# The Effects of Situational Demands on Gaze, Speech and Gesture Input in the Vehicle

**Florian Roider**[1,2], **Sonja Rümelin**[1], **Bastian Pfleging**[3], **Tom Gross**[2]
[1]BMW Group, Munich, Germany
[2]Human-Computer Interaction Group, University of Bamberg, Bamberg, Germany
[3], LMU Munich, Munich, Germany
{florian.roider, sonja.ruemelin}@bmw.de, bastian.pfleging@ifi.lmu.de, tom.gross@uni-bamberg.de

## ABSTRACT
Various on-the-road situations can make additional demands on the driver that go beyond the basic demands of driving. Thereby, they influence the appropriateness of in-vehicle input modalities to operate secondary tasks in the car. In this work, we assess the specific impacts of situational demands on gaze, gesture and speech input regarding driving performance, interaction efficiency and subjective ratings. An experiment with 29 participants in a driving simulator revealed significant interactions between situational demands and the input modality on secondary task completion times, perceived suitability and cognitive workload. Impairments were greatest when the situational demand addressed the same sensory channel as the used input modality. This was reflected differently in objective and subjective data depending on the used input modality. With this work, we explore the performance of natural input modalities across different situations and thereby support interaction designers that plan to integrate these modalities in automotive interaction concepts.

### Author Keywords
Gaze interaction; gesture interaction; natural interaction; resource competition; situational demand; speech interaction

### CCS Concepts
•**Human-centered computing** → **HCI theory, concepts and models; Interaction design theory, concepts and paradigms;** *Laboratory experiments; Pointing; Gestural input;* User studies;

## INTRODUCTION
Many in-vehicle interaction concepts integrate multiple input modalities such as touch, speech, gestures or gaze to control the in-vehicle information system while driving. Concepts typically aim to enhance interaction regarding naturalness, efficiency, robustness and flexibility. The latter can be achieved by providing alternative input modalities. They allow drivers to choose freely, which input modality to use for interaction with the vehicle. This bears a number of advantages: First, drivers can choose input modalities according to their personal preferences. They can pick whichever modality they feel most comfortable with [9]. Second, they can alternate between modes to avoid physical overexertion for any individual modality [10]. Third, drivers can choose adequate modalities depending on the task they want to execute. It has been shown that there is a tight connection between the type of a task and the suitability of a specific input modality. Spatial tasks (e.g. interaction with a virtual map) are better achieved using manual input, whereas verbal tasks (e.g. address input for navigation) are easier to complete using speech input [16]. Finally and most importantly for this paper, alternative modalities allow the driver to accomplish interactions using the modality that is most appropriate to the driving situation [9].

**Figure 1. The experiment setup included a head-up display (A), which displayed icons that instructed a requested category. According categories were selected on the central information display (B) in the center stack. In this illustration, the participant uses gesture input to select the category "contacts" for the mail icon.**

### Varying Demands on the Driver
Driving a car is typically a dual task situation: Besides driving the car (primary task), the driver performs additional non-driving related tasks such as operating the information, enter-



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


ISBN 978-1-4503-5150-8/17/09...$15.00
DOI: https://doi.org/10.1145/3122986.3122999

tainment and comfort functions of the car (secondary tasks). Both tasks pose demands on the driver, who has to share his attention between them.

Cognitive science showed that people have separate pools of attentional resources that refer to the different sensory modalities [14]. The most important resource for driving a car is the visual resource, since the majority of all information that drivers use is obtained visually. This is followed by the manual resource, which is needed for steering the vehicle. Finally, auditory and speech resources can be classified least important for driving [17]. Besides these basic demands, on-the-road situations can impose additional demands on the driver's resources. Bad sight, conversations with co-drivers, and curvy or bumpy roads are just some examples for situations that occur on the road. They add up to the demands of basic driving and thereby change the proportion of demanded resources for the driving task. There are situations that put additional demands on the driver's visual, auditory, or manual resource. A distinct assignment of a real world situation to one demanded resource is mostly not applicable, nevertheless there is usually one resource that is primarily demanded by a certain situation.

Accordingly, alternative input modalities allow to change the proportion of demanded resources for operating the secondary task. Although the full range of potential demands of an interface needs to be considered, there is typically one sensory channel that is mainly addressed [8]. Gaze input mainly addresses the visual resource, gesture input the manual resource and speech input the auditory resource. We note that these channels match the demanded resources from the on-the-road situations described in the paragraph before (see Figure 2). In this work, we focus on the natural input modalities speech, gestures, and gaze (input via haptic elements and touch input are not natural in our sense).

### Interference Between Resources

There is a potential interference between the demanded resources for the driving task and those for operating secondary tasks while driving. Strayer et al. identified three sources of driver distraction: visual, manual and cognitive interference [13]. In their experiment, cognitive interference is created by having a conversation over a cell phone, which is related to the auditory resource. According to their differentiation, interaction with different devices can lead to competition from one, two or all three sources.

The competition between resources is also adressed in multiple resource theory, which describes the influence of different resources for the prediction of dual task interference [15]. In simplified terms, the theory implies that time-sharing two tasks with separate modalities (e.g. one visual task, one auditory task) has advantages over using the same modality for both tasks (e.g. both visual) due to resource competition. The amount of interference due to resource competition for a given resource is depending on the amount of competition for the resource and the relative importance of that resource in performing the task [16].

