SmartKom-Public

Axel Horndasch[1], Horst Rapp[2], and Hans Röttger[3]

[1] Sympalog Voice Solutions GmbH, Erlangen, Germany
horndasch@sympalog.de
[2] MediaInterface Dresden GmbH, Dresden, Germany
rapp@mediainterface.de
[3] Siemens AG, CT IC 5, München, Germany
hans.roettger@siemens.com

Summary. SMARTKOM-Public is the result of consistent development of traditional public telephone booths for members of a modern information society in the form of a multimodal communications booth for intuitive broad-bandwidth communication.

1 Introduction

The SMARTKOM-Public communications booth is provided with a wide palette of modern communication appliances. SMARTKOM-Public offers a higher degree of privacy, much higher communication bandwidth and optimal quality with regard to input and output results than is available to mobile systems for technical reasons. High rates of data flow and high-resolution screens lead, in contrast to a personal digital assistant (PDA), to increased comfort and better ergonomics. This communication space offers the simultaneous and comfortable use of language, pictorial and data services. A document camera is integrated for the transmission and processing of documents. An interesting further possibility concerns the saving of the user's data in the communications booth. Upon biometric identification, users are able to access their personal data. The idea of a virtual communications workstation that is capable of traveling with the users and adapting to their needs becomes reality. As the product at hand deals with a communication space for the public, along with supplying services, intuitive usability is of primary importance for acceptance. This is particularly so as even people who do not have access to a home computer with similar services will also make use of SMARTKOM-Public. The human–machine interface is designed to replicate normal human means of communication. This includes the broadest range of free speech dialogue possible and the use of natural gestures. Consideration must be made that conventional uses, and not only those based on multimodal communication, can be performed from the communication booth. Therefore our project started with a user study. The results are presented in the next section, followed by a description of the implemented functionality and some detailed information on selected realization topics in subsequent sections.

2 User Study

2.1 Introduction

SMARTKOM-Public demonstrates fundamental techniques of a user-friendly, multimodal human–machine communication in public areas. In order to have a close connection between the project specification and the behavior and the wishes of the potential users, the project was accompanied by a user study from the beginning. This study was to give answers to following questions:

- What are the expectations of potential users concerning a multimodal communication cell?
- Which kinds of human–machine communication does the user accept?
- Which communication services have to be provided?
- Which fears and application thresholds are connected with these new techniques?

2.2 Milestones

The data of the user study were obtained by questionnaires in two phases. In a first phase (9/99–5/00) a questionnaires from the technical point of view was elaborated. The questionnaires were adjusted among the SMARTKOM partners. Around 100 persons took part in the first part – mainly ,,insiders,“ persons who were directly involved in SMARTKOM or who had a very good knowledge about the project.

In the second phase (6/00–12/00) a more structured questionnaire was developed in collaboration with the Institute for Communication, Information and Education (KIB) of the University of Applied Sciences Zittau/Görlitz. The questions were asked in more detail, and the answers had to be given on a 5-point scale. It took the participants around 30–60 minutes to fill in the questionnaire. The questionnaire was provided in an Internet version, and owing to different marketing strategies, like mailings actions and a quiz, around 400 persons took part in the second phase.

2.3 Questions and Results

2.3.1 Demographic Data (“Who?”)

- Gender: 50 % female, 50% male
- Age: 40% below 30 years
  50% between 30 and 50 years
  10% above 50 years

The mean age was 34 years.

- Education: ca. 33% school
  ca. 33% colleges
  ca. 33% universities

- Concerning their professions, there was a broad variety among the participants with a certain concentration of employees.
- Their computer skills were assessed as follows:
  60% good
  30% average
  10% bad

Thus it can be concluded that the study achieved a well-balanced and representative mixture of potential users concerning sex, age, education and skills.

2.3.2 Application Domains ("What?")

With these questions the preferred application domains of the potential users for SMARTKOM services were investigated. The participants were asked to give their opinions on four domains:

- information (timetables, weather, holiday planning, cultural and sport events, tourist and cultural locations, stock values, navigation and complex information)
- reservations (hotel, cultural events)
- orders and booking (tickets, products, travels)
- financial services (bank and stock business)

The potential users gave the following assessments to the application domains:

- Information concerning timetables and routes found a high degree of acceptance.
- Information about locations, events and complex information was also accepted.
- Information about weather, travel and especially about stock values found only a low acceptance.
- Reservations in cinemas and hotels were accepted.
- From the application domains directly connected with financial transactions, ordering of tickets was the only accepted area.
- Ordering of products and the booking of holiday trips were refused, more or less.
- For financial services the SMARTKOM-Public communication cell was not found to be suitable.

2.3.3 Media Services ("Which?")

Within these questions the technological equipment of a multimodal communication cell had to be estimated. The following media could be chosen:

- telephone
- video telephone
- fax
- Internet
- Access to personal mailboxes

The answers coincided in a high degree:

- All the media offered were accepted.
- The telephone as the most usual and best-known means of communication found the highest acceptance.

