

# Measurement and Analysis of Interactive Behavior in Tutoring Action with Children and Robots

Anna-Lisa Vollmer

Diplom Mathematikerin Anna-Lisa Vollmer  
AG Angewandte Informatik  
Research Institute for Cognition and Robotics (CoR-Lab)  
Technische Fakultät Universität Bielefeld  
email: avollmer@techfak.uni-bielefeld.de

Abdruck der genehmigten Dissertation zur Erlangung  
des akademischen Grades Doktor-Ingenieur (Dr.-Ing.).  
Der Technischen Fakultät der Universität Bielefeld  
am 06.07.2011 vorgelegt von Anna-Lisa Vollmer,  
am 31.08.2011 verteidigt und genehmigt.

Gutachter:

Dr.-Ing. Britta Wrede, Universität Bielefeld

Dr.-Ing. Jannik Fritsch, Honda Research Institute Europe, Offenbach/Main

Prof. Angelo Cangelosi, University of Plymouth

Prüfungsausschuss:

apl. Prof. Dr.-Ing. Stefan Kopp, Universität Bielefeld

Dr.-Ing. Britta Wrede, Universität Bielefeld

Dr.-Ing. Jannik Fritsch, Honda Research Institute Europe, Offenbach/Main

Prof. Angelo Cangelosi, University of Plymouth

Dr.-Ing. Thies Pfeiffer, Universität Bielefeld

Gedruckt auf alterungsbeständigem Papier nach ISO 9706.

# Measurement and Analysis of Interactive Behavior in Tutoring Action with Children and Robots

Der Technischen Fakultät der Universität Bielefeld  
zur Erlangung des Grades

*Doktor der Ingenieurwissenschaften*

vorgelegt von

Anna-Lisa Vollmer

Bielefeld, Juli 2011



## Abstract

Robotics research is increasingly addressing the issue of enabling robots to learn in social interaction. In contrast to the traditional approach by which robots are programmed by experts and prepared for and restricted to one specific purpose, they are now envisioned as general-purpose machines that should be able to carry out different tasks and thus solve various problems in everyday environments. Robots which are able to learn novel actions in social interaction with a human tutor would have many advantages. Unexperienced users could “program” new skills for a robot simply by demonstrating them.

Children are able to rapidly learn in social interaction. Modifications in tutoring behavior toward children (“motionese”) are assumed to assist their learning processes. Similar to small children, robots do not have much experience of the world and thus could make use of this beneficial natural tutoring behavior if it was employed, when tutoring them.

To achieve this goal, the thesis provides theoretical background on imitation learning as a central field of social learning, which has received much attention in robotics and develops new interdisciplinary methods to measure interactive behavior. Based on this background, tutoring behavior is examined in adult-child, adult-adult, and adult-robot interactions by applying the developed methods. The findings reveal that the learner’s feedback is a constituent part of the natural tutoring interaction and shapes the tutor’s demonstration behavior.

The work provides an insightful understanding of interactional patterns and processes. From this it derives feedback strategies for human-robot tutoring interactions, with which a robot could prompt hand movement modifications during the tutor’s action demonstration by using its gaze, enabling robots to elicit advantageous modifications of the tutor’s behavior.

## Acknowledgements

I gratefully acknowledge the financial support from Honda Research Institute Europe for the project “Acquiring and Utilizing Correlation Patterns across Multiple Input Modalities for Developmental Learning”.

# Contents

<b>List of Figures</b>	<b>ix</b>
<b>List of Tables</b>	<b>xi</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Robots for Household Use . . . . .	1
1.2 Robot Learning . . . . .	4
1.3 Main Goals and Objectives . . . . .	4
1.4 Outline . . . . .	5
<b>2 Social Interaction: Imitation Learning</b>	<b>7</b>
2.1 The Question of What to Imitate . . . . .	8
2.2 To Imitate or to Emulate—Definitions and Evidence from Neuro-Science and Behavioral Science . . . . .	9
2.3 Imitation Learning Approaches in Robotics . . . . .	12
2.4 Conclusion . . . . .	14
<b>3 Tutoring Behavior</b>	<b>15</b>
3.1 Behavior Modifications in Infant-Directed Interaction . . . . .	15
3.1.1 Motherese . . . . .	15
3.1.2 Motionese . . . . .	16
3.2 Operationalizing Motionese . . . . .	17
3.2.1 Methodology of Qualitative Data Analysis . . . . .	17
3.2.2 Annotations . . . . .	18
3.2.3 Visualizations . . . . .	20
3.2.4 Quantitative Measures for Motionese . . . . .	22

