

Multimodal Information Access Using Speech and Gestures

Norbert Reithinger

norbert.reithinger@dfki.de



Overview



- Natural access to digital resources using multimodality
- SmartKom: Intuitive multimodal interaction
- Generic technologies and language (in-)dependence

Natural Access to Digital Resources

- Speech dialog systems already ready-to-market
- Next step: multimodal interfaces
- Add modalities beyond speech – Enable interactions using
 - Gestures
 - Pointing
 - Haptics...

Advantages of Multimodality for the Users



- Natural use of preferred modalities
- Mutual disambiguation using multiple modalities under adverse conditions
- Environment may elicit preferred interaction modality
- Product realization e.g. as multimodal information kiosk
- Example: SmartKom

SmartKom: Intuitive Multimodal Interaction



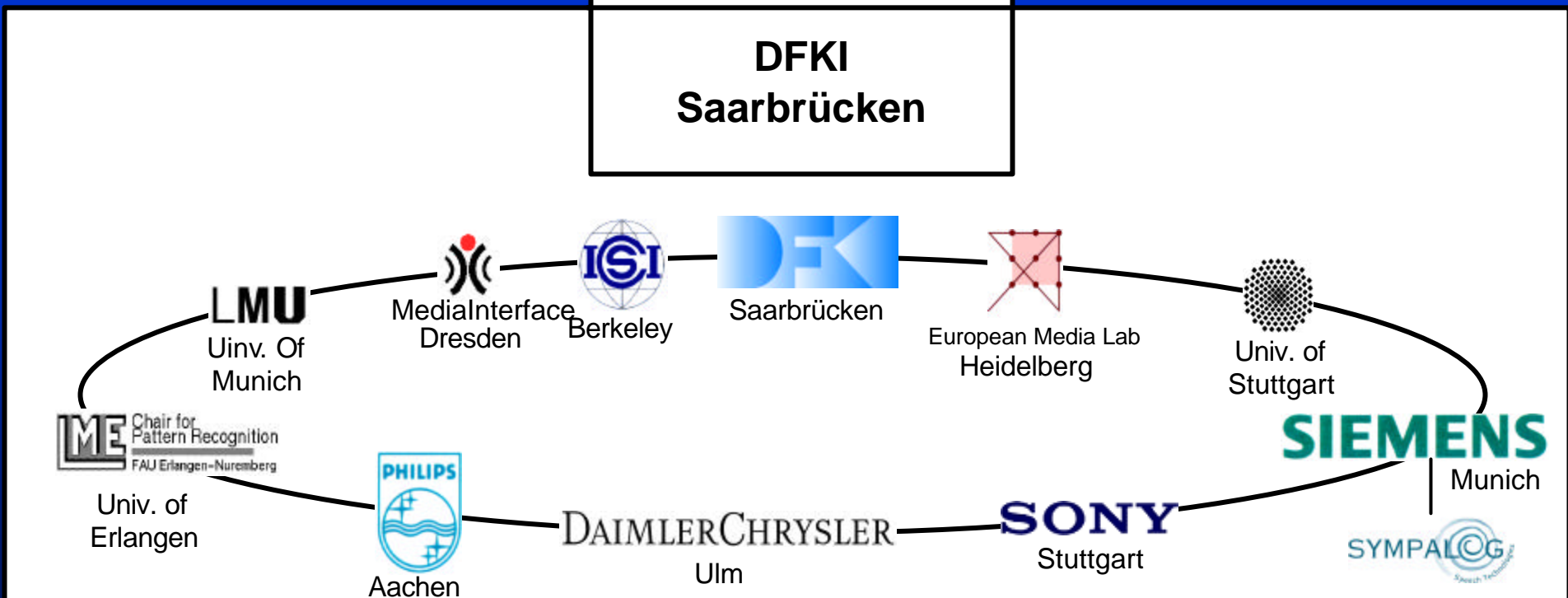
Project Budget: € 25.5 million (partly funded by the German Ministry of Education and Research - BMBF)

Project Duration: 4 years (September 1999 – September 2003)