2.3.4 User Groups and Places of Installation ("Where?")

From these questions it should be concluded which places of installation would be preferred by different user groups. The participants of the questionnaire were offered the following situations:

- business traveler in an unknown city
- vacationer in an unknown city
- inhabitant in their home city
- employee in an office

2.3.5 Interaction and Technology ("By what?")

With these questions the modalities and possibilities of interaction were assessed. The participants had to give their preferences concerning:

- input options (speech, gestures, free input by pencil, mimic, document input)
- output options (text and pictures on a screen, speech, video telephone, printer)
- others (use of personal devices, chipcards for identification and payment)

The participants gave the following adjustments:

- High acceptance rates for input by speech and hand gestures.
- Input by pencil and document camera was also accepted.
- Mimic as input modality was refused. (Here it can be presumed that the importance of mimic for the dialogue was not understood sufficiently. Further, the mimic input requires a permanent video recording of the face, which was not accepted.)
- Among the output modalities speech and visual output were preferred.
- The printer was estimated as very useful (for faxes, roadmaps, receipts for reservations, etc.)
- The chipcard reader and the plug-in possibility for personal devices were also accepted.

2.3.6 Verification ("How?")

For purposes of verification the following items had to be adjusted:

- personal identification number (PIN)
- signature
- voice
- hand contour

- finger print
- face

The results were as follows:

- The well-known methods with chipcards or PIN were accepted.
- Among the biometrical methods, the fingerprint found the highest acceptance.
- Signature and face recognition were refused in many cases.

2.3.7 General Requirements for a Communication Cell

Different questions about general requirements for a communications cell gave the following views:

- There were high expectations concerning the intelligence of the system. More than 70% of the participants expected a successful dialogue with only a few clearing steps.
- The communication cell and the services were regarded more as an office employee. Only a third of the participants saw the technical system in the foreground.
- Discretion and anonymity were expected by the majority of potential users. Alternatively to the loudspeaker, a telephone was demanded. The video recording for the mimic recognition was refused in most of the cases. Around 75% of the participants would not make use of the services if they were connected with personal identification.
- There was a high consensus about the fact that better technological possibilities should be offered at constant prices. The only additional costs accepted were taxes comparable to information services and other special services like fax.

2.3.8 Correlation Analysis

In order to investigate the relations between the judgments of the potential users, their answer were analyzed in a paired correlation matrix. For example, there were high correlations in the answers for questions like “information for holiday travel” and “booking of holiday travel”. This seems logical — if somebody wants to book travel, of course they want to get information about it beforehand. From the variety of such logical relations, two main conclusions can be obtained:

- The correlation matrix gives indications for producers of future devices on which applications are compatible with each other and can be integrated into one product.
- The answers were well chosen, which underlines the fact of representative nature of the study.

2.3.9 User Group Analysis

From the analysis according to the gender it can be concluded that female and male participants tended to give the same answers. Women showed less interest in technological details, and their expectations concerning the intelligence of the system were not so high. On the other hand, they were more restrictive in the face recognition problems. The user groups “below 30 years” and “between 30 and 50 years” tended to give the same answers in tendency. The user group “above 50 years” had different opinions on some questions. Elderly people want make less use of the multimodal communication cell in general. They were less interested in the technical possibilities, and they are less restrictive in security problems. The comparison between “insiders” and “others” showed that the answers also were given with the same tendency. The “insiders” however were slightly more optimistic and in agreement with the scenario.

2.3.10 User Suggestions

The questionnaires asked the participants to notify individual suggestions, opinions, apprehension or ideas. Fortunately, this possibility was used in a high degree. In summary, a catalogue of useful ideas containing details of technical question and general requirements could be obtained. Some of the most interesting statements follow:

Applications

- information and booking for actual events in the city (exhibitions, sport, cinema, conferences, etc.)
- restaurant guide with detailed information concerning the menu and quality assessments
- service for results of sport events
- social and political information about the city/the country
- virtual city tours, information about sights
- regional and global news service
- entertainment for waiting times, Internet games
- city map and city information with guide to locations of cultural interest, administration, service
- information about working hours of offices
- substitution of ways to the office (for registration of cars, new passports, etc.)
- flat and property service
- job and qualification announcements

Locations

- villages and small towns for elderly and less mobiles persons
- central places and transportation hubs for business and tourism
- inner city and important/big buildings (post and other offices, shopping malls. . .)
- medical centers, parking areas, truck stops, schools and universities

General Requirements

- interaction results should not be obvious to any but current user
- safety from vandalism
- anonymity, only PIN accepted
- guarantee of data security
- ease of use for physically challenged people
- service free of charge in the tourist information centers
- special services (fax, etc.) at normal charges
- clearly structured with a good overview in each application domain
- simple and intuitive handling

2.4 Summary

In the study, a representative variety of potential users took part, thus the results are valuable. The judgments of the application confirmed the SMARTKOM Demo scenarios. The most important locations are tourist and transportation centers. The different input and output modalities were accepted; only the mimic recognition was refused. User verification by PIN and fingerprint was more accepted than by signature or face recognition. The intelligence of the system was expected to be very high in general. The requirements concerning discretion, anonymity and data security were emphasized with a high degree of importance. The criterion "ease of use for physically challenged people" played an important role. The high number of logical correlations proves the representative nature of the study. They give indications for producers of future devices on which applications are compatible with each other and can be integrated into one product. From the point of view of user groups there were significant differences between female and male users. Elderly people showed less interest in SMARTKOM itself and its technological possibilities. The "insiders" of the SMARTKOM scenario showed a more optimistic behavior compared to the "general user." With the verbal notices a valuable collection of ideas concerning application domains and technological improvements is available.

