

---

# An Exemplary Interaction with SmartKom

Norbert Reithinger and Gerd Herzog

DFKI GmbH, Stuhlsatzenhausweg 3, D-66123 Saarbrücken  
{reithinger,herzog}@dfki.de

**Summary.** The different instantiations of the SmartKom demonstration system offer a broad range of application functions and sophisticated dialogue capabilities. We provide a first look at the final SmartKom prototype from the point of view of the end user. In particular, a typical interaction sequence will be presented in order to illustrate the functionality of the integrated multimodal dialogue system.

## 1 Introduction

The three different usage scenarios of SmartKom and its various application functions allow for a wide range of possible interactions. Aiming at flexible and natural, multimodal dialogues, we need to define the intended behaviour of the system as well as its general look and feel. In particular, we have to lay out basic dialogue steps, which the user can freely combine during his or her interaction with the system. This design task leads to an iterative process, which takes the initial project definition as a starting point. System developers, scenario experts and prospective users (e.g. through wizard-of-oz experiments) collaborate to design and refine the capabilities of the multimodal system to be build. Following the paradigm of scenario-based design [2, 12], the initial focus of the design activity is not a formal functional specification but a description of how people will use the system to accomplish work tasks and other activities.

In the next section, we first present the dialogue descriptions we use to define and document the basic interactions in SmartKom as well as the illustrative dialogue protocols that can be generated automatically from the extensive log data resulting from a system run. Then we walk through an original sample dialogue between a test user and the system prototype to provide an insight into the capabilities of the SmartKom demonstrator. A presentation on paper, of course, can only provide a rough sketch of a multimodal interaction. The SmartKom Web site located at [www.smartkom.org](http://www.smartkom.org)

provides a comprehensive video that complements the description of the integrated SmartKom prototype given here.

## 2 From Dialogue Drafts to Dialogue Protocols

For a large, distributed project it is necessary to coordinate the design and development efforts on various levels. One important task is to agree on those dialogue steps and discourse phenomena that the system should be able to process. At the beginning, the different scenario experts performed user studies [1, 9, 11]. The goal was to come up with preferred interaction metaphors and interaction sequences that should be realised. In addition, the wizard-of-oz data collection [13] provided important insights concerning natural interaction sequences.

With this in mind, we designed a template for dialogue descriptions, which were the basis for the communication between interface designers and scenario experts. The idea was to collect all relevant information for turns, i.e., input/output sequences between end user and system, and turn sequences that were to be realised.

Figure 1 shows a typical excerpt from such a document and presents one turn for the English Mobile scenario. A turn description starts with possible inputs from the user. The list of utterances is not exhaustive and contains category types like names. The explanatory notes section below contains remarks about the processing of these input utterances. In the example, it contains all entries of the English base lexicon for sights in the town of Heidelberg. This is also the place to document limitations of the system. The turn description ends with possible verbal system reactions and example screenshots.

Initially, we took the interaction descriptions as defined by the scenario experts and wrote one document for each functionality. The documents have been made available via the SmartKom intranet for project-wide discussion and potential enhancements. During the realisation phase, the information in the descriptions has been further augmented with real processing results from the system. The scenario experts, who had sometimes no immediate access to the latest version of the development system, could then comment on the results and provide feedback. To ensure consistency, the head of the system integration group was in charge of all additions and changes.

Of course it is difficult to argue about a dialog-based interaction, unless you have a video of the interaction or some other sort of protocol. The SmartKom testbed [7] is able to trace all data communication between the various modules of the system. This option is very useful during system development to debug the system on various levels. As the log contains the results from speech recognition, the modality analysis components, and the presentation modules, it provides all information necessary to automatically create a protocol of a particular interaction sequence. Even though it does not contain the animations, important changes in the screen display and the final screenshot are

SmartKom Dialog Description english.doc  
 (Gerd Herzog Page 5 of 18 12.12.2003)

Example dialog for the English demonstrator, which is based on the mobile scenario with integrated route planning.

**SMA:** = SmartKom system, [*system action*] , **USR:** = User, <*user action*>

**6. Turn**

**USR:**

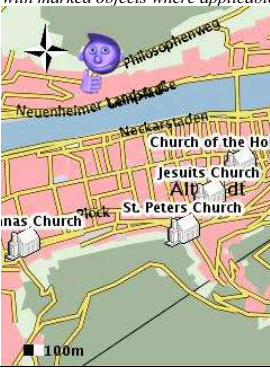
- a) Where is a Category?
- b) Show me a Category.
- c) Show me Category.
- d) What Category are there?
- e) Show me a map with Category, please.
- f) I am looking for a Category.
- g) I'm looking for a Category.

