



# Adaptive and Flexible Multimodal Access to Multiple Applications

Norbert Reithinger

Jan Alexandersson, Tilman Becker, Anselm Blocher,  
Ralf Engel, Markus Löckelt, Jochen Müller, Norbert Pfleger,  
Peter Poller, Michael Streit, Valentin Tschernomas

[sk-info@dfki.de](mailto:sk-info@dfki.de)



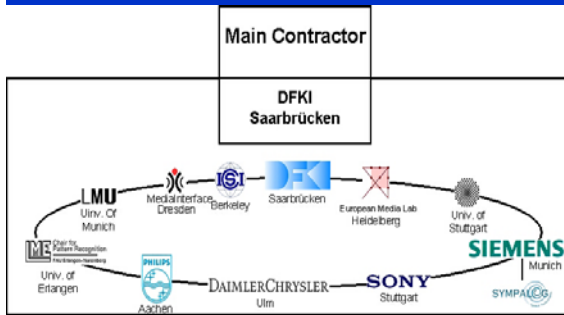
## Overview



- SmartKom's goals and functionalities
- Tour through the dialog backbone
  - Modality fusion
  - Discourse processing
  - Action planning
  - Dynamic help
  - Modality fission and presentation

# The SmartKom Project

- **Research goal:** Uniform multimodal dialog interface to applications in three scenarios (Home, Public, Mobile)
- **Hypothesis:** Multimodal interaction facilitates interaction
- **Task:** Design and implement the dialog backbone
- **Framework:** German Human Computer Interaction Program (1999 – 2003)



- 12 project partners
- Budget: ca. 25 Mio €
- Project partly funded by the German Ministry of Education and Research (BMBF)

06.11.2003 ICMI-PUI 2003/NR

3

# Interaction in SmartKom

- Input (user independent):
  - Speech
  - Gestures
  - Facial expressions
- Output:
  - Intuitive presentation, addressable through gestures and speech
  - Anthropomorphic interaction agent *Smartakus* uses
    - Speech
    - Gestures
  - Agent provides feedback about processing state