3 Functionality

The SMARTKOM-Public demonstrator provides a wide range of applications considering the results of the user study. The focus is on communication and information services. Since it was not possible to implement all user-suggested services, some exemplary applications were selected:

- phone call as the most common and best-known means of communication, here supported by an electronic address book
- fax and e-mail services to realize the transmission of written documents
- a cinema information and booking system combined with a city map
- multimodal biometric authentification to access personal data

Fig. 1. The former German federal president using SMARTKOM-public e-mail

Altogether, SMARTKOM-Public offers more than 20 functionalities that can be combined in many ways.

These functionalities are accessed through an easy to use multimodal user interface that combines speech, gesture and mimic input with symmetric output. The optical output is projected on an interaction surface, where it can be referenced by natural pointing gestures. This part of the system is based on the Siemens Virtual Touchscreen (SiViT) system (Maggioni and Röttger, 1999). A second dedicated camera scans documents placed on the interaction area for further processing. As demanded by potential users, all applications can be accessed with only a few clearance steps. Although in the user study potential users expressed objections to mimic input, it was integrated into the demonstrator for scientific reasons and to evaluate possible applications.

Another compromise was made on the physical design of the communication booth. On one hand, potential users stressed their need for discretion and anonymity, demanding a traditional closed telephone booth design. On the other hand, SMARTKOM-Public is a showcase to demonstrate new technologies to the interested public, which requires an open design. We decided to use an open information terminal design for the second reason. For the same reason we used hands-free communication instead of a traditional phone handset.

The next sections describe the basic SMARTKOM-Public applications in detail. For more information on the gesture and mimic input, see Racky et al. (2006) and Frank et al. (2006).

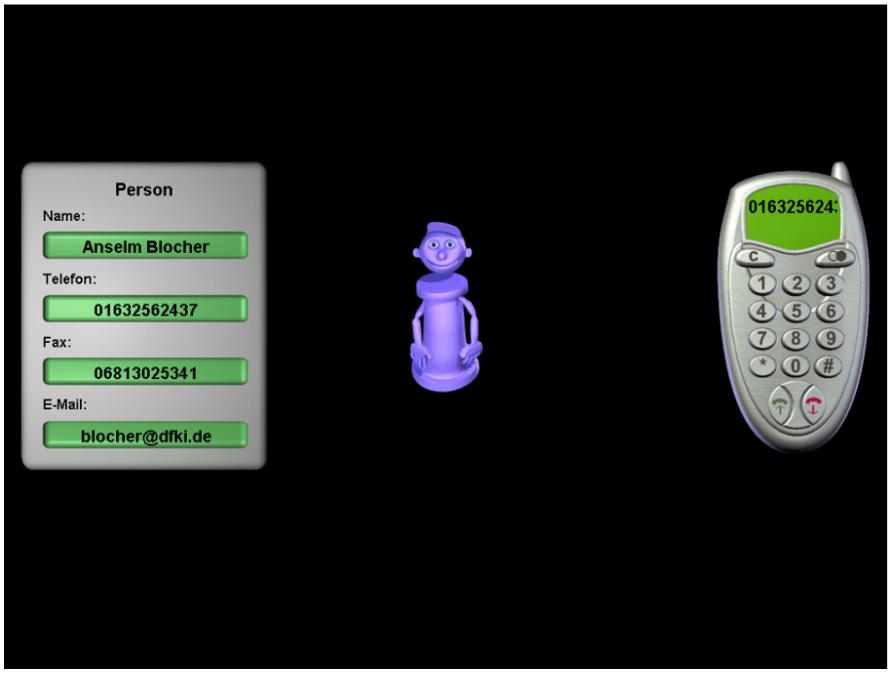

Fig. 2. Phone application in SMARTKOM-Public

3.1 Phone and Address book

Just like a traditional public telephone booth SMARTKOM-Public allows users to make standard telephone calls. The life-like character "Smartakus," visible on the projected output, serves as an interface agent to meet users need. Thus SMARTKOM-Public's interaction style breaks radically with the traditional desktop metaphor. It is based on the situated delegation-oriented dialogue paradigm (SDDP (Wahlster et al., 2001). The user simply tells Smartakus to establish a phone connection. This can be done in two different ways:

- dialing manually on a projected phone keyboard
- dialing by a phone book reference

In contrast to other scientific demonstrators, SMARTKOM-Public establishes real phone connections based on the European ISDN standard and thus has to handle signalization events correctly. It also uses hands-free communication based on the SMARTKOM audio modules and provides an address book based on the standard application module (see Sect. 4 for technical details).

