Interacting with the Ambience: Multimodal Interaction and Ambient Intelligence

W3C Workshop on Multimodal Interaction

19/20 July 2004,

Sophia Antipolis, France

Kai Richter (ZGDV) & Michael Hellenschmidt (Fh-IGD)

Fraunhofer Institut Graphische Datenverarbeitung

Zentrum für Graphische Datenverarbeitung e.V.

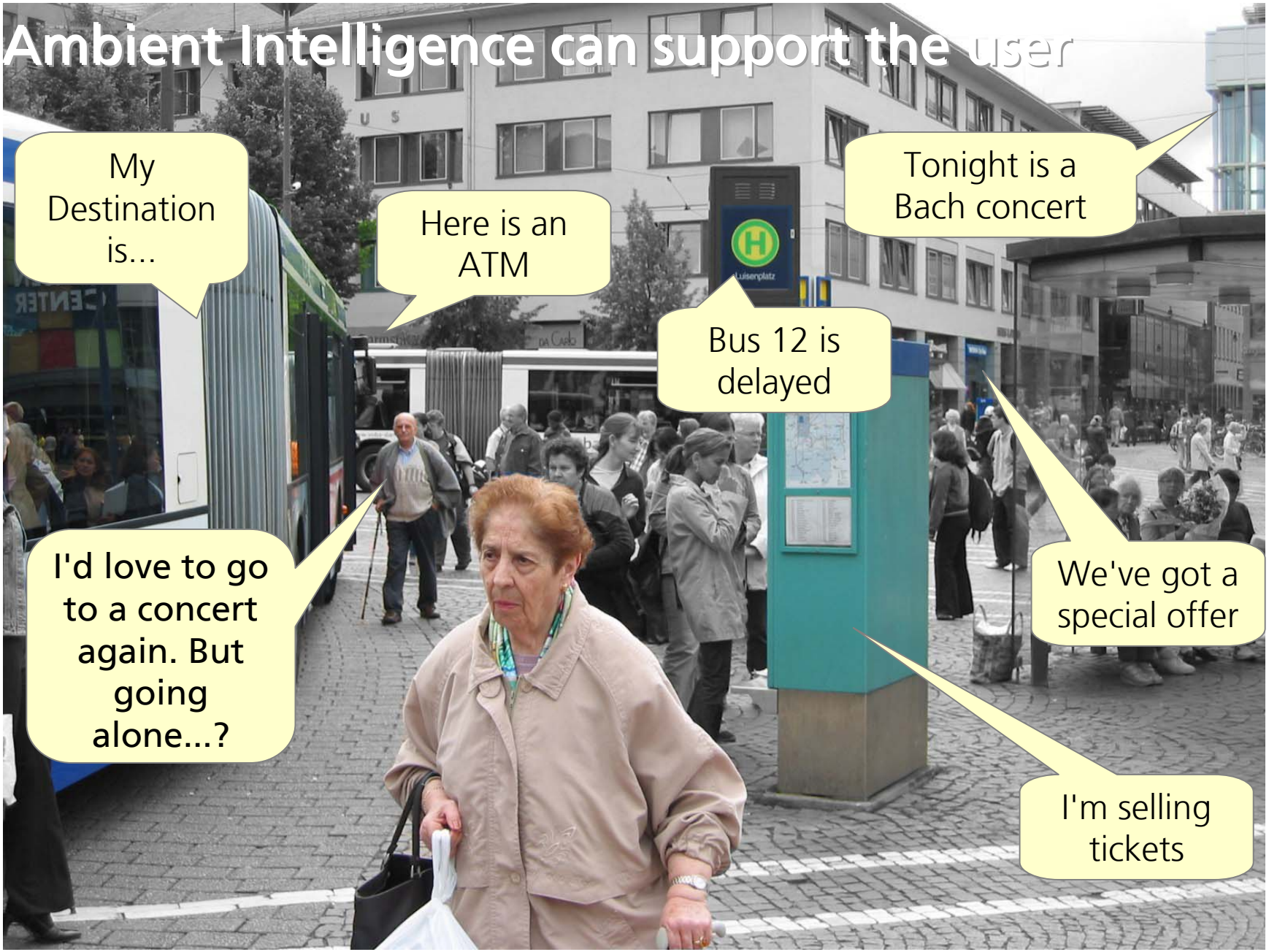
Member of INI-GraphicsNet

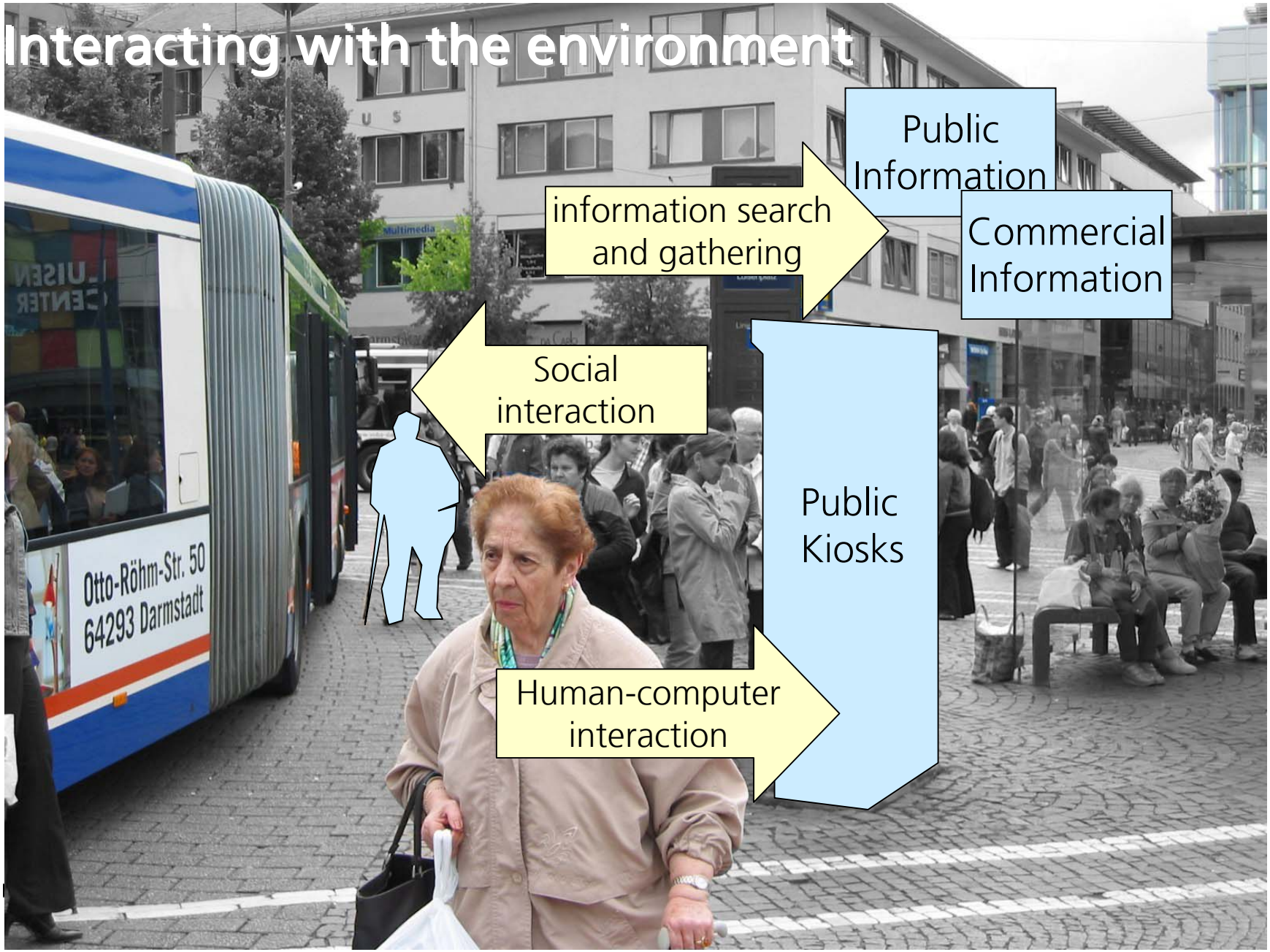
Overview

- Motivation
- Multimodal access for disabled persons
- Evaluation results and further requirements
- Middleware approach for dynamic device configurations
- Current and future work

The User and his environment

Ambient Intelligence can support the user




Interacting with the environment
Public Information
information search and gathering
Commercial Information
Social interaction
Public Kiosks
