Interpreting Symptoms of Cognitive Load in Speech Input

André Berthold and Anthony Jameson
Department of Computer Science
University of Saarbrücken, Germany
http://w5.cs.uni-sb.de/~ready/ (Slides, etc.)

Table of Contents

3

Overview

Content

Why can it be important to recognize cognitive load?

What features of speech input are indicators of cognitive load?

Methodology

How can various kinds of empirical data be combined in the development of a user modeling component?

How can you evaluate the *data-limitedness* of a user modeling component?

Problem and Approach Taken

Why Recognize Cognitive Load?

4

Characterization of present situation

Primary task

Using Acrobat Reader on laptop

Secondary task

Using Remote Commander on PalmPilot

Situational distraction

Giving plenary conference talk

Claim

A user's situationally determined cognitive load can affect interaction more strongly than her knowledge, preferences, etc.

So it's one more thing that a system $\mathcal{S}$ may need to adapt to

5 Situation Considered Here

Duration of interactions

The system ($\mathcal{S}$) in general interacts only once with each user ($\mathcal{U}$)

E.g., $\mathcal{S}$ is a computer hotline

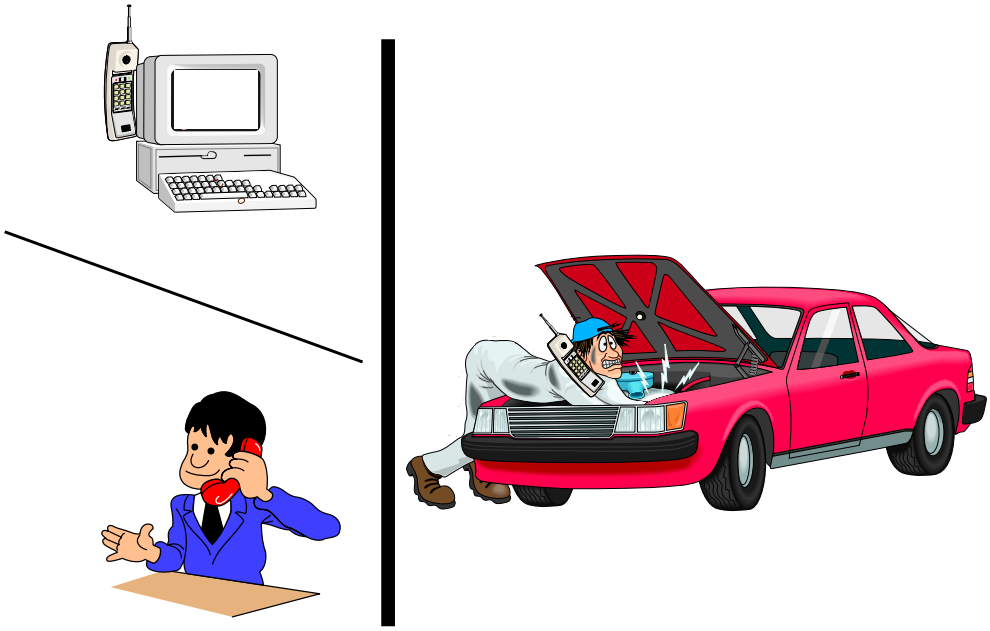
Þ Gradual, long-term learning about $\mathcal{U}$ is not possible

Available evidence

Speech is the primary input medium

Try It Yourself 6

Raise your hand when you recognize high cognitive load
in the taped dialog.



7 Straightforward Machine Learning? (1)

Straightforward approach

1. Create samples of speech with known cognitive load
2. Encode their features
3. Use features as input to supervised, off-line learning algorithm
4. Cross-validate the learned performance component
5. Apply to new users

Example of successful application

System for recognizing emotions on basis of speech

(See Valery Petrushin, UM99, for demo)

Straightforward Machine Learning? (2) 8

Complications

Features

Which features should you use?

How should they be defined?

How can they be extracted automatically and in real time?

Performance component?

How can $\mathcal{S}$'s inferences be made comprehensible?