Let us close the phone section with a look at two sample phone dialogues:

Example 1: Dialing manually on a projected phone keyboard

USER:	I'd like to make a call.
SMARTAKUS:	(presents phone keys)
	Please dial a number.
USR:	(dials number)
SMARTAKUS:	(establishes connection)

Example 2: Dialing by a phone book reference

USER:	I'd like to call Mr. Blocher.
SMARTKOM	(presents entry for Mr. Blocher)
	I've found an entry for Mr. Blocher.
	Is this correct?
USR:	Okay.
SMARTAKUS:	(establishes connection)

3.2 Facsimiles and E-Mail

Users can send facsimiles (fax) and e-mails with SMARTKOM-Public as well. Smartakus receives the users' requests and guides them through the necessary steps.

Transmission of pictorial information is mainly used for documents related to business or original artifacts that cannot be described in words. Therefore SMARTKOM-Public provides the possibility to capture prepared information with a high-resolution document camera and to send it as a picture to the receiver. For example, it is possible to send an image of your just-purchased new tie — do not try this with a standard fax machine! If the user decides to send an e-mail, the original is captured as a true-color image and sent to the selected e-mail address as an e-mail attachment. For a fax the color image is converted into a black and white version and sent as a real fax to the selected ISDN phone number.

The image transmission functionality is a complex task combining several modules into one service (see Sect. 4 for technical details).

Here is a typical fax dialogue to illustrate the interaction steps:

USER:	I'd like to send a fax to Mr. Wahlster.
SMARTAKUS:	(presents scan area)
	Please place your document here.
USER:	(places the document)
SMARTKOM:	(captures the document)
	Please remove the document.
USER:	(removes document)
SMARTKOM:	The document was captured.
	I've found an entry for Mr. Wahlster.
	Is this correct?

Fig. 3. Fax application in SMARTKOM-Public

USER: Yes.
SMARTAKUS: (sends fax to Mr. Wahlster)

3.3 Cinema

The cinema application was chosen to demonstrate the advantages of the SDDP in a complex information service environment. It was also favored in the user study. The cinema application provides a wide range of choices:

- the current cinema program
- the movie schedule in specific theaters
- information on specific movies
- information on movie actors
- locations of the theaters and directions to those

As in all SMARTKOM-Public applications, the presented information is real and always up to date. A module for external information services scans the Internet for current data. The vocabulary for the speech recognition is updated dynamically as well. Thus it is even possible to ask questions on completely unknown actors, for

Fig. 4. Cinema application in SMARTKOM-Public

example, “Is there a movie with Leonardo DiCaprio?”. Here is more serious sample dialogue:

USER:	I’d like to go to the movies tonight.
SMARTAKUS:	(presents a list of movies and a map)
	Here is a selection of tonight’s movies.
	I’ve marked the cinemas on the map.
USER:	(points to the cinema “Kamera” on the map)
	What’s on here?
SMARTAKUS:	(presents the program of the cinema “Kamera”)
	These are tonight’s movies at the “Kamera”.
USER:	(points to *Matrix: Reloaded*)
	Tell me more about this movie.
SMARTAKUS:	(presents additional information)
	Here is more information on *Matrix: Reloaded*.
USER:	I’d like two tickets for this film at 8 pm.
SMARTAKUS:	(presents seat plan of the “Kamera”)
	Where do you want to sit?
USER:	(encircles specific seats)
	Here.

