

IDEALAB!
WHU Vallendar, 9.11.2001



Multimodal Dialogue Processing

Tilman Becker
Norbert Reithinger



Deutsches Forschungszentrum für Künstliche Intelligenz GmbH
Stuhlsatzenhausweg 3, Geb. 43.1 - 66123 Saarbrücken
Tel.: (0681) 302-5271/5346
Email: {becker,bert}@dfki.de
www.smartkom.org

Overview

- **Introduction**
- **Wizard of Oz Experiments**
- **Architectures for Multi-Modal Systems**
- **Mensch-Technik Interaktion Lead Projects**
- **SmartKom as an Example System**
- **DARPA Communicator: Infrastructure for Dialogue Interaction**
- **Interface Languages: XML and beyond**
- **Structure of Dialogs**
- **Research Roadmap for Multimodality**

Introduction

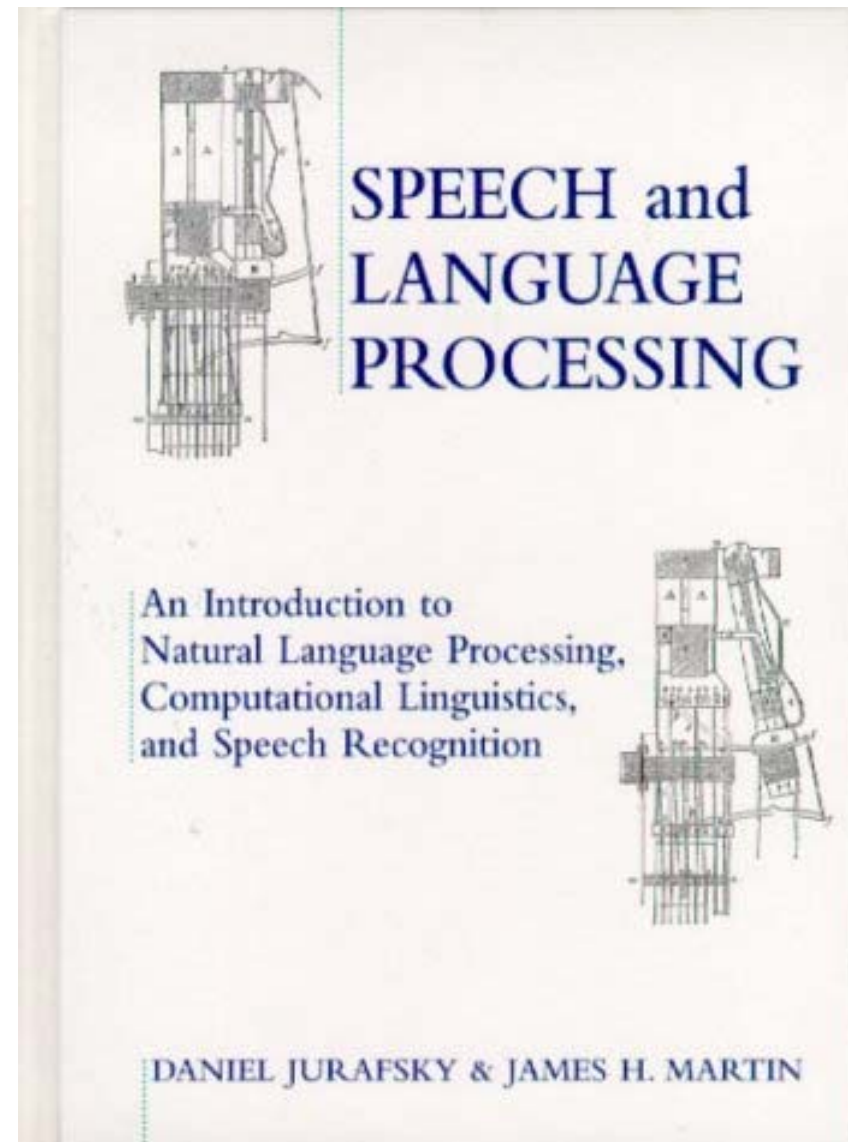
- **Speech dialog systems already ready-to-market**
- **Next step: multimodal interfaces**
 - Add modalities beyond speech
 - Enable interaction using gestures, pointing, haptics...
 - Dialog with one (or more) virtual agents
 - Delegation of tasks to the agent
 - Realization e.g. on PDAs, UMTS phones
- **This talk presents an overview of**
 - architectures
 - components
 - methods

The Current Standard Text Book

Daniel Jurafsky and James H. Martin

**Speech and Language Processing:
An Introduction to Natural Language
Processing, Computational
Linguistics and Speech Recognition**

US-\$ 66.00



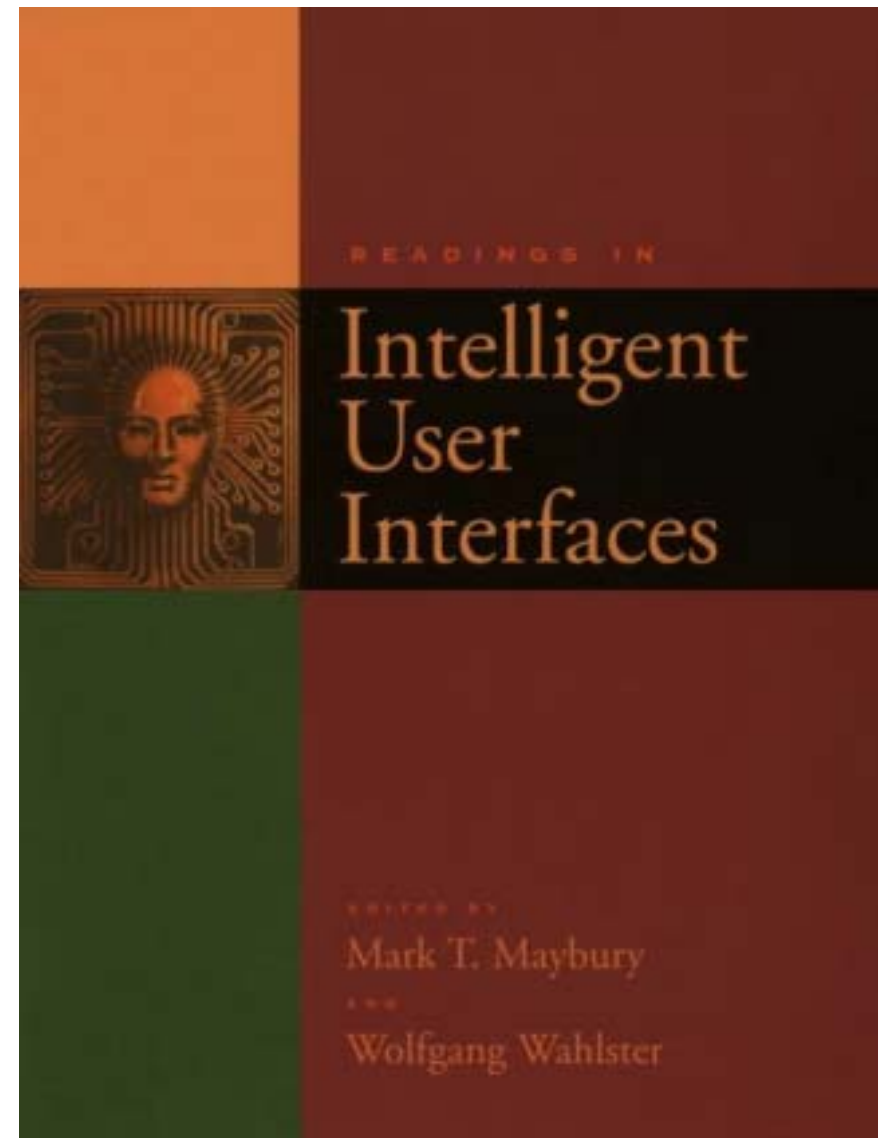
More About Intelligent User Interfaces



Mark T. Maybury and Wolfgang
Wahlster (Editors)

