Berlin

THEME TANK

5 June 2010

# Future Spoken Dialog Systems:

## Multimodal, Multilingual, Multiparty, Multitask

**Wolfgang Wahlster**

German Research Center for Artificial Intelligence
DFKI
Saarbrücken, Kaiserslautern, Bremen, Berlin
e-mail: wahlster@dfki.de
WWW:http://www.dfki.de/~wahlster

# 13 Trends for Spoken Dialog Systems

1. From Unimodal
   to Multimodal Dialogs

2. From Monolingual
   to Multilingual Systems

3. From Single Task
   to Multitask Dialogs

4. From Dyadic Dialogs
   to Multiparty Conversations

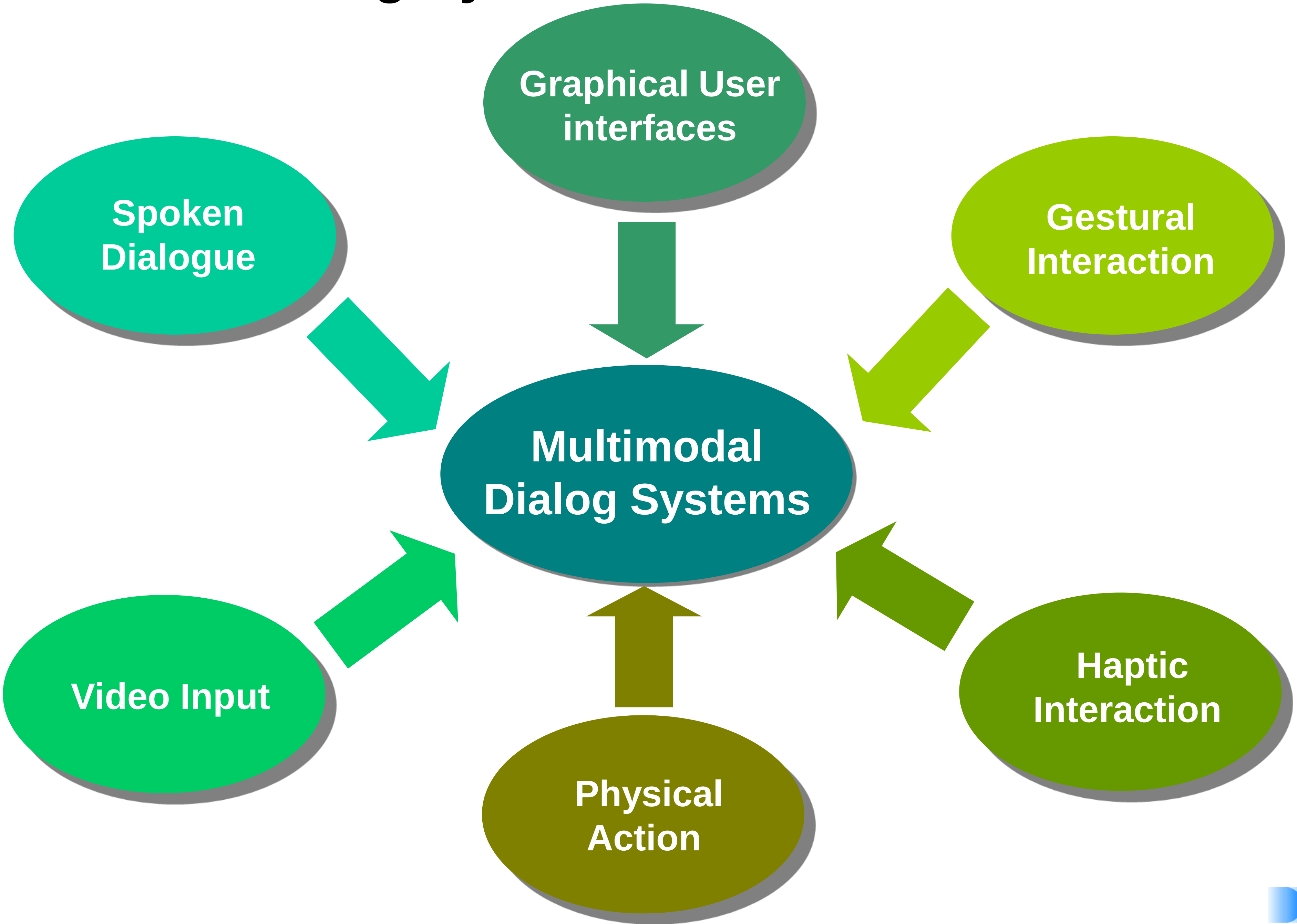
# 13 Trends for Spoken Dialog Systems

5. From Close Speaking
   to Microphone Arrays for Distant Speaking
6. From Cooperative Speech
   to Spontaneous Speech
7. From Stationary
   to Mobile Spoken Dialog Systems
8. From Hosted Voice Portals
   to Cloud-based Speech Solutions