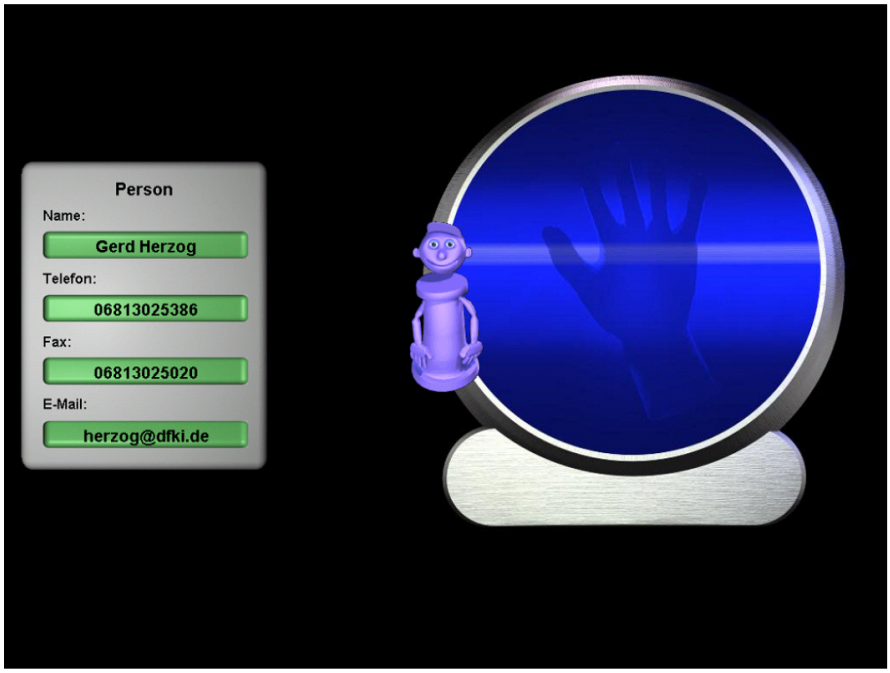

Fig. 5. Multimodal biometrics in SMARTKOM-Public

SMARTAKUS: (highlights the selected seats on the plan)
Is this correct?
USER: Exactly.
SMARTAKUS: I reserved the tickets.
Your reservation code is 42.

3.4 Biometrical Authentification

As mentioned in the Introduction, SMARTKOM-Public provides personalized services. Therefore it is essential to determine the identity of the user. This is done by multiple biometrics (see Grashey and Schuster (2006) in this volume). To guarantee comfort and safety, three different biometrics have been integrated: speaker verification, signature verification and hand contour verification. Authentification can be actively initiated by the user with the phrase “My name is” or can be demanded as the situation calls for:

SMARTAKUS: This action requires an authentification.
What’s your name?

USER:	My name is “Herzog”.
SMARTAKUS:	I’ve found an entry for Gerd Herzog. Hand contour is the default biometry. Please place your hand with spread fingers on the marked position.
USER:	(puts his hand on the interaction area)
SMARTAKUS:	Authentification accomplished.

4 Technical Realization

As is the case for all SMARTKOM demonstrators, SMARTKOM-Public is based on the common SMARTKOM system architecture handling the dialogue with the user. The SMARTKOM system architecture consists of several independent modules connected via communication pools. Most of these modules have been described in the previous chapters. The following sections describe SMARTKOM-Public–specific modules:

- the communication module supported by the audio modules, realizing phone calls and data transfers on standard ISDN lines
- the module for standard applications, realizing address book, e-mail and calendar functionalities
- the virtual module for sending documents, as a combination of document camera and communication module.

4.1 Communication (TK-Module)

The TK-Module in SMARTKOM realizes the telephone and fax services on a Microsoft Windows PC. It works on the hardware base “Fritz!Card PCI” manufactured by AVM, which provides an ISDN S0 connection with two channels.

The software of the TK-Module is realized in a layer structure. The hardware driver is controlled by CAPI2.0. This is enclosed in a library that provides the functionalities typical for telecommunication. The most important features are the support of n-channel telephone calls and fax, the support of ISDN functions and the conversion between the audio formats PCM (in SMARTKOM) and A Law (in ISDN). The highest layer consists of a console application. It controls the internal workflow and allows the parameter input.

In the “telephone” operation mode the TK-Module works together with the modules AudioIn and AudioOut, which are controlled by the “function modeling.” During a telephone call AudioOut is occupied exclusively and thus it cannot submit other system outputs.

Each task for the TK-Module contains the telephone number to be dialed and a name of the connection. This name is also used to identify all the data streams of the connection.

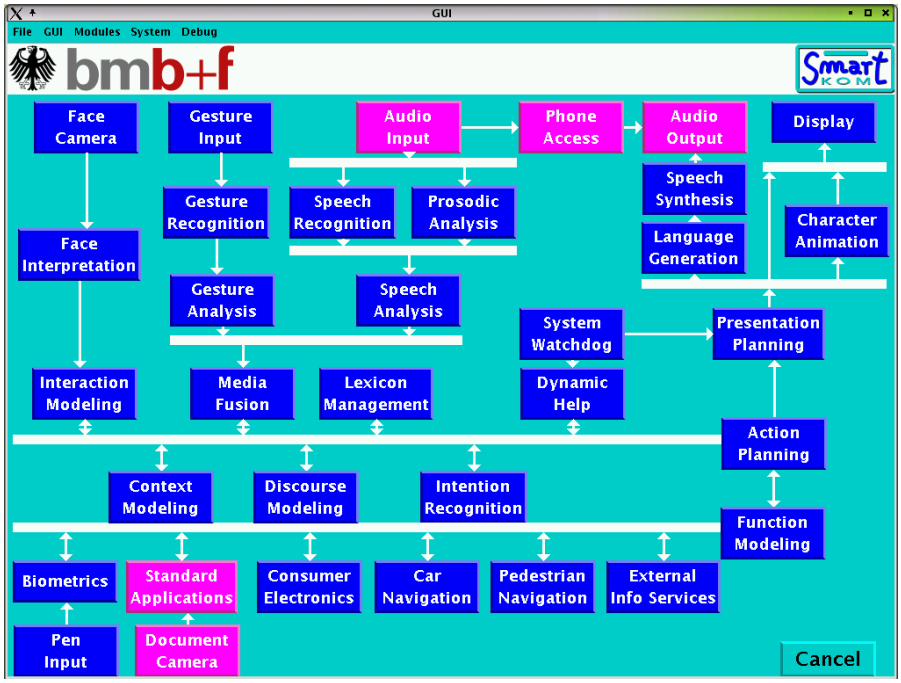

Fig. 6. SMARTKOM-Public module overview

In the “fax delivery” operation mode the TK-Module works together with the “realDocument” module controlled by the “function modeling.” The TK-Module receives the documents to be sent in the structured fax file (SFF) format together with the number to be dialed and the name of the connection.