Readings in Intelligent User Interfaces

US-\$ 69.00

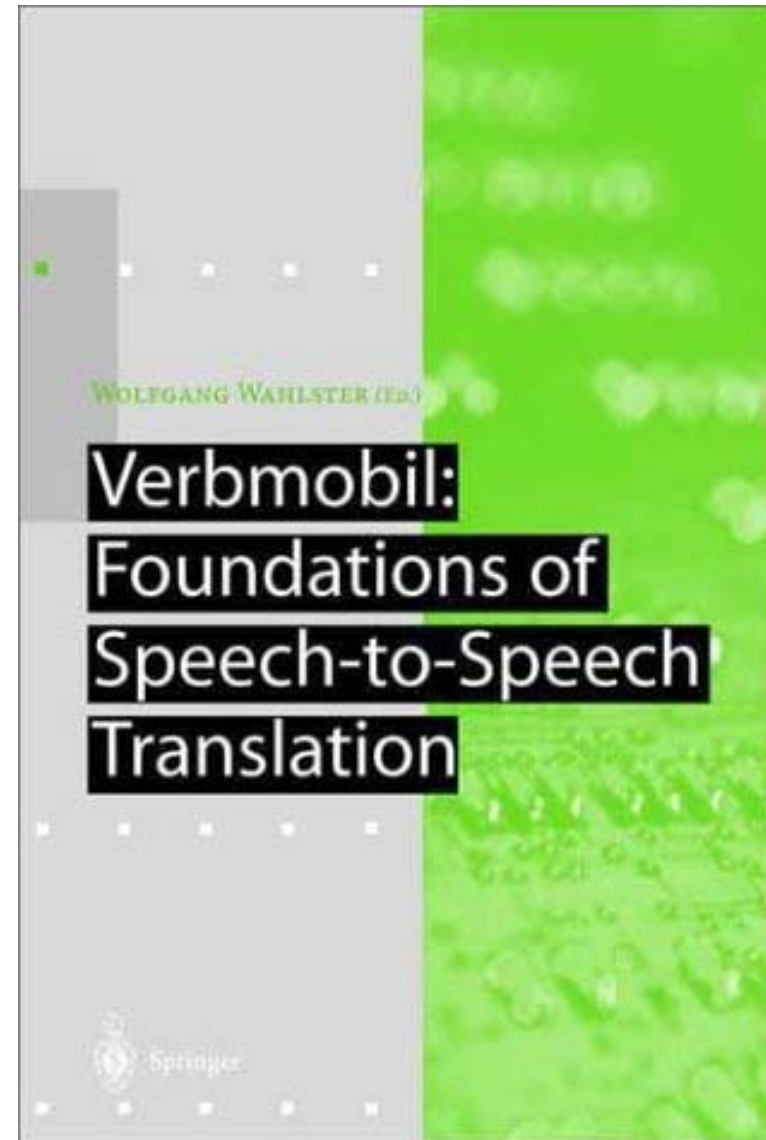


More About Speech-To-Speech Translation

Wolfgang Wahlster (Editors)

Verbmobil : Foundations of Speech-To-Speech Translation

US-\$ \$69.95



Before starting: Wizard of Oz Experiments



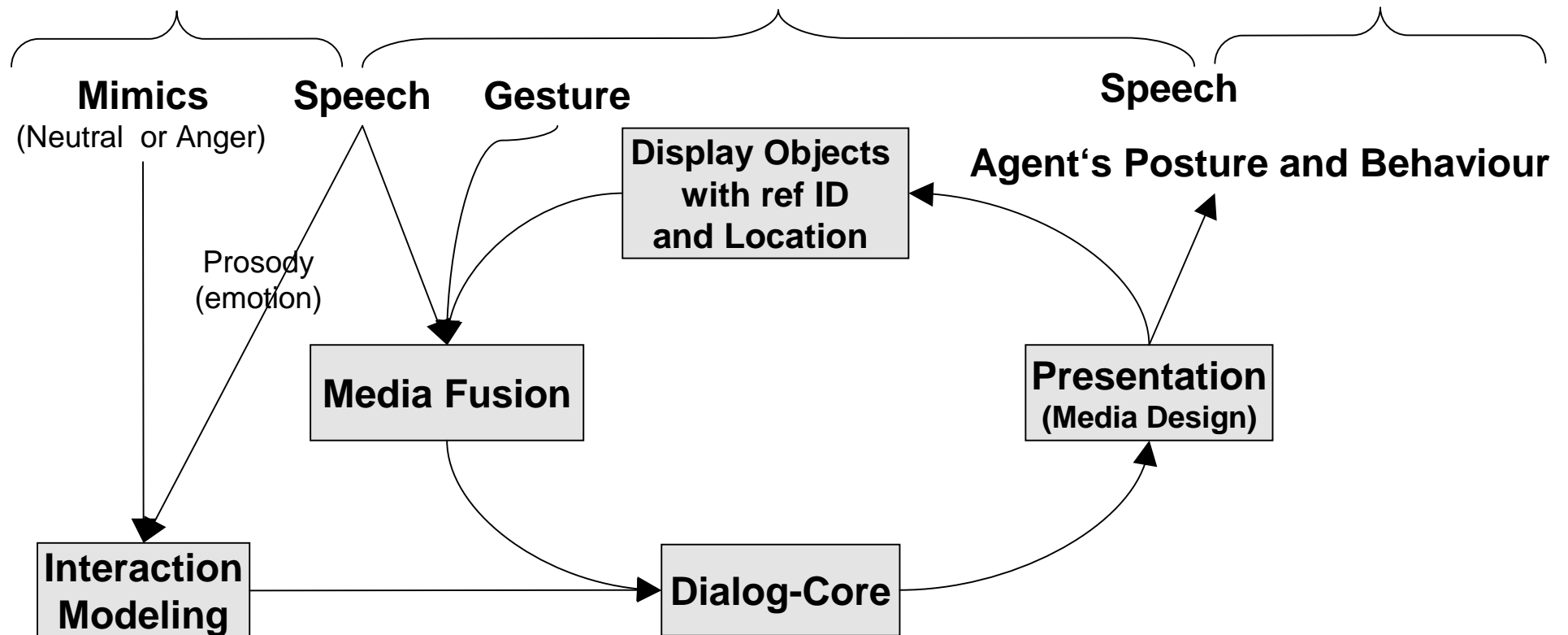
- **Simulate the envisioned system**
- **Method:**
 - Define envisioned use cases
 - Set up an experimental mock-up of the system
 - Test with naive subjects
 - Record the sessions
- **Necessary to test, verify, and modify the system's appearance, behaviour, and usability**
- **Side effect: data collection for training and test of system components**

