# SiAM – Situation-Adaptive Multimodal Interaction for Innovative Mobility Concepts of the Future

Monika Mitrevska*, Mohammad Mehdi Moniri*, Robert Neßelrath*,
Tim Schwartz*, Michael Feld*, Yannick Körber*, Matthieu Deru* and Christian Müller*
*German Research Center for Artificial Intelligence
Saarbrücken, Germany
Email: monika.mitrevska@dfki.de

***Abstract*—What does situation-adaptive technology mean for car drivers and how can it improve their lives? Why is multimodal interaction in the cockpit a critical ingredient? This contribution summarizes several important technological results of the three-year research project SiAM, which investigated these questions. Motivated by the story of an urban commuter, we illustrate three use cases for situation adaptivity: multimodal control of car functions, cognitive load aware interaction with the environment, and a persuasive intermodal travel assistant.**

## I. Introduction

Cars today contain large array of sensors. Altogether, they are potentially able to provide a very complete picture of the current driving situation. The upcoming car2x technology provides a source for the car to obtain additional context information. Being always connected, cars also provide access to external services for an even greater level of knowledge. However, much of the potential of these developments is not used today. To reach a new degree of driver (and passenger) assistance, we need to combine all those sources and allow the car to utilize this knowledge to enhance comfort and safety. The result is a vehicle that is aware of its current situation. Situation awareness in this sense means interpreting incoming information from all sources, understanding the implications for the car and passengers and reacting accordingly. It is a term from psychology; when we consider the technological realization for the in-car platform, we refer to this concept as situation adaptivity: the ability of a system to behave dynamically based on who is using the system and in which context.

In SiAM, a three-years research project, we have investigated several ways of achieving situation adaptivity. Our approach combines knowledge from various sources from inside and outside the car to proactively take action. When the user needs to be involved, the car attempts to provide the optimal type of interaction method in the current situation in order to minimize the distraction. Since interaction with the user - either system or user initiated - plays a crucial role, a dialogue platform was created that covers all interaction aspects. This platform provides access to new modalities that can also be combined via multimodal fusion.

In this paper, we describe a scenario that illustrates many facets of situation adaptivity resulting into SiAM. It features three working applications which have been demonstrated live on several occasions. In the following text and the accompanying video, we present a prototypical user who is using his car on a daily basis. Picking out two days, we show a few cases where situation adaptivity can make a difference, and illustrate the technology behind.

## II. Scenario

Imagine commuter Max on his daily way to work with his own car. Organizing a dinner for his friends the next day, he passes by an inviting restaurant that has recently opened. He wants to know what the restaurant is offering and how the prices are. Since his car is equipped with a context-aware multimodal dialog system, he can get the desired information simply by asking "what is the menu for today?". The system uses its context knowledge, which is derived from Max' focus of attention (a combination of eye-gaze and head-pose), current position and an environment model, to automatically conclude that Max is talking about the restaurant he was looking at earlier. Because the system keeps track of the dialog discourse, Max can inquire additional information in a natural dialog fashion, and finally reserve a table.

While driving to work and back, Max is able to switch from traditional buttons and switches in the cockpit to multimodal operation of basic car functions, if the situation demands it. Windows, sunroof, mirrors and even turning lights can be controlled through a set of speech commands, micro-gestures and head orientation, which can also be combined almost arbitrarily. For example, Max can open the window by saying "open this window" while looking slightly to the front right to open the passenger window.

Today, there is a long traffic jam on his way to work. The system recognizes that he would be late because of this. Because Max has agreed to try out the Persuasive Intermodal Trip Assistant that aims to improve traffic conditions in the city, his car takes the opportunity to suggest taking a bus for the remaining part of the travel after checking the live schedule and verifying that bus stops with parking nearby are available. Because of priority lanes, the bus will arrive in time and be comparable in cost.

A few days later, Max finds himself in a similar situation. The system again checks the available options, but this time proposes to park the car and take an e-bike to the destination. Because of the context – good weather conditions and a nearby e-bike sharing station with free bikes – the bike is preferred over the bus this time, as it is even cheaper and more ecological. In order to help him with the decision, additional information about the money and time saved is presented. Max

978-1-4673-6654-0/15 $31.00 © 2015 IEEE
DOI 10.1109/IE.2015.39

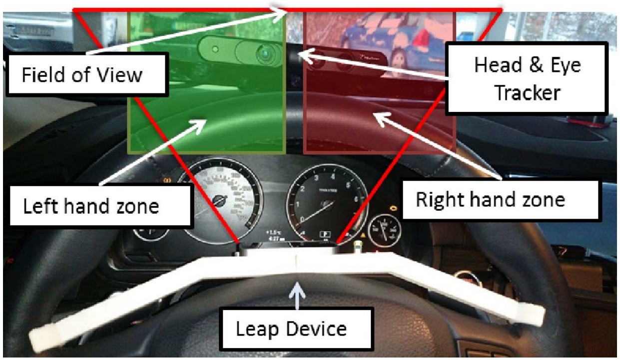


Fig. 1. Input devices in the car

agrees to the suggestion and is guided to the bike sharing station, while the system reserves the bike. After leaving the car, the support continues on Max's private device, so he can locate and unlock the bike using his smart phone and smart glasses.

