A Design Space for Conversational In-Vehicle Information Systems

Michael Braun
BMW Group
85748 Garching, Germany
michael.bf.braun@bmw.de

Bastian Pfleging
LMU Munich
80333 München, Germany
bastian.pfleging@ifi.lmu.de

Nora Broy
BMW Group
85748 Garching, Germany
nora.nb.broy@bmw.de

Florian Alt
LMU Munich
80333 München, Germany
florian.alt@ifi.lmu.de

MobileHCI '17, September 4–7, 2017, Vienna, Austria
ACM 978-1-4503-5075-4/17/09.
https://doi.org/10.1145/3098279.3122122

Abstract

In this paper we chart a design space for conversational in-vehicle information systems (IVIS). Our work is motivated by the proliferation of speech interfaces in our everyday life, which have already found their way into consumer electronics and will most likely become pervasive in future cars.

Our design space is based on expert interviews as well as a comprehensive literature review. We present five core dimensions – assistant, position, dialog design, system capabilities, and driver state – and show in an initial study how these dimensions affect the design of a prototypical IVIS.

Design spaces have paved the way for much of the work done in HCI including but not limited to areas such as input and pointing devices, smart phones, displays, and automotive UIs. In a similar way, we expect our design space to aid practitioners in designing future IVIS but also researchers as they explore this young area of research.

Author Keywords

Automotive User Interfaces; Speech Interaction; Natural Language Interfaces; Design Space.

ACM Classification Keywords

H.5.2 [Information interfaces and presentation]: User Interfaces

Introduction

Speech interfaces have found their way into our everyday life. We ask our smart phones to call our friends, we talk to our laptop computer to enter search queries, and we tell our TV which movie to load from the streaming platform of our choice. Such interfaces are in most cases limited to rather simple, user-initiated requests and do not exploit the opportunities of conversations, where dialogs can take more complex forms. We see particular potential for such interfaces in cars, where it is important to minimize driver distraction which usually occurs due to the need for directing visual attention towards an in-car display. With voice, the user could simply request information such as the weather forecast, the estimated time of arrival, new messages, or upcoming appointments. Such a system could also initiate a dialog to keep drivers alert in situations where they get bored due to a monotonous drive [6].

Designing speech interfaces is difficult. Sometimes it is not easy to make a specific request in such a way that it can be easily understood by the machine. Often users may need to interrupt the current conversation due to a situation on the road that requires their attention. In this case it may be difficult to remember where the dialog was interrupted and users may require the be supported of additional visualizations. Additionally, it is not clear what effect the use of such a system has on driving performance.

To understand possible impacts on how such a conversational IVIS can be used in a safe and efficient manner, we chart a design space based on a thorough literature review and expert interviews. We envision this design space to contribute to the field of HCI (1) by fostering research in the area since it points at interesting challenges and opportunities and (2) by establishing common ground for designers of IVIS regarding important decisions in the design process.

Our work is complemented by a description of how the design space can be used to create a conversational IVIS. We report on the design and implementation of the system and present the results from an initial user study.

Related Work

When designing automotive user interfaces, standards and guidelines provide valuable input on how to structure these interfaces [17] and minimize distraction [7, 8]. At the same time, design spaces provide a foundation for ideation on new combinations of modalities and devices and thereby enable possible system improvement not through regulation, but through innovation.

A design space represents an abstraction of existing and plausible point designs which can be used to create new systems through rearrangement of components [3]. First design spaces by Buxton and Card et al. focused on the classification of input devices for HCI [2, 4] and with emerging technologies, design spaces for various upcoming fields have been presented and refined over the years. For example, Card and Mackinlay published a design space for information visualization based on first point designs [3] and Chi later improved on the taxonomy [5]. Design spaces for public displays [14], input on hand-held devices [1] and general multimodal interaction [15] have equally attracted attention in their respective communities as publications in the realm of automotive UIs, which we would like to contribute to.

Notable publications in the automotive context include Kern and Schmidt's design space for driver-based automotive user interfaces, which describes fundamental input and output modalities, positioning and graphical representations for in-vehicle UIs [12]. Rümelin et al. present a classification of interaction areas for drivers and passengers and comment on the potential of collaboration to reduce workload [20].