How can evidence from speech be combined with other evidence available to $\mathcal{S}$?

$\mathcal{U}$'s task

Properties of $\mathcal{U}$

Other behavior of $\mathcal{U}$

Approach taken here

Get features from experimental psycholinguistic literature

Check potential utility with off-line analyses of realistic dialogs

Then do machine learning

Use Bayesian network that explicitly represents causal relationships

Embed this network in a larger one that includes other variables

Possible Symptoms

9 Overview of Psycholinguistic Results

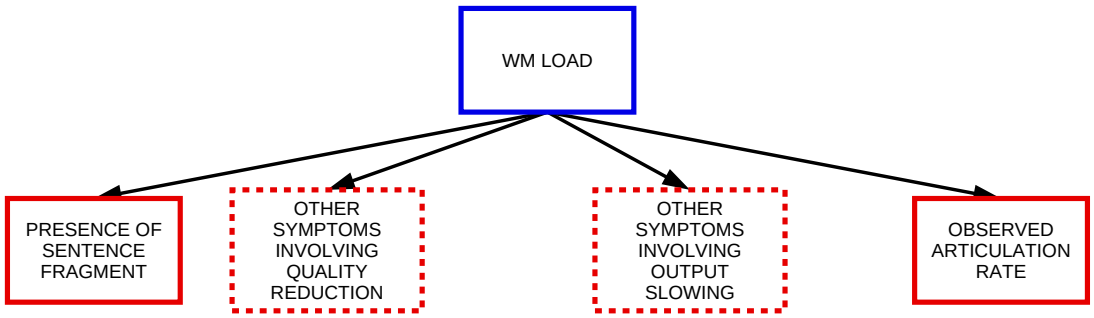
Symptoms involving output quality

Feature	*Trend*	*Tally*
Sentence fragments	+	4/5
False starts	+	2/4
Syntax errors	+	1/1
Self-repairs	+, −, 0	2, 1, 4
Amount of detail	−	4/5
Redundancy	+	2/2

Symptoms involving output rate

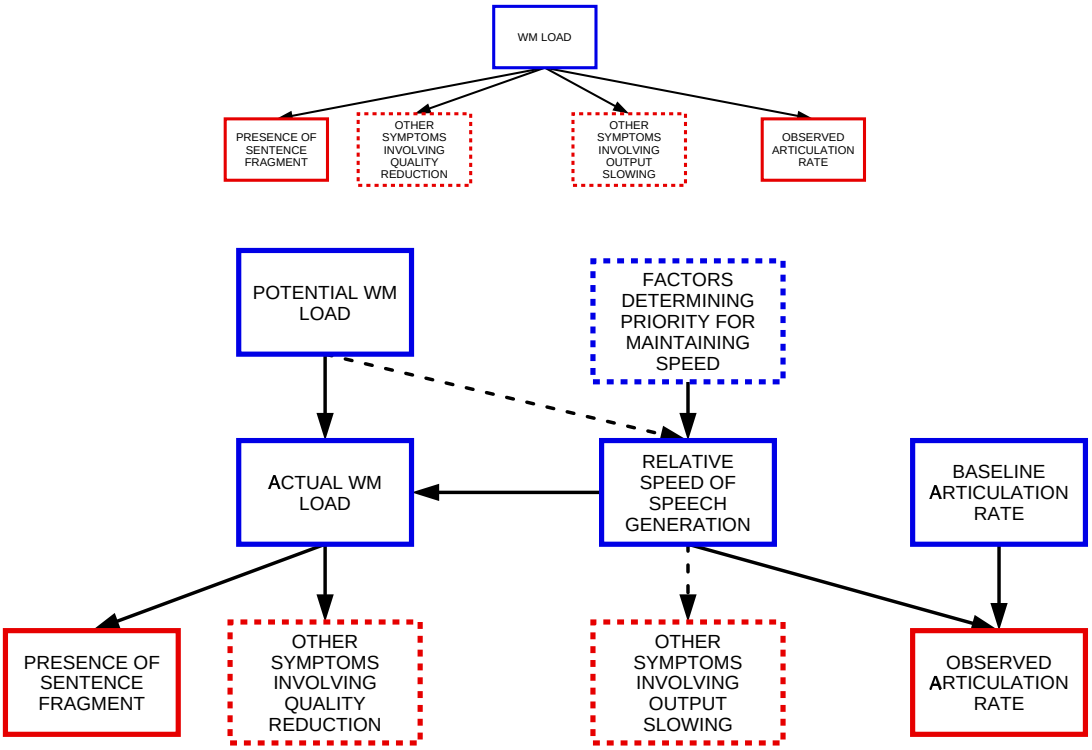
Feature	*Trend*	*Tally*
Articulation rate	−	7/7
Speech rate	−	7/7
Onset latency	+	9/11
Silent pauses (number)	+	4/5
Silent pauses (duration)	+	8/10
Filled pauses (number)	+	4/6
Filled pauses (duration)	+	1/2
Repetitions	+	5/6