Human-computer interaction
Otto-Röhm-Str. 50
64293 Darmstadt

Problem

- Diversity of devices and services
- Diversity of user's needs and capacities
- Diversity of interaction devices and user interfaces

therefore...

*Ambient Intelligence targets to "[…] improve the quality of life of **people** by creating the desired atmosphere and functionality via intelligent, personalized inter-connected systems and services."*

Philips

Application scenario: Access for disabled users

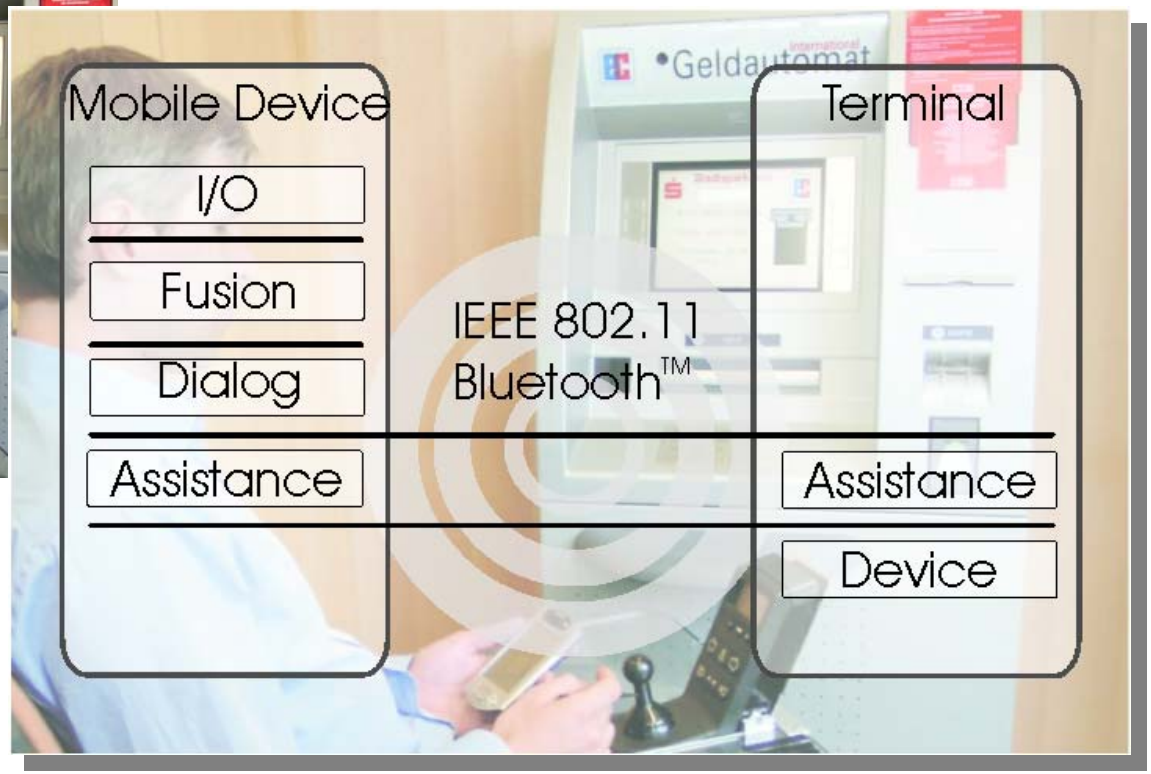
- How can public kiosk systems be made accessible to the great variety of users with special needs?
  - Provide a mobile device to access ambient infrastructure
  - Equip mobile device with interaction devices supporting those modalities which are optimal for the user
  - implement additional personal assistance on the mobile device

