Situated Interaction with a Virtual Human

Perception, Action, and Cognition

Situierte Generierung

Universität Potsdam, WiSe 2011/12
Anne Schumacher

Outline

1. Introduction
2. Modeling Dialogue Behaviour
3. Architectural Approach
4. Situated Interaction Management

Previous work – SFB 360

- 1996 -2005 in Bielefeld „Situated Artificial Communicators“
- Goal: to model what a person performs when, with a partner, he cooperatively solves a simple assembly task in a certain situation.
- Which intelligent abilities are necessary for the accomplishment of an assembly task?
  (with restriction to situated tasks)

Previous work – SFB 360

- Virtual human must be able to process…
  - acoustic (spoken) input
  - visual input of the partner
  - visual input of the objects involved
  - logics of what is going on in the situation
- Virtual human must be able to perform…
  - understanding of what he perceived
  - formulation of own utterances
  - planning of own actions
- Situatedness is a prerequisite for a more exact examination of the intelligence abilities!

Previous work – SFB 360

- setting in SFB 360:
  - human = instructor
  - communicator = executive constructor

- new project:
  - interaction is guided by the user‘s wish of building a certain assembly
  - <u>But:</u> roles (instructor/ constructor) can switch any time

  → more interaction and higher flexibility

Motivation

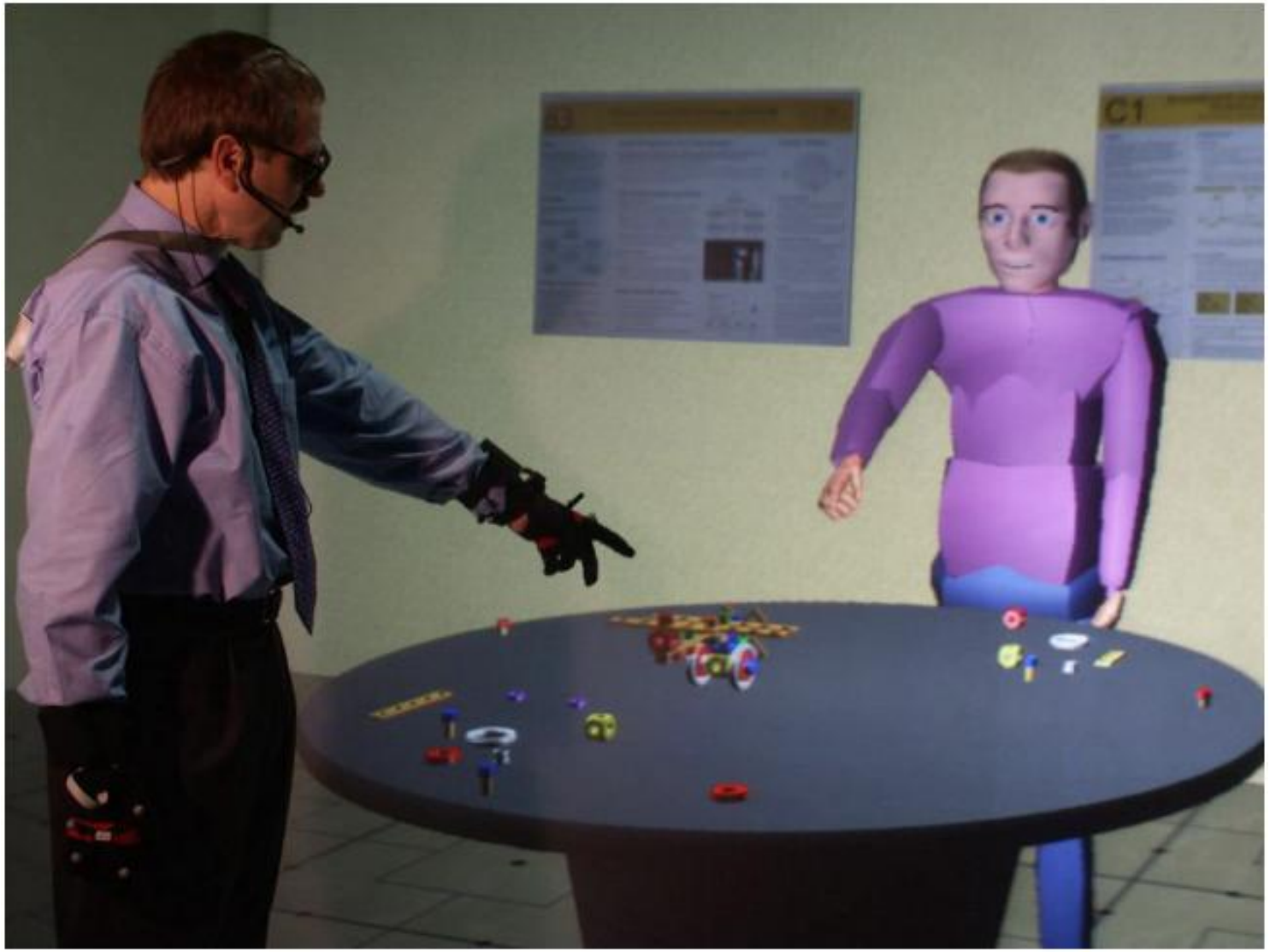
- Modelling a virtual communicator can lead to a better understanding of what is needed in natural communication.
- Virtual humans can be used in collaboration tasks.
- Development of intelligent robots which can be used in various fields.

