



Human-Robot Interaction through Natural Language Dialogue

Ozan Özdemir

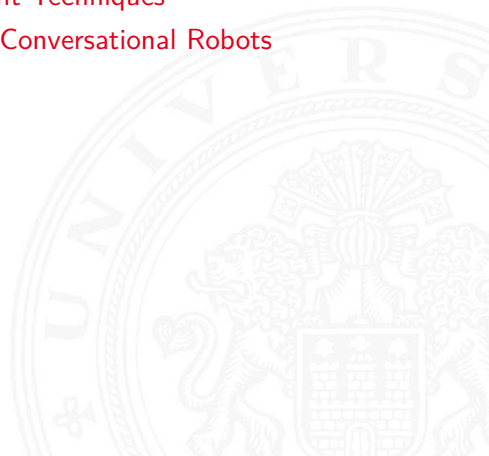


University of Hamburg
Faculty of Mathematics, Informatics and Natural Sciences
Department of Informatics
Technical Aspects of Multimodal Systems

26 November 2018



1. Motivation and Introduction
2. Communication via Spoken Dialogue Systems
3. Working Flow of Dialogue Systems
4. Different Dialogue Management Techniques
5. Common Limitations of Early Conversational Robots
6. A Path to Follow
7. A Recent HRI Implementation
8. Conclusion



Can robots *fluidly converse with humans in natural language*?

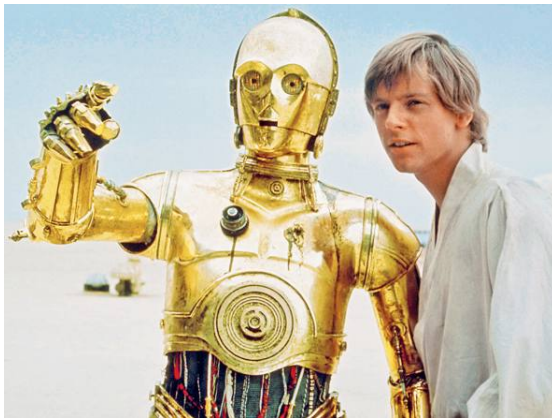


Figure: C-3PO and Luke Skywalker from Star Wars,

<http://www.calto.info/topics/3po-luke-skywalker-on.html>



What is HRI?

“Human–Robot Interaction (HRI) is a field of study dedicated to understanding, designing, and evaluating robotic systems for use by or with humans.” (Goodrich et al, 2008 [1])

- ▶ Remote interaction
- ▶ Proximate interaction
 - ▶ Physical interaction
 - ▶ ***Social interaction***: social, emotive and cognitive aspects of interaction

- ▶ Flexible manufacturing robots
- ▶ Lab or household robotic assistants
- ▶ Assistive robotics
- ▶ Robotic receptionists
- ▶ Robotic educational assistants
- ▶ Museum robots
- ▶ And many more...



www.slideshare.net/seokhwankim7/natural-language-in-humanrobot-interaction



www.blogcdn.com/www.engadget.com/media/2007/09/pic-servicerobot1.jpg

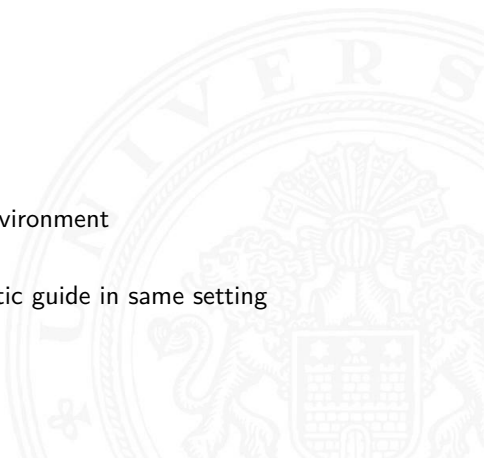


www.qries.com/assets/images/141941-1506423749.jpg



The first pioneering robots with natural language conversational abilities in the early 1990s [2]. Examples:

1. MAIA [3, 4]:
 - ▶ Mobile assistant robot
2. RHINO [5]:
 - ▶ Museum guide robot
3. AESOP [6]:
 - ▶ Surgical robot
4. Polly [7, 8]:
 - ▶ Robotic guide in an office environment
5. TJ [9]:
 - ▶ Slightly more advanced robotic guide in same setting





