Adapting Spoken and Visual Output for a Pedestrian Navigation System, based on given Situational Statements

Saarland University
Rainer Wasinger, Dominika Oliver, Dominik Heckmann,
Bettina Braun, Boris Brandherm, Christoph Stahl

Presented by Rainer Wasinger

German Research Center for Artificial Intelligence GmbH

UNIVERSITÄT DES SAARLANDES

Motivation

- Pedestrian Navigation System
- User
  - Changing needs:
    - time pressure, cognitive load.
  - Environment, dynamically changing:
    - indoor/outdoor, noisy/quiet, rainy/dry.
  - Devices, differing:
    - Knowledge databases, hardware characteristics (e.g. processing power, display size).

- What is needed?
  - ”User adapted visual and audio output”.

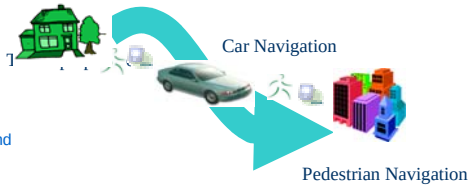
What this talk is on?

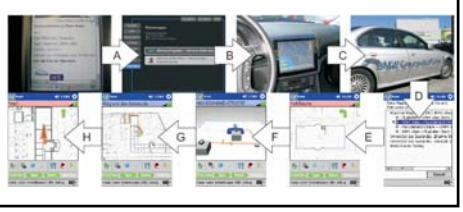
- Adapting mobile device output, to create user-tailored presentations.
- Identifying input sources:
  - User, mobile device, environment.
- Identifying media output parameters:
  - Speech, graphics, sound, and text.
- Proposed architecutre for such a system.
- Two examples illustrating our initial insight into the modelling of causal relationships between input and output parameters.

Pedestrian Navigation System

Mobile Multimodal Interaction (M3I) - Indoor and outdoor pedestrian navigation and exploration.

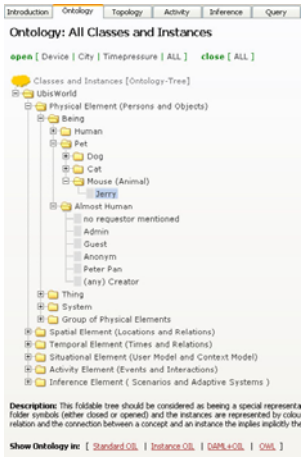
- Target users:
  - Business users, tourists.
- Input modalities:
  - Speech-gesture combined interaction.
- Output modalities:
  - Speech (e.g. instructions, descriptions), and graphics (e.g. street and landmark maps).
  - Sounds and text.
- Supporting infrastructure (aside from the PocketPC):
  - Snavigation server, GPS, IR beacons, mobile phone.
  - Proposed: Usermodelling & BN server.
- Supporting software (for the PocketPC):
  - Graphics: ParallelGraphics' Cortona VRML viewer.
  - Speech: IBM's Embedded ViaVoice.





Adaptation Input

- Modelled in Ubisworld.
- Statements and queries are represented uniformly as Situational ML.
- The data is currently statically defined, but we do hope to incorporate techniques for updating input data dynamically in the future.
  - Less relevant for static variables (e.g. gender), and more relevant for dynamic variables (e.g. emotions).



Adaptation Input – User Model

User modelling:

- Role of user: tourist, business person.
- Age: young, middle-aged, elderly.
- Gender: male, female.
- Native language: german, english.
- Walking speed: slow, normal, fast.
- Eyesight and hearing.
- Emotions: indifferent, angry, distressed, happy.
- Cognitive load: low, medium, high.
- Time pressure: low, medium, high.
- Social environment: ability to look at PDA, lookin at PDA.
- User interests: shopping, travel, destination.
- Preferences

Adaptation Input – Device & Environment Modelling

Device modelling:

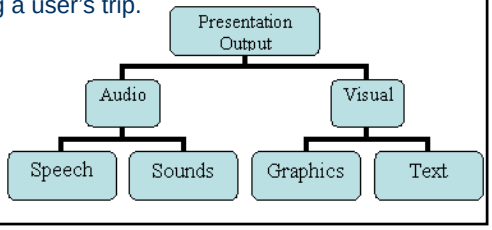
- Remaining battery: 2 percent, 30 percent, 70 percent.
- GPS coverage: 1to2, 3to4, 5to6.
- Screen size: desktop, PDA, phone.

Environment modelling:

- Noise level: low, medium, high.
- Light level: low, medium, high.
- Route surface: sealed, non-sealed, grass.
- People density: low, medium, high.
- Weather conditions: indoor, outdoor sun, outdoor rain, outdoor wind.

