Why and how to model multi-modal interaction for a mobile robot companion

Shuyin Li, Julia Peltason and Britta Wrede

Bielefeld University
Germany
Outline

- Introduction to Human-Robot Interaction (HRI)
- Observations in a user study
- A multi-modal interaction framework
- Summary
Introduction: HRI with a personal robot

Robot characteristics

- situated

Requirements for the system

- situation-awareness

Robots should be aware of environmental changes
Introduction: HRI with a personal robot

Robot characteristics

- situated

Robots should be aware of environmental changes

Requirements for the system

- situation-awareness

- social behaviors

Users expect human-like behaviors
Introduction: HRI with a personal robot

Robot characteristics

- situated
  - Robots should be aware of environmental changes
- anthropomorphic
  - Users expect human-like behaviors
- embodied
  - Both users and robots have visual access to their interaction partner's body

Requirements for the system

- situation-awareness
- social behaviors
- multi-modal interaction
Introduction: HRI with a personal robot

**Robot characteristics**
- situated
- anthropomorphic
- embodied

**Requirements for the system**
- Robots should be aware of environmental changes
- Users expect human-like behaviors
- Both users and robots have visual access to their interaction partner’s body
- situation-awareness
- social behaviors
- multi-modal interaction
Introduction: HRI with a personal robot

Robot characteristics

- situated: Robots should be aware of environmental changes
- anthropomorphic: Users expect human-like behaviors
- embodied: Both users and robots have visual access to their interaction partner's body

Requirements for the system

- situation-awareness
- social behaviors
- multi-modal interaction

Li, Peltason and Wrede, March 2007
Outline

- Introduction to Human-Robot Interaction (HRI)
- **Observations in a user study**
  - BIRON@Home
  - Quiet speakers
  - Meta-commentators
- A multi-modal interaction framework
- Summary
The user study: BIRON@Home
The user study: BIRON@Home

Experimental setup with BIRON
14 subjects, each interaction about 7 min.

Non-task behaviors of BIRON:
1. situation awareness
2. social behavior:

Only output-modality of BIRON: speech
The user study: situation awareness

- Face recognition
- Sound source detection
- Human leg detection
Observation I: quiet speakers
Observation I: quiet speakers

- **The problem**
  
  No means to communicate pre-interaction attention

- **Possible reason**
  
  inappropriateness of speech modality
  
  (“legs detected, face detected, face lost again, face detected, ...”)

- **Solution**
  
  use non-verbal modalities
  
  (because they are suitable to represent static information which is only occasionally updated)
The user study: social behavior

... 
User: Follow me.
BIRON: OK, I'm following you.
User: This is a cup.
BIRON: It's nice.
You are really doing very well!
User: (laugh)
BIRON: Come here.
Observation II: meta-commentators
Observation II: meta-commentators

- Problem
  users reply to social comments using out-of-vocabulary words
- Possible reason
  reciprocity and obtrusiveness of the speech modality
- Solution
  making remarks using non-verbal modalities
  (because they are unobtrusive and do not impose strong obligation to reply)
Outline

- Introduction to Human-Robot Interaction (HRI)
- Observations in a user study
- A multi-modal interaction framework
  - Currently popular approaches
  - Our approach
- Summary
Currently popular approaches

address differences between multi-modal information by grouping it into categories and handle different categories separately.
Interaction framework: existing works

Cassell: generic architecture for embodied conversational agents

Li, Peltason and Wrede, March 2007
Interaction framework: existing works

Traum: dialog model for multi-modal, multi-party dialog in virtual world
(based on information-state theory)

- Contact
- Attention
- Conversation
  - Participant
  - Turn
  - Initiative
  - Grounding
  - Topic
  - Rhetorical
- Social commitments (obligations)
- negotiation
Interaction framework: multi-modal grounding

Our approach

address a common feature of multi-modal information: evocative functions
Evocative functions of conversational behaviors (CBs)

- **Definition**: CBs evoke a reaction from the interaction partner
- **Validity**: for both propositional and interactional info
Interaction framework: multi-modal grounding

Grounding

- **Definition:** the process of establishing shared understanding during a conversation.

- **Basic idea:** for each account (Presentation) issued in a conversation, there needs to be a feedback (Acceptance) from the interaction partner

- **Application area:** traditionally adopted to model evocative functions of propositional information → to be extended!
Interaction framework: multi-modal grounding

Our approach: extending grounding

- Modeling both propositional and interactional contribution with Interaction Unit
- Organizing IUs based on the principle of grounding
Interaction framework: multi-modal grounding

Interaction Unit (IU)

- Behavior Layer
- Motivation Layer
Interaction framework: multi-modal grounding

Interaction Unit (IU)

- verbal generator
- non-verbal generator

(motivation conception)
Interaction framework: multi-modal grounding

Grounding

<table>
<thead>
<tr>
<th>speech</th>
<th>gesture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motivation</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>speech</th>
<th>gaze</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motivation</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>speech</th>
<th>facial expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motivation</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>...</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motivation</td>
<td></td>
</tr>
</tbody>
</table>
Interaction framework: multi-modal grounding

- **Grounding models:**
  - [Clark92]
  - [Traum94]
  - [Cahn&Brennan99]
- **Our approach:** [Li2006]

Pre-interaction attention: solving the quiet-speaker-problem

User: (shows legs)

U1

uninstantiated  shows legs

unintentional motivation
Interaction framework: multi-modal grounding

Pre-interaction attention:
solving the quiet-speaker-problem

User: (shows legs)

BIRON: (opens eyes)

Uninstantiated shows legs

Uninstantiated opens eyes

Unintentional motivation

Provides acceptance to user IU

Li, Peltason and Wrede, March 2007
Interaction framework: multi-modal grounding

Pre-interaction attention: solving the quiet-speaker-problem

User: (shows face)

BIRON: (raises head)

<table>
<thead>
<tr>
<th>U3</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>uninstantiated</td>
<td>shows face and legs</td>
</tr>
</tbody>
</table>

unintentional motivation

<table>
<thead>
<tr>
<th>U4</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>uninstantiated</td>
<td>raises head</td>
</tr>
</tbody>
</table>

provides acceptance to user IU
Interaction framework: multi-modal grounding

Making social comments: solving the meta-commentator-problem

User: This is a cup.

U5

“This is a cup”   deictic gesture

shows BIRON a cup
Interaction framework: multi-modal grounding

Making social comments: solving the meta-commentator-problem

User: This is a cup.

BIRON: I beg your pardon?

**U5**

| "This is a cup" | deictic gesture |

shows BIRON a cup

**U6**

| "I beg your pardon" | uninstantiated |

initiates conversational repair
Making social comments: solving the meta-commentator-problem

User: This is a cup.

BIRON: I beg your pardon?

BIRON: (looking embarrassed)

"This is a cup" deictic gesture
U5
shows BIRON a cup

"I beg your pardon" uninstantiated
U6
initiates conversational repair

uninstantiated looking embarrassed
U7
shows social awareness

Li, Peltason and Wrede, March 2007
Interaction framework: multi-modal grounding

Implemented on systems BIRON and BARTHOC
Summary

- Two cases studies revealing the importance of multi-modality in situatedness and social behaviors of a robot
- A multi-modal interaction framework that addresses the evocative function of conversational behaviors