Media Processing: The Data Flow

User State

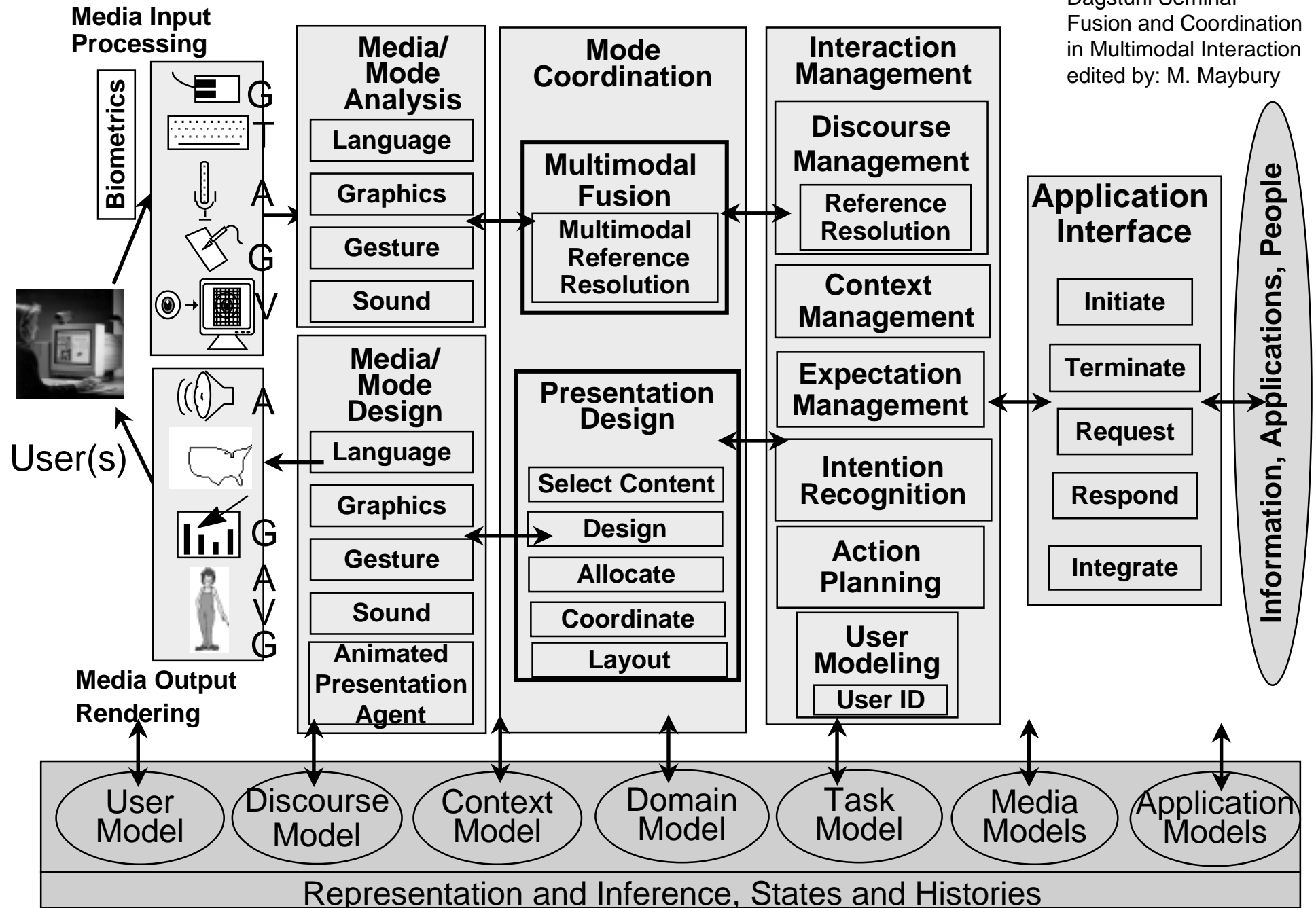
Domain Information

System State



Reference Architecture for Multimodal Systems

2 Nov. 2001
 Dagstuhl Seminar
 Fusion and Coordination
 in Multimodal Interaction
 edited by: M. Maybury



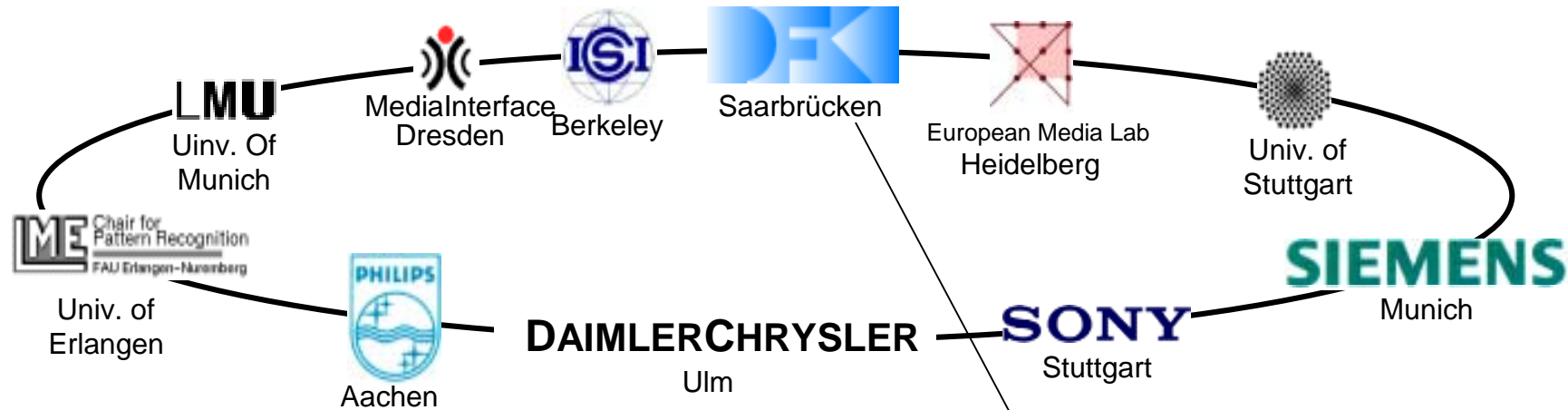
Mensch-Technik Lead Projects

Project	Topic	Coordinator	Funding Period
<u>INVITE</u>	Intuitive Mensch-Technik-Interaktion für die vernetzte Informationswelt der Zukunft	ISA GmbH, Stuttgart	07/99 - 06/03
<u>MORPHA</u>	Intelligente anthropomorphe Assistenzsysteme	Delmia GmbH, Fellbach	07/99 - 06/02
<u>EMBASSI</u>	Elektronische Multimediale Bedien- und Service-Assistenz	Grundig GmbH, Fürth	07/99 - 06/03
<u>ARVIKA</u>	Augmented Reality für Entwicklung, Produktion und Service	Siemens AG, Nürnberg	07/99 - 06/03
<u>SMARTKOM</u>	Dialogische Mensch-Technik- Interaktion durch koordinierte Analyse und Generierung multipler Modalitäten	DFKI GmbH, Saarbrücken	09/99 - 09/03
<u>MAP</u>	Multimedia Arbeitsplatz der Zukunft	AlcatelSel AG, Stuttgart	04/00 - 03/03

SmartKom: Intuitive Multimodal Interaction



The SmartKom Consortium:

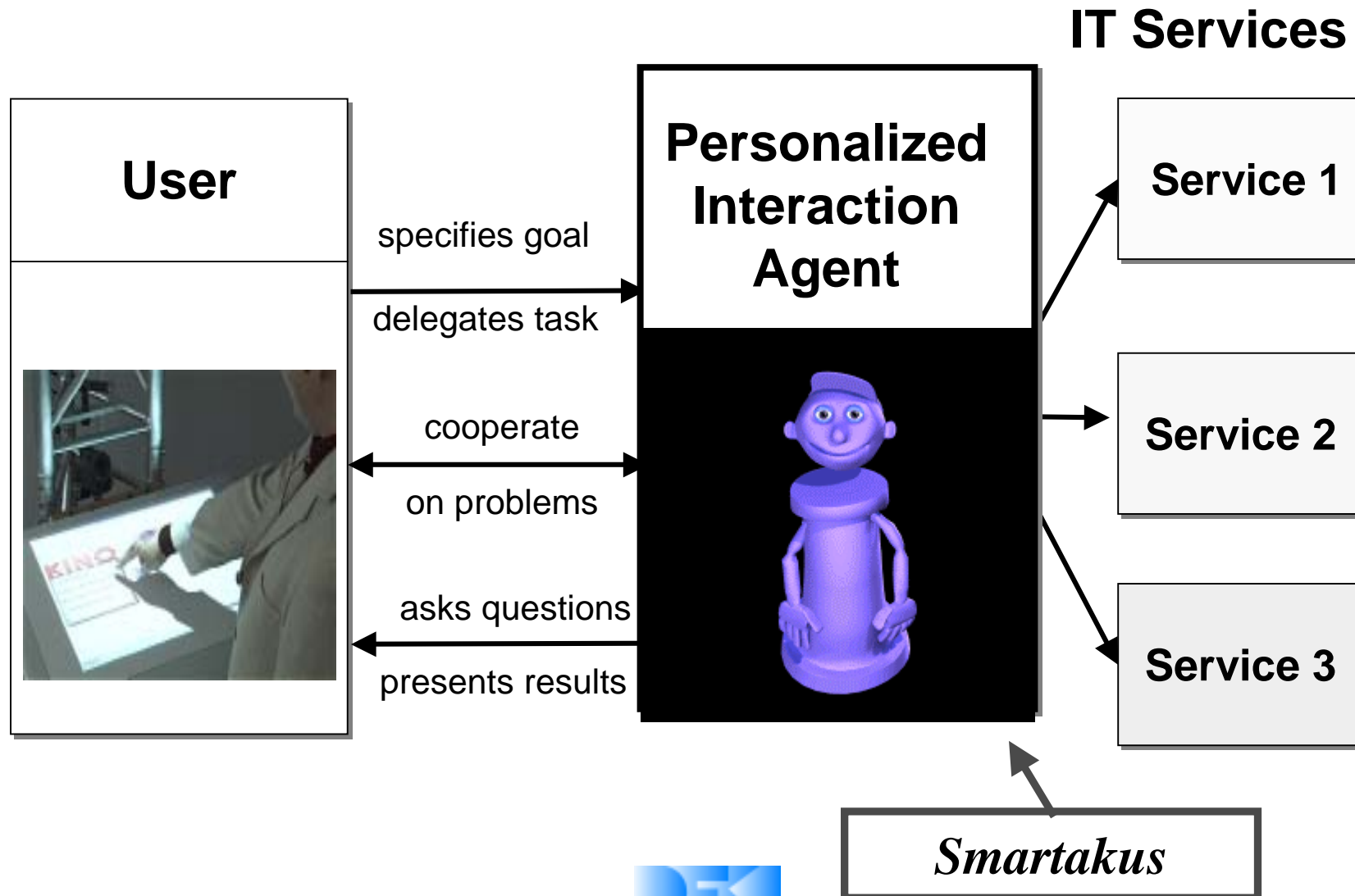


Project Budget: \$ 34 M
 Project Duration: 4 years

Main Contractor
Project Management
Testbed
Software Integration
DFKI GmbH Saarbrücken



