Free-hand Pointing for Identification and Interaction with Distant Objects

Sonja Rümelin[1,2], Chadly Marouane[2], Andreas Butz[2]
[1]BMW Group Research and Technology, Munich, Germany
[2]University of Munich (LMU), Munich, Germany
{sonja.ruemelin, chadly.marouane, butz}@ifi.lmu.de

ABSTRACT

In this paper, we investigate pointing as a lightweight form of gestural interaction in cars. In a pre-study, we show the technical feasibility of reliable pointing detection with a depth camera by achieving a recognition rate of 96% in the lab. In a subsequent in-situ study, we let drivers point to objects inside and outside of the car while driving through a city. In three usage scenarios, we studied how this influenced their driving objectively, as well as subjectively. Distraction from the driving task was compensated by a regulation of driving speed and did not have a negative influence on driving behaviour. Our participants considered pointing a desirable interaction technique in comparison to current controller-based interaction and identified a number of additional promising use cases for pointing in the car.

Categories and Subject Descriptors

D.3.3 [**Infomation Interfaces and Presentation**]: User Interfaces – *interaction styles.*

General Terms

Design, Human Factors.

Keywords

Pointing; Gesture interaction; Camera-based tracking.

1. POINTING IN CARS

Our parents told us not to point a finger at people. However, this is often the quickest way to identify something in human conversations, and its potential for Human-Computer interaction has been shown extensively, starting with [3]. In the vehicle, the driver should not be distracted from the street [14], but concentrate on the primary driving task. For pointing, this means that the time, during which the driver's hand leaves the steering wheel, as well as the visual demand for performing the task, must be minimized. Therefore, we are looking for a set of short, yet meaningful interactions, which do not restrict the user to just trivial tasks. We identify several use cases, in which pointing in the car supports direct interaction with the environment and enables meaningful new applications.

Figure 1. In this paper, we present how a user can point from the inside of a vehicle to objects outside to easily select them. Subsequently, an interaction with this object can take place.

When we walk or drive, we have plenty of points of interest (POIs) around us. This includes not only museums or restaurants, but generally everything that may be interesting for the user, e.g., a nice spot or the house of a friend. Current navigation systems can only search for POIs around a position or along a route, but not for results in a certain direction, such as *"I know back there, there is this little café"*. A search for cafés might bring up many unneeded results, most of which could have been excluded. When we can point towards a certain direction, this serves as an effective filter to adequately narrow down the results. We can also point to get more information about a POI, or select it to perform further interaction with it, e.g. make a call with an associated number, save the address for later use, mark a POI with a tag such as *"accident"* or *"nice spot"*, or use the location to *"check in"* with social services like Foursquare or Facebook.

The contribution of this paper is an analysis of pointing in the vehicle. First, we looked at the specific requirements of pointing detection. We then designed an interaction concept (see Figure 1), that covers several scenarios in which pointing is used to select objects, directions or areas. A subsequent in-situ evaluation of our concept examines the use of pointing while driving and how it supports natural interaction with objects in the environment.

2. RELATED WORK

2.1 Gestural interaction in cars

The use of gestural interfaces for in-car interaction can reduce visual demand and thus increase safety [18]. Althoff et al. [1] investigated 17 different hand and six different head gestures and used them for complex interactions with the in-car infotainment system, such as skipping between individual music tracks or navigating in a map. To increase the expressive power of gestures even more, their function was partially context-dependent. Mahr et al. [12] designed a reduced set of three micro-gestures, which could be applied while keeping the hands on the steering wheel. However, all of these approaches only replace existing controls in the menu structure. They do not introduce new use cases that only

become possible with gestural interaction, such as the direct interaction with objects in the environment.

2.2 Pointing gesture applications

Baudel et al. [2] used a DataGlove to control applications with gestures. They emphasize the potential of pointing gestures as a natural means for defining a location. Moreover, they highlight the potential for a short as well as substantial interaction and that pointing supports direct interaction with the intended objects.

In 1980, Bolt [3] used pointing for the selection of objects on a distant screen. He used speech commands such as "*Put that there*" to specify the interaction. Moreover, he used pointing to identify an area on the screen to which an object should be moved. Instructions by gesture have subsequently been applied in robotics [9, 16]. Often, speech commands are added to specify the intention of the gesture. Richarz et al. [17] instructed a robot to walk towards the pointed direction with "*There you go*". In mobile applications, a direction can be used to specify the spatial relation to the environment. Robinson et al. [19] presented an approach in which users receive tactile feedback as soon as they point their navigational device in the destination's direction. Moreover, pointing can be enhanced by other measures of direction. NaviRadar [20] used the current walking direction to adjust constant feedback about the overall direction towards a destination. VisionWand used pointing with a smart device to interact with distant areas on large displays [5].

Pointing is also used in augmented reality applications. Information services such as Wikitude[1] use position and orientation of a smart device to augment its camera image with additional, virtual objects based on location data. Moreover, pointing is used in virtual reality (VR) environments to select objects or to control virtual menus in 3D space [8].