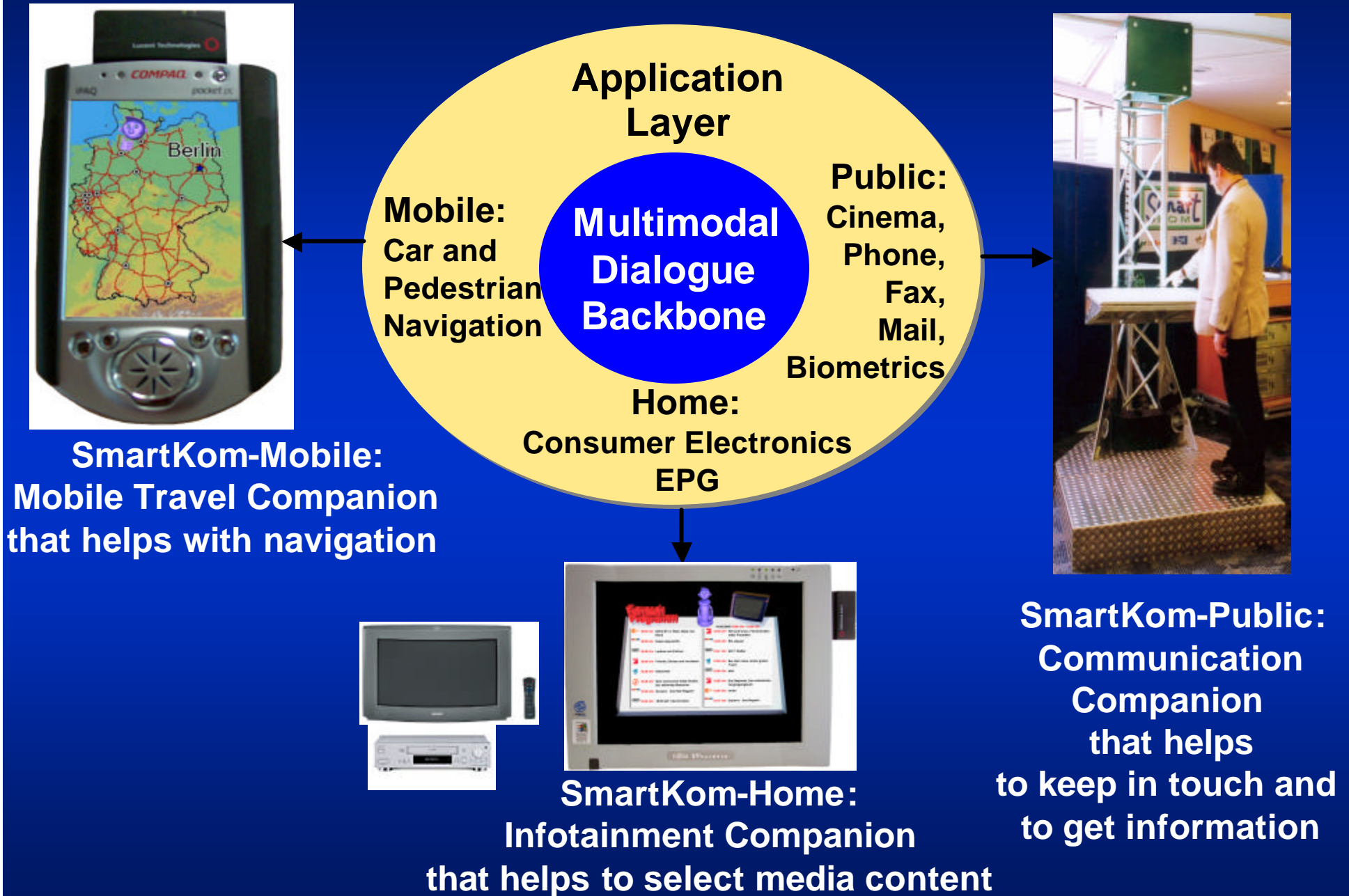
The SmartKom Consortium:

Main Contractor

DFKI
Saarbrücken



SmartKom: The Three Scenarios



SmartKom's Multimodal Input and Output Devices



Multimodal Control of TV-Set

Multimodal Control of VCR/DVD Player

3 dual Xeon 2.8 Ghz processors with 1.5 GB main memory



Infrared Camera for Gestural Input, Tilting CCD Camera for Scanning, Video Projector