Situated Delegation-oriented Dialog Paradigm



SmartKOM-PUBLIC: A Multimodal Communication Booth



Room microphone

Face-tracking camera

Virtual touch screen
protected against vandalism

Multipoint video conferencing



Loudspeaker

Smart card/
Credit Card
for authentication
and billing

Docking station
for PDA/Notebook/
Camcorder
high speed and broad
bandwidth Internet
connectivity

High-resolution
scanner

SmartKom-Mobile: A Handheld Communication Assistant



SmartKom-Home/Office: A Versatile Agent-based Interface



SpeechMike

Virtual Touch screen

Natural Gesture Recognition



Multimodal Interaction in SmartKom

Scenario:

public (mobile, home)

Application:

movie information
(EPG, email, phone, fax,
address book,
TV and VCR control,
routing/tourist info)

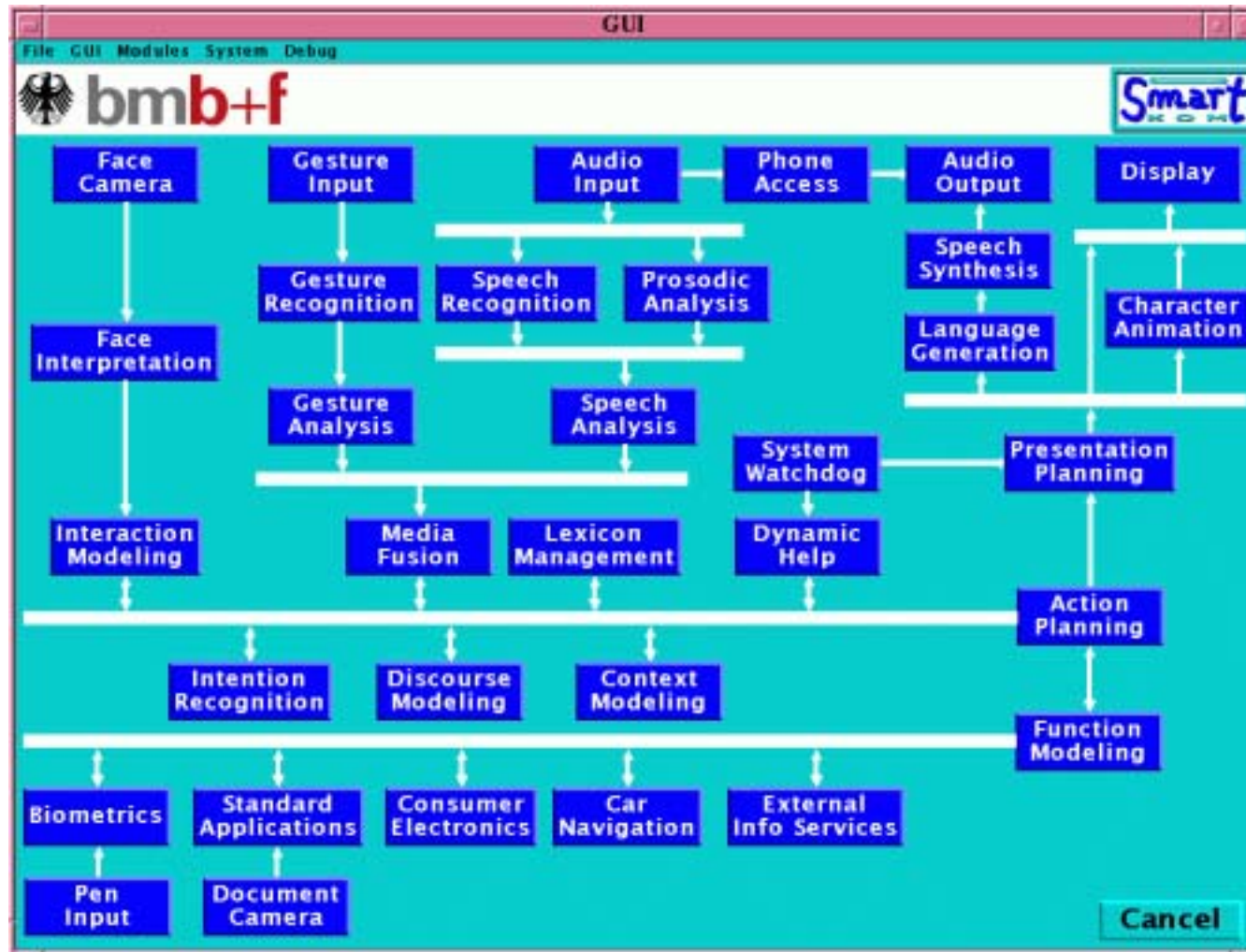


U: *I want to make a reservation in (↑) this movie theater*

S: This theater does not take reservations

U: *Then a different one, (↑) this one perhaps*