Engström and colleagues presented results that show radically different effects on driving performance between visual and au-

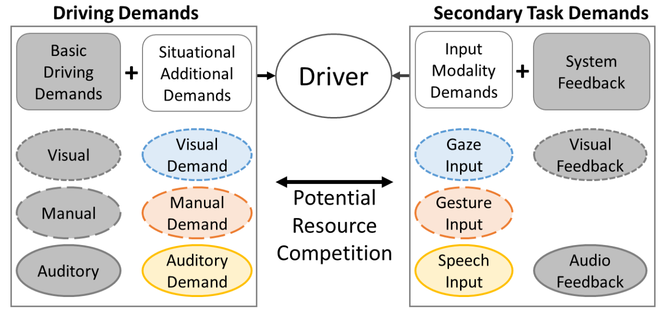


**Figure 2. This figure illustrates demands of a driving task and a secondary task on the driver. Rounded elements represent demands on driver's visual, manual and auditory resource. Gray elements remained unchanged in our study setup.**

ditory secondary tasks [3]. The visual task led to a reduction of speed and large steering corrections, whereas the auditory task led to increased lane keeping quality. This variance of effects for visual and auditory tasks suggest that manual secondary tasks may also result in different impacts on the driver.

Several studies in the last years investigated the application of natural input modalities to operate in-vehicle information systems using either standardized tests, such as the Lane Change Task (LCT) (e.g. [5, 12]), or miscallaneous driving simulations (e.g. [6, 7]) to simulate driving demands. While these studies produced valuable results they are related to one specific driving scenario. It is not clear how natural input modalities perform in driving situations with different demands. Being aware of the mutual influence is essential to correctly assess results regarding secondary task performance because results might vary across studies if driving task impose different demands on participants. For example, gesture input (manual resource) is likely to perform worse in a driving simulation that requires frequent steering (also manual resource) compared to gesture input on a straight road while listening to the radio (auditory resource). Accordingly, a driving simulation that has many detailed and rich sounds of the vehicle and the environment might discriminate speech input, due to resource competition.

In this work we investigate the specific effects of using gaze, gesture, and speech input for operating a secondary task in different driving situations. In a driving simulator-based experiment, we measured the impacts on driving performance, secondary task performance and cognitive workload. Thereby, we aim to provide a basis for the development of future situation-aware interaction concepts, as well as for the interpretation of results from (past and future) experiments that investigate natural input modalities in the car.

## EXPERIMENT

### Design

We used a within-subject design for our experiment in the driving simulator. Participants experienced four driving scenarios in a balanced layout. Three of these scenarios contained additional events that increased the demand on either the visual, manual, or auditory resource of the driver, whereas one scenario represented the same driving scene, without any additional demands. A more detailed explanation of how these demands were induced is given in the following subsection. During each scenario participants performed three trials of a secondary task via either gaze, gesture or speech input in a balanced layout. During each trial of 90 seconds they tried to make as many selections as possible. Consequently, our independent variables are the induced additional demand (visual, manual, auditory, none) and the used input modality (gaze, gesture, speech). We measure the effects of these two variables and their mutual influence with the help of the following dependent variables: driving performance (deviation of distance and lane keeping quality) and secondary task performance (secondary task completion time and subjective rankings). For subjective ratings, we used the driver activity load index (DALI) [11] and a 6-point likert scale to assess suitability from 0 (not suitable for the interaction) to 5 (very suitable for the interaction).

### *Situational Demands*

The different demands of traffic situations that occur on the road are the basis for the idea of applying additional demands in our study. However, we decided against a completely realistic implementation of situations for two reasons: The first reason is the complexity of a realistic technical implementation and the resulting problem of controlling the situations. The use of artificial events gave us more control over the study setup and allowed us to provide consistent demands for all participants. The second reason is the possibility to separate the induced demands more clearly with the help of artificial events. Our goal was to design events that have a realistic connection to real-world driving scenarios on the one hand, but address only one specific resource on the other hand. Figure

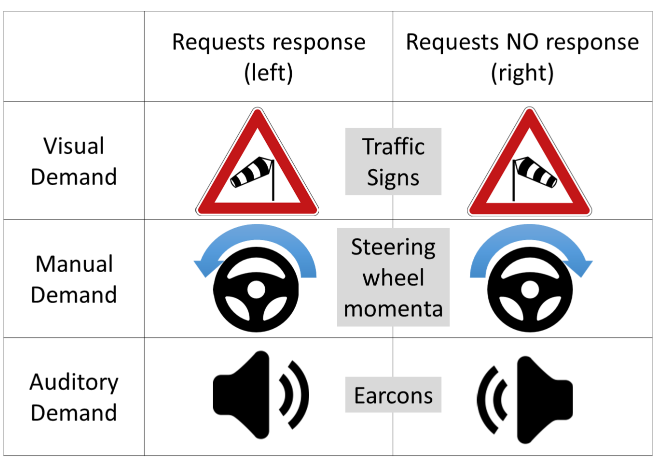


**Figure 3. Visual demand was induced by traffic signs showing wind directions. Manual demand was created by applying a momentum on the steering wheel to either the left or the right side. Acoustic signals on the left side and the right side of the driver increased the auditory demand.**

3 illustrates the events that were used to create the additional demands during the experiment: traffic signs at the roadside (visual demand), ear-cons in the cockpit (auditory demand) and temporary steering wheel momenta (manual demand). The latter was achieved by applying a force on the steering wheel that turned the wheel either to the left or to the right for one second. Participants would describe this event like a gust of wind that hits the car from the side. The distances between events were randomly distributed between 160 and 200 meters, which corresponds approximately to one event every six to seven seconds during the drive with 100 kilometers per hour. To make sure that participants really perceived and cognitively processed the meaning of the events, every event had two directions, left and right. Participants had to press a marked button on the steering wheel when a) the traffic sign shows to the left b) the sounds comes from the left c) the steering wheel turns to the left (compare Figure 3). This way, events stayed consistent regarding their demand of cognitive processing and response and only differed in the perceptual modality.