## CONTENTS

---

<b>4</b>	<b>Analyzing Tutoring Behavior</b>	<b>29</b>
4.1	The Motionese Corpus . . . . .	30
4.2	Motionese Compared to Modifications in Tutoring Robots . . . . .	31
4.2.1	Data . . . . .	33
4.2.2	Method . . . . .	35
4.2.3	Results . . . . .	37
4.3	Motionese Toward Children of Different Age . . . . .	42
4.3.1	Data . . . . .	42
4.3.2	Method . . . . .	43
4.3.3	Results . . . . .	43
4.4	Discussion . . . . .	44
<b>5</b>	<b>Analyzing Learner Behavior</b>	<b>47</b>
5.1	Feedback: Children’s Contribution to Tutoring Interactions . . . . .	48
5.1.1	Data . . . . .	49
5.1.2	Method . . . . .	49
5.1.3	Group 1: Prelexical Infants (8 to 11 months) . . . . .	49
5.1.4	Group 2: Early Lexical Infants (12 to 24 months) . . . . .	54
5.1.5	Group 3: Lexical Infants (25 to 30 months) . . . . .	56
5.2	Discussion . . . . .	56
<b>6</b>	<b>The Interactional Account of Motionese</b>	<b>59</b>
6.1	On the Loop of the Tutor’s Action Modifications and the Learner’s Gaze	59
6.1.1	Data . . . . .	60
6.1.2	Method . . . . .	60
6.1.3	Starting Point: Variability of Hand Trajectories . . . . .	62
6.1.4	Empirical Observations on the Interplay between the Tutor’s Hand Motions and the Learner’s Gaze . . . . .	64
6.1.5	Systematization: From Empirical Observations to Formal De- scription . . . . .	65
6.2	Discussion . . . . .	75
<b>7</b>	<b>A Human-Robot Interaction Study of Feedback in an Imitation Learn- ing Scenario</b>	<b>79</b>
7.1	Motivation to Investigate Online and Turn-based Feedback in a Demonstration- Action Loop . . . . .	79
7.2	Design and Realization . . . . .	81
7.2.1	Subjects . . . . .	82
7.2.2	Setting and Experimental Conditions . . . . .	82
7.2.3	Hypotheses . . . . .	93



## CONTENTS

---

7.2.4	Technical Realization . . . . .	94
7.3	Data Analysis . . . . .	96
7.3.1	Methods . . . . .	96
7.3.2	Results . . . . .	97
7.4	Discussion . . . . .	102
7.4.1	Feedback Strategies . . . . .	103
<b>8</b>	<b>Conclusion</b>	<b>105</b>
	<b>References</b>	<b>109</b>
<b>A</b>	<b>Conventions for Transcription</b>	<b>115</b>
<b>B</b>	<b>Table Overviewing the Results of the Analysis of Tutoring Behavior</b>	<b>121</b>
<b>C</b>	<b>Questionnaire for Human-Robot Interaction Study</b>	<b>123</b>
<b>D</b>	<b>Interview for Human-Robot Interaction Study</b>	<b>127</b>

## CONTENTS

---

# List of Figures

1.1	KUKA industrial robot . . . . .	2
1.2	ASIMO humanoid robot . . . . .	3
2.1	Call and Carpenter’s theoretical terms for reproduction behavior . . . .	10
2.2	Call and Carpenter’s theoretical terms for reproduction behavior . . . .	10
3.1	Three different eye gaze directions . . . . .	19
3.2	Segmentation of the cup-nesting action . . . . .	20
3.3	Example for visualization of individual hand trajectories . . . . .	21
3.4	Example for visualization of hand trajectories across all tutor’s . . . . .	22
3.5	Example for visualization of hand trajectories and learner’s gaze . . . . .	23
4.1	Objects and task instructions for the Motionese Corpus . . . . .	30
4.2	AAI and ACI setting . . . . .	31
4.3	Two ways of nesting cups . . . . .	34
4.4	The robot simulation “Aka-chan” . . . . .	34
4.5	ARI setting . . . . .	35
4.6	Motionese results in bar charts . . . . .	39
4.7	Contingency results in bar chart . . . . .	40
5.1	Course of the demonstration . . . . .	50
5.2	Example fragment group 1 . . . . .	51
5.3	Example fragment group 1 . . . . .	52
5.4	Child gaze . . . . .	54
5.5	Pointing . . . . .	57
6.1	Example trajectories . . . . .	63
6.2	Normalized trajectories . . . . .	63
6.3	Example trajectories including orienting devices . . . . .	69
6.4	Example trajectories including orienting devices functioning as repair activity. . . . .	69