2.3 Recognition approaches

There are different ways to detect pointing. For hands-only pointing, Nickel and Stiefelhagen identified three phases [15]. First, the hand moves towards the target. Then it remains still in this pointing position and finally, it moves away from it. For the middle phase, Mine [13] proposed to use either the current orientation of the user's hand or the vector from the user's head through his hand as the pointing direction. The intersection of this vector with the environment describes the object pointed to. Nickel and Stiefelhagen [15] compared the head-hand-vector to the orientation of forearm or head. In most cases, the direction described by head and hand delivered the best results. However, the vector that describes the intended direction best might also depend on the actual pointing gesture [6]. Droeschel et al. suggest basing the decision on the posture of the pointing arm, which can be either bent or stretched out for targets further away [6].

Wong and Gutwin [21] investigated pointing accuracy with a laser pointer (turned off during the pointing gesture) in a collaborative virtual environment. Targets were located in a distance of 3 and 6 meters. Gestures were translated into VR and interpreted with a mean angular error of less than 5°. Zhao et al. [22] examined pointing with a mobile phone in a distance from a few to several hundred meters. They observed a constant offset horizontally as well as vertically. Overall, they found a horizontal accuracy of 9.9°, and a vertical accuracy of 3.9°.

3. RESEARCH APPROACH

We decided for an iterative design process. We started by specifying the project goal to cover the identification and interaction with distant objects via pointing gestures. We then focused on identifying the best algorithm to detect the pointing direction, keeping in mind the specific characteristics of a car environment. Having participants at hand performing pointing gestures to verify our approach, we collected ideas for possible use cases. In a next step, those inspired brainstorming sessions in which we derived several scenarios where pointing can be applied meaningfully. Having shown that our recognition approach of using head and hand position can be used to identify the pointing direction, the next step was to conduct an in-situ-study where we focused on acceptance and the effect on driving behaviour. Overall, we cover different aspects side by side and and provide an overview of the conducted and future work as well as implications for the community.

4. POINTING RECOGNITION

For our implementation, we adapted existing recognition algorithms, as there are specific conditions in the car that help robust pointing detection. Movements of people are restricted as they are sitting and only move their hands, arms, shoulders and head. When pointing at an object outside the car, the pointing direction is restricted to the windows. Inside the car, we restricted pointing to an area on the co-driver's side of the cockpit.

4.1 Detecting pointing gestures

To robustly detect the gesture, we assume that the user points towards predefined areas (windows and cockpit) and that the hand remains stable in the pointing position for a certain time. We then detect the head position and hand position to calculate a vector in 3D space that describes the direction relative to a stationary sensor. We used a Microsoft Kinect for depth and image recognition. It was positioned in the outer right corner of the front window (see Figure 2). In a first attempt, we used the skeleton recognition provided by the OpenNI framework[2] that separates moving objects from the background and identifies them as a certain body part. For this method, it is sufficient to only see the upper body.
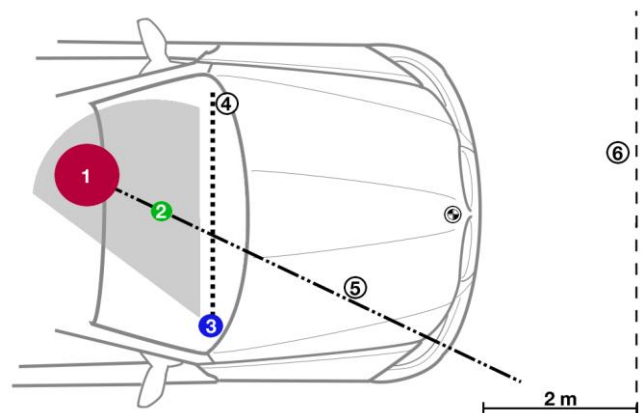

Figure 2. Recognition setup in the car. The inside dimensions of the vehicle were also replicated in the lab setting. (1. driver's head 2. pointing hand 3. depth camera 4. window area 5. pointing direction 6. projection plane)

However, as soon as fore- and background cannot be separated correctly, the method fails. The second attempt isolated head and hand tracking. For the head position, we built upon the work of [4, 7] and used the characteristics of a human head (size of about 20 cm x 15 cm x 25 cm) to identify it in a predefined area of the depth image. For the hand position, we used a method provided by OpenNI that uses optical flow detection initiated by a waving gesture. These procedures require periodical recalibration, so we chose a third approach in which we used the previous head tracking. By scanning through the depth image, we classify all neighboring pixels with depth values similar to the head as *user pixels*. A pointing hand can be identified based on the height and width of the recognized blobs, and the pixels' distance from the head. This procedure is computationally intensive, but independent of frameworks or hardware and can, depending on the resolution of the depth sensor, be used to even detect forefinger details.

4.2 Evaluation of pointing

We conducted a first user study to test our algorithm and to get initial feedback about pointing in the car.