4.2 Audio Modules AudioIn and AudioOut

The AudioIn and AudioOut modules are responsible for the input and output of audio data into and out of SMARTKOM, not only to realize phone calls but also for speech input and output. The modules are realized in different objects:

WIN-32 Console Application: This application realizes the communication of the audio modules with the SMARTKOM system using the SMARTKOM pool architecture supervised by the module manager. It controls the internal workflow and allows the parameter input. It is also responsible for the time synchronization of the audio data.

DLL Object: This multithread capable dynamic link library contains all functions for the communication of SMARTKOM with different audio sources and audio consumers. The most important features are the support of *n* channels, the mixer functionality and the handling of different PCM audio formats (sample rate, etc.). Audio data can be submitted and delivered local area networks (LAN) and also

by wireless LAN. The connection to the local audio devices is realized by the Multimedia-Microsoft-Win32-API. The remote connection is based on the RTP protocol. The devices for audio input and output can be switched in runtime by parameters.

The audio modules are controlled by the "function modeling" module. The tasks from "function modeling" contain the actual parameters, the purpose and a unique identification of the data stream. Both modules support the synchronous mechanism of SMARTKOM. The actions controlled by the modules can be activated by external events, and the modules themselves can send events related to certain system states.

AudioIn Module: The AudioIn module receives the audio data from the input devices and puts it into chunks of defined formats and sizes. The format definitions are made by the function modeling module. So, for instance, in the operation modes "telephone" and "speaker verification" the data chunks contain 2-KB audio samples with 8-kHz sample rate — thus speech data of 256 ms. In the operation mode "normal" the chunks contain typically 8-KB data sampled with 16 kHz.

The audio data are sent in regular intervals in XML documents. The documents contain time stamps with the absolute time of the first sample of each data chunk. Thus the information of different modalities can be synchronized within the whole system.

In addition, the AudioIn module provides the functionalities "silence detection" and "echo compensation". In the case of silence, the transmission of audio data is going on but with empty data chunks. The echo compensation provided by partner DaimlerChrysler removes the SMARTKOM system messages via loudspeakers from the microphone signals.

AudioOut Module: The AudioOut module provides the output samples for the output devices at the given sample rates. In order to be able to synchronize the audio stream with other output streams like video data, the AudioOut module can start a certain audio output at an absolute time given in the audio document.

4.3 Standard Applications

4.3.1 Introduction

The interface for users of SMARTKOM is intuitive and offers many possibilities, as described in the previous sections. Every interaction with the system can lead to numerous complex activities involving devices and services especially designed and programmed for SMARTKOM. The ability to use standard IT applications within SMARTKOM is, however, also essential.

Groupware applications to manage e-mail, address books or calendars are typical examples of such services. Standard software offering these features has been integrated into SMARTKOM instead of duplicating their functionality within the system. Currently, the supported program suites are Lotus Domino Server and Lotus Notes, which include an application programming interface (API) capable of controlling all necessary groupware components. Other common software products can, however, also take their place.

4.3.2 Lotus Domino Server and Lotus Notes — A Short Overview

Lotus stores all data in objects which can be accessed using an API that exists for different programming languages. These databases, also known as *Notes Storage Facility* or *NSF files*, contain views and/or documents, which are the equivalent of, for example, an address book, an e-mail folder or a calendar. From a technical point of view the Java API that was used to integrate Lotus offers two possible ways of reading or manipulating Lotus data:

- The *local access pattern*, which allows the programmer to open NSF files without the server running in the background. The disadvantage is that the code has to run on the machine on which Lotus is installed.
- The *remote access pattern*, which hides a lot of tedious implementation details. Its drawback is that all methods are invoked through CORBA's *Internet Inter-ORB Protocol (IIOP)* interface, which means that processing is done remotely. Even if no traffic between different computers on the network is created, the bits have to travel through the network adapter of the local machine.

With performance being one of the main issues during the development of SMARTKOM, the local access pattern represented the natural choice for the implementation.

Apart from the possibility of retrieving data from a database or modifying it through its programming interface, Lotus offers standard e-mail protocols such as POP3, IMAP and SMTP, enabling users of the SMARTKOM system to communicate with the outside world.

4.3.3 Lotus Applications Modeled in SmartKom

To satisfy personal communication needs, the following Lotus applications were made available to SMARTKOM users:

- searching the public address book
- sending e-mails
- using the personal calendar

During the course of the multimodal dialogue with SMARTKOM, the system internally generates a plan to meet the demands of the user. The standard applications are employed to perform parts of the overall task through the abstract layer, which models the functionality of the different devices and services. For example, if the user wishes to send e-mail, the retrieval of the address as well as the task of sending the e-mail itself are building blocks of the entire process.

