FLEXIBLE MULTIMODAL HUMAN-MACHINE INTERACTION IN MOBILE ENVIRONMENTS

Dirk Bühler, Wolfgang Minker
DaimlerChrysler - Research and Technology
Ulm/Donau, Germany

Jochen Häußler, Sven Krüger
European Media Lab
Heidelberg, Germany

ABSTRACT

This article describes a prototype system for multimodal human-machine interaction in mobile environments that is being developed within the German national SmartKom research project. The system is novel in that it a) aims at combining the flexibility of state-of-the-art handheld digital devices with the computing power of standard PC machinery, and b) connects two substantially different mobile environments, namely pedestrian and car driver.

Motivated by the specific safety and privacy considerations and requirements in both environments, we present a framework for flexible modality control. A characteristic feature of our framework is the insight that both user and system may independently and asynchronously initiate a modality transition. We conclude with a brief discussion of further issues and research questions.

1. INTRODUCTION

SmartKom [1], the successor of the Verbmobil project [2], is a long-term research effort funded by the German Federal Ministry for Education and Research (BMBF). Started in 1999, the project aims at developing advanced concepts for intuitive human-centered computer interfaces.

These concepts are currently being tested and demonstrated as running systems in three major application scenarios:

SmartKom Home/Office: The system will serve as a personal device comparable to a traditional desktop computer, enhanced with multimodal interaction.

SmartKom Public: The system will be made available as an advanced multimedia communication center in central public places such as airports, train stations, or administrative buildings.

SmartKom Mobile: A small yet powerful portable device will act as a permanent digital companion that provides the systems’ functionality and services to the user in mobile environments such as in a car or outdoors.

The three SmartKom scenarios share a core functionality. A set of “standard applications” has therefore been defined. These include communication (via email and telephone), personal assistance (address-book, agenda), etc. In addition, for each scenario there exists a number of exclusive applications (e.g., incremental and location-aware navigation in SmartKom Mobile).

The most important modalities for human-machine interaction considered in the SmartKom project are speech recognition and synthesis, gesture recognition (such as pointing on the screen or free gestures), recognition of facial expressions using a face tracker, and graphical display of text, animations, and maps.

In this article, we will focus on the SmartKom Mobile scenario and present a first prototype system. Section 2 describes this application scenario and its differences with respect to the other SmartKom scenarios. Section 3 presents a framework for flexible modality control motivated by the identification of certain characteristic interaction modes in SmartKom Mobile. Section 4 describes the implementation platform and a first version of the SmartKom Mobile prototype. Finally, section 5 summarizes the presented concepts and designs and provides an outlook to further development and research issues.

2. THE SMARTKOM MOBILE APPLICATION SCENARIO

The SmartKom Mobile application scenario applies to two related yet significantly different user environments:

Pedestrian environment: The user (walks around and) carries the SmartKom Mobile device as a permanent and location-aware digital companion.

Driver environment: By installing the SmartKom Mobile device into the docking and synchronization station inside the car the system can be used as a car-based digital assistant that takes advantage of a in-car hands free audio system and car-specific functionality, such as real-time information about the car state.
As outlined above the general goal of the SmartKom project is to make available to the user some standard applications and common way of interaction, while flexibly enriching the system with functionality specific to a particular scenario.

For illustration, we outline the navigation functionality as the salient application domain in SmartKom Mobile. It includes:

- **Trip planning**: A trip may consist of multiple route segments to be covered either by walking or driving. This requires specifying start and destination locations, as well as optional stops. The user may also wish to specify different route types (e.g., shortest distance) and other properties of the trip. He may be interested in some information on the trip in advance, e.g., distance, estimated duration, details of points of interest nearby.

- **Trip execution**: This functionality includes incremental guidance and monitoring. It involves timed presentation of driving directions as well as processing incoming positioning information. In the pedestrian environment, the functionality also involves pro-activeness, e.g., by presenting additional navigation-unrelated information available from digital maps.

While the first part of the functionality may also be useful in other scenarios (e.g., for planning and making reservations at home), trip execution is apparently only available within the mobile scenario.

