Benefits of Personalization in the Context of a Speech-Based Left-Turn Assistant

Dennis Orth, Nadja Schömig, Christian Mark, Monika Jagiellowicz-Kaufmann, Dorothea Kolossa, Martin Heckmann

2017

Preprint:

This is an accepted article published in Proceedings of the 9th ACM International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI ’17). The final authenticated version is available online at: https://doi.org/10.1145/3122986.3123004

Benefits of Personalization in the Context of a Speech-Based Left-Turn Assistant

Dennis Orth[1,3], Nadja Schömig[2], Christian Mark[2], Monika Jagiellowicz-Kaufmann[2], Dorothea Kolossa[1], Martin Heckmann[3]
Institute of Communication Acoustics, Ruhr-Universität Bochum, Bochum, Germany[1]
Wuerzburg Institute for Traffic Sciences, Veitshöchheim, Germany[2]
Honda Research Institute Europe GmbH, Offenbach am Main, Germany[3]
dennis.orth@rub.de, schoemig@wivw.de, mark@wivw.de, jagiellowicz@wivw.de, dorothea.kolossa@rub.de, martin.heckmann@honda-ri.de

ABSTRACT
We have previously introduced a novel Assistance On Demand (AOD) concept in the context of an urban speech-based left-turn assistant which supports the driver in monitoring and decision making by providing recommendations for suitable time gaps to enter the intersection. In a first user study participants showed a clear preference for the AOD system, yet also frequently mentioned that the recommended gaps did not fit their driving behavior. In the user study we present here, we investigate in how far the acceptance and efficiency of the AOD system can be increased by a personalization of the recommended gaps to the individual driver. For this purpose, we estimate individual drivers' gap acceptance from observations of their manual driving and use it to evaluate a default and a personalized variant of the AOD system. Results reveal a clear preference for the personalized assistant compared to the default one and to driving manually.

Author Keywords
ADAS; critical gap; gap acceptance; personalization; speech-based

CCS Concepts
•**Human-centered computing** → **User models;** *Natural language interfaces; User studies;*

INTRODUCTION
Due to the increasing number of advanced driver-assistance systems (ADAS) and thus the increasing variety of information and options delivered by these systems, the user can be overwhelmed, distracted and finally annoyed [1]. Therefore, more and more approaches which apply personalization to such systems to increase their efficiency and acceptance are being proposed [4].

AutomotiveUI '17, September 24–27, 2017, Oldenburg, Germany

ISBN 978-1-4503-5150-8/17/09…$15.00
DOI: https://doi.org/10.1145/3122986.3123004

In [7], [8] we have proposed the concept of Assistance On Demand (AOD) in the context of a speech-based left-turn assistant for urban intersections. It can be activated on demand and controlled via speech commands by the driver. While waiting at the intersection, coming from a subordinate road, the AOD system informs drivers about suitable time gaps, which should enable them to perform the left turn and vacate the intersection. We have compared the voice-controlled AOD system to driving manually and to a HUD-based (Head-Up Display) assistance system, which was always active at the intersection. The latter showed colored arrows to the user, which indicated approaching traffic from the left and/or right side on the superordinate road. The participants reported that their perceived workload decreased and their perceived safety and driving quality increased for the drive in which they used the AOD system. In addition, the AOD system was preferred to the HUD system and higher usefulness ratings were given in situations with increased traffic density. Even if a high acceptance of the AOD system was indicated, the system still had some limitations. A fixed and rather conservative time gap was used in the recommendations to the drivers. Many participants felt that the system's recommendations did not match their driving behavior. Consequently, participants remarked that they would like to be able to adjust the time gap recommended by the system.

Based on these observations, in a subsequent study we investigated the inter-individual differences of drivers' left-turning behavior [6]. In a driving simulator several participants repeatedly turned left at an intersection with traffic approaching from the left and right side. The results revealed that the gaps participants accepted showed a strong variation from participant to participant. In [6] and in [5] we furthermore developed the methodology to predict drivers' individual left-turning behavior, i.e., which gaps they accepted, from few observations of their previous driving. In the study we present here, we evaluated the hypothesis that the acceptance and efficiency of the AOD system can be increased by using personalized gap recommendations, which better fit to the individual driver's left-turning behavior.

In the next section we will introduce the concept of the critical gap, which we use as a measure for driver left-turning

behavior. Afterwards the methodology used for this user study will be described. Finally, the results of the user study are presented and analyzed, particularly with regard to the desired personalization of the AOD system.

CRITICAL GAP ESTIMATION ALGORITHM

The so-called *critical gap* is a parameter, which is often used to calculate the capacity and delay of a minor road, especially for unsignalized T-intersections. It signifies how large a gap minimally has to be for the driver to accept it and make the turn, thus vacating the minor road. Previously, this value was primarily used to measure the capacity of a specific intersection and hence was calculated from a large pool of drivers all passing this intersection [3]. In our case, however, we use it as a criterion to quantify the behavior of one individual driver. We use it to predict if a driver will make the turn given a particular intersection layout and traffic density.