Presentation Output

- 4 Sets of output parameters:
  - 2 primary sets: speech, and graphics.
  - 2 support sets: sounds, and text.
- Can aid presentation *generally*, i.e. for all user groups, but in varying environment contexts (e.g. emphasis on keywords), or for a single user group *specifically* (e.g. volume for elderly).
- Goal: Improve quality and intelligibility of presentation, for given contexts, and for different locations along a user's trip.



Presentation Output Parameters

Speech:

- Language: german, english_UK, english_US.
- Language formality: formal, informal.
- Voice gender: male, female.
- Voice chars, pitch: low, middle, high.
- Voice chars, speed: slow, medium, fast, very fast.
- Voice chars, volume: quiet, medium, loud.
- Prosody, emphasis: none, normal, heavy.
- Prosody, pauses: small, medium, long.
- Speaking style: whisper, normal.

Presentation Output Parameters

Graphics

- Zoom: zoomed in, not zoomed, zoomed out.
- Filter: landmarks, POIs, streets.
- Toolbar size: small, middle, large.
- Colour scheme: normal, maximum contrast.
- Menu scheme: beginner, advanced.

Sound

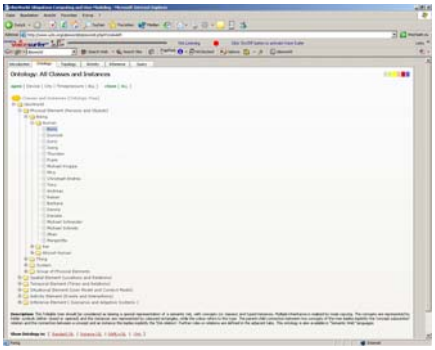
- Landmark: yes, no.
- Distance: far, middle, close, off.

Text

- Font size: small, medium, large.
- Presentation: scrolling, dot_ended.

Procedure

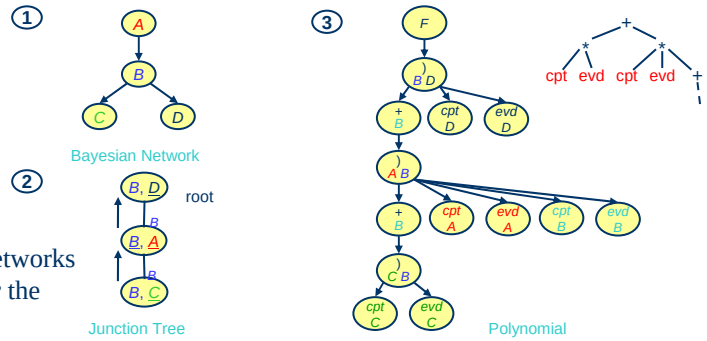
1. Parameters are statically modelled in Ubisworld. Bayesian Networks (BNs) are then also statically modelled using the same parameters.



Ubisworld

Procedure (cont.)

Idea: Compute a polynomial, that represents the Bayesian Network. This polynomial can be simplified or specialized for selected queries, so that it is suitable for small devices

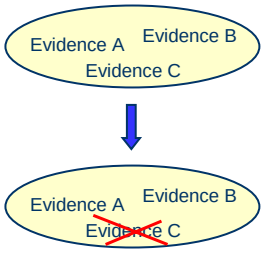


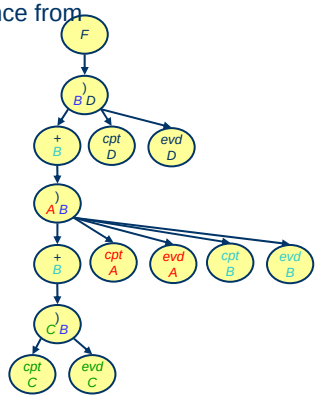
Bayesian Networks modified for the Pocket PC.

Procedure (cont.)

Taking into account impossible combinations of evidence

Suppose we have a polynomial that handles evidence from the nodes A, B and C





If we know that one type of evidence does not occur with other types of evidence, we can simplify the formulas.

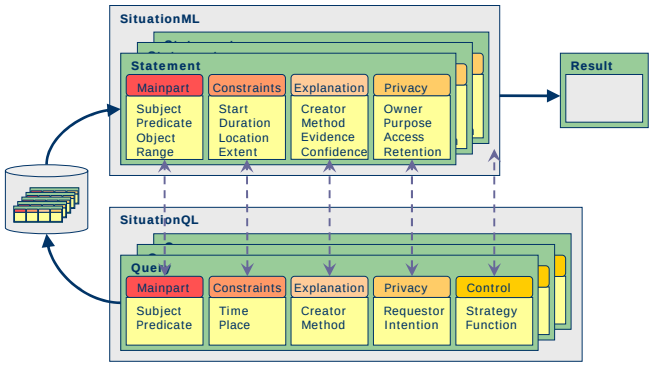
Procedure (cont.)