> *"[…] multimodal interfaces have the potential to accommodate a broader range of users than the traditional interfaces"*
>
> Oviatt, 2003

The EMBASSI Project

EMBASSI

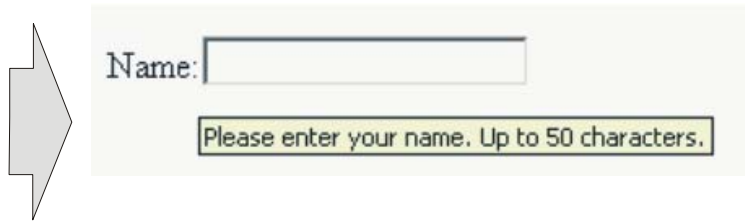
- Mobile Multimodal Assistants



eXtensible User Interfaces Language

- XML User Interface Language based on W3C XForms

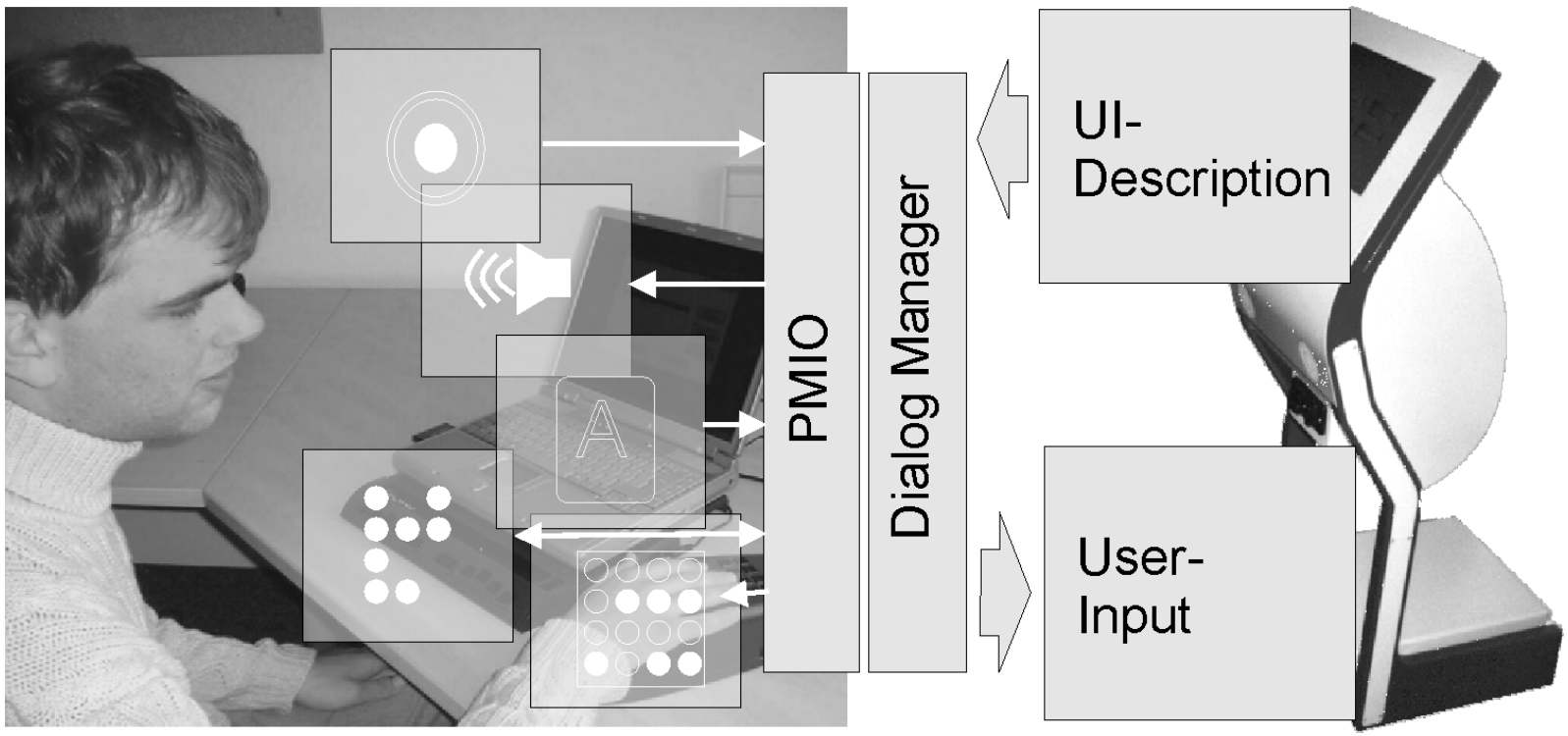
```
<?xml version="1.0" encoding="UTF-8"?>
<input id="field1" lang="DE"
    navIndex="2" accessKey="s">
    <caption>
        <text>Name:</text>
    </caption>
    <hint>Please enter your name.
        Up to 50 characters</hint>
    <help>http://www.embassi.de</help>
    <model>
   <instance/>
   <schema xmlns:xs="...XMLSchema">
    <xs:restriction base="xs:string">
      <xs:length value="50"/>
    </xs:restriction>
   </schema>
  </model>
  <style>
        <x_pos>0.1</x_pos>
        <y_pos>0.2</y_pos>
        <height>0.1</height>
        <width>0.1</width>
    </style>
</input>
```



System (normal voice): "Name"
User: "What?"
System (higher voice): "Please enter your name. Up to 50 characters"
User: "C-A-R-L"

Several prototype implementations

- Example: Device for visually impaired



Device Configurations

- Mobile device for non-disabled
  - Compaq iPAQ with HTML UI
- Mobile device for visually impaired
  - Braille keyboard and line, keyboard, tactile display, voice output (3 voices for different types)
- Mobile device for physically handicapped
  - Control unit for devices with few degrees of freedom, voice input, virtual keyboard in combination with sentence completion software
- Mobile device for elderly people
  - Tablet PC with speech I/O, large scale UI-components and conventional I/O (pen)

Empirical Evaluation

- Prototypes have been tested by users:
  - Physically handicapped users were able to operate an example shopping terminal with the mobile assistant more accurately than non disabled without a mobile device.
  - Physically handicapped users rated the system as very easy to learn, they had fun while operating on it and they experienced a good support from the system.