One of the most frequently used methods to estimate the critical gap is the one of Troutbeck [10],[11]. He assumes that a driver's critical gap lies somewhere between the largest rejected gap and the gap that was accepted by the driver. To model the behavior of the drivers, [11] uses a log-normal distribution for the critical gaps. The transformation of the observed gaps to the logarithm domain allows the use of a normal distribution for the core computations. The critical gap can not be observed directly. It can only be observed that it lies in the semi-closed interval $(r_i, a_i]$ where r_i indicates the logarithm of the largest gap rejected by the driver at intersection i and a_i is the logarithm of the gap accepted by the driver at intersection i. The hidden parameters μ and σ of the distribution of the logarithm of the individual driver's critical gap μ_c can then be found by maximizing the log-likelihood:

$$\mathscr{L} = \sum_{i=1}^{N} \ln \mathsf{p}(r_i < \mu \leq a_i | \mu, \sigma) \,, \tag{1}$$

for all observations N. Eq. 1 can then be written as [2]:

$$\mathscr{L} = \sum_{i=1}^{N} \ln[\mathsf{F}_c(a_i) - \mathsf{F}_c(r_i)] \,, \tag{2}$$

where $\mathsf{F}_c()$ denotes the cumulative distribution function of the normal distribution with parameters μ and σ. Hence the parameters can be determined via:

$$(\mu_c, \sigma_c) = \arg\max_{\mu,\sigma} \sum_{i=1}^{N} \ln[\mathsf{F}_c(a_i) - \mathsf{F}_c(r_i)] \,, \tag{3}$$

where σ_c is the standard deviation corresponding to the individual driver's critical gap μ_c. As a final step, the critical gap t_c and variance s^2 in linear scale are computed according to

$$t_c = \mathrm{e}^{\mu_c + 0.5\sigma_c^2}, \qquad s^2 = t_c^2(\mathrm{e}^{\sigma_c^2} - 1). \tag{4}$$

Please note that the index i in the tuple (r_i, a_i) indicates the i-th intersection of the current driver. This is in contrast to the original formulation in [11], where the index i is used to denote the i-th driver at the current intersection.

Figure 1. Static driving simulator with a full-car mockup. The ego car is standing at the four-way intersection, while traffic cars are arriving from the right and left side on the superordinate road.

METHODOLOGY

To assess the possible benefits of the personalization of the gap recommendations of the AOD system, we have performed a user study in a static driving simulator.

Participants

A total of $N = 25$ participants took part in this study, 12 of them were male. The mean age was 42 years (standard deviation 13.1 years) with driver ages ranging from 22 to 64 years. The mean number of years of driving experience was 22.8 with a standard deviation of 12.0. The mean number of traveled kilometers in the previous year was 18 660 km with a standard deviation of 12 375 km.

Study environment

The study took place in the static driving simulator of the Würzburg Institute for Traffic Sciences (WIVW; see Fig. 1). The simulator is based on a full car mock-up of an Opel Insignia, for which outside rear-view mirrors are replaced by LCD displays. The scenery is projected onto five screens. The steering wheel has an integrated steering force simulator. The mock-up interior includes two integrated LCD-displays, one replacing the speedometer, the other in the center console to display optional additional information.

System specification

In the implementation for this study, the AOD system was restricted to monitoring traffic arriving from the right. Therefore, all system outputs only refer to traffic from the right side, so that traffic from the left still has to be monitored by the drivers themselves. While the driver is approaching the intersection, the driver's request (e.g. "Please watch right") activates the system. In the simulator study, the final activation of the system was triggered by a button pressed by the experimenter (this was the only manual action of the experimenter). The system confirms the successful activation by answering "Okay - I will watch." When the driver reaches the intersection, the AOD system starts giving recommendations.

If the time distance of the closest vehicle from the right to the center of the intersection is above 10 s, the system will interpret this as no vehicle being present and it triggers the output "no vehicle from the right." It was deliberately decided not to announce "right is free," as this could be interpreted

as a permission to drive without further monitoring the actual traffic and could consequently lead to hazardous situations. If a vehicle is approaching from the right and the time distance to the intersection falls below the current adjusted critical gap, while simultaneously another vehicle is following with a time gap smaller than the critical gap, the system interprets this as a sequence of vehicles, which does not allow entering the intersection for turning. The linked speech output is "vehicle from the right."

If a vehicle from the right is expected to reach the intersection within 3 s and if the time gap to the next oncoming vehicle is larger than or equal to the critical gap, the system interprets this as a suitable time gap for entering the intersection. The system output is given before the previous vehicle has passed the intersection, to create a certain preparation time. Hence, the speech output is "gap after approaching vehicle." If the time gap has elapsed and the next vehicle is approaching under the same conditions, the output is replaced by: "Gap after next vehicle." This implementation of the AOD system is identical to the one we already used in [8, 7].

