Journal on Multimodal User Interfaces
https://doi.org/10.1007/s12193-019-00301-2

ORIGINAL PAPER

Visualizing natural language interaction for conversational in-vehicle information systems to minimize driver distraction

Michael Braun[1,2] · Nora Broy[1] · Bastian Pfleging[2,3] · Florian Alt[2,4]

Received: 19 March 2018 / Accepted: 13 March 2019

Abstract
In this paper we investigate how natural language interfaces can be integrated with cars in a way such that their influence on driving performance is being minimized. In particular, we focus on how speech-based interaction can be supported through a visualization of the conversation. Our work is motivated by the fact that speech interfaces (like Alexa, Siri, Cortana, etc.) are increasingly finding their way into our everyday life. We expect such interfaces to become commonplace in vehicles in the future. Cars are a challenging environment, since speech interaction here is a secondary task that should not negatively affect the primary task, that is driving. At the outset of our work, we identify the design space for such interfaces. We then compare different visualization concepts in a driving simulator study with 64 participants. Our results yield that (1) text summaries support drivers in recalling information and enhances user experience but can also increase distraction, (2) the use of keywords minimizes cognitive load and influence on driving performance, and (3) the use of icons increases the attractiveness of the interface.

Keywords Human–computer interaction · Natural language interfaces · Automotive user interfaces

✉ Michael Braun
michael.bf.braun@bmw.de

Nora Broy
nora.nb.broy@bmw.de

Bastian Pfleging
b.pfleging@tue.nl

Florian Alt
florian.alt@unibw.de

1 BMW Group Research, New Technologies, Innovations, Munich, Germany

2 LMU Munich, Munich, Germany

3 Eindhoven University of Technology, Eindhoven, Netherlands

4 Bundeswehr University, Munich, Germany

Published online: 19 March 2019

Journal on Multimodal User Interfaces (Springer Nature)
(Pre-Print)

Visualizing Natural Language Interaction for Conversational In-Vehicle Information Systems to Minimize Driver Distraction

Michael Braun[1,2], Nora Broy[1], Bastian Pfleging[2,3] and Florian Alt[2,4]

Accepted: March 2019

Abstract In this paper we investigate how natural language interfaces can be integrated with cars in a way such that their influence on driving performance is being minimized. In particular, we focus on how speech-based interaction can be supported through a visualization of the conversation. Our work is motivated by the fact that speech interfaces (like Alexa, Siri, Cortana, etc.) are increasingly finding their way into our everyday life. We expect such interfaces to become commonplace in vehicles in the future. Cars are a challenging environment, since speech interaction here is a secondary task that should not negatively affect the primary task, that is driving. At the outset of our work, we identify the design space for such interfaces. We then compare different visualization concepts in a driving simulator study with 64 participants. Our results yield that (1) text summaries support drivers in recalling information and enhances user experience but can also increase distraction, (2) the use of keywords minimizes cognitive load and influence on driving performance, and (3) the use of icons increases the attractiveness of the interface.

Keywords Human-Computer Interaction · Natural Language Interfaces · Automotive User Interfaces

[1]BMW Group Research, New Technologies, Innovations
Munich, Germany
E-mail: michael.bf.braun@bmw.de
E-mail: nora.nb.broy@bmw.de

[2] LMU Munich
Munich, Germany

[3] Eindhoven University of Technology
Einhoven, Netherlands
E-mail: b.pfleging@tue.nl

[4] Bundeswehr University
Munich, Germany
E-mail: florian.alt@unibw.de

1 Introduction

In 1982 the general public made contact with the embodiment of futuristic cars: *K.I.T.T.* – the talking, autonomous supercar in the TV series *Knight Rider*. In contrast to state-of-the-art interfaces in cars, K.I.T.T.'s interface allows for interaction using natural language, rather than issuing specific commands that provide access to the car's features, e.g., accessing the navigation system and providing the destination.

Recent developments in natural language understanding [4, 22] and the rise of digital assistants like *Siri* and *Alexa* will make it possible in the future to interact with vehicles in the form of a natural spoken conversation where both the driver and the car participate in a similar manner. At the same time, this conversation will always be a secondary task that takes place as the user is focussing on another task, that is driving. As a result, drivers will constantly re-allocate their cognitive resources based on the current driving situation, ultimately resulting in a trade-off between how well they can follow the conversation and safely maneuver the car. As a solution to this, speech interaction can be supported by a visualization of the conversation, which allows the driver to follow up on the current conversation after interrupting it for handling a complex driving situation.

In this work, we focus on the design of such visualizations. At the outset, we conducted a literature review from which we derived a design space for visualizing natural language interaction in vehicles. In a subsequent driving simulator study (N=64) we then compared different visualizations (for example, full text, keywords), comparing how they impact cognitive load, driving performance, as well as the driver's experience. Our results show that (a) a text-based summary of the visualization in the form of keywords leads to better recognition and reduced cognitive load while not negatively influencing driving performance and (b) that the use of icons increases the attractiveness of the interface and hence the driver's experience. From our findings we derive a number of design recommendations. As speech interfaces are becoming ubiquitous, this research is relevant for both researchers and practitioners who are working with interfaces enabling natural language interaction in cars.

The contribution of this paper is threefold. Firstly, we introduce a design space for conversational in-vehicle information systems (IVIS) and determine viable implementations of visualizing text in such systems. Secondly, we present results of a driving simulator study with 64 participants to expose effects of different visualizations on understanding, distraction, driving performance, workload, and user experience. Thirdly, the results are translated into design recommendations for text visualizations in conversational IVIS, based on the study findings and feedback from automotive designers.

2 Background and Related Work

The first cars consisted of not much more than wheels, an engine, and a steering wheel. More than a century later we look at the very same concept, but vehicles evolved to fast, connected – even partially intelligent – allrounders. Thereby the human-car interface underwent a shift from being solely used to maneuver the car towards concurrently providing access to a myriad of multimedia and entertainment functions. As a result, user interfaces of modern cars need to support the driver during access to these functions while at the same time being optimized for safe use while driving [20], for example by minimizing distraction and eyes-off-the-road time.

2.1 Tasks in Automotive User Interfaces

The fundamental idea behind automotive user interfaces lies in matching more prominent positions with more important tasks and placing subsidiary tasks offside. Bubb defines three main types of tasks in the driving environment [8]:

Primary tasks are used to maneuver the car, to control direction and speed and to assess distances to surrounding objects or people. The common control devices for these kind of tasks are the steering wheel, the pedals for acceleration, clutch and brake, and the windshield [32,33].

Secondary tasks are additional safety-relevant functions like the speedometer, turn signals, windshield wipers, or control buttons for electronic safety systems. While we see a steady rise in functionalities in modern cars, we also experience an ongoing automation of the secondary task and digital instrument displays offer new grounds for innovation.

Tertiary tasks are made up of functionalities connected to multimedia, information and comfort features. Many of those are nowadays bundled in in-vehicle information systems (IVIS, [24]). These systems get more and more comprehensive with technological advancement, which amplifies the need for research on the minimization of distraction through such interfaces.

In-vehicle information systems are not constituent in many modern legislative texts due to their volatile and quickly changing nature. However, a need for regulations is present as those systems have weighty impact on road safety. Organizations such as NHTSA, ISO, and AAM issue guidelines based on research. AAM standards on distraction for example demands that IVIS should not induce single glance durations exceeding 2 seconds and drivers should not require more than 20 seconds of total glance time to displays and controls to complete a task [21]. We base our concepts on such standards for automotive user interfaces, e.g. [13,10,20] in order to provide a safe and efficient experience.

2.2 Distraction & Cognitive Load

With ongoing advances in automotive information and entertainment systems, more and more potential sources of distraction are introduced into the car while distraction is found to be involved in a majority of serious injury crashes [3]. White et al. state that drivers generally underestimate the degree of distraction and, thus, continue to multi-task while driving without concern for consequences [59], although any activity that diverts the driver's attention away from driving is potentially dangerous [49]. Competing activities to the driving task can usually be sorted into visual, manual, auditory, vocal, or cognitive distractions [17]. Drivers often have to deal with multiple tasks consisting of combinations of numerous stimuli at the same time, thus appropriate management of available attention ressources is highly important. With higher utilization rates of cognitive processing capacities, higher cognitive load ensues [45].

Common sources of elevated cognitive load while driving are visual distractions and demands connected to conversations. Engstrom et al. found lower ratings for own driving performance as a common effect of both auditory and visual load. However drivers recognized their worse driving behaviour less accurately during an auditory task than during a visual task [15]. This is especially dangerous as the risk of a collision quadruples during auditory distractions such as conversations on the phone [50]. An increase in risk also takes effect no matter if a hands-free device is used to make the call [36, 40]. Even conversations between driver and passenger impair the driver's situation awareness, especially for the positions of vehicles behind their own cars [26]. In general, distractions like cell phone usage while driving result in longer reaction times [9] and a higher risk for minor and major accidents [35].

2.3 Interaction with IVIS

Car manufacturers are striving to minimize distraction in the interaction process by combining new modalities to established approaches [41]. For example, the introduction of gestures can reduce distraction compared to haptical input [18]. Other approaches to multimodal interaction include gaze tracking to activate certain functionalities [44] or tactile interfaces [1]. Affective interfaces can gather information through sensors to select calming music or suggest a stress-free route [27], to recognize user frustration and act upon it [29], or to estimate the driver's cognitive load and adjust the interface accordingly [43].

Speech interfaces are another promising way to minimize distraction. Lo & Green summarize the outcome of 9 studies with the overall conclusion that driving performance is generally better and distraction is lower when using speech interfaces compared to visual-manual interfaces [38]. Tsimhoni et al. tested speech interfaces versus touch screen input for text entry tasks and found shorter task completion times and less degradation of vehicle control when using speech recognition [55]. However, speech interfaces can also lead to a high level of visual engagement when text is displayed [51] and to a decrease in task completion efficiency with audio-only systems [58, 28]. We expect an ideal compromise between multiple modalities can lead to an interaction concept with a minimum of distraction.