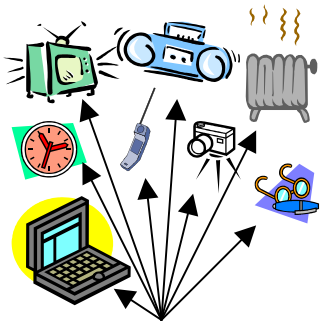


Challenges

- Cash Dispenser
- Vending Machine
- Ticket Machine
- etc.

- Graphical User Interface
- Synthesized Speech
- Virtual Characters
- Braille Output
- Gesture Interfaces
- Speech Input / Recognition
- etc.

Shortcomings of the present approach



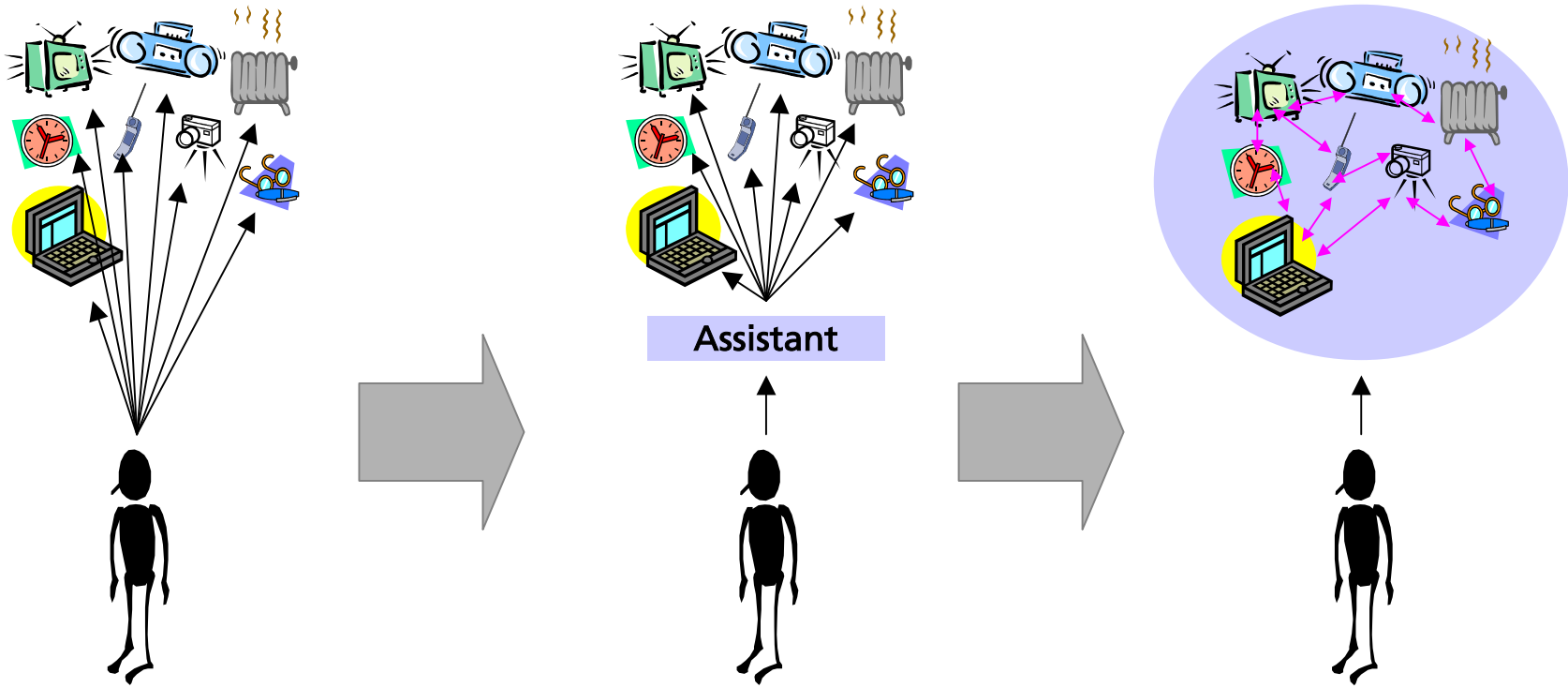
Controlling and monitoring by a Central Component

There are almost infinite combinations of multimodal input and output devices

For every combination of multimodal input and output devices different (pre-configured) assistants are needed.

Disadvantages:
- how to cover all possible combinations ?
- handcrafted by software engineers !

Interacting with the Ambience



Controlling and Monitoring by the User

Controlling and monitoring by a Central Component

Controlling and Monitoring by Self-Organization

Middleware-Requirements

- Extensibility (with new components)
- Independency (of components)
- Avoidance of central components (what will happen, if a central assistant drops out?)
- Exchangeability of components

Strategies needed:

How are tasks completed in such a dynamic system?