Simple Conception of Causal Relationships 10



11

Speed-Quality Tradeoff?



Example Symptom: Sentence Fragments

12

Example

"Yes, that's ... uh, just keep repeating."

General relationship to cognitive load (from experiments)

When the speaker is performing a secondary task,

sentence fragments are about 3 times as frequent, on the average

Role in dialog situations (from our own analyses)

Frequency

7% of dialog turns

Complications

Sometimes due to factors not present in experiments

(e.g., interruptions)

13 Example Symptom: Articulation Rate (1)

Example

<uh> ... *In the ... inside under the steering wheel ... to the left ... there's a fuse box.*

Definition

$$\frac{\text{Number of syllables articulated}}{\text{Total duration of articulated syllables}}$$

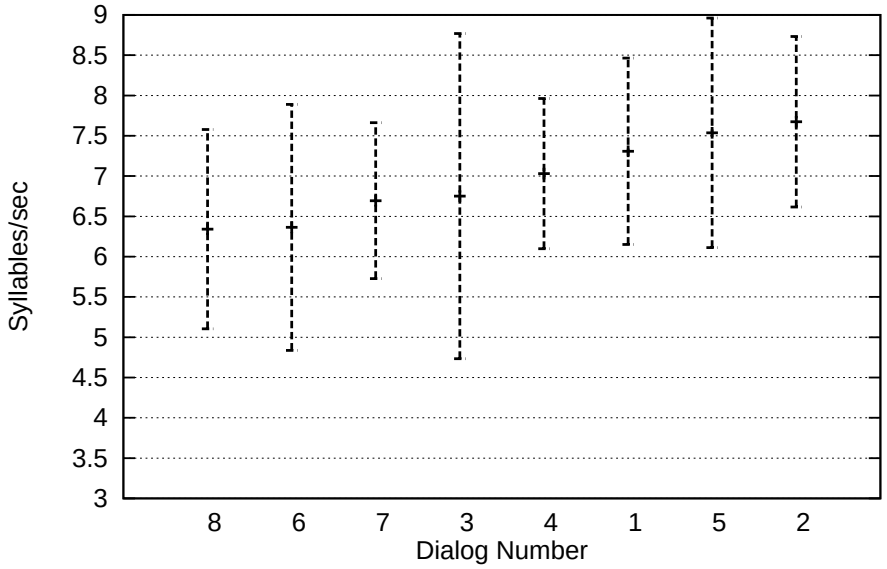
General relationship to cognitive load (from experiments)

- About 14% lower given fairly high cognitive load, on the average
- Considerable individual differences

Example Symptom: Articulation Rate (2) 14

Role in dialog situations (from our own analyses)

- Measurement problematic when number of syllables < 4
- Means and SDs for different callers (number of syllables > 3):



Checking Data-Limitedness

15 Accuracy- and Data-Limited User Models

Ultimate question

"OK, but can you really use these symptoms to recognize cognitive load?"