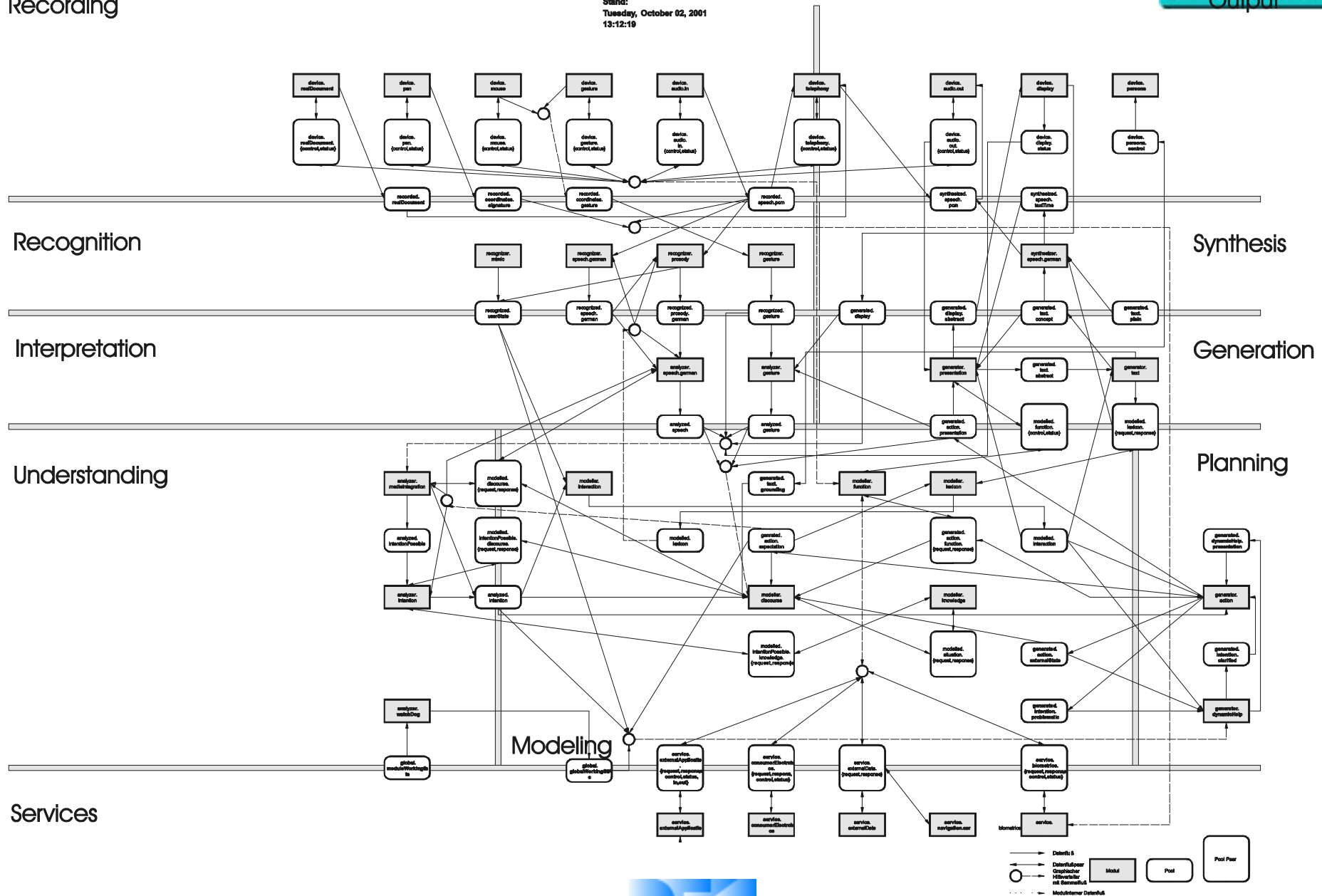
An Overview: The Components of Smartkom



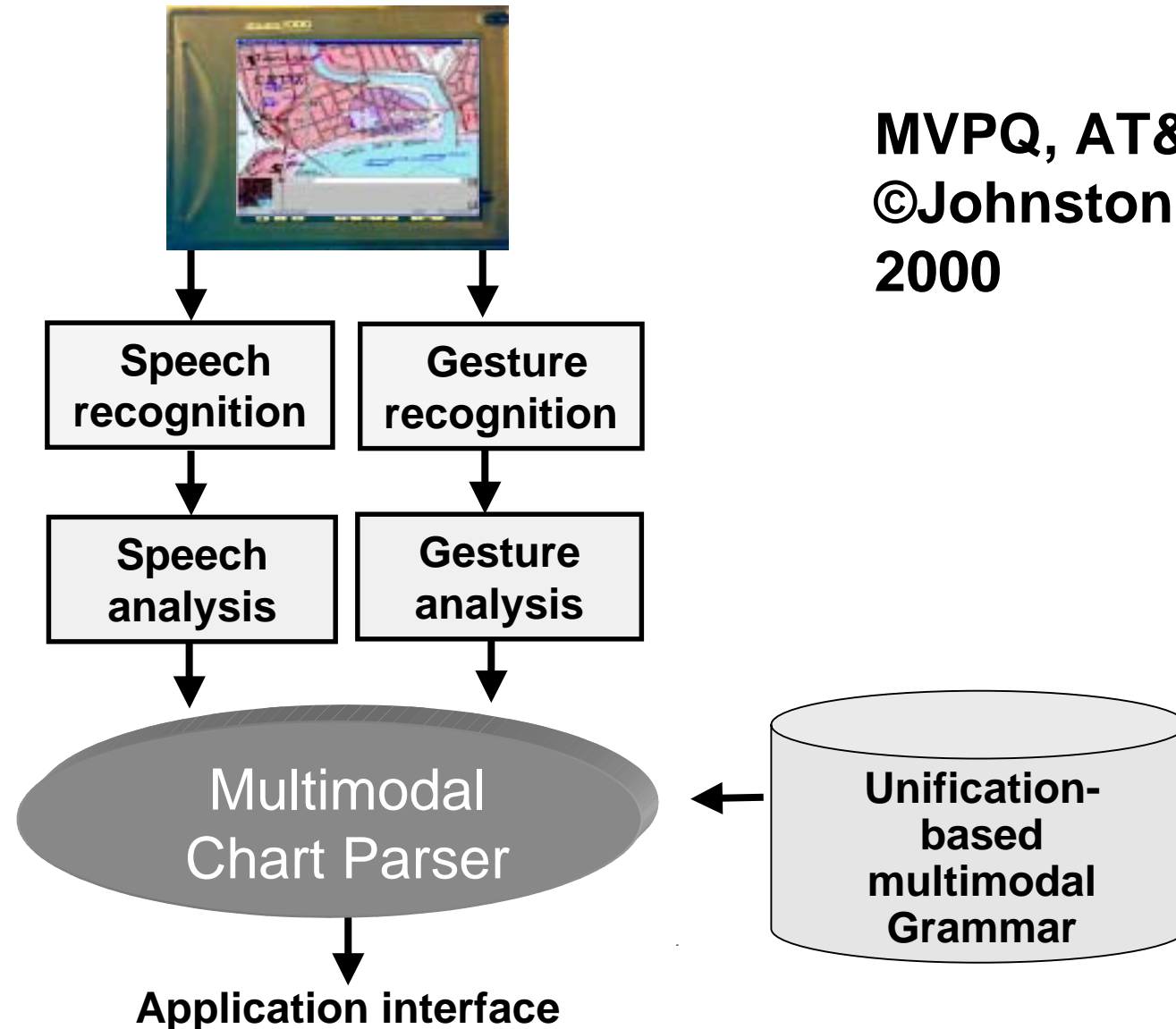
The Real Picture

Recording

Stand:
Tuesday, October 02, 2001
13:12:19



Unification-based Integration of Speech and Gesture



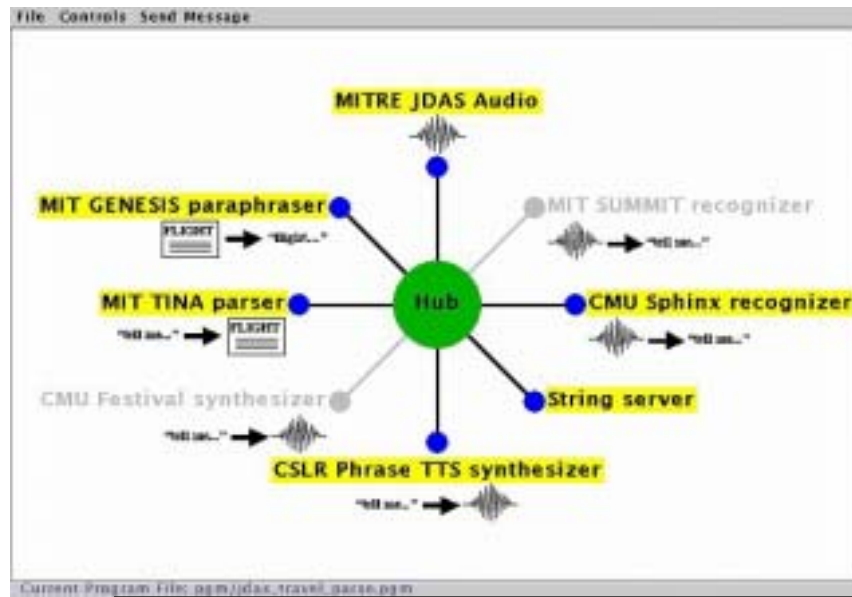
MVPQ, AT&T
©Johnston
2000





Communicator: <http://www.darpa.mil/ito/research/com/projlist.html>

MITRE: Lynette Hirschman

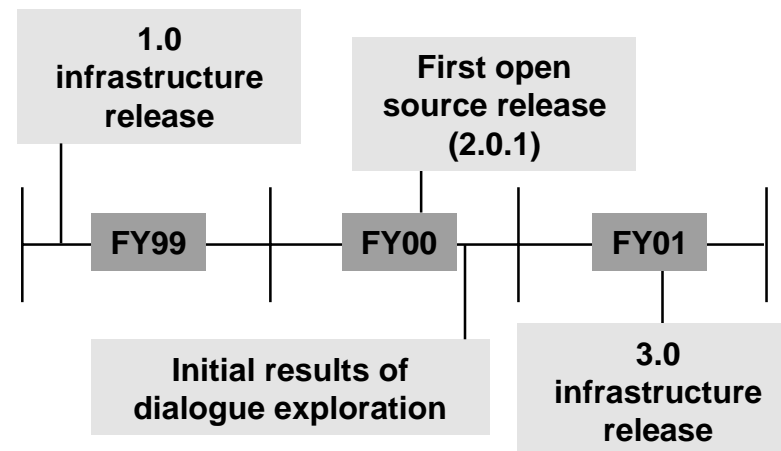


New Ideas

- Robust, open-source cross-platform infrastructure for dialogue interaction
- Mining of human-human and human-computer dialogue data to enhance dialogue design
- Human-factors-based metrics for evaluation of human-computer dialogue

Impact

- Time to plug-and-play components now hours instead of days or weeks
- Hundreds of infrastructure downloads worldwide lowers bar to entry for researchers and product developers
- Leveraging human-human dialogue data for analysis will improve usability of human-computer dialogue strategies