### *Secondary task design*

The secondary task was a simple selection task. There were three icons displayed on the central information display (CID) that represented typical categories of infotainment systems: contacts, navigation and settings. An icon on the head-up display (HUD) instructed which category to select (see Figure 1). Selections could be made via speech, gestures and gaze. Selection via speech required participants to name the categories. Category names were displayed above the according elements on the CID. Selections via gestures were made by pointing in the direction of the category to highlight it and then make a tap in the air with the index finger to select it. Gaze input was implemented using a dwell time approach. Participants had to focus the icon of the requested category with their gaze. The focused icon was selected when participant's gaze stayed on the icon for 600ms. This duration was visualized by filling the icon with a lighter color over the duration of the dwell time. For all three modalities, the system provided visual and acoustic feedback when categories where hovered or selected. The dwell time approach represents the standard approach in purely gaze controlled systems [4]. The combinations of gaze input with haptic buttons might produce better results in the automotive context [5]. However, we deliberately decided against such an approach in order to keep input modalities clearly separated. Moreover, our study focuses on the relative effects for each modality, rather than absolute interaction times.

## Participants

Originally, 36 participants took part in the study, but we had to exclude data of six participants. For those participants the eye-tracking system did not work properly, which resulted in non-representative secondary task times. Furthermore, one participant had to abort the experiment, because of simulator sickness.

This resulted in a total number of 29 participants (21 male, 8 female) with a mean age of 25.3 years ($SD = 4.9$). All of them possessed a valid drivers license. The majority of participants was driving regularly, 18 of them reported to drive more then 20,000 kilometers per year. All but two had experiences with speech input either in cars or on their mobile phones. 24 participants reported that they had never used any form of gaze input before.

### Apparatus

The experiment was conducted in a static driving simulator. The driving scene was projected on a 180 degree canvas in front of a vehicle mock-up. There were three displays in the cockpit (see Figure 1): the instrument cluster with a speedometer and the CID and HUD to display the secondary task. A *Leap Motion*[1] gesture controller was placed below the center stack facing upwards to enable gesture recognition in the area in front of the CID. Speech recognition was achieved using a clip-on microphone that was attached to participants' clothing to provide good audio quality. Gaze recognition was accomplished using *Dikablis Live Essentials Eye Tracking Glasses*[2], a head mounted eye-tracker. The system includes one forward-facing camera and a second camera filming the participant's left pupil enabling an exact determination of users' gazes in the cockpit. The software for the secondary task was implemented in *Unity3D*. The examiner controlled the trials from an adjacent examiner room. Data capture was managed automatically by the software. Additionally, the software offered the examiner to act as a *Wizard-of-Oz* for speech input, in case that speech recognition did not work reliably for some participants.

### Procedure

The examiner welcomed the participants in the simulator room and gave them a brief introduction about the course of the study. Participants signed a consent form and completed a questionnaire for demographic data. After that, they adjusted the seat position to their needs and put on the eye-tracking glasses. The examiner explained the secondary task and the interaction with gaze, gesture and speech input. After the calibration of the eye tracking system, participants had a few minutes to practice the selection via gaze, gestures and speech (without driving). They had another five minutes to familiarize with the driving simulator before they entered the motorway and followed a leading vehicle with 100 kilometers per hour at a distance of 50 meters. Participants were instructed to keep in the middle of the right lane and were not allowed to overtake the leading vehicle. After a few minutes of driving the examiner explained the additional demands and demonstrated them to the participants. It was made sure that participants experienced all additional demands (traffic signs, acoustic signals and steering wheel momenta) and were able to distinguish between both directions (see Figure 3) before continuing with the trials. At the beginning of each trial, participants were informed about the upcoming type of additional demand and instructed to prioritize the primary task of driving when simultaneously executing the secondary task. The secondary task started with the instructions on the HUD. After the selection of a category, a new random icon was displayed. Participants continued selecting categories for the duration of the trial. Between trials they stopped the vehicle and completed the DALI questionnaire before continuing with the next trail. This procedure was repeated for all conditions.

[1]https://www.leapmotion.com/

[2]http://www.ergoneers.com/eye-tracking/

## RESULTS

The main goal of our experiment was to investigate the influence of additional demands on different input modalities. Our dwell time approach for gaze input added a constant amount of time to every gaze selection. Times for speech input include the processing time of the system for automated speech recognition. Gesture input, in contrast, is more immediate and does not include any constant duration. For these reasons, absolute measures only represent our specific implementation for speech, gaze and gesture interaction. A more suitable approach to compare effects also across modalities is to describe the effect of additional demands in relation to the mean values of an input modality over all driving scenarios. Accordingly, we applied z-transformations on data from each input modality. Standardization provides a means to compare values from different normal distributions, by focusing on relative effects. All statistical calculations in the following section were made based on z-scores. At this point, we want to emphasize that, for absolute measures, there occurred significant differences between modalities and that the z-scores are only a means to provide a theoretical comparability here. Figures in this section display both, absolute values and standardized values, in order to support the awareness of actual differences.