Potential problems

Accuracy limitations

Network structure and/or probabilities are seriously wrong

Data limitations

Given the limited diagnostic value of the symptoms, there won't be enough data available to permit an accurate assessment

How to check for both types of limitation at once

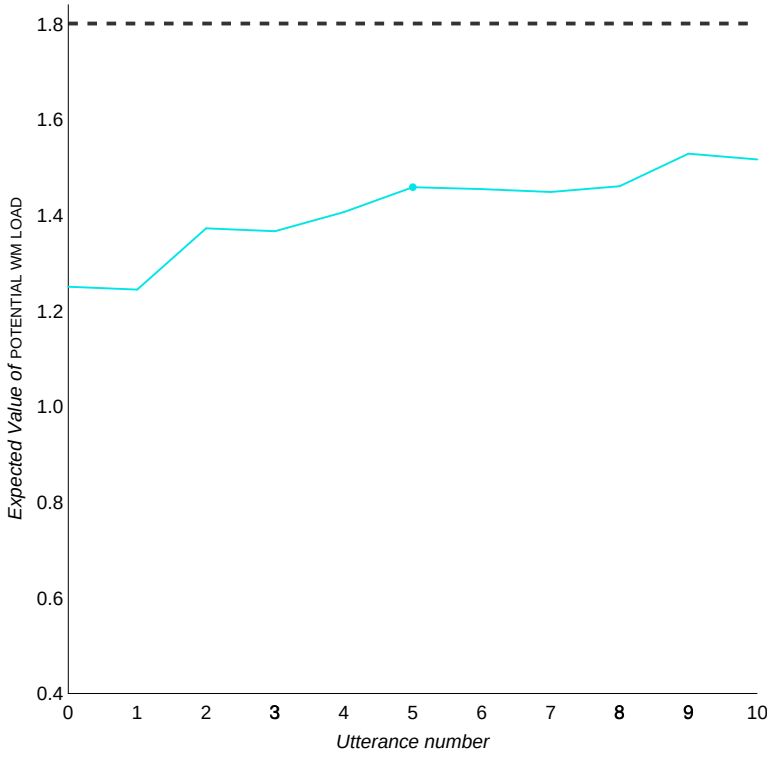
- Collect speech data while manipulating cognitive load
- Learn the Bayesian network
- Try to classify new users

Þ "Current Work"

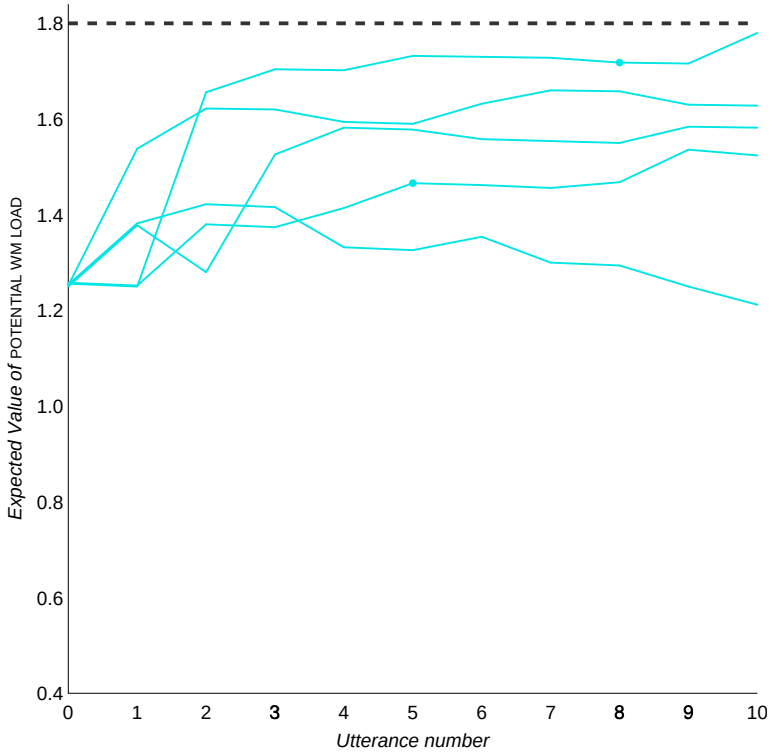
How to check just the data limitations

- Assume there are no accuracy limitations
- Generate input data from hypothetical users accordingly
- See if $\mathcal{S}$ can classify the "users" successfully