## LIST OF FIGURES

---

6.5	Example trajectories: Anticipating gaze . . . . .	73
7.1	HRI setting . . . . .	83
7.2	Objects and task instructions . . . . .	84
7.3	Interaction loop . . . . .	84
7.4	Example sequence of imitation of a manner-crucial action . . . . .	85
7.5	Example sequence of imitation of a goal-crucial action . . . . .	86
7.6	Example sequence of emulation of a manner-crucial action . . . . .	87
7.7	Example sequence of emulation of a goal-crucial action . . . . .	88
7.8	Conditions . . . . .	89
7.9	Social gaze imitation . . . . .	90
7.10	Social gaze emulation . . . . .	90
7.11	Random gaze . . . . .	92
7.12	Static gaze . . . . .	92
7.13	Technical setup . . . . .	94
7.14	Event knowledge results for speed in bar charts . . . . .	98
7.15	Event knowledge results for range in bar charts . . . . .	99
7.16	Turn-based feedback results in bar charts . . . . .	100
7.17	Online feedback results in bar charts . . . . .	101
C.1	Questionnaire . . . . .	124
D.1	Interview . . . . .	128

# List of Tables

4.1	The subjects of the three different age groups of the Motionese Corpus.	31
4.2	The subjects considered in the analysis for adult-child (ACI), adult-adult (AAI), and adult-robot interaction (ARI).	35
4.3	Description of means and standard deviations for the groups. $F$ and $p$ values illustrate results of the ANOVA with hypothesis $df = 2$ and error $df = 29$ for all measures.	41
4.4	The subjects considered of the three different age groups.	43
5.1	The subjects of the three different age groups.	49
6.1	The subjects of age group 1.	61
7.1	The subjects of the different age groups and assigned robot online feedback condition.	82
A.1	Conventions for transcription used for manual annotations of video data.	115
B.1	Short summary of the results of Section 4.2.	122

## LIST OF TABLES

---

# 1

## Introduction

Robots are currently mainly applied in industrial settings. In factories they are reliable workers, stacking boxes and sorting products without getting tired. They do specific tasks very precisely, accurately, efficiently and always identically. They are fast, strong, but restricted to one specific purpose. Figure 1.1 shows an example of an industrial robot. Robots have also found their way in people's homes. There are robot vacuum cleaners, floor washers and lawn mowers helping by fulfilling simple, but disliked tasks in an increasing number of homes to date. Service robots are envisioned to soon take orders for even more complex tasks in every household.

### 1.1 Robots for Household Use

If a robot, which I ordered to wash the dishes and clean the windows, were delivered to my home, I would want the robot to be able to fulfill the tasks right away. A robot designed and programmed to put together cars would only be able to do the exact same movement as in the factory if I switched it on. It would not function at all or worse destroy my kitchen because it was only build and programmed for this one task of picking up a car part and placing it where it belongs in the body. Neither its program, nor its physical appearance permits the robot to carry out other tasks. Trying to use it in the household would be equivalent to wanting to vacuum-clean with a toaster.

For working in a household, the robot thus has to have an embodiment, which enables it to perform a variety of tasks. Because working in the household and helping humans mostly involves work, which normally humans would do, one approach is to let the robot have a human-like body, which enables it to cope with the household environment build for humans. Such robots with a human-like body are called "humanoids". See Figure 1.2 for an example of a humanoid robot in a household environment.

## 1. INTRODUCTION

---



**Figure 1.1: KUKA industrial robot** - An example of an industrial robot handling drink cartons. Image source: KUKA Roboter GmbH. Reprinted with permission.





**Figure 1.2: ASIMO humanoid robot** - An example of a humanoid robot in a household setting. Image source: Denise Cross, flickr. Reprinted with permission.

The tasks of doing the dishes and cleaning the windows for me seem to be easy, but the robot needs to have a lot of skills. Also, if it can do the one, it cannot do the other yet. Given that the robot can move like a human, the robot needs to be able to detect dirty dishes anywhere in my home, tell them apart from clean ones, know how to grasp and transport them without breaking anything, know where to find cloths or a sponge and dishwashing liquid in my home, know how to turn on the faucet, how to fill the sink with water, that it works better with hot water and so on. Thus, the robot has to have skills specially fit to the properties of my home and also to the current situation. The dirty dishes can never be found at exactly the same places and sometimes the cat is in the way.

It is impossible that I thought of everything before I placed my order, so that programmers would have prepared the robot for each possible situation it could encounter in my home by programming every detail. This would be simply unfeasible. Therefore, the robot needs to have the ability to learn new skills while in my home and generalize them to different situations.

But how should a robot learn? How should a programmer give the robot the ability that something, which was given to the robot, evolves, grows and emerges?

The field of machine learning deals with the development of algorithms to solve these questions. So far current approaches generally only allow for learning of statistical relations, like for example balancing a pole, and rely on a large set of examples as training data. The generalization of a task, which is more complex or sequential and involves

## 1. INTRODUCTION

---

far more uncertainties and chaos has not been done yet.