Microphone

Camera for Facial Analysis

Projection Surface

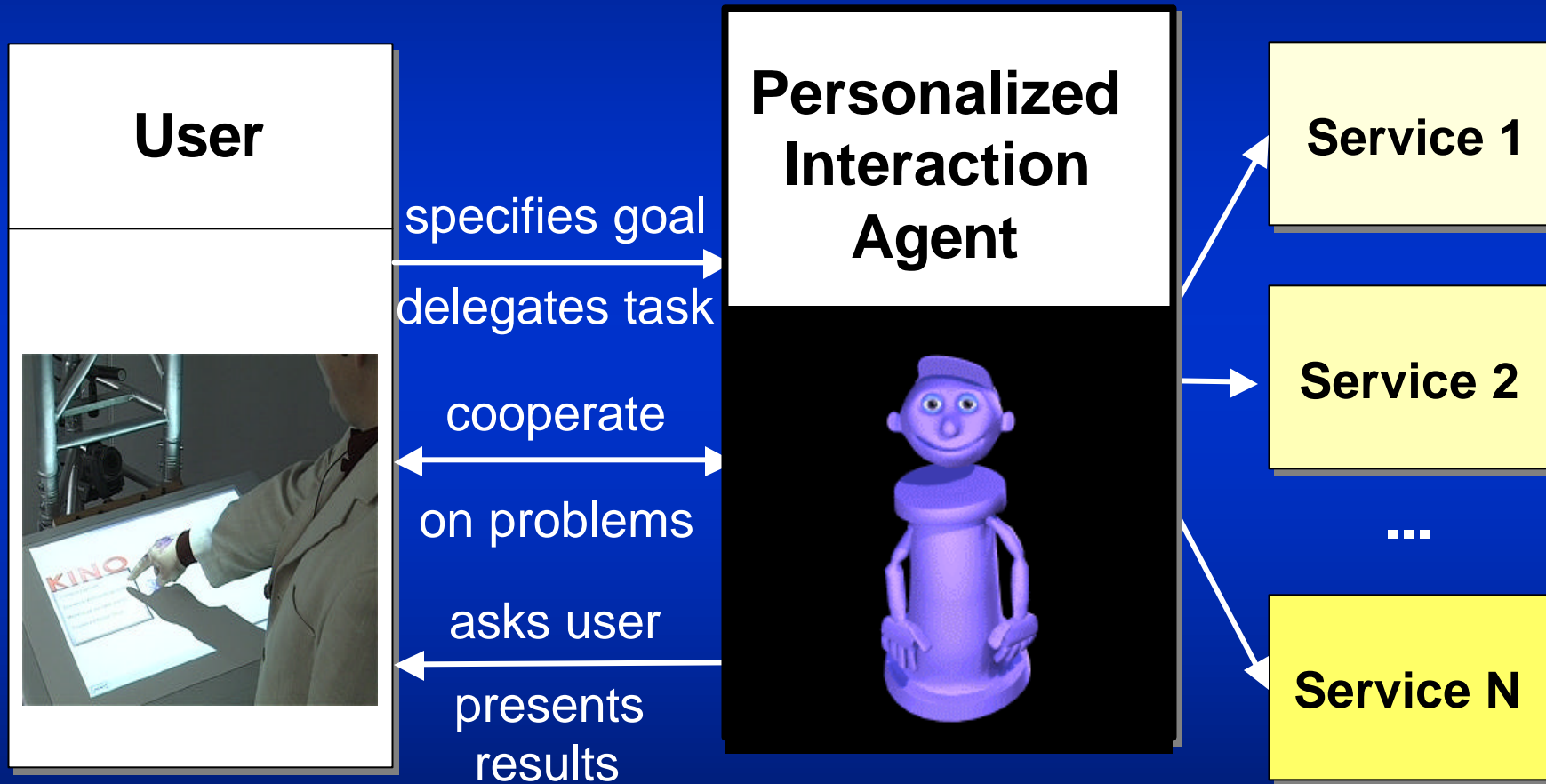
Speakers for Speech Output

SmartKom`s SDDP Interaction Metaphor

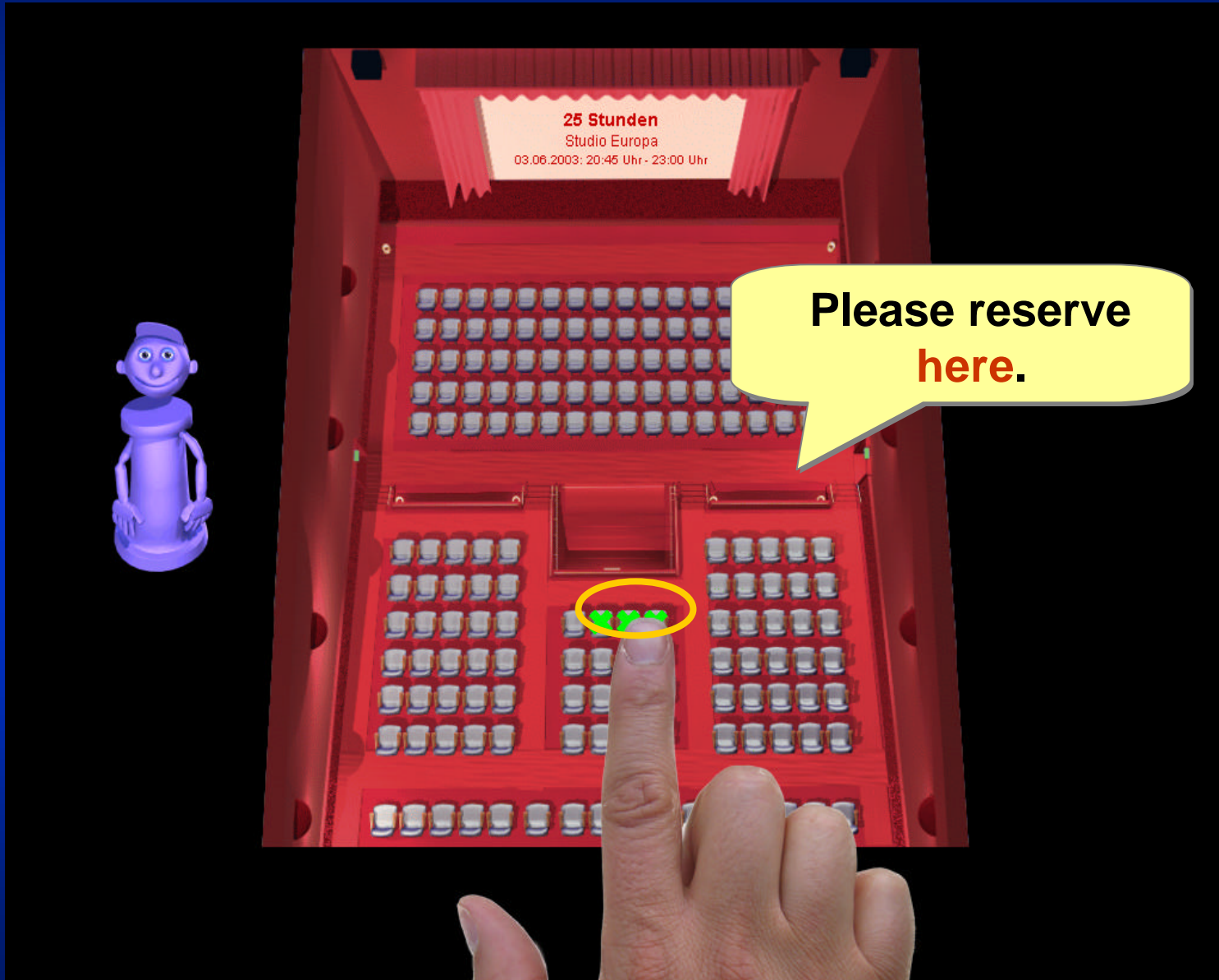


SDDP = Situated Delegation-oriented Dialogue Paradigm

IT Services



SmartKom Understands Multimodal Input



SmartKom:

14 applications
52 functionalities

(Source: Reithinger et al:
SmartKom - Adaptive and Flexible
Multimodal Access to Multiple
Applications. In ICMI '03)



Home			
EPG (Electronic-Programming Guide)	General program	Information for one broadcast	
	Channel selection	Time-based operations	
	Channel information	Help functions for genres	
	Selection based on genre		7
TV	On/off	Channel selection	2
VCR control	On/off	Wind/rewind	
	Record	Programming using EPG and the calendar	
	Play		6
	Pause		
Lean-Forward/ Lean Backward	Select Lean-Backward	Context aware presentations	
	Deactivate Lean-Backward		3
Total Home			18
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	
	Movie information	Cinema location	4
Total Public			20
Mobile			
Car navigation	Selection of start und goal city	Selection of parking garage	
	Route type selection	Information about parking garages	
	Car route computation		5
Pedestrian navigation	Selection of map type	Selection of points of interest	
	Selection of start und goal	Information for points of interest	
	Route computation	Integrated car and pedestrian route planning	6
Map manipulation	Resize	Change viewpoint	
	Help functions for map interactions		3
Total Mobile			14
Total System			52