2.4 Conversational Assistants

Progress in the field of artificial intelligence will most likely lead to an enhancement of natural language recognition in voice-controlled IVIS, and enable the development of an intelligent personal assistant. In real life, human assistants possess the convenient ability of decreasing their boss's workload by doing the work for them. The same concept powers the urge for assistants in the car. When we look at the workload of drivers we can see this might be for the best. A study by Ehsani et al. finds 83% of teenage drivers engage in electronic device use while driving [14]. Rümelin et al. tested the concept of transferring work to an assistant, in their study embodied by the co-driver, and reported not only minimized workload for the driver but also an increased level of control over the situation for both driver and co-driver [52]. Furthermore, Nafari & Weaver report that interactions between systems and users are faster, more efficient, less error-prone, and easier to learn when the system translates queries into natural language output [42], just like in a conversation with an assistant.

Previous work by Yan et al. [60] and Large et al. [37] provides insights on how a conversational interface can be structured in order to limit workload, for example, through good error handling. We contribute to the spectrum of system design by finding an ideal interplay of speech and visual output to maximize driver safety. Drivers should be able to abandon any conversation when their full attention is needed on the road and then be able to return into the conversation seamlessly. Visual summaries could help the driver to refocus faster and make the interaction more efficient and safe.

2.5 Commercial Systems

While command-based speech interfaces have been state-of-the-art for quite a while now (at least in premium cars), interaction in the form of advanced conversations has not yet arrived in production vehicles. Speech recognizers today can differentiate between global and functional commands given in natural language, for example, 'play the Knight Rider Theme Song' is interpreted as global music mode and functional song choice. These systems come to use in cars by BMW, Audi, Mercedes-Benz, Tesla, and many more. However, real conversations, like in the Ford Concept Model U [48], have not yet been put on the road by car manufacturers. Other companies try to fill this gap by bringing their digital assistants into cars. Both Google and Apple are pushing for better integration of smart phones into car HMI, and of course Apple CarPlay comes with the assistant Siri[1] and Android Auto includes the speech assistant Google Now[2]. Tesla has been the first car manufacturer to support this connection by providing a public API for car functionalities, allowing users to, for example, 'summon' their vehicle by talking to their smart watch[3].

[1] `https://www.apple.com/ios/carplay/`, last access: 2017-08-12

[2] `https://www.android.com/auto/`, last access: 2017-08-12

[3] `https://itunes.apple.com/de/app/remote-s-for-tesla/id991623777`, last access: 2017-08-12

Amazon's assistant Alexa now is also available as add-on hardware for the dashboard and integration in cars has been announced by Ford, Volkswagen, Hyundai, and Volvo[4].

2.6 Summary

Our review of related literature shows that despite speech-based interfaces becoming ubiquitous, it is currently still limited in that speech interaction is often based on specific commands. At the same time, advances in natural language processing open new opportunities for making conversation with the car more natural, similar to talking to other passengers. Little is known as to how such conversations impact on the cognitive load of the driver and how well drivers can remember the conversation after interrupting it, for example, to handle a dangerous or demanding driving situation.

To close this gap, our research investigates different visualization strategies capable of supporting the driver as they shift their attention from the conversation to the driving situation and vice versa.

3 Dialog Visualization Concepts

In the following, we explore how a natural language interface can be designed for and be integrated with the car. In particular, we are interested in how such an interface can be designed in a way such that it minimizes driver distraction and cognitive load while at the same time optimizing perception.

3.1 Research Approach

To begin with, we analyzed related literature and existing approaches on conversational assistants regarding interaction techniques and channels. This data was then clustered in a workshop involving automotive designers, resulting in a design space for conversational IVIS which we iterated upon through several rounds of feedback. We then built an early design to test interesting variables with users which helped us identify variables for our final concepts and study.

3.2 Design Space

Natural language interaction has been introduced in various operating systems for smartphones and home automation. These implementations, as well as depictions from science fiction (for example, HAL 9000, Her) are mostly optimized for interactions with unlimited attention, as they rely on spoken words as main feedback channel. In cars, however, conversations are frequently interrupted by more important tasks, resulting in a need to safely pick up the conversation at a later point in time, and need to be designed with that in mind.

Previous work introduced design spaces with the goal to help designers understand the complexity of the design process as well as to identify directions for future research. Design spaces are available for many domains, including mobile phones [2] and large displays, but also for vehicles. For example, Kern and Schmidt [32] introduced a design space on automotive user interfaces, and Häuslschmid et al. [23] presented a design space for windshield displays. We believe our design space [6] to be similarly valuable for researchers and practitioners as they design natural language-based conversational interfaces for cars.

3.2.1 Methodology

This design space is the result of a search for meaningful variables of natural speech interaction systems in the automotive context. It can be seen as ground work for our concept as it sets dimensions and within them defines possible adjustments we can later tweak to build good interfaces.

For our work on a conversational in-vehicle information system (IVIS), we searched for mobile and automotive interfaces, speech interaction, affective and multimodal in-car UIs, and digital assistants within the repositories of digital libraries, resulting in 85 relevant publications (for example, [4,22,28,54]). Furthermore, we examined existing smart phone and home automation assistants (Apple Siri, Google Now, Microsoft Cortana, Amazon Alexa, SoundHound Hound) regarding interaction techniques and feedback channels. Next, we looked at the current landscape in speech-enabled IVIS (for example, BMW Voice Control System) and concept studies of conversational IVIS (for example, Nissan Pivo, Audi AIDA, Ford Model U) and consulted design spaces from other domains for inspiration and methodological advice. We then brought our collection of items into a focus group consisting of 6 automotive interface designers of BMW who identified groupings and worked on the arrangement following grounded theory. Our approach consisted of clustering germane items and combining closely related topics to then discuss their placement in the design space or to dismiss them if they were found inept.

[4] `https://www.theguardian.com/technology/2017/feb/07/amazon-alexa-car-logitech-zerotouch-voice-services-assistant`, last access: 2017-08-12

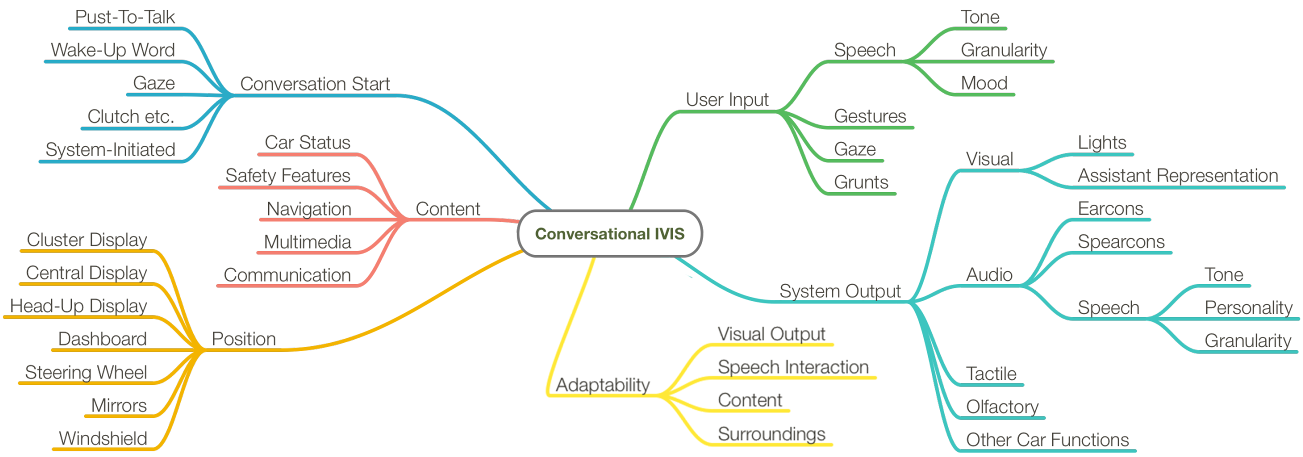

Fig. 1 Design space for conversational in-vehicle information systems

Building upon these ground works we compiled a first version of the design space which we then iterated upon. The resulting arrangement depicted in Figure 1 reflects the current landscape of possibilities for conversational IVIS. The design space is composed of 5 dimensions explained in the following section.

3.2.2 Dimensions of the Design Space

User Input: In conversations between humans, we mostly rely on speech to transfer messages. However gaze behavior, gestures, and body language also play a role. In interactions with IVIS, speech is also a suitable modality to transmit information and systems can extract additional data such as tone and moods from the sound of a voice [16]. Assumptions on driver emotions and affirmation or negation can even be derived from non-lexical sounds like grunts [57]. Other input modalities such as gestures and gaze control can also enhance the speech dialog in a multi-modal setting [41].

System Output: The output is another important topic to address within the dialog, as we need to think about a sensible way of conveying information without distracting the driver. We can, for example, support the driver by combining ambient lights, tactile feedback and smells with visual output such as status animations and audio feedback like speech or earcons [21]. Car functionalities like air conditioning, window movement and tint, or exhaust sound are other examples. The system might also have a personality of its own, so the speaker's choice of words, personality, and tone can play a role in how it is perceived.

Adaptability: A conversational IVIS can adapt its behavior, like timing and visual or auditory output, to the driver's cognitive load [30] or to traffic situations and the resulting emotions [5]. It can adapt its content to the audience and its surroundings, or allow for different input modalities depending on situational demands [7].

Conversation Start: Current IVIS mainly rely on push-to-talk buttons to activate voice control, while Siri and Alexa can detect their names to start an interaction. Other possibilities like look-to-talk, and gesture-to-talk could also enhance the experience, especially in an environment where the conversation is a tertiary task [44]. With intelligent assistants the initiative might even come from the system itself.

Content: Available contents shape the use cases for every interface and therefore need to be included in the design process. Information on the car's status, safety features, navigation and multimedia are common contents in modern cars. Increasing availability of cheap communication and online services now also allow for inclusion of, for example, video chats, home automation or social media.

Position: The Content of automotive UIs is mostly displayed in the central information display, on head-up displays, or on a digital instrument cluster [52]. Positioning an IVIS or a representation of the assistant in a novel position such as the windshield, in the mirrors, or on top of the dashboard can help establishing the assistant as a standalone companion in the car.