4.2.1 Experimental design

During development, we found that the setup in the actual car caused problems for the Kinect's depth recognition, as the space inside is narrow and several objects are positioned between the driver and the camera (rear mirror, parts of the cockpit). Therefore, we decided to split the study into a lab and an in-situ part. We chose a split-plot design and applied a between-subjects design for the independent variable **setup** (*lab*, *car*). The inside dimensions of the car were replicated exactly in a lab setup, without the disturbing objects. In this way, we could gather data on the robustness of tracking in the lab, but also conduct other tests in the target environment. Figure 3 shows the setup inside the car. We used three **street scenes** with 5 highlighted **POIs** each. The order of the scenes was counterbalanced for a repeated-measures design using a latin square, the order of POIs was randomized. In a second part of the study, we investigated the tolerable **pointing duration**, needed to unambiguously detect the gesture (*600, 800, 1000 and 1200 ms*). We also investigated pointing performance to the co-driver's side of the cockpit. We used 8 predefined **areas** in randomized order. Audio feedback was given when a pointing gesture was detected. It consisted of a short sound played every time a gesture was detected. It was loud enough to be heard while talking, but not disturbing.

4.2.2 Procedure

18 participants, not involved in the current research, took part in the study individually. Their age ranged from from 23 to 33 (mean = 27) and they had an average body height of 175.3cm (sd = 6.7). First, they were introduced to the topic of the study. This was followed by a short training phase of pointing towards the window and cockpit. After that, they were presented with the first part of the study, in which three city scenes were projected to the wall in front of them. Buildings were highlighted one after another and participants were asked to point to them. In the second part, they experienced different recognition durations while pointing at a static object in a city scene. Third, participants were asked to point at numbered areas on the cockpit surface. Numbers were prompted by the investigator. The study was closed with a semi-structured interview about subjective impressions, perceived limitations and potentials of pointing gestures in the car. All steps were videotaped for later analysis. The study took about 30 minutes.

Figure 3. View from inside the vehicle. Participants were presented street scenes via projections (background). Inside the car, cockpit areas were marked with numbers (right).

4.2.3 Results

Two data sets had to be removed due to logging errors. We report the results of the remaining 16 participants. Table 1 and Figure 4 present the results of pointing accuracy in the street scene. As expected, tracking in the car setup caused a high rate of failure due to the limited performance of the Kinect with narrow and cluttered foregrounds. In the lab setup, an average hit rate of 95.8% could be achieved, indicating that the calculation of the pointing direction via head position and hand position works when the sensor is tracking the positions correctly.

		Scene 1	Scene 2	Scene 3
Car setup	Hits	22/40	15/40	22/40
	Hit rate	55.0%	37.5%	55.0%
Lab setup	Hits	39/40	38/40	38/40
	Hit rate	97.5%	95.0%	95.0%

Table 1. Results of pointing performance in the street scene.

Figure 4. Visualization of pointing performance outside the car. This example shows the hits for scene 1 in the lab setup.

Participants rated the tolerable duration of remaining in a stable position while pointing (middle pointing phase) (Table 2). We found a significant correlation between duration and tolerance (rho = 0.30). Hold times need to be long enough to ensure unambiguous pointing [11], which was around 800 ms for our soft-/hardware setup. It is obvious that the shorter the holding phase, the better for the user. Our results serve as an orientation and show our hold time should not influence the results negatively.

600 ms	800 ms	1000 ms	1200 ms
2.00	2.11	2.55	2.77

Table 2. Subjective feedback on tolerable pointing duration. (1 = very tolerable, 5 = not tolerable)

For the interior pointing task to the cockpit dashboard, in only 24% of trials, the correct area was detected. Due to the proximity of the hand to the camera, depth image values were not usable. Nevertheless, participants got an impression of what in-car pointing gestures could look like and gave valuable subjective feedback discussed below.

4.2.4 Conclusion
Despite the tracking problems in the actual car, we could verify our approach for pointing (to detect the direction via head position and hand positions) in a car-like setup. Moreover, interviews confirmed the usefulness of audio feedback that informed participants if a gesture was detected. We learned that the shorter the hold time, the higher is the tolerance. However, if this prevents unintended system reactions, a certain hold time was found to be acceptable. Feedback on the task of selecting an object in a street scene was positive. Pointing was described as a natural interaction modality.

Moreover, different ideas for use cases emerged during the study. Pointing to objects in a street scene to get additional information on associated location data was said to be a valuable use case. Additionally, participants named street signs, other vehicles, and scenes of accidents or other lanes as possible pointing targets. Subsequent actions are the gathering of more information, storage of a POI for later or tagging it for social communities. Furthermore, abstract associations such as pointing towards the sky to get weather forecast were named. Concerning pointing in the interior, the co-driver's area was mainly associated with storage for rare or later use. This could be used for POI information of outside objects, but also for digital content from the vehicle's infotainment system.