### 1.2 Robot Learning

The most obvious thing to do is to look to human development. How do children learn new skills? It is not easy to directly ask them, but it is clear, that they are able to explore their environment and learn from their own experience and from tutors. The general research topic of this work is aiming at studying this capability, namely to develop robots which can learn the way children learn.

Infants do not only learn alone, they receive a lot of support from their social environment, and also the robot is not alone, but in my home I know where to find things and how to do the dishes. This knowledge needs to be transferred to the robot.

To teach the robot, thus, it has to learn in a way, which is understandable for me, or at least so that I can naturally provide it with the necessary information without having to study informatics or read at least one book of manual and instructions.

### 1.3 Main Goals and Objectives

The general goal of this work is to let robots learn the way children learn; especially it focuses on how robots could learn new manipulative actions in social interaction.

But how do children acquire new skills in social interaction? And what exactly supports their learning? Assuming that children receive support from their social environment and are tutored by their caregivers, this goal is very challenging and involves several open questions.

- What constitutes a natural tutoring interaction?
- How can complex human behavior in naturalistic interactions be analyzed, especially with computational means on large sets of data?
- Children particularly learn new skills by observing others perform and by copying their behavior. If a robot had this capability, how would it know what is important about the shown action and what to copy?

This work addresses these questions and will present findings of detailed analyses of adult-child interaction—in a first step, focussing on the input that infants receive from their caregivers, which has been found to contain significant modifications in different modalities—, develops novel methods of analyzing human behavior, and proposes ways of how robots could learn in social interaction.

## 1.4 Outline

For the goal of enabling robots to learn in social interaction, in a first step, the concept of imitation learning is introduced in Chapter 2 and difficulties for robotic systems are mentioned. Imitation learning is discussed from the viewpoint of neuro-science and behavioral science regarding the question of what to imitate and imitation learning approaches in robotics are overviewed.

Chapter 3 explores the tutor's behavior as it might support learning. The two terms "motherese" and "motionese" are introduced describing certain tutoring behavior in adult-child interaction. The chapter also describes the developed methods of analyzing human interactional behavior.

In Chapter 4, these methods are used for the first analysis and comparison of tutoring behavior in adult-child, adult-adult and adult-robot interaction and the analysis of tutoring behavior toward children of different age. The findings reveal differences in the tutor's behavior and suggest important implications for human-robot tutoring interactions.

They lead to further analysis of the learner's behavior in tutoring situations presented in Chapter 5. The analysis concerns adult-child interactions with children of different age and investigates if the learners' contributions to the tutoring interactions differed according to the learners' age and abilities. Again important findings for human-robot tutoring interactions are discussed.

In Chapter 6, an interactional perspective is taken to bring together both the tutor's and the learner's behavior. In this chapter, interactional patterns are identified and reasons and effects of each participant's actions are determined.

Chapter 7 reports a human-robot interaction study, which aims at investigating the insights drawn from the previous analyses by employing two different kinds of movement reproduction behavior in an imitation learning scenario.

The thesis concludes with a summary of the presented analyses and discusses shortcomings, as well as research questions open for future consideration in Chapter 8.

## 1. INTRODUCTION

---

## 2

# Social Interaction: Imitation Learning

The goal of learning in social interaction is to acquire a novel skill, which can then be reproduced and also applied to new situations. Human children master this endeavor with ease. From an evolutionary point of view, copying the behavior of others has subserved the ability of tool-use, but also allows for engaging in social and collaborative interactions (Carpenter and Call, 2007; Tomasello, 1999; Tomasello et al., 2005). Humans are able to learn complex object-related skills by observing others perform. This way of learning by imitation is safer and more efficient than trial and error learning behaviors for example. Many different terms have been used in cognitive, developmental and neuro-science to describe the phenomenon of imitation learning, but they have not been properly defined and are not easily distinguished. In this chapter, definitions of imitation and related terms are discussed from the perspective of different disciplines. Having obtained much attention in behavioral science and neuro-science, the term and its underlying processes have gained increasing importance in robotics as well. If robots were able to learn from demonstration, non-expert users would be able to “program” the robot by teaching it new skills and not every single detail would have to be pre-programmed. Humans would likely find this efficient way of teaching natural. Thus the following question arises: Which mechanisms are necessary for robotic systems to be able to learn novel skills in social interaction with a human tutor?

The current chapter focusses on one major issue, the question of what to imitate, meaning how to understand which aspects are important of a demonstrated action to achieve the intended outcome of the task. It provides an overview over definitions of the term imitation and related terms (e.g., emulation) and findings in the fields of neuroscience and behavioral science. Finally, current approaches to solving this issue for robotic systems are discussed identifying several open questions.

