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

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>[Title Page]</td>
<td>1</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>2</td>
</tr>
<tr>
<td>Overview</td>
<td>3</td>
</tr>
<tr>
<td>Problem and Approach Taken</td>
<td>4</td>
</tr>
<tr>
<td>Why Recognize Cognitive Load?</td>
<td>4</td>
</tr>
<tr>
<td>Situation Considered Here</td>
<td>5</td>
</tr>
<tr>
<td>Try It Yourself</td>
<td>6</td>
</tr>
<tr>
<td>Straightforward Machine Learning?</td>
<td>7</td>
</tr>
<tr>
<td>Possible Symptoms</td>
<td>9</td>
</tr>
<tr>
<td>Overview of Psycholinguistic Results</td>
<td>9</td>
</tr>
<tr>
<td>Simple Conception of Causal Relationships</td>
<td>10</td>
</tr>
<tr>
<td>Speed-Quality Tradeoff?</td>
<td>11</td>
</tr>
<tr>
<td>Example Symptom: Sentence Fragments</td>
<td>12</td>
</tr>
<tr>
<td>Example Symptom: Articulation Rate</td>
<td>13</td>
</tr>
<tr>
<td>Checking Data-Limitedness</td>
<td>15</td>
</tr>
<tr>
<td>Accuracy- and Data-Limited User Models</td>
<td>15</td>
</tr>
<tr>
<td>Hypothetical Users With Very High Load</td>
<td>16</td>
</tr>
<tr>
<td>Users With Low and Very High Load</td>
<td>20</td>
</tr>
<tr>
<td>Current Work</td>
<td>21</td>
</tr>
<tr>
<td>Experiment</td>
<td>21</td>
</tr>
<tr>
<td>Conclusions</td>
<td>24</td>
</tr>
<tr>
<td>Summary of Contributions</td>
<td>24</td>
</tr>
</tbody>
</table>
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?

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 $S$ may need to adapt to
Situation Considered Here

Duration of interactions
The system ($S$) in general interacts only once with each user ($U$)
E.g., $S$ is a computer hotline
Gradual, long-term learning about $U$ is not possible

Available evidence
Speech is the primary input medium

Try It Yourself
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)

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 S’s inferences be made comprehensible?
How can evidence from speech be combined with other evidence available to S?
U’s task
Properties of U
Other behavior of 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

#### Overview of Psycholinguistic Results

<table>
<thead>
<tr>
<th>Symptoms involving output quality</th>
<th>Feature</th>
<th>Trend</th>
<th>Tally</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentence fragments</td>
<td>+</td>
<td>4/5</td>
<td></td>
</tr>
<tr>
<td>False starts</td>
<td>+</td>
<td>2/4</td>
<td></td>
</tr>
<tr>
<td>Syntax errors</td>
<td>+</td>
<td>1/1</td>
<td></td>
</tr>
<tr>
<td>Self-repairs</td>
<td>+, −, 0</td>
<td>2, 1, 4</td>
<td></td>
</tr>
<tr>
<td>Amount of detail</td>
<td>−</td>
<td>4/5</td>
<td></td>
</tr>
<tr>
<td>Redundancy</td>
<td>+</td>
<td>2/2</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Symptoms involving output rate</th>
<th>Feature</th>
<th>Trend</th>
<th>Tally</th>
</tr>
</thead>
<tbody>
<tr>
<td>Articulation rate</td>
<td>−</td>
<td>7/7</td>
<td></td>
</tr>
<tr>
<td>Speech rate</td>
<td>−</td>
<td>7/7</td>
<td></td>
</tr>
<tr>
<td>Onset latency</td>
<td>+</td>
<td>9/11</td>
<td></td>
</tr>
<tr>
<td>Silent pauses (number)</td>
<td>+</td>
<td>4/5</td>
<td></td>
</tr>
<tr>
<td>Silent pauses (duration)</td>
<td>+</td>
<td>8/10</td>
<td></td>
</tr>
<tr>
<td>Filled pauses (number)</td>
<td>+</td>
<td>4/6</td>
<td></td>
</tr>
<tr>
<td>Filled pauses (duration)</td>
<td>+</td>
<td>1/2</td>
<td></td>
</tr>
<tr>
<td>Repetitions</td>
<td>+</td>
<td>5/6</td>
<td></td>
</tr>
</tbody>
</table>

### Simple Conception of Causal Relationships

```
WM LOAD

- PRESENCE OF SENTENCE FRAGMENT
- OTHER SYMPTOMS INVOLVING QUALITY REDUCTION
- OTHER SYMPTOMS INVOLVING OUTPUT SLOWING
- OBSERVED ARTICULATION RATE
```
**Possible Symptoms**

**Speed-Quality Tradeoff?**

---

**Example Symptom: Sentence Fragments**

**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)
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)

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):

![Graph showing articulation rate by dialog number](image-url)
Checking Data-Limitedness

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
"Current Work"

How to check just the data limitations
• Assume there are no accuracy limitations
• Generate input data from hypothetical users accordingly
• See if \( S \) can classify the "users" successfully

Hypothetical Users With Very High Load (1)

![Graph showing Expected Value of Potential WM Load over Utterance number]
17 Hypothetical Users With Very High Load (2)

18 Hypothetical Users With Very High Load (3)
Hypothetical Users With Very High Load

Users With Low and Very High Load
Current Work

Experiment (1)

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

Experiment (2)

"Is there ... uh ... Where can I ... change my baby's diapers?"
Experiment (3)

Independent variables
- Cognitive load
  Does the user 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
Frank Wittig, Doctoral Consortium, today, 6:15 pm

Conclusions
Summary of Contributions

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