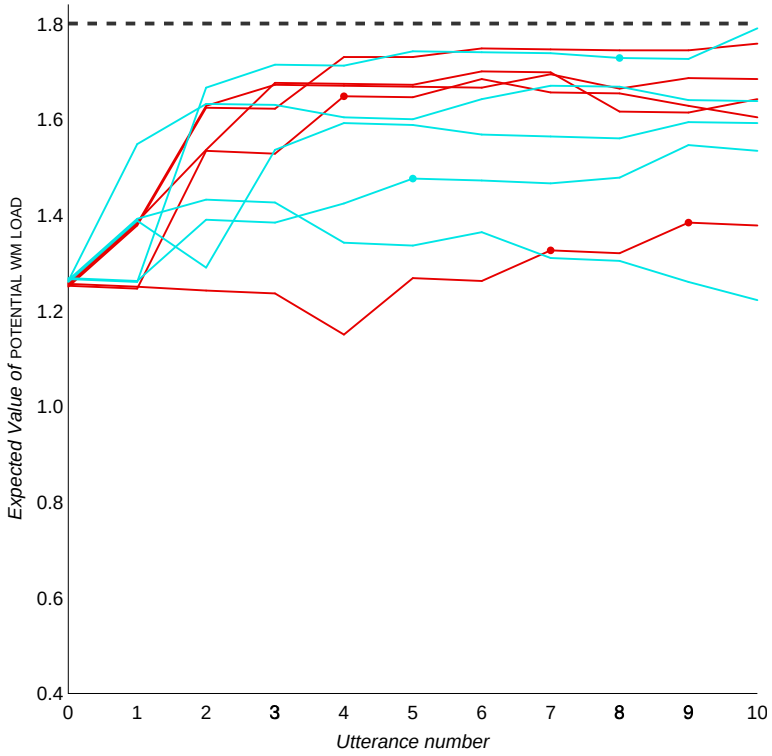
Hypothetical Users With Very High Load (1) 16



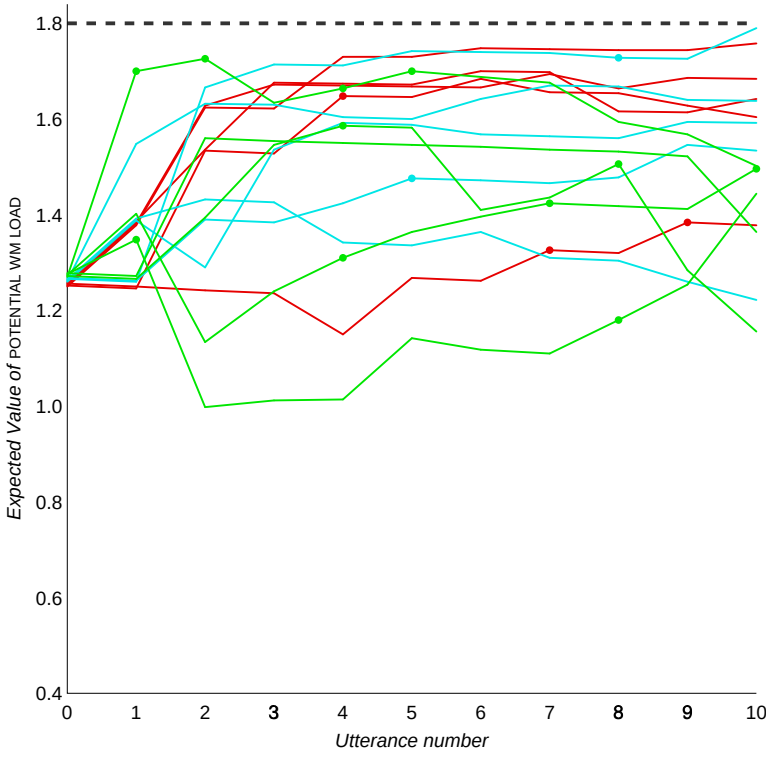
17 Hypothetical Users With Very High Load (2)