Communication via Spoken Dialogue Systems

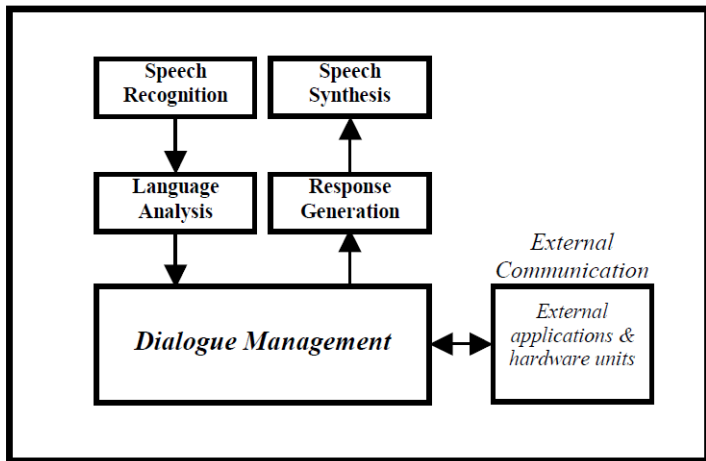


Figure: Architecture of dialogue systems [10]

Simplified Architecture of Dialogue Systems

Communication via Spoken Dialogue Systems

HRI through Natural Language Dialogue

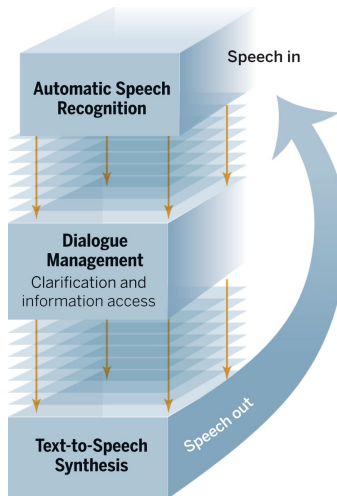


Figure: Simplified architecture of spoken dialogue systems [11]



Working Flow of Dialogue Systems

1. Speech Recogniser
 - ▶ Responsible for speech-to-text conversion
2. Language Analyser
 - ▶ Responsible for building a logical representation
3. Dialogue Manager
 - ▶ Responsible for communicating with robot's controller and creating a follow-up message
4. Response Generator
 - ▶ Responsible for creating response in written form
5. Speech Synthesizer
 - ▶ Responsible for text-to-speech conversion



Different Dialogue Management Techniques

- ▶ ***State-based:***
 - ▶ Most popular and simplest dialogue management technique [10]
- ▶ ***Frame-based:***
 - ▶ Frames instead of series of states [10]
- ▶ ***Plan-based:***
 - ▶ Identification of the user's plan and contribution to the execution of the plan [10]
 - ▶ Dynamic process: dynamic perception of the plan and system's contribution [10]

State-based Dialogue Manager

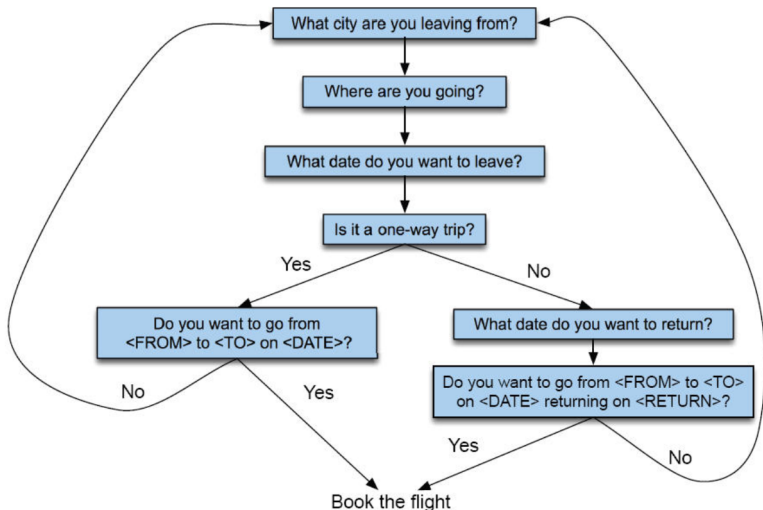


Figure: State-based Dialogue Manager Example [12]

Frame-based Dialogue Manager

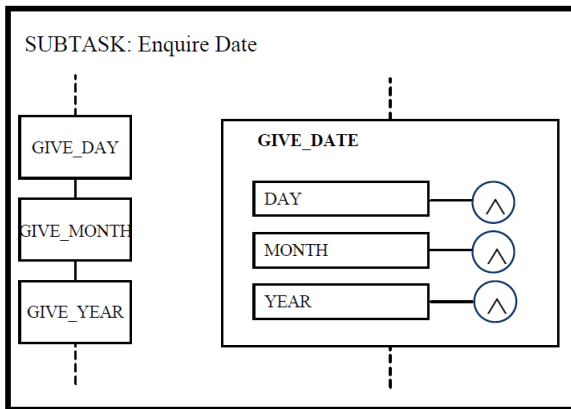


Figure: Frame-based Dialogue Manager Example [10]