### 2.1 The Question of What to Imitate

To learn novel skills in social interaction is a challenging endeavor. Human children master it with ease, but for robotic systems several issues and problems become apparent. Four main questions have been formulated in robotics research recently (Breazeal and Scassellati, 2002; Nehaniv and Dautenhahn, 2000):

- Who to imitate
- When to imitate
- How to imitate
- What to imitate

The question of *who to imitate* is on the perceptual side of the system. The human teacher addressing the robot and also the onset of the action presentation have to be detected. Which human should a robot attend to, when there are several humans present? How should a robot distinguish if a human is tutoring it or if he/she is only normally interacting with it? The question of *when to imitate* is related to the previous question. When is the robot being tutored and when is it the robot's turn to reproduce the shown action? The question of *how to imitate* is on the motor side of the system. For example, the robot needs to know if a movement should be mirrored, when it is reproducing it. This is also a question of generalizing the action to be able to reproduce it in a different situation. Additionally, when learning from demonstration, the robot cannot profit from the tutor's sensory-motor information and thus has to map the tutor's movements onto its own body. This issue is also called the correspondence problem. The fourth question of *what to imitate* is on the perceptual side again and is concerned with what is necessary or important about a presented action and what is unnecessary or incidental. When reproducing a presented action, the robot would know which parts are important, which it should try to copy exactly, and which ones allow for more modifications and alterations in the reproduction.

The question of what to imitate is especially crucial in order to generalize a learned skill to a new situation and is in the focus of this chapter. To know what to imitate is equivalent to understanding the goal (i.e., the intended outcome) of a demonstrated action, cf. Section 2.2.

Children face the same question of what to imitate from a tutor's demonstration and thus, how to infer the goal of a demonstrated action. Brugger, Lariviere, Mumme, and Bushnell investigated the mechanism of determining what is important to achieve the goal of the task and what is not as one of the main mechanisms necessary for learning by imitation in studies with children and found that 14 to 16-months-old infants rely on their knowledge of causality in the physical world, but also exploit the tutor's social

## 2.2 To Imitate or to Emulate—Definitions and Evidence from Neuro-Science and Behavioral Science

---

signals (Brugger et al., 2007). Furthermore, Nagai and Rohlfing argue that motionese behavior, special behavior modifications in infant-directed action described in Section 3.1.2, could help learners find what is important about a demonstrated action (Nagai and Rohlfing, 2007).

## 2.2 To Imitate or to Emulate—Definitions and Evidence from Neuro-Science and Behavioral Science

Many different terms have been used in cognitive, developmental and neuro-science to describe the phenomenon of imitation learning, but they have not been properly defined and are not easily distinguished. According to Thorpe, true imitation is to acquire behavior by copying a demonstrator’s behavior (Thorpe, 1956). This is only the case, if i) the imitated behavior is a new behavior for the learner, ii) the same actions the demonstrator employed are reproduced, and iii) the learner understands the demonstrator’s intention and achieves the same goal (Tomasello et al., 1993). Employing Tomasello’s approach of contrasting social learning behaviors on the basis of the types of information and the sensitivity to intentions (Tomasello, 1990), Call and Carpenter further identify that a tutor’s demonstration reveals three elements to the learner: *goals*, *actions*, and *results* (Call and Carpenter, 2002):

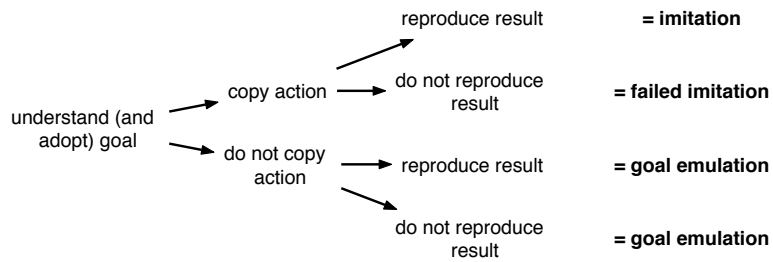
[...] a demonstrator’s model releases at least three products: goals, actions, and results.

The *goals* are the intended outcome of the task. The *actions* are the demonstrated movements or “motor patterns” with which a certain *result* (i.e., an effect or change in the physical environment) is obtained. Obviously the goal of an action is most difficult to infer, because it is not as easily observable as action and result. Accordingly, Call and Carpenter summarize and define concepts of reproduction behaviors based on which of the three elements it comprises, see Figure 2.1 and 2.2.

