INCREMENTAL DIALOGUE SYSTEMS
The work of David Schlangen, Gabriel Skantze and Anna Hjalmarsson presented by Fai Greeve
**Today:**

- Dialogue systems: Why? Which are out there? What difference does incrementality make?

- Incremental dialogue systems: Which topology, how does the information flow and how do the modules process information?

- Incremental dialogue systems: Implementation examples. How do they perform?
DIALOGUE SYSTEMS

Watch it at: http://www.youtube.com/watch?v=YBuaZH29Ko
AN IN-CLASS EXPERIMENT

System 1:
- 2 7 9 45 89 67 0 9

System 2:
- 64 3 6 9 22 19 9 2

- Did you succeed in the task?
- How well were you understood?
- How responsive was the system?
- Did the system behave as expected?
- How humanlike was the system?
- Which system did you prefer?
AN EXAMPLE DIALOGUE BETWEEN A USER AND THE NUMBERS SYSTEM

U.01 two five three -
S.02 okay
U.03 one forty, five -
S.04 sorry, do you mean one and then forty and then five?
U.05 no, one forty five
S.06 okay, one forty five
U.07 yeah, and six eight four -
S.08 okay
U.09 seven twelve nine.
S.10 so, from the start, two five three -
U.11 yeah
S.12 one forty five six six four -
U.13 no, six eight four
S.14 okay, six eight four and then seven twelve nine.
U.15 that’s right
INCREMENTALITY

- Proceed in steps

- Processing starts before the input is complete, and the first output is produced as soon as possible

- Trade-off: responsiveness vs output quality
NUMBERS

Speech Synthesis

Action Management

Contextualiser

Interpreter

Speech Recognition & Prosodic Analysis
Topology

- Pipeline Topology
- Pipeline Topology + Alternative Paths
- Star Topology
PIPELINE TOPOLOGY
PIPELINE TOPOLOGY + ALTERNATIVE PATH
STAR TOPOLOGY
PARALLEL INFORMATION
Incremental Unit

Identifier: ID567
Same Level Link: ID566s (nine)
Grounded in: ID476
Confidence: 0.8
Commited: T
Seen: SR
Payload: two
LEFT BUFFER – PROCESSOR – RIGHT BUFFER
CONNECTED AXIOMS
**Module Behaviour**

**Update**
- Frequency
  - \( f: \text{in} = \text{out} \)
  - \( f: \text{in} \geq \text{out} \)
  - \( f: \text{in} \leq \text{out} \)

**Connectedness**

**Completeness**
- \( c: \text{in} = \text{out} \)
- \( c: \text{in} \geq \text{out} \)
- \( c: \text{in} \leq \text{out} \)
Module Behaviour

**Update**
- Frequency
- \( f: \text{in} = \text{out} \)
- \( f: \text{in} \geq \text{out} \)
- \( f: \text{in} \leq \text{out} \)

**Connectedness**

**Completeness**
- \( c: \text{in} = \text{out} \)
- \( c: \text{in} \geq \text{out} \)
- \( c: \text{in} \leq \text{out} \)
**Module Behaviour**

**Update**
- Frequency
  - $f: \text{in} = \text{out}$
  - $f: \text{in} \geq \text{out}$
  - $f: \text{in} \leq \text{out}$

**Connectedness**
- $c: \text{in} = \text{out}$
  - $c: \text{in} \geq \text{out}$
  - $c: \text{in} \leq \text{out}$

**Completeness**
- $c: \text{in} = \text{out}$
  - $c: \text{in} \geq \text{out}$
  - $c: \text{in} \leq \text{out}$
Module Operations: Update Purge Commit
NUMBERS

Speech Synthesis

Action Management

Contextualiser

Interpreter

Speech Recognition & Prosodic Analysis
NUMBERS
SPEECH RECOGNITION & PROSODIC ANALYSIS