## III. Multimodal Control of Car Functions

In this part, we illustrate the support of multimodal interaction in the SiAM dialogue platform [1]. It comprises the control of several in-car actuators like the windows, outside mirrors, or the turning lights. The main idea of the adaptive approach is that the various input modalities can synergistically or alternatively be employed dependent on the actual situation or the preferences of the driver.

### A. In-Car Setup

We integrated three input devices and an interface to the car's CAN (Controller Area Network) bus into the dialogue application (see Figure 1):

*EyeVIUS* is a research prototype that provides third-party applications with the possibility to integrate information about the driver's attention in their existing platforms [2]. It consists of different modules to monitor the driver's eye-gaze, head-pose, and mimic. This information can then be provided to other applications for interaction or analysis purposes.

*Micro Gestures* are directly performed with the hands on the steering wheel. Finger movements are recognized by a Leap Motion Controller (https://www.leapmotion.com), a device for tracking hand motion using a generated 3D pattern of infrared (IR) dots and an IR camera. The Leap device is placed at the center of the steering wheel, for which we have designed a special holder produced with a 3D printer.

*Speech Input* is supported by accessing the off-the-shelf Microsoft Speech API (SAPI). The SAPI provides a grammar-based speech recognition interface.

*CAN* is a vehicle bus standard that allows micro controllers and devices to communicate without a host computer. The bus is widespread in modern cars and internally is used for the control of car actuators. We access this interface for the control of the in-car functions.

### B. Uni/Multimodal combinations

Our interaction strategy distinguishes between two types of contribution that an input can provide. First is the *context* or the concrete *actuator instance* the driver intends to control, e.g. the front right window or the left turning light. Second is the *function* the driver wants to execute, e.g. opening (a window) or folding (the outside mirrors). Both types of input can be provided by a single input device, but also as a multimodal combination of two devices. We support the following combinations:

**Speech Only:** Both context and function can be communicated in a single utterance from the driver. Examples of possible utterances are *"Open the right windows"* or *"Fold the outside mirrors"*. Furthermore, content can be subsequently contributed. E.g. the first utterance *"the front right window"* introduces the actuator instance, and the following contribution *"open it"*, which contains an anaphoric reference to the previous utterance, specifies the function.

**Gestures Only:** We developed an interaction concept that associates each type of contribution with one hand. The left hand specifies the context. One stretched finger indicates the turning lights, two fingers the front windows, and three fingers the outside mirrors. The right hand specifies the function. Here, two fingers of the hand are stretched to a plain level zero position. Relative to this plain level we also define an upper and a lower position. The concrete function is dependent on the context and summarized in Table I.

TABLE I. Summary of Finger Gestures

| Context: \ Finger Level: | Lower | Zero | Upper |
|---|---|---|---|
| Turning Light | Flash Right | Deactivate | Flash Left |
| Front Windows | Open | Stop | Close |
| Outside Mirrors | Fold | - | Unfold |

**Speech & Gesture:** In this combination, each type of content is provided by a different modality. The dialogue platform then fuses these inputs. One option is that the actuator is introduced by a left hand micro gesture and the function is given by speech. The other one is to introduce the actuator by speech and to specify the function by right hand motion.

**Speech & Focus-of-Attention:** Since the information about the driver's focus is comparable to the information from a deictic gesture, it is more suitable for introducing actuators than functions. It is possible that more than one actuator is located on a line of sight, e.g. if the driver looks through a window at the right outside mirror. In this case, the semantics of a spoken utterance like *"fold"* help to resolve the ambiguity since it is possible to fold an outside mirror, but not a window.

## IV. Cognitive Load aware Interaction with the Outside Environment

In this part of the scenario, we take the step from in-car interaction to interaction with the outer world, which requires a slightly different approach for reference resolution. We combine multimodal interaction involving several dialogue phenomena with another aspect of situation adaptivity, namely the prediction of cognitive load inflicted by user interaction.

Our environment reference approach is an extension of [3]. The outside environment was scanned with a professional 3D laser scanner. For this purpose we scanned more than three hectares of the university campus with centimeter accuracy. The resulting point cloud was then used as basis for a 2.5D polygon model. We used this 2.5D model together with a GPS map-matching algorithm to position the vehicle in the environment in real-time. In addition, this model also contained different buildings as well as other smaller objects, such as bus stops and traffic signs. The information from EyeVIUS about the driver's attention was then used together with this 2.5D model and the vehicle's position to identify which object in the environment is in the focus of the driver.

