Speech improves human-automation cooperation in automated driving

Frederik Naujoks[1], Yannick Forster[2], Katharina Wiedemann[2] and Alexandra Neukum[1]

Center for Traffic Sciences (IZVW), University of Wuerzburg[1]
Wuerzburg Institute for Traffic Sciences (WIVW)[2]

Summary

During highly automated driving, upcoming automated manoeuvres (e.g., lane changes) should be communicated to the driver in order to ensure system transparency. As driving time can be used for non-driving-related tasks (NDRT), such as office work or in-vehicle entertainment, drivers might prefer to be informed in a non-distracting way as interruptions of ongoing NDRTs may be perceived as a nuisance (e.g., when drivers are required to retrieve information from attention demanding displays). In this paper, the potential for using speech output to improve human-automation communication is explored. A sample of 17 drivers participated in a simulated automated driving experience in a motion-based driving simulator, which replicated different situations for the participants that required communication between the automation and the driver (e.g., lane changes, avoiding of obstacles). Compared to generic auditory output (i.e., standard information tones), communicating upcoming automated manoeuvres by speech led to a decrease in self-reported visual workload and reduced interference with NDRTs. Participants clearly favoured the speech-based output.

1 Introduction

Highly automated driving (HAD) provides the opportunity to disengage from the task of driving for longer time periods without the need for continuous monitoring of the driving situation (Gasser & Westhoff, 2012). However, unlike fully automated driving, occasional manual intervention may be necessary because of operational system limits (e.g., missing lane markings) or system failures (e.g., sensor malfunctions). In these cases, successful human-automation cooperation requires fast and effective communication of the need for manual intervention (e.g., Gold, Damböck, Lorenz & Bengler, 2013; Naujoks, Mai & Neukum, 2014; Naujoks et al., 2015). However, highly automated vehicles may also be capable of managing certain non-critical driving situations, such as overtaking or adaptation of the host vehicle's

Veröffentlicht durch die Gesellschaft für Informatik e.V. 2016 in
B. Weyers, A. Dittmar (Hrsg.):
Mensch und Computer 2016 – Workshopbeiträge, 4. - 7. September 2016, Aachen.

http://dx.doi.org/10.18420/muc2016-ws08-0007

speed, without any driver intervention. In these cases, it is necessary to communicate the actions that are initiated by the automation in an unambiguous way to avoid distrust and unnecessary manual interventions (Parasuraman & Riley, 1997). Notifying the driver about upcoming automated manoeuvres may, on the other hand, be perceived as a nuisance if this interferes with non-driving related activities that are carried out during the highly automated drive, such as office work or entertainment. This may be especially the case if visual attention has to be directed away from the activity. In view of these challenges in the design of human-machine interfaces for highly automated vehicles, the current study investigates whether human-automation cooperation can be improved by means of speech output. Using a motion-based driving simulator, participants completed a highly automated drive while performing a common non-driving related task (NDRT). During the drive, several situations either required manual intervention or were carried out by the automation independently. In this paper, the drivers' engagement in the NDRT during system-initiated manoeuvres as well as the drivers' subjective evaluations of the human-machine interface are analysed. The human-machine interface (HMI) was designed and evaluated in a previous study (Naujoks et al, 2016; Forster et al., 2016) and consisted of a visual-auditory interface that either used generic auditory output (i.e., standard warning and notification tones) or additional speech output. It was expected that additional speech would enhance the human-automation cooperation and that the participants would be less inclined to interrupt the NDRTs. However, the additional speech output could be perceived as unnecessary and annoying. The design of the study is presented in the next chapter.

2 Method

2.1 Participants and driving simulator

The original sample consisted of *N* = 20 drivers (*M* = 29.0; *SD* = 8.1; *Min* = 22; *Max* = 56). Due to technical problems, three drivers could not complete the test drive. The study was conducted in the motion-based driving simulator at the Wuerzburg Institute for Traffic Science using the simulation software SILAB.

2.2 Human-machine interface

The visual part of the HMI is shown in Figure 1. Blue lane symbols in the centre of the HMI indicate that the lateral guidance is carried out by the HAD function. The length of a blue rectangle shows the set distance to vehicles ahead. This part of the proposed HMI resembles that of existing HMI solutions for ACC with additional steering assistance (e.g., Naujoks et al., 2015). The set speed (1a) and current speed (2) are displayed. If the driver changes the set speed, the new set speed is depicted. If a traffic event, such as an upcoming speed limit, requires speed adaptation, this is displayed to the driver in advance by a message box on top of the HMI (3) that includes a symbolic representation of the traffic event (4) and the distance to the traffic event (5). Automated speed adaptation is depicted by marking a line through the set speed (1b) until the speed limitation event is over.