Scenarios

Each experimental drive consisted of a set of several scenarios, all containing an urban intersection. As the basic layout for this scenario, a four-way intersection was chosen with the ego vehicle approaching from the subordinate road. Despite the four-way layout, we considered only vehicles arriving from the left and the right. For one drive, 13 different scenarios were put together into one driving course, meaning that the driver drove from one intersection to the next by always turning left. Yield signs were placed at the roadside. A stop line should assure that all drivers stop at a comparable distance from the entrance of the intersection. The surroundings at the intersection were created in such a way that the drivers could not see the arriving vehicles on the superordinate road while approaching the intersection. When having stopped at the intersection, the line-of-sight was about 8 s to the right and 10 s to the left (taking 50 kmh^{-1} as a basis). The instruction asked the driver to turn left at the intersection.

The participants were asked to drive under three conditions: a "manual drive" where the AOD system was not activated, a "default AOD" system drive and the "personalized AOD" drive, during which the participants were required to activate and use the left-turn assistant. In the default AOD system drive a fixed critical gap of 5.5 s was set in the system. For the personalized AOD system drive, a critical gap was calculated from the recorded data of the manual drive according to the algorithm proposed above.

Scenario definition

We have previously observed that the usefulness of the AOD system is perceived the higher, the more challenging the situation is [8]. Therefore, it was necessary to make the scenarios challenging for the participants. We can achieve this by simultaneously presenting traffic from both sides with small to moderate time gaps. Traffic arriving only from one side would have been easy to handle and would not require assistance for an experienced driver. In addition, to strengthen, on one hand, the participant's perception of different critical gaps but, on the other hand, to preserve a possible effect of personalization, only gaps of 5 different lengths for the traffic from the right were presented to the participants: 2 s, 4.5 s, 5 s, 5.5 s and 6 s. The gap of 2 s was expected to be too small for all participants. Its purpose was to allow the insertion of traffic from the left without changing the effective length of the target gaps from the right. The gaps from the left were consequently timed such that the gap to the left was always larger than the gap to the right when there was a target gap on the right. Smaller gaps from the left, and in particular the passing of the cars from the left in front of the participant, coincided with the 2 s gaps from the right. Hence, the gaps relevant to the participants for making their decision to turn were only the gaps from the right. As a result, for each drive there was traffic approaching the intersection from the left and the right side. At most 10 vehicles passed from either side. Consequently, even if none of the presented gaps were suitable for the participants, they did not have to wait too long at the intersection. Details on each intersection are given in Table 2. To force the participant to stop at the intersection, the first few gaps were smaller than 3 s. The target gaps either remained fixed for an intersection or were chosen randomly but fixed for each participant from the range of target gaps. Intersection number 13 was an exception. For this intersection, no traffic from the left was present and the gaps from the right were increasing monotonously starting from 4.5 s to 9.0 s in 0.5 s increments. The purpose of this was to create an additional, yet rather limited, possibility to measure a participant's critical gap: the gap the participant takes in the sequence. All traffic cars were driving at a constant speed of 50 kmh^{-1}.

Intersection	*Distribution*	*Range*
1	const.	5.5 s
2	const.	4.5 s
3	const.	5.0 s
4	const.	6.0 s
5	variable	4.5 −6.0 s
6	variable	4.5 −6.0 s
7	variable	4.5 −6.0 s
8	variable	4.5 −6.0 s
9	variable	4.5 −6.0 s
10	variable	4.5 −6.0 s
11	variable	4.5 −6.0 s
12	variable	4.5 −6.0 s
13	variable	4.5 −9.0 s

Table 1. Distributions of the time gaps between the traffic cars on the superordinate road.

The critical gap μ_c of a participant was calculated from the observed driving behavior at intersections 1 to 12 based on the algorithm described above and verified based on the gap taken in intersection 13. From the critical gap, the gaps the system will advertise to a specific participant were calculated. As the critical gap is the minimum gap necessary for the driver to make the turn, the system announced only those gaps which were at least as large as the calculated critical gap.

Experimental plan

All 25 participants had a 5-minute introductory drive to familiarize themselves with the driving simulation and the simulation environment. Then, the manual drive without any assistance was performed. It usually took around 10 min. Following this, the participants had to fill out a first questionnaire regarding this drive, which took around 5 min. Before conducting the drives with the activated AOD system, the participants had a 5 min introduction into the AOD functionality followed by a practice drive with the activated AOD system. After this, both AOD system drives (personalized and default) were performed. The sequence of the default AOD system drive and the personalized AOD system drive was permuted and counterbalanced and drivers were assigned at random to one of the two sequence orders. The participants were not informed on the differences between them. They were only informed on the difference between the manual and the AOD drive. The latter drives were referred to as "AOD drive 1" and "AOD drive 2". Immediately after these drives the participants again filled in a questionnaire. Finally, another questionnaire with a direct comparison between the two AOD drives was answered by the participants. In addition to the questions after every drive, the participants were asked to give feedback about the appropriateness of the time gaps after every intersection. Later in this work we refer to these answers as "online judgments."