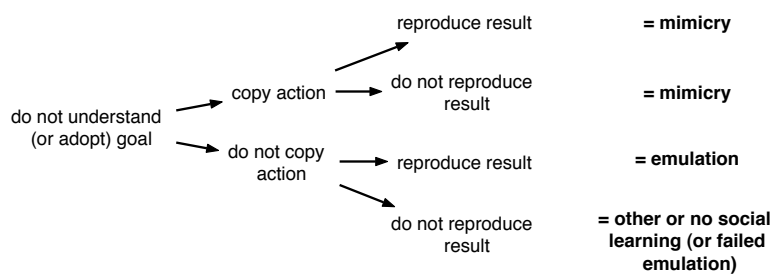
Here, for imitation the learner’s understanding of the goal is required. *Imitation* thus involves copying of the action, reproducing the result, and understanding the goal. If the result is not reproduced in an imitation attempt, the theoretical term the authors use is *failed imitation*. If the action is not copied, but the goal is understood, the behavior is termed *goal emulation*, independent of whether the result is reproduced or not. It thus describes the intention rather than the effect of the behavior (Whiten and Ham, 1992). Without goal understanding, a copied action is called *mimicry*, not necessarily comprising the reproduction of the result (Tomasello et al., 1993). A behavior without goal understanding, in which the action is not copied, but only the result or the ends is/are reproduced, is called *emulation* (Tennie et al., 2006; Tomasello, 1990, 1999; Tomasello et al., 1987).

## 2. SOCIAL INTERACTION: IMITATION LEARNING

---



**Figure 2.1: Call and Carpenter's theoretical terms for reproduction behavior -**  
 The three sources of information reproduced determine the corresponding term: when the goal is understood. Figure adapted from (Call and Carpenter, 2002)



**Figure 2.2: Call and Carpenter's theoretical terms for reproduction behavior -**  
 The three sources of information reproduced determine the corresponding term: when the goal is not understood. Figure adapted from (Call and Carpenter, 2002)



## 2.2 To Imitate or to Emulate—Definitions and Evidence from Neuro-Science and Behavioral Science

---

Let's consider the example of dusting off a bookshelf with a feather duster. The products of a demonstration of this task would be the result of dust-free books, the action of waving the feather duster over the books, and the goal of cleaning the bookshelf. If the learner blows the dust off the books, the result of the task is achieved, because there is no more dust on the books. This behavior is called emulation. If the learner additionally has understood that the bookshelf should be cleaned (i.e., the goal of the task has been understood), but the action with the feather duster is not reproduced, the behavior is called goal emulation. If the learner has observed the demonstration and reproduces the task by only copying the action, thus doing the waving motion with the feather duster over the books without understanding the underlying goal, and possibly not even freeing the books from dust, this is called mimicry. For imitation, the learner has to reproduce both, action and result, and have an understanding of the intention of cleaning the bookshelf.

This work adopts Call and Carpenter's definitions of imitation, emulation and mimicry.

Evidence for imitation has been reported for human infants and even newborns, which led some researchers to argue that imitation is an innate human behavior (Meltzoff and Moore, 1983). Most animals are not able to imitate, which leads to the conclusion that imitation behavior is an expression of human higher intelligence (Schaal, 1999). Several studies have analyzed imitation behavior in primates and have shown that primates generally rather emulate than imitate (Call and Tomasello, 1994; Nagell et al., 1993; Tomasello et al., 2005). They also seem to attend more to the results of a demonstrated task rather than to the actions (Call and Tomasello, 1994). Inculturated primates however have been reported to also copy actions additionally to copying results (Buttelmann et al., 2007; Tomasello et al., 1993).

In the neurological research, so-called "mirror neurons" have been found in the F5 brain region of primates (Di Pellegrino et al., 1992; Fogassi et al., 1998; Gallese et al., 1996; Rizzolatti et al., 1996). Mirror neurons are characterized by being active when a specific behavior is observed in others, but also when it is executed. There is evidence that in humans a kind of mirror neuron system exists as well (Decety, 1996; Decety et al., 1994; Fadiga et al., 1995) and that this system involves an area in the human brain so far only associated with speech production, Broca's area (Rizzolatti and Arbib, 1998). Rizzolatti and Arbib thus argued that imitation could have helped to promote the development of communication skills (Rizzolatti and Arbib, 1998).

Human infants are neither exclusive imitators, nor are they exclusive emulators. Ten- nie, Call and Tomasello found in their study that twelve-months-old children rather emulated demonstrated actions, but older children of age 18 and 24 months imitated the action (Tennie et al., 2006). This suggests that during human development children