# 13 Trends for Spoken Dialog Systems

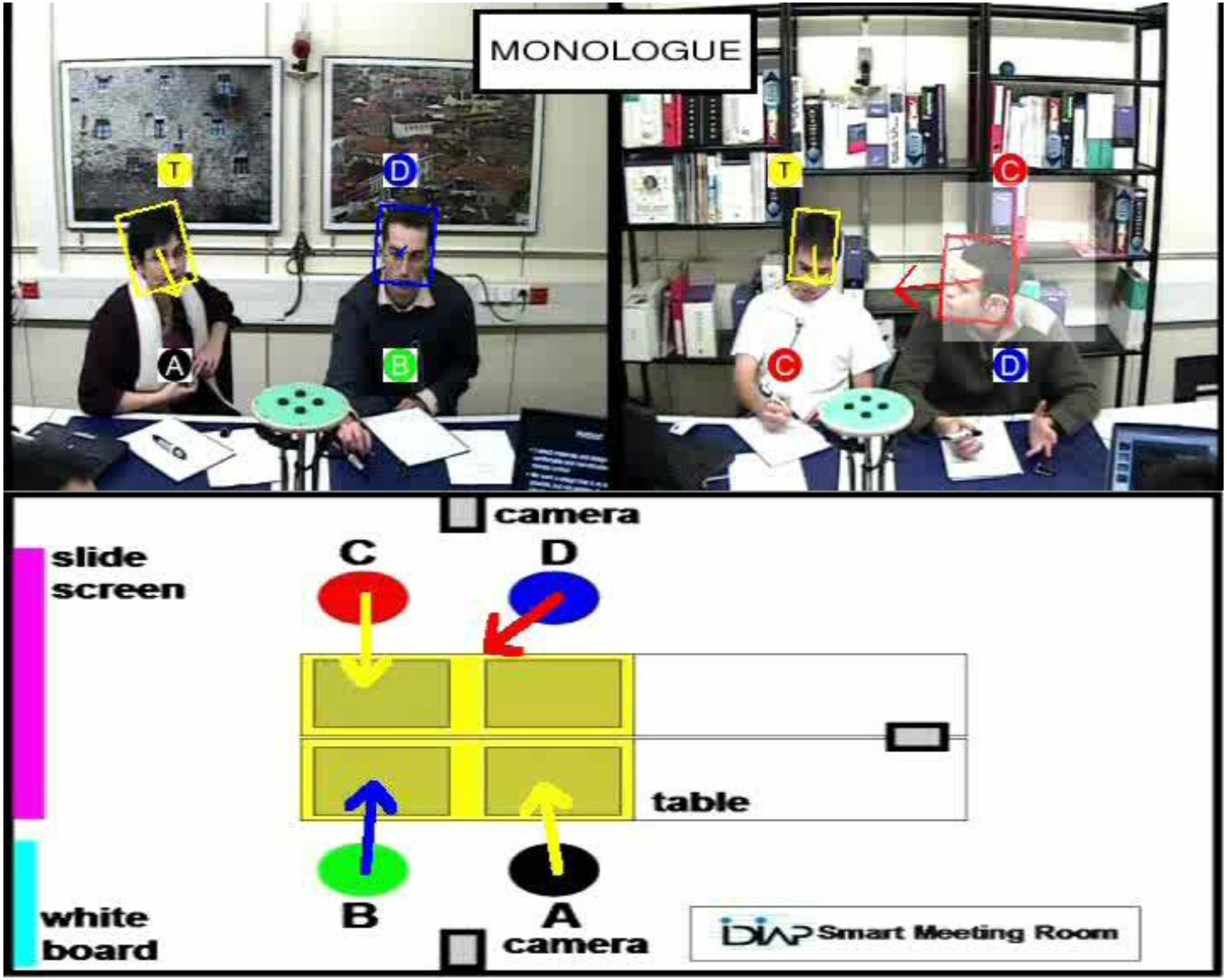
9. From Client-Server Spoken Dialog Systems to Embedded Systems
10. From Database Transactions to Problem Solving Dialogs
11. From Access to the Web of Information to the Internet of Services
12. From Generic to Personalized Voice User Interfaces
13. From Human-Machine to Human-Environment-Interaction