Measures

In the questionnaires administered after each of the three drives, participants rated the required attention, difficulty, perceived demand, riskiness, feeling of safety and perceived performance. A 16-point rating scale with verbal categories from "not at all" (0), "very low" (1-3), "low" (4-6), "medium" (7-9), "high" (10-12), to "very high" (13-15) was used. After deciding for one verbal category, drivers were requested to further define their rating by numbers between 0 and 15.

For the online judgments, the 16-point-rating scale was also used. After the third and last drive, the question was posed whether they had "recognized a change of gap sizes" between those that were recommended in the first AOD drive and those that were recommended in the second AOD drive.

In a larger evaluation, also after the third drive, the participants additionally rated certain aspects of driving with the AOD system variants on a 5-point-rating scale (0 "does not apply at all" ... 4 "does fully apply"). Those aspects were: *usefulness* (useful, intuitive, easy to use, reliable, relied on system, time gaps comfortable), *workload* (relieved, facilitated monitoring, facilitated decision), *affective evaluation* (liked to use, not annoying) and *safety* (felt safe in usage, increased traffic safety). In addition, they had to judge whether the recommended time gaps during the AOD drives were too long or too short. For this, also a 5-point-rating scale (-2 "too long" ... 2 "too short") was provided.

Finally, the participants were asked to give a ranking about how comfortable they found the manual and the first and second AOD drive.

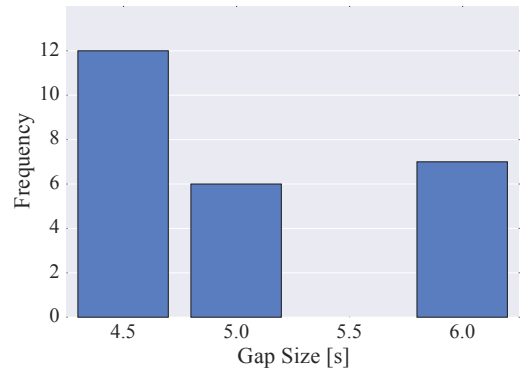

Figure 2. Distribution of individual critical gaps.

Statistical analysis

In the text and in the figures the sample statistics we state are sample mean and standard deviation. For nominal scale data we applied a Chi-squared test. Depending on the number of independent variables, we have used either a paired-sample t-test or a repeated measures ANOVA for the analysis of Likert-type scaled and continuous data. To report effect sizes we use partial eta-squared (η_p^2). For statistical significance we assume a significance level α of .05.

RESULTS

In the following, we will first look at the distribution of the driver-specific critical gaps, which have been calculated from the drivers' behavior, recorded in the manual drive. After that, we will go into the participants' subjective ratings of the system variants and look at objective measures.

Note that, since some of the questionnaires were not completely filled in by the participants, the total number of votes for the ranking and the degrees of freedom for the paired-samples t-test vary. Yet in all cases we received at least 24 out of 25 answers.

Driver-specific critical gaps

The evaluation of the calculated driver-specific critical gaps reveals that the full range of critical gaps that were possible in our scenario also occurred (compare Fig. 2). Almost half of the participants showed a critical gap of 4.5 s, the smallest possible critical gap in our scenario design. On the other end, one fourth of them showed the largest possible critical gap of 6.0 s. None of the participants showed the critical gap we had chosen as default: 5.5 s. Even though we did expect a large share of the drivers to take the 5.5 s gap, these results are overall in line with our previous investigations into individual critical gaps and suggests a high potential benefit from the personalization [6]. A comparison between the estimated critical gap and the gaps drivers took for the last intersection, where gap sizes were increasing, showed that the estimates were highly plausible.

Subjective assessment of personalized and default drive

76 % of the 25 participants recognized a change between the default and the personalized time gaps. This is a first indica-

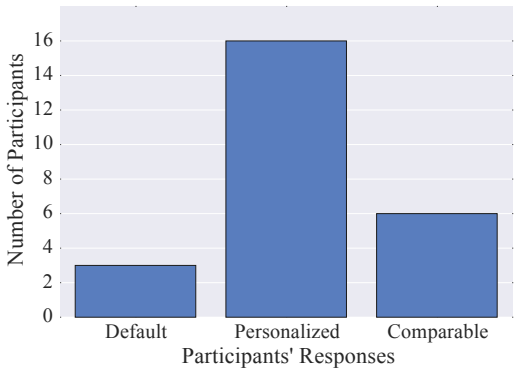

Figure 3. Number of answers to the question: "Which recommended gap size was more appropriate?"