On the back-end side we interact with Internet services instead of local actuators inside the car. In the demonstrator, we address a service that returns information about a restaurant's menu and offers the possibility to reserve a table in this restaurant. However, this service is representative for an arbitrary Internet service and can easily be replaced.

The second focus lies on the demonstration of the dialogue platform's capability to handle certain dialogue phenomena:

- *Deictic References:* An expression that points to an entity, which is introduced by a pointing modality, e.g. the restaurant in the scenario that is introduced by the EyeVIUS system.
- *Temporal References:* A relative temporal reference applies to a temporal frame. In the scenario, the temporal reference *"today"* and *"tomorrow"* are relative to the current time and must be resolved by the dialogue application.
- *References on Discourse Context:* The user references to entities that have previously been introduced by responses of the system.
- *Ellipsis:* In an ellipsis, the utterance omits one or more words but nevertheless makes sense in the current discourse context. In the scenario the stand-alone utterance "And tomorrow?" is hard to interpret, but a retrospection on the previous question *"What is the menu for today?"* gives it a meaning.

Finally, this part showcases a new cognitive load model for drivers, developed in SiAM. This model was originally introduced in [4] and is based on [5]. Unlike other approaches, which measure the effects of cognitive load (e.g. stress), its focus is on predicting the effects of the user's dialogue interaction. In short, the model attempts to split user input and system output into primitive interaction tasks, which can be button presses, scrolling through a list, or entering a number. There is a fixed list of such primitive tasks, which are based on the way users typically interact with them. Most complex interactions can be broken down into these primitives. For each task and modality, a cognitive demand was determined through experimentation. The actual demand may depend also on the type of widget and presentation style (e.g. length of speech utterance). Following the Wickens model of multiple resource theory [6], the demand is assigned to different processing channels (visual, auditive).

In order to develop a better understanding of these effects, we have created a design tool that visualizes the estimated cognitive load according to this model. In our scenario, dialog turn is visualized with its demand and summed up to the overall cognitive load. This total may also include external influences such as traffic, which are visualized separately.

## V. Persuasive Intermodal Travel Assistant

Persuasive technology is employed in the car to help drivers optimize their driving behavior: by driving safer, more ecologically, saving money on fuel, or reducing travel time. A persuasive system focuses on the long-term effects of a change in driver behavior, while exploiting situation adaptivity to give the right suggestions at the time when they apply.

Going towards the concept of mobility as a service, we developed a persuasive intermodal travel assistant that raises the driver's awareness of the transportation possibilities outside his own car. It also pursues the goal of reducing car traffic in the city center. Learning from driver's past behaviour in correlation with the driver's agenda, current position, weather and traffic conditions, the travel assistant makes context-aware suggestions for alternative trips e.g. in case of traffic jams on the route or a lack of parking places around the destination. It finds alternative solutions by combining the trip with public transport or car/bike sharing services available along the selected driving route. Real-time information on public transport or bike sharing services are acquired through external web services. Part of the context, e.g. weather conditions and parking availability, is also provided by external web services.

For this part of the demonstration, we have selected two use-cases to show the functionality that is provided by the persuasive component. In the first use-case, a traffic jam situation is resolved by proposing the user to take the next bus to his destination. Correlation of bus line and schedule information with underlying route attributes (in this case a bus priority lane on the affected route) and weather conditions results in a context-aware decision that the bus is the best travel alternative for the driver.

Sometimes drivers do not follow a certain supposedly better behavior because they have objections or good reason not to. But in many cases, they simply lack the background information to understand the advantage of the alternatives. Therefore, a persuasive system should provide valuable information to assist the driver's decision. This strategy can sometimes conflict with the goal of avoiding distraction while driving. Showing a lot of background information in the car, especially on a small screen, may have adverse effects and possibly even reduce the effectiveness of the advice. Therefore, the SiAM Persuasive Intermodal Trip Assistant combines the following ideas: (1) It attempts to use the in-car screen space efficiently by providing the most valuable information; (2) Information is presented both visually and through speech output, thereby allowing the user to choose the most suitable modality depending on the situation; (3) In addition, we include an external roadside display (electronic billboard) for further complementary information.

In our scenario, we display additional information about the parking availability around the selected bus stop and the ticket price on the main in-car screen. As a result, the driver receives a suggestion to take a specific bus from the closest bus stop, gets information about the arrival time and ticket price, information about the saved travel time and parking options.