Plan-based Dialogue Manager

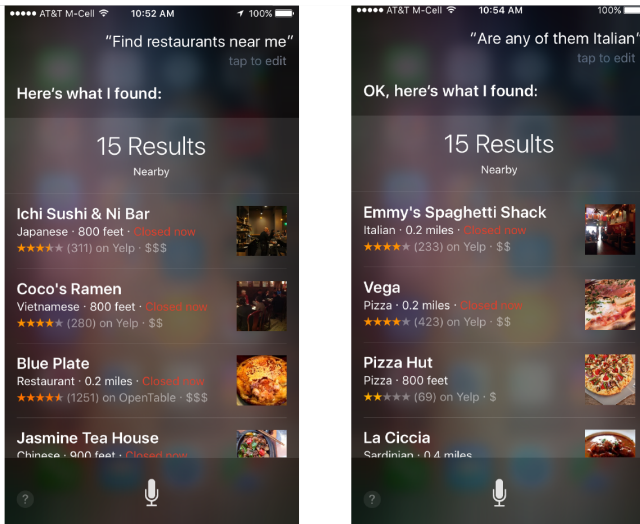


Figure: Plan-based Dialogue System Example: Siri [12]



Common Limitations of Early Conversational Robots [2]

Common Limitations of Early Conversational Robots

HRI through Natural Language Dialogue

1. Recognition of only *simple commands* and response with *canned answers*
2. Handling requests only in terms of *speech acts*
3. Mostly *human initiative* dialogues, no flexibly mixed-initiative dialogues
4. No support for *situated language*
5. No recognition of *affective speech*: no recognition or generation of emotional speeches
6. Almost no *non-verbal* communication capability such as gestures, gait and facial expressions
7. Usually *stimulus-response* dialogue systems (no actual speech planning or purposeful dialogue generation)
8. No *real learning*: preprogrammed verbal behaviour



1. Mimic the human developmental pathway and build robots that can handle situated language [2]
2. Move to a wider spectrum of linguistic abilities [2]

The levels of increasing abstraction and detachment from concrete language to wider spectrum [2]:

- ▶ 1st Level: “**Here and now**”, existing concrete things
- ▶ 2nd Level: “**Now, existing concrete things**”, not restricted to “here”.
- ▶ 3rd Level: “**Past or present, existing concrete things**”, not restricted to “now”
- ▶ 4th Level: “**Imagined or predicted concrete things**”, not limited to actuality
- ▶ 5th Level: “**Abstract things**”, not restricted to “concrete things”



A Recent HRI Implementation

- ▶ A model of cognitive interaction for service robots by Lemaignan et al [13]
- ▶ Main assumption: internal adaption of human-level semantics paves the way for human-level interaction
- ▶ Recognition, understanding and participation in communication
 - ▶ Explicitly (Verbal)
 - ▶ Implicitly (Pointing)
- ▶ *Situated, natural and multi-modal dialogue*

Architecture of the System

A Recent HRI Implementation

HRI through Natural Language Dialogue

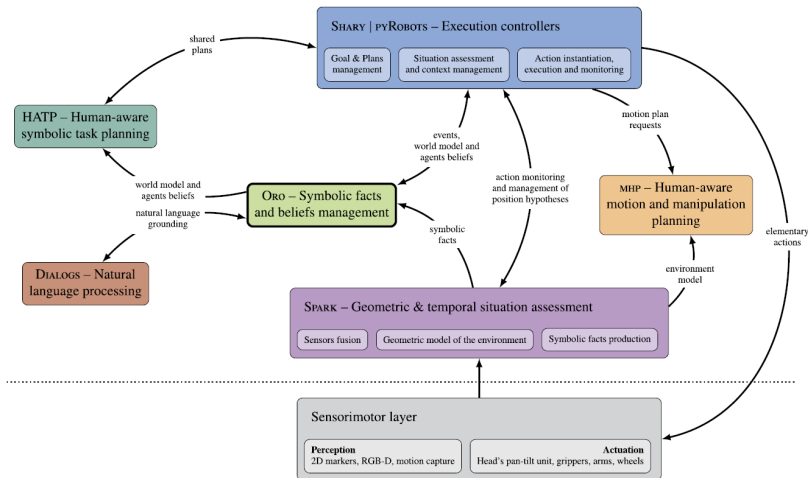


Figure: [13]'s Architecture



Main Modules of the System

A Recent HRI Implementation

HRI through Natural Language Dialogue

- ▶ Active knowledge base (ORO): semantic blackboard that connects most of the modules
- ▶ Geometric reasoning module (SPARK): quickly produces symbolic assertions of the environment and its changes over time
- ▶ *Language processing module (DIALOGS)*: queries knowledge base and writes back assertions
- ▶ Symbolic task planner (HATP): uses the knowledge base to initialise planning domain and returns a symbolic plan to execution controller
- ▶ Execution controller (SHARY/PYROBOTS): executes plans and monitor them