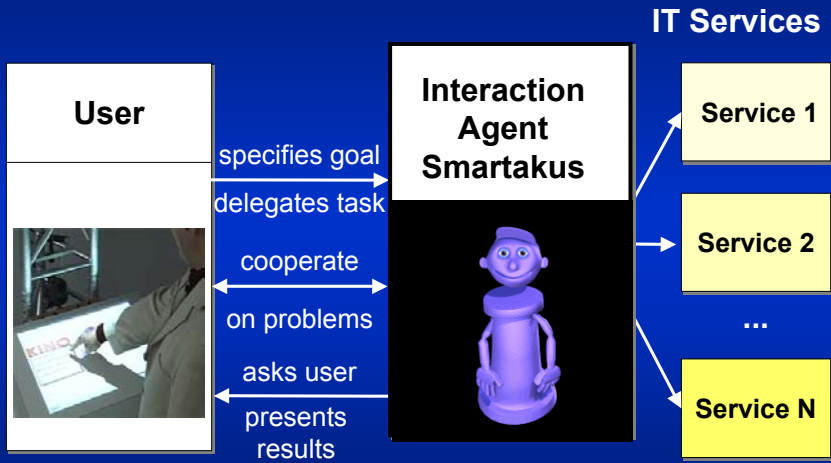
06.11.2003 ICMI-PUI 2003/NR

4

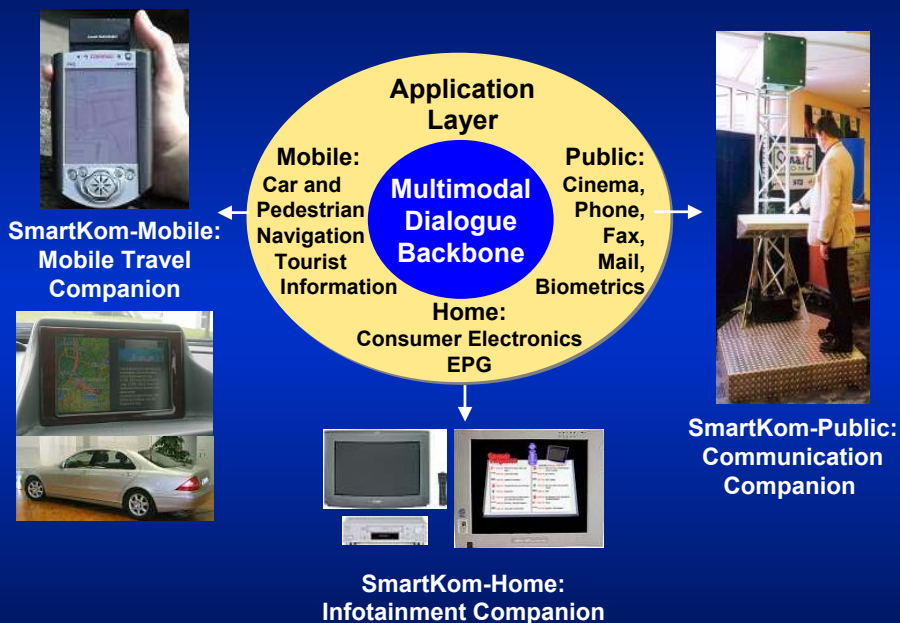
# SmartKom's Interaction Metaphor



Situated Delegation-oriented Dialogue Paradigm



# SmartKom: The Three Scenarios



# Applications and Functionalities

Home			
EPG (Electronic-Programming Guide)	General program	Information for one broadcast	
	Channel selection	Time-based operations	
TV	Channel information	Help functions for genres	7
	Selection based on genre		
VCR control	On/off	Channel selection	2
	Record	Wind/rewind	
	Play	Programming using EPG and the calendar	
	Pause		6
Lean-Forward/ Lean-Backward	Select Lean-Backward	Context aware presentations	3
	Deactivate Lean-Backward		
<b>Total Home</b>			<b>18</b>
Public			
Telephone	Manipulative key operations	Audio handling	
	Telephony functions	Address book	4
Hand contour biometry	Selection of biometry type	Presentation and camera control	
	Hand biometry	Address book (see above)	3
Voice biometry	Presentation and audio control	Address book (see above)	
	Voice biometry	Selection of biometry type (see above)	2
Signature biometry	Presentation and tablet control	Address book (see above)	
	Signature biometry	Selection of biometry type (see above)	2
Fax	Presentation and interaction	Address book (see above)	
	Fax handling	Camera control	3
E-Mail	Presentation and interaction	Address book (see above)	
	E-Mail handling	Camera control (see above)	2
Cinema	General program	Seat reservation	4
	Movie information	Cinema location	
<b>Total Public</b>			<b>20</b>
Mobile			
Car navigation	Selection of start and goal city	Selection of parking garage	
	Route type selection	Information about parking garages	5
Pedestrian navigation	Car route computation		
	Selection of map type	Selection of points of interest	
	Selection of start and goal	Information for points of interest	
Map manipulation	Route computation	Integrated car and pedestrian route planning	6
	Resize	Change viewpoint	
	Help functions for map interactions		3
<b>Total Mobile</b>			<b>14</b>
<b>Total System</b>			<b>52</b>

06.11.2003 ICMI-PUI 2003/NR

## Design Principles

### 1. General solutions for general problems:

- Coherent use of a knowledge based approach
- No shortcuts or application specific procedural dialog processing steps

### 2. No processing and presentation without representation:

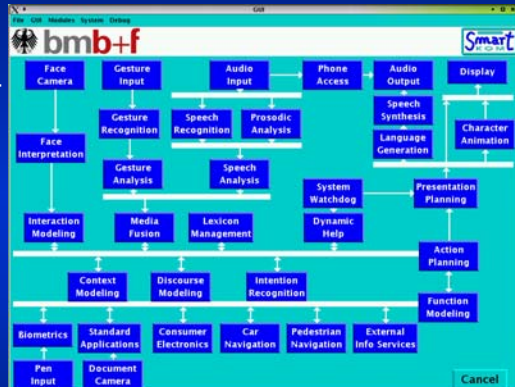
- Common representation approach for all multimodal inputs and outputs (MultiModal Markup Language M3L, set of XML Schemata)
- Ontology defined off-line with OIL, representing everything the user and the system can talk about (approx. 700 concepts and 200 relations)
- Automatically transformed into an M3L XML Schema