# Multimodal Dialog Systems

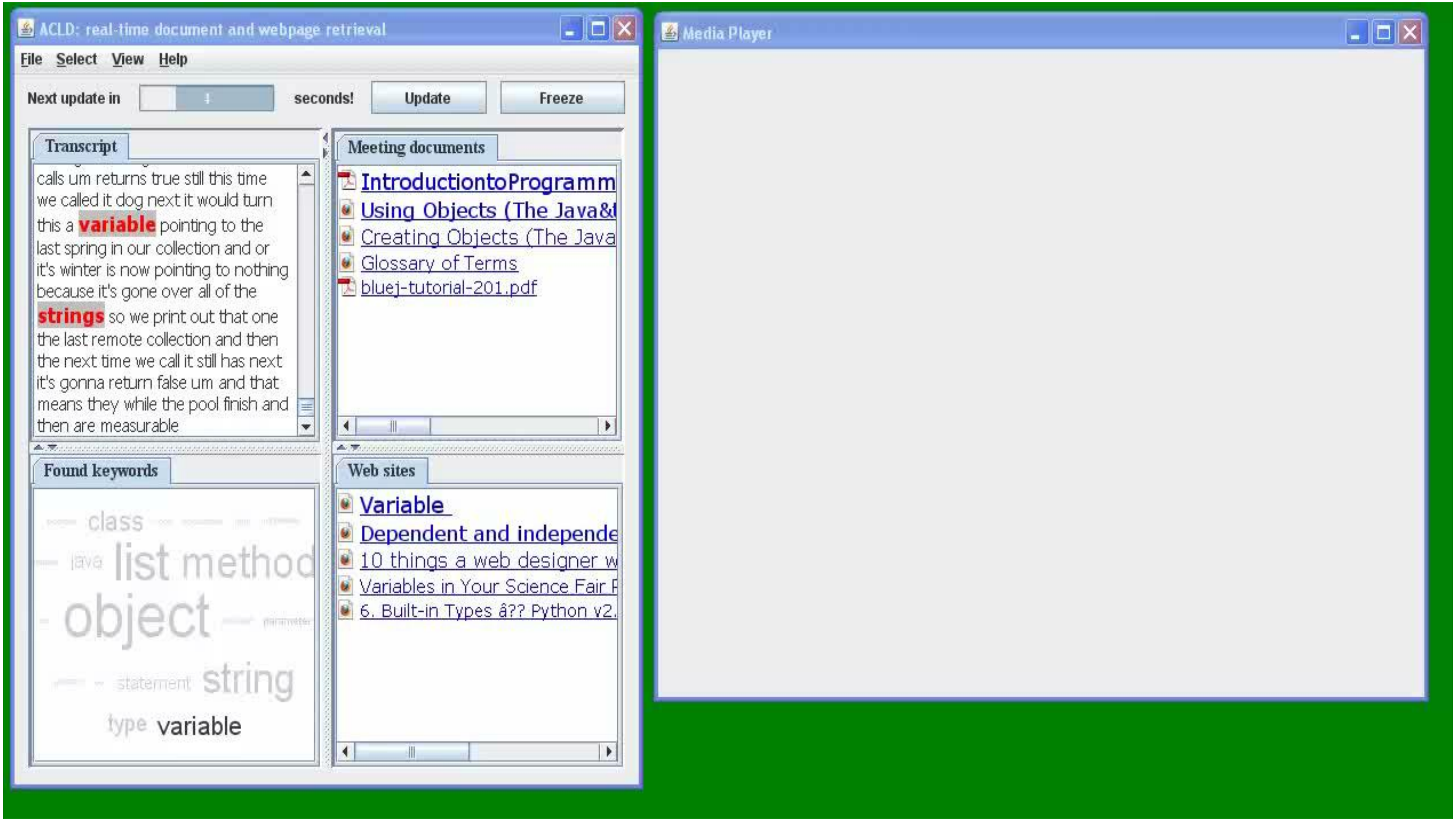


# Multiparty Speaker Diarization and Tracking
# Who spoke When, with Whom , Where about What?

**Overlapped, non-native accented and spontaneous Speech**



# Just-in-time Access to Relevant Documents or Fragments of Past Recorded Meetings



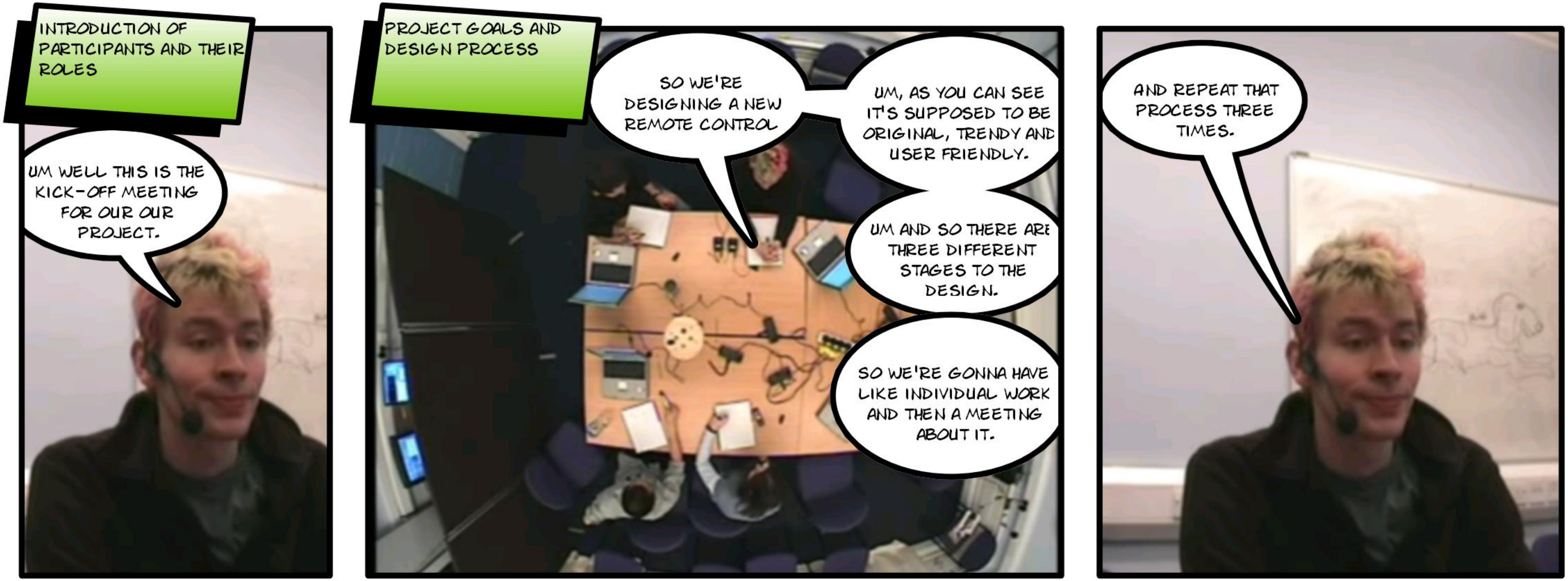
calls um returns true still this time we called it dog next it would turn this a **variable** pointing to the last spring in our collection and or it's winter is now pointing to nothing because it's gone over all of the **strings** so we print out that one the last remote collection and then the next time we call it still has next it's gonna return false um and that means they while the pool finish and then are measurable

- IntroductiontoProgramm
- Using Objects (The Java&
- Creating Objects (The Java
- Glossary of Terms
- bluej-tutorial-201.pdf

- Variable
- Dependent and independe
- 10 things a web designer w
- Variables in Your Science Fair F
- 6. Built-in Types â?? Python v2.

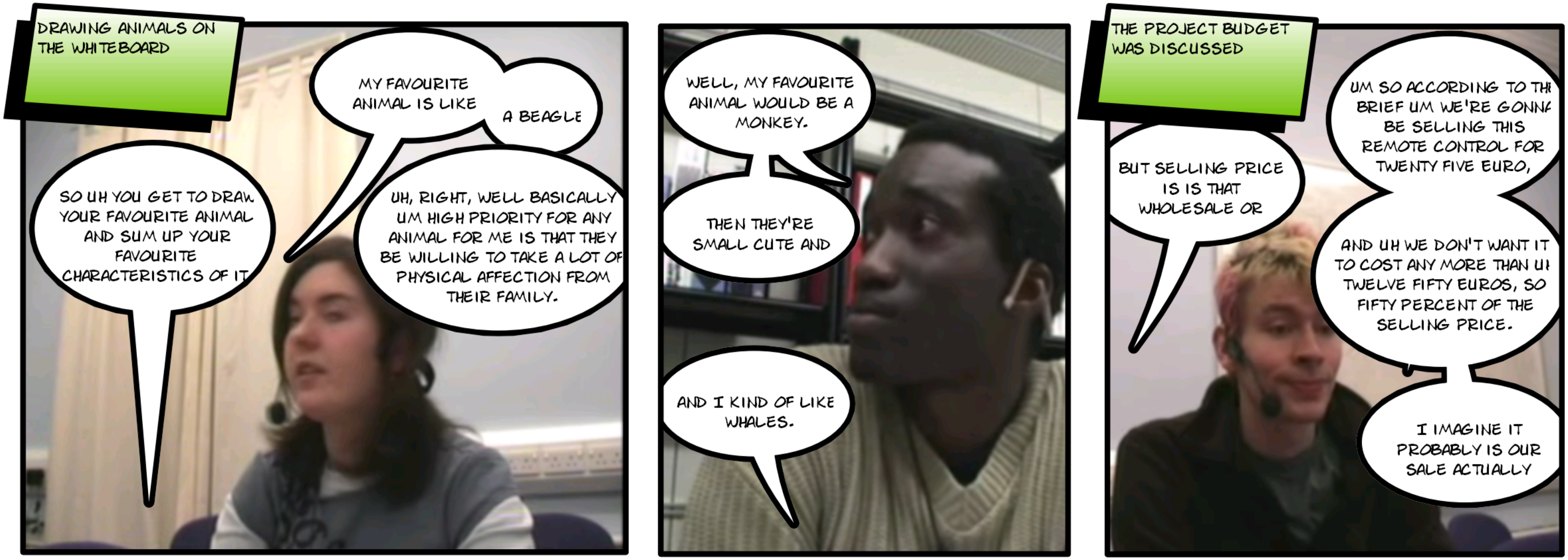
**Killer App for Call Centers: Just-in-Time Answer Retrieval during the Conversation between an Agent and a Customer by Parallel Speech Understanding**

# SuVi: The Generation of Meeting Summaries as Story Boards in Cartoon Style


INTRODUCTION OF PARTICIPANTS AND THEIR ROLES
UM WELL THIS IS THE KICK-OFF MEETING FOR OUR OUR PROJECT.
PROJECT GOALS AND DESIGN PROCESS
SO WE'RE DESIGNING A NEW REMOTE CONTROL
UM, AS YOU CAN SEE IT'S SUPPOSED TO BE ORIGINAL, TRENDY AND USER FRIENDLY.
UM AND SO THERE ARE THREE DIFFERENT STAGES TO THE DESIGN.
SO WE'RE GONNA HAVE LIKE INDIVIDUAL WORK AND THEN A MEETING ABOUT IT.
AND REPEAT THAT PROCESS THREE TIMES.