The proposed HMI for displaying automated manoeuvres is shown in Figure 2. For example, these could signal an in-lane avoiding manoeuvre (upper part of Figure 2) or a lane change manoeuvre (lower part of Figure 2). When approaching the manoeuvre, the situation is announced to the driver (Figure 2, left part). To clearly communicate that no manual intervention is needed, the same blue colours used in the normal operating state are used. The type of traffic event (Naujoks & Neukum, 2014; Wiedemann et al., 2015) and the remaining distance to the event are also announced to the driver. The preparation stage (Figure 2, middle) informs the driver about the specific manoeuvre the system plans to carry out. The turquoise arrow and the text message above the main state indicate that the automation is planning to execute the manoeuvre. Subsequently, the execution of the manoeuvre is also communicated by a text message and blue colouring of the situation specific arrow (Figure 2, right). Visual information is provided in a head-up display (HUD).

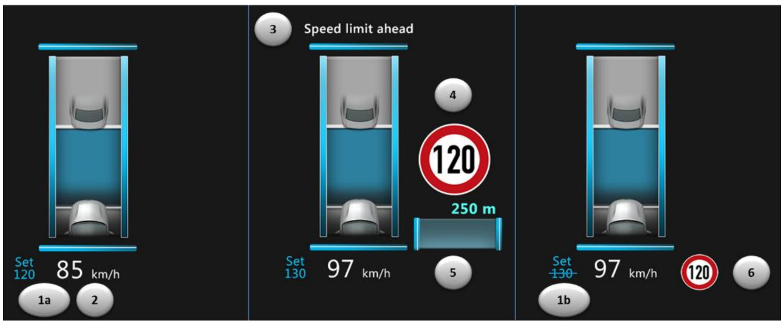

Figure 1. HMI for communication of speed limits.

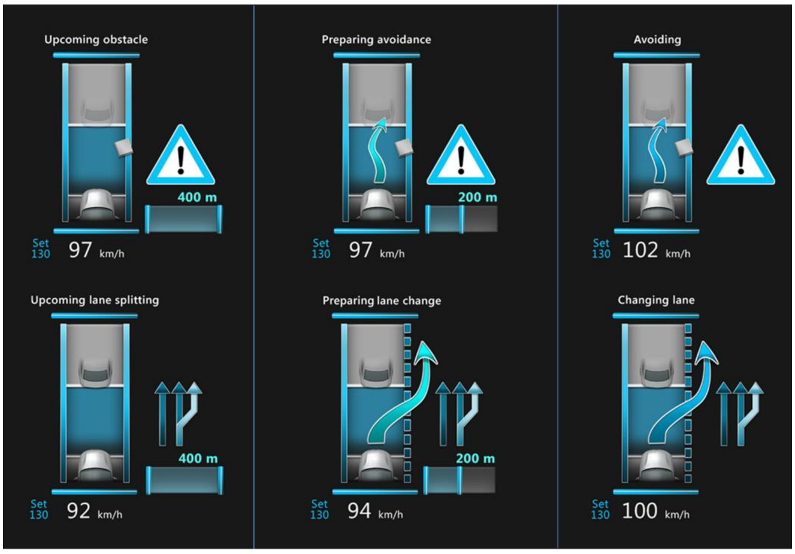

Figure 2. HMI for communication of driving manoeuvres. Announcement stage (left), Preparation stage (middle) and Execution stage (right)

In addition to the visual display, generic auditory output was presented together with the visual announcement of the traffic event. The generic auditory output (condition: “generic”) consisted of two tones (duration: 150 ms; frequency: 1000 Hz, interval: 150 ms). In another experimental condition, this generic output was also accompanied by speech output (condition: “speech + generic”). Instead of generating machine-based speech output (e.g., text-to-speech), a female voice (cf. Bazilinskyy & de Winter, 2015) was recorded using a Dictaphone. The speech output verbalized the information provided by the visual HMI. At this point, it is important to emphasize that the usefulness of the visual HMI was previously demonstrated and further improved in a prior study (Forster, Naujoks & Neukum, 2016; Naujoks, Forster, Wiedemann & Neukum, 2016). Speech output was originally recorded in the German language. The exact wording translated from German into English is shown in Table 1.