## 2. SOCIAL INTERACTION: IMITATION LEARNING

---

first focus on the results and then this focus changes to the actions at around the age of 18 months. Other research suggests that it is not the children’s age alone, which is consequential of the children imitating or emulating a shown action, but the difficulty of the task coupled with the children’s attention capabilities causes them to imitate or emulate (Bauer and Kleinknecht, 2002). Gergely, Bekkering and Kiraly found that 14-months old children choose the most effective means to reach a certain goal (Gergely et al., 2002) according to “the principle of rational action” (Csibra and Gergely, 1998; Gergely and Csibra, 2003), which is conform with findings of studies with older children (three to six year-olds), which suggest that imitation of children is goal directed (Bekkering et al., 2000). Gergely, Bekkering and Kiraly tested children in two conditions. In the first condition an experimenter pressed a light box with the forehead even though her hands were free to move in order to switch on the light. In the second condition the experimenter also switched on the light with her forehead, but this time she pretended to be cold and could not use her hands, which were hidden under a blanket she was covered in. 69% of the children *imitated* in condition 1 (i.e., they also pressed the light box with their forehead), but in condition 2, children rather *emulated* (i.e., pressed the light box with a hand), because the constraint does not apply to them, and only 21% *imitated* in this condition, which is significantly less than in condition 1. Call and Carpenter report findings suggesting that autistic individuals on the other hand tend to attend and reproduce results rather than actions (Call and Carpenter, 2002). They argue that this could lead to disadvantages in the development of social skills and the ability to understand others or it might—the other way around—be caused by them.

### 2.3 Imitation Learning Approaches in Robotics

Robots being able to learn in social interaction would have many advantages. Unexperienced users could teach robots in a natural way what would otherwise have to be implemented by experts. Therefore, robots should generalize skills from few demonstrations and for that infer the goal of the demonstrated action. To imitate actions previously shown by a human tutor poses many problems (see Section 2.1). For a robot, imitating a demonstrated movement is difficult, because the robot’s situation is never exactly the same as the one of the demonstrator. Situation here does not only include the positions of objects and actors in the physical world, but additionally for example embodiment, sensors, and perception. A large body of work on imitation learning exists (for an overview refer to (Argall et al., 2009; Schaal, 1999)), but in the field of robotics the term imitation does not have a clear definition either.

Approaches face problems, which can be divided into two categories: the perceptual and the motor side of the system (Schaal, 1999). On the perceptual side, the action

## 2.3 Imitation Learning Approaches in Robotics

---

that should be imitated can have several sources as categorized by Argall et al. (Argall et al., 2009). Research exists on demonstrations using techniques like teleoperation (Pook and Ballard, 1993) (e.g., a human remotely controls the robot or operates the robot by moving its limbs and putting the robot through the task, “kinesthetic teaching” (Billard et al., 2006)), or shadowing (Demiris and Hayes, 2002; Nicolescu and Mataric, 2001) (i.e., the robot “shadows” the behavior of the tutor and for example follows a tutor robot through a maze). These two techniques have the advantage that the robot can record the execution of the task using its own sensors and thus, the correspondence problem, mentioned in Section 2.1, on the motor side, does not need to be solved. Other techniques involve sensors on the teacher (Ijspeert et al., 2002), which record the movement as accurately as possible, and imitation from external observations (Atkeson and Schaal, 1997; Billard and Matarić, 2001), which is typically vision-based and involves the highest degree of uncertainty and the most sources of errors. Combinations of the latter two types of approaches have also been applied (Lopes and Santos-Victor, 2005)

The latter two data sources are in the focus here. To replicate a demonstrated movement, a model of the skill has to be created (i.e., a representation), which could serve as a metric for the system’s replication performance. As representation, high-level approaches suggest symbolic encodings using sets of predefined actions (Nicolescu and Matarić, 2005; Pardowitz et al., 2007; Saunders et al., 2006) and low-level approaches suggest trajectory-based encodings (Calinon et al., 2005; Mühlig et al., 2009), which are either on joint level, on task space level or hybrid variants. Regarding the motor side, replicating a movement on joint level is very difficult, as it again involves the correspondence problem, which is especially difficult to solve for systems with a high degree of freedom and complex motor control (e.g., humanoid robots) if it is not solved for the robot beforehand. Task space level approaches represent trajectories commonly in Cartesian coordinates, thus, reducing the dimensionality and easing or avoiding the correspondence problem. This leads to reproductions which can differ from the original presentation, since there can be several solutions of the inverse kinematics. There exists attempts to let the system select the task space autonomously (Gienger et al., 2010; Mühlig et al., 2009), but these also require a predefined pool of task spaces, which the system can select from, but generally the relevant task space is predefined by programming.

In the field of robotics, the term imitation is for the most part used for all kinds of behavior replication (Atkeson and Schaal, 1997; Dautenhahn, 1995; Hayes and Demiris, 1994). Thus, current imitation learning approaches are mainly about observing and replicating movements, but do not consider nor distinguish any variants of imitation as for example emulation or mimicry as defined in the previous section.

To perceive the important aspects of a demonstrated action and to infer its goals has

## 2. SOCIAL INTERACTION: IMITATION LEARNING

---

been addressed in only few works. In most approaches, the goal is given to the robotic system beforehand, but progress has been made. Recent approaches attempt to extract relevant information from the demonstrated action including the tutor's social signals (e.g., extracting the tutor's line of sight and follow it, recognizing facial expressions, and detecting affective vocalizations) (Breazeal et al., 2004). Other approaches aim at extracting important elements of the demonstration by observing the tutor perform multiple demonstrations of the same action and calculating which parts of the movement do not allow for variability (Calinon and Billard, 2007; Mühlig et al., 2009).