**Still pictures extracted from video capture, cartoon-style speech balloons for spoken dialog contributions and text boxes for the results of topic detection**

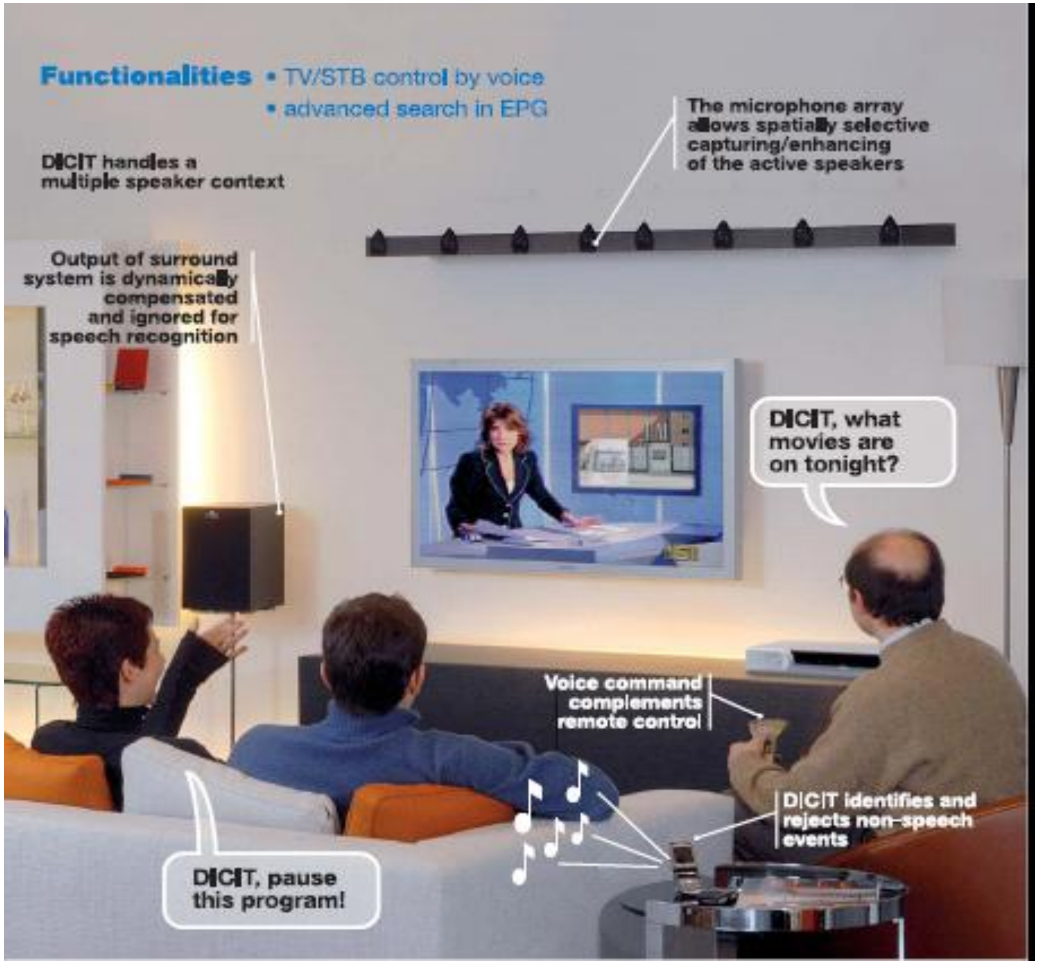
# SuVi: The Generation of Meeting Summaries as Story Boards in Cartoon Style


DRAWING ANIMALS ON THE WHITEBOARD
MY FAVOURITE ANIMAL IS LIKE
A BEAGLE
SO UH YOU GET TO DRAW YOUR FAVOURITE ANIMAL AND SUM UP YOUR FAVOURITE CHARACTERISTICS OF IT
UH, RIGHT, WELL BASICALLY UM HIGH PRIORITY FOR ANY ANIMAL FOR ME IS THAT THEY BE WILLING TO TAKE A LOT OF PHYSICAL AFFECTION FROM THEIR FAMILY.
WELL, MY FAVOURITE ANIMAL WOULD BE A MONKEY.
THEN THEY'RE SMALL CUTE AND
AND I KIND OF LIKE WHALES.
THE PROJECT BUDGET WAS DISCUSSED
UM SO ACCORDING TO THI BRIEF UM WE'RE GONNA BE SELLING THIS REMOTE CONTROL FOR TWENTY FIVE EURO,
BUT SELLING PRICE IS IS THAT WHOLESALE OR
AND UH WE DON'T WANT IT TO COST ANY MORE THAN UI TWELVE FIFTY EUROS, SO FIFTY PERCENT OF THE SELLING PRICE.
I IMAGINE IT PROBABLY IS OUR SALE ACTUALLY

**Still pictures extracted from video capture, cartoon-style speech balloons for spoken dialog contributions and text boxes for the results of topic detection**