Table 1. Auditory HMI speech output for system manoeuvres (translated from German into English).

Situation	**Wording**
Speed Limit	“Speed Limit ahead. I will adapt the speed.”
Avoiding	“Obstacle ahead. I will adapt the lane position and avoid it.”
Lane Change	“Highway intersection ahead. I will change the lane and follow the navigation.”

2.3 Driving situations and instructions

Participants drove in the highly automated mode on a three-lane highway with moderate traffic. The drive lasted approximately 15 minutes and included three driving events that required communication between the HAD function and the driver:

- *Avoiding*: Lost cargo on the right lane. The HAD system adapts its lateral position and avoids the obstacle on the road
- *Speed limit*: Adaptation of the host vehicle’s set speed (from 120 km/h to 80 km/h) due to a speed limit change
- *Lane change*: HAD system recognizes highway intersection ahead and changes lane to the right to follow the route

In addition, one take-over scenario was included in the drive. The results pertaining to the take-over scenario are not part of this paper and will be reported elsewhere. A within-subject design was used. All participants completed the simulator drive twice, with and without speech output. The order of the drives was balanced. Within the drives, the different driving situations were encountered in randomised order. Participants were instructed to complete the simulator drive using the HAD function. They were instructed on how to activate and deactivate the system, but they were not given any advance information about the HMI. The participants were instructed that the automated driving function would carry out the driving tasks completely and that they would be informed by the automation if manual intervention was necessary. They were not given any information about the automated driving manoeuvres that they were about to experience during the drive. They were also asked to read articles in a magazine during the automated drive that were selected from a weekly German news magazine. To increase participants' motivation to carefully read the articles, the participants were told that their knowledge of the articles' content would be tested after the drive.

2.4 Dependent measures

The dependent measures are listed in Table 2. The main objective was to investigate whether drivers would benefit from the additional speech output during the automated manoeuvres. The behaviour of the drivers was assessed by observing if they would interrupt or even stop working on the NDRT. In addition, participants were asked to evaluate the usefulness of the visual and auditory output of the HMI, as well as the visual workload of retrieving information from the HMI. Assessments of usefulness and visual workload were recorded just after completing the respective scenario during the drive. A 15-point scale ranging from 1 (very little) to 15 (very much) with an additional category (0 = not at all) was used. Lastly, acceptance of both variants of the HMI was assessed by asking the drivers whether they preferred the HMI with or without speech output during a follow-up interview.

Table 2: Dependent measures and description of categories/questions.

Measure	**Description**	**Range**
Level of interference with non-driving related activity	Coding of video recordings during automated manoeuvres ▪ Not distracted: Driver does not perform task (1) ▪ Continuation of NDRT: – No reaction, continuation of NDRT (2) – Short glance ahead, continuation of NDRT (3) ▪ Alternating of NDRT and looking ahead (4) ▪ Interruption of NDRT until situation is completed: – Interruption of NDRT and looking ahead, magazine in hand (5) – Interruption of NDRT and looking ahead, putting magazine aside (6)	[1-6]
Assessment of usefulness	How helpful was the HUD? How helpful was the auditory output?	[0-15]
Visual workload	How much attention did you pay to the HUD?	[0-15]
Acceptance	Which system do you prefer?	-

3 Results

Interruption of non-driving related task. Figure 3 shows the frequency of behavioural observations for both “speech + generic” and “generic” auditory output during the system manoeuvres. In tendency, the NDRT was less often interrupted in the “generic + speech” condition across all scenarios (19 vs. 27 task interruptions with coding category 5 or 6, χ^2 =

2.53, p = .082, one-sided). On a descriptive level, this trend was also visible when the different scenarios were considered:

- *Lane Change*: most drivers alternate between NDRT and looking ahead (category 4) in the "speech + generic" condition, whereas the NDRT is most often interrupted in the "generic" condition (category 5: interruption of NDRT, looking ahead, magazine in hand).
- *Avoiding:* most observations fall into the category 5 (interruption of NDRT, looking ahead, magazine in hand) in the "speech + generic" condition, but into category 6 in the "generic" condition (putting magazine aside, interruption of NDRT).
- *Speed Limit*: most participants reacted by glancing briefly ahead but continuing the NDRT in the "speech + generic" condition (category 3). When driving in the "generic" condition, however, a bimodal distribution was found with a peak in category 3 (short glance ahead, continuation of NDRT) and another peak in category 5 (interruption of NDRT and looking ahead, magazine in hand) which also shows a greater interference with the NDRT in this condition.