06.11.2003 ICMI-PUI 2003/NR

8

# Software Integration Platform: Multiplatform

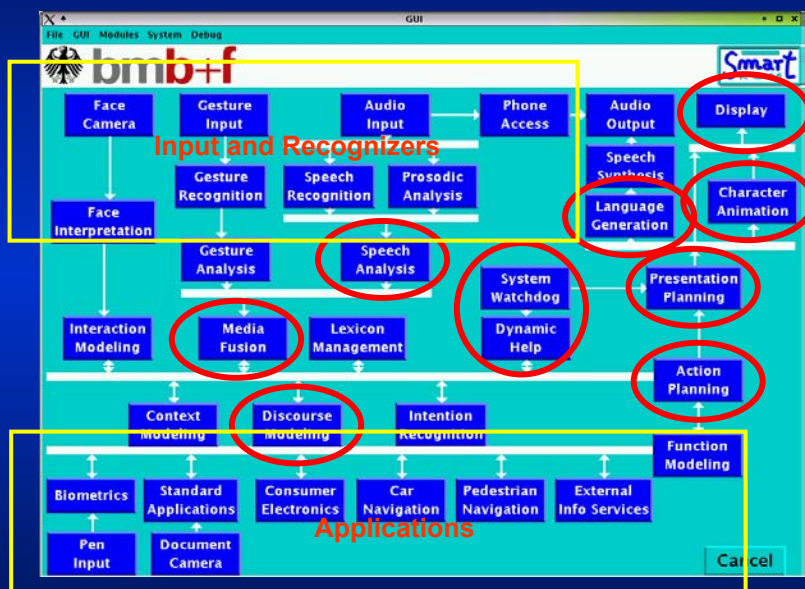
- Open, flexible, and scalable software architecture (Linux & Windows)
- High performance blackboard-like publish/subscribe messaging based on PVM/PCA
- Data-centric interfaces for module interaction
- Lean API to connect modules for C, C++, Java and Prolog
- Elaborated infrastructure for logging and visualizing processing
- Predecessor already used in Verbmobil