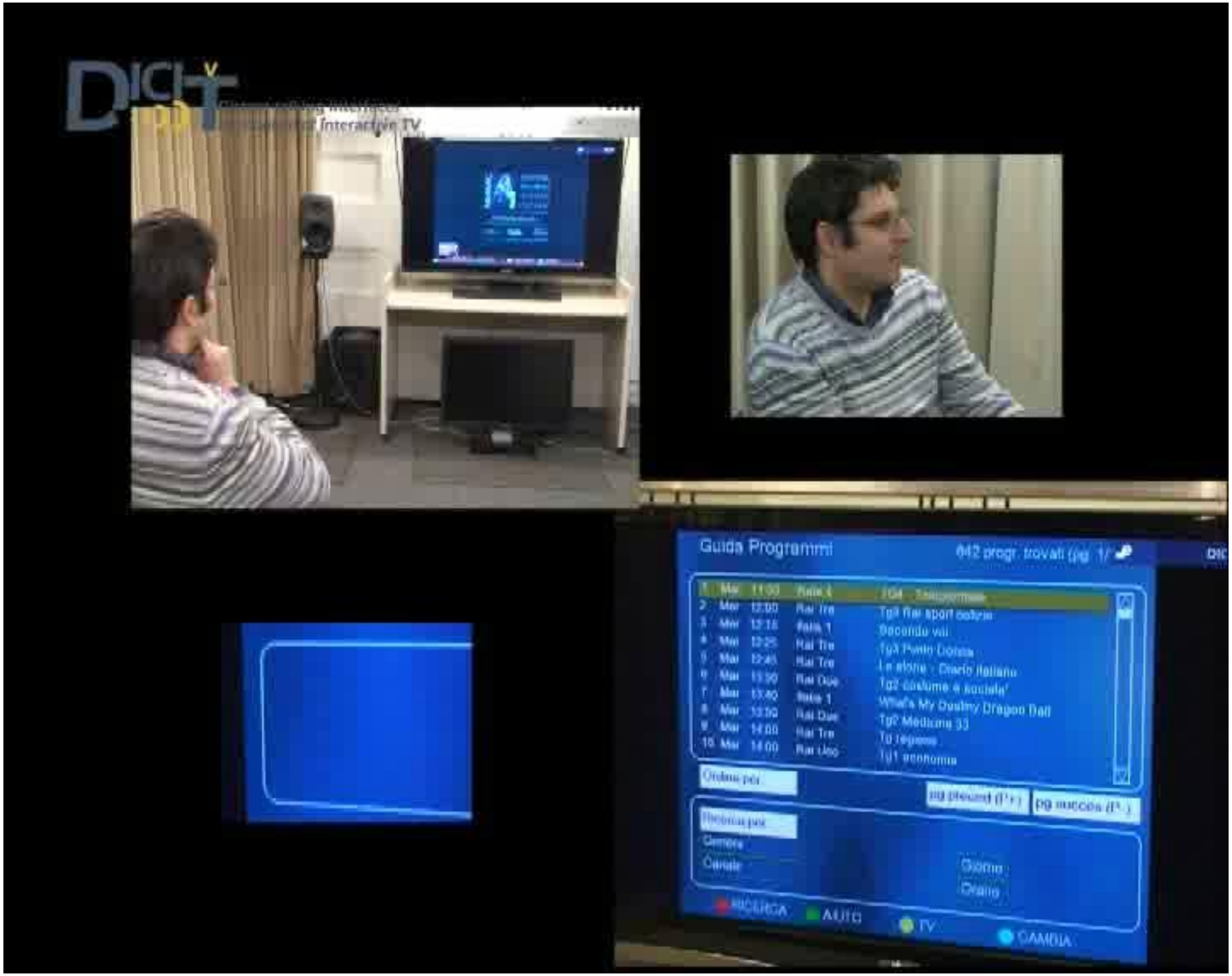
# DICIT (Distant-talking Interfaces for Control of Interactive TV) EC project

- Coordinated by FBK
- Goal: voice control of TV and related devices
- Robustness against noise and audio interferences
- Smart processing also including a real-time multi-speaker localization
- Understanding of voice input queries
- Multimodal spoken dialog management


Functionalities • TV/STB control by voice
• advanced search in EPG
DICIT handles a multiple speaker context
The microphone array allows spatially selective capturing/enhancing of the active speakers
Output of surround system is dynamically compensated and ignored for speech recognition
DICIT, what movies are on tonight?
Voice command complements remote control
DICIT identifies and rejects non-speech events
DICIT, pause this program!

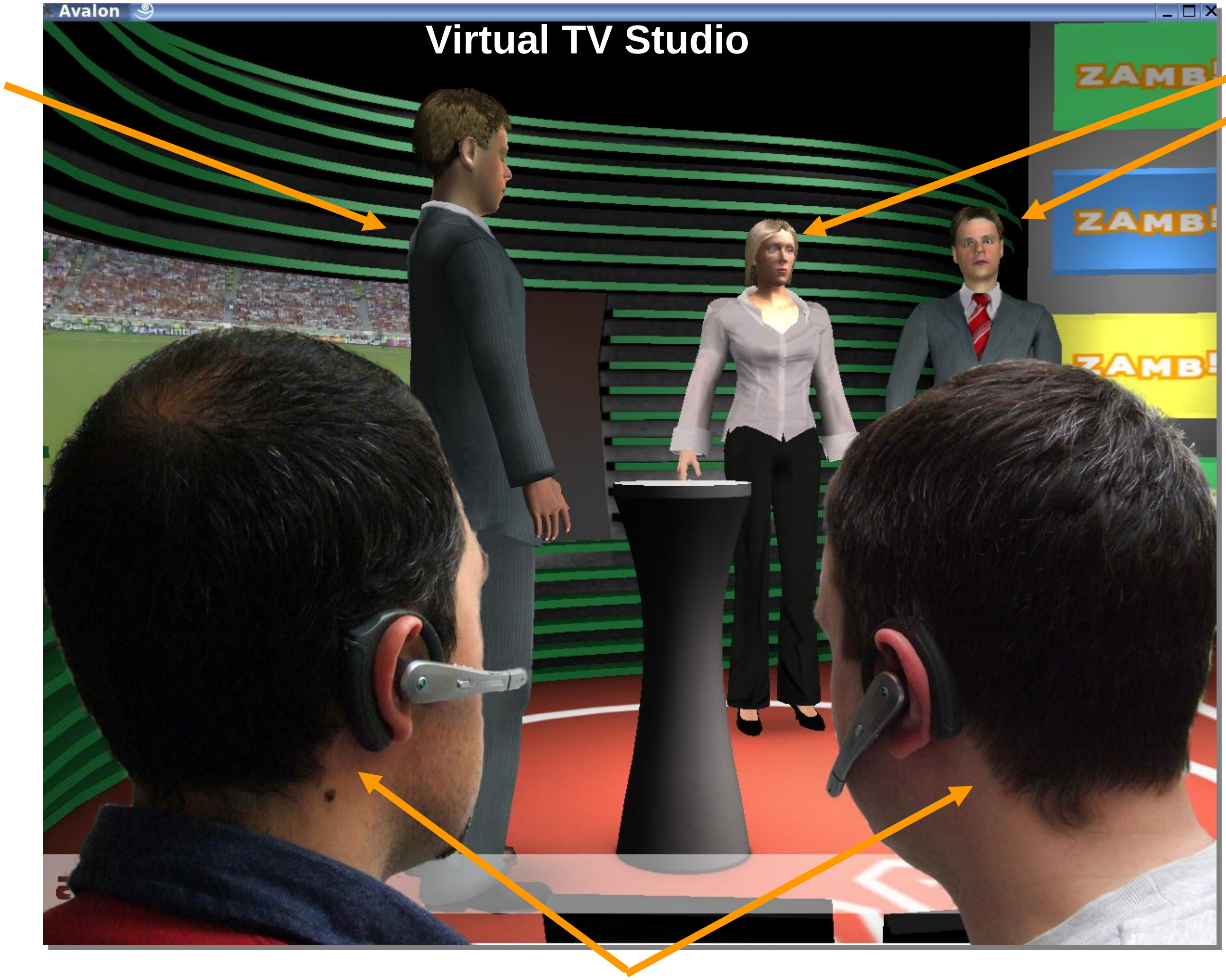
For more details: http://dicit.fbk.eu

# Multilingual Access to a Electronic Program Guide (EPG) with Distant Speech



| | | | | |
|---|---|---|---|---|
| 1 | Mar | 11:00 | Rete 4 | TG4 - Telegiornale |
| 2 | Mar | 12:00 | Rai Tre | Tg3 Rai sport notizie |
| 3 | Mar | 12:15 | Rete 4 | Secondo voi |
| 4 | Mar | 12:25 | Rai Tre | Tg3 Punto Donna |
| 5 | Mar | 12:45 | Rai Tre | Le storie - Diario italiano |
| 6 | Mar | 13:00 | Rai Due | Tg2 costume e societa' |
| 7 | Mar | 13:40 | Italia 1 | What's My Destiny Dragon Ball |
| 8 | Mar | 13:50 | Rai Due | Tg2 Medicina 33 |
| 9 | Mar | 14:00 | Rai Tre | Tg regione |
| 10 | Mar | 14:00 | Rai Uno | Tg1 economia |