Step 2.

```
# REQUEST:
<?xml version="1.0" encoding="ISO-8859-1"?>
<SituationQL>
  <QUERY subject="rainer_wasinger" predicate="role"/>
  <QUERY subject="rainer_wasinger" predicate="gender"/>

</SituationQL>

# RESPONSE:
<?xml version="1.0" encoding="ISO-8859-1"?>
<SituationML>
  <STATUS>
    <STATUSCODE>0</STATUSCODE>
    <STATUSDESCRIPTION>OK</STATUSDESCRIPTION>
  </STATUS>

  <STATEMENT subject="rainer_wasinger" predicate="role" object="business"/>
  <STATEMENT subject="rainer_wasinger" predicate="gender" object="male"/>
  ... ... ...
</SituationML>
```

Situation ML

2. PPC connects to the Ubisworld service and requests values for the input variables relevant to pedestrian navigation.



Procedure (cont.)

3. PPC connects to the BN service, specifying:
   - BN (e.g. fast, accurate).
   - Current values for the known input variables.
   - Output variables for which it would like to obtain results.

4. PPC then executes the desired adaptation methods.

```
# REQUEST:
<SMLRequestForBNValues>
  <BN value="quick" subject="rainer_wasinger" service="ped_nav_sys">
    <INPUT predicate="role" object="business"></STATEMENT>
    <INPUT predicate="gender" object="male"></STATEMENT>
    ... ... ...
    <OUTPUT predicate="language">
    <OUTPUT predicate="formality">
    ... ... ...
  </BN>
</SMLRequestForBNValues>

# RESPONSE:
<SMLResponseForBNValues>
  <STATUS>
    <STATUSCODE>0</STATUSCODE>
    <STATUSDESCRIPTION>OK</STATUSDESCRIPTION>
  </STATUS>
  <BN value="quick" subject="rainer_wasinger" service="ped_nav_sys">
    <OUTPUT predicate="language" object="de">
    <OUTPUT predicate="formality" object="informal">
    ... ... ...
  </BN>
</SMLResponseForBNValues>
```

Step 3.

```
M3I.Speech_SetLanguage("de");
M3I.Speech_SetFormality("informal");
... ... ...
```

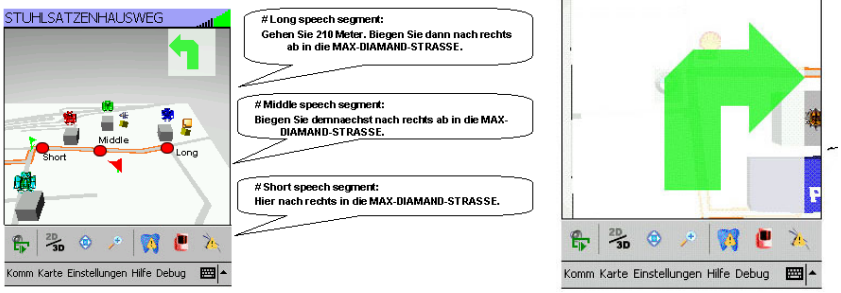
Step 4.

Adapted Output (Example 1 - General)

- Speech enhancements targeted at all user groups, in a varying environment context.
  - **Description**: “Das Gebäude 36.1 ist der Sitz des Lehrstuhls Professor Wahlster.”
    <Original> <Enhanced><Extreme>
    - Tempo of known tokens is increased (e.g. Das Gebäude), and heavy emphasis is placed on keywords (e.g. Wahlster).
  - **Instruction**: “”Gehen Sie 210 Meter, biegen Sie dann nach rechts ab in die Max-Diamand-Strasse.”
    <Original> <Enhanced><Extreme>
    - Similar to above, but the pitch is also lowered to attain a more factual tone.

Adapted Output (Example 2 - Specific)

- Audio & visual enhancements targeted at people with a high cognitive load.
  - Heavy reliance on speech output.
  - Very large graphics to show the way.



Future Work & Thankyou

- Modularize the current implementation to allow for easier manipulation of the multimodal parameters.
- Research the causal relationships between all input and output parameters.
- Conduct user studies to help model the Bayesian networks.

Questions?