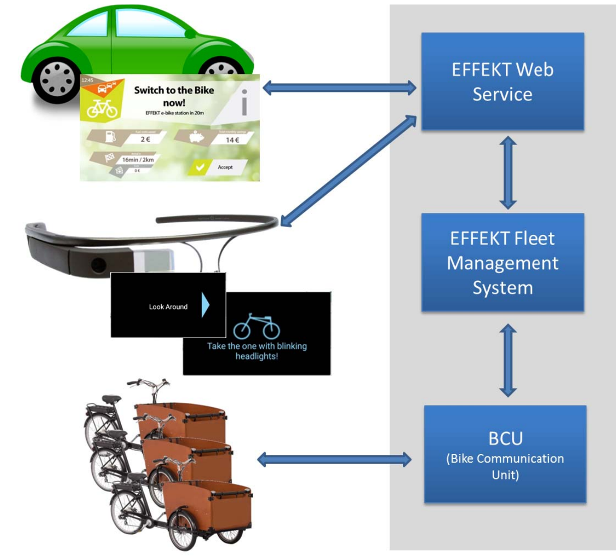


Fig. 2. EFFEKT – Architecture of the migrational bicycle service

On the outside billboard, we show real-time bus schedule information so that the driver could further inform himself about alternative options.

In the second use-case, we show how a traffic jam situation is resolved by proposing to use a bike-sharing service. In this case, the availability of a bike sharing service, weather conditions and terrain information result in a suggestion to the driver to continue the trip with a bicycle. Information on the saved money on fuel and the arrival time are used as persuasion methods. Real time bicycle availability is acquired by our bike sharing platform EFFEKT, which grants a real-time overview on the availability of its sensor-equipped bicycles via a bike sharing web service. Through a restful API, applications can integrate the following bike sharing functions: (1) check for available bicycles around the current position of the user or in a given area; (2) reserve a bicycle; (3) unlock a bicycle; (4) check out a centralized reservation number from several applications/devices (in our case the in-vehicle system and smart glasses); (5) return a bicycle; (6) get the current state of the bicycle sensors.

From the user's perspective, different travel modes can also mean leaving the comfort of their own vehicles and exposing themselves to new, probably inconvenient solutions instead. We argue that in order to persuade people to consider different travelling modes, the services offered during the whole trip have to be comparable with the comfort people have in their own vehicles. Therefore, EFFEKT together with SiAM develops a migrational bicycle sharing service, so that drivers can continue receiving travel support on their smart phone or a wearable device once they leave the car.

EFFEKT's main components and the integration with external systems is shown in Figure 2. The in-vehicle system and the smart glasses both connect to the EFFEKT web service using the provided APIs. A central fleet management system brings together the front-end applications and the bicycles. EFFEKT's bicycles have an on-board communication unit that allows data collection from sensors as well as direct communication between bicycles and user devices. In this demonstration, the service is used by the in-vehicle system to reserve a bicycle for the driver. The driver later uses his Google Glass to locate the the reserved bicycle. Blinking front lights indicating the reserved bicycle represents one use-case for the bicycle communication unit.

## VI. Conclusion

We described a system for smart vehicles that incorporates a multimodal dialog system in combination with different modules for detecting micro gestures and assessing the driver's focus of attention and cognitive load. The first two modules provide the opportunity to refer to objects outside the car or control in-car functions in a safe manner while driving. Furthermore, the system includes a persuasion module which is connected to a back-end infrastructure that offers the driver environment friendly transportation modes by suggesting alternative options, such as riding an e-bike or taking a bus, in suitable traffic and weather conditions.

## Acknowledgment

This work was funded by the German Ministry of Education and Research (project SiAM: Situation-Adaptive Multimodal Interaction for Innovative Mobility Concepts for the Future; grant number 01IW11004 in collaboration with the Software Campus project - EFFEKT).

## References

[1] R. Neßelrath and M. Feld, "SiAM-dp: A platform for the model-based development of context-aware multimodal dialogue applications," in *IE'14: Proceedings of the 10th International Conference on Intelligent Environments*. Shanghai, China: IEEE, July 2014.

[2] M. M. Moniri and C. Müller, "Eyevius: Intelligent vehicles in intelligent urban spaces," in *Adjunct Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. New York, NY, USA: ACM, 2014, pp. 1–6.

[3] M. M. Moniri, M. Feld, and C. Müller, "Personalized in-vehicle information systems: Building an application infrastructure for smart cars in smart spaces," in *Proceedings of the 8th International Conference on Intelligent Environments IE'12*. Guanajuato, Mexico: IEEE, 27-28 June 2012, best Video/Demo Award.

[4] R. Neßelrath and M. Feld, "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*, 2013.

[5] Christoph Endres, "PresTK: Situation-Aware Presentation of Messages and Infotainment Content for Drivers," Ph.D. dissertation, Saarland University, 2012. [Online]. Available: http://www.dfki.de/˜endres/CEndres_36.pdf

[6] C. D. Wickens, "Multiple resources and performance prediction," *Theoretical issues in ergonomics science*, vol. 3, no. 2, pp. 159–177, 2002.