More than just words
Building a Chatty Robot

Emer Gilmartin
Nick Campbell
Speech Communications Lab
Trinity College Dublin
• SFI FASTNET Project @ TCD
• Focus on Actions in Social Talk
• Investigating acoustic, visual, and content aspects of natural conversation
  – Speech Communications Lab
  – Phonetics and Speech Lab
  • Trinity College Dublin
Outline

• What is chat?
• Building a spoken chat dialogue
• Building the robot platform
• Data collection and Wizard of Oz trials
• Future work
Introduction

• Conversation meets ecological needs.
  • communicate propositional content / perform a task.
  • build and maintaining social bonds: platform for co-presence.

• Task-based dialogue has been widely studied but chat has not been as extensively analysed or synthesised.

• Dialogue systems use information transfer to perform tasks

• Systems using chat and task-based dialogue more closely model human-human communication and could improve HMI.
Language ≠ Speech ≠ Conversation
Conversation is multimodal
Spoken Language

Core message transmitted using syntax/lexicon/phonetic/prosodic layers
Conversational speech

- Gesture calls
- Gesticulation
- Prosody / voice quality
- Quotable gestures
- Quotable vocalisations
- Spoken language
# Chat vs Task

<table>
<thead>
<tr>
<th></th>
<th>Task-based</th>
<th>Chat</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Goal</strong></td>
<td>Accomplish task discrete short term, clearly defined</td>
<td>Establish co-presence, service interpersonal bonds, maintain channels of communication may not be explicit</td>
</tr>
<tr>
<td></td>
<td>Booking a flight</td>
<td>Daily chat with hall porter, corner shop owner</td>
</tr>
</tbody>
</table>

30/11/2012
# Chat vs Task

<table>
<thead>
<tr>
<th>Task-based</th>
<th>Chat</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Content</strong></td>
<td></td>
</tr>
<tr>
<td>information transfer vital may be positive, negative or controversial</td>
<td>light</td>
</tr>
<tr>
<td>accomplishment of task may override social considerations</td>
<td>non-controversial</td>
</tr>
<tr>
<td></td>
<td>often phatic</td>
</tr>
</tbody>
</table>
# Chat vs Task

<table>
<thead>
<tr>
<th></th>
<th>Task-based</th>
<th>Chat</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Goal</strong></td>
<td>Accomplish task discrete short term, clearly defined</td>
<td>Establish co-presence, service interpersonal bonds, maintain channels of communication may not be explicit</td>
</tr>
<tr>
<td><strong>Content</strong></td>
<td>information transfer vital may be positive, negative or controversial accomplishment of task overrides social considerations</td>
<td>light non-controversial often phatic</td>
</tr>
<tr>
<td><strong>Structure</strong></td>
<td>adjacency pairs - often nested to several layers</td>
<td>same building blocks - shallower structure</td>
</tr>
<tr>
<td><strong>Cognitive Processing</strong></td>
<td>feedback control important Deeper, slower processing</td>
<td>feed forward or predictive control shallow, fast processing</td>
</tr>
</tbody>
</table>
Chat vs Task

- Contrast
- Dual process theory
- “a continuing state of incipient talk”
Research Questions

1. What features distinguish *spoken* chat and task-based elements of dialogue?
2. How to implement *spoken* chat in HMI?
3. Will *spoken* chat enhance HMI?
Building the dialogue

• Chat
  – Lessons from Chatbots
  – stateless
  – system-controlled
  – ‘stanzas’
  – light, friendly content

• Voice
  – high-pitched, lively
  – ‘infectious’ laugh

• Timing

my name is hur-mi
H E R M E
what's your name?
how old are you?
really
I'm nearly seven weeks old

are you from Dublin?
really
I'm from the Speech Communication Lab here in Tee See Dee

I like your hair

do you like the exhibition?
really
owe
why?
Building the Robot Platform

- Lego NXT with Python
- The LEGO robot, HERME
  1. HD web-cam / noise-cancelling microphones.
  2. light sensor
  3. sound sensor
  4. distance sensor.
Building the Robot Platform

• Cameras
• Microphones
• Presence detector
Running the dialogue

• Science Gallery TCD Human+
• Wizard of Oz scenario
  • Wizard-controlled from another building via Skype
• Only dof – timing
• 3 months in Science Gallery, TCD
• Participants: members of public (casual museum visitors)
• 483 participants signed consent form as part of their chat with robot, agreeing to research use of the recordings - 1.5 Tb audio and video
• **Scripted robot-led dialogue – several ‘volleys’**
• Statement / question – wait for response – feedback – wait – feedback
• Robot controlled conversation – give impression of polite listening
• Light, friendly, jokey content
• **Physical Presence**
• Always turn to ‘look’ at the closest person
• Recoil automatically if someone tries to touch or gets too close.
• Display face circles on the screen to indicate focus of attention.
• **Automatic and wizard mediated**
• Wizard-controlled from another building via Skype
• Used insights from wizard sessions to improve timing on automatic system
Current work

• Herme 2
• Corpora
• Chat vs task experiments