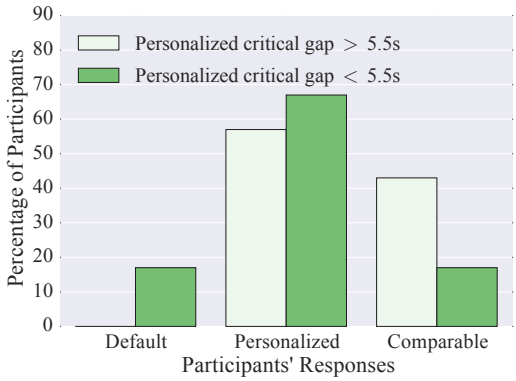

Figure 4. Percentage of answers to the question: "Which recommended gap size was more appropriate?" split into participants showing a critical gap > 5.5 s **and <** 5.5 s**.**

tion that changing the gap size has an effect for most drivers. Note that, as mentioned above, the order of these drives was counterbalanced and randomly assigned. Furthermore, most of the drivers preferred the system with personalized critical gaps (Fig. 3). A chi-square goodness of fit test confirmed that the underlying population of the sample is significantly different from a uniform distribution ($\chi^2(2) = 11.12, p = .004$). If we split up the drivers into the group with a personalized critical gap larger than the default 5.5 s and a group with a personalized critical gap smaller than 5.5 s, we see that for both groups the personalized critical gap is rated as more appropriate (Fig. 4). All drivers with the larger personalized critical gaps prefer the personalized system or rate both systems as at least comparable. A chi-squared test showed that we can not reject the hypothesis that the distributions of both gap size groups are identical ($\chi^2(2) = 2.68, p = .262$).

Online judgments

In both drives, recommended time gaps were rated as very appropriate. The appropriateness of the recommended gaps of the default system were rated as 10.2 ($SD = 2.3$) on a 0 to 15 scale. However, with a rating of 11.8 ($SD = 1.6$), the recommended gaps of the personalized system were rated significantly more appropriate. The corresponding values of the paired-samples t-test show significant differences: $t(24) = -3.67, p = .001, \eta^2 = 0.359$.

System evaluation

In general, the system was rated to be very useful, to have positive effects on perceived workload, participants have a positive attitude towards the system, and the system is perceived to have positive effects on safety aspects (Fig. 5). In comparison to driving with the default AOD system, participants perceived the personalized AOD system to

- be more reliable ($t(23) = -3.25, p = .004, \eta_p^2 = .314$),
- facilitate monitoring of the traffic situation ($t(23) = -2.23, p = .036, \eta_p^2 = .178$),
- facilitate the decision for entering an intersection ($t(24) = -3.09, p = .005, \eta_p^2 = .285$),
- increase the comfort of the recommended time gaps ($t(23) = -2.19, p = .038, \eta_p^2 = .167$),
- like it more to use it ($t(23) = -2.82, p = .009, \eta_p^2 = .249$).

In both AOD drives, recommended gap sizes were, on average, rated to be just right ($M = 0.33, SD = 0.87$ for the default system and $M = -0.13, SD = 0.61$ for the personalized system on a scale from -2 "too long" to +2 "too short"). The better fit of the gap sizes of the personalized system, as evident from the smaller absolute value, was statistically significant ($t(23) = 2.30, p = .031, \eta_p^2 = .187$).

Overall Ranking & Evaluation of drives

Fig. 6 shows that, when asked to rank the different drives, most drivers ranked the personalized AOD system first and only a few put the manual driving on the first rank. In detail, 68.75 % ranked the personalized AOD system, 18.75 % the default AOD system and 12.5 % the manual drive first. Clearly, there is a strong preference for the personalized AOD system. In total, 87.5 % ranked one of the AOD systems first, which can be seen as a confirmation of the results from our previous AOD study. The reason for the occurrence of the non-integer values is that some participants ranked both, the personalized AOD drive and the default AOD drive, first. These votes were split between the two alternatives.

The evaluation of the drive assessment (Fig. 7) shows differences for perceived attentiveness (ANOVA: $F(2,48) = 3.91, p = .027, \eta_p^2 = .140$) and demandingness (ANOVA: $F(2,48) = 3.18, p = .050, \eta_p^2 = .117$). Compared to manual driving, drivers reported to have felt less attentive in the default drive ($t(24) = 2.50, p = .020$) and perceived the personalized drive as less demanding ($t(24) = 2.67, p = .013$). The other two comparisons to manual driving showed reductions in attentiveness and demandingness but they were not statistically significant.

Evaluation of objective driving data

We evaluated two types of objective driving behavior. First we looked at the commitment of the drivers to the system, i.e., if the drivers followed the system's recommendation. Next we