Main Features of DIALOGS

A Recent HRI Implementation

HRI through Natural Language Dialogue

- ▶ Retrieval of *speech input* from human through an Android-based interface, which relies on the Google speech recognition API for speech-to-text and feeds *the textual transcript* into robot.
- ▶ Parsing the text into a *grammatical structure* by a heuristics-based parser
- ▶ Resolution of the resulting pieces with the help of ORO for *grounding concepts* such as objects and actions

Interactive Grounding (Experiment)

A Recent HRI Implementation

HRI through Natural Language Dialogue



Figure: Interactive Grounding in a Messy Environment [13]

1. The person asks the robot to pass him a video tape
2. DIALOGS processes the sentence, queries the ontology to identify the object that the person refers to.
3. Two video tapes are visible to the robot: one on the table, the other in the cardboard box.
4. Since only the tape on the table is visible to the person, NL processor recognises that human is referring to the tape on the table.



Disambiguation through Pointing (Experiment)

A Recent HRI Implementation

HRI through Natural Language Dialogue



Figure: Disambiguation through Pointing [13]

1. Another person asks the robot: "What's in the box?"
2. Since two boxes (toolbox, cardboard box) are visible to both the robot and the person, it needs to find which box is referred.
3. Robot responds back with a question: "Which box, toolbox or cardboard box?"
4. Person responds by pointing out at the cardboard box. SPARK identifies that the person referred to the cardboard box.



- ▶ HRI: intriguing subfield of robotics, with its own *characteristics and challenges*.
- ▶ Natural languages: probably the most complex thing that humanity has ever created.
- ▶ Conversations with *allusions, metaphors* etc.
- ▶ Very unlikely to have robots with *human-level* communication capabilities from today to tomorrow
- ▶ Dialogue agents with the likes of Siri and Alexa are promising
- ▶ Possibility of robots with human-level natural language capabilities to be part of our everyday lives *in the coming decades*



Thank you for your attention.
Any questions?



- [1] M. A. Goodrich, A. C. Schultz, *et al.*, "Human-robot interaction: a survey," *Foundations and Trends® in Human-Computer Interaction*, vol. 1, no. 3, pp. 203–275, 2008.
- [2] N. Mavridis, "A review of verbal and non-verbal human-robot interactive communication," *Robotics and Autonomous Systems*, vol. 63, pp. 22–35, 2015.
- [3] G. Antoniol, R. Cattoni, M. Cettolo, and M. Federico, "Robust speech understanding for robot telecontrol," in *Proceedings of the 6th International Conference on Advanced robotics*, pp. 205–209, 1993.
- [4] G. Antoniol, B. Caprile, A. Cimatti, and R. Fiutem, "Experiencing real-life interactions with the experimental platform of maia," in *In Proceedings of the 1st European Workshop on Human Comfort and Security*, Citeseer, 1994.
- [5] W. Burgard, A. B. Cremers, D. Fox, D. Hähnel, G. Lakemeyer, D. Schulz, W. Steiner, and S. Thrun, "The interactive museum tour-guide robot," in *Aaai/iaai*, pp. 11–18, 1998.
- [6] L. Versweyveld, "Voice-controlled surgical robot ready to assist in minimally invasive heart surgery," *Virtual Medical Worlds Monthly*, 1998.
- [7] I. Horswill, "Polly: A vision-based artificial agent," in *AAAI*, pp. 824–829, 1993.
- [8] I. Horswill, "The design of the polly system," *The Institute for the Learning Sciences, Northwestern University, Tech. Rep.*, 1996.
- [9] M. C. Torrance, "Natural communication with mobile robots," *Master's thesis, Massachusetts Institute of Technology, Cambridge, MA*, 1994.
- [10] D. Spiliotopoulos, I. Androutsopoulos, and C. D. Spyropoulos, "Human-robot interaction based on spoken natural language dialogue," in *Proceedings of the European workshop on service and humanoid robots*, pp. 25–27, 2001.
- [11] J. Hirschberg and C. D. Manning, "Advances in natural language processing," *Science*, vol. 349, no. 6245, pp. 261–266, 2015.
- [12] D. Jurafsky, "Conversational Agents." <https://web.stanford.edu/class/cs124/lec/chatbot17.pdf>, 2017 (accessed November 6, 2018).
- [13] S. Lemaignan, M. Warnier, E. A. Sisbot, A. Clodic, and R. Alami, "Artificial cognition for social human-robot interaction: An implementation," *Artificial Intelligence*, vol. 247, pp. 45–69, 2017.