SpeeT: A Multimodal Interaction Style Combining Speech and Touch Interaction in Automotive Environments

Bastian Pfleging*, Michael Kienast#, Albrecht Schmidt*
VIS, University of Stuttgart
Pfaffenwaldring 5a
70569 Stuttgart, Germany
+49-711-685-60069 / -60048
*{firstname.lastname}@vis.uni-stuttgart.de
#{firstname.lastname}@studi.informatik.uni-stuttgart.de

Tanja Döring
Paluno, University of Duisburg-Essen
Schützenbahn 70
45127 Essen, Germany
+49-201-183-2955
tanja.doering@uni-due.de

ABSTRACT
SpeeT is an interactive system implementing an approach for combining touch gestures with speech in automotive environments, exploiting the specific advantages of each modality. The main component of the implemented prototype is a speech-enabled, multi-touch steering wheel. A microphone recognizes speech commands while a wheel-integrated tablet allows touch gestures to be recognized. Using this steering wheel, the driver can control objects of the simulated car environment (e.g., windows, cruise control). The idea is to use the benefits of both interaction styles and to overcome the problems of each single interaction style. While touch input is well suited for controlling functions, speech is powerful to select specific objects from a large pool of items. The combination simplifies the problem of remembering possible speech commands by two means: (1) speech is used to specify objects or functionalities and (2) in smart environments - particularly in cars - interaction objects are visible to the user and do not need to be remembered. Our approach is specifically designed to support important rules in UI design, namely: provide feedback, support easy reversal of action, reduce memory load, and make opportunities for action visible.

Categories and Subject Descriptors
H5.2 [Information interfaces and presentation (e.g., HCI)]: User Interfaces – Interaction Styles (e.g., commands, menus, forms, direct manipulation).

General Terms
Human Factors.

Keywords
Gesture, speech, multimodal interaction, automotive user interfaces, smart environments.

1. INTRODUCTION
Multimodal technologies offer a great potential to reduce shortcomings of single modalities for interaction. Although quite some research on multimodality has been conducted and some general guidelines have been shaped (e.g., [3]), no specific patterns or interaction styles for an appropriate integration of different modalities have emerged yet. SpeeT (Speech+Touch) is implemented to evaluate the concept of combining speech and touch gesture input for interaction with real objects. The concept was designed based on design rules and usability heuristics, centered around a formative study to gather user-elicited speech and gesture commands.

AutomotiveUI'11, November 29-December 2, 2011, Salzburg, Austria
Adjunct Proceedings

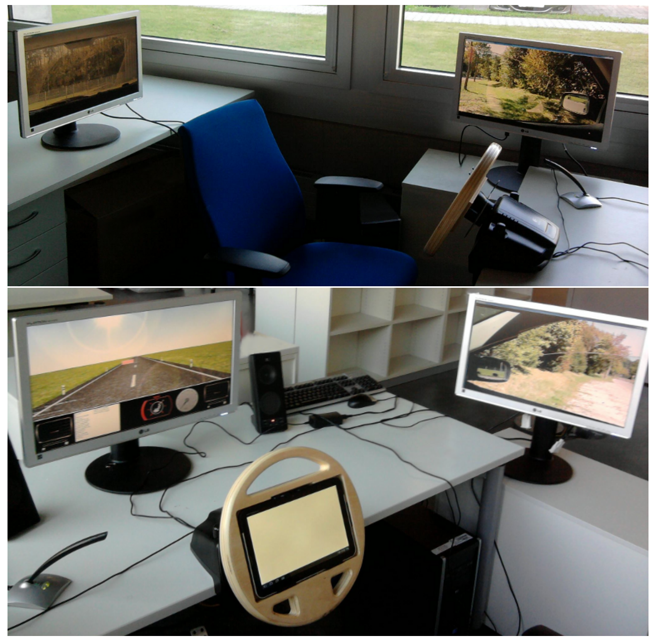

Figure 1: Prototypical implementation of our multimodal interaction style – a microphone captures speech commands while gestures are performed on a multi-touch steering wheel. Screens around the driver's seat simulate the back window, driver/passenger windows and the exterior mirrors. The front screen displays the driving simulator and a virtual dashboard.

Speech input is very powerful in selecting functions and objects by naming them without a need for hierarchical structures and explicit navigation. Gestures instead support fine-grained control of functions very well and permit easy reversal of action. For the automotive domain, previous research has shown that gestures are powerful as they require minimal attention and can be performed without taking the eyes off the road [1], whereas interaction with (graphical) menus and lists is visually much more demanding and results in more driver distraction. Further results are that finding intuitive gesture commands to select functions can be difficult [1]. Based on these insights, we developed a multimodal interaction style where the object to be interacted with and the function that is performed are selected by voice and the actual control is done by a gesture. We expect that the presented interaction style will reduce the visual demand during interaction. In this paper, we report on the design of a first prototypical implementation.