06.11.2003 ICMI-PUI 2003/NR

9

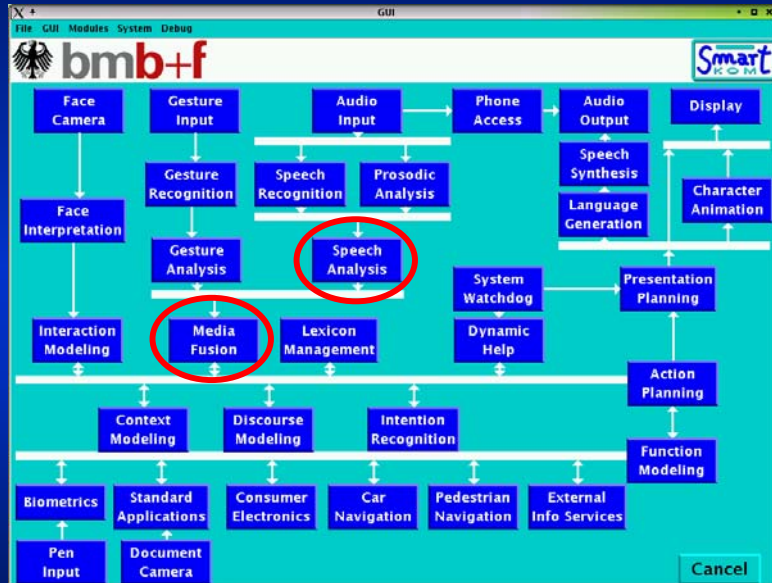
## The Tour Through the Dialog Backbone



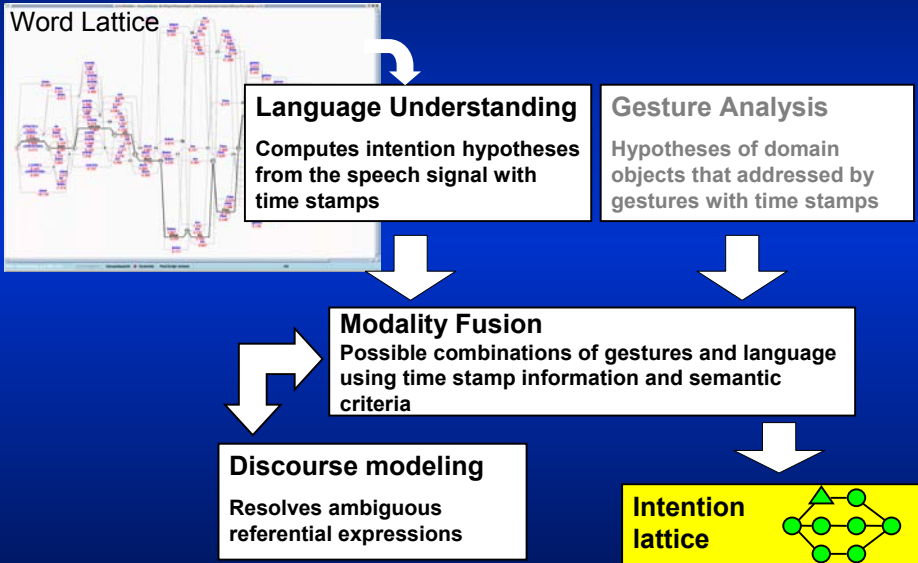
06.11.2003 ICMI-PUI 2003/NR

11

# Speech Analysis and Modality Fusion



# Data Flow In the Analysis Modules



# Natural Language Understanding



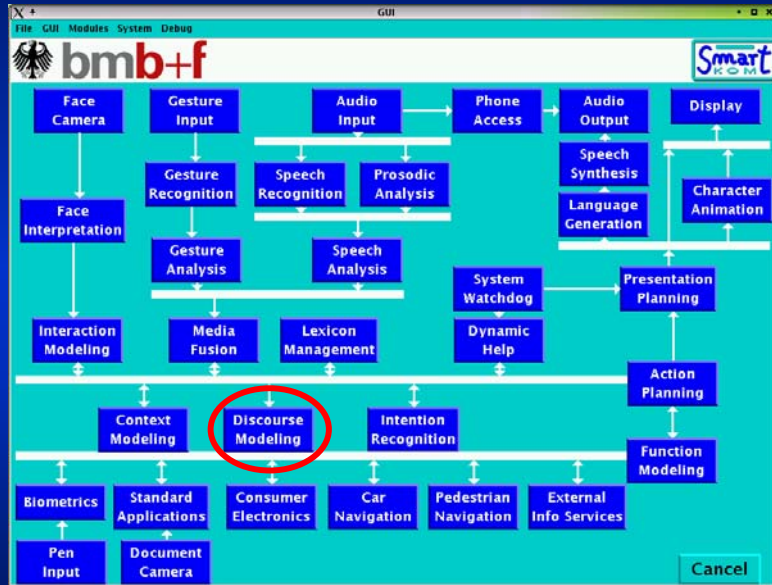
- Template based semantic parser (SPIN)
- Set based matching
  - Simplifies processing of free word order languages
  - Increases robustness significantly
  - Semantic based segmentation included
- Powerful template language (Type 0)
  - Allows more elegant and compact templates
- Predefined template application order
  - Determined automatically offline (dependency graph)
  - Provide high performance

# Modality Fusion



- Slot filling approach
- Combines analyzed speech and gestures and generates all possible solutions
- Scoring function used to determine n-best solutions
- Approach supports
  - Multiple referring expressions and gestures in one utterance
  - Selection of multiple objects (encircling gestures)
  - Processing of ambiguous or underspecified referring expressions and gestures

# Discourse Processing

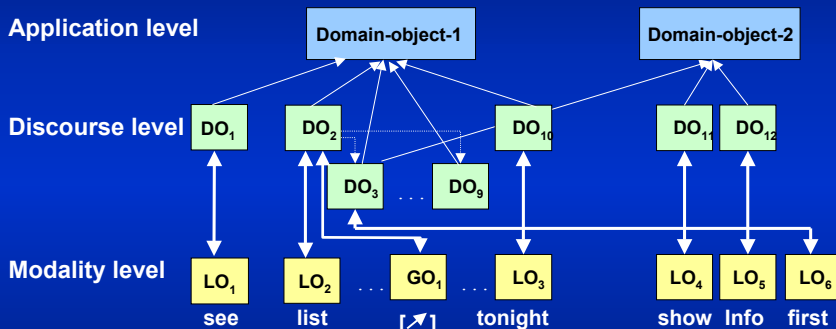


06.11.2003 ICMI-PUI 2003/NR

16

# Multimodal Discourse Memory

Three level representation with global and local focus



System: Here [↗] you see a list of movies that are on TV tonight

User: Please show me infos for the first

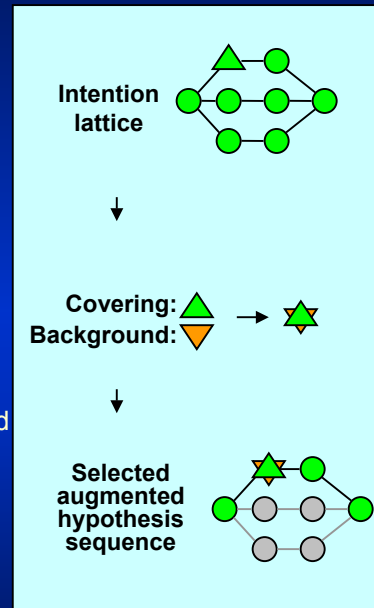
DO = Discourse Object, LO = Linguistic Object  
GO = Gesture Object