3.3 Early Design & Evaluation

Before running our extensive driving simulator study, we identified aspects to focus on and determined if there were widely popular or disliked concepts with negligible novelty. Therefore, we implemented a preliminary system design and let 10 users interact with it. The goal was to collect qualitative feedback to select the most promising concepts and reduce the number of experimental conditions for the main experiment.

3.3.1 System Design

We created seven distinct visualization concepts which implemented different characteristics of the design space, namely status position, dialog visualization (full text, keywords, icons) and granularity of the speech output (keywords, natural language) as shown in Table 1.

These variables were selected because related work did not contain satisfactory statements on how to implement status and text visualizations, nor how speech output should be structured to avoid distraction. The chosen combinations cover the most interesting variables and the reduction of time per participant allowed us to include more participants which we felt to be important to get better insights.

We used a within-subject design and collected subjective feedback from ten participants aged 29.2 ± 8.7 years. All participants work in automotive research.

3.3.2 Apparatus

We designed an interface which understands spoken natural language and renders the recognized text in real time. It can also identify keywords relevant to the use cases within the text and highlight them in full-text mode, respectively display them in keyword mode. In addition, the system responds with a human voice in keywords or natural language, depending on the text output. The spoken system output is visualized similar as the recognized speech input by the user. We define five states for the status visualization: idle, listening, processing, talking, and waiting, which we illustrate in a minimalist design (see Figure 2).

3.3.3 Procedure

As primary task, participants performed the Critical Tracking Task, simulating a measure of workload [47]. The secondary task consisted of talking to the conversational IVIS prototype, following a predefined agenda. The visual output of the prototype was shown at the position of the CID. Users experienced the different visualizations in permuted order. The conversation agenda included four use cases with conversation starts alternating between user and assistant. For example, the assistant would say that the car was running out of gas soon. The user would then ask for the location of the nearest gas station and get an answer from the system. Subsequently, the user had to ask a follow-up question, in this case how far the gas station was away which was again answered by the system. Once participants had experienced all concepts, they rated each tested variable (Figure 1) on personal preference.

3.3.4 Measures

After each condition, participants reported on user experience (AttrakDiff Mini [25]) and gave verbal feedback on the interaction. At the end of their turn they additionally performed a ranking on the experienced variables and were interviewed on their preferences.

3.3.5 Results

Users evaluated visualizations with keywords as more practical than concepts with full text. However, they found full text output more appealing to use than only keywords. Icons were judged as both practical and appealing. Displaying text was generally desired as it simplified error detection. In interviews led by the examiner, 7 out of 10 participant said to prefer visualizations of the output text and the system status during the conversation. 8 found keywords more suitable for the application than full text and 6 of them also liked the additional icons. Half of them each said they want the user's input text to be visualized and that full text display is more distracting than keywords. Differences in voice output (long / short mode) were mostly not recognized. Other participant feedback mainly concerned the user interface which should be modern, colorful, and more daring.

At the end of the experiment, participants had to rank for each investigated dimension of the design space the three different approaches that we presented in our experiment. Rank 1 was defined as best, rank three as worst. Figure 3 shows a boxplot of all the rankings except for the variable voice output. The results for voice output were very unambiguous as all participants favored full sentences over short answers ($Z = -2.8, p = 0.005$). The ranking for speech input visualization did not produce statistically significant results ($\chi^2 = 2.6, df = 2, p = 0.273$) but shows a general tendency towards displaying minimal or no input text. For the variable speech output visualization, the combination of keywords with icons scored the highest, keywords without icons placed second and full text came in last ($\chi^2 = 11.4, df = 2, p = 0.003$). The ranking of positions for the status visualization did not produce statistically significant results ($\chi^2 = 4.2, df = 2, p = 0.122$), however a preference for the visualization in general can be determined.

		1	2	3	4	5	6	7
Speech Input Visualization	Full Text	×	×					
	Keywords			×	×			
	None					×	×	×
Speech Output Visualization	Full Text	×		×		×		
	Keywords		×		×		×	
	Keywords & Icons							×
Voice Output	Long Mode	×	×	×		×	×	
	Short Mode				×			×
Assistant Status Visualization	Right	×		×				
	Left					×		×
	None		×		×		×	

Table 1 Variable combinations per visualization concept in the pre-study setup

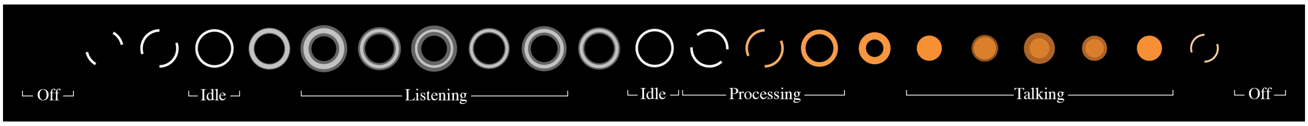

Fig. 2 States and animation frames for the status visualization. State changes are indicated by movement, output by color.

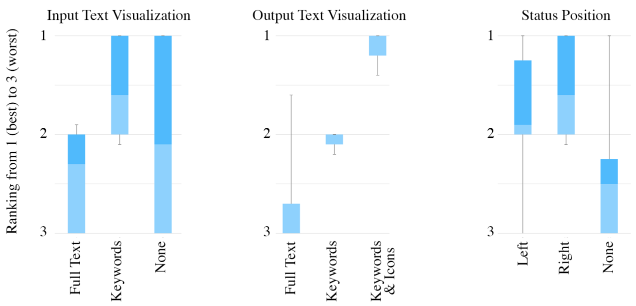

Fig. 3 Pre-study rankings for input and output text visualization and status position

3.3.6 Summary

Based on the findings from this preliminary study, we decided to focus primarily on the granularity of the text visualization and the dialog design for our main study. We derive from the pre-study that keywords seem interesting but full text also has its merits. Furthermore, participants clearly preferred the presence of a status visualization as well as naturalistic speech over spoken keywords.

3.4 Concept Development

With the feedback we went back to the drawing board to define the focus of the main study where we enable drivers to hold naturalistic conversations with their IVIS while keeping their attention on the road. Our designs contain different granularities of text visualizations which we envision to help the driver understand the conversation's flow after being occupied with driving for a while. Status visualization and spoken feedback in natural language were taken over from early designs, as they were unanimously liked by users. The concepts were refined with help of automotive designers at BMW and with our design space in mind.

3.5 Final Visualization Concepts

Figure 4 shows the final visualization concepts based on the preliminary user study. The concept *Status* acts as a baseline in our experiment. It holds all basic components like speech synthesis and a status visualization. However, the system does not display any text or other contents. Every other concept is built upon this platform. The concept *Full Text Chat* displays the conversation in full text mode. This design is inspired by messenger apps known from consumer electronics: User input is displayed in a box on the left side which is closest to the driver in left-hand drive vehicles, system output is displayed on the right side. This way input and output are spatially separated and users can easily recognize who said what. The boxes and text are additionally colored according to the sender of the message (user input is white, system output is orange) and key information is highlighted to further simplify understanding. In concept *Keyword Chat*, the only difference to concept *Full Text Chat* is the display of text in keyword mode. Only the most important information of the conversation is displayed within the boxes, separated by a dot to distinguish individual bits of information. We kept this concept similar to the one before to ensure direct comparability. This way we can investigate whether the granularity of displayed text is an important factor for distraction. Finally, the concept *Keyword Cards* also displays text in keywords and adds contextual icons to the visualization. This concept does not use the chat arrangement, related information is instead grouped into concise cards with the request on top and the response below. The text is again color-coded like in the previous concepts, yet not spatially separated. An icon on the left of the card symbolizes the domain when a request is detected and it can be updated to represent the content of the response if useful.

3.6 Hypotheses

We use the visualizations described above to test our hypotheses and to derive recommendations for future conversational IVIS. We infer the need for further guidelines from the fact that interaction in the car is getting more complex with intelligent personal assistants, compared to traditional IVIS. Furthermore, design principles from other consumer electronic devices can not directly be applied to automotive interfaces, since road safety can seriously be affected through driver distraction [54].

Generally, visualizing conversations with an assistant in the car bears the risk of introducing additional workload which can distract drivers from their primary task, thereby produce increased cognitive load [22,34] and cause decreased driving performance [15,26,56]. On the other hand, the presence of a written log could relax time constraints within a spoken dialog and thus improve the general quality of the conversation, help drivers to understand the system better, and generate a more positive UX [28,32,39]. We also have reason to assume that negative consequences can be prevented or at least mitigated by thoughtful design choices [11,39]. In our experiment we examine the following hypotheses:

- H_1 The visualization of text leads to better recognition of information than speech only. For example, previous work showed how visual support [31] can help users in reengaging with a navigation task that was interrupted due to the driving situation; however, with text it is not clear if drivers look at the text and how easy it is for them to perceive it.
- H_2 The visualization of text enhances user experience over a conversational interface without written text. State-of-the-art concepts such as Amazon Alexa omit the display of text for the sake of simplicity, while others (Siri, Google Now) extend the speech output with visual information. Both approaches have potential for the usage in an environment of diverted attention like driving, yet we are not sure if less visual distraction outweighs the additional context.
- H_3 Interactions with an audio-visual IVIS produce more cognitive load than with speech-only interfaces. Strayer et al. showed that voice-only assistants can lead to high cognitive loads [54], yet speech interfaces are considered safe compared to interacting with screens while driving. We assume that added information results in more overall load, although the distribution between channels could also lead to lower loads.
- H_4 The visualization of text while driving leads to deterioration of driving performance compared to an interface without text. Similar to H_3, we assume that added information diverts the driver from the primary task. However the allocation of information in voice and text could also allow for a more relaxed way of driving.
- H_5 Distracting effects of text visualizations can be mitigated by reducing the amount of displayed information. We propose that there is a 'sweet spot' for distracting effects of displayed text in accordance with NHTSA Guidelines [12].