# Multiparty Dialog between Virtual & Human Football Experts: Discussing the UEFA EURO 2016 in France



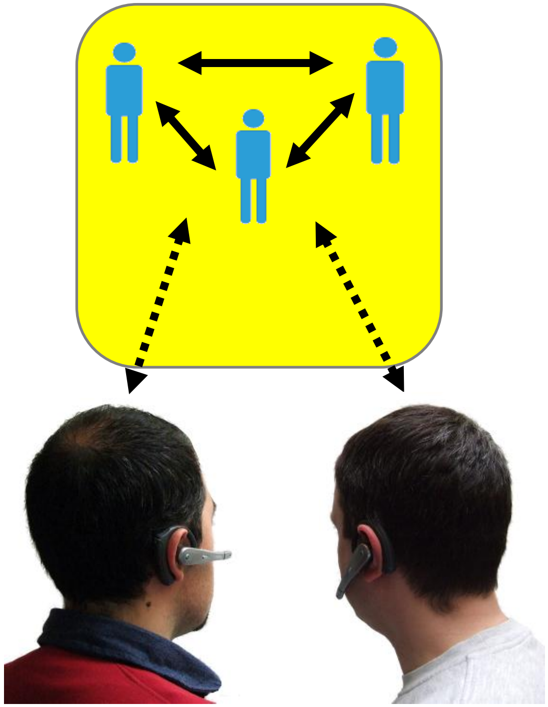
Multilingual Virtual Moderator

N > 2 Virtual Experts

n > 2 Human Football Fans from Different EU Member States Speaking their Mother Tongues

# Discussing the Best of European Football in Your Mother Tongue 2016

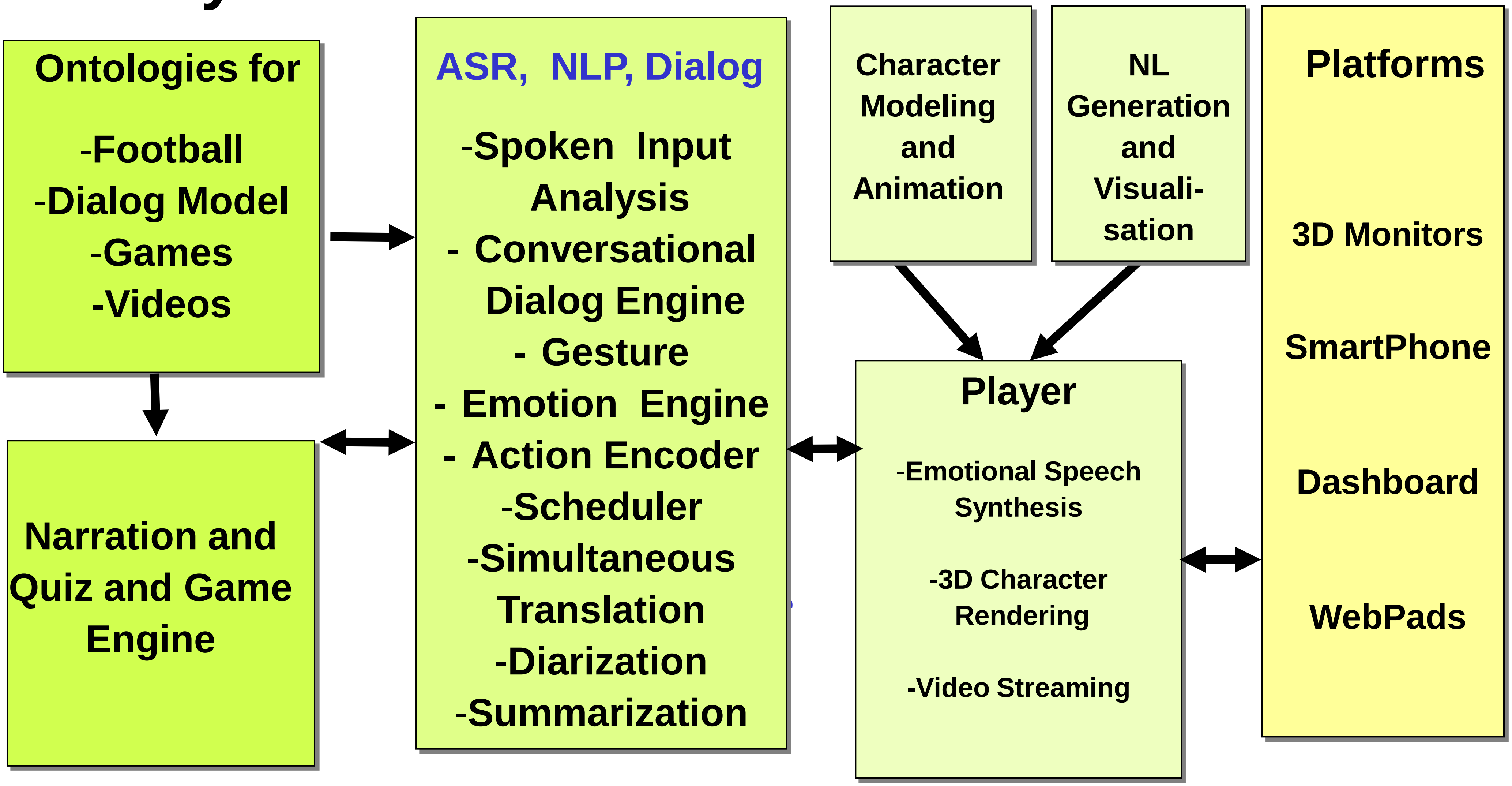
multiparty spoken dialog between virtual & human agents



1. on your mobile with football fans from all over Europe
2. with spontaneous speech translation, diarization, simultaneous cross-lingual multimodal content linking
3. 24 languages of 24 European teams
4. quiz and game shows, defining your own teams, virtual coaching