4.3.4 Storing SmartKom User Profiles

All the applications mentioned above make it necessary to log on to the Lotus Domino Server with a user name and a password. Because it is not acceptable to request this information after the user has been authenticated, it has to be stored

somewhere within the SMARTKOM system. Since all users who want to use the standard applications have to be registered on the Lotus Domino Server anyway, it makes sense to store the information relating to the person in the public address book. This also makes it possible to store other pieces of information about users, for example, their preferred method of authentication. Within SMARTKOM these parts of a user profile are only available to the administrator of the Lotus Domino Server through a metasearch in the public address book. When SMARTKOM is initialized, the login information of this metauser is read from the global configuration file. Returning to the example of an e-mail being sent by a SMARTKOM user, it now becomes clear that the standard applications are involved when restoring the user profile, authenticating the person and fulfilling the request of sending the e-mail.

4.3.5 The Import Tool for User Profiles and Address Book Entries

Once Lotus Domino Server and Lotus Notes have been installed, user profiles and contacts which should be accessible to SMARTKOM have to be created on the server. This task is handled by an import tool with a graphical user interface, which lets the SMARTKOM administrator add and edit the relevant data. The offline tool, which supports comma-separated value (CSV) files as input, connects to the server and makes all the desired entries in the address book.

4.3.6 The Interface to Standard Applications

Since all communication within SMARTKOM is carried out using the Multimodal Markup Language (M3L), which is based on XML, it was necessary to design an M3L interface and implement a middleware to pass on the functionality of the groupware applications to the SMARTKOM system. This interface accepts messages with the following content as input:

- an identifier for the data stream
- login information of the user
- the data necessary to complete the request

The data for the request are organized so they can be easily mapped to the data structures of the groupware product.

The output for all requests consists in the simplest case of the following control information:

- the identifier of the data stream that was processed
- the request status

Normally the request status is *finished* or *fail*. In the case that the interaction exceeds a certain time limit, it is set to *pending*. A further description of the output generated requires a consideration of two different kinds of request. First, there are messages that do not expect an answer from the server, apart from status information. Creating, changing or deleting calendar entries are requests of that type, as is

the sending of an e-mail. Under special circumstances it may also be necessary to return unique identification numbers for the data items processed on the server. These numbers represent groupware documents and are referred to as *note ID* or *universal ID* in Lotus (N.B. this is not the same as the identifier of the data stream). An example for the use of such an ID is a user utterance such as "please move the appointment to 7 pm," which leads to the modification of an entry in the calendar that was created earlier during a SMARTKOM session but is still held in the system. The second kind of request is characterized by the fact that it results in the generation of data output on top of the metainformation. The retrieval of contacts from the address book or entries from a personal calendar are requests of that type. These searches are modeled much like an SQL query: restrictions (the WHERE-clause) such as the last name of a person, and the desired answer format (the SELECT-clause) for the resulting data have to be specified. A typical request for example would be "I need John Doe's telephone number" (for the M3L code of this request, see the next section). For the calendar it is also possible to ask for appointments that take place during a certain period of time. This supports the processing of speech input such as "show me all my appointments between 4 and 6 pm".

4.3.7 Address Book Search: An Example

The following example shows a valid M3L representation of a request to the standard applications module for the telephone number of *John Doe* by a user that is registered on the Lotus Domino Server with the name *TestUser*. The format for the result of the query in the address book contains the telephone number and the first and last name of the persons found. Please note that for better legibility the header information has been deleted from all example messages in this section.

<externalApplicationRequest> \ *a request to*
<dataStream> \ *standard applications*
<name> \
abook_1\
</name> \
<purpose> \
queryAddressBook\
</purpose> \
</dataStream> \
<smartkomUserID> \ *user information*
TestUser\
</smartkomUserID> \
<action> \
<addressBookSearch> \ *the type of the request*
<personAnswerFormat> \ *the desired answer format*
<name> \ *(SELECT part of query)*
<firstName> \
</firstName> \
<lastName> \

```
</lastName> \
</name> \
<contact> \
<telephoneNumber> \
</telephoneNumber> \
</contact> \
</personAnswerFormat> \
<searchPerson> \ the restrictions
<name> \ (WHERE part of query)
<firstName> \
John\ look for 'John Doe'
</firstName> \
<lastName> \
Doe\
</lastName> \
</name> \
</searchPerson> \
</addressBookSearch> \
</action> \
</externalApplicationRequest> \
```

The answer containing the data matching the request is the following (the request status information, which is also produced as output, has been omitted for the sake of clarity):

```
<externalApplication> \ data from standard applications
<dataStream> \
<name> \
abook_1\
</name> \
<purpose> \
queryAddressBook\
</purpose> \
</dataStream> \
<addressBook> \ the type of the data returned
<smartkomUserID> \ user information
TestUser\
</smartkomUserID> \
<person> \ all address book entries
<name> \ matching the restrictions
<firstName> \
John\
</firstName> \
<lastName> \
Doe\
</lastName> \
</name> \
```

