

Multimodal Dialogue System

Navigation systems, climate control units, electronic entertainment – today's cars provide comfort but also a confusing diversity of operating elements. Voice-activated systems address this problem. However, current spoken dialogue systems have strong limitations. The new system SAMMIE, which was developed by DFKI (German Research Center for Artificial Intelligence), the Saarland University Saarbrücken, BMW and Bosch as part of the EU-funded research project TALK (Tools for Ambient Linguistic Knowledge), allows natural, multimodal interaction in the car.

Intuitive, unrestricted control of the infotainment system

Integrated into a test car, the system demonstrates the next generation of multimodal dialogue systems. It is called SAMMIE (Saarbrücken MultiModal Interaction Experiment) and enables intuitive, unrestricted control of the infotainment system in the car by means of natural, speech-centred, multimodal dialogues. It understands speech input, manual input by the iDrive controller or combinations of both modalities. The output of the system is also multimodal by using SVOX synthesised speech output as it responds to user requests and graphical presentations on the in-

formation display and performs the requested commands on the integrated multifunctional MP3 player. Users neither have to learn special speech commands to operate the system nor do they need to thread their way through a pre-defined multi-level menu. Rather, the system allows the users to formulate requests in their own words, for example, "Please play Yesterday by the Beatles." It will interpret the request regarding the respective dialogue context and either play the song or take the initiative and ask for clarification, for example by asking "From which of these three albums:

'One', 'Red Album' and 'Number One Hits'?", also displaying the album titles on the dashboard screen. The system behaviour is adapted to the dialogue context and user behaviour in several aspects to make sure, for both safety reasons and user comfort, that all information is redundantly provided in all output modalities at all times. For example, users may explicitly formulate specific output modality requests ("show all albums" vs. "read out all albums"). It prefers this requested output modality as long as the user does not explicitly change it. The system operates intuitively at all times, that is, every kind of interaction can be performed at any time in the dialogue in any input modality. The system behaviour is further context-sensitive, as the user and system initiative alternate according to the dialogue context. Moreover, the graphics and speech output of the system is also adapted to the dialogue context and user behaviour. The system understands German and English, even allowing English song titles or album names in a German dialogue.



SAMMIE system in the car.

Flexible Modular Architecture

Possible system interactions with the MP3 system range from browsing, searching and playing music from the MP3 database to the administration of playback lists. This is reflected in the multifunctional GUI, which combines a series of graphical output elements

with visual system information, Figure 1. Database browsing output is shown on the left-hand side with an interactive tabular presentation supplemented by contextual information about the table contents, a scrollbar indicating the sense of rotation and a line selection cursor. On the right-hand side, graphical information about the microphone and player status is shown. The system is implemented in a flexible modular architecture in line with the multi-modal dialogue architecture, comprising components for recognition, analysis, dialogue modelling, generation and output rendering, Figure 2. For speech recognition and speech synthesis, commercial software from speech solution partners was used. The remaining components, especially the dialogue and discourse modeller, are based on the “Extended Information State” (EIS) (an advancement of the “Information State Update” (ISU) approach), which has been combined with the “Collaborative Problem Solving” (CPS) model of dialogue in order to realize a generic, application-independent and thus easily reusable dialogue shell. In addition to that, all dialogue-processing components are implemented in Java and based on a common rule interpreter, DFKI’s PATE system (a production rule system based on activation and typed feature structure elements). The latest version runs in real time on a standard PC with an Intel Pentium 4 processor and 1 GB of RAM. The average reaction time ranges between 0.4 and 1.5 sec. The system was integrated into a test car using standard dashboard devices (display, iDrive, steering wheel buttons, speakers) as input and output devices, except for the Andrea microphone, which has been additionally installed right behind the steering wheel.

The system is a demonstration of how techniques originally built from a scientific perspective (in this case the Information State Update (ISU) approach) can also excel in meeting commercially-oriented requirements. System development involved several development cycles over a period of three years, including Wizard-of-Oz experiments and multiple

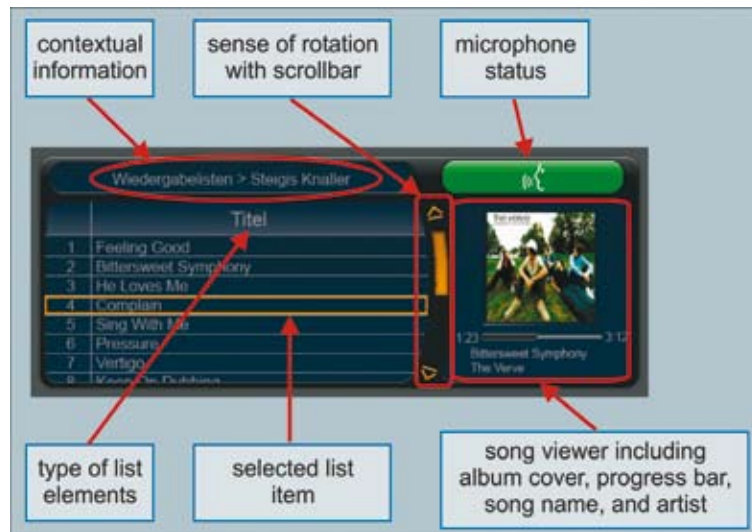


Figure 1: Multifunctional MP3 GUI.

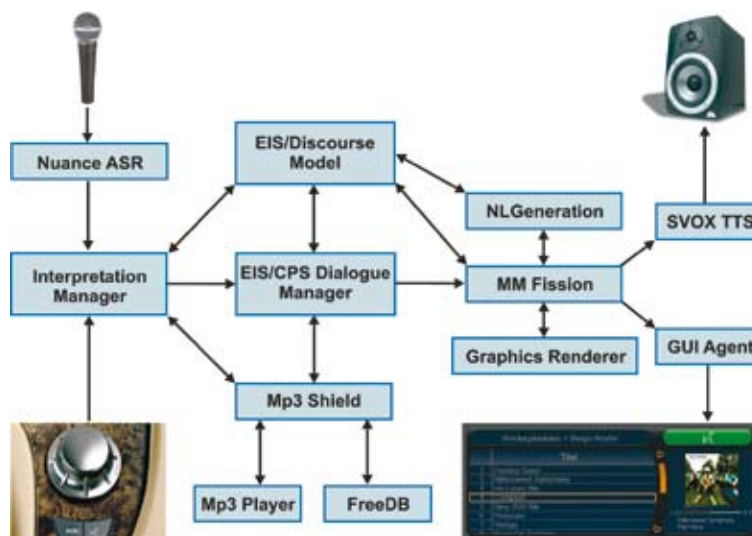


Figure 2: SAMMIE system architecture

usability experiments that lead to the present „Final Showcase“ being integrated in the test car.

The Future

With the system, there now is an architecture and generic tools that can be used to cover all (non safety-related) applications in the car. A version of the system for city-wide restaurant information has already been developed, including multi-modal interaction with maps, icons and a flexible search of the database. Ongoing work is adding more applications and improving individual modules. The next challenge will be the integration of a large number of applications into a common, speech-centred multi-modal interface based on the systems’s architecture. (red)