4 Prototype

We implemented a prototype called *Conversational Assistant Interface Tester* (CAIT) to test the different concepts and evaluate them in a realistic manner. The application is connected to the driving simulation and accepts natural language speech input. Speech processing as well as speech synthesis for the agent were realized through third-party cloud platforms (Microsoft Cognitive Services Speech SDK[5], IBM Watson Text-to-Speech[6]). Intent recognition is implemented for each defined use case with a dictionary approach to remain flexible in case of changes. The visualization concepts are being displayed on the central information display (CID) inside the car while the examiner can control the experiment through a separate Wizard-of-Oz GUI. This way, a remote operator can perform the logical decisions of the system, without being observable for the participant. While CAIT is capable of real-time speech recognition and keyword detection and hyphenation within the implemented use cases, the sometimes complex flow of conversation is controlled by the examiner to enable more flexible reactions.

The user can interact with the system through speech input only. Each use case starts with a spoken explana-

[5] `https://azure.microsoft.com/de-de/services/cognitive-services/speech/`, last access: 2017-05-14

[6] `https://www.ibm.com/watson/developercloud/text-to-speech.html`, last access: 2017-05-14

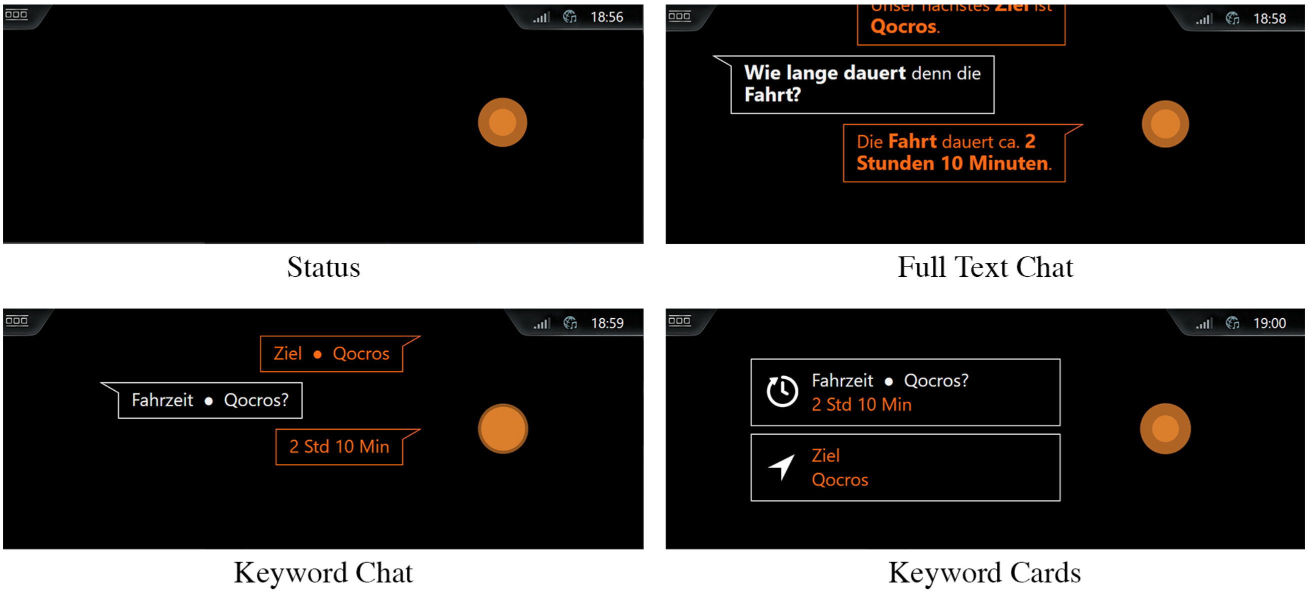

Fig. 4 Visualization concepts for the main study, displaying the feedback for the command to navigate to Qocros and an ensuing question on how long the ride will take.

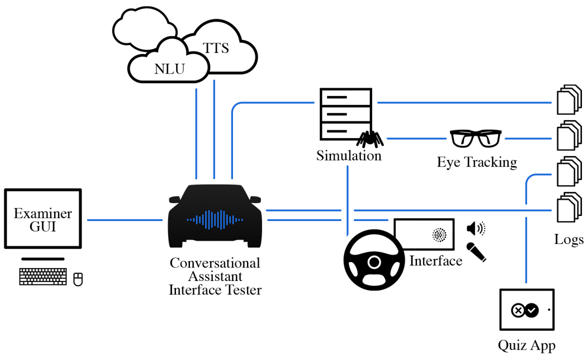

Fig. 5 System architecture for the experiment with CAIT

tion on what kind of information the user has to find out and is followed by an earcon [21] which signals that the system is ready to listen. We also built a mobile app for the quiz and questionnaires so users could conveniently answer them on a tablet inside the car. Figure 5 shows the relations between system components.

5 Main Study

As identified in the pre-study and its interviews, we identified text granularity and the general dialog presentation as the most interesting variables for a detailed investigation. Based on these learnings we designed a simulator-based experiment to investigate how to visualize a conversation with an intelligent voice-based assistant in the car.

5.1 Study Design

We conducted the study as a mixed between- and within-subject design, incorporating two levels of demand in the experimental rides as between-subject factor (stressful ride, $n = 32$; simple ride, $n = 32$). Each participant experienced four similar rides with the previously introduced visualization concepts as within-subject variable. Sequence permutation resulted in a 4×4 latin square and, thus, in 4 groups of each 8 participants for the within factor. Each driver experienced four different conditions.

5.2 Apparatus

The study was conducted in a state-of-the-art static driving simulator, consisting of a BMW 5-series mockup in front of a curved canvas which was used to project the driving scene on. Additional screens were mounted to reflect the driving situation behind the driver to the rear view mirrors. Figure 6 shows the actual simulator setup.

To enable speech interaction, the driver used a lavalier microphone which was connected to the CAIT prototype. Visualizations of the conversation with the assistant were shown on the CID. To monitor the driver's gaze behavior, the participants wore a mobile eye-tracker (Ergoneers Dikablis[7]) and gaze behavior was logged in the accompanying software D-Lab.

[7] `http://www.ergoneers.com/eye-tracking/dikablis-glasses/`, last access: 2017-05-13

Fig. 6 Car mockup in front of the curved screen

The examiner observed the participants from a nearby control room (outside the participant's field of view) where (s)he could operate the Wizard-of-Oz application and monitor the simulation, eye tracking, and voice recognition systems. A camera inside the cockpit and a two-way audio connection enabled participant surveillance and communication between control room and vehicle.

5.3 Tasks

During each experimental condition, the participants experienced typical in-car situations: They had to perform the *driving task* in a driving simulator and experienced the conversation with the intelligent voice-based assistant as an additional *non-driving-related task*.

Driving Task As stated before, the driving task was a between-subject factor with two levels (simple ride or stressful ride). The driving task for the *simple ride* consisted of following a dedicated car with a constant distance of 50 m and keeping the own car in the center of the lane. The setting for the simple ride was a straight highway with low traffic and a constant speed of 100 km/h without overtaking. For the *stressful ride* participants had to follow a car in about 50–70 m distance with the main goal not to loose contact with the car. The stressful ride was based on the same features as the simple ride, but the lead vehicle varied its speed between 110 and 120 km/h and overtook regularly in denser traffic. Additionally, third party vehicles acted as stress inducing factors by impeding the driver's attempts to catch up to the dedicated car. The participants were instructed to consider the driving task as the activity with highest priority.

Non-Driving Task The participants were given a non-driving task to handle while they were driving. This task had lower priority, meaning it should only be payed attention to when the driving situation allowed for it. The assignment was to converse with the car's assistant about a given use case and memorize certain information.

Use Cases To test the visualizations we defined 7 different use cases (see Figure 7). The user had to ask the assistant for information, resulting in an answer by means of audio feedback and the respective visualization format. The order of use cases was the same for every participant, answers by the system (i.e., names, locations, songs, etc.) varied for each ride.

5.4 Data Collection

In order to analyze the effect of the different visualization concepts, we collected different measurements during the experiment.

Recognition quiz After each condition, participants took a quiz on the information they were given by the conversational system. This was supposed to provide us with information on how well a visualization is suited for supporting information delivery. The quiz was implemented as an application running on a tablet PC. Questions were asked in the same order as the information was given in the use cases and participants had to identify the right one out of six possible answers. We queried recognition because drivers have to recognize information, e.g. signposts, more frequently than purely remember them.

Subjective questionnaires We used the *AttrakDiff Mini* [25] questionnaire, a standardized test for perceived user experience and the *Driving Activity Load Index* [46] to measure the experienced, subjective workload. Both questionnaires were incorporated in the aforementioned application, subsequent to the recognition quiz. We also asked the participants for feedback on the visualizations immediately after each ride.

Objective user performance An eye tracking system (Dikablis, see Figure 9) was used to gather data on gaze behavior during the tasks. In addition, we measured how well subjects performed the driving task with regard to lane keeping and distance control (deviation of headway variability) in order to make a statement about their distraction throughout the ride. The analysis of driving performance was not applicable for the stressful ride since the car in front moved on an unpredictable trajectory.