06.11.2003 ICMI-PUI 2003/NR

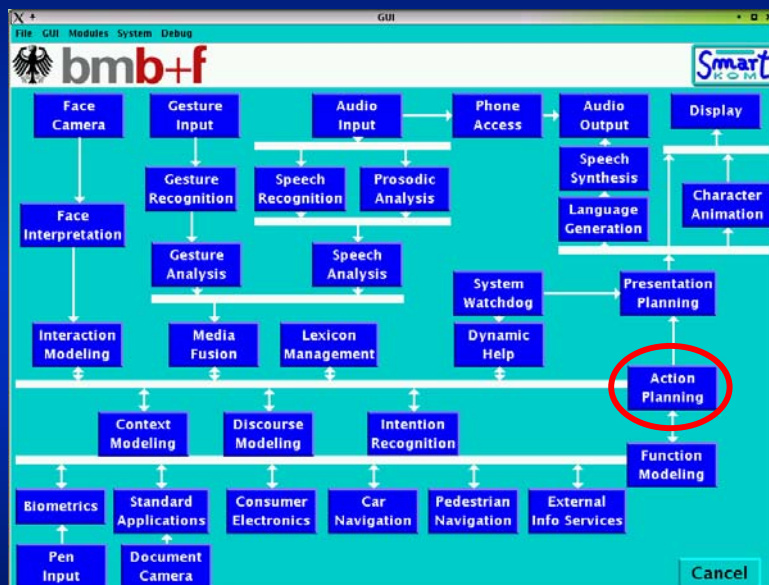
17

# Discourse Processing

- Lattice of intention hypothesis sequences from Modality Fusion
- Enrichment and Validation
  - Compare and enrich with a selected number of discourse states:
    - Fill in consistent information
    - Compute a score
  - For each hypothesis - background pair:
    - Overlay (covering, background)
    - Score fit between covering/background
- Scores of overlay, speech interpretation etc. give overall score
- Augmented hypothesis sequence with best score is passed to Action Planning