Figure 1: Assistant representation of MIT SENSEable City Lab Affective Intelligent Driving Agent (AIDA) [13]

Figure 2: Status visualizations used for Microsoft Cortana – used with permission from Microsoft

Recently, Riener et al. focussed on the compatibility of multiple in-car systems related to gesture input [19]. They show an interaction space which helps manufacturers define generally valid gesture sets without affecting other domains. Thus, they provide input for possible future norms on consistent gesture classification. Another novel design space, focusing on automotive windshield displays, is presented by Häuslschmid et al. [10]. They analyzed existing literature and reviewed patents to summarize what has been done already and give an outlook on future opportunities.

Methodology

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 (e.g., [9, 11, 21]). 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 (e.g., BMW Voice Control System) and concept studies of conversational IVIS (e.g., Nissan Pivo, Audi AIDA) 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 iterated upon 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. The resulting 5 categories, discussed in the following section, cover most of today's design issues of naturalistic conversational user interfaces, with a focus on automotive applications.

Design Space

This design space is the result of a search for expressive variables for future conversational interaction systems. It reflects the current landscape of possibilities and should be extended continuously as innovations relevant to the field emerge. The following passages describe the design space and its categories.

Assistant

In a conversational IVIS, users talk to a digital assistant which adds a further layer between user and machine. The assistant acts as a conversational partner which can mediate between driver and car in natural language. Its speech behavior plays a major role in how it is perceived and should be modelled intently. Like human input, assistants can vary in speech granularity and personality, and they can take different genders in their speech synthesis settings. These attributes affect the assistant's personality, which could become a distinguishing feature for future markets.

Independent of the assistant's personality, designers have to think about how to convey the system's status to the user. This can be achieved through personification or abstraction, through graphics, lights and many more. Even omitting the display is an option. It could make sense to either express system states (e.g., listening, processing, talking) in the representation or to translate the assistants emotional state into a graphical interpretation as Microsoft did with Cortana (see Figure 2).

Position

Currently, IVIS are mostly displayed on screens in the central dashboard (central information display, CID) or on head-up displays (HUD) [20] and sometimes on the instrument cluster (IC) behind the steering wheel. Other positions like on the steering wheel, in the windshield or the mirrors, or on

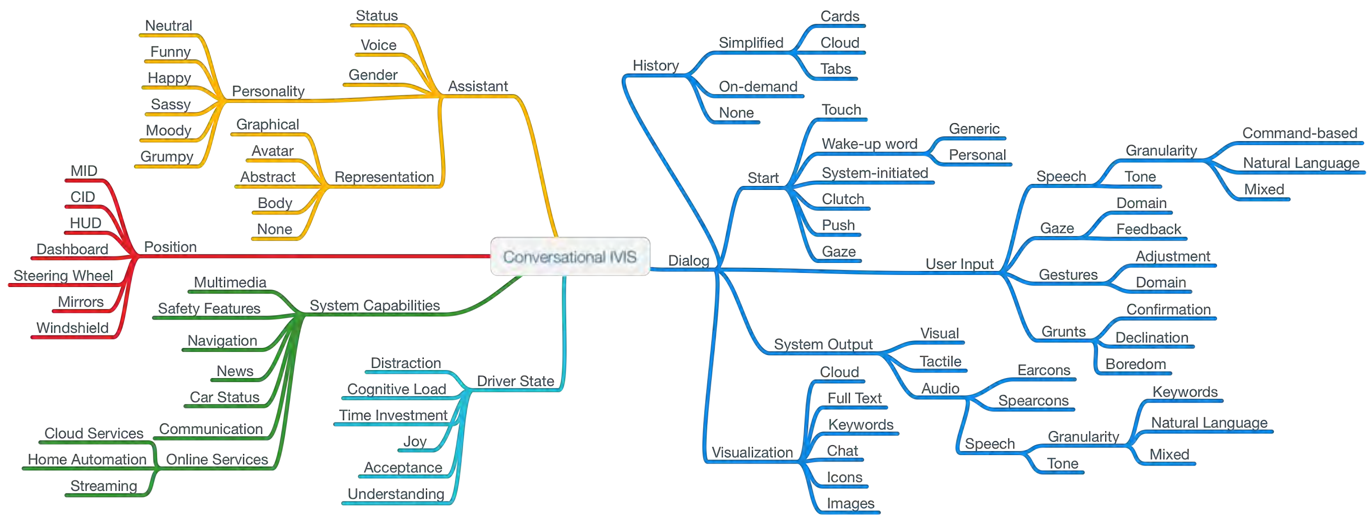

Figure 3: Graphical representation of the design space for conversational in-vehicle information systems

Figure 4: Plausible display positions for a conversational IVIS

top of the dashboard are potential locations designers can choose from. By assigning the assistant to a dedicated spot it might even be possible to establish a mental model where it is separated from the rest of the car. Figure 4 shows possible positions for a standalone assistant in the cockpit.

