Challenges in Spoken and Multimodal Dialogue Systems: Architectures, Components, Tasks

Massimo Romanelli
romanell@dfki.de
Overview

• Generic Multimodal Interfaces

• Case Studies:
  – W3C MMI (Stress on Protocols and Communication)
  – Dialog Backbone (Stress on Discourse/Fusion)
  – Smartweb (Stress on Interaction with Web Data)
  – Theseus (Stress on Presentation)
  – IKS Aml Case (Stress on Sensors & Devices)
Overview

• Generic Multimodal Interfaces

• Case Studies:
  – W3C MMI (Stress on Protocols and Communication)
  – Dialog Backbone (Stress on Discourse/Fusion)
  – Smartweb (Stress on Interaction with Web Data)
  – Theseus (Stress on Presentation)
  – IKS AmI Case (Stress on Sensors & Devices)
Generic Multimodal Interface

Dagstuhl Seminar on Coordination and Fusion in Multimodal, Working Group 3 Software Architectures for Multimodal Systems (2001)

Generic Multimodal Interface
Generic Multimodal Interface
Generic Multimodal Interface
Generic Multimodal Interface
Generic Multimodal Interface

[Diagram showing the components of a generic multimodal interface, including media input processing, media mode analysis, multimodal fusion, mode coordination, interaction management, discourse management, context management, lexicon management, and intention recognition.]

User(s) -> Media Input Processing -> Media/Mode Analysis (Language, Graphics, Gesture, Sound) -> Presentation Design (Select Content, Design, Allocate, Coordinate, Layout) -> User Modeling (User ID) -> Application Interface (Initiate, Terminate, Request, Respond, Integrate)
Generic Multimodal Interface
Generic Multimodal Interface
Generic Multimodal Interface
Generic Multimodal Interface
Overview

• Generic Multimodal Interfaces

• Case Studies:
  – W3C MMI (Stress on Protocols and Communication)
  – Dialog Backbone (Stress on Discourse/Fusion)
  – Smartweb (Stress on Interaction with Web Data)
  – Theseus (Stress on Presentation)
  – IKS Aml Case (Stress on Sensors & Devices)
W3C: MMI-Framework
Protocols and Communication (W3C)

Input/Output
- SRGS: “Speech Recognition Grammar Specification”, (Hunt and McGlashan, 2004);
- SISR: “Semantic Interpretation for Speech Recognition” (Van Tichelen and Burke, 2007);
- SSML: “Speech Synthesis Markup Language” (Burnett et al., 2004);

Management
- MRCP: “Media Resource Control Protocol”, (Shanmugham et al., 2006).
- CCXML: “Voice Browser Call Control eXtensible Markup Language”, (Auburn, 2007);
- SCXML: “State Chart XML state machine notation for control abstraction”, (Barnett et al., 2009);
- EMMA: "Extensible multimodal Annotation markup language", (Michael Johnson et al., 2009)

Presentation
- SMIL 2.1: "Synchronized Multimedia Integration Language”, (Bulterman et al., 2005).
- XHTML 1.0: "The Extensible HyperText Markup Language", (Pemberton et al., 2004).
Overview

• Generic Multimodal Interfaces

• Case Studies:
  – W3C MMI (Stress on Protocols and Communication)
  – Dialog Backbone (Stress on Discourse/Fusion)
  – Smartweb (Stress on Interaction with Web Data)
  – Theseus (Stress on Presentation)
  – IKS Aml Case (Stress on Sensors & Devices)
Seminar (SS11): Challenges in Spoken and Multimodal Dialogue Systems

Dialogue Back-Bone

Analysers
- Speech
- Gesture
- ...

Modality Fusion

Discourse Modelling DiM

Dialogue Manager DM

Generators
- Speech
- Graphics
- Gesture

Presentation Manager

External Services

Communication pools
Main data flow
Context information
Dialog Backbone

- Lattice of user intentions hypotheses (MF)
- Compare and enrich with a selected number of discourse states (discourse history)
- Determination of steps for goals reaching
- Produce multimodal result
Example for Overlay

User: "What films are on TV tonight?"

System: [presents list of films]

User: "That's a boring program, I'd rather go to the movies."