5.5 Procedure

After arriving at the lab, participants started with a generous familiarization period to get familiar with the simulation, the prototype, the use cases, and the related

Fig. 7 Use cases for the visualization assessment – participants had to ask for information within the active category

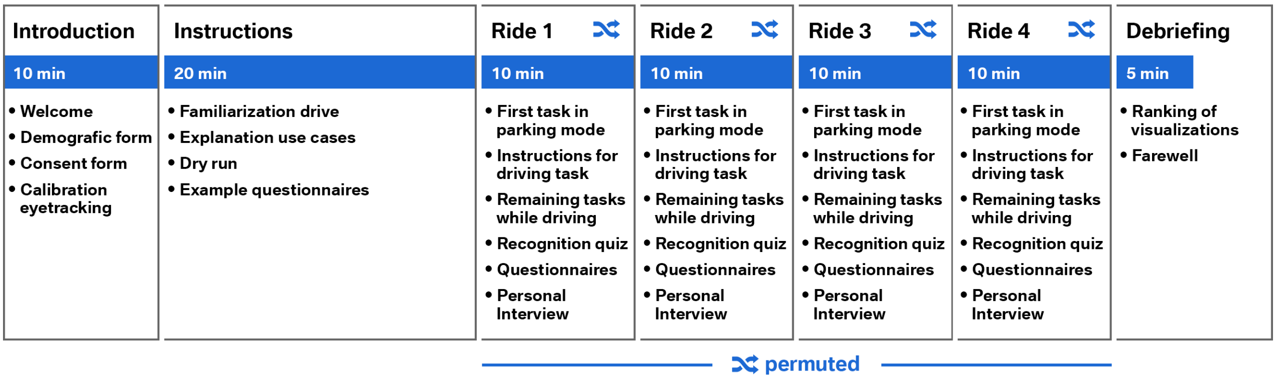

Fig. 8 Experiment schedule: Introductions and 4 assessment rides in permuted order

Fig. 9 Participant with tracking glasses, prototype implementation

questionnaires. Each condition started with the first use case in parking mode, so the user had a chance to see the visualization at rest. In addition, the participants received (a repetition of the) driving instructions. Each use case was preceeded with a spoken instruction of what the participant is expected to do in this use case. Except the first use case, the other six use cases were conducted while the car was in motion on the highway, with about 15 seconds break between each, resulting in a ride duration of approximately 6 minutes. The order of visualizations (i.e., the different runs) was determined by the experiment group. After the ride participants answered the recollection quiz, the questionnaires, and a personal interview. In the end they were presented with screen shots of the experienced variations, they could give feedback on what they saw, and finally they had to rank the visualizations in order of personal appreciation. Figure 8 shows the schedule of the experiment. Participants were not financially compensated for the experiment and in case of simulator sickness, the experiment was immediately terminated.

5.6 Participants

We performed sessions with 66 participants of whom 2 had to cancel halfway through the rides due to simulator sickness, which resulted in a total of $N = 64$. The 15 women and 49 men were between the age of 20 and 59 years ($M = 30.95$, $SD = 9.73$) and held a driver's license for $M = 13.36$ years ($SD = 9.37$). All but five participants stated that they had prior experience with voice interaction, of which 30 people said that they also used voice controls in the car and 16 categorized themselves as frequent users. 42 subjects also stated negative experiences with voice interaction, with bad recognition quality being the most mentioned cause.

5.7 Limitations

One limiting factor within this study was the recruiting process which only included employees of BMW in Munich, Germany. The sample consisted of more men than women and many participants indicated they had an academic and/or engineering background. This distribution does not represent a global entirety. However, it fits reasonably well to the demographics of premium mobility customers. Because of their affiliation to the company, all participants were familiar with the general operation of the test vehicle. Hence, we can rule out effects arising from participants having to get used to the car. As none of the participants were involved in the development of the concepts, their connection to the company should also not have had an influence on the outcomes of the study. We incorporated subjec-

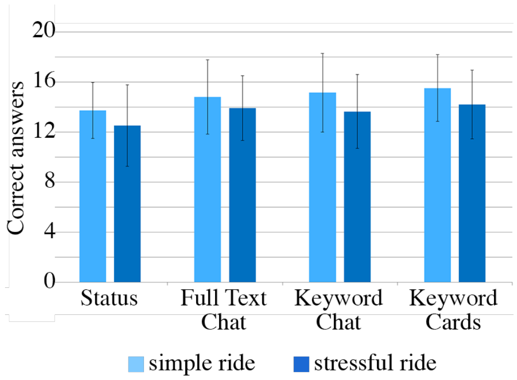

Fig. 10 Recognition rates for the quiz: users performed worse during the stressful ride, text display favored better performance

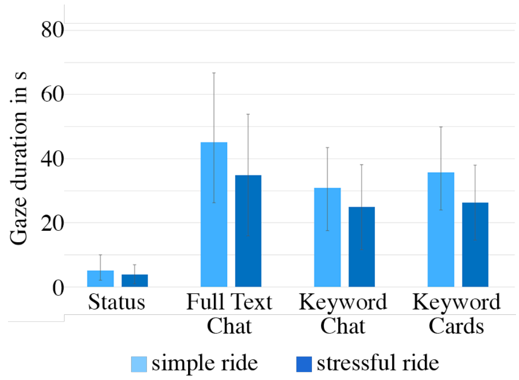

Fig. 11 Total gaze duration with standard deviation

	Single Glance Duration	Total Task Glance Time
Status	0.6 s	0.753 s
Full Text Chat	1.08 s	6.735 s
Keyword Chat	0.96 s	4.684 s
Keyword Cards	0.96 s	5.232 s

Table 2 85th percentile for single glance duration and mean total glance time per task

tive workload measurements and gaze tracking to assess the driver's cognitive load during interactions. Applying a standardized detection response task as used by Strayer et al. [54] might have been a worthwhile additional objective measure to support our findings. This study measured recognition success with simple use cases, future systems might be capable of more complex dialogues than what we have tested.

6 Results

During the course of our experiment, we collected data on the driver's gaze behavior and driving performance. In addition, we assessed their performance on recognizing information after interacting with the different visualization concepts and we asked them how attractive and how distracting they experienced the tested user interfaces. If applicable, a two-way ANOVA was used to determine between-subject effects as well as interaction effects of visualizations and ride difficulty. Results were regarded as statistically significant for $p < .05$.

6.1 Recognition Quiz

After each condition, participants took a quiz on the information they were given by the system during the ride. The quiz contained a total of 20 questions. For each question, the system offered 6 comparable answers among which the participant had to chose the correct answer. The data supports hypothesis H_1: visualizing text yields better results than no display of text ($F_{(3,\ 186)} = 7.334, p < 0.001$), with *Keyword Cards* performing best by a slight margin (see Figure 10). A pairwise comparison of p-values for quiz performance (t-test with bonferroni correction) shows better performance for the 3 visualizations incorporating written text than for the variant *Status* (no text displayed). There were no differences in performance for experienced or first users.

6.2 Gaze Data

The Dikabilis eye-tracking system enabled us to observe the participants' gaze behavior during the ride. The values for mean gaze duration to the CID ($F_{(3,\ 144)} = 84.578$, $p < .0001$), as well as gaze count ($F_{(2.4,\ 115.1)} = 96.997$, $p < .0001$) and total gaze duration ($F_{(2.7,\ 131.5)} = 87.19$, $p < .0001$) all deliver significant effects for the tested visualizations. As expected, visualizations which display more information also attract more and longer gazes, visible as the visualization concept *Status* shows significantly shorter glances than the other tested concepts (see Figure 11). The stressful ride allowed for less and shorter glances than the simple ride condition. Pairwise comparisons show no significant differences for mean gaze duration when text was displayed, keywords however led to a significantly lower total gaze count compared to full text.

The Alliance of Automobile Manufacturers provides guidelines for automotive user interfaces in which they state that automotive user interfaces should not induce glance durations exceeding 2 seconds and task completion should require no more than 20 seconds of total glance time [21]. Table 2 shows that all visualization concepts fulfill these requirements.

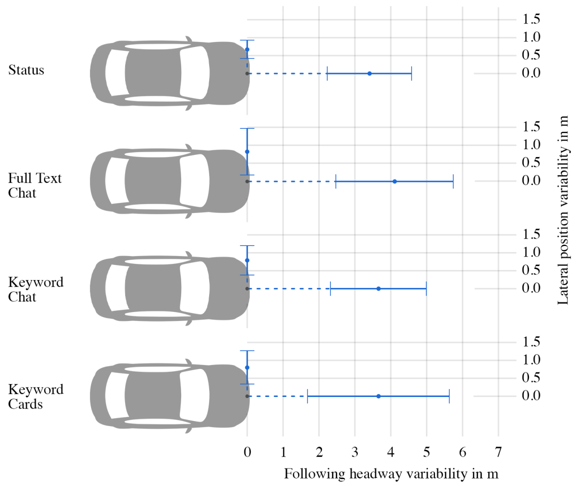

Fig. 12 Standard deviation of lateral position (SDLP) and headway variability for each visualization

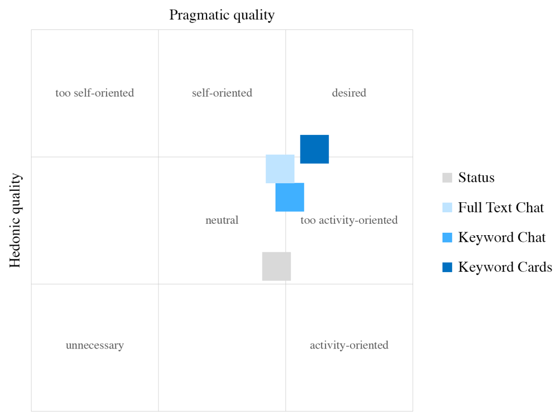

Fig. 13 Portfolio chart of ratings in the *AttrakDiff Mini* questionnaire

6.3 Driving Data

32 of 64 participants experienced a simple ride, meaning the driver was not put under a lot of stress by traffic or lane changes but only had to steer the car in the center of the lane and keep constant distance to the car in front. The prime objective of this ride was to provide comparable data, which we would not have gotten from our second, more demanding ride.

The first value we want to take a look at is the *standard deviation of lateral position (SDLP)*, or in other words the driver's performance in lane keeping [19,53]. Here, we measured performance for each task and calculated means over all of them for the final values. We found no significant effects between visualizations ($F_{(2.3,\ 70.3)} = 1.040$, $p = .368$). Another measure is the *following headway variability*, which was determined in the same manner. We found effects between visualizations ($F_{(2.5,\ 74.9)} = 3.587$, $p = .024$), more precisely there were significant differences ($p = .010$) between the concept *Status* and *Full Text Chat* with the latter causing higher variability (see Figure 12). Visualizations with only keywords however did not lead to a change in driving performance compared to no display of text.