### 2.4 Conclusion

In the current chapter, different terms describing imitation learning have been presented to clarify the phenomenon. Among the terms, *imitation* and *emulation* were distinguished: Imitation was defined as reproducing the action and result and understanding the goal of a presented task, (goal-)emulation was defined as only reproducing the result. Advantages of robots possessing the ability to learn by imitation have been mentioned as well as issues in realizing this ability. The chapter focussed on the question of how learners—robots as well as children—could know what to imitate of an action presented by a tutor. Robotics research generally does not distinguish between different forms of reproduction. Children appear to imitate as well as emulate (Bauer and Kleinknecht, 2002; Gergely et al., 2002; Tennie et al., 2006). According to the principle of rational action, they choose the most effective means to reach the goal of the presented task (Gergely et al., 2002). To find what is important about the task, children seem to rely on their knowledge of causality and social cues given by the tutor (Brugger et al., 2007). Special tutoring behavior modifications observed in adult-child interaction, motionese, might as well facilitate this task (Nagai and Rohlfing, 2007) not only for children, but also for robots, which—similar to small children—have limited knowledge about the world. It is often argued that robots could induce the parental action modifications in tutoring interactions (Nagai et al., 2008).

Furthermore, a brief overview over robotic systems aiming at solving the issues arising when answering the question of what to imitate has been given.

# 3

## Tutoring Behavior

The previous chapter has pointed out that learning in human children is not only a concern of an individual. It is a social endeavor and children receive support from their social environment on multimodal levels.

Robots do not have the same experience and cognitive as well as physical abilities as humans, but infants do not have the same background and knowledge as grown-ups either. The idea is that robots could benefit from the tutoring behavior children are supported with and learn in social interaction with a human tutor.

In the current chapter, therefore, findings of research on adult-child interaction, which report behavior modifications when tutoring young children, are presented. These behavior modifications in different modalities, “motherese” and “motionese”, are introduced and features and recent research are discussed in Section 3.1. In order to further study this tutoring behavior for action learning, objective measures are presented in Section 3.2, which also are a first step toward making modifications available online for processing in a robotic system.

### 3.1 Behavior Modifications in Infant-Directed Interaction

When tutoring young children, adults modify their speech and also their motions. Vocal modifications are known as “motherese” and “motionese” is the term used for the modifications in infant-directed action.

#### 3.1.1 Motherese

The term “motherese” denotes all infant-directed speech (Newport, 1975). Motherese is modified compared to normal speech (Masataka, 2003). In (Masataka, 2003) and (Fernald, 1985) motherese has been reported to use a higher pitch, an exaggerated intonation (Fernald and Simon, 1984; Garnica, 1977), a simplified lexicon (Ferguson,

### 3. TUTORING BEHAVIOR

---

1964), longer pauses between utterances, and slower tempo (Fernald and Simon, 1984). It uses shorter utterances, fewer words per utterance, more repetition, simpler structure (Goodluck, 1991; Rohlfing et al., 2006). Infants seem to prefer motherese to adult-directed speech (Cooper, 1997; Fernald, 1985; Zangl and Mills, 2007). The exaggerated pitch contour was argued to be the most salient property of motherese for children's perception (Fernald and Kuhl, 1987). Not only mothers and fathers use motherese, but also nonparent adults (Masataka, 2003). Despite this evidence, there are many individual differences (Shute and Whezldall, 1995). Concerning the effect of motherese on infant's development, Fernald and Simon claim that motherese helps the infant parse the speech stream (Fernald and Simon, 1984). Furthermore, motherese seems to emphasize new information (Gleitman, L., & Wanner, 1984) and mark turn-taking phases (Snow, 1977). It is argued to elicit and maintain infants' attention, engage in interaction and communicated affect (Fernald, 1984; Stern et al., 1982) and thus, is argued to be beneficial for language acquisition (Fernald et al., 1989).

#### 3.1.2 Motionese

Additional to modifications in infant-directed speech, modification in movement can also be observed. These are called "motionese" (Brand et al., 2002). They range from modifications in posture and facial expressions (Chong et al., 2003; Stern, 1974) over a modified sign-language in deaf parents (Masataka, 1996) to modifications in gestures toward infants (Iverson et al., 1999). Brand and colleagues defined eight parameters to measure the degree of modification in a study of parental object demonstrations toward their infants (Brand et al., 2002). They reported that motionese compared to adult-directed action exhibited closer proximity to the partner, higher interactiveness, more enthusiasm and more repetition, was slower and simpler, had high exaggeration as range of the movements, and