2. USER EXPERIENCE & INTERACTION
Sitting behind a multi-touch steering wheel, the user can manipulate several objects/functions in the (simulated) car by using the proposed multimodal interaction style that combines speech and

gesture interaction (see Figure 2): First, *speech* is used to *select* and *quantify* one or multiple *objects* and their *function* to be controlled. If an object offers only one function to be modified, the selection process can be shortened by just calling the object and implicitly choosing its function, e.g., "cruise control". If multiple instances of an object type exist (e.g., windows), the desired objects need to be quantified (e.g., "passenger window"). The interaction can also be started by just saying the object's name ("window") – if the selection is ambiguous, the system will ask for a suitable quantification until the object and function selections are well-defined. Similarly, for objects offering more than one function, the user clearly has to select the desired function. An integrated disambiguation cycle (see Figure 2) assures an explicit selection of object(s) and function by providing speech prompts to refine the selection if necessary. As the objects are visible in the corresponding environment, it is easy to remember the items of the interaction space and to comply with the visibility principle. Thus, even a larger number of items can be addressed without an increased memory load. After the selection of the interaction object(s) and function, a modality switch takes place and the user performs gestures to modify the desired parameters. This allows for a fine-grained control and an easy reversal of actions. As the action is executed immediately, a direct feedback is given by means of manipulated objects.

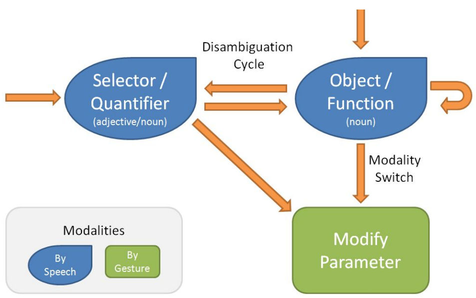

Figure 2: Diagram of the speech-gesture interaction style – Interaction objects are selected by speech and then manipulated by (touch) gestures.

3. STUDY AND IMPLEMENTATION

In a first study conducted with 12 people (2 female, 20–39 years, avg. age 28,25 years; avg. driving experience 10 years) we validated the hypotheses that users can easily identify objects and (secondary and tertiary [2]) functions in the car by speech input without prior training and that users have similar expectations with regard to gestural control. At the same time, the study was used to gather user-elicited speech and gesture commands. In this study, the users first had to identify objects/functions of the car that were presented as pictures on our touch-enabled steering wheel. As a second step, the users should propose a touch gesture to control the corresponding parameter. In 97.8% of the presented scenarios people were able to easily find appropriate terms to name visible and known objects (82.1%) and/or their functions (15.7 %). The study showed, that it is crucial to realize a broad set of different voice commands for a single functionality and that the denotations of visible objects have potential for intuitive voice commands. Further, the study revealed a high agreement on touch gesture commands among participants. Overall, they did not have problems to think of touch gestures and chose very similar and simple gestures to control different functions. 86.5% of the recorded gestures were simple and easily transferrable directional gestures (up/down/left/right). These gestures are based on embodied conceptual metaphors and seem well suited to control the parameters of most objects' functions.

The setup for the study included a speech-enabled multi-touch steering wheel, which was connected to a driving simulator (CARS[1] open source driving simulation) that ran on a PC with a 24" screen. As a steering wheel, we used a commercial Logitech G27 racing wheel base and replaced the actual wheel by a self-made wooden steering wheel containing an integrated Android-based tablet (see Figure 1). An Android application was designed for the study to gather verbal commands and gestural input by presenting different objects and functions in the car.

Encouraged by the results of the study, SpeeT was constructed as an iteration of the prototype. While keeping the steering wheel and driving simulator, a virtual dashboard is included on the front screen (driving simulator) showing gauges for air vents, the current speed (cruise control) and the seat heating. Screens on the left and right side show the driver and passenger windows and the exterior mirrors. A fourth screen simulates the rear window and wiper. By using the proposed interaction style all aforementioned objects can be controlled. A USB microphone captures voice input while gestural input is gathered by an Android app on the tablet and broadcasted via UDP to a control application on the PC. There, all events are processed and mapped to the available functions. This allows the driver to use voice commands to select objects / functions and to conduct gestures to modify the related parameter. Speech interaction is initiated either by using a button on the back of the steering wheel or saying the word "command" to fully activate the speech recognition system. If a speech command is ambiguous, speech prompts (disambiguation cycle) ask to refine the object selection. Direct visual feedback is given to the user by providing the corresponding simulation like an opening window. Figure 1 illustrates the prototypical implementation.

4. CONCLUSION

In this paper, we present a first prototype implementing our novel interaction style to combine speech and (multi-) touch gestures for multimodal input. Speech is used to select the function based on visible objects in the environment. Touch is used to control parameters, providing immediate feedback and easy reversal of action. With this interaction style the advantages of both modalities are exploited and the drawbacks are reduced. A first study revealed that this interaction style is understandable to users. As a next step, we will investigate how visual demands change by using this interaction style compared to existing interaction styles.

5. ACKNOWLEDGMENTS

The project is funded from the German Research Foundation within the Cluster of Excellence in Simulation Technology (EXC 310/1) at the University of Stuttgart

6. REFERENCES

[1] Döring, T., Kern, D., Marshall, P., Pfeiffer, M., Schöning, J., Gruhn, V., and Schmidt, A. Gestural interaction on the steering wheel: reducing the visual demand. In Proc. of CHI'11, ACM, 2011, 483-492.

[2] Geiser, G. Man machine interaction in vehicles. ATZ, 87, 1985, 74–77.

[3] Reeves, L. M., Lai, J., Larson, J.A., Oviatt, S., Balaji, T. S., Buisine, S., Collings, P. Cohen, P., Kraal, B., Martin, J.-C., McTear, M., Raman, T. V., Stanney, K. M., Su, H., and Wang, Q. Y. Guidelines for multimodal user interface design. Commun. ACM 47, 1, 2004, 57-59.

[1] http://cars.pcuie.uni-due.de (last access 2011-10-14).