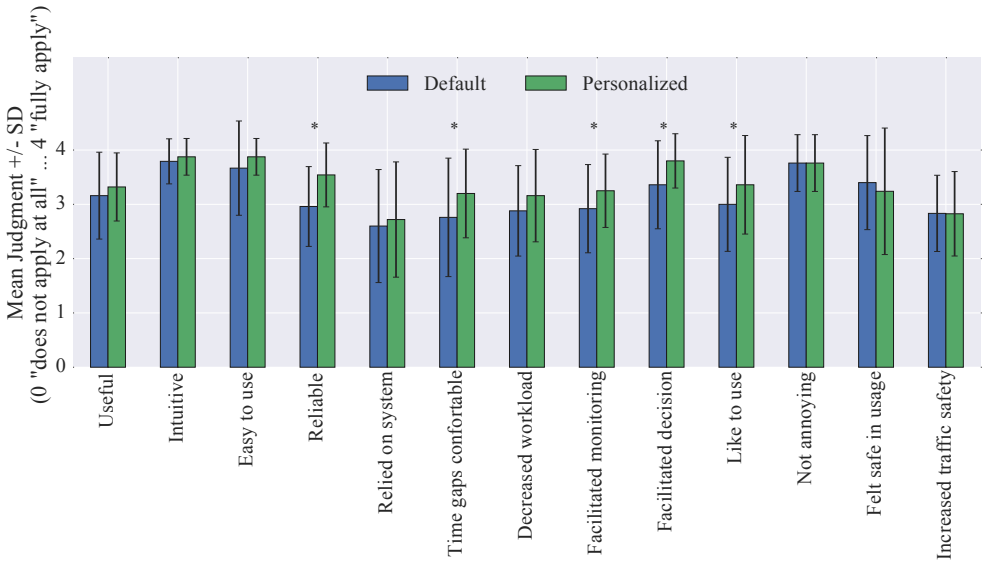

Figure 5. Participants' system evaluation.

Figure 6. Ranking of the three systems.

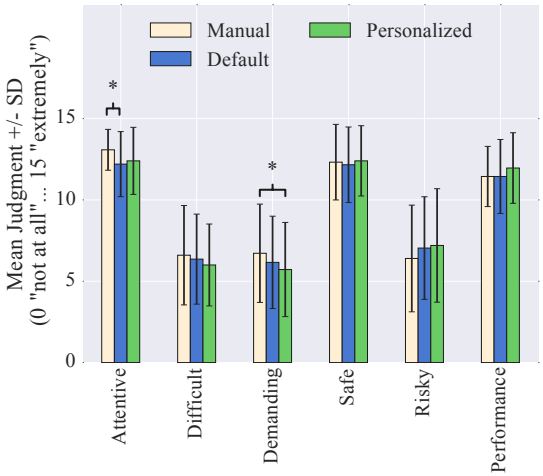

Figure 7. Participants' drive evaluation.

calculated the time headway (THW) of the ego vehicle to the traffic vehicles at the moment the driver started entering the intersection.

Commitment to the system

In the default AOD system drive, on average at three out of 13 intersections, drivers ignored the system message "vehicle approaching from right" (*NO-GOs*) and entered the intersection nevertheless (Fig. 8). In the personalized AOD system drive, in contrast, the commitment to the system was nearly 100 % at all intersections. This difference is also statistically significant ($t(24) = 4.18, p < .001, \eta_p^2 = .421$). Together with the previously reported increased appropriateness of the personalized gaps, this is a strong indication that personalized gaps improve the participants' compliance with the AOD system's recommendation.

Fig. 9 shows the mean number of "GO"s that were ignored by the participants, i.e., the number of cases in which the AOD system informed the participant that a suitable gap had appeared yet the participant did not make the turn. As mentioned previously, due to our scenario design, all of these were gaps that drivers could have taken, as no car from the left could have been interfering. The results show that there were very few drivers, who did not enter the intersection despite the system's recommendation. But there is no significant difference between the two system alternatives ($t(24) = 0.77, p = .452, \eta_p^2 = .024$). Most of the time, it was the first intersection, where the recommendation was ignored, possibly because the drivers may have needed a certain adaptation phase. Taking these results into account, it can be argued that there is in general a high acceptance of the system.

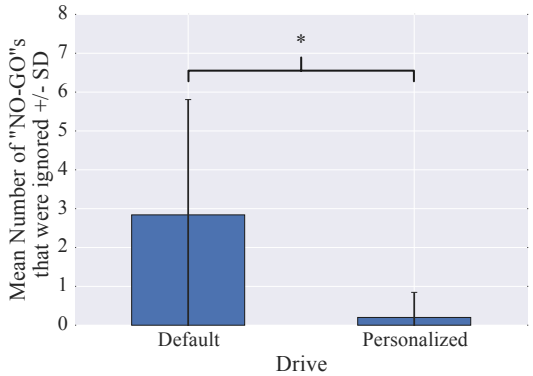

Figure 8. Mean number of situations in which a participant entered the intersection contrary to the system recommendation.

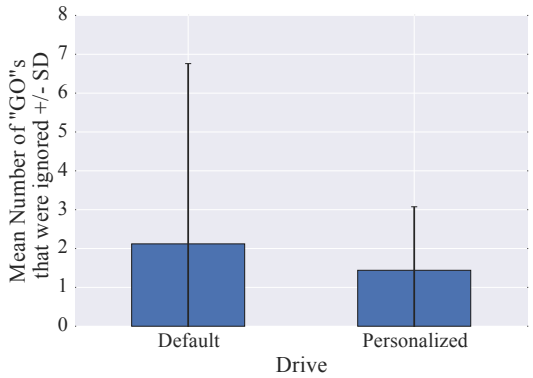

Figure 9. Mean number of situations in which a participant waited at the intersection contrary to the system recommendation.