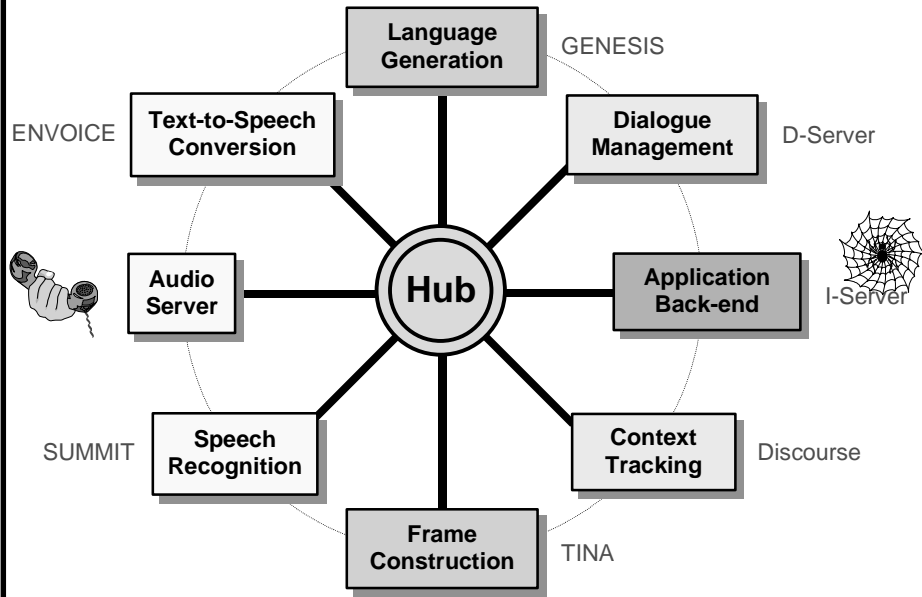


Communicator: Participants

AT&T Labs	Telephone Architecture
BBN Technologies	Dialog Interaction
Carnegie Mellon University	Travel task with architecture
Carnegie Mellon University	Lingwear: Wearable linguistic assistance for foreign language conversation
Corporation for National Corporation	Gateway Design
Daimler Chrysler	Spoken Dialogue Technology
Dragon Systems	Dialog Interaction
HRL Laboratories, LLC	OnTheMove: Distributed Spoken Language Technology for Ground Vehicle Operators
Interactive Drama Inc.	Dialog Interaction
International Business Machines Corporation	Extraction of Linguistic Substructure for Translingual Information Retrieval and Robust Input to Phrase Translation
Johns Hopkins University	Dialog Workshop
Lockheed Martin	LCS Marine
Lucent Technologies Inc.	Dialogue and Acoustic Development
Marine Acoustics	DARPA One-way integration
Massachusetts Institute of Technology (MIT)	Communicator Architecture



Mobile Information Access Using Spoken Dialogue



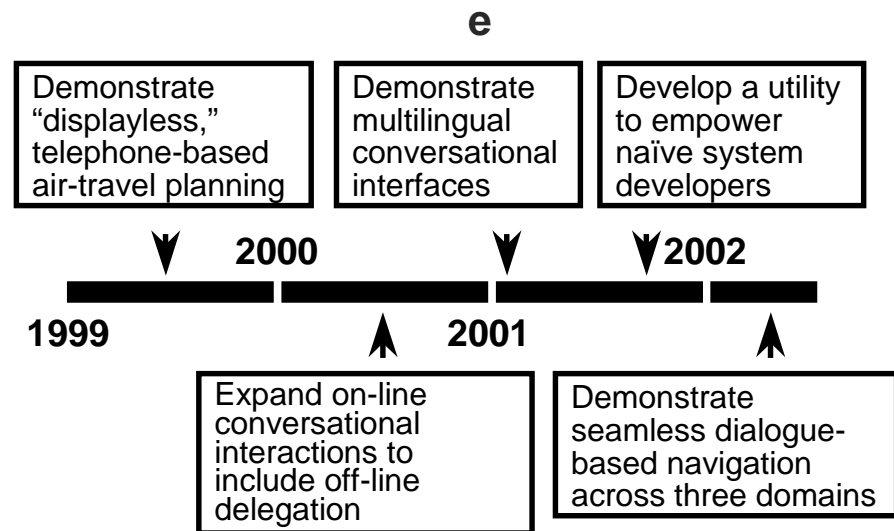
New Ideas

- **Segment-based speech recognition** leads to more explicit utilization of speech knowledge
- **Robust, probabilistic natural language processing** provides constraints for speech recognition, deals with variabilities in spoken language, and derives a meaning representation
- **Mixed-initiative dialogue modelling** facilitates natural interactions between a user and a computer
- **Real application backends** illuminates real research issues and facilitates technology transfer
- **Client-server architecture, Web-based launching, and phone-based input** promote affordability and mobility

Impact

- **Productivity Enhancement:** Change the human-machine interaction paradigm from **programming** to **conversation**, where human and machine cooperate to solve problems better and faster (in applications such as transportation planning, decision support, and interactive training)
- **Enabling Technology:** Revolutionize access to on-line resources for novices, non-programmers, the handicapped, and people without PCs
- **Multiple Modality:** Speech-based interfaces are well-suited for hands-busy, eyes-busy applications (e.g. air-traffic control, navigational assistance, and equipment maintenance)

Schedule





Interfaces and Standards

- **XML**
 - Human readable (ASCII vs. binary)
 - Should nevertheless be managed with tools!
- **Many, many dialects:**
 - VoiceXML, TemicGDML, M3L, SSML, JSPG, ...
 - Tendency towards standardization (see HTML)
 - ⇒ Exchange of tools and knowledge sources
- **XML makes testing and debugging easier**
- **SmartKom uses XML also in all *internal* interfaces and most knowledge sources**

<title>Mrs.</title>
00100110



Modelling Dialogues

- **Dialog Definition Languages (DDL):**
 - Specialized description languages that are tailored to the task: dialogue concepts like *prompts* and *grammars*

Variants:

- Text based: Philips-HDDL, VoiceXML, TemicGDML
- Graphics based: flow charts, dialog state charts

Disadvantage:

- Support only a limited number of constructs

Advantage:

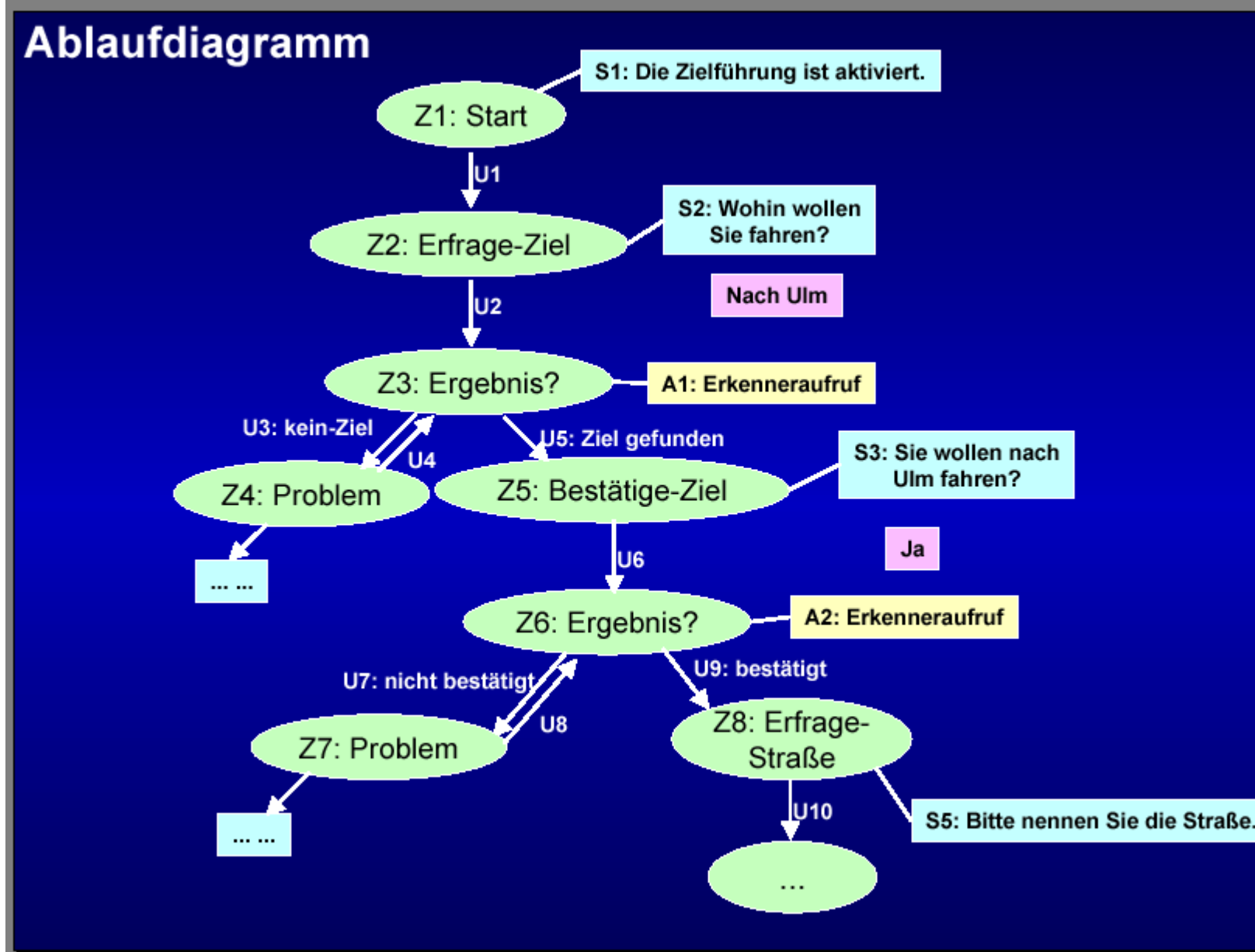
- Clear, structured
- Easily changed, reused
- Can be used by non-experts

Why Dialog Definition Languages?