5. IN-SITU EVALUATION
In the main in-situ study, we focused on acceptance and applicability in a real life scenario.

5.1 Introducing the Scenarios
We picked three different scenarios. In the **first scenario** users had a certain idea of what to look for – a restaurant with a decent rating. They had to drive through a certain quarter and find a nice looking restaurant with a rating of 4 stars or higher. Moreover, they should make a reservation for the next evening by calling the associated number of a given restaurant. The **second scenario** gave more flexibility; participants should point at different sights in order to prepare a sightseeing tour for a visiting friend. Here, not only the name, but also further information such as opening times and entrance fees are presented and influence the decision of integrating it into the tour. The **third scenario** was to memorize certain buildings. When looking for a new flat, people can use real estate websites and set various filters to get a list of available objects. A different way to set these filters could be to drive around the target area and simply point at appealing houses. You can then get instant information about these buildings and use their characteristics to filter your online search. A second interaction step is needed to save an identified object after pointing. We offered two methods: users can either use a menu on the central display in combination with a multifunctional controller, or use a follow-up gesture by pointing towards the glove compartment on the co-driver's side, to "store" the object for later use.

5.2 Participants
15 participants (13 male, 2 female) with a mean age of 27 took part in the study. They are working for BMW Group Research and Technology but are not involved in gesture research. On average, they are driving 10.000 km per year, and predominantly characterize their driving style as calm. All of them had to have undergone an internal driver training that ensured they could react safely to potential issues emerging during the study. Two of them were left-handed. All except one participant were already familiarized to the use of a central display in combination with a multifunctional controller. About half of them are using location-based services on a regular basis to search for specific locations such as shopping or parking spaces or in general for interesting places.

5.3 Experimental design
The main factors we wanted to examine were the acceptance and applicability of pointing in the car in the three scenarios above. Moreover, we wanted to gain insights into the effects on driving behaviour as well as patterns of glances that appear when pointing at something outside and inside the car while driving. From the results of our first study, we had found that pointing recognition in the narrow cockpit area of a vehicle did not work properly with our inexpensive depth sensor. A further restricting factor was the Kinect's vulnerability to the sunlight's infrared components that could not easily be filtered out in our test car. Therefore, we decided for a wizard-of-Oz study in which the tracking module of the application was replaced by the investigator in the rear part of the car. Attention was paid to make sure the wizard's recognition performance was similar to the results of the first lab study. Different typical types of errors of detection were included into the tracking to imitate a realistic scenario. Every participant experienced each error at least twice. We decided to use one route layout for all participants (see Figure 5). Our goal was to gather qualitative insights rather than deciding if one scenario is better than the other, hence the lack of counterbalance.

The route included a section in which participants got used to driving and to the pointing gesture. Each scenario was fitted to a specific section of the route where restaurants, sights and nice houses, respectively, could be found easily, and that included parking lots for the interviews after the completion of each scenario. For the second part of the third scenario, in which participants should either use gesture or controller input, the order was alternated. The quantitative dependent variables were pointing behavior and duration as well as driving speed. However, the main focus was on the qualitative evaluation.
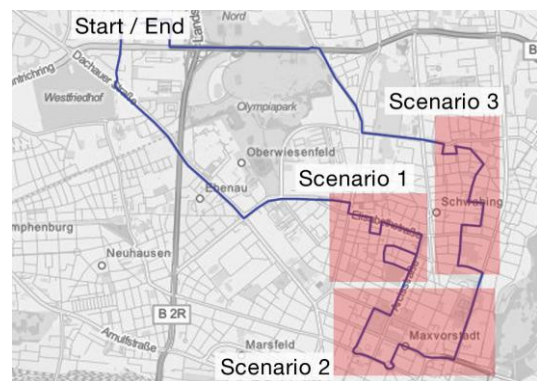

Figure 5. Route layout used for the in-situ evaluation.

Figure 6. Screenshots of the prototype. Top left: The system is continuously searching for a new pointing gesture. Top right: As soon as a pointing gesture is detected, a geodatabase like Qype is checked for available object data. Bottom left: Available information such as opening hours can be presented. Bottom right: The object can be tagged and forwarded.

5.4 Study setup

The study was conducted while driving, with the instructor sitting in the rear part of the car. We used a BMW 7 series vehicle and, according to the wizard-of-Oz methodology, set up the hardware components for gesture recognition as if everything was working (Figure 7). Next to the Kinect sensor on the front right dashboard, we installed a video camera to record the driver from the front. Moreover, the driver was equipped with a Dikablis eye-tracking system from Ergoneers. Its glasses integrate two cameras. One is directed at the pupil of the right eye to track the viewing direction. The other one is directed towards the field of view of the person wearing it. As a result, it gives back a live stream of images in which the spot the person is looking at is highlighted in the front view camera image. Informal preliminary tests had shown that with this video stream and the view from the back of the car, the wizard could identify the objects chosen for pointing accurately. They were then selected by him on a specially tailored map application that communicated with the in-car system. Information for this object was presented in the standard central information display, with a resolution of 1280 x 480 pixels. Since we wanted to integrate our application seamlessly into the on-board system to make it feel like a realistic feature of the car, we decided to match layout and main interaction features to the latest iDrive system (see Figure 6). The main interaction during the study was performed via pointing. In all other use cases, rotating and pressing of the controller in the center console was needed.
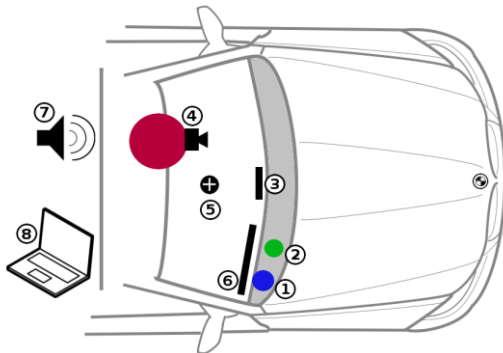