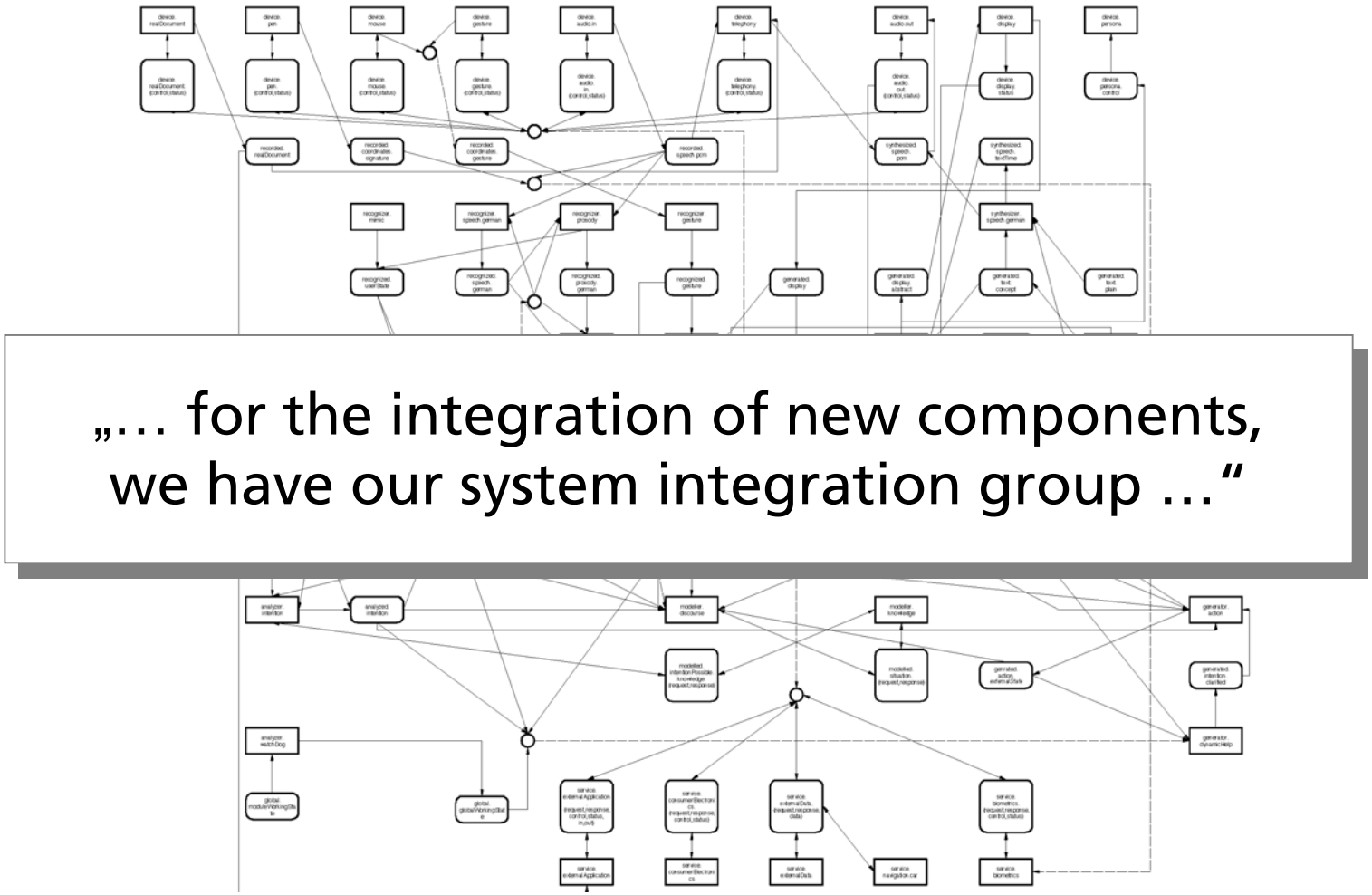
- Applicability of Distributed Problem Solving Strategies (e.g. presentation of system output with different complement components – graphical output together with voice output)

What if two components are competing?

- Applicability of Conflict Resolution Strategies (if different components compete for same tasks)

→ Where are those strategies located?

What we do not want to see ...

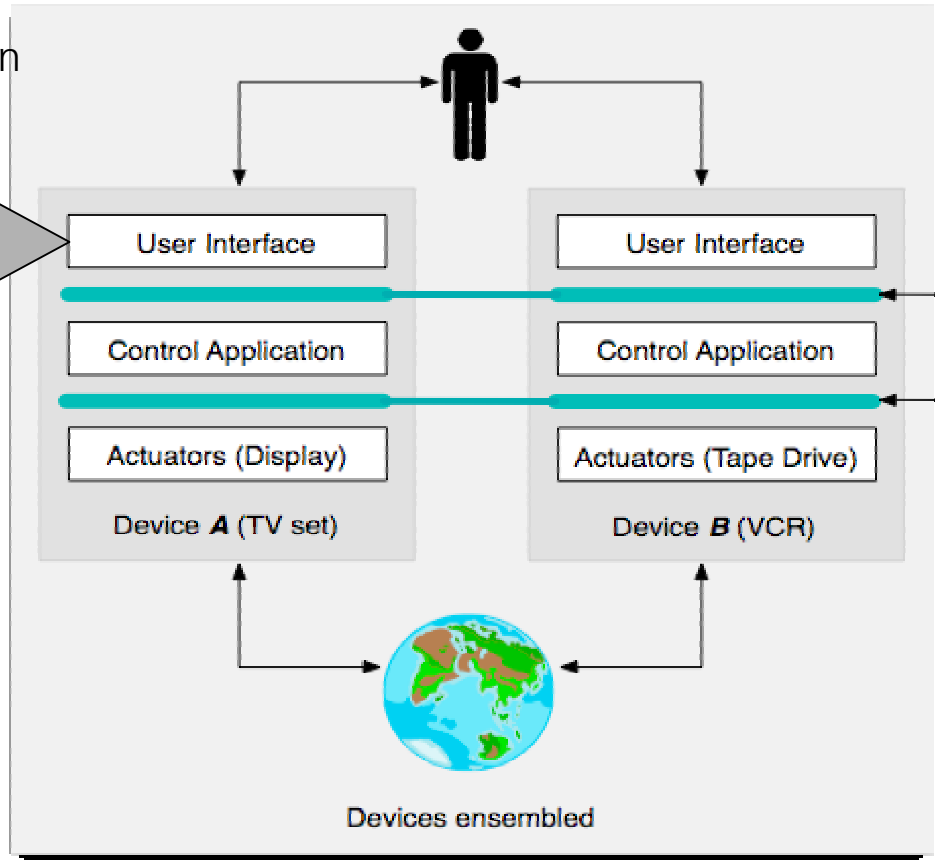


„… for the integration of new components, we have our system integration group …“