### Driving Performance

#### *Deviation of Distance*

Figure 4 shows the standardized mean deviations of the requested distance of 50 meters to the leading vehicle. Highest mean deviations occurred during the visual demand conditions across all modalities. This indicates that it was most challenging for participants to keep the requested distance while they had to keep an eye on traffic signs on the same time. A two-way repeated measures ANOVA (additional demand: visual, manual, auditory, none; input modality: gaze, gesture, speech) was performed on these data. There was a significant main effect of additional demand $F(3,84) = 6.36, p < .01, \eta_p^2 = .185$. Bonferroni corrected post-hoc tests show that visual demand caused significantly higher deviation than all other conditions (all $p < .05$). According to the observation that visual demand caused highest deviation over all modalities there was no significant interaction between additional demand and input modality.

#### *Lane Keeping Quality*

We did not find significant effects on lane keeping quality. A two-way repeated measures ANOVA did not reveal any significant effects neither for the type of additional demand or the input modality and there was no interaction between both factors.

### Secondary Task Performance

#### *Task Completion Times*

Figure 5 illustrates the z-scores of secondary task completion times. We only used data of correct selections for the calculations. The most noticeable impairment due to resource competition can be observed for gaze input. The impact on gestures was less pronounced and not clearly visible for speech. A two-way repeated measures ANOVA reveals that the additional demand had a significant effect on task completion time, $F(3,84) = 6.51, p = .02, \eta_p^2 = .131$. We applied

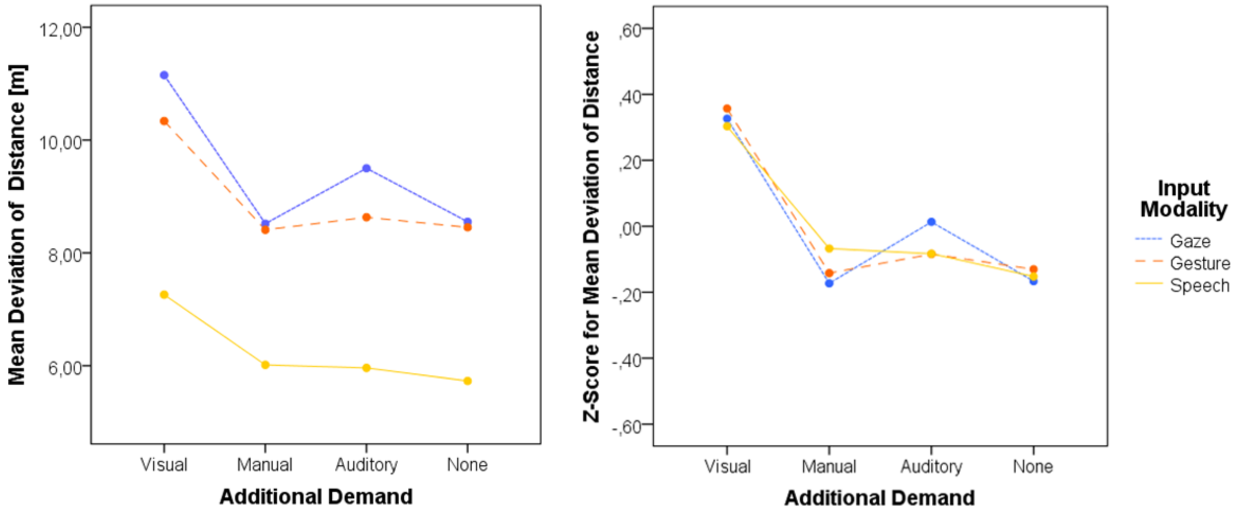


**Figure 4. Mean deviation of requested distance to the leading vehicle. Visual additional demand caused highest deviations of distance over all input modalities.**