Generic Technologies Used



- Speech and gesture recognition
- Language and gesture understanding
- Modality fusion
- Dialog processing
- Information extraction/retrieval, e.g. from Internet sources
- Biometry
- Presentation planning
- Answer generation
- Speech synthesis
- Interactive presentation

Interactive Biometric Authentication by Hand Contour Recognition



Please place
your hand with
spread fingers on
the marked area.

Person

Name:
Gerd Herzog

Telefon:
06813025386

Fax:
06813025020

E-Mail:
herzog@dfki.de

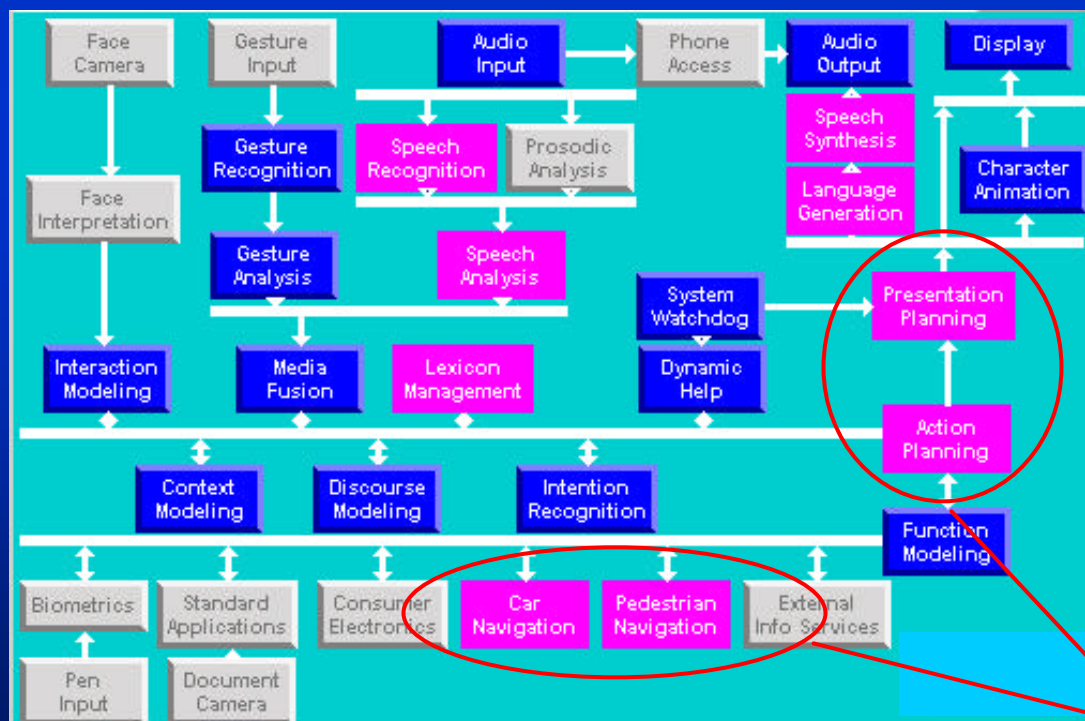


Adaptation to Another Language: SmartKom Mobile English



SmartKom's modular architecture encapsulates language specific knowledge in few language processing modules

SmartKom system overview:



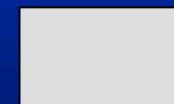
Module was ...



Modified



Not modified



Not used

Minor modifications

Conclusion

- Multimodal interaction enables natural access to digital resources
- Advantageous for many users
- SmartKom realizes an exemplary multimodal information kiosk
- Adaptation to different languages relatively easy

Thank you very much for your attention!

- Please find more information at <http://www.smartkom.org>
- Other multimodal projects with participation of DFKI
 - MIAMM (EU): Multidimensional Information Access using Multiple Modalities: <http://www.miamm.org>
 - COMIC (EU): COnversational Multimodal Interaction with Computers: <http://www.hcrc.ed.ac.uk/comic>
 - VirtualHuman (BMBF): Virtual agents for education <http://www.virtual-human.org>