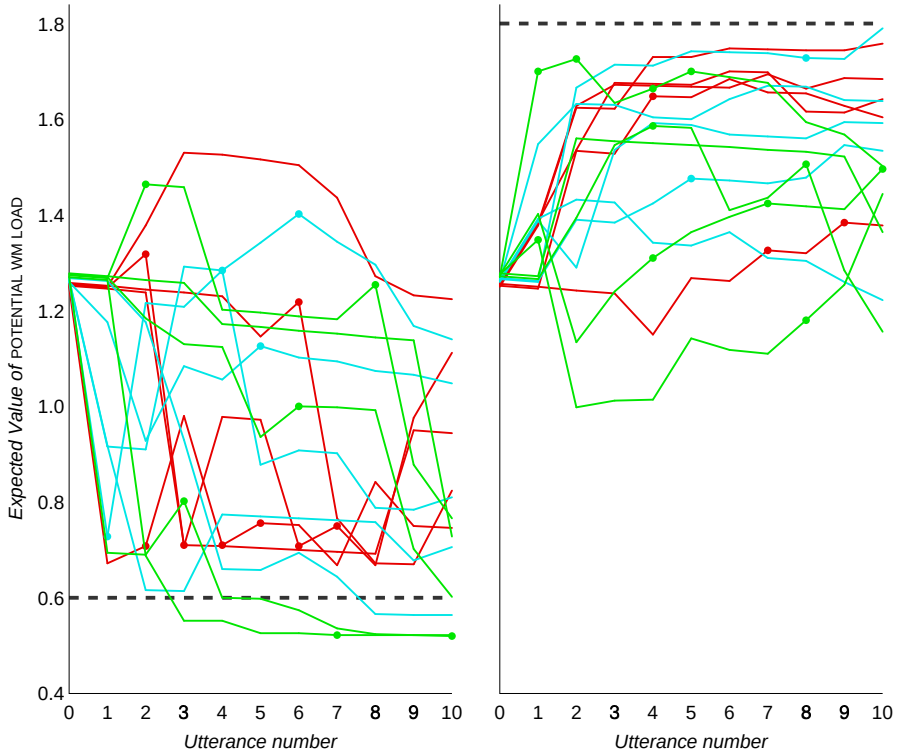
Hypothetical Users With Very High Load (3) 18