Explanatory Notes:  
 All objects of the given type will be shown on the map. In addition to the given types, specific location names may be used as well.  
 The English base lexicon already contains the following sights: *Philosophers' Walk, Bridge, Saint Peters Church, Grain Market, Palais Morass, Castle* and the following names for object categories: *church, square, cinema, museum, café, pub, park, theater, garage, parking, restaurant, churches, squares, cinemas, museums, cafes, pubs, parks, restaurants, theaters, garages.*

**SMA:**

- a) Here is a map with Category.
- b) There is no information. Here is the city map.

**Screen:** [*Presents city map, with marked objects where applicable*]



Additional Notes:

Fig. 1. An example page of the English dialogue description

included in the log file and can be extracted together with the textual output of the system. The information in the trace file is encoded in M3L [8, 6], the XML-based language which is used for data exchange between SmartKom software components. XSLT style sheets (see e.g. [3]) are employed to automatically extract and format the relevant information from an interaction log. As a result, a browsable HTML page and a self-contained document in PDF format are created, which contain the condensed information for the specific interaction. This documentation of an interaction sequence is used to discuss the implemented dialogue functionality and may initiate a new development iteration, which can also lead to changes in the corresponding dialogue description.

Figure 2 shows a page of the dialogue protocol for the example dialogue to be presented in the next section. Each entry starts with the internal message number of the corresponding data exchange. On the top of the page, there are two screenshots which are the result of the previous user interaction, where the user asked for the cinema programme. The next message in the protocol contains the best hypothesis of the speech recogniser, followed by the word chain that has been selected as the basis for input interpretation in the system. The chosen hypothesis can be different from the best chain since the speech analysis module is able to parse the complete word lattice in order to select the most appropriate semantic interpretation. The next entry represents the derived user intention, which is used by the action planning component to select a suitable system reaction. In this case, the style sheet transformations create a compact predicate-argument structure from the original M3L description, which condenses the sometimes rather lengthy XML markup. Finally, the system output is printed as a text string, followed by the final screenshot of the animated presentation.

In the end, the final dialogue descriptions and meaningful protocols of current interactions provide a comprehensive overview of the detailed capabilities of the multimodal dialogue system.

### 3 An Extended Example Dialogue

In this section, we will walk through an example dialogue to provide the reader with a feeling how the actual system works. As the public information kiosk integrates most of the functionality and features of SmartKom (c.f. [1, 9, 10, 11]) the presentation will use the scenario SmartKom Public instead of the English Mobile system (see [4]). The dialogue interaction in SmartKom Public is based on German, so we will also provide English translations.

The assumed location of the system installation is in the main railway station in the city of Heidelberg and the current date is September 6, 2003. Imagine a user who just arrived in Heidelberg. She accesses the system and is presented with the initial display (see Fig. 3 on the left). To activate the