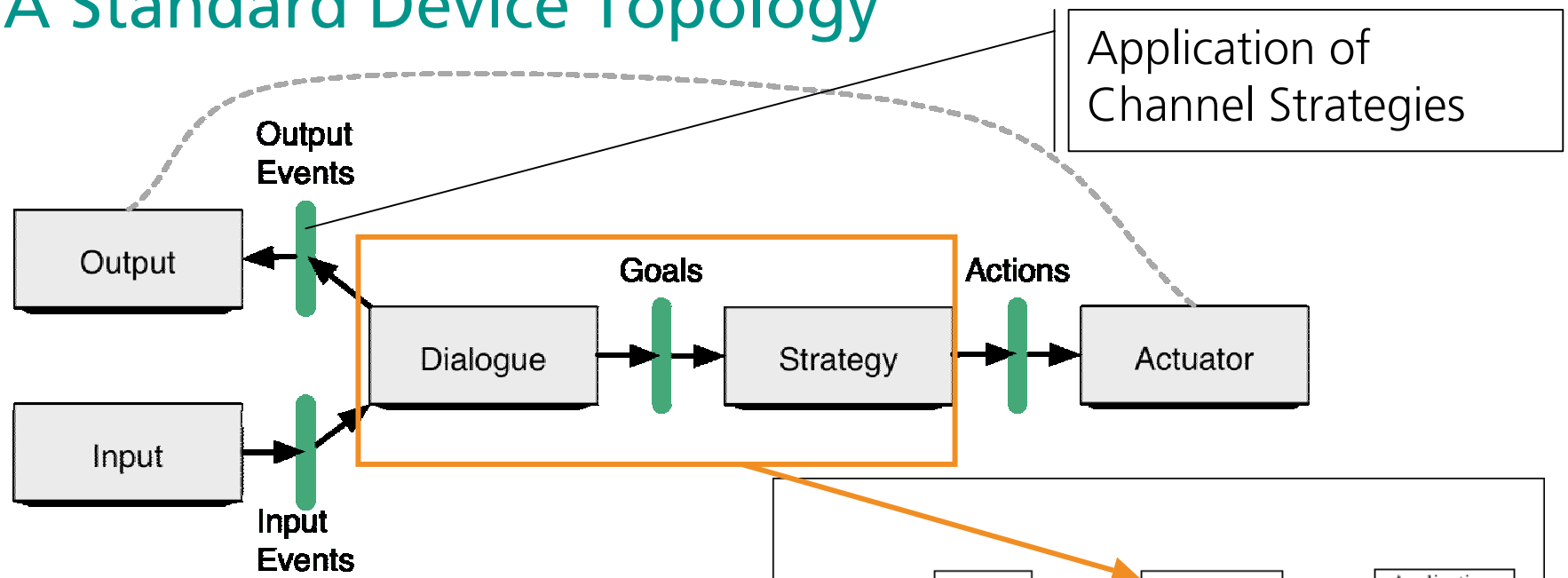
Architectural Considerations

Within Level:
Fine-grained self-organization of functionally similar components

Between Levels:
Coarse-grained self-organization based on a data-flow partitioning



A Standard Device Topology



... bears some resemblance to ....