19 Hypothetical Users With Very High Load (4)



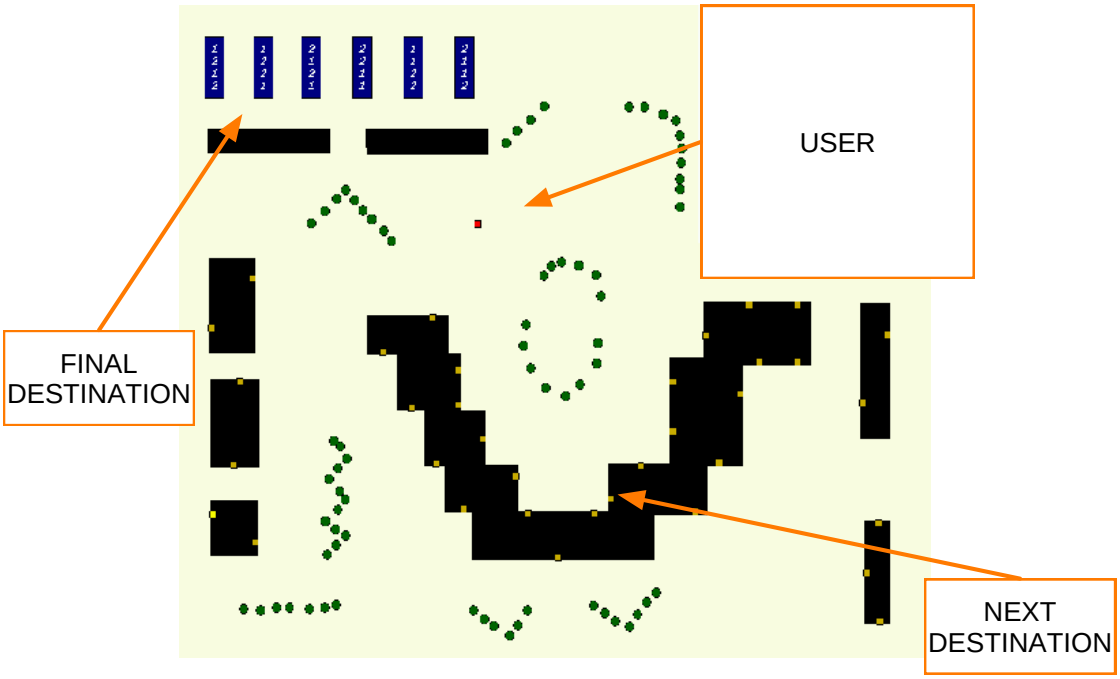
Users With Low and Very High Load 20



Current Work

Experiment (1)

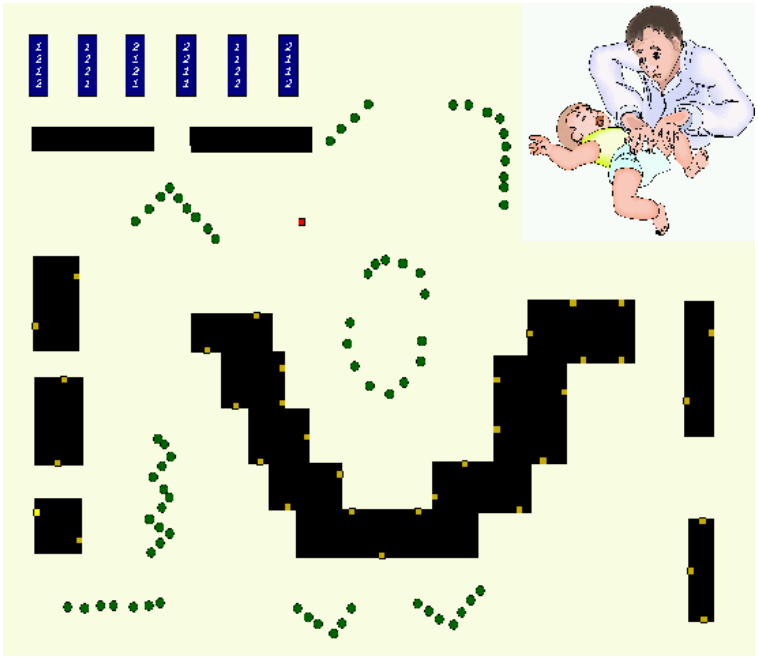
21



$\mathcal{U}$ is navigating through Frankfurt Airport

Experiment (2)

22



"Is there ... uh ... Where can I ... change my baby's diapers?"

23

Experiment (3)

Independent variables

- Cognitive load
  Does $\mathcal{U}$ have to navigate?
- Time pressure
  Reward for speed?

Dependent variables

Various symptoms of cognitive load

Use of data

Basis for learning of Bayesian network with specified structure and hidden variables

Þ Frank Wittig, Doctoral Consortium, today, 6:15 pm

Conclusions

Summary of Contributions

24

Content

- Overview of known symptoms of cognitive load
- Hypothesis about relationships between symptoms
- Discussion of diagnostic value and interpretation problems for two example symptoms

Methodology

- Way of synthesizing previously published experimental data and more naturalistic studies
- Method for analyzing data-limitedness of a user modeling component