Speech Synthesis

Action Management

Contextualiser

Interpreter

Speech Recognition & Prosodic Analysis

28-04-2011
Fai Greeve
NUMBERS
ACTION MANAGEMENT

Speech Synthesis

Action Management

Contextualiser

Interpreter

Speech Recognition & Prosodic Analysis
NUMBERS
SPEECH SYNTHESIS

Speech Synthesis

Action Management

Contextualiser

Interpreter

Speech Recognition & Prosodic Analysis

28-04-2011
Fai Greeve
Visual Demonstration NUMBERS

Watch it at: http://www.youtube.com/user/gskantze?blend=4&ob=5#p/u/3/_rDkb1K1si8
INCREMENTAL SPEECHPLAN

28-04-2011
Fai Greeve
# Self Repairs

<table>
<thead>
<tr>
<th>Covert Segment Repair</th>
<th>Overt Segment Repair</th>
</tr>
</thead>
<tbody>
<tr>
<td>you are right</td>
<td>it is blue</td>
</tr>
<tr>
<td>you are right</td>
<td>they are blue</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Covert Unit Repair</th>
<th>Overt Unit Repair</th>
</tr>
</thead>
<tbody>
<tr>
<td>you are right</td>
<td>it is blue</td>
</tr>
<tr>
<td>you are wrong</td>
<td>it is red</td>
</tr>
<tr>
<td>sorry</td>
<td>you are wrong</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Covert Unit Repair</th>
<th>Overt Unit Repair</th>
</tr>
</thead>
<tbody>
<tr>
<td>you are right</td>
<td>it is blue</td>
</tr>
<tr>
<td>you are wrong</td>
<td>it is red</td>
</tr>
<tr>
<td>sorry</td>
<td>you are wrong</td>
</tr>
</tbody>
</table>
AN EXAMPLE DIALOGUE BETWEEN A USER AND THE JINDIGO SYSTEM

S.1  [welcome] [how may I help you]
U.2  I want to buy a doll
S.3  [eh] [here is] [a doll]
U.4  how much is it?
S.5  [eh] [it costs] [120 crowns]
U.6  that is too expensive how much is the teddy bear?
S.7  [well] [you can have it for] [let’s see] [40 crowns]
U.8  I can give you 30 crowns
S.9  [you could have it for] [37 crowns]
U.10 I can give you 10 crowns
S.11 [let’s say] [or, I mean] [that is way too Little]

Watch it at: hhtp://www.youtube.com/watch?v=cQQmgItIMvs
A wizard-of-Oz Experiment

Speech Synthesis

Action Management

Contextualiser

Interpreter
RESPONSE TIME AND LENGTH

<table>
<thead>
<tr>
<th></th>
<th>inc</th>
<th>non</th>
</tr>
</thead>
<tbody>
<tr>
<td>start</td>
<td>0.5</td>
<td>3.0</td>
</tr>
<tr>
<td>end</td>
<td>5.5</td>
<td>5.0</td>
</tr>
<tr>
<td>length</td>
<td>4.5</td>
<td>3.5</td>
</tr>
</tbody>
</table>
## User Experience

<table>
<thead>
<tr>
<th></th>
<th>diff</th>
<th>z-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>preferred</td>
<td>0.23</td>
<td>-1.24</td>
<td>0.214</td>
</tr>
<tr>
<td>human-like</td>
<td>0.15</td>
<td>-0.76</td>
<td>0.445</td>
</tr>
<tr>
<td>polite</td>
<td>0.40</td>
<td>-2.19</td>
<td>0.028*</td>
</tr>
<tr>
<td>efficient</td>
<td>0.29</td>
<td>-2.08</td>
<td>0.038*</td>
</tr>
<tr>
<td>intelligent</td>
<td>0.11</td>
<td>-0.70</td>
<td>0.484</td>
</tr>
<tr>
<td>faster response</td>
<td>0.26</td>
<td>-1.66</td>
<td>0.097</td>
</tr>
<tr>
<td>feedback</td>
<td>0.08</td>
<td>-0.84</td>
<td>0.400</td>
</tr>
<tr>
<td>when to speak</td>
<td>0.35</td>
<td>-2.38</td>
<td>0.017*</td>
</tr>
</tbody>
</table>
WRAP UP

- Dialogue systems: Why? Which are out there? What difference does incrementality make?

- Dialogue systems: Which topology, how does the information flow and how do the modules process information?

- Dialogue systems: Implementation examples. How do they perform?


http://www.speech.kth.se/~gabriel/