6.4 User Experience

The AttrakDiff questionnaire produces two individual quality ratings, one being pragmatic quality, which varies significantly between visualizations ($F_{(2.8,\ 171.8)} = 13.47$, $p < 0.001$). In particular, the visualization concepts *Keyword Chat* and *Keyword Cards* are rated equally high whereas *Status* and *Full Text Chat* achieve a lower score. The second quality rating is hedonic quality which describes the perceived product quality on basis of personal needs. We can see meaningful differences between visualizations ($F_{(3,\ 186)} = 54.242$, $p < 0.001$) with the visualization concept *Status* scoring worse than each other visualization while *Full Text Chat* and *Keyword Cards* also outperform *Keyword Chat*. Neither pragmatic ($F_{(1,\ 62)} = 0.192$, $p = .663$) nor hedonic quality ratings ($F_{(1,\ 62)} = 0$, $p = 1$) were influenced by ride difficulty. Overall, the most attractive visualization turns out to be *Keyword Cards* while *Status* is rated worst (see Figure 13). Values for attractiveness showed the same tendencies, with significant differences between all variants except *Keyword Chat* and *Keyword Cards*.

6.5 Subjective Workload

The DALI questionnaire evaluates the cognitive load drivers are exposed to during the ride. We found significant results for visualizations ($F_{(2,\ 121.7)} = 5.713$, $p = .004$) as well as for rides ($F_{(1,\ 62)} = 7.219$, $p = .009$) in overall workload with the visualization concept *Full Text Chat* generating the highest workload while keyword visualizations resulted in significantly lower loads and *Status* produced no substantial differences to either variant (see Figure 14).

Interestingly, we found interaction effects between the rides ($F_{(2,\ 121.7)} = 3.501, p = 0.034$). This means answers for the stressful ride were noticeably more evenly allocated through concepts than for the simple ride. For the simple ride we found the lack of text visualization in the concept *Status* to result in more mental load than with Keyword approaches and in extremely more auditive load than with any other concept, yet it induced almost no visual load, quite understandably. The visualization *Full Text Chat* caused higher visual load than other concepts and also lead to a higher perception of interference between the primary and secondary task (this means applicants felt more distracted from the driving task through interaction with the system).

In the stressful ride we can only see the effects on visual and auditory load, the remaining values have very likely been blanketed be the high demand of the stressful ride. More precise data on the analysis of variance for all the sections can be seen in Table 3.

6.6 Interview

In the final interview, participants could voice their personal opinions about the four tested concept. *Status* was classified as least distracting by half of the people, however 62.5% also perceived it harder to remember the information for the quiz. 40.6% rated the visualization *Full Text Chat* as too distracting for the driving context. Feedback for the keywords concepts was more ambiguous, the main takeaway being that 34.4% liked the addition of icons.

6.7 Ranking

In the end participants had the chance to rank the tested concepts according to their liking. The majority put the visualization concept *Status* as least appealing, followed by *Full Text Chat* on third place and *Keyword Chat* on second. The most popular visualization turned out to be *Keyword Cards* (see Figure 15). A Friedman test shows the results are significant ($\chi^2 = 84.71, df = 3, p < 0.0001$).

7 Discussion and Design Recommendations

We present results of a driving simulator study with 64 participants with the goal of resolving questions concerning effects of different visualizations on understanding, distraction, driving performance, workload, and user experience. We can accept 4 of the 5 hypotheses we initially wanted to prove and with them we present recommendations for the design of conversational IVIS based on our findings.

H_1: The visualization of text leads to better recognition of information than speech only We initially proposed that visualizing text in conversational IVIS leads to better recognition of information. In fact, the performance in the quiz indicates better recognition of information when text was displayed. Hackenberg reports a similar effect [22]. We can say that additional icons do not benefit recognition significantly, although some participants remarked that they found them helpful for remembering certain information, for example, the weather or movie ratings. **Recommendation: showing a history of past interactions can help users to easily discover recently retrieved content.**

H_2: The visualization of text enhances user experience over a conversational interface without written text Results of the user experience questionnaire showed a higher pragmatic quality for text as keywords than for full text or no text, so keywords are found to be most useful for the task itself. Text display was also evaluated positively in regards to hedonic quality, with *Keyword Cards* being rated best. Overall, the combination of keywords and icons was considered most attractive and both chat concepts were also liked more than the system without text. This is also verified in the final ranking, so we can confidently accept hypothesis H_2. **Recommendation: display key information in a short and concise manner and enhance displayed text with visual context information such as icons or maps.**

H_3: Interactions with an audio-visual IVIS produce more cognitive load than with speech-only interfaces Another assumption we made before the study was that interactions with an audio-visual IVIS produce more cognitive load than with speech-only interfaces. Results from the DALI questionnaire actually suggest a high load for the full-text interface. The concept without text, however, scored comparably. The best overall workload score was achieved by concepts which used keywords to display text. This can be explained when we have a closer look at the single dimensions of the DALI: full text induced a high visual load in the driver, while the blank concept caused a high auditory and mental load and increased time pressure. Most likely this is due to the fact that users could not review the information if they did not pay attention at the very moment. In summary, both speech-only and extensive audio-visual interfaces cause rather high cognitive loads because they do not distribute the resources of the working memory efficiently between the processing units for visual and auditory information [45]. Lower workload was achieved with a keyword-based interface. Thus hypothesis H_3 cannot be accepted. On the contrary, audio-visual IVIS can even reduce cognitive load. **Recommendation: combine time-sensitive output (for example, speech) with persistent methods (for example, visualization) to avoid stressful multi-tasking situations.**

H_4: The visualization of text while driving leads to deterioration of driving performance compared to an interface without text. Regarding driving performance, we

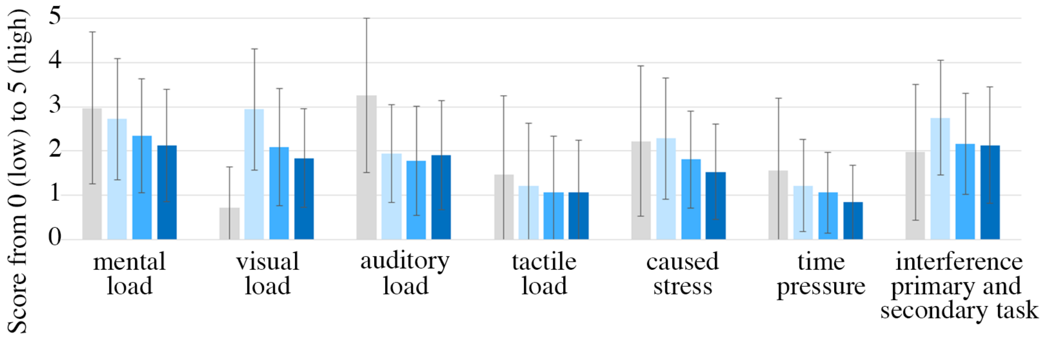

Fig. 14 Mean score and standard deviations of questions included in the *DALI* questionnaire

	Mental load	Visual load	Auditory load	Tactile load	Caused stress	Time pressure	Interference between primary and secondary task
df	2.37	3	2.248	2.601	2.322	2.324	2.514
df_{error}	146.911	186	139.378	161.255	143.962	144.108	155.657
F_{vis}	3.311	64.019	25.22	1.593	5.043	2.128	7.149
p_{vis}	0.032	0	0	0.199	0.005	0.115	0
$p_{vis*ride}$	0.146	0.043	0.12	0.157	0.322	0.034	0.487
F_{ride}	1.898	2.777	2.099	1.798	1.158	3.259	0.779
$p_{(1,62)ride}$	0.146	0.043	0.12	0.157	0.322	0.034	0.487

Table 3 Analysis of variance for each criterion in the DALI questionnaire

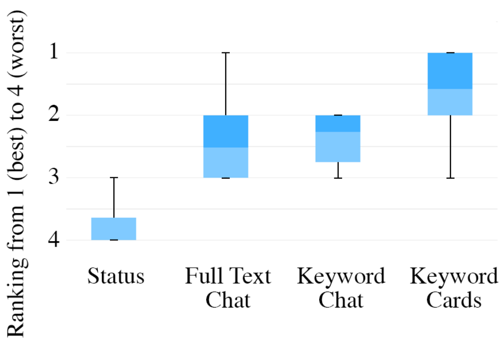

Fig. 15 Ranking of visualizations

can agree with hypothesis H_4 for full-text visualization, as this concept led to more headway variability than the interface without text. However, text displays in keyword format did not negatively affect any driving performance indicators, thus the general assumption that interactions with audio-visual IVIS impair driving performance has to be rejected. NHTSA guidelines propose a displayed text maximum of 30 characters per task for safely controlling a vehicle [12]. Based on our findings we support the limitation of displayed text in cars. **Recommendation: adhere to a limit of 30 characters when displaying text to the driver.**

H_5: Distracting effects of text visualizations can be mitigated by reducing the amount of displayed information

The final hypothesis we made was that distracting effects of visualizations can be mitigated by reducing the amount of displayed information. This can be affirmed based on the results of the DALI questionnaire mentioned above as well as by the gaze data we collected during the rides. Full text attracted a higher number of looks to the screen than keywords and in direct comparison, total glance time per task was 30% lower for the keyword chat than for the full text concept. **Recommendation: compress natural language into concise keywords to minimize visual distraction.**

Summary Alltogether, the tested speech interfaces with or without text visualizations would be appropriate for usage in vehicles according to AAM gaze standards [21]. However, the full-text visualization frequently surpasses the aforementioned limit of 30 characters per task and would, thus, not be found fit for in-vehicle use by NHTSA standards [12].

We recommend a system visualization which displays text in a short, structured way. Concept *Keyword Cards* is an example for such an approach. A positive impact on the user experience can additionally be achieved through the display of limited additional information (e.g. icons) and the separation of user and assistant text through different colors and/or spatial separation.

8 Conclusion & Future Work

In this paper, we present an analysis and experiment with conversational in-car interfaces, based on different approaches of visualizing natural speech. To understand the effect of different visualizations, we evaluated the concepts in a mixed design driving simulator study.