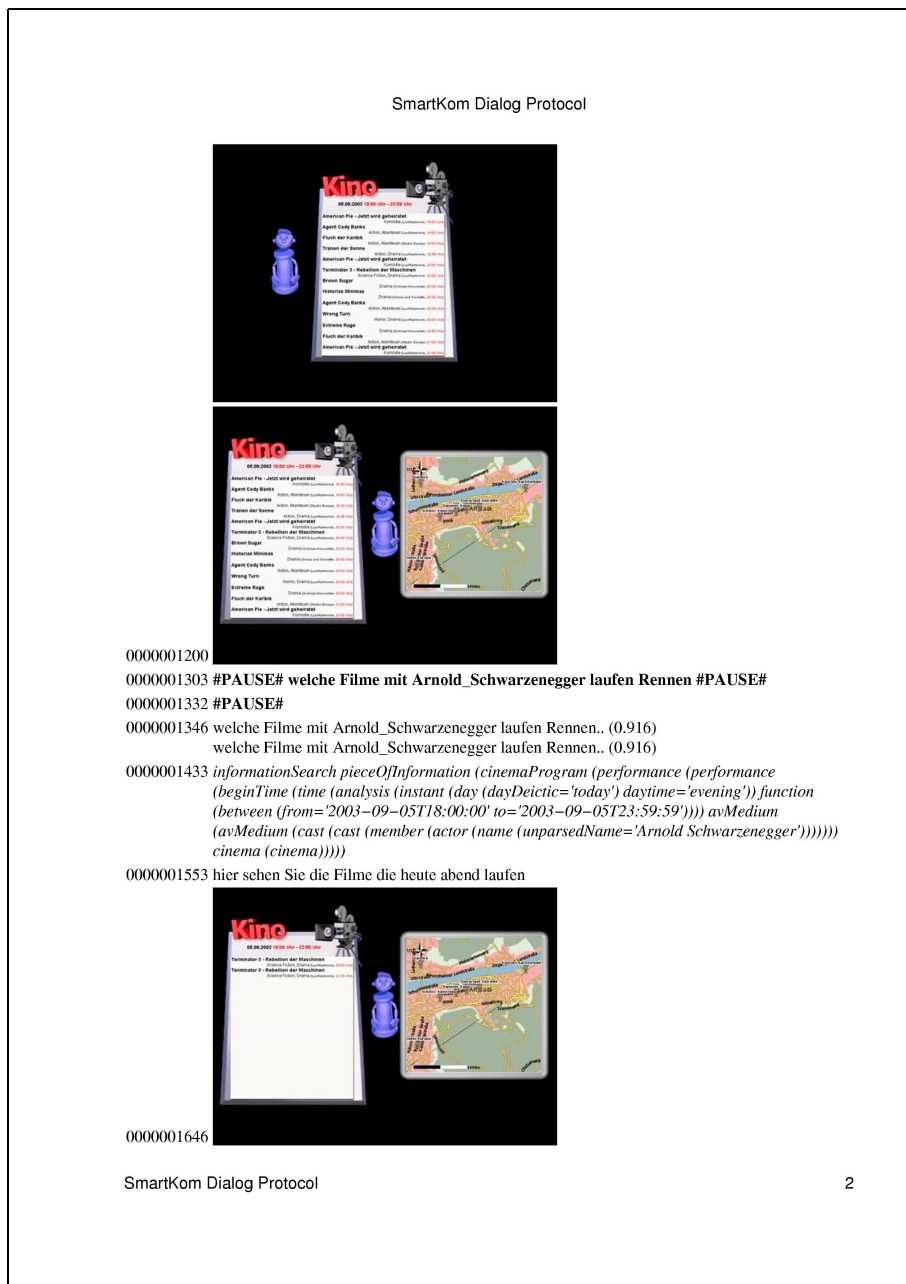


Fig. 2. A page of the interaction protocol for the example dialogue

system, the user can, for example, place a hand in the focus of the gesture recognition camera.



**Fig. 3.** Start screens of the SmartKom system

The interaction agent Smartakus appears on the screen (see Fig. 3 on the right) and greets the user

- (1) SYS: Herzlich willkommen beim SmartKom-System. Ich bin Smartakus wie kann ich Ihnen helfen?  
*(Welcome to the SmartKom information system. I am Smartakus. How may I help you?)*

The user considers to go to the movies tonight and therefore asks

- (2) USR: Was kommt heute Abend im Kino?  
*(What is playing at the cinema tonight?)*

Figure 4 shows the processing results: On the left hand side the cinema programme, and on the right hand side a city map of Heidelberg with cinema locations. Note that SmartKom anticipated that the user might be interested to know where the cinemas are. This is in particular important, if for example a movie is playing at various theatres. Smartakus provides as spoken information

- (3) SYS: Hier sehen Sie die Filme die heute abend laufen. Auf dieser Karte sind die Kinos markiert.  
*(These are the movies playing tonight. The cinemas have been marked on the map.)*

Our user recalls that there is a new movie with Arnold Schwarzenegger, so she asks

- (4) USR: Welche Filme mit Arnold Schwarzenegger laufen denn?  
*(Are there any movies featuring Arnold Schwarzenegger?)*



pointing to one of the movie titles. SmartKom replaces the movie information frame and displays the description for this specific film (see the enlarged display in Fig. 5 on the right), telling the user

- (7) Sys: Ich habe diese Informationen zu “Terminator 3 – Rebellion der Maschinen”.  
*(I have got this information about “Terminator 3: Rise of the Machines”.)*

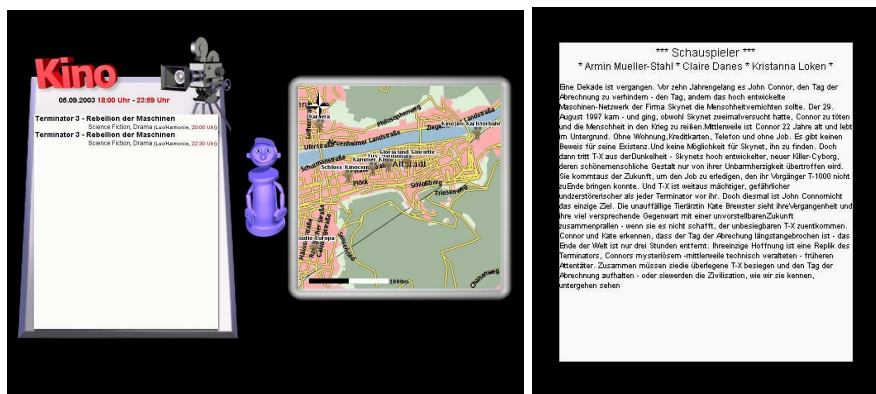


Fig. 5. Information about the film “Terminator 3”