Figure 7. Hardware setup for the in-situ study. (1. Kinect sensor 2. video camera 3. central information display 4. eye-tracker 5. central controller 6. storage area of the cockpit 7. speaker for audio feedback 8. laptop for wizard-of-Oz input)

5.5 Study procedure

Participants took part individually. They were introduced to the topic of the study and the test environment in the car. They put on the eye-tracking system, and it was calibrated. Moreover, to make the tracking credible, the gesture recognition setup was calibrated. After that, participants could explore menu navigation and pointing. Moreover, they were given audio feedback at the recognition of a gesture. Before starting the tour, participants were instructed to think aloud, to get "live" feedback on situations. We used those comments to enrich the post-study interview. The study lasted for about 1½ hours. In the first part of the route, participants could get used to driving. Participants were given navigation instructions by the investigator for the whole study.

Then, the **first scenario**, finding a restaurant with a high rating, started. After that, a short structured interview was conducted. The second part of this scenario was to find a given crêperie and initiate a phone call to make a reservation. The **second scenario** took place in the museum quarter of Munich. After selecting up to ten POIs, another structured interview phase took place. Before starting the next scenario, the participant should navigate to the list of POIs and forward the last selected sight tagged as "worth seeing" to Facebook. In the **third scenario**, participants were driving through a residential area with large buildings and could look for places they would want to live. For saving a POI, they either used a pointing gesture to the glove compartment area of the cockpit, or the central controller for menu navigation. After the subsequent structured interview, they were asked to delete the first entry. The additional parts of each scenario were added to clarify possible use cases for pointing gestures and should highlight that the system is, at least for now, integrated in a controller-navigated system. Moreover, we wanted to investigate distraction during a complex subsequent interaction.

The study was closed with a semi-structured interview about the system as a whole and gave room for general questions of the participants. Positive and negative experiences during the study were collected and a short version of the AttrakDiff [10] as well as a demographic questionnaire was completed.

5.6 Results and Discussion

Results are reported at a .05 level of significance. If not stated otherwise, the questionnaire used 6 point Likert scales.
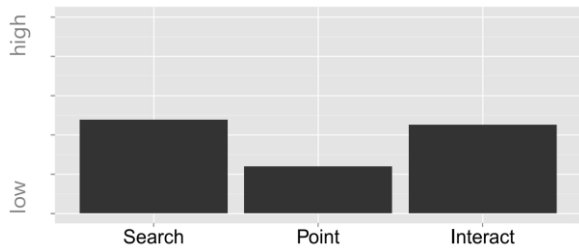
Figure 8. Perceived distraction of the single interaction steps during a pointing gesture over all scenarios.

5.6.1 Subjective Distraction

In the interviews immediately after the scenarios, participants rated their subjective impression on distraction while pointing. They commented that especially the second scenario, looking for sights, was very close to reality and allowed a spontaneous interaction. Participants rated the second and third scenario less distractive than the first which may be caused either by the more realistic tasks or by the order we could not counter balance. Moreover, restaurants in the first scenario were sometimes hidden in a row of shops and stores. In contrast, the second and third scenario required the selection of arbitrary large objects that might have been easier to localize. Therefore, the searching process before pointing might have been more difficult and distracting in the first scenario. In addition, in the second and third task, participants were already more accustomed to pointing. A more detailed analysis of pointing in all scenarios showed significant differences between the single interaction steps and their potential for distraction ($\chi^2(2) = 2.71$) (see Figure 8). A follow-up Bonferroni test showed that the first step, *searching a POI*, causes significantly more distraction than the second step, *pointing at the POI*. A possible explanation for the high value of searching is the study setting. Participants did not act self-determined and pointed when they saw an interesting POI. They were rather forced to find certain POIs depending on the scenario.

5.6.2 Comparison of Pointing and Controller

In the real estate task, participants either used a pointing gesture to the glove compartment area or menu interaction with the central controller for saving a building. Every participant used each option three times subsequently, the starting order was counterbalanced. Figure 9 shows the rating of perceived distraction for each modality. A Wilcoxon test showed a significant difference (T = 5.00), indicating that gestural interaction is less distracting than controller interaction. When asked which modality they would favor, participants agreed on the pointing gesture (100%), although most of them are experienced with controller interaction. Despite the lack of haptic feedback when pointing, and the limited functionality (only saving of the selected object was available with the gesture), the main positive
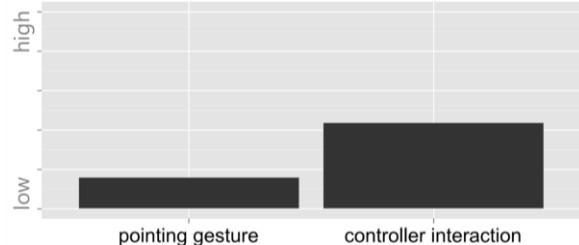