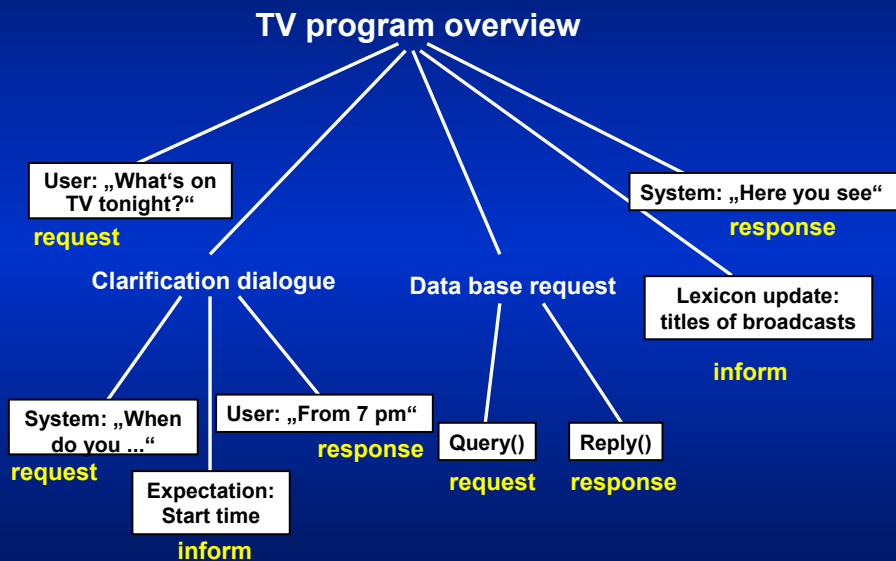
# Action Planning



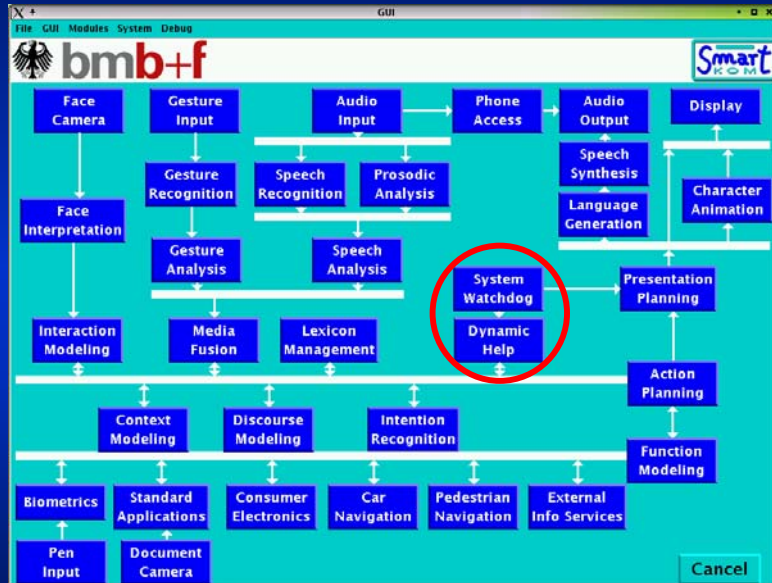
# Action Planning

- Requirements:
  - Maintain a meaningful dialog with the user
  - Address multiple interdependent applications and configurations
  - Enables synchronous or asynchronous communication
- Method: Backward-chaining planner
  - Plan operators define application functionalities (based on the ontology)
  - Performs communicative games with all actors in the system
  - Provides dialog expectations to discourse and lexicon modeller
- Open architecture to support dynamic extension and reuse

# Example Communicative Game



# Dynamic Help And System Watchdog



06.11.2003 ICMI-PUI 2003/NR

22

## Supervising the Processing – WatchDog

- Supervises the processing flow – reflection on the system's internal state
- Infers processing states
- Informs Dynamic Help
- Initiates feedback activities of Smartakus
- Visualizations:



Understood

Computing

06.11.2003 ICMI-PUI 2003/NR

23

# Dynamic Help



Recognizes and handles problematic events and meta-request (rule based)

## Some elicitors for help:

- Underspecified input
- Explicit help requests
- Negative/positive feedback
- Emotions (exemplary)
- Processing breakdown

## Reaction types:

- Presentation of input options
- Query relaxation/repetition
- Error messages
- Preference adjustment

# Dynamic Help



- Problem specific methods and knowledge sources e.g.
  - Type completion with restrictions
  - Relaxation rules
- Subdialog generator specifies multimodal help interactions

## Query repetition

Ich habe folgendes verstanden:

- Ort
- Name: Kornmarkt

I understood:

- Place
- Name: Grain Market

## Presentation of options for actions

Stadtplan

Fußweg

Autokarte

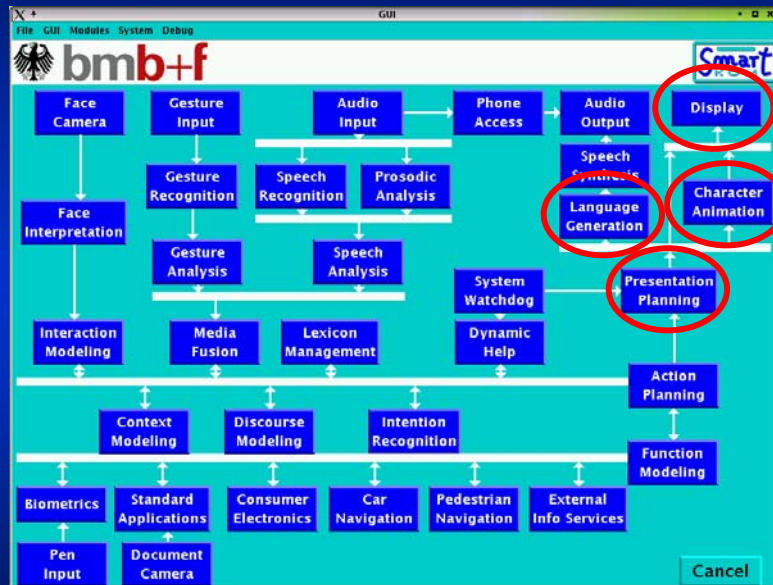
Autoroute

city map  
pedestrian route  
car map  
car route

## Presentation of domain options (e.g. TV)

Sendungsformate	Inhaltsmerkmale
Nachrichten	Action
Magazin	Krimi
Sportler	Science Fiction
Serie	Mehr zu Action
Show	Sport
Geeignet für	Fußball
Jugendliche	Mehr zu Sport
Kinder	Musik
	Charts
	Mehr zu Musik