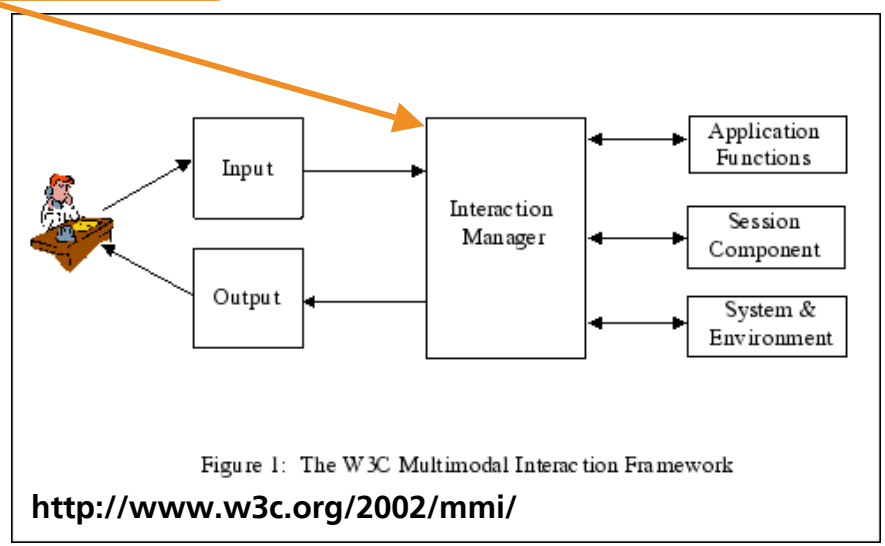
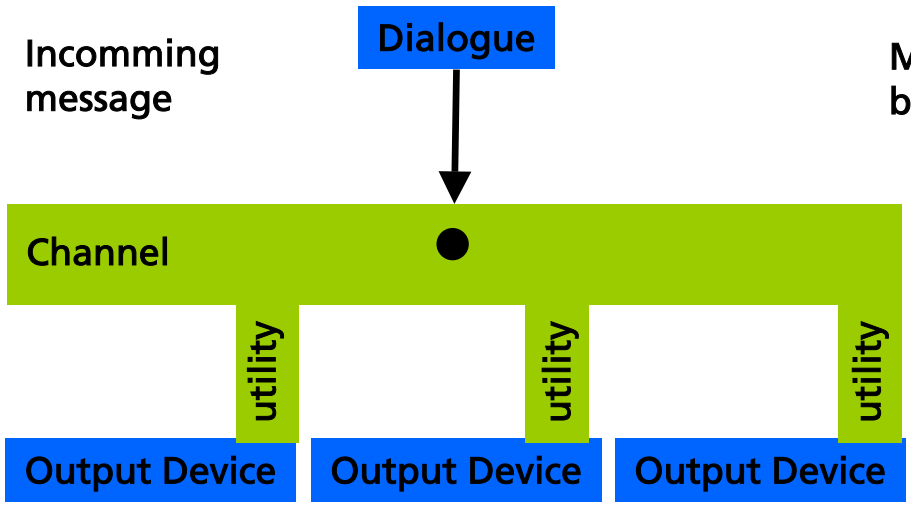
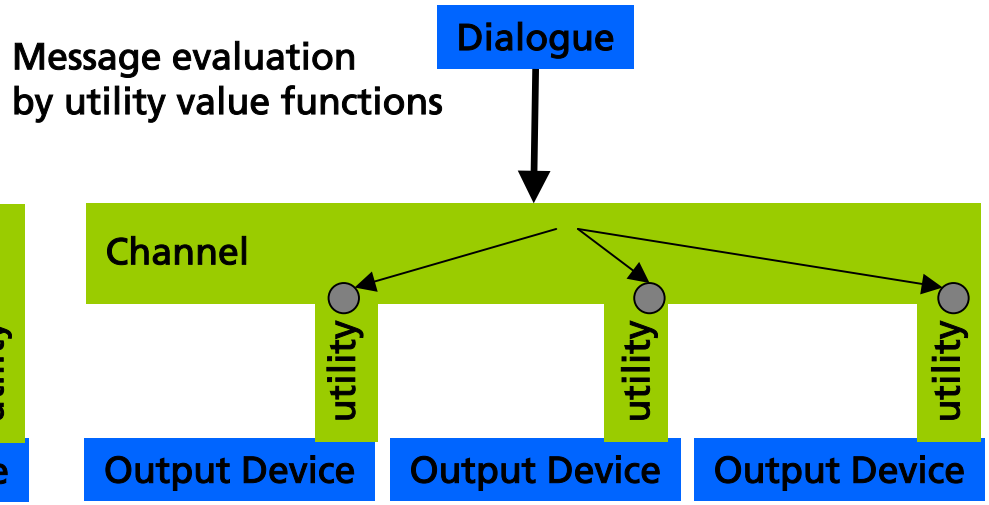


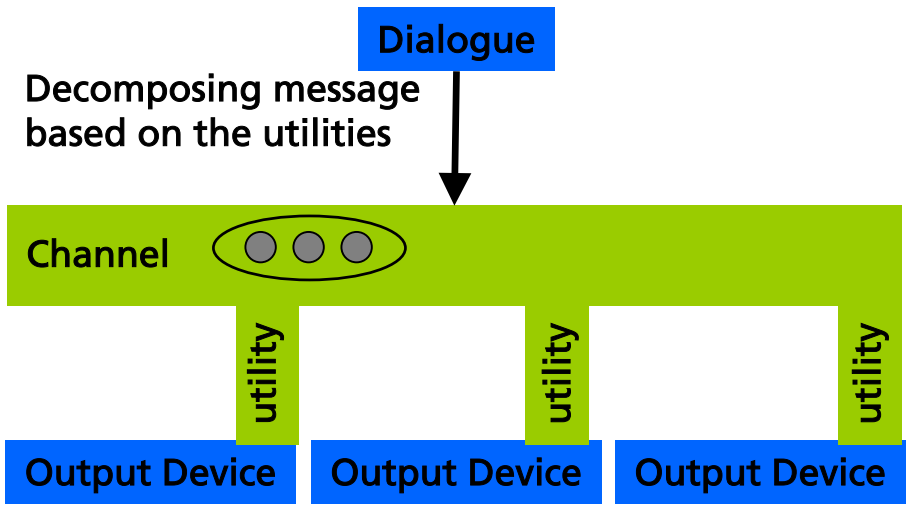
Figure 1: The W3C Multimodal Interaction Framework