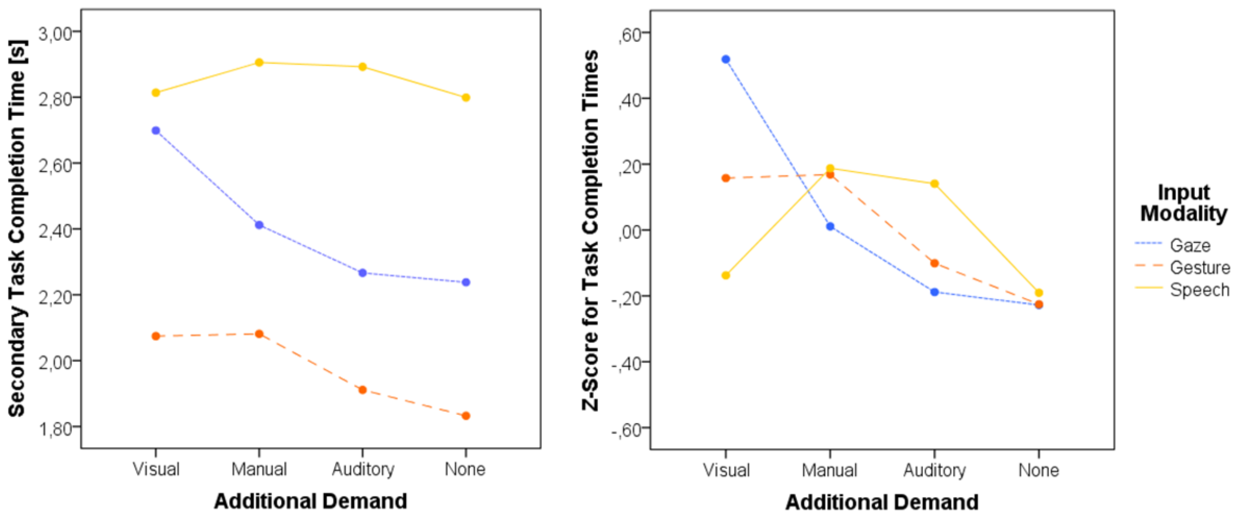


**Figure 5. Mean secondary task completion times for input modalities during additional demands. Largest impairments due to resource competition can be observed for gaze input.**

correction based on Greenhouse-Geisser, because Mauchly's test of sphericity indicated that the assumption of sphericity was not met for this test. Bonferroni corrected post-hoc tests show that selections during trials without additional demands were completed faster than during additional visual or manual demand (both $p < .05$). Additional auditory demand, in contrast, did not lead to a significant increase of task completion times. More interesting, there was a significant interaction between additional demand and input modality, $F(6,168) = 3.20, p = .02, \eta_p^2 = .103$, showing that there is a specific influence of additional demands on task completion times depending on the input modality.

#### *Error Rates*

Error rates were similar for gaze ($M = 6.1\%, SD = 0.01$), gesture ($M = 6.4\%, SD = 0.01$) and speech input ($M = 3.2\%, SD = 0.01$). A two-way repeated measures ANOVA of z-scores did not reveal a significant effect for the type of additional demand nor was there an interaction between both factors.

### Subjective Ratings

#### *Cognitive Workload*

The dimensions of the DALI refer to cognitive, temporal and perceptive components [11]. Results regarding the perceptive components would not add value at this point (e.g. there was higher auditory demand during additional auditory demand). More interesting are those dimensions, which refer to cognitive workload: effort of attention, interference and stress [11]. We took the average ratings of those three dimensions to assess the cognitive workload during each trial. Figure 6 shows the impact of different types of demand on cognitive workload. Participants perceived relatively highest cognitive workload when input modalities addressed the same modality as the

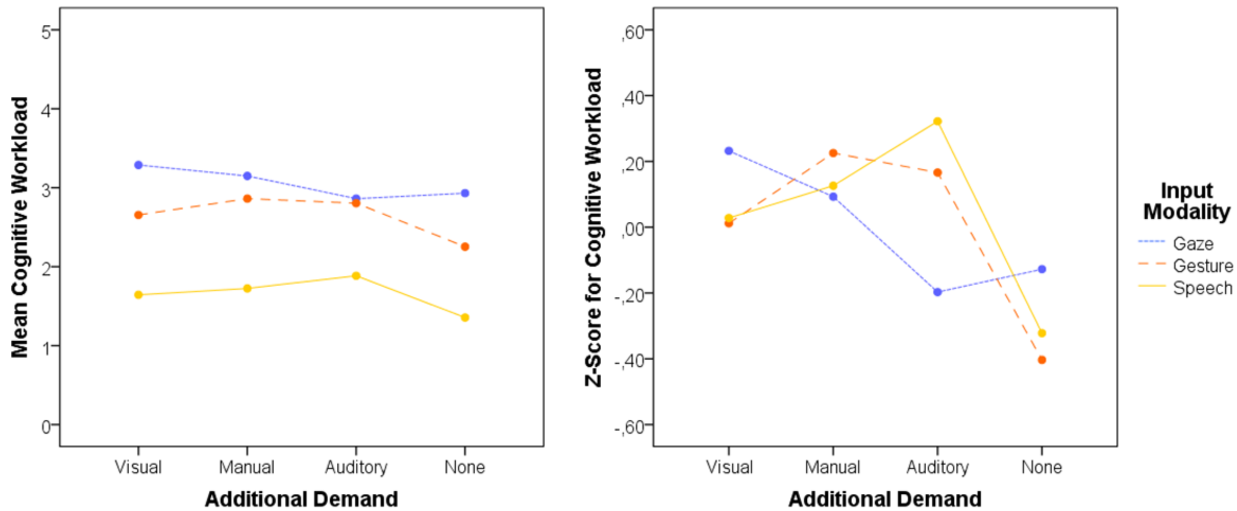


**Figure 6. Mean rankings of cognitive workload were calculated by the average ratings for effort of attention, interference and stress for each trial. Within each modality, resource competition led to the greatest increases of the perceived cognitive workload.**