The user wants to visit this cinema showing with her friend and tells Smartakus

- (8)USR: Bitte 2 Karten für diesen Film reservieren.  
*(Please reserve 2 seats for this film.)*

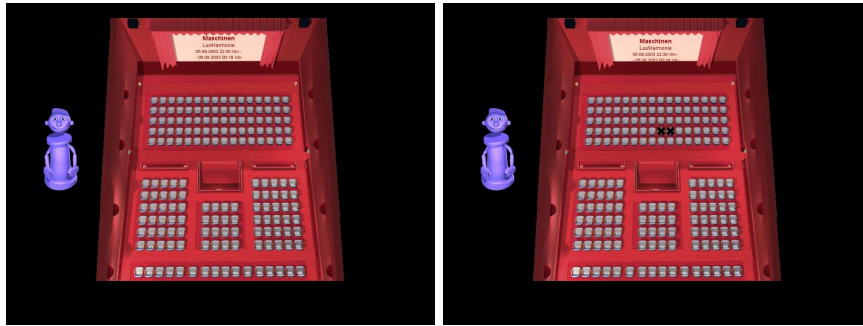
In the context of the ongoing dialogue it is not ambiguous which performance she addresses and the system presents the floor plan of the selected cinema (see Fig. 6 on the left), asking

- (9) Sys: Zeigen Sie mir wo Sie sitzen wollen.  
*(Show me where you want to sit.)*

The user approximately encircles the desired area and says

- (10)USR: Dort möchte ich sitzen.  
*(I want to sit there.)*

Again, the discourse context provides the information that she wants two seats and thus a proposal can be made. Smartakus marks two convenient seats (see Fig. 6 on the right) and asks



**Fig. 6.** Floor plan of the cinema

- (11) SYS: Ist das so richtig?  
*(Is this correct?)*

The user is convinced, agrees, and gets the reservation number from Smartakus.

- (12) USR: Ja, genau.  
*(Yes, perfect.)*

- (13) SYS: Ich habe die gewünschte Reservierung für Sie durchgeführt. Ihre Reservierungsnummer ist 20. Sie können die Karten bis spätestens eine halbe Stunde vor Vorstellungsbeginn an der Kinokasse abholen.  
*(I reserved your seat. Your reservation number is 20. Please get your tickets not later than 30 minutes before the movie starts.)*

Now she wants to know the route to the cinema and asks

- (14) USR: Wie komme ich von hier zum Kino “Studio Europa”?  
*(Can you show me the way from here to “Studio Europa” cinema?)*

Originally, the functionality for pedestrian navigation has been conceived for the scenario SmartKom Mobile. The consistent, knowledge-based approach of SmartKom, however, facilitated the inclusion of this application also in other scenarios. So, later on, the pre-trip planning has simply been added to SmartKom Public. The system knows its own location from its general context knowledge source and is thus able to resolve the adverb “here”. The public information kiosk is located at the main station in Heidelberg and SmartKom can compute the pedestrian route from there to the selected cinema. Figure 7 shows the map with the route inserted. Since the computation of the route takes some time, Smartakus also informs the user to be patient and says

- (15) SYS: Die Route wird berechnet. Hier sehen Sie die Karte mit der Route von Hauptbahnhof zu Kino “Studio Europa”.  
*(The route is being computed. Here you can see the map with the route from main station to cinema “Studio Europa”.)*



Fig. 7. Pedestrian route from main station to cinema

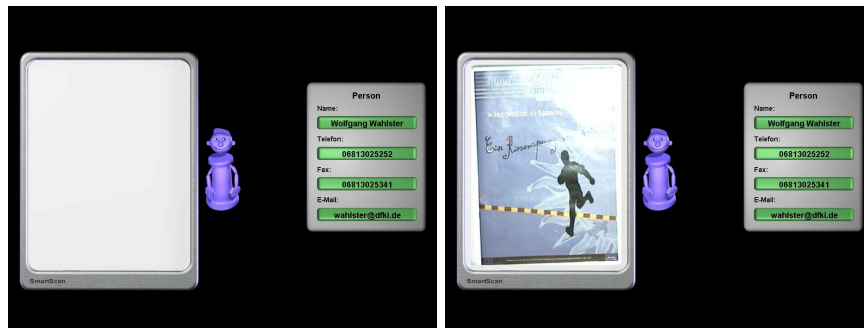
After she has taken care of today's evening entertainment, our user recalls that there is an important document that needs to be sent to her director. She tells Smartakus