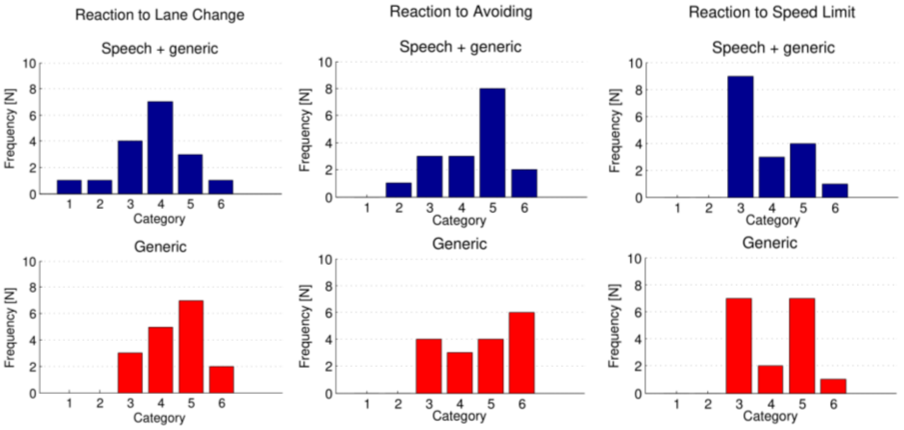

Figure 3. Behavioural observations during system manoeuvres, top: "speech + generic", bottom: "generic". Categories: (1) not distracted, (2) no reaction, continuation of NDRT, (3) short glance ahead, continuation of NDRT, (4) alternating of NDRT and looking ahead, (5) interruption of NDRT and looking ahead with magazine in hand, (6) interruption of NDRT and looking ahead, putting magazine aside.

Self-reported usefulness, visual workload and acceptance. Subjective assessments of the usefulness of the visual display (Figure 4, left) were on a high level with a trend towards higher ratings in the "generic" condition ("speech + generic": M = 10.29, SD = 3.28; "generic": M = 11.59, SD = 3.34, F(1,15) = 3.05, p = .101). It thus appears that the redundant information provided by speech output caused drivers to rely less on the visual information provided by the HMI. Usefulness ratings of the auditory output (Figure 4, middle) were also on a high level ("speech + generic": M = 12.92, SD = 3.50; "generic": M = 11.18, SD = 3.29, F(1,16) = 1.97, p = .180). Participants tended to report a higher level of usefulness of the auditory "speech + generic" output in the *avoiding* (t = -1.74, df = 16, p = .102) and *lane change* (t = -1.77, df = 16, p = .096) situations, but not in the *speed limit* situation (t = -0.65, df = 16, p = .527; interaction situation*condition: F(2,15) = 3.85, p = .045).

The visual workload (Figure 4, right) was rated on a high level in the “generic” condition (*M* = 11.31, *SD* = 3.29) and on an intermediate level in the “speech + generic” condition (*M* = 8.70, *SD* = 4.57; *F*(1,15) = 10.40, *p* = .006), suggesting a relief of visual workload when presenting situation-specific information with speech output. In tendency, the *avoiding* situation was considered to be more visually demanding in both conditions in comparison to the *speed limit* and the *lane change* situations (*F*(1,14) = 3.21, *p* = .071). Participants clearly favoured “speech + generic” (*n* = 16) over “generic” output (*n* = 1). The only participant in favour of the “generic” output indicated that speech output could be annoying over time when occurring too frequently.

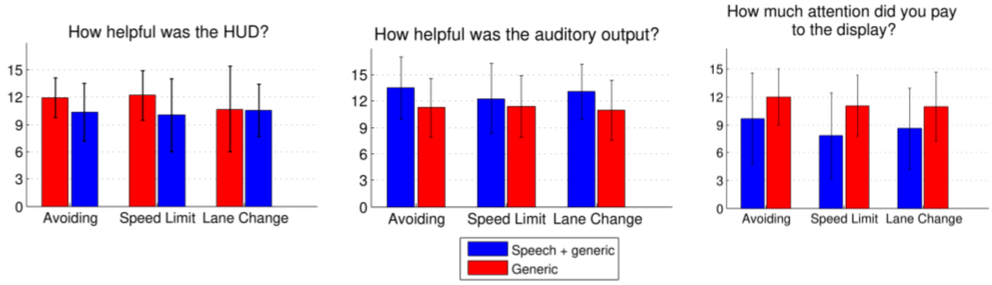

Figure 4. Usefulness ratings of the visual and auditory components of the HMI and visual workload (right to left).