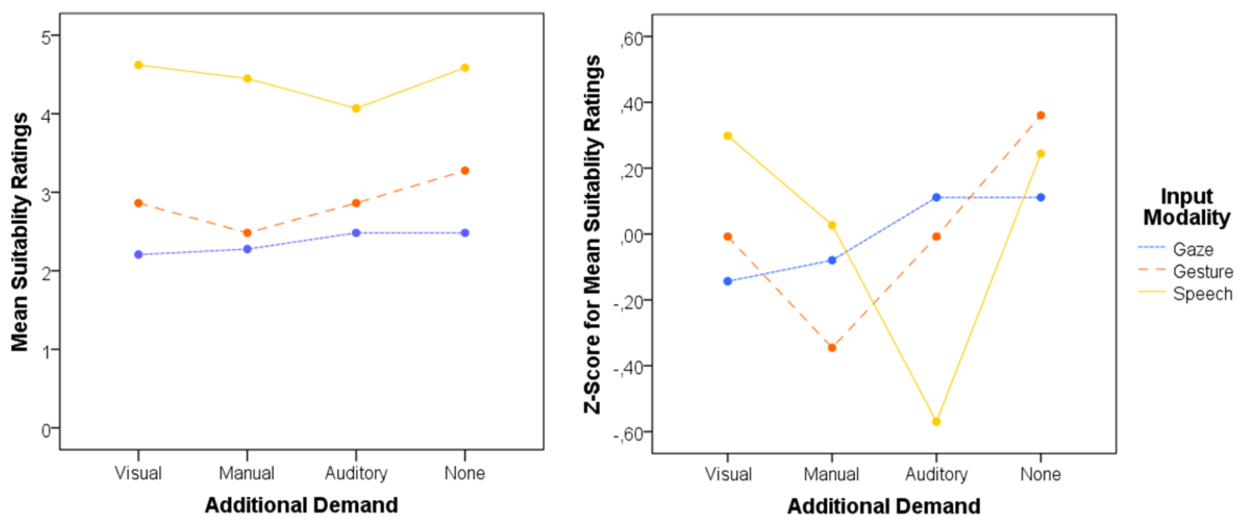


**Figure 7. Mean rankings of suitability for each input modality during additional demands. Note that lower values are negative here since they indicate a lower perceived suitability. Resource competition led to impairments of the perceived suitability for all input modalities.**

additional demand. A two-way repeated measures ANOVA reveals a significant effect of the type of additional demand on cognitive workload, $F(3,84) = 9.90, p < .01, \eta_p^2 = .261$. Bonferroni corrected post-hoc test show that the ratings for each additional demand are significantly worse than without additional demand (all $p < .01$). There was a significant interaction between both factors (additional demand and input modality), $F(6,168) = 2.77, p < .01, \eta_p^2 = .119$. The dimension regarding temporal demand showed the same interaction pattern as cognitive workload, yet the effect was not significant.

*Suitability Rating*

The effects of additional demands on perceived suitability of secondary task interaction is illustrated in Figure 7. We find the same interconnection as proposed in rankings of cognitive workload. Participants rated the suitability of the input modality lowest during trials in which the additional demand referred to the same resource. The impact on speech input best reflects this observation (see Figure 7). During trials with additional auditory demand, speech input was rated significantly worse than during other trials (all $p < .05$). A two-way repeated measures ANOVA showed a significant effect of additional demand $F(3,84) = 6.70, p < .01, \eta_p^2 = .193$. Bonferroni corrected post-hoc tests show that there was a significant difference between trials with no additional demand and those with manual and auditory demand (both $p < .01$). There was a significant interaction between additional demand and input modality ($F(6,168) = 4.87, p < .01, \eta_p^2 = .148$). This supports the observation that participants rated each input modality least suitable when it matched the applied additional demand (gaze input during visual demand, gesture input during manual demand, speech input during auditory demand).

| *Input modality* | *TCT* | *Suitability* | *Cog. Workload* |
|---|---|---|---|
| Gaze | .153 | ns. | .123 |
| Gesture | .106 | .128 | .194 |
| Speech | ns. | .285 | .216 |

**Table 1. Estimated effect sizes for the effect of additional demands on task completion times (TCT) and subjective measures for each modality.**

### Combined Impact

Inspection of Figures 5, 6 and 7 implies that additional demands induce significant effects of different sizes on secondary task completion times and subjective ratings for each modality. Therefore, we conducted one-way repeated measures ANOVAs for objective and subjective data to examine the effects of additional demand for each input modality independently. For *gaze input*, the effect on task completion times was significant. Suitability was not significantly affected and the effect on cognitive workload was only weak. For *speech input* in contrast, there was no effect on task completion times, but greater effects on suitability and cognitive workload. Finally, *gesture input* was significantly influenced in all three categories, at the cost of smaller effect sizes. Table 1 summarizes the estimated effect sizes ($\eta_p^2$) of additional demands on task completion times and subjective ratings. Consequently, overall conclusions regarding the specific effects of situational demands on secondary task controls must be drawn from the impact of both, objective measures and subjective ratings. In Figure 8 we illustrate the summed z-scores of task completion times and of the average from cognitive workload and suitability ratings. It illustrates that the combined impact on secondary task performance was comparable for all three modalities. We observe that there were also some smaller cross-modal impairments, but the greatest magnitude of effects was caused by the direct competition of resources.

## DISCUSSION

The results section described interaction effects between additional demands and input modalities on secondary task performance and subjective rankings. For driving performance, we did not find an interaction effect between the additional demand and input modality neither on the deviation of distance nor on lane keeping quality. This can be explained by the fact that participants were instructed to prioritize the primary task. In a potential case of resource competition between driving task and secondary task, participants neglected the later one in order to fulfill the driving task. Therefore, their driving performance was not influenced by specific combinations of input modality and additional demand. Instead, we observed a significant impairment on distance keeping for all input modalities during additional visual demand. Since the basic task of driving is mainly visually demanding (over all trials), there is already a resource competition between visual additional demand and the basic driving task [17]. Other studies observed similar behaviour and concluded that additional visual distraction often leads to reduction of speed in order to reduce the primary task demand (e.g. [1]). This explains greater deviations of requested distance to the leading vehicle during trials with additional visual demand.