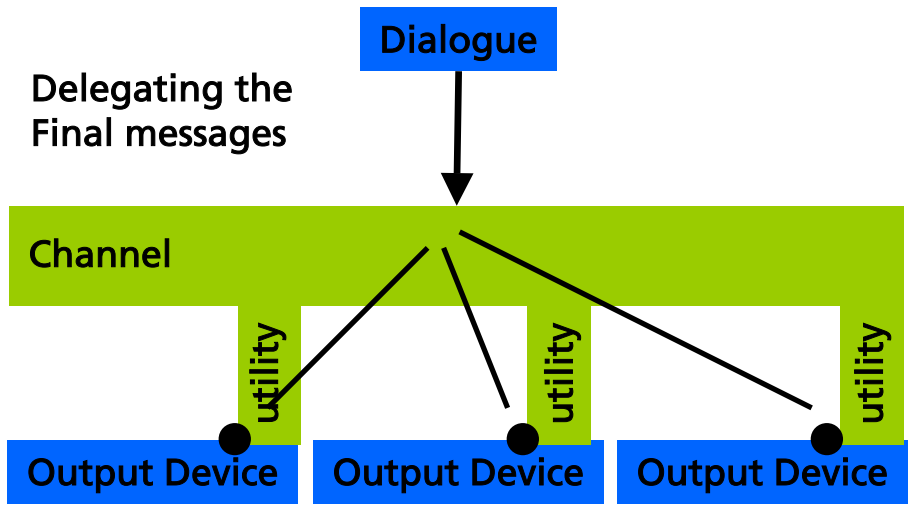
http://www.w3c.org/2002/mmi/

Channel Strategies









Fraunhofer Institut Graphische Datenverarbeitung

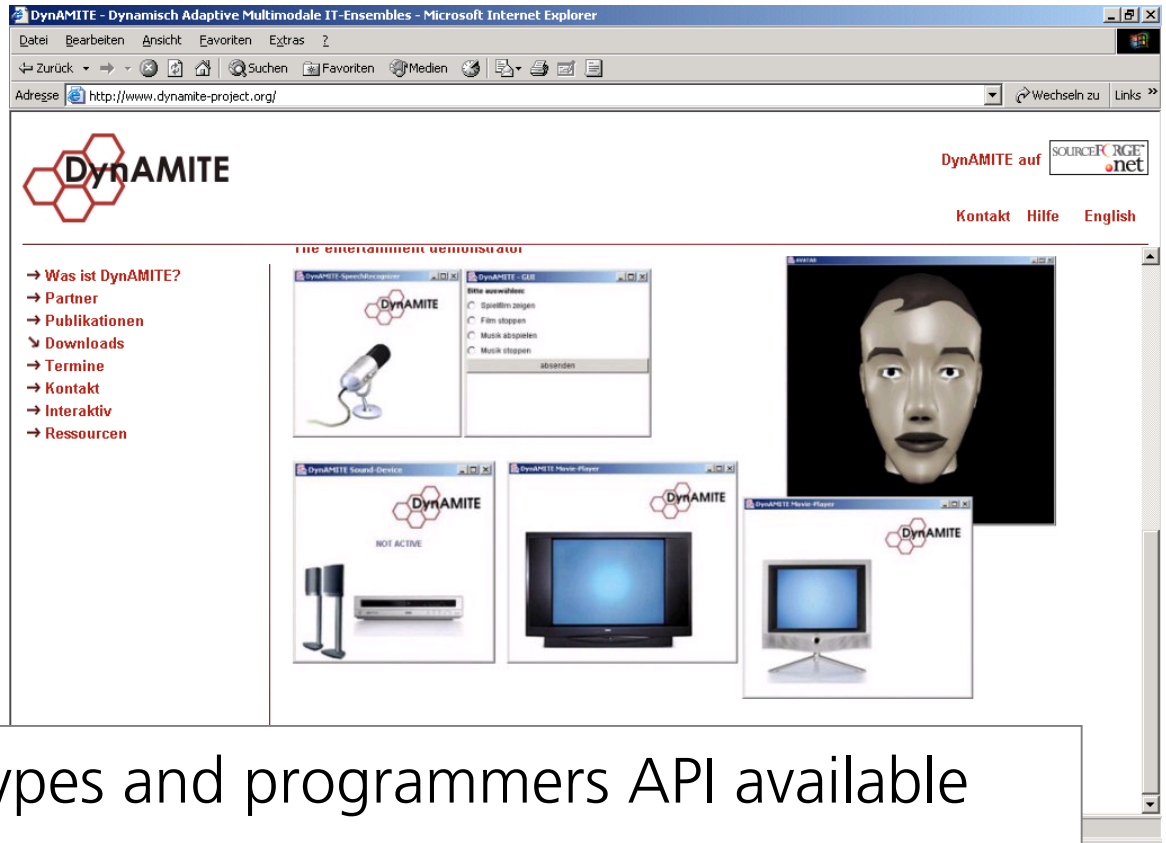
Zentrum für Graphische Datenverarbeitung e.V.

Member of INI-GraphicsNet

Current State

- **Middleware Model:** SodaPop – Self-Organizing Data-flow suPporting Ontology-based Problem Decomposition
- **Prototype** of the Middleware (Haskell, Java 1.3.1 or higher)
- Some Channel **Strategies**
  - Multimodal Presentation Strategy: supports graphical user interfaces, synthesized speech output and virtual characters
  - Incoming events / sequences of events

DynAMITE



Software / Prototypes and programmers API available from:

The DynAMITE project homepage: http://www.dynamite-project.org

Fraunhofer IGD Institut Graphische Datenverarbeitung

Zentrum für Graphische Datenverarbeitung e.V.

Member of INI-GraphicsNet

Summary

- Multimodality is not limited to text and speech
- Multimodality has the potential to provide disabled users with access to information technologies.
- Rather individually configuration and personalization on a personal mobile device than equipping the kiosk systems with some assistive technology.
- Dynamically configurable systems need flexible middleware architecture.
- Distributed Middleware, basing on “publish-subscribe-mechanism”, that evaluates utility-value-functions guarantees more flexibility than “hard-wired” solutions.

Future Work

- Facilitating multi-modal and multi-device development
  - Graphical development tools with design support (tweaking)
  - Encapsulating MMUI within APIs supporting conventional development.
  - Quality measures for generated presentations (e.g. Cross-device-consistency)
- W3C EMMA could be a suitable format for propagating channel messages.
- Develop presentation strategies optimised for users with certain disabilities and for a broader diversity of output devices

Additional Information

- EMBASSI: http://www.embassi.de
- DynAMITE: http://www.dynamite-project.org
- Ambient Intelligence at the INI-GraphicsNet
  - http://ami.inigraphics.net (effective from: 2004 / 08 / 01)
  - http://www.igd.fhg.de/igd-a1/amiatini/index.html (for the present)

Thank you for your attention.

Fraunhofer Institut Graphische Datenverarbeitung

Zentrum für Graphische Datenverarbeitung e.V.

Member of INI-GraphicsNet