# The Presentation Modules



# Presentation Requirements

- Uniform look and feel across scenarios
- Modality availability
  - User preference
  - Scenario restrictions
- Agent animation
  - Gestures
  - Posture
  - Movements
  - Lip synchronisation
- Scenario dependent output devices

Projection surface (Public) PDA (Mobile)

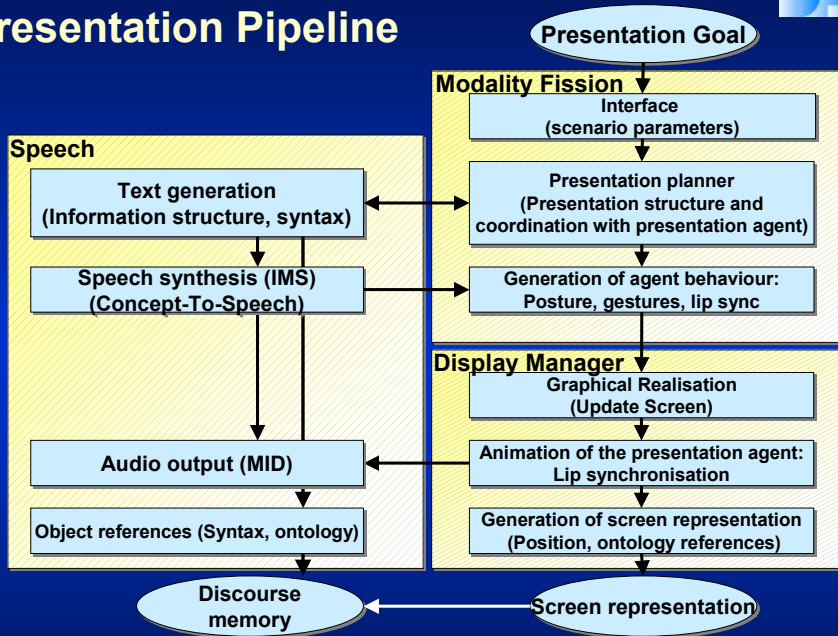


Webpad (Home)

Car display (Mobile)



# Presentation Pipeline



# Modality Adapted Presentation



Lean forward: the user pays attention to the graphical presentation

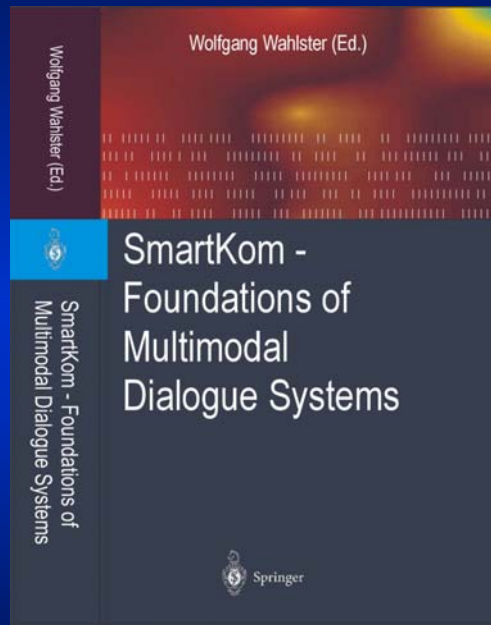
Lean backward: the user e.g. watches TV

For example, these broadcasts start at 6 pm: SOKO 5513 on ZDF, Guten Abend RTL on RTL, Lenßen und Partner on SAT1





To appear in Spring 2004  
Approx. 490 pages



## Thank you very much for your attention!

- Project information (including videos of the system) at <http://www.smartkom.org>
- Projects reusing SmartKom's technology
  - MIAMM (EU): Multidimensional Information Access using Multiple Modalities: <http://www.miamm.org>
  - COMIC (EU): COntersational Multimodal Interaction with Computers: <http://www.hcrc.ed.ac.uk/comic>
  - VirtualHuman (BMBF): Communicating virtual agents: <http://www.virtual-human.org>