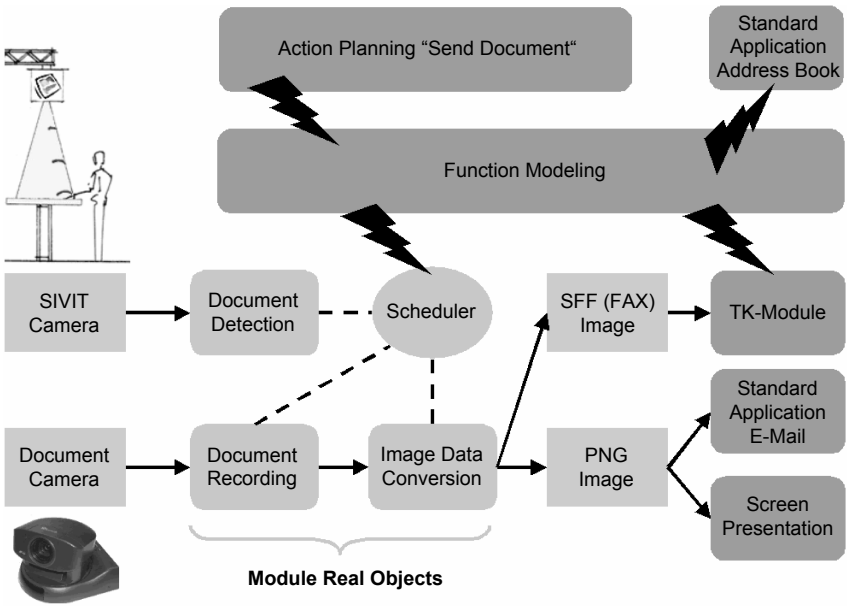

Fig. 7. SMARTKOM-Public fax and e-mail functionalities

```
<contact> \
<telephoneNumber> \
16508451000\
</telephoneNumber> \
</contact> \
</person> \
</addressBook> \
</externalApplication> \
```

4.3.8 Summary: Standard Applications in SmartKom

All standard applications that are available to SMARTKOM users are provided by a middleware that offers an M3L interface and models certain functionalities of a groupware product such as the Lotus Domino suite. Single tasks, like the retrieval of an e-mail address, are small building blocks of plans that are generated based on a multimodal dialogue between the SMARTKOM user and the system.

The standard applications are used to store pieces of information about SMARTKOM users, for example, their preferred method of authentication, whether voice, signature or hand. The different tasks generate a flow of information that is separated into a data and a control flow conforming to the architecture of SMARTKOM. Unique IDs enable back-referencing to data objects already held in the system.

4.4 Composed Applications: Fax and E-Mail

Among the most complex applications in the SMARTKOM system are the fax and e-mail functionalities. A lot of SMARTKOM modules are involved to produce the desired results. Although the internal operating sequence is complex, the user gets the impression of an easy to use functionality. But now we go into the details. Let us start with the previous Fax example user utterance "I'd like to send a fax to Mr. Wahlster." Let us also assume that the action planning module of SMARTKOM already understood the sense of the user's utterance and sent the order "send document to Wahlster fax" to the function modeling module, which coordinates the next steps. What is happening internally?

1. The ISDN fax number of Mr. Wahlster has to be determined.
   (Standard Application Address Book)
2. The Screen presentation "Document scanning" is shown.
   (Module Display)
3. SMARTKOM waits for a document on the interaction area.
   (Module Real Objects analyzes SiViT gesture camera input)
4. The high-resolution image is captured by the document camera.
   (Module Real Objects performs Panoramic Imaging)
5. The image is converted into a display format and rendered on the screen.
   (Module Real Objects and Module Display)
6. The image is converted into a SFF black and white Fax image.
   (Module Real Objects)
7. Image generated in step 7 is sent to the Fax number determined in step 1.
   (TK-Module).

If the user wants to send an e-mail the steps are slightly different: instead of the ISDN fax number the e-mail address is determined, and instead of the black and white image a PNG color image is generated and sent as an attachment to the selected e-mail address.

References

C. Frank, J. Adelhardt, A. Batliner, E. Nöth, R.P. Shi, V. Zeißler, and H. Niemann. The Facial Expression Module, 2006. In this volume.

S. Grashey and M. Schuster. Multiple Biometrics, 2006. In this volume.

C. Maggioni and H. Röttger. Virtual Touchscreen – A Novel User Interface Made of Light — Principles, Metaphors and Experiences. In: *Proc. 8th Intl. Conf. on Human-Computer Interaction*, pp. 301–305, Munich, Germany, 1999.

J. Racky, M. Lützeler, and H. Röttger. The Sense of Vision: Gestures and Real Objects, 2006. In this volume.

W. Wahlster, N. Reithinger, and A. Blocher. SmartKom: Multimodal Communication with a Life-like Character. In: *Proc. EUROSPEECH-01*, vol. 3, pp. 1547–1550, Aalborg, Denmark, September 2001.