THW when entering the intersection

The mean THW of the ego vehicle to the vehicles approaching from either left or right at the moment the ego vehicle entered the intersection are shown in Fig. 10. We show the THWs from the left merely for completeness. As we mentioned previously, we designed the scenario in such a way that the traffic from the left side did not interfere with the gaps presented from the right. Consequently, the THWs from the left should always be very large. This can be also seen from Fig. 10. A comparison of the THWs from the right for the manual and the system drives reveals that for the latter the means and in particular the standard deviations are smaller. An ANOVA confirms that these differences are significant (ANOVA: $F(2,48) = 4.84, p = .012, \eta_p^2 = .168$). Furthermore, a post hoc paired sample t-test shows a statistical difference between manual driving and the personalized AOD system ($t(24) = 2.63, p = .014, \eta_p^2 = 0.225$) and between the two system variants $t(24) = 2.52, p = .019, \eta_p^2 = 0.210$) yet not between manual driving and the default AOD system ($t(24) = 1.77, p = .089, \eta_p^2 = 0.116$).

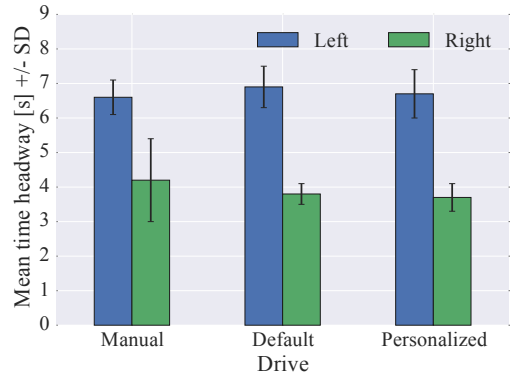

Figure 10. Time headway for vehicles approaching from the right or left side at the moment the ego vehicle started entering the intersection. Bars show mean and standard deviation.

CONCLUSION

In our previous study [8] we have already shown that the AOD concept is, in the context of an urban intersection left-turn assistant, highly attractive for the drivers and in general preferred to driving without any assistance.

In this work we have further developed the AOD system and shown how the use of driver-specific critical gaps yields an appropriate personalization of the system. The participants first drove manually and then twice with the AOD system activated. The first drive served to estimate the individual driver's left-turning behavior, i.e., the corresponding critical gap. In one AOD system drive, a default critical gap was used and in another the calculated personalized critical gap. The order of those two AOD system drives was counterbalanced for the participants. The participants of the presented study reported that this personalization was not only noticeable but also enhanced the monitoring of the traffic situation and the decision of entering an intersection compared to driving with the non-personalized AOD system. Consequently, 68.75 % of the participants have ranked the personalized system first, compared to only 18.75 % ranking the non-personalized AOD system most highly. Furthermore, 87.5 % of the participants preferred either the personalized or default AOD over unassisted driving. Additionally, the evaluation of the objective driving data has shown that the participants followed the recommendations of the personalized AOD system much more closely than those of the default AOD system. One could hypothesize that the reason for the good performance of the personalized AOD system was that the time gaps recommended in the default system were inadequate in general. Yet the evaluation of the appropriateness ratings split between participants showing a critical gap larger or smaller than 5.5 s in Fig. 4 rules this out. If the default time gaps were inadequate in general only either those showing a critical gap larger or smaller than the default critical gap would consider the personalized system to be an improvement. However, Fig. 4 clearly shows that both groups see it as a clear improvement.

The analysis of the THWs from the right revealed reduced THWs for the system drives compared to the manual drive.

This is accompanied with significantly reduced standard deviations. From this we conclude that the reduced THWs are the result of a habituation effect. The manual drive was always the first drive. Here participants were more insecure in their gap choices. This is reflected in the much higher standard deviation. In the following system drives participants more consistently chose the gaps best fitting to them. This is not primarily an effect of the system recommendation as can be seen from the high number of "NO-GO"s ignored in the default drive. Furthermore, the personalized system did also not motivate them to take smaller gaps. The recommendations they received were based on their driving behavior in the manual drive. They did then not deviate much from these recommendations as illustrated by the very small number of "NO-GO"s ignored in the personalized AOD system drive. Another indication that the reduced THWs are not primarily an effect of the AOD system is that the differences between the two system variants are marginal. Overall we conclude that both AOD system variants do not lead to more risky driving of the participants.

Driving simulators are in general not able to replicate the driving experience exactly. Particularly relevant for our study could be the incorrect, i.e., two-dimensional visual representation of the environment. This might lead to misjudgment of distances, particularly those close to the driver. Consequently, in our study the estimation of the critical gaps could be impaired by this effect. However, it has to be taken into account that for our study not the absolute validity but only the relative validity is of relevance. This means that, despite absolute differences to the expected real driving behavior, the relative differences between the different scenarios should be identical to those in real driving. The relative validity of driver behavior at intersections was investigated in [9]. The authors compared the number and type of driving errors while crossing intersections in a real drive and a simulator drive, with real intersections replicated in the simulator study. They found no significant interaction effects between the used method (real driving vs. simulator) and the turning direction (left vs. right) and therefore attest the simulator a relative validity. From this we conclude that it can be expected that using a simulator instead of real driving also only has an effect on the absolute validity of our experiment, but that the relative results, in particular the preference for one of the system variants, is not affected by this.