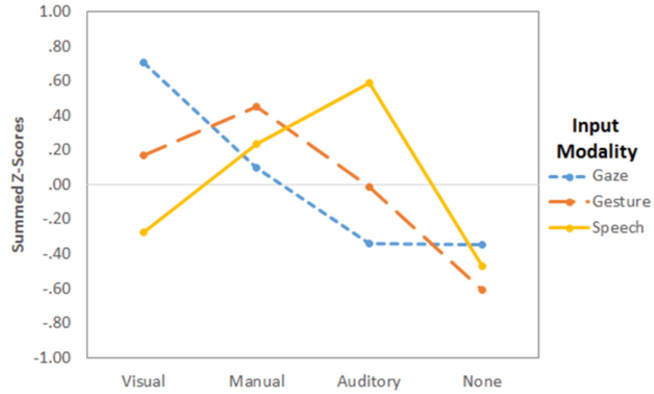


**Figure 8. Combined impact of additional demands on input modalities. Illustrated values represent the average of z-scores from secondary task completion times and the average of subjective ratings.**

In contrast to driving performance, participants' secondary task performance suffered from the effects of resource competition. Significant interaction effects on secondary task performance show that there is a differentiated influence on secondary task performance depending on the input modality and additional demand. This applies for task completion times, as well as for perceived cognitive workload and suitability. For each input modality, task completion times were longest, cognitive workload was greatest and suitability was lowest when the input modality and the additional task addressed the same resource. The estimated effect sizes of interaction on task completion times ($\eta_p^2 = .103$), suitability ($\eta_p^2 = .148$) and cognitive workload ($\eta_p^2 = .119$) indicated small effects of resource competition according to Cohen [2]. However, literature also suggests that demands of higher intensity (e.g. higher frequency of events) will result in an even greater disruptive effect of resource competition [16]. Accordingly, the potential advantage of using non-competing input modalities for interaction is likely to be even greater when the magnitude of demands on the driver increases.

Moreover, we found that the magnitude of impairments on objective and subjective measures differed between gaze, gesture and speech input. Therefore, it is crucial to include objective and subjective measures for investigation of natural interaction modalities in the car. For example, regarding task completion times only, there is no significant effect of additional demands on speech input (see Table 1). Concluding that speech input is not affected by additional demands of specific situations would be wrong, since there is a significant effect on subjective ratings that reflects the specific influence of additional auditory demand on speech input. Figure 6 shows that absolute ratings for cognitive workload of speech input was on a relatively low level compared to gesture and gaze. We assume that participants could maintain task completion times, even during resource competition, at the cost of increased cognitive workload and resulting impairments on perceived suitability. For gaze input, cognitive workload was already at a relatively high level without additional demand (see Figure 6). Therefore, participants could not make this compensation at the cost of higher cognitive workload, which resulted in an impairment on task completion times. For gesture input, cognitive workload was rated slightly below gaze input. A part of the impact of resource competition could be compensated by increased cognitive workload, while the rest resulted in increased task times. Despite these differences, inspection of Figure 8 suggests that the combined impact of objective and subjective data was comparable for all three modalities. This suggests that measurable impacts on task completion times only occur after a compensable level of cognitive workload is exceeded.

### Limitations

All figures showing absolute values depict that the input modality had a significant effect on our measures. However, these measures strongly depend on our specific implementation of gaze-, speech- and gesture input. Other implementations (e.g. gaze input with a different dwell time, faster speech recognition systems) might lead to differing measures (e.g. generally shorter completion times), or even switched rankings (e.g speech is faster than gestures). While this is a point to keep in mind, we did not want to compare the absolute performances of gaze, gesture and speech input for the selection task. Instead, we illustrated the effects on task completion time in relation to the mean values for each input modality by using z-scores. This enhances the generalizability of results and achieves a comparability between the effects on individual input modalities.

Speech recognition did not work perfectly for some participants who did not speak loud enough to cover the sounds in the driving simulation. For those participants we included an additional *Wizard-of-Oz control* that allowed the examiner to select the element himself, according to what participants said. The examiner imitated the typical response time of the system, in order to keep data for speech recognition consistent.

## CONCLUSION

We conducted a user study in a driving simulator with the aim to assess the performance of natural input modalities in different driving situations. Our results show that gaze, gesture and speech input perform worst during additional demands that address the associated resources. This underpins the necessity to differentiate between visual, auditory, and also manual demands that influence the driver in order to provide the best suited in-vehicle interaction.

Beyond that, we showed that gaze, gesture and speech input differ in *how* they are influenced by additional demands in the car. The effect of additional demands on gaze input mainly led to an increased of task completion times, whereas effects on speech input were reflected only in subjective ratings and effects in gesture input were split up between both factors. This knowledge is especially relevant for interaction designers that integrate gaze, gestures and speech in automotive concepts. They need to understand how situational demands can influence the efficiency and perceived demands of these input modalities. Furthermore, the results of our experiment help to interpret results of other studies that examine gaze, gesture or speech input by respecting the driving scenario that was used.

In this respect, our results support the integration of alternative input modalities. A greater variety of input modalities, which can be used alternatively, allows to avoid resource competition for a wider range of situations that occur on the road. Ideally, the driver can always pick one channel for interaction with the vehicle that is least claimed by environmental influences. This corresponds to the flexible way how people choose communication channels in human interaction depending on their situation. In this sense, alternative input modalities also contribute to a more natural interaction between driver and vehicle.