- **Historically:** dialogue systems are self-contained programs (C, C++, ...)
- **Advantage:**
 - no restrictions
- **Disadvantage:**
 - error-prone, because no restrictions
 - no structure
 - no concepts of dialogues
 - no separation of generic/task-specific data
⇒ adaptation to new applications costly
 - no support in reusing models
 - expert know-how needed

The structure of a dialog

- **Common concepts:**
 - Dialog States
 - State transitions (e.g. flow chart)
 - Branching into sub-dialogues
 - Prompts
 - Recognition grammars
 - Error handling
- **Example:**
 - Next slide



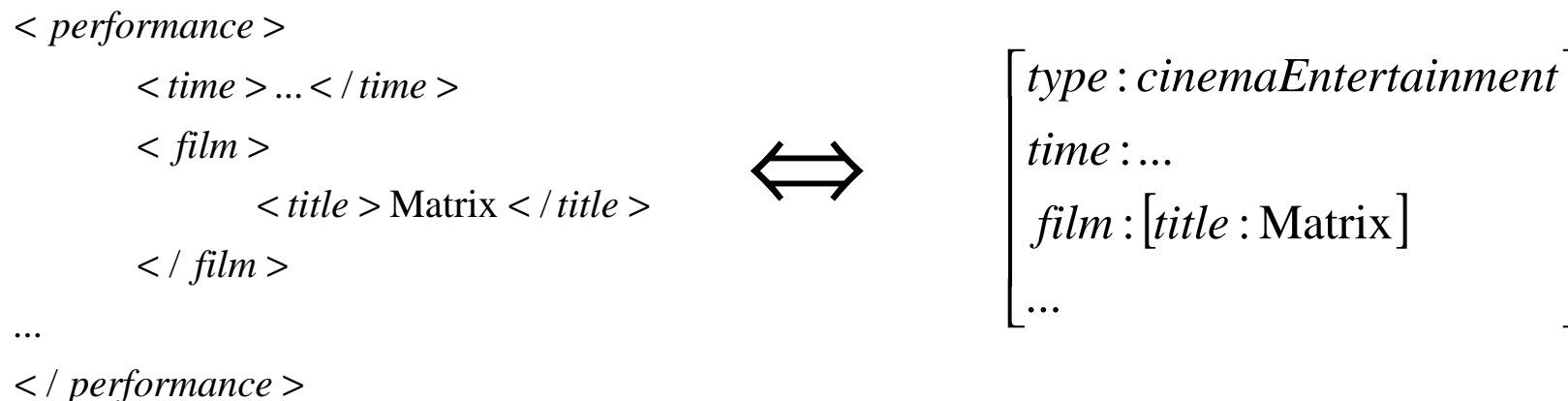
Source: Anke Kölzer, DaimlerChrysler



XML and typed feature structures



- Hypotheses are XML documents as defined in the domain modeling
- XML documents can be viewed as typed feature structures
 - advantage: use results from tfs
 - advantage: contribute to tfs



```

<concept sequence="11">
  <discourseElement id="9011" discourseRelation="simple">
    <sentence id="tsen-9011" sentenceMode="declarative">
      [...]
      <syntaxElement case="acc" argumentStatus="Object" syntaxCategory="NP">
        <syntaxElement syntaxCategory="Det">
          <lexicalElement partOfSpeechTag="ART">
            <text> die </text>
          </lexicalElement>
        </syntaxElement>
        <syntaxElement syntaxCategory="N">
          <lexicalElement partOfSpeechTag="NN">
            <text> Anfangszeiten </text>
          </lexicalElement>
        </syntaxElement>
      </syntaxElement>
    </sentence>
  </discourseElement>
</concept>

```

Result of Natural Language Generation

```

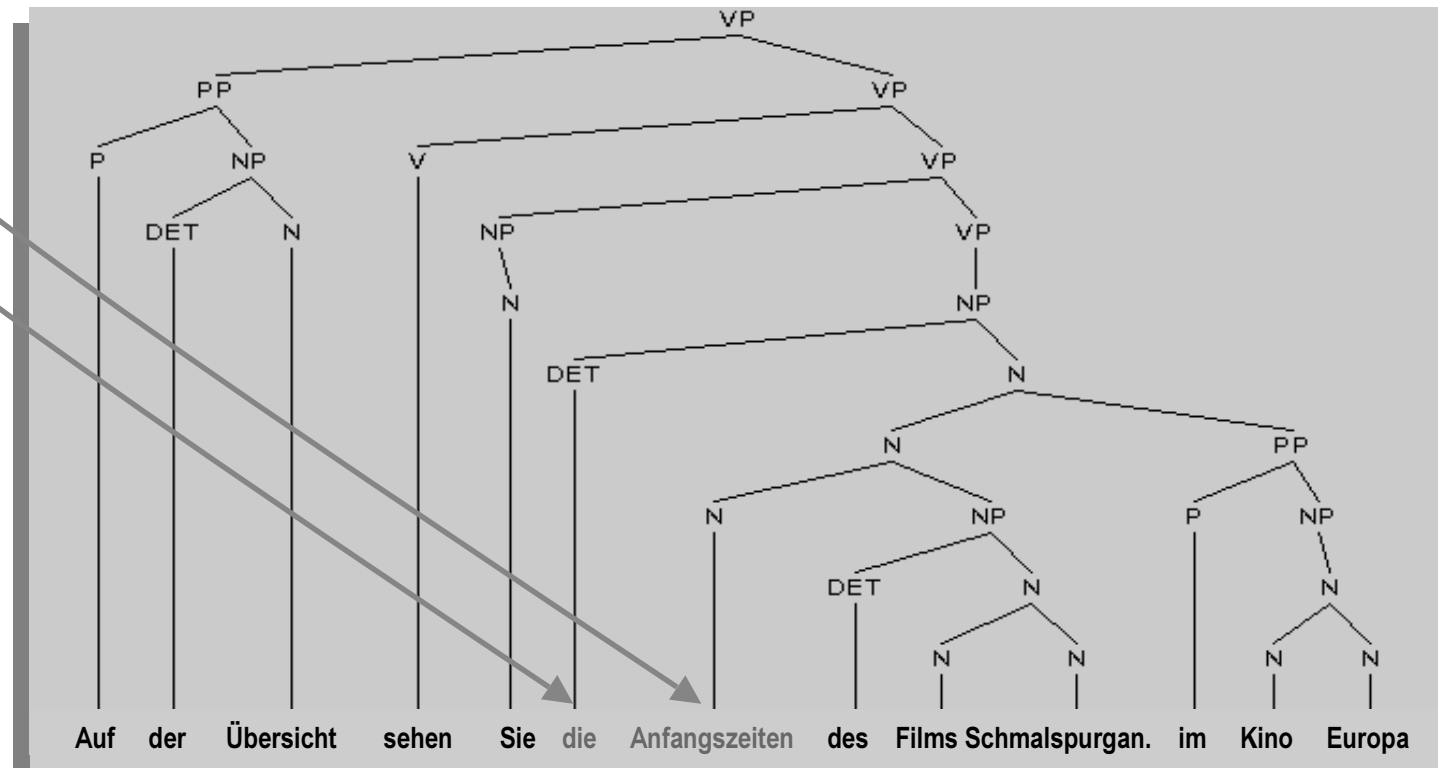
[...]
</syntaxElement>
</sentence>
</discourseElement>
</concept>

```



XML Format

graphical representation 



Don't forget the Tools!