As the time the participants could spend in the simulator was limited, we restricted the scenario in several ways. First of all, the intersection layout was identical for all intersections. Similarly, we also did not change road and weather conditions. Consequently, we were not able to investigate how these factors influence the acceptance of the AOD system in general and its personalization in particular. In the current study, we also kept the traffic density constant. However, in our previous study [8] we investigated different traffic densities, including traffic arriving only from one side. There we saw that the AOD system is perceived the more useful the more difficult the situation is for the driver, i.e., the higher the traffic density and the more directions it is arriving from. As the personalized AOD system was reported to facilitate monitoring the traffic and the decision process for entering the intersection it can be expected that the usefulness of the personalized AOD system will be even higher for these challenging situations. This view is supported by the observed significantly higher commitment to the personalized compared to the default AOD system. A second restriction caused by the limited available time of the participants was that the number of observations available to estimate their driving behavior and hence perform the personalization was also small. In [5] we observed that a reliable estimation of a driver specific critical gap with the method of Troutbeck requires approximately 15 observations. This number was too high for our scenario. As a consequence, we only presented four different gap sizes to the drivers to make the estimation of the driver specific critical gaps easier. This strongly limited the adaptation potential of the personalization. At the same time, drivers were not aware of these limitations as they only experienced these scenarios. This means that with the current study we were not able to answer the question of how the precision of the personalization influences the acceptance of the system.

While in this work, only the observed driving behavior was used for a personalization of the system, in future work, we also plan to consider other aspects of the system that have a potential for personalization, such as the frequency and timing of the system recommendations or the inclusion of other useful information regarding the traffic situation.

ACKNOWLEDGMENTS

We would like to thank Heiko Wersing for fruitful discussions. Many thanks go to our participants for their patience and effort.

REFERENCES

1. L. Biester. 2009. *Cooperative automation in automobiles*. Ph.D. Dissertation. Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät II.

2. W. Brilon. 2016. Some Remarks Regarding the Estimation of Critical Gaps. *Transportation Research Record: Journal of the Transportation Research Board* 2553 (2016), 10–19. DOI: http://dx.doi.org/10.3141/2553-02

3. W. Brilon, R. Koenig, and R. J. Troutbeck. 1999. Useful estimation procedures for critical gaps. *Transp. Research Part A* 33, 3-4 (1999), 161–186.

4. M. Hasenjäger and H. Wersing. Personalization in Advanced Driver Assistance Systems and Autonomous Vehicles: A Review. In *2017 IEEE 20th Int. Conf. on Intelligent Transportation Systems (ITSC)*. Yokohama, Japan.

5. D. Orth, D. Kolossa, and M. Heckmann. Predicting Driver Left-Turn Behavior from Few Training Samples Using a Maximum-A-Posteriori Method. In *2017 IEEE 20th Int. Conf. on Intelligent Transportation Systems (ITSC)*. Yokohama, Japan.

6. D. Orth, D. Kolossa, M. Sarria Paja, K. Schaller, A. Pech, and M. Heckmann. A Maximum Likelihood Method for

Driver-Specific Critical-Gap Estimation. In *2017 IEEE Intelligent Vehicles Symposium (IV)*. 553–558.

7. N. Schömig, M. Heckmann, H. Wersing, C. Maag, and A. Neukum. 2016. Assistance-On-Demand: A Speech-Based Assistance System for Urban Intersections. In *Adjunct Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '16 Adjunct)*. ACM, New York, NY, USA, 51–56. DOI: http://dx.doi.org/10.1145/3004323.3004328

8. N. Schömig, M. Heckmann, H. Wersing, C. Maag, and A. Neukum. to appear. "Please watch right" - Evaluation of a speech-based on-demand assistance system for urban intersections. *accepted for: Transportation Research F* (to appear).

9. O. Shechtman, S. Classen, K. Awadzi, and W. Mann. 2009. Comparison of driving errors between on-the-road and simulated driving assessment: a validation study. *Traffic injury prevention* 10, 4 (2009), 379–385.

10. Z. Tian, M. Vandehey, B. W. Robinson, W. Kittelson, M. Kyte, R. J. Troutbeck, W. Brilon, and N. Wu. 1999. Implementing the maximum likelihood methodology to measure a driver's critical gap. *Transportation Research Part A: Policy and Practice* 33, 3-4 (1999), 187–197.

11. R. J. Troutbeck. 1992. *Estimating the Critical Acceptance Gap from Traffic Movements*. Physical Infrastructure Centre, Queensland University of Technology.