Setting

- Virtual reality: Max (Multimodal Assembly eXpert) and Baufix pieces are projected

- human is equipped with:
  - microphone
  - stereo glasses
  - optical position trackers
  - data gloves

→Everything the human does is being perceived and processed by the system.



Kopp, Jung, Leßmann, Wachsmuth (2003) *Max – A Multimodal Assistant in Virtual Reality Construction*. Gesellschaft für Informatik KI, 4/03, Seite 11-17

Challenges for the Virtual Human

- goal detection & concretisation
- action planning
- communication vs. manipulation
- mixed initiative & turn taking
- failure detection/ correction
- constantly changing environment
- real-time processing and acting

2. Modeling Dialogue Behaviour

Interaction Model

Which information does one need to participate in the interaction?

layers of the Interaction Model:

- initiative
- turn
- goals
- content
- grounding
- discourse structure
- partner model

Interaction Moves (IM)

- interaction move can be ***communicative*** or ***manipulative***
- For processing and planning IMs, the system needs lots of information → ***slots*** of an interaction move:
    1. action
    2. goal
    3. content
    4. surface form
    5. turn-taking
    6. discourse function
    7. agent
    8. adressee

Filling the Interaction Model

What information do we put into those slots??

1. action → performative types:
   - inform
   - query
   - request
   - propose

2. goal: specific goal of the IM, depending on the performative

Filling the Interaction Model

3. content: facts conveyed by the move
4. surface form: either the words spoken or the action performed
5. turn-taking: *take*, *want*, *yield*, *give*, *keep*
6. discourse function: *start-segment*, *contribute*, *close-segment*
7. agent: the one performing the IM
8. adressee: the one receiving the IM (if communicative)

Interaction Move	*"Let us build a propeller."*	*"Ok."*	*"First, insert a bolt in the middle of a bar."*
Action	propose.action	inform.agree	request.order
Goal	(Achieve (Exists prop))	(Perform (Inform.agree))	(Achieve (Connected $s $b $p1 $p2))
Content	(Build prop we)	(Build prop we)	(Connect $s $b $p1 $p2) (Inst $s bolt) (Inst $b bar) (Center_hole $b $p2))
Surface form	<words>$_t$	<words>$_t$	<words>$_t$
Turn-taking	take | give	take | keep	| give
Discourse function	start-segment (DSP=prop)	contribute	start-segment (DSP=prop-s1)
Agent	User	Max	Max
Addressee	Max	User	User

Leßmann, Kopp, Wachsmuth (2006) *„Situated Interaction with a Virtual Human – Perception, Action and Cognition"* in *Situated Communication*, Berlin: Mouton de Gruyter, 2006, 287-323

3. Architectural Approach

Architecture

- Memories and knowledge bases:
  - storage of all initial and gained knowledge about the world and partner‘s beliefes
  - storage of all that had been perceived and when it was added to the memory

- Perception:
  - visual sensors
  - infrared tracking of human‘s position
  - data gloves
  - auditory sensors & speech recognizer

Architecture

- Reasoning and Deliberation:
  - Belief-Desire-Intetion model (BDI) = control architecture for choosing from possible actions
  - beliefs = knowledge about the world
  - desires = actions that want to be performed/ states that want to be achieved
  - intention = current goal with plan to achieve it

Architecture

- Planning:
  - plan library that stores all possible actions
  - plans can be transformed into intentions

- Acting and Reacting:
  - acting area is in charge of triggering the performance of manipulative/ communicative actions