Dialog Design

The first step for a dialog is an activating action, such as establishing eye contact between humans or saying a wake-up word to your smart phone. There are many options for initiating a chat with a machine, such as push-to-talk, talk-to-talk, look-to-talk, and gesture-to-talk [16]. Variations can also occur in the choice of placement. Push-to-talk can be, for example, assigned to multiple touch sensitive areas with different domains mapped to them for quick access or it can be incorporated into an unused space, like the far left foot area in cars with automatic transmission [12]. Ultimately, with intelligent systems the initiative to start a dialog might come from the assistant itself. The system could remark on points of interests – if the situation permits – or propose a break when it senses clues for fatigue.

Next, some form of input is needed to tell the system about the users' intentions. Speech is a natural input modality for humans, especially in modern times where computers can understand complex sentences. Although natural language might be preferred in many scenarios, it makes sense to keep the opportunity for different speech granularities, e.g., keywords, in mind. Natural language also comes in different tonalities which a system might want to pay attention to in conversations to detect the user's intentions. Additionally, we often communicate in non-lexical sounds, or grunts, which can be used to identify driver states such as emotions or moods, and affirmation or declination [22].

Figure 5: Dialog concepts with status visualization and different text variants embedded in an IVIS

Figure 6: Exemplary system capabilities for a modern car: (1) communication with IoT devices, (2) smart home controls, (3) environment perception through sensors

Other input modalities such as gestures and gaze control can enhance the speech dialog. Be it hand movements, facial expressions, or general body movements, they can be utilized as input to select domains, to control settings, or as feedback channels for the system. For example, the 2017 BMW 7series incorporates hand gestures to switch through songs or control the volume. These modalities can be combined with speech input in conversational interfaces to create a more natural experience by, for example, confirming commands with a frugal nod or choosing which mirror to adjust by simply looking at it.

The dialog visualization is another important topic to address, especially in the automotive setting, as drivers may not be able to unrestrictedly follow the conversation at all times. Therefore we need to think about a sensible way of displaying information without it becoming too much of a distraction to the driver. In an accompanying study we explored different ways to format and visualize the dialog: a chat metaphor found in many smart phone apps, a condensed view with the text summarized in keywords, and a combination of keywords and additional icons representing the domain in focus (cf. Figure 5). We can also support the user by displaying a history, either on-demand or in some kind of simplified form like tabs or a tag cloud. Beside the visualization of the spoken input and output, graphics (e.g., a map), charts and diagrams can convey information that is hard to express solely by speech.

System Capabilities
System capabilities will influence the way the interface is perceived by its users and therefore need to be included in the design process, even if they cannot be altered in certain settings. They are the foundation for use cases and their integration can have substantial influence on the success of the system. Possible capabilities include multimedia, safety and navigation features, online services, car status visualizations, smart home connectivity, and all kinds of other extensions to current systems' range of function. We claim that the fidelity of the speech output as well as the level of the assistant's personality should correlate with the system capabilities. For example, if the conversational IVIS is characterized by a human-like speech output and a multifaceted personality, the user would expect to access various functions through the speech UI and not only a specific subset of features.

Driver State
Understanding and controlling the driver state is important to reduce driver distraction and mental workload, and to design an appropriate user experience [6]. Besides evaluating how different interfaces affect these aspects, the driver state can also be a separate dimension of the design space. The conversational IVIS can be adapted according to the current driver state. As outlined by Coughlin et al. (based on the Yerkes-Dodson law), the driver's performance is ideal at an intermediate level of workload and decreases both with fatigue or inattention and active distraction or overload [6]. Thus, the goal is to keep the driver in this ideal range. When driving on an empty and tiring highway, a system-initiated dialog might help to keep the driver alert and, thus, maintain ideal driving performance.

Application of the Design Space
To showcase the potential of the presented design space, we applied it in a user study on GUI concepts for conversational IVIS. We developed several concepts based on the dimensions *assistant* and *dialog design* in order to investigate user preferences for the display of conversation contents while driving.