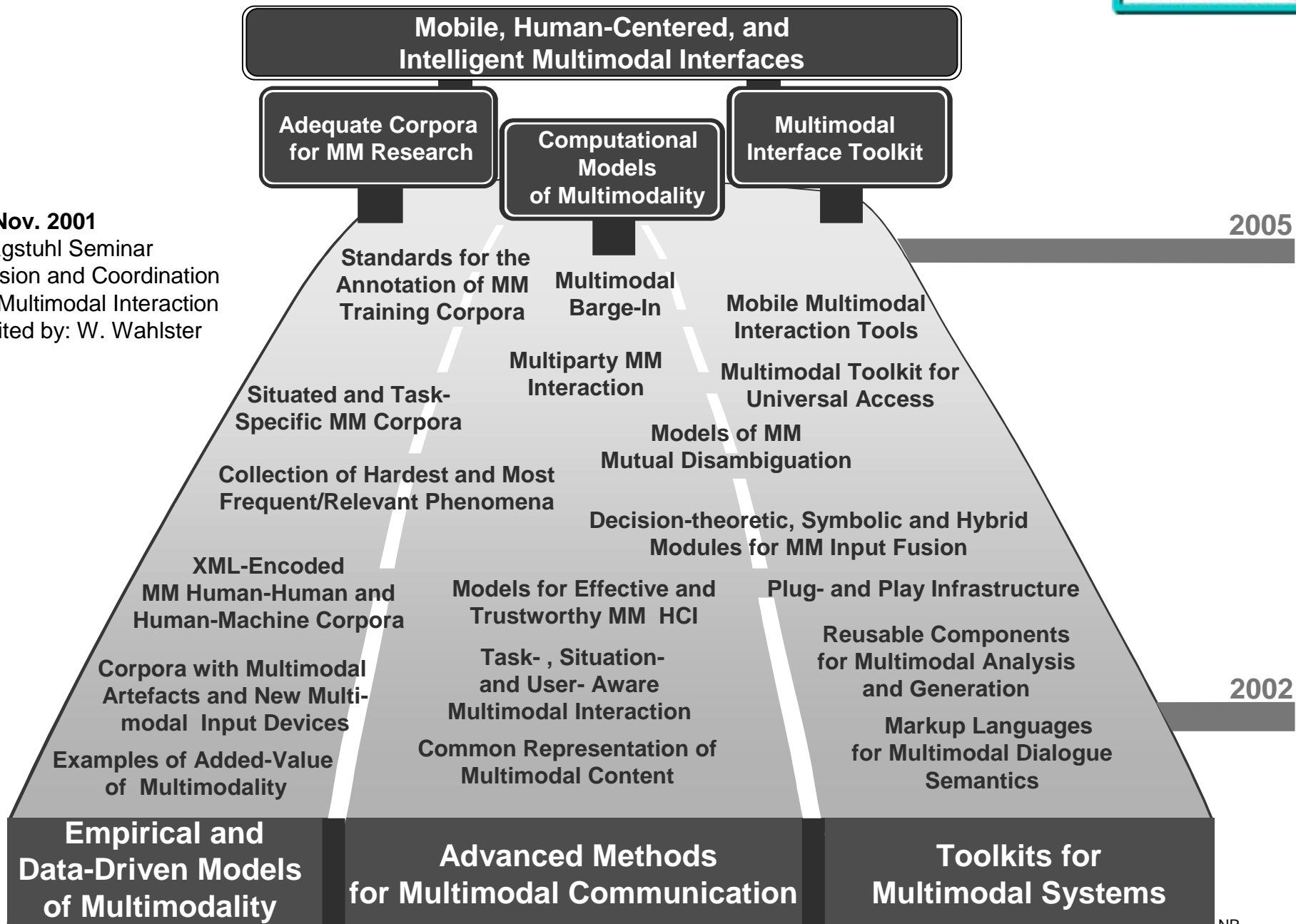
- Example: Graphical editor for grammars
- Implemented in Java
 - Interfaces
 - XML

The screenshot displays a graphical user interface for editing grammars. On the left, a parse tree for the sentence "Wollen Sie sehen" is shown. The root node is "S", which branches into "V" (Wollen) and "VP". The "VP" node further branches into "NP" (Sie) and "VP" (sehen). The "NP" node is highlighted in red. The interface includes a "file help" menu, a "redraw" button, and an "add new root" button. A file browser window is open, showing a list of grammar files, with "Wollen Sie sehen" selected. The file browser also shows a list of actions: "add tree family", "remove tree family", "rename tree family", "add tree", "copy tree", "remove tree", and "edit tree".



Research Roadmap of Multimodality 2002-2005

2 Nov. 2001
 Dagstuhl Seminar
 Fusion and Coordination
 in Multimodal Interaction
 edited by: W. Wahlster

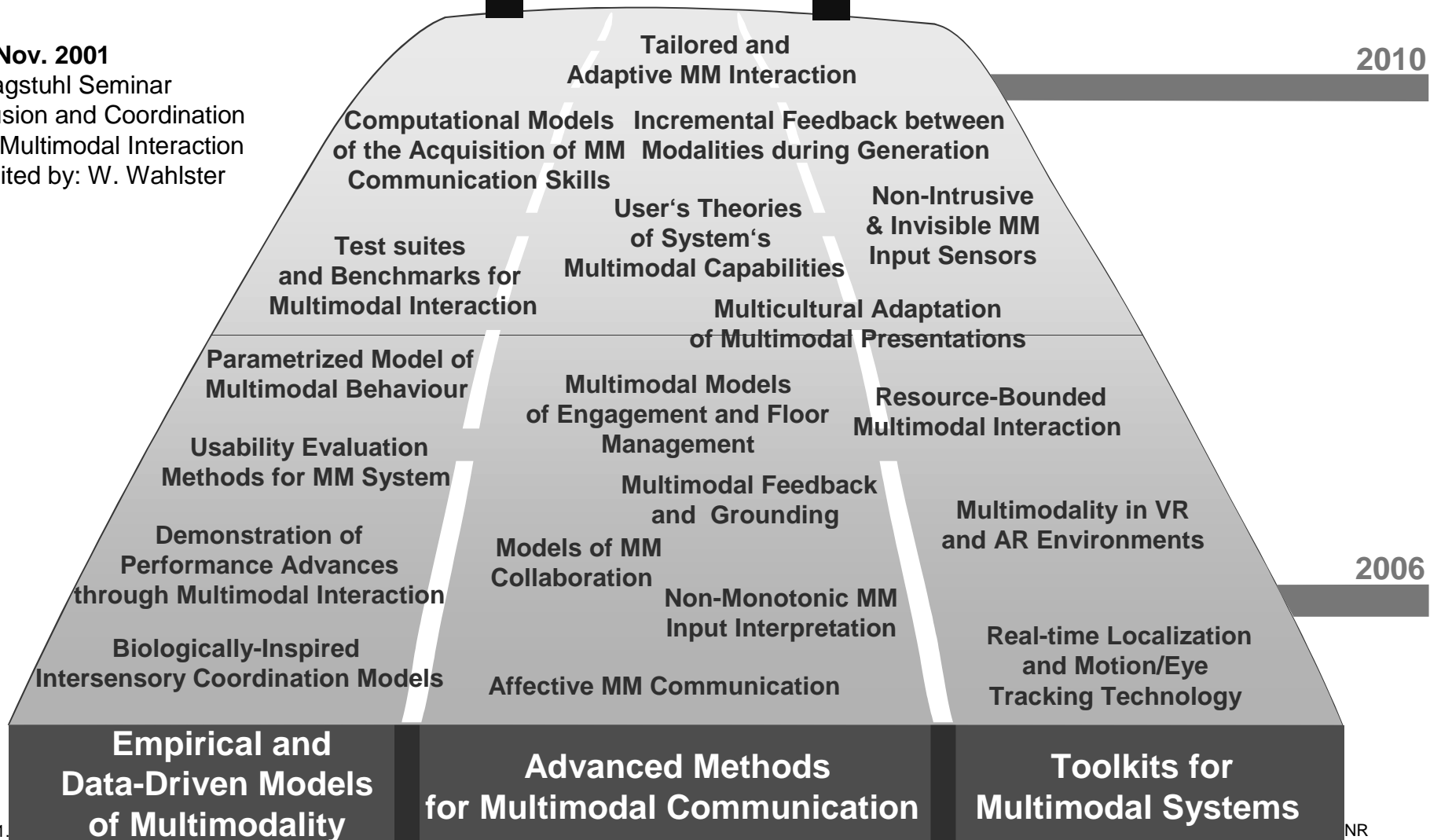


Research Roadmap of Multimodality 2006-2010



Ecological Multimodal Interfaces

2 Nov. 2001
 Dagstuhl Seminar
 Fusion and Coordination
 in Multimodal Interaction
 edited by: W. Wahlster



Research Roadmap of Multimodality 2001-2010

Enabling Technologies and Important Contributing Research Areas



2 Nov. 2001
 Dagstuhl Seminar
 Fusion and Coordination
 in Multimodal Interaction
 edited by: W. Wahlster

Multimodal Input	Multimodal Interaction	Multimodal Output
<ul style="list-style-type: none"> ● Sensor Technologies ● Vision ● Speech & Audio Technology ● Biometrics 	<ul style="list-style-type: none"> ● User Modelling ● Cognitive Science ● Discourse Theory ● Ergonomics 	<ul style="list-style-type: none"> ● Smart Graphics ● Design Theory ● Embodied Conversational Agents ● Speech Synthesis

- Machine Learning
- Formal Ontologies
- Pattern Recognition
- Planning



Conclusion

- **Multi-modal Dialog Systems are close to market**
- **Have a clear design of your application**
 - Wizard of Oz, Application Design (use cases, demo dialogues, evaluation, ...)
- **Think about multimodality:**
 - gestures, user state (emotion), written text, haptic buttons, jog dials etc.
- **Architecture design is converging**
- **Use XML and upcoming standards for modelling and interfaces**