### Future Work

Interaction concepts in future vehicles could detect the types of demand that have an effect on the driver in certain situations and propose suited input modalities accordingly. Further research topics emerge from this idea. First, the development of an automated and reliable assessment of the level and types of demand that currently influence the driver. Second, the exploration of effective, yet non-obtrusive ways to propose a suited modality for interaction.

This will also be relevant in the context of highly automated driving. Demands on the driver will change fundamentally when the role of the driver switches from driving the vehicle to observing the vehicle and only intervening when needed. While driving manually, the visual and manual resource are claimed primarily. In an autonomous vehicle, the omission of the typical driving task leads to a relief of the visual and manual resource resulting in an increased potential of gaze and gesture input.

## REFERENCES

1. Jonathan F. Antin, Thomas A. Dingus, Melissa C. Hulse, and Walter W. Wierwille. 1990. An evaluation of the effectiveness and efficiency of an automobile moving-map navigational display. *International Journal of Man-Machine Studies* 33, 5 (1990), 581–594. DOI: http://dx.doi.org/10.1016/S0020-7373(05)80054-9

2. Jacob Cohen. 1988. Statistical power analysis for the behavioral sciences. (1988). DOI: http://dx.doi.org/10.1234/12345678

3. Johan Engström, Emma Johansson, and Joakim Östlund. 2005. Effects of visual and cognitive load in real and simulated motorway driving. *Transportation Research Part F: Traffic Psychology and Behaviour* 8, 2 SPEC. ISS. (2005), 97–120. DOI: http://dx.doi.org/10.1016/j.trf.2005.04.012

4. Anke Huckauf and Mario H. Urbina. 2011. On Object selection in gaze controlled systems. *ACM Transactions on Applied Perception* 8, 4 (2011), 1–14. DOI: http://dx.doi.org/10.1145/1870076.1870081

5. Dagmar Kern, Angela Mahr, Sandro Castronovo, Albrecht Schmidt, and Christian Müller. 2010. Making use of drivers' glances onto the screen for explicit gaze-based interaction. In *Proceedings of the 2nd International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI*

*'10*. ACM, New York, NY, USA, 110–116. DOI: http://dx.doi.org/10.1145/1969773.1969792

6. Sang Hun Lee, Se-One Yoon, and Jae Hoon Shin. 2015. On-wheel finger gesture control for in-vehicle systems on central consoles. In *Adjunct Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '15*. ACM, New York, NY, USA, 94–99. DOI: http://dx.doi.org/10.1145/2809730.2809739
7. Keenan R. May, Thomas M. Gable, and Bruce N. Walker. 2014. A Multimodal Air Gesture Interface for In Vehicle Menu Navigation. In *Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '14*. ACM, New York, NY, USA, 1–6. DOI: http://dx.doi.org/10.1145/2667239.2667280
8. Bruce Mehler, David Kidd, Bryan Reimer, Ian Reagan, Jonathan Dobres, and Anne McCartt. 2016. Multi-modal assessment of on-road demand of voice and manual phone calling and voice navigation entry across two embedded vehicle systems. *Ergonomics* 59, 3 (mar 2016), 344–367. DOI: http://dx.doi.org/10.1080/00140139.2015.1081412
9. Christian Müller, Garrett Weinberg, and Anthony Vetro. 2011. Multimodal input in the car, today and tomorrow. *IEEE Multimedia* 18, 1 (2011), 98–103. DOI: http://dx.doi.org/10.1109/MMUL.2011.14
10. Sharon Oviatt. 1999. Ten myths of multimodal interaction. *Commun. ACM* 42, 11 (1999), 74–81. DOI: http://dx.doi.org/10.1145/319382.319398
11. Annie Pauzié. 2008. A method to assess the driver mental workload: The driving activity load index (DALI). *IET Intelligent Transport Systems* 2, 4 (2008), 315. DOI: http://dx.doi.org/10.1049/iet-its:20080023
12. Bastian Pfleging, Stefan Schneegass, and Albrecht Schmidt. 2012. Multimodal interaction in the car - combining speech and gestures on the steering wheel. In *Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '12*. ACM, New York, NY, USA, 155–162. DOI: http://dx.doi.org/10.1145/2390256.2390282
13. David L. Strayer, Jason M. Watson, and Frank A. Drews. 2011. *Cognitive Distraction While Multitasking in the Automobile*. Vol. 54. 29–58 pages. DOI: http://dx.doi.org/10.1016/B978-0-12-385527-5.00002-4
14. Christopher D. Wickens. 1980. The Structure of Attentional Resources. *Attention and Performance VIII* 8 (1980), 239–257.
15. Christopher D. Wickens. 2002. Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science* 3, 2 (2002), 159–177. DOI: http://dx.doi.org/DOI10.1080/14639220210123806
16. Christopher D. Wickens, Diane L. Sandry, and Michael Vidulich. 1983. Compatibility and resource competition between modalities of input, central processing, and output. *Human factors* 25, 2 (1983), 227–248. DOI: http://dx.doi.org/10.1177/001872088302500209
17. Walter W. Wierwille. 1993. Demands on driver resources associated with introducing advanced technology into the vehicle. *Transportation Research Part C: Emerging Technologies* 1, 2 (jun 1993), 133–142. DOI: http://dx.doi.org/10.1016/0968-090X(93)90010-D