The SmartKom system aims at providing a common and uniform dialogue-based interaction model in the three application scenarios. Therefore, one of the major challenges is to accommodate the various peculiar restrictions that each of these application scenarios pose on the system. This concerns in particular SmartKom Mobile since the user interacts with the system in two substantially different environments. In many situations, the user may also be concerned with tasks other than communicating with the system (e.g., driving) and he is not immobilized in front of the system (i.e., the user frequently moved his head and body). Finally, due to traffic safety considerations, pointing gestures are prohibitive if the system is used whilst driving.

Consequently, for a given application the different dialogue flows need to be adapted across the different application scenarios and also within SmartKom Mobile itself.

### 3. A FRAMEWORK FOR FLEXIBLE MODALITY CONTROL

SmartKom Mobile is different from the Home/Office and Public scenarios in that the control of the interaction modality is a major concern.

Based on experience in developing and testing speech-based user interfaces in mobile environments [3] we have identified five major combinations of modalities (interaction modes) that seem characteristic to the SmartKom Mobile scenario (Figure 1):

![Interaction modes](image)

- **Default**: Mainly used in the pedestrian environment and when privacy or disturbing others is not an issue, all modalities should be enabled in this mode. Stemming from safety considerations (discussed below), this mode is available in the driver environment only when the car is not moving.

- **Listener**: In this mode the user accepts spoken system output but he does not use speech as an input modality (i.e., the user is the listener here). Spoken output is useful for quickly providing concise summaries or extended background information to the user whose focus of attention is placed on tasks other than interacting with the system. The user may employ ear phones in order to protect his privacy. The listener mode may also prove useful for avoiding confusion by off-talk. In this mode, the system does not try to excite spoken input from the user. It should therefore generate prompts like “Show me ...” rather than “Tell me ...”.

- **Silent**: This mode is useful when spoken language human-machine dialogue is problematic, for instance in certain public places or in public transportation. The interface should be similar to traditional graphical user interfaces.

- **Car**: This mode is a restricted version of the default mode for use mainly in the driver environment. Speech is the dominant communication modality, while graphical displays are used only for presenting additional (“non-vital”) information, such as maps for presenting a route while driving.

### Fig. 1. Interaction modes as combinations of communication modalities in SmartKom Mobile. The inter-modal transitions are described in the text.

Speech-Only (suspend)
use of a speech-only dialogue is mandatory.

In the driver environment an emergency requires the human-machine dialogue to be temporarily suspended. This **Suspend** situation (cf. Figure 1) occurs when the driving situation becomes dangerous and the entire user attention is required. Although the system suspends any interactivity, it should be prepared to resume the dialogue, especially after very short interruptions. Temporarily suspending interaction may also be useful in the pedestrian environment when the user is engaged in tasks others than communicating with the system, such as crossing a street or talking to another person.

Having identified these five interaction modes we need to determine, under which circumstances inter-modal transitions are possible.

We distinguish between **user-** and **system-driven** transitions. A user-driven transition is the result of an explicit user command to use or not to use a specific modality. In turn, system-driven transitions are triggered by a system without any user interaction.

The user should be able to switch between modalities anytime, although some modalities may temporarily be disabled for safety reasons.

A user-driven modality transition may be performed by the following actions (cf. Figure 1):

- **The user may suspend/resume the operation with the system.** In the pedestrian environment this operation could be performed by activating a button of the portable device. In the driver environment this could be realized by a Push-to-Activate button or lever.

- **In the pedestrian environment speech input and output may be suspended by activating a button on the portable device.**

- **In the pedestrian environment, the user may toggle display operation.** A request for turning on the display could be recognized by the touch sensitive screen of the portable device.