- (16) USR: Ich will ein Dokument an Wolfgang Wahlster verschicken.  
*(I would like to send a document to Wolfgang Wahlster.)*

Sending a document requires the address of the recipient. SmartKom has therefore to combine two different applications to satisfy a single user request. It looks up its address database for the coordinates of Wolfgang Wahlster. Being successful, Smartakus presents a scanning area and the address information (see Fig. 8 on the left) saying

- (17) Sys: Ich habe den Eintrag von Wolfgang Wahlster gefunden. Bitte legen Sie das Dokument auf die dafür vorgesehene Fläche.  
*(I found the address for Wolfgang Wahlster. Please put the document on the scanning area.)*

Following the request of the system, the user places the document on the area. The document camera detects the object and scans it. At the end of the scan process, Smartakus tells the user



**Fig. 8.** Document scanning area and address information

- (18) Sys: Bitte nehmen Sie es jetzt wieder weg.  
*(Please remove it now.)*

The system displays the scanned document in the scanning area (see Fig. 8 on the left) and finally sends it via electronic mail.

- (19) Sys: Das Dokument wurde aufgenommen. Das Dokument wurde abgeschickt.  
*(The document has been scanned. The document has been sent away.)*

The user is satisfied with SmartKom and leaves the system to meet her friend, saying

- (20) USR: Tschau.  
*(Good bye.)*
- (21) Sys: Auf Wiedersehen.  
*(See you again.)*

## 4 Conclusion

In this contribution, we gave an impression of how the multimodal dialogue capabilities of SmartKom are defined, starting from dialogue description of scenario experts, resulting in detailed documentation of dialogues and their individual turns. The recorded example dialogue presented here provides an idea of one particular interaction sequence, which combines seven different applications—cinema programme, city information, seat reservation, pedestrian navigation, document scanning, address book, and email—in a coherent, seamless interaction.

The subsequent chapters of this volume will provide a thorough discussion of the underlying methods and techniques that are required to achieve this kind of advanced multimodal dialogue functionality.

## References

1. A. Berton, D. Bühler, and W. Minker. Mobile. In this volume.
2. J. M. Carroll. *Making Use: Scenario-Based Design of Human-Computer Interactions*. MIT Press, Cambridge, MA, 2000.
3. J. R. Gardner and Z. L. Rendon. *XSLT and XPATH: A Guide to XML Transformations*. Prentice Hall PTR, 2001.
4. D. Gelbart, J. Bryant, A. Stolcke, R. Porzel, M. Baudis, and N. Morgan. SmartKom English: From Robust Recognition to Felicitous Interaction. In this volume.
5. S. Goronzy, S. Rapp, and M. Emele. The Dynamic Lexicon. In this volume.
6. G. Herzog and A. Ndiaye. Building Multimodal Dialogue Applications: System Integration in SmartKom. In this volume.
7. G. Herzog, A. Ndiaye, S. Merten, H. Kirchmann, T. Becker, and P. Poller. Large-scale Software Integration for Spoken Language and Multimodal Dialog Systems. *Natural Language Engineering*, 10, 2004. Special issue on Software Architecture for Language Engineering.
8. G. Herzog and N. Reithinger. The SmartKom Architecture: A Framework for Multimodal Dialogue Systems. In this volume.
9. A. Horndasch, H. Rapp, and H. Röttger. SmartKom-Public. In this volume.
10. R. Malaka, J. Häußler, H. Aras, M. Merdes, D. Pfisterer, M. Jöst, and R. Porzel. Intelligent Interaction with a Mobile System. In this volume.
11. T. Portele, S. Goronzy, M. Emele, A. Kellner, S. Torge, and J. te Vrugt. Smartkom-Home: The Interface to Home Entertainment. In this volume.
12. M. B. Rosson and J. M. Carroll. Scenario-based Design. In J. A. Jacko and A. Sears, editors, *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*, pages 1032–1050. Lawrence Erlbaum, Mahwah, NJ, 2003.
13. F. Schiel and U. Türk. Wizard-of-Oz Recordings. In this volume.