Participants performed better at recognizing information when it was not only said but also displayed. The users also preferred visualizations with text more than concepts without displayed information. A combination of keywords and icons in particular was most popular and induced the lowest workload. Our findings suggest that conversational IVIS benefit from text visualizations through enhanced speech recognition and improved user experience, as long as it does not distract the driver. Approaches which overstrain modalities can be seen as problematic, for example, audio-only or audio and full text compared to audio and keywords-only. We show that the driver's subjective workload can be reduced by distributing the task of information reception to auditory and visual channels. Overall, we found that distracting effects of visualizations can be mitigated by reducing the amount of displayed information and attractiveness can be increased by presenting well-structured information and additional enrichment, for instance through icons or limited media content.

Looking at future work, one interesting challenge could be the extraction of keywords in a use case independent environment. The sheer amount of data needed to cover all possible keywords could not be handled with conventional methods. With machine learning, however, the question arises if keyword detection is still a reasonable approach for finding context.

Other questions, for example, about the assistant's nature, arise form the design space presented in Chapter 3.2: how can we modify the assistant's speech characteristics, for example, tone, to fit the driver's mood? Would an assistant which communicates with non-lexical grunts for approval or negation appear more human? How well is such an assistant accepted in reality, in particular if the system initiates conversations autonomously? This design space was assembled from the related work we are familiar with and see as applicable for conversational in-vehicle information systems. It has potential for improvements, e.g. a more detailed breakdown of all branches, or the addition of context influences and design cues for conversational behavior.

After investigating speech and visual modalities, combinations of gaze and gestures are also plausible. But how would a tactile notification system work in the automotive setting? We can imagine a vibrating steering wheel, a tap on the shoulder by the seat or a tightening seat belt. Olfactory stimuli might also be used to communicate, be it gentle scents to lighten the mood or heavy smells to wake up the driver, there is a variety of possibilities.

Finally, research of positioning the virtual assistant in a distinct location could grant insights on the mental model users have of their car and assistant: is the assistant a personification of the car or is it operating its functions? Should it sit inside a display, switch positions, or live in the peripheral location of the rear-view mirror?

With this research we took a first step towards conversational in-vehicle information systems becoming ubiquitous. Beyond evidence on how the design of such systems – in particular a visualization of the conversations – impacts on the driver, we identified interesting directions for future research which we believe to be valuable for the automotive UI community. In addition, the derived recommendations are meant to support both researchers and practitioners as they design future conversational IVIS.

References

1. Asif, A., Boll, S.: Where to turn my car?: Comparison of a tactile display and a conventional car navigation system under high load condition. In: Proceedings of the 2nd International Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI '10, pp. 64–71. ACM, New York, NY, USA (2010). DOI 10.1145/1969773.1969786. URL `http://doi.acm.org/10.1145/1969773.1969786`
2. Ballagas, R., Borchers, J., Rohs, M., Sheridan, J.G.: The smart phone: a ubiquitous input device. IEEE Pervasive Computing **5**(1), 70–77 (2006). DOI 10.1109/MPRV.2006.18
3. Beanland, V., Fitzharris, M., Young, K.L., Lenné, M.G.: Driver inattention and driver distraction in serious casualty crashes: Data from the australian national crash in-depth study. Accident Analysis & Prevention **54**, 99–107 (2013). DOI 10.1016/j.aap.2012.12.043. URL `http://www.sciencedirect.com/science/article/pii/S000145751300047X`
4. Bellegarda, J.R.: Spoken language understanding for natural interaction: The siri experience. In: Natural Interaction with Robots, Knowbots and Smartphones: Putting Spoken Dialog Systems into Practice, pp. 3–14. Springer, New York, NY, USA (2014). DOI 10.1007/978-1-4614-8280-2. URL `http://link.springer.com/10.1007/978-1-4614-8280-2`
5. Braun, M., Pfleging, B., Alt, F.: A survey to understand emotional situations on the road and what they mean for affective automotive uis. Multimodal Technologies and Interaction **2**(4) (2018). DOI 10.3390/mti2040075. URL `http://www.mdpi.com/2414-4088/2/4/75`
6. Braun, M., Pfleging, B., Broy, N., Alt, F.: A design space for conversational in-vehicle information systems. In: Proceedings of the 19th International Conference on Human-computer Interaction with Mobile Devices and

Services, MobileHCI '17. ACM, New York, NY, USA (2017)
7. Braun, M., Völkel, S.T., Wiegand, G., Puls, T., Steidl, D., Weiß, D., Alt, F.: The smile is the new like: Controlling music with facial expressions to minimize driver distraction. In: Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia, MUM '18. ACM, New York, NY, USA (2018). DOI 10.1145/3282894.3289729. URL http://dx.doi.org/10.1145/3282894.3289729
8. Bubb, H.: Fahrerassistenz - primär ein beitrag zum komfort oder fuer die sicherheit? In: Der Fahrer im 21. Jahrhundert : Anforderungen, Anwendungen, Aspekte für Mensch-Maschine-Systeme; Tagung Braunschweig, 2. und 3. Juni 2003, VDI-Berichte 1768, pp. 25–44. VDI-Verlag, Düsseldorf, Germany (2003)
9. Caird, J.K., Willness, C.R., Steel, P., Scialfa, C.: A meta-analysis of the effects of cell phones on driver performance. Accident Analysis & Prevention **40**(4), 1282–1293 (2008)
10. Commission of the European Communities: Commission recommendation of 22 december 2006 on safe and efficient in-vehicle information and communication systems: Update of the european statement of principles on human machine interface (2007/78/ec). Official Journal of the European Union **50**(L32), 200–241 (2007). URL http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2007.032.01.0200.01.ENG
11. Cooper, J.M., Strayer, D.L.: Mental workload of voice interactions with 6 real-world driver interfaces. In: Proceedings of the Eighth International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design, DA '15, pp. 2–8 (2015). URL http://drivingassessment.uiowa.edu/sites/default/files/DA2015/papers/002.pdf
12. Department of Transportation, National Highway Traffic Safety Administration: Visual-manual nhtsa driver distraction guidelines for in-vehicle electronic device. Federal Register **78**(81), 24,817–24,890 (2013). URL https://federalregister.gov/a/2013-09883
13. on Driver Distraction, I.T.F.G.: Report on user interface requirements for automotive applications. Tech. rep., International Telecommunication Union (2013)
14. Ehsani, J., Li, K., Simons-Morton, B.G.: Teenage drivers portable electronic device use while driving. In: Proceedings of the Eighth International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design, DA '15, pp. 219–225 (2015). URL http://drivingassessment.uiowa.edu/sites/default/files/DA2015/papers/034.pdf
15. Engström, J., Johansson, E., Östlund, J.: Effects of visual and cognitive load in real and simulated motorway driving. Transportation Research Part F: Traffic Psychology and Behaviour **8**(2), 97–120 (2005). DOI 10.1016/j.trf.2005.04.012. URL http://www.sciencedirect.com/science/article/pii/S1369847805000185. The relationship between distraction and driving performance: towards a test regime for in-vehicle information systems In-vehicle information systems
16. Eyben, F., Wöllmer, M., Schuller, B.: Opensmile: the munich versatile and fast open-source audio feature extractor. In: Proceedings of the 18th ACM international conference on Multimedia, pp. 1459–1462. ACM (2010)
17. Foley, J.P., Young, R., Angell, L., Domeyer, J.E.: Towards operationalizing driver distraction. In: Proceedings of the Seventh International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design, pp. 57–63 (2013). URL http://drivingassessment.uiowa.edu/sites/default/files/DA2013/Papers/010_Foley_0.pdf
18. Geiger, M., Zobl, M., Bengler, K., Lang, M.: Intermodal differences in distraction effects while controlling automotive user interfaces. In: Proceedings of HCI International 2001 Volume 1: Usability Evaluation and Interface Design: Cognitive engineering, intelligent agents and virtual reality., HCII '01. Lawrence Erlbaum, New Jersey, NJ, USA (2001)
19. Green, P.: Standard definitions for driving measures and statistics: Overview and status of recommended practice j2944. In: Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI '13, pp. 184–191. ACM, New York, NY, USA (2013). DOI 10.1145/2516540.2516542. URL http://doi.acm.org/10.1145/2516540.2516542
20. Green, P.A.: Driver interface/hmi standards to minimize driver distraction/overload. In: Convergence 2008 Conference Proceedings. Society of Automotive Engineers, SAE International, Warrendale, PA, USA (2008)
21. Group, A.D.F.T.W.: Statement of principles, criteria and verification procedures on driver interactions with advanced in- vehicle information and communication systems. Tech. rep., Alliance of Automobile Manufacturers (2006)
22. Hackenberg, L., Bongartz, S., Härtle, C., Leiber, P., Baumgarten, T., Sison, J.A.: International evaluation of nlu benefits in the domain of in-vehicle speech dialog systems. In: Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI '13, pp. 114–120. ACM, New York, NY, USA (2013). DOI 10.1145/2516540.2516553. URL http://doi.acm.org/10.1145/2516540.2516553
23. Haeuslschmid, R., Pfleging, B., Alt, F.: A design space to support the development of windshield applications for the car. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pp. 5076–5091. ACM (2016)
24. Harvey, C., Stanton, N.A., Pickering, C.A., McDonald, M., Zheng, P.: In-vehicle information systems to meet the needs of drivers. International Journal of Human-Computer Interaction **27**(6), 505–522 (2011). DOI 10.1080/10447318.2011.555296. URL http://dx.doi.org/10.1080/10447318.2011.555296
25. Hassenzahl, M., Koller, F., Burmester, M.: Der user experience (ux) auf der spur: Zum einsatz von www.attrakdiff.de. In: H. Brau, S. Diefenbach, M. Hassenzahl, F. Koller, M. Peissner, K. Röse (eds.) Tagungsband UP08, pp. 78–82. Fraunhofer Verlag, Stuttgart (2008)
26. Heenan, A., Herdman, C.M., Brown, M.S., Robert, N.: Effects of conversation on situation awareness and working memory in simulated driving. Human Factors **56**(6), 1077–1092 (2014). DOI 10.1177/0018720813519265. URL http://dx.doi.org/10.1177/0018720813519265. PMID: 25277018
27. Hernandez, J., McDuff, D., Benavides, X., Amores, J., Maes, P., Picard, R.: Autoemotive: Bringing empathy to the driving experience to manage stress. In: Proceedings of the 2014 Companion Publication on Designing Interactive Systems, DIS Companion '14, pp. 53–56. ACM, New York, NY, USA (2014). DOI 10.1145/2598784.2602780. URL http://doi.acm.org/10.1145/2598784.2602780
28. Jakus, G., Dicke, C., Sodnik, J.: A user study of auditory, head-up and multi-modal displays in vehicles. Applied