**multimodal, multilingual, multiparty, multitask**

# The Basic Architecture of the 4M EURO 2016 System



Ontologies for

-Football

-Dialog Model

-Games

-Videos

Narration and Quiz and Game Engine

ASR, NLP, Dialog

-Spoken Input Analysis

- Conversational Dialog Engine

- Gesture

- Emotion Engine

- Action Encoder

-Scheduler

-Simultaneous Translation

-Diarization

-Summarization

Character Modeling and Animation

NL Generation and Visuali-sation

Player

-Emotional Speech Synthesis

-3D Character Rendering

-Video Streaming

Platforms

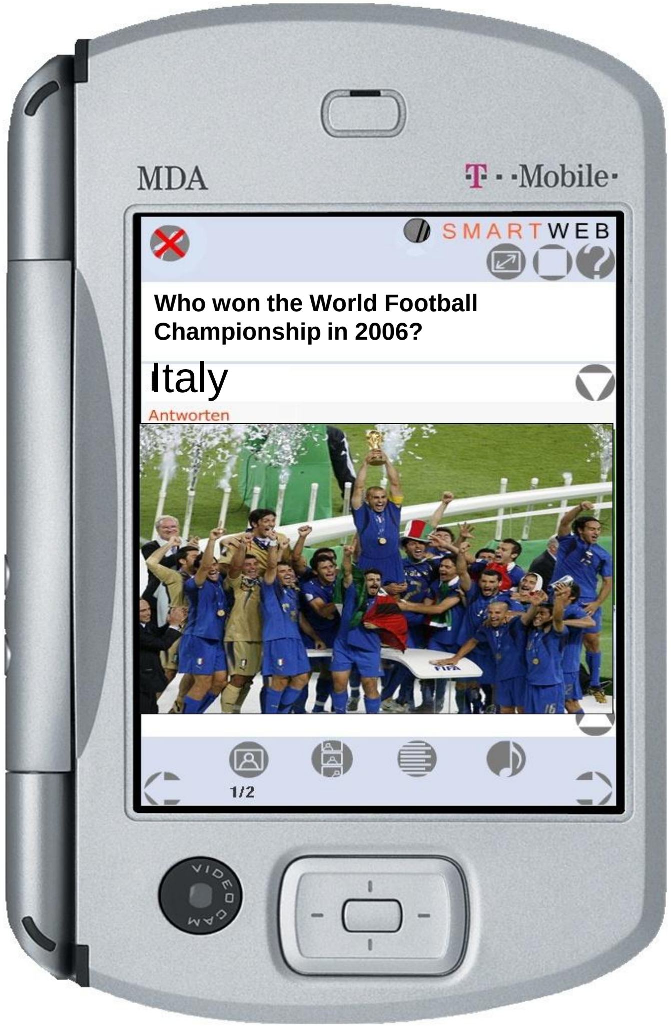
3D Monitors

SmartPhone

Dashboard

WebPads

# SmartWeb: Getting Answers on the Go



Who won the World Football Championship in 2006?