	Input Viz	Output Viz	Speech Granularity	Assistant Repres.
1	Full Text	Full Text	Natural	Abstract
2	Full Text	Keywords	Natural	None
3	Keywords	Full Text	Natural	Abstract
4	Keywords	Keywords	Comm.	None
5	None	Full Text	Natural	Abstract
6	None	Keywords	Natural	None
7	None	Key+Icons	Comm.	Abstract

Figure 7: Variants of the prototype tested in the experiment

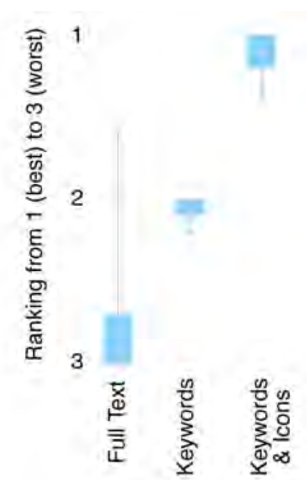

Figure 8: Ranking on system output visualizations

Apparatus
The prototype consists of an application that can process natural language input and respond through synthesized speech. For the evaluation, we created 7 distinct variants of the prototype which differed in 3 sub-dimensions of the design space, namely dialog visualization (full text, keywords, keywords & icons, see Figure 5), assistant representation (none, abstract), and granularity of the speech output (keywords, natural language) as shown in Figure 7.

Study Design
We used a within subject design and collected subjective feedback from ten participants aged 29.2 $\pm$ 8.7 years. The group included only 1 woman and all of the subjects had a background in the automotive industry. As the primary assignment, the participants performed the Critical Tracking Task simulating a measure of workload [18]. 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.

Setup and Procedure
Participants experienced the different visualizations in permuted order. The conversation agenda included 4 use cases including turn-taking between the user and the 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 away the gas station was, which was also answered by the system. Once the participants experienced all of the concepts, they ranked them according to personal preference.

Results
Users rated the variants with fewer text as less distracting. However, displaying text was generally desired as it simplified error detection. The dialog visualization using keywords with accompanying icons was ranked highest, followed by keywords alone, while full text display was rather disliked ($\chi^2 = 11.4$, $df = 2$, $p < .003$, see Figure 8). All participants opted for natural language speech synthesis compared to a shortened audio mode ($Z = -2.8$, $p = .005$). Seven out of 10 users stated they prefered an abstract representation of an assistant, after no representation at all.

Conclusion
In this paper, we propose a design space for conversational user interfaces in cars. We derived the design space from analyzing conversational UIs of modern consumer electronics devices and cars. It consists of 5 major dimensions: the *design of the assistant*, which constitutes the dialog partner for the user, its *position*, the *dialog design*, and the *system capabilities* have to be chosen with the *driver's state* in mind. We claim that this design space helps automotive UI designers to come up with novel conversational UI concepts in cars and opens exciting questions for research. E.g., how people talk to a graphical or physical assistant compared to an assistant without any representation.

In the future, we plan to use this design space to investigate, among other things, dialog visualizations in different driving contexts (e.g., bored or stressed drivers) building upon the initial study presented above. As natural language interfaces are an emerging field, we hope to encourage discussions in the HCI community and expect this design space to be heavily used and extended as new technologies emerge.

REFERENCES

1. Rafael Ballagas, Michael Rohs, Jennifer G. Sheridan, and Jan Borchers. 2008. The Design Space of Ubiquitous Mobile Input. *Handbook of Research on User Interface Design and Evaluation for Mobile Technology* I, 11 (2008), 1172. DOI: http://dx.doi.org/10.4018/978-1-59904-871-0

2. William Buxton. 1983. Lexical and Pragmatic Considerations of Input Structures. *SIGGRAPH Computer Graphics* 17, 1 (1 1983), 31–37. DOI: http://dx.doi.org/10.1145/988584.988586

3. Stuart K. Card and Jock Mackinlay. 1997. The structure of the information visualization design space. In *Proceedings of the 1997 IEEE Symposium on Information Visualization (INFVIS '97)*. IEEE, 92–99. DOI: http://dx.doi.org/10.1109/INFVIS.1997.636792

4. Stuart K. Card, Jock D. Mackinlay, and George G. Robertson. 1991. A Morphological Analysis of the Design Space of Input Devices. *ACM Trans. Inf. Syst.* 9, 2 (April 1991), 99–122. DOI: http://dx.doi.org/10.1145/123078.128726

5. Ed H. Chi. 2000. A taxonomy of visualization techniques using the data state reference model. In *Proceedings of the IEEE Symposium on Information Visualization 2000 (INFOVIS 2000)*. IEEE, 69–75. DOI: http://dx.doi.org/10.1109/INFVIS.2000.885092

6. Joseph F. Coughlin, Bryan Reimer, and Bruce Mehler. 2011. Monitoring, Managing, and Motivating Driver Safety and Well-Being. *IEEE Pervasive Computing* 10, 3 (7 2011), 14–21. DOI: http://dx.doi.org/10.1109/MPRV.2011.54