Ergonomics **46, Part A**, 184–192 (2015). DOI 10.1016/j.apergo.2014.08.008. URL http://www.sciencedirect.com/science/article/pii/S0003687014001471
29. Kapoor, A., Burleson, W., Picard, R.W.: Automatic prediction of frustration. International Journal of Human-Computer Studies **65**(8), 724–736 (2007). DOI 10.1016/j.ijhcs.2007.02.003. URL http://www.sciencedirect.com/science/article/pii/S1071581907000377
30. Kennington, C., Kousidis, S., Baumann, T., Buschmeier, H., Kopp, S., Schlangen, D.: Better driving and recall when in-car information presentation uses situationally-aware incremental speech output generation. In: Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI '14, pp. 7:1–7:7. ACM, New York, NY, USA (2014). DOI 10.1145/2667317.2667332. URL http://doi.acm.org/10.1145/2667317.2667332
31. Kern, D., Marshall, P., Schmidt, A.: Gazemarks: gaze-based visual placeholders to ease attention switching. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 2093–2102. ACM (2010)
32. Kern, D., Schmidt, A.: Design space for driver-based automotive user interfaces. In: Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI '09, pp. 3–10. ACM, New York, NY, USA (2009). DOI 10.1145/1620509.1620511. URL http://doi.acm.org/10.1145/1620509.1620511
33. Kim, S., Dey, A.K.: Simulated augmented reality windshield display as a cognitive mapping aid for elder driver navigation. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '09, pp. 133–142. ACM, New York, NY, USA (2009). DOI 10.1145/1518701.1518724. URL http://doi.acm.org/10.1145/1518701.1518724
34. Kountouriotis, G.K., Wilkie, R.M., Gardner, P.H., Merat, N.: Looking and thinking when driving: The impact of gaze and cognitive load on steering. Transportation Research Part F: Traffic Psychology and Behaviour **34**, 108–121 (2015). DOI 10.1016/j.trf.2015.07.012. URL http://www.sciencedirect.com/science/article/pii/S1369847815001163
35. Laberge-Nadeau, C., Maag, U., Bellavance, F., Lapierre, S.D., Desjardins, D., Messier, S., Saidi, A.: Wireless telephones and the risk of road crashes. Accident Analysis & Prevention **35**(5), 649–660 (2003)
36. Lamble, D., Kauranen, T., Laakso, M., Summala, H.: Cognitive load and detection thresholds in car following situations: safety implications for using mobile (cellular) telephones while driving. Accident Analysis & Prevention **31**(6), 617–623 (1999)
37. Large, D.R., Burnett, G., Anyasodo, B., Skrypchuk, L.: Assessing cognitive demand during natural language interactions with a digital driving assistant. In: Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Automotive'UI 16, pp. 67–74. ACM, New York, NY, USA (2016). DOI 10.1145/3003715.3005408. URL http://doi.acm.org/10.1145/3003715.3005408
38. Lo, V.E.W., Green, P.A.: Development and evaluation of automotive speech interfaces: Useful information from the human factors and the related literature. International Journal of Vehicular Technology **2013**, 924,170:1–924,170:13 (2013). DOI 10.1155/2013/924170
39. Luger, E., Sellen, A.: "like having a really bad pa": The gulf between user expectation and experience of conversational agents. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16, pp. 5286–5297. ACM, New York, NY, USA (2016). DOI 10.1145/2858036.2858288. URL http://doi.acm.org/10.1145/2858036.2858288
40. McEvoy, S.P., Stevenson, M.R., McCartt, A.T., Woodward, M., Haworth, C., Palamara, P., Cercarelli, R.: Role of mobile phones in motor vehicle crashes resulting in hospital attendance: a case-crossover study. Bmj **331**(7514), 428 (2005)
41. Müller, C., Weinberg, G.: Multimodal input in the car, today and tomorrow. IEEE MultiMedia **18**(1), 98–103 (2011). DOI 10.1109/MMUL.2011.14
42. Nafari, M., Weaver, C.: Augmenting visualization with natural language translation of interaction: A usability study. Computer Graphics Forum **32**(3pt4), 391–400 (2013). DOI 10.1111/cgf.12126. URL http://dx.doi.org/10.1111/cgf.12126
43. Neßelrath, R., Feld, M.: Towards a cognitive load ready multimodal dialogue system for in-vehicle human-machine interaction. In: Adjunct Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI Adjunct '13, pp. 49–52 (2013). URL http://www.auto-ui.org/13/docs/aui_adjunct_proceedings_final.pdf\#page=49
44. Oh, A., Fox, H., Van Kleek, M., Adler, A., Gajos, K., Morency, L.P., Darrell, T.: Evaluating look-to-talk: A gaze-aware interface in a collaborative environment. In: CHI '02 Extended Abstracts on Human Factors in Computing Systems, CHI EA '02, pp. 650–651. ACM, New York, NY, USA (2002). DOI 10.1145/506443.506528. URL http://doi.acm.org/10.1145/506443.506528
45. Paas, F., Tuovinen, J.E., Tabbers, H., Van Gerven, P.W.M.: Cognitive load measurement as a means to advance cognitive load theory. Educational Psychologist **38**(1), 63–71 (2003). DOI 10.1207/S15326985EP3801_8. URL http://dx.doi.org/10.1207/S15326985EP3801_8
46. Pauzié, A.: A method to assess the driver mental workload: The driving activity load index (dali). Intelligent Transport Systems, IET **2**(4), 315–322 (2008). DOI 10.1049/iet-its:20080023. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4706518&tag=1
47. Petzoldt, T., Bellem, H., Krems, J.F.: The critical tracking task: A potentially useful method to assess driver distraction? Human Factors **56**(4), 789–808 (2014). DOI 10.1177/0018720813501864
48. Pieraccini, R., Dayanidhi, K., Bloom, J., Dahan, J.G., Phillips, M., Goodman, B.R., Prasad, K.V.: Multimodal conversational systems for automobiles. Commun. ACM **47**(1), 47–49 (2004). DOI 10.1145/962081.962104. URL http://doi.acm.org/10.1145/962081.962104
49. Ranney, T.A., Garrott, W.R., Goodman, M.J.: Nhtsa driver distraction research: Past, present, and future. In: Proceedings of the 17th International Technical Conference on the Enhanced Safety of Vehicles. U.S. Department of Transportation, National Highway Traffic Safety Administration (2001). URL http://www-nrd.nhtsa.dot.gov/pdf/esv/esv17/proceed/00177.pdf
50. Redelmeier, D.A., Tibshirani, R.J.: Association between cellular-telephone calls and motor vehicle collisions. New England Journal of Medicine **336**(7), 453–458 (1997)
51. Reimer, B., Mehler, B.: The effects of a production level "voice-command" interface on driver behavior: Summary findings on reported workload, physiology, visual attention, and driving performance. Tech. Rep. White Paper 2013-18A, MIT AgeLab - Massachusetts Institute of Technology (2013). URL

http://agelab.mit.edu/files/MIT_AgeLab_White_Paper_2013-18A_(Voice_Interfaces).pdf
52. Rümelin, S., Siegl, P., Butz, A.: Could you please... investigating cooperation in the car. In: Adjunct Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI Adjunct '13, pp. 61–64 (2013). URL http://www.auto-ui.org/13/docs/aui_adjunct_proceedings_final.pdf\#page=63
53. SAE International: Operational definitions of driving performance measures and statistics. Tech. Rep. SAE J2944, SAE International (2013). URL http://standards.sae.org/wip/j2944/. Proposed Draft
54. Strayer, D.L., Cooper, J.M., Turrill, J., Coleman, J.R., Hopman, R.J.: The smartphone and the driver's cognitive workload: A comparison of apple, google, and microsoft's intelligent personal assistants. Tech. rep., AAA Foundation for Traffic Safety, Washington, DC, USA (2015). URL https://www.aaafoundation.org/sites/default/files/strayerIIIa_FINALREPORT.pdf
55. Tsimhoni, O., Smith, D., Green, P.: Address entry while driving: Speech recognition versus a touch-screen keyboard. Human Factors **46**(4), 600–610 (2004). DOI 10.1518/hfes.46.4.600.56813. URL http://dx.doi.org/10.1518/hfes.46.4.600.56813. PMID: 15709323
56. Vollrath, M., Maciej, J., Niederée, U.: In-car distraction study - final report. Tech. rep., Kognitions-und Ingenieurpsychologie, Technische Universität Carolo-Wilhelmina zu Braunschweig, Braunschweig, Germany (2008)
57. Ward, N.: Non-lexical conversational sounds in american english. Pragmatics & Cognition **14**(1), 129–182 (2006)
58. Weinberg, G., Harsham, B., Medenica, Z.: Investigating huds for the presentation of choice lists in car navigation systems. In: Proceedings of the Sixth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, DA '11, pp. 195–202 (2011)
59. White, M.P., Eiser, J.R., Harris, P.R.: Risk perceptions of mobile phone use while driving. Risk Analysis **24**(2), 323–334 (2004). DOI 10.1111/j.0272-4332.2004.00434.x. URL http://dx.doi.org/10.1111/j.0272-4332.2004.00434.x
60. Yan, B., Weng, F., Feng, Z., Ratiu, F., Raya, M., Meng, Y., Varges, S., Purver, M., Lien, A., Scheideck, T., Raghunathan, B., Lin, F., Mishra, R., Lathrop, B., Zhang, Z., Bratt, H., Peters, S.: A conversational in-car dialog system. In: Proceedings of Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations, NAACL-Demonstrations '07, pp. 23–24. Association for Computational Linguistics, Stroudsburg, PA, USA (2007). URL http://dl.acm.org/citation.cfm?id=1614164.1614176