4 Conclusions

The current study investigated whether speech output could improve human-machine cooperation in the area of HAD. While communication of transitions from HAD to manual driving has received considerable research interest, enhancing system transparency by communicating upcoming automated manoeuvres has not been studied extensively yet. *N* = 17 participants completed the same driving simulator course twice while interacting with a system that incorporated speech output (“speech + generic”) and a system that only applied generic auditory feedback (“generic”). We investigated whether additional speech output would facilitate human-automation cooperation by effectively informing the driver about upcoming automated driving manoeuvres and would, therefore, cause less interference with the execution of a NDRT. During automated manoeuvres, a trend towards continuation of the NDRT in the “speech + generic” condition was observed. When only “generic” auditory output was used, a higher interference with the NDRT was observed. This suggests that additional speech output can facilitate behavioural responses that are in accordance with the system’s intentions. Drivers can stay more focused on NDRT and do not have to monitor the visual component of the HMI and the traffic situation as much as with “generic” auditory output. Accordingly, participants reported lower visual workload originating from reading and interpreting the visual HMI when speech output was presented. When considering the perceived usefulness of the auditory feedback, the “speech + generic” output was rated as more useful compared to the “generic” auditory output. In contrast to the unspecific “generic” feedback, the semantic speech output apparently facilitated retrieving information that was relevant for understanding the system's intentions. Consequently, the visual component was considered less useful in the "speech + generic" condition because important information about the upcoming manoeuvre could be derived from the semantic speech output. Finally, a strong preference for the system with speech output was found, which emphasizes the advantages of semantic feedback. However, it should be noted that the benefits of speech output might decrease when drivers are engaged in other NDRTs than the one investigated in this study, especially when they draw on the driver's auditory attention. It should also be emphasised that the decreased interference with NDRTs might turn into a disadvantage if drivers fail to notice system malfunctions during automated manoeuvres timely enough as a result of insufficient system monitoring. Future studies should also investigate whether speech is still favoured over generic output after longer periods of system usage, especially when drivers are already familiar with the system.

Acknowledgements. This work results from the joint project Ko-HAF –Cooperative Highly Automated Driving and has been funded by the Federal Ministry for Economic Affairs and Energy based on a resolution of the German Bundestag.

References

Bazilinskyy, P. & de Winter, J. (2015). Auditory interfaces in automated driving: an international survey. *PeerJ Computer Science, 1*, e13.

Forster, Y., Naujoks, F. & Neukum, A. (2016). Your turn or my turn? Design of a Human-Machine Interface for second generation Highly Automated Driving, *8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. Ann Arbor.

Gasser, T. & Westhoff, D. (2012). *BASt-study: Definitions of automation and legal issues in Germany.* Paper presented at the Workshop on the Future of Road Vehicle Automation, Irvine, CA.

Gold, C., Damböck, D., Lorenz, L. & Bengler, K. (2013). "Take over!" How long does it take to get the driver back into the loop? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 57*, 1938-1942.

Naujoks, F., Forster, Y., Wiedemann, K. & Neukum, A. (2016). A Human-Machine Interface for Cooperative Highly Automated Driving, *7th AHFE Conference*. Orlando.

Naujoks, F., Mai, C. & Neukum, A. (2014). The Effect of Urgency of Take-Over Requests During Highly Automated Driving Under Distraction Conditions. In T. Ahram, W. Karowski & T. Marek (Eds.), *Proceedings of the 5th AHFE Conference* (Vol. 7, pp. 2099-2106). Krakau: AHFE Conference.

Naujoks, F. & Neukum, A. (2014). Specificity and timing of advisory warnings based on cooperative perception. In A. Butz, M. Koch & J. Schlichter (Eds.), *Mensch & Computer 2014 - Workshopband* (pp. 229-238). Berlin: De Gruyter Oldenbourg.

Naujoks, F., Purucker, C., Neukum, A., Wolter, S. & Steiger, R. (2015). Controllability of Partially Automated Driving functions–Does it matter whether drivers are allowed to take their hands off the steering wheel? *Transportation Research Part F, 35*, 185-198.

Parasuraman, R. & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39*(2), 230-253.

Wiedemann, K., Schömig, N., Mai, C., Naujoks, F. & Neukum, A. (2015). Drivers' monitoring behaviour and interaction with non-driving related tasks during driving with different automation levels, *6th AHFE Conference*. Las Vegas, USA.