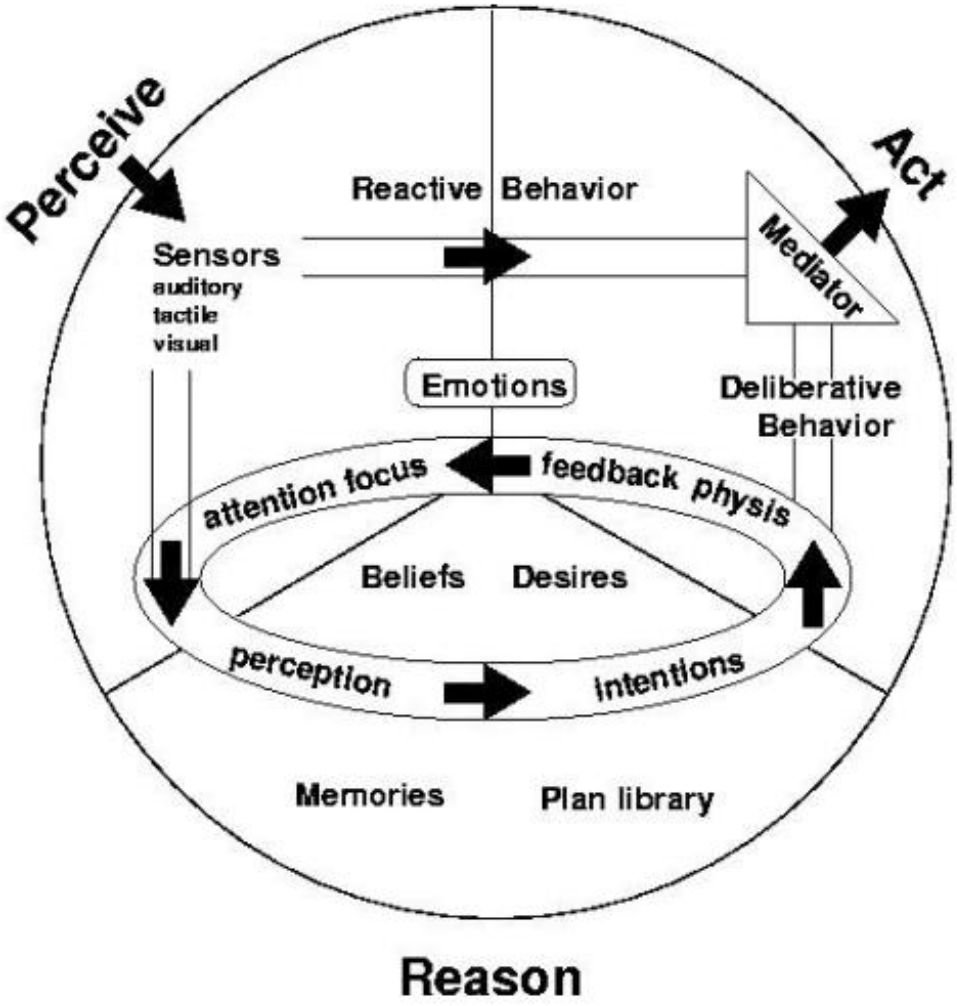


Figure 3: Outline of the architectural framework of Max.

Leßmann, Kopp, Wachsmuth (2006) *„Situated Interaction with a Virtual Human – Perception, Action and Cognition“* in *Situated Communication*, Berlin: Mouton de Gruyter, 2006, 287-323

4. Situated Interaction Management

Dealing with the Input

- physical input:
  - touch detection
  - gesture recognition

- Speech input:
  - parsing
  - keyword-spotting
  - semantic analysis
  - reference resolution

Planning actions

What must be taken into account to plan an action?

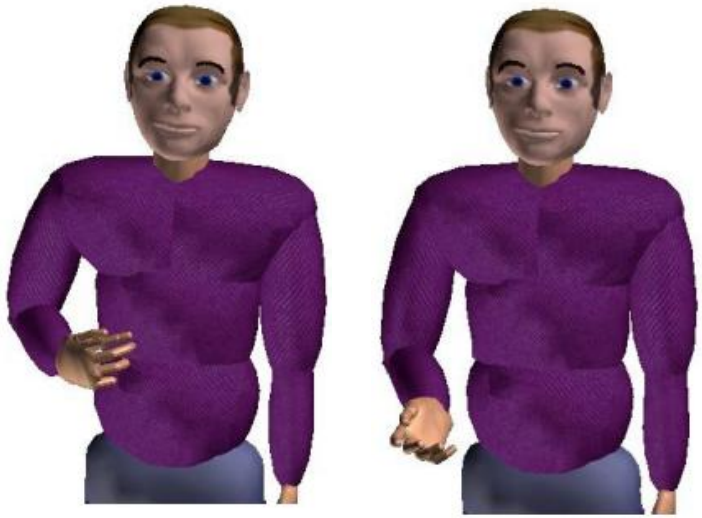
- current goals
- discourse history
- performative of incoming action
- dominance relation between partners
- turn-taking model
- communicative vs. manipulative action
- direct reaction and interruption vs. analysing

Planning communicative moves

- content selection
  - deriving the performative from the intended act
  - determining information needed for the content
- discourse planning
  - determining the discourse function
  - lookup discourse history
- sentence planning & realization
  - generating referring expressions
  - constructing grammatical sentences
  - applying prosody

Physical moves

→Physical behaviour is directly derived from communicative goals.



- turn-taking signals
- pointing gestures

- signals to underscore utterances:
  - eyebrow raise
  - head nod
  - posture shift
  - facial expressions

Summing up…

1. Input is being received by different sensors.
2. All information are being processed and stored in different data bases.
3. Based on the existing (initial and learned) data the agent can plan own actions.
4. Actions are either executed according to the plan or are being derived from the reactive behaviour.

Sources

- Leßmann, Kopp, Wachsmuth (2006) *„Situated Interaction with a Virtual Human – Perception, Action and Cognition“* in *Situated Communication (*287-323), Berlin: Mouton de Gruyter

- Kopp, Jung, Leßmann, Wachsmuth (2003) *Max – A Multimodal Assistant in Virtual Reality Construction*. Gesellschaft für Informatik KI, 4/03, Seite 11-17

- http://www.sfb360.uni-bielefeld.de

- http://www.techfak.uni-bielefeld.de/~skopp/max.html