. 2 signs were missed when performing touch gestures and

consequently 2 were missed performing mid-air gestures. Removing these 4 participants for the overall results yields a different average mdev for the STs, as they are reduced to 0.49 (ST A) and 0.50 (ST B) respectively.
More interesting observations can be made when looking at group T and M individually. Figure 6 presents the results with respect to the median deviation for each of the four test drives for the 9 participants from group T.

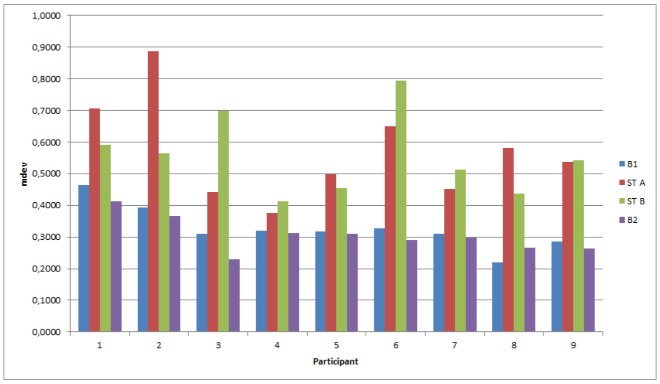

Fig. 6. Mdev results for the 9 participants in Group T (starting with touch gestures as secondary task).

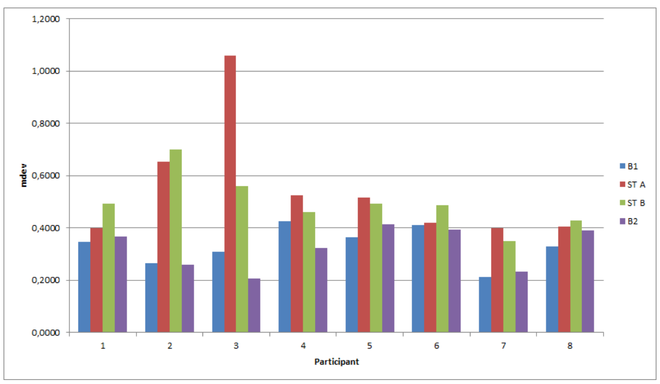

Fig. 7. Mdev results for the 8 participants in Group M (starting with mid-air gestures as secondary task).

First of all, an improvement is measurable when considering the performance of the Baseline 1 (B1) versus Baseline 2 (B2) for group T, as the median deviation from the optimal drive is reduced for all but participant 8. The same effect cannot be determined for group M as a slight decrease in baseline performance is measurable for 4 from 8 participants (cf. Figure 7). However, two things have to be noted: Firstly, in those cases in which a decrease is measured, it is of relatively low significance ($<$0.05). The performance for the baseline drive for group T improves from an average 0.33 to 0.31 and for group M from 0.33 to 0.32 meaning a slightly improved PT performance is measurable however not above a significant threshold. For the ST several observations can be made. In group T, 3 participants show a strong increase in performance and 1 shows a moderate improvement (see Fig.6). These are the four participants less distracted by mid-air gestures but rather strongly distracted by the gaze diversion during touch interaction. More specifically, if a person handles the cognitive load of learning new poses well, driving performance declines only slightly during mid-air interaction. Moreover, 3 participants with mdev values around between 0.4 and 0.5 for touch gestures (P4, P7, P9) show only minor decrease in performance during mid-air interaction. As can be seen from Fig. 7, only P3 had problems with mid-air interaction. After the test, the participant concluded that the additional effort of having to memorize the hand gestures was too high, simultaneously adding that she/he is very much used to touch interaction in the car on a daily commute basis. For the other participants the mdev values for touch and mid-air interaction are similar. This is also reflected in the overall mdev values for touch and mid-air interaction, as, when removing the cases where the participants were confused, the decrease in driving performance is similar for both STs. In this case the mdev value for performance during touch interaction overall is around 0.50 and during mid-air interaction around 0.49, a slightly better performance.

VII. Conclusion

Aside of the obvious fact that the ability to change and keep lane is strongly influenced by both means of interaction, the results show slight superiority in terms of performance when interacting with touch gestures. However, this is no longer the case if the four participants who made mistakes during the test are removed from the analysis, as mid-air gestures then seem to be less distracting. Missing a sign during the LCT results in greatly increased mdev scores, even greater if the wrong lane is chosen, as occurred once during this study. This is significant, since, as mentioned before, none of the participants has had experience with mid-air hand gesture interaction. However, all participants are frequently interacting with different kinds of touch devices, i.e. this is a well-established concept for them. Therefore, it is reasonable to assume that after a longer introduction period the performance during interaction via mid-air hand gestures is likely to improve, possibly way beyond the performance during touch interaction. This is further corroborated by the fact that all participants, except for 1 who is an avid touch interface user, found the interaction with hand gestures more intuitive and less distracting, especially emphasizing the fact that no gaze had to be redirected during the test drives. Moreover, the task of mid-air gesture interaction was more demanding for the participants as 6 new hand gestures had to be learned and employed after a brief introduction period. Conversely, interaction by touch gestures is a well-known principle for so many people nowadays that it is nearly impossible to find uninfluenced participants, hence the perceived additional load in this case was near zero as was confirmed by all participants. Conclusively, one can make the assumptions, that the measured decrease of performance during touch interaction must stem from the fact that the driver is forced to redirect gaze from the main task of following the road to the ST of interaction with the infotainment system. Simultaneously, the decrease in performance during mid-air gesture interaction stems from the additional effort imposed by having to learn a novel interaction technique. The authors of this contribution are confident that this will manifest itself in improved scores when conducting further, more extensive tests, letting the participants get properly acquainted with the modern technology of mid-air hand gestures in a car scenario. This makes, in our opinion, a strong argument for the integration of hand gestures into the automotive environment. It furthermore provides a good basis for further research in this direction as mid-air hand gestures seem to be an intuitive, lightweight and easy-to-learn means of interaction. Building on the findings in this contribution, further studies will address the exact influencing factors stemming from the positional change of the infotainment system compared to those stemming from the change of interaction means. Future work will also address the intuitiveness of mid-air hand gestures and its measurability.

References

[1] M. Zobl, M. Geiger, K. Bengler, and M. Lang, "A usability study on hand gesture controlled operation of in-car devices," *Abridged Proceedings, HCI*, pp. 5–10, 2001.

[2] B. Pfleging, S. Schneegass, and A. Schmidt, "Multimodal interaction in the car: combining speech and gestures on the steering wheel," in *Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM, 2012, pp. 155–162.

[3] S. Loehmann, M. Knobel, M. Lamara, and A. Butz, "Culturally independent gestures for in-car interactions," in *Human-Computer Interaction–INTERACT 2013*. Springer, 2013, pp. 538–545.

[4] M. Geiger, M. Zobl, K. Bengler, and M. Lang, "Intermodal differences in distraction effects while controlling automotive user interfaces," in *Proc. HCII*, 2001, pp. 263–267.

[5] T. Kopinski, S. Geisler, L.-C. Caron, A. Gepperth, and U. Handmann, "A real-time applicable 3d gesture recognition system for automobile hmi," in *Intelligent Transportation Systems (ITSC), 2014 IEEE 17th International Conference on*. IEEE, 2014, pp. 2616–2622.

[6] C. A. Pickering, K. J. Burnham, and M. J. Richardson, "A research study of hand gesture recognition technologies and applications for human vehicle interaction," in *3rd Conf. on Automotive Electronics*. Citeseer, 2007.

[7] T. Döring, D. Kern, P. Marshall, M. Pfeiffer, J. Schöning, V. Gruhn, and A. Schmidt, "Gestural interaction on the steering wheel: reducing the visual demand," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2011, pp. 483–492.

[8] T. Kopinski, A. Gepperth, and U. Handmann, "A time-of-flight-based hand posture database for human-machine interaction," in *International Conference on Automation, Robotics and Computer Vision*. IEEE, 2016, p. to appear.

[9] ISO, "Road vehicles – ergonomic aspects of transport information and control systems – simulated lane change test to assess driver distraction." International Organization for Standardization, Geneva, Switzerland, Tech. Rep. ISO 26022:2010, 2010.