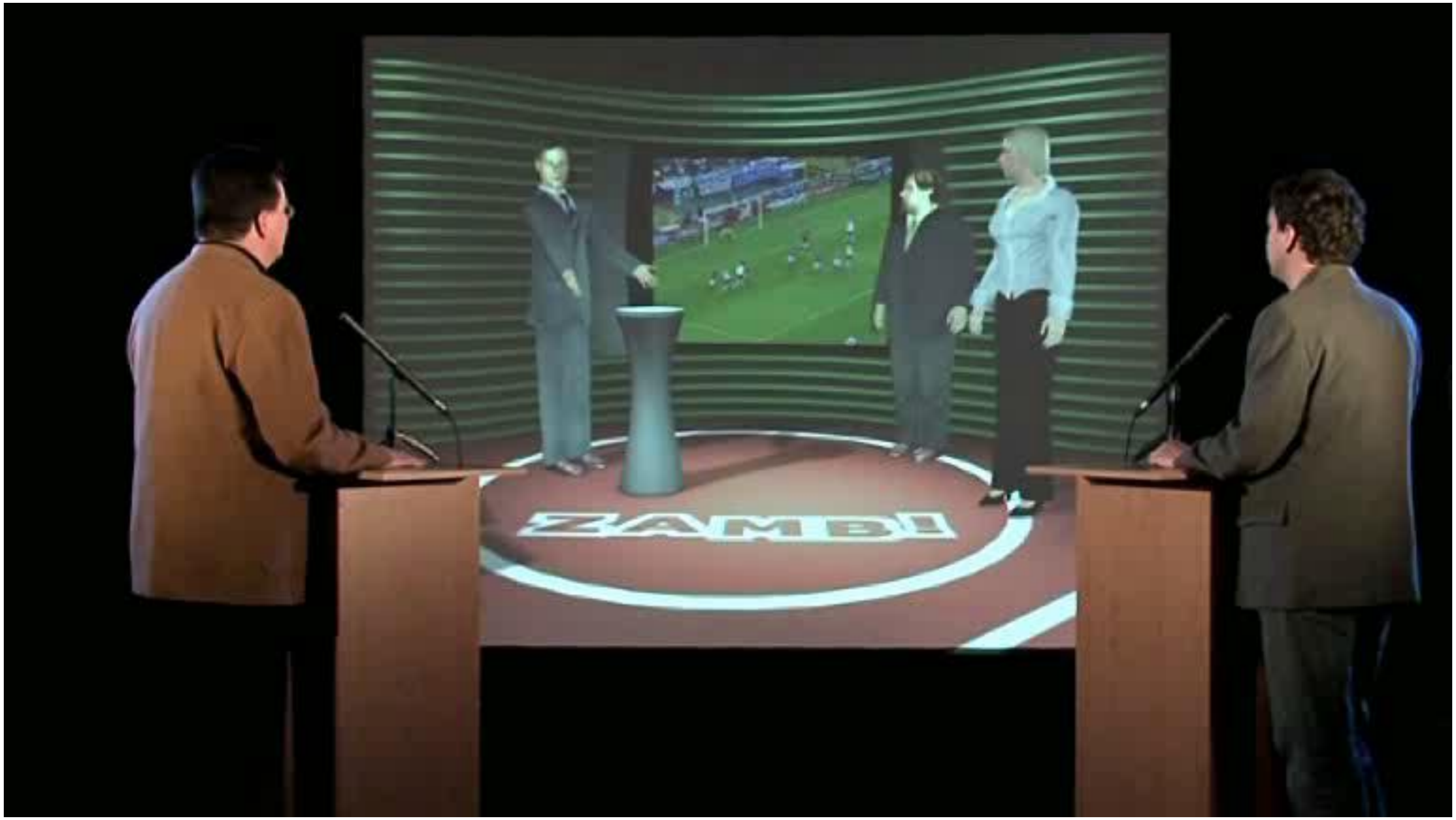
Italy

Antworten

Italy

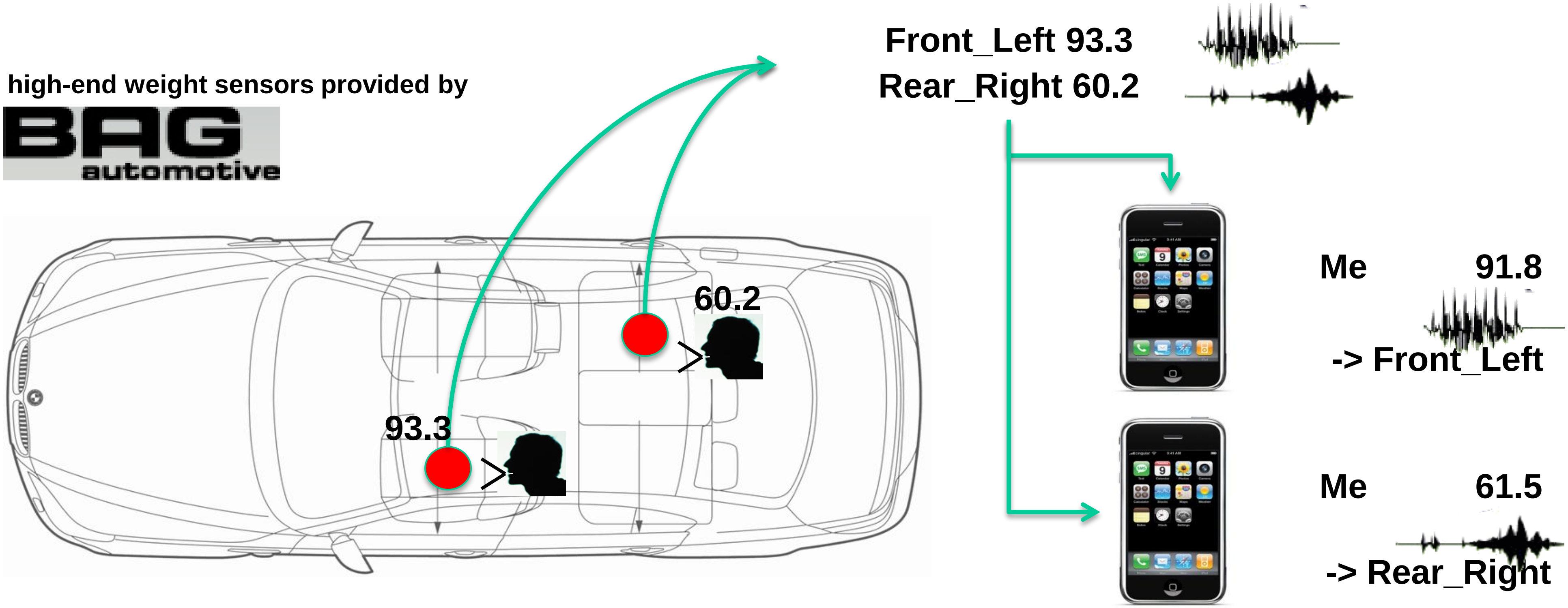
Personal guide for the FIFA world cup

# Monolingual Multiparty Football Quiz and Game Show at DFKI



# Multitask Games with Multimodal Dialogs

# Multimodal Computing: Speech and Car Sensors



high-end weight sensors provided by

BAG automotive

Front_Left 93.3

Rear_Right 60.2

Me 91.8

-> Front_Left

Me 61.5

-> Rear_Right

- Weight sensors and microphones in the car take measures / capture speech on the respective seats
- Values and speech features are broadcasted and received by personal (nomadic devices)
- Speaker models and weights are stored on personal device.
- Personal devices “decode” the position information and decide, which service is allowed to use it

# Multiparty Conversation and Speaker Identification in the Car

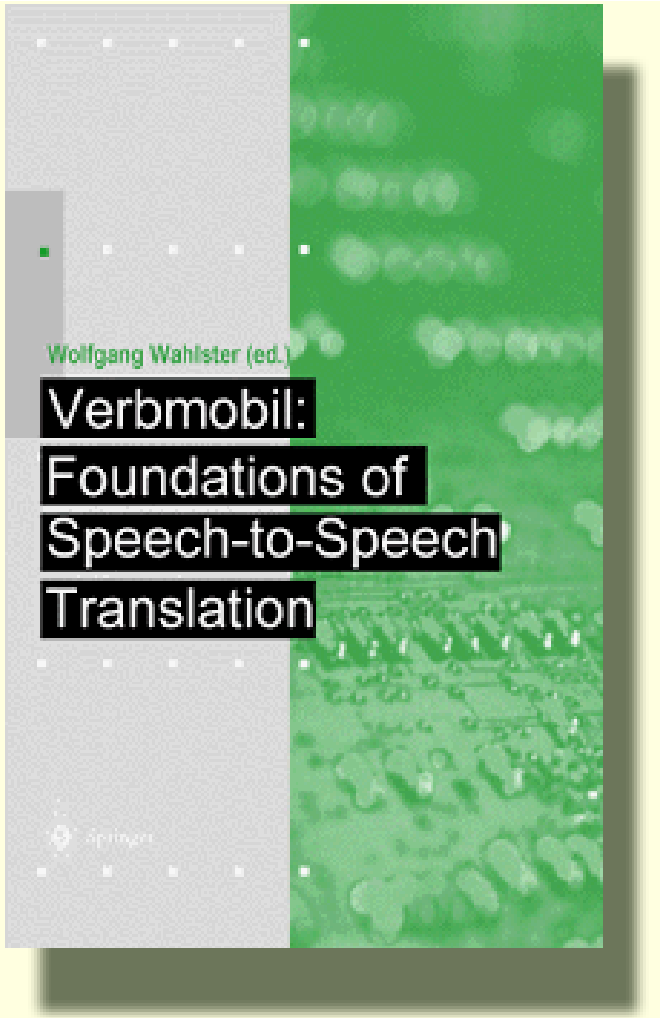
# There Are Many Open Problems for the Next 6 Years:

- Integrating top-down context and dialog knowledge into low- level speech recognition processes

- Exploiting more knowledge about human communication and translation strategies including psycho- and neurolinguistic inspirations.

- Avoiding expensive data collections and cognitively unrealistic training data for machine learning.

# 10 Years after Verbmobil +
# 5 Years after SmartKom/SmartWeb



Wolfgang Wahlster (ed.)

Verbmobil: Foundations of Speech-to-Speech Translation

Springer

+

Wolfgang Wahlster (Ed.)

SmartKom: Foundations of Multimodal Dialogue Systems

Springer

**15 and 16 November 2010, Saarbrücken: 10 Years Verbmobil**

**Looking Back and Looking Ahead**

# Football Tournaments Create Emotions: Emotional Speech, Emotional Facial Expressions

ZAMB!
ZAMB!
MB!

# Realistic Facial Expressions combined with Emotional Speech Synthesis

## Jules: the Robotic Head by Hanson Robotics used by a Team at the University of Bristol

# Thank you very much for your attention.

Design by R.O.