Figure 9. Perceived distraction while saving an object in scenario 3 by pointing gesture or controller interaction.
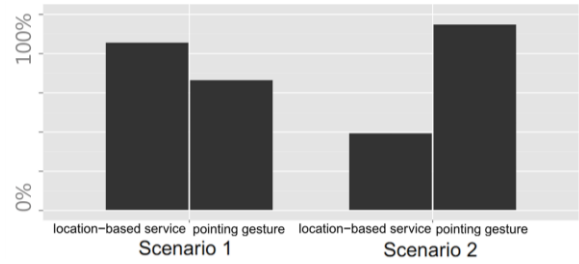

Figure 10. Comparison of interaction with either existing location-based service pointing gesture.

arguments for pointing were 1) the direct connection to the selection process, 2) that there was no need to change the modality between selecting and saving, and 3) easy learnability, because the idea of storing something in the given area seemed obvious.

5.6.3 Comparison with Location-Based Services

To assess whether the unique feature of interacting directly with the environment provides an advantage over existing applications, participants were asked to compare pointing to the location-based services (LBS) that are integrated into latest car generations already. Taking the example of *Google Local Search* which was familiar to all participants and which was demonstrated via screenshots, participants rated the usefulness of both pointing and the LBS. Figure 10 shows that the result depends on the use case, and a combination of both was suggested. An LBS was favored when planning ahead, as in the POI search in the first scenario. Pointing was favored in the spontaneous interaction with an object nearby as in the second scenario, the sightseeing task. Here, the direct interaction possibly helps to associate information with a POI and memorize it.

5.6.4 Error Acceptance

Participants were presented with three error types during the study. First, pointing gestures were *not detected* despite their correct execution. Second, *no object* was presented after pointing. Third, *multiple objects* were presented as a result of an ambiguous gesture. When asked whether an error would discourage them from using the system, participants showed a high fault tolerance (see Figure 11). When multiple objects were recognized, a list was displayed. Participants suggested applying personalized filters so that only one object remains, or in case of multiple results, to add images to the list view.

5.6.5 User Experience

The analysis of the results of the AttrakDiff questionnaire shows that the overall score is located in the top right corner *desired* (see Figure 12).
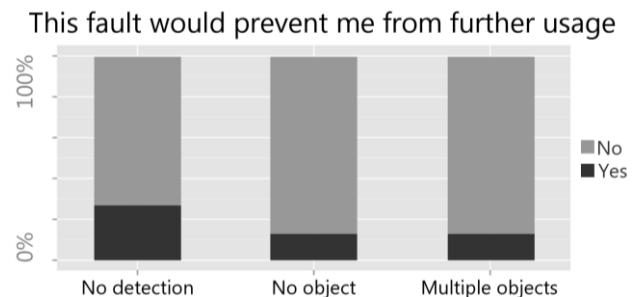

Figure 11. Fault tolerance for different possible errors in pointing interaction.
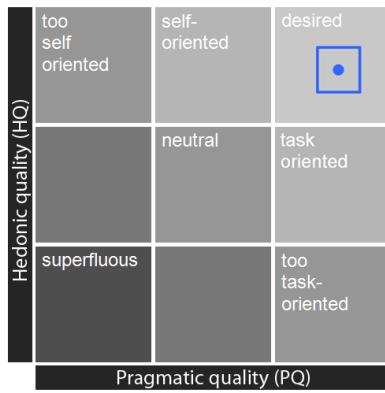
Figure 12. AttrakDiff. The participants rated the idea of pointing for interaction with contrary pairs of adjectives to evaluate the perceived hedonic and pragmatic quality

Pragmatic quality (PQ), which explains whether or not the system fulfills functional goals by providing useful and usable means is rated high, as is hedonic quality (HQ), which describes whether it helps to fulfill individual needs, such as the desire to improve oneself or to communicate with others.

5.6.6 Driving Behaviour

To measure the effect of pointing interaction on driving speed *before*, *during* and *after* a gesture, speed samples were taken from CAN every 200ms. Results are based on the second scenario in which participants were more accustomed to pointing than in the first scenario. In the third scenario, overall driving speed was very low, since participants were driving in a quiet residential area. Overall, 97 gestures are included in the analysis. Pointing is mostly executed while standing (see Figure 13).
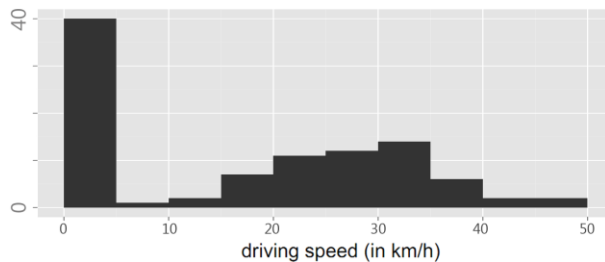

Figure 13. Speed distribution during pointing gestures in scenario two (number of occurrences).