- **Re-opening the speech channel by uttering spoken commands re-enables spoken language dialogue, e.g., when the system is suspended or when graphical output is exclusively used.**

- **Requesting graphical output by spoken commands (like “Show me the map”) enables graphical output, unless this modality is disabled for safety reasons (cf. speech-only mode).**

In addition to the user, the system may initiate a modality transition in one of the following ways:

- **As outlined above a specialized software running on the car PC may detect an emergency in the driver environment. To this end, the software has access to the information about the state of the car (e.g., driving speed, state of brakes, etc.) through the CAN bus.**

- **Analogously, starting the car is detected by the car PC and interpreted as a transition into the speech-only mode. Likewise, stopping the car leads to the default mode.**

- **In the pedestrian environment, i.e., when graphical output is available, repeatedly failing to understand spoken input from the user should lead the system to infer that the current situation is not suitable for speech recognition (e.g., through strong background noise). The system therefore switches to the listener mode in order to rely not exclusively on speech input.**

### 4. THE SMARTKOM MOBILE PROTOTYPE

A prototype version of the SmartKom mobile system has been designed and implemented. The iPAQ H3600 Series Pocket PC has been chosen as the primary interface because it allows innovative modes of interaction and flexibility.

For the driver environment, a hands-free audio system and a color display is available in the car.

Our primary concern relies on connecting different mobile environments rather than specific technological issues (such as, for instance, speech recognition constrained by the limited resources and computing power of current PDAs). Therefore, all the data processing is performed on the PCs whereas the iPAQ and the car interior hardware are used as mere periphery clients responsible for recording and displaying the input and output data, respectively. The communication of the data is done via a high-speed WaveLAN connection, as shown in Figure 2.

![Fig. 2. Hardware and software components used in the SmartKom Mobile scenario.](image-url)
SMA: Displays initial display
USR: I would like to travel to Heidelberg.
SMA: Displays initial display
Where would you like to start?
USR: Saarbrücken.
SMA: Displays car route
Here you can see a map with the way from Saarbrücken to Heidelberg.
USR: Yes.
SMA: Displays map with points of interest
Here you can see a map of Heidelberg.
USR: Points to Peterskirche
SMA: Slide show about Peterskirche
Here is some information about the Peterskirche.
USR: How can I get to the Peterskirche?
SMA: Displays pedestrian route
The route is displayed in the map.

Fig. 3. Integrated planing of car and pedestrian routes in a single dialogue using the SmartKom mobile prototype system. (a) Multi-modal interaction in the default mode; (b) handheld device showing a pedestrian route.

In the car, the iPAQ will be installed in a docking station. There, it will take advantage of the car interior infrastructure, such as built-in speakers, a microphone array, and a GPS device.

With the current prototype, the SmartKom Mobile system for the pedestrian environment, the user is able to perform integrated trip planning (pedestrian and car) using a simple multi-modal interaction (i.e., speech and pointing gesture). The system displays graphical output in the form of maps (route information) or slide shows (sight information). Using synthesized speech an animated agent provides additional information.

Figure 3 shows an example interaction and a possible output of the system. For now limited to the default mode, we are currently adapting the multi-modal dialogue strategy to the framework described in section 3.

5. SUMMARY AND OUTLOOK

With the SmartKom Mobile prototype system we have developed a new architecture and described a conceptual framework for flexible multimodal human-machine interaction in mobile environments. For research purposes we have combined the flexibility of a digital handheld device with the computing power of standard personal computers. The system is novel in that it allows an integrated trip planning (pedestrian and car) using multimodal interaction.

Stemming mainly from safety and privacy requirements in these environments, we have discussed a framework for flexible modality control. It consists of a state model for modality combination, describing possible interaction modes in different situations. We conclude that both the user and the system should be able to initiate transitions across interaction modes.

A number of issues remain to be discussed. One question is how the graphical displays and the system in general can convey a uniform “Look & Feel” whether or not spoken language is available.

The system should be adaptive (or adaptable) to user preferences and to his experience with the system. Novice users often require system guidance (e.g., in the form of a tutorial mode.) Experienced users may wish to avoid unnecessary verbosity by graphical or spoken short cuts. In addition, memorization and statistical models could be used to automatically update the user profile or infer preferences.

We are currently adapting the multimodal dialogue strategy of SmartKom Mobile to this framework. One of the next steps also consists of the integration of SmartKom Mobile in the car and to enable an incremental route planning for pedestrians.

6. REFERENCES