How do we inherit “tonight”?
• Non-monotonic unification-like operation
• Inherit (non-conflicting) background information
  – if background and covering both contain information recursion
• two sources of conflicts:
  – conflicting atomic values
    ➔ overwrite background with covering
  – type clash
    ➔ assimilate background to the type of covering; recursion
Type Hierarchy of TFS

- **Entertainment**
  - `beginTime` : Time
  - `title` : String

- **Broadcast**
  - `channel` : Channel

- **Performance**
  - `cinema` : Cinema

- **A named entertainment at some time**
- **A named TV program at some time on some channel**
- **A named Movie at some time at some cinema**
Unification Simulation

Films on TV tonight
- Broadcast
  - beginTime
    - "tonight"
  - channel: any

Go to the movies
- Performance
  - beginTime

Fail – type clash
Overlay Simulation

Films on TV tonight

Broadcast

startTime : Time

channel : Channel

Performance

startTime : Time

tonight

channel : any

Performance

startTime : Time

tonight

cinema : any

Performance

startTime : Time

tonight

cinema : any

Assimilation

Entertainment

startTime : Time

Title : String
Overview

- Generic Multimodal Interfaces

- Case Studies:
  - W3C MMI (Stress on Protocols and Communication)
  - Dialog Backbone (Stress on Discourse/Fusion)
  - Smartweb (Stress on Interaction with Web Data)
  - Theseus (Stress on Presentation)
  - IKS Aml Case (Stress on Sensors & Devices)
SMARTWEB General Architecture

PDA Client
- Local VXML
- Local Control
- GUI Control
- Camera
- GPS

Dialog Server
- Multi-Modal Recog.
- Speech Synthesis
- Multimodal Dialog Manager
- GEN
- REAPR
- SPIN
- IHUB
- FADE
- EUP

Semantic Web
- Q&A System
- Web Services
- Agent Access
- Knowledge Server
System/User Feedback

ASR & Correction

Semantic Paraphrase

Werner war 1990 Weltmeister?

Suche nach:

Weltmeister Mannschaft oder Land
Im Jahr 1990
in der (Sportart) Fußball der Männer

Concept Icons
Pattern language for QA Interaction Design

Interaction Design Patterns

- Progressive Disclosure
  - Incremental Display
  - More active correction mode
  - Display unrecognised, low confidence tokens
- Correction Possibilities
  - Flash Query Editor
- Layout Patterns
  - Content Frame
    - Display Paraphrase
    - Display Result
    - Factoid Answer
  - Additional multimodal information
  - Multimodal Answer
Pattern language for QA Interaction Design

- Inform
- Disambiguate
- Present

Interaction Design Patterns

- Progressive Disclosure
  - Incremental Display
  - More active correction mode
  - Display unrecognised, low confidence tokens
  - Additional multimodal information

- Correction Possibilities
  - Flash Query Editor
  - Display Paraphrase
  - Factoid Answer

- Layout Patterns
  - Content Frame
  - Display Result
  - Multimodal Answer
Interaction System Behavior

- plan interactive behavior autonomously
- can initiate interaction

- respond immediately to interruptions
- clarification questions

- anticipate the user's needs
- adopt the user's goals
- provide unsolicited comments & suggestions
Overview

• Generic Multimodal Interfaces

• Case Studies:
  – W3C MMI (Stress on Protocols and Communication)
  – Dialog Backbone (Stress on Discourse/Fusion)
  – Smartweb (Stress on Interaction with Web Data)
  – Theseus (Stress on Presentation)
  – IKS Aml Case (Stress on Sensors & Devices)
THESEUS General Architecture


Example application: BabbleTunes:
Concepts for RDF Stores Presentation

Arc view

Nobel Prize in physics 1921

3D Cube

Search

Word Tree

Star constellation view

Einstein

was a German-born theoretical physicist.

received the 1921 Nobel Prize in Physics

published over 300 scientific works and over 150 non-scientific works

and Mileva Maric had a daughter, Lieserl Einstein, born in early 1902

published "Über die Entwicklung unserer Anschauungen über das Wesen und die Konstitution der Strahlung" ("The Development of Our Views on the Composition and Essence of Radiation"),

became an associate professor at the University of Zurich.

traveled to New York City in the United States for the first time on 2 April 1921

published a paper in Nature in 1940 entitled "Science and Religion"
MEDICO Storyboard
Overview

• Generic Multimodal Interfaces

• Case Studies:
  – W3C MMI (Stress on Protocols and Communication)
  – Dialog Backbone (Stress on Discourse/Fusion)
  – Smartweb (Stress on Interaction with Web Data)
  – Theseus (Stress on Presentation)
  – IKS AmI Case (Stress on Sensors & Devices)
IKS Aml Case
IKS AmI Case: Interactive Points

(1) IK Point Mirror

(2) IK Point eScreen

(3) IK Point Shower
IKS AmI Case: Architecture
Processing of Content in IKS AmI Case

Context: Monday morning, 7 a.m., user in front of mirror

(1) gets sensor data
(2) broadcasts sensor results
(3) broadcasts context change
(4) Requests situational part of KR (Pre-Artifacts)
(5) Selects type of situation
(6) Broadcasts type of situation → weather information required
(7) Requests generic weather information
(8) stores weather information in contextual part of KR (AmI ODPs)
(9) Requests weather information for specific user and location
(10) Presentation of weather on mirror
Thank you 😊