Most of the time, participants slowed down before or during a gesture (52%) or they kept their current speed (33%). Figure 14 visualizes a segment of log data in which the decrease of speed at the moment of pointing becomes apparent. In only 15% of cases, speed increased slightly. However, with a maximum speed of 47 km/h, participants always remained below the current speed limit (50 km/h). These results suggest that when pointing gestures are used, drivers adjust their speed accordingly if they are not already standing during interaction to ensure safety. Participants commented that *concentrating* on objects outside the car was the distracting part during the tasks. However, this referred mainly to the *search phase* and was said to be similar than when looking for an object without the possibility to point. According to this, slowing down is already part of the everyday experience for following traffic and does not introduce new issues to solve. Participants did not perceive that their driving performance degraded while pointing because the glances needed for pointing did not tear their visual attention off the road *"for more than a*

short glance". This indicates that adjusting speed is a subconscious action not disturbed by the pointing gesture itself.
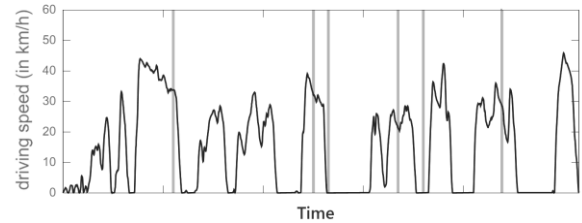

Figure 14. Varying driving speed with pointing events (vertical lines). Example of log data in second scenario.

5.6.7 Glance Behavior

By tracking eye movements, we could analyze pointing behavior in detail. Data from the third scenario was taken, in which the saving task caused additional distraction. 63 samples are included in the analysis. Similar to the three phases of hand movement in [15], we found a repeated pattern in the execution of a pointing gesture. It consists of three phases (see Figure 15), the last of which is not always executed. An *information glance* is happening before a gesture. The driver is looking away from traffic to the objects that will be selected. Then, the arm is brought into the pointing position. As soon as the arm is in a fixed position, the driver looks back to the front. Next, a *control glance* is executed to ensure the position. For this, the focus moves to the object outside again. If necessary, the position is adjusted. Then, the driver looks back towards the front. In 43% of the gestures, the interaction was finished here. In the rest of the samples, a *further control glance* was executed that followed the same procedure as the first one. In the saving step in the third scenario, there were differences in glancing behavior regarding gesture or controller interaction. 75% of the gestures were confirmed with one control glance. In the remaining 25% participants did not look at all when pointing in the interior. In contrast, controller interaction caused mainly one or two control glances (48% each).

5.6.8 Comments and Observations

In general, pointing was considered a natural way of interaction. Drivers point at POIs already and this might only become critical when other people are located in the pointing direction and could see and misunderstand it. It did not appear to be a problem during the study that objects are moving during the pointing procedure, because of high speed or close distance. People could quickly enable knowledge they have from pointing in other situations.

6. LIMITATIONS

The prototype in the study supported neither speech input nor output. This was seen as a promising addition leading to a potentially even more natural interaction. In contrast, the combination of natural gestures and conventional controller interaction was regarded as negative by some. Since we used the Kinect as a depth sensor, we were not able to realize robust gesture recognition in the car. Therefore, we focused on the effect on driving and subjective feedback. For the realization of tracking in the car, it is crucial to use advanced sensors such as time-of-flight cameras [6], to achieve similarly good sensor data quality as in the lab. We expect our detection algorithm to work even better with such a high resolution depth sensor. Since it was not possible for the wizard to distinguish several directions targeted at the cockpit reliably from the rear seat, we had decided to use only one pointing area in the cockpit for interior pointing. Here, a depth camera with a low minimum distance is required.

Figure 15. Procedure of pointing. Participants are first looking towards the object they want to select (left). Then they bring their arm into the pointing position and hold it there (middle). A second and sometimes third glance is used to ensure the correct arm position (right). After that, the arm is brought back to the steering wheel.

7. CONCLUSION AND FUTURE WORK

This paper introduced an interaction concept that uses a pointing gesture to identify distant objects and allows for subsequent interaction in an automotive environment. First, we investigated different detection setups and algorithms for the special case of pointing in a car. A lab user study with a stationary setup yielded valid algorithms with a recognition rate of about 96%. Moreover, we gained first insights into use cases for pointing gestures and, on this basis, developed several scenarios. Because of technical problems with the sensor used, we decided for a wizard-of-Oz study to further investigate our interaction concept. The main user study revealed the potential and some limitations of pointing interaction while driving. Overall, we showed that the basic idea of pointing was considered useful in three use cases. It was used to select POIs and get further information, select buildings for a sightseeing tour, and choose favored real estate objects. Our results on acceptance for gestural interaction, i.e. the need to provide additional benefit like making use of the pointed direction, and the seamlessness of subsequent interactions, can serve as a basis for future development. Moreover, our results give insights into compensation strategies when pointing while a primary driving task. Participants were able to issue the gesture while driving and self-regulated safety issues, related to distraction from their primary driving task, by decreasing driving speed. Analyzing eye-tracking data revealed that for most pointing gestures, two short glances are directed towards the selected objects, before and after positioning the hand, with an occasional third glance to verify the result. The time in-between was used to monitor traffic.