7. Department of Transportation and National Highway Traffic Safety Administration. 2013. Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Device. *Federal Register* 78, 81 (2013), 24817–24890. https://federalregister.gov/a/2013-09883

8. Paul A. Green. 2008. Driver Interface/HMI Standards to Minimize Driver Distraction/Overload.. In *Convergence 2008 Conference Proceedings*. Society of Automotive Engineers, SAE International, Warrendale, Pennsylvania, USA, Article 2008-21-0002, 19 pages. http://papers.sae.org/2008-21-0002/

9. Linn Hackenberg, Sara Bongartz, Christian Härtle, Paul Leiber, Thorb Baumgarten, and Jo Ann Sison. 2013. 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)*. ACM, New York, NY, USA, 114–120. DOI: http://dx.doi.org/10.1145/2516540.2516553

10. Renate Haeuslschmid, Bastian Pfleging, and Florian Alt. 2016. 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 (CHI '16)*. ACM, New York, NY, USA, 5076–5091. DOI: http://dx.doi.org/10.1145/2858036.2858336

11. Grega Jakus, Christina Dicke, and Jaka Sodnik. 2015. A user study of auditory, head-up and multi-modal displays in vehicles. *App. Erg.* 46, Part A (2015), 184–192. DOI: http://dx.doi.org/10.1016/j.apergo.2014.08.008

12. Dagmar Kern and Albrecht Schmidt. 2009. 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)*. ACM, New York, NY, USA, 3–10. DOI: http://dx.doi.org/10.1145/1620509.1620511

13. MIT SENSEable City Lab. 2009. AIDA - Affective Intelligent Driving Agent. (2009). http://senseable.mit.edu/aida

14. Jörg Müller, Florian Alt, Daniel Michelis, and Albrecht Schmidt. 2010. Requirements and Design Space for Interactive Public Displays. In *Proceedings of the 18th ACM International Conference on Multimedia (MM '10)*. ACM, New York, NY, USA, 1285–1294. DOI: http://dx.doi.org/10.1145/1873951.1874203

15. Laurence Nigay and Joëlle Coutaz. 1993. A Design Space for Multimodal Systems: Concurrent Processing and Data Fusion. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems (CHI '93)*. ACM, New York, NY, USA, 172–178. DOI: http://dx.doi.org/10.1145/169059.169143

16. Alice Oh, Harold Fox, Max Van Kleek, Aaron Adler, Krzysztof Gajos, Louis-Philippe Morency, and Trevor Darrell. 2002. 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)*. ACM, New York, NY, USA, 650–651. DOI: http://dx.doi.org/10.1145/506443.506528

17. ITU-T Focus Group on Driver Distraction. 2013. *Report on User Interface Requirements for Automotive Applications*. Technical Report. International Telecommunication Union. https://www.itu.int/en/ITU-T/focusgroups/distraction

18. Tibor Petzoldt, Hanna Bellem, and Josef F. Krems. 2014. The Critical Tracking Task: A Potentially Useful Method to Assess Driver Distraction? *Human Factors* 56, 4 (2014), 789–808. DOI: http://dx.doi.org/10.1177/0018720813501864

19. A. Riener, A. Ferscha, F. Bachmair, P. Hagmüller, A. Lemme, D. Muttenthaler, D. Pühringer, H. Rogner, A. Tappe, and F. Weger. 2013. Standardization of the In-car Gesture Interaction Space. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '13)*. ACM, New York, NY, USA, 14–21. DOI:http://dx.doi.org/10.1145/2516540.2516544

20. Sonja Rümelin, Peter Siegl, and Andreas Butz. 2013. 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 '13 Adjunct)*. 61–64. http://auto-ui.org/13/docs/aui_adjunct_proceedings_final.pdf#63

21. David L Strayer, Joel M Cooper, Jonna Turrill, James R. Coleman, and Rachel J. Hopman. 2015. *The Smartphone and the Driver's Cognitive Workload: A Comparison of Apple, Google, and Microsoft's Intelligent Personal Assistants*. Technical Report October. AAA Found. for Traf. Saf. https://www.aaafoundation.org/sites/default/files/strayerIIIa_FINALREPORT.pdf

22. Nigel Ward. 2006. Non-lexical conversational sounds in American English. *Prag. & Cog.* 14, 1 (2006), 129–182. DOI:http://dx.doi.org/10.1075/pc.14.1.08war