In our future work, we want to implement gesture detection in the actual car with a sensor that is more robust to sunlight and disturbing objects. To further assess the potential of pointing, we need to investigate how it can be integrated into existing infotainment systems and conduct long-term studies to evaluate the use cases in their daily use, and in combination with other modalities like speech. We are positive that pointing interaction can be unambiguously tracked even in a car and that it can serve as a promising modality for various existing and future use cases, not only in an automotive environment.

8. REFERENCES

[1] Althoff, F., Lindl, R., and Walchshäusl, L. Robust Multimodal Hand- and Gesture Recognition for Controlling Automotive Infotainment Systems. In *Proc. VDI: Fahrer im 21. Jh.*, (2005).

[2] Baudel, T. and Beaudouin-Lafon, M. CHARADE: remote control of objects using freehand gestures. *Comm. ACM, 36*(7): 28-35, 1993.

[3] Bolt, R. A. 'Put that there': Voice and Gesture at the Graphics Interface. In *Proc. SIGGRAPH 1980*, ACM (1980), 262 - 270.

[4] Burgin, W., Pantofaru, C., Smart, W. D. Using depth information to improve face detection. In *Proc. HRI 2011*, ACM (2011), 119–120.

[5] Cao X. and Balakrishnan, R. VisionWand: interaction techniques for large displays using a passive wand tracked in 3D. In *Proc. UIST 2003*, ACM (2003), 173–182.

[6] Droeschel, D., Stückler, J., and Behnke, S. Learning to interpret pointing gestures with a time-of-flight camera. In *Proc. HRI 2011*, ACM (2011), 481.

[7] Garstka, J. and Peters, G. View-dependent 3D Projection using Depth-Image-based Head Tracking. In *Proc. Workshop on Projector-Camera Systems*, IEEE (2011), 52–58.

[8] Heidemann G., Bax I., and Bekel, H. Multimodal interaction in an augmented reality scenario. In *Proc. ICMI 2004*, ACM (2004), 53–60.

[9] Holzapfel, H., Nickel, K., and Stiefelhagen, R.. Implementation and evaluation of a constraint-based multimodal fusion system for speech and 3d pointing gestures. In *Proc. ICMI 2004*, ACM (2004), 175–182.

[10] Hassenzahl, M. and Monk, A.. The Inference of Perceived Usability From Beauty. *Human-Computer Interaction*, 25(3): 235–260, 2010.

[11] Kortum, P. HCI Beyond the GUI: Design for Haptic, Speech, Olfactory, and Other Nontraditional Interfaces. Morgan Kaufmann Publishers Inc, San Francisco, CA, USA, 2008.

[12] Mahr, A., Endres, C., Müller, C., and Schneeberger, T. Determining Human-Centered Parameters of Ergonomic Micro-Gesture Interaction for Drivers Using the Theater Approach. In *Proc. AutoUI 2011*, ACM (2011), 151–157.

[13] Mine, M. Virtual environment interaction techniques. UNC Chapel Hill Computer Science Technical Report TR95-018. 1995.

[14] National Highway Traffic Safety Administration. Visual-Manual NHTSA Driver Distraction Guidelines. Feb. 2012.

[15] Nickel, K. and Stiefelhagen, R.. Pointing gesture recognition based on 3d-tracking of face, hands and head orientation. In *Proc. ICMI 2003*, ACM (2003), 140–146.

[16] Park, C.-B. and Lee, S.-W. Real-time 3d pointing gesture recognition for mobile robots with cascade hmm and particle filter. *Image and Vision Computing, 29*(1): 51–63, 2011.

[17] Richarz, J., Martin, C., Scheidig, A., and Gross H.-M.. There You Go! - Estimating Pointing Gestures In Monocular *Images For Mo*bile Robot Instruction. In *Proc. ROMAN 2006*, IEEE Computer Society (2006), 546–551.

[18] Riener, A. Gestural Interaction in Vehicular Applications. *Computer*, 45 (4): 42-47, 2012.

[19] Robinson, S., Eslambolchilar, P., and Jones, M. Evaluating haptics for information discovery while walking. In *Proc. BCS HCI 2009*, ACM (2009), 93-102.

[20] Rümelin, S., Hardy, R., and Rukzio, E. NaviRadar: A Tactile Information Display for Pedestrian Navigation. In *Proc. UIST 2011*, ACM (2011), 293–302.

[21] Wong, N. and Gutwin, C. Where are you pointing? In *Proc. CHI 2010*, ACM (2010), 1029–1036.

[22] Zhao, Y., Chakraborty, A., Hong, K.W., Kakaraddi, S., and St. Amant, R. Pointing at responsive objects outdoors. In *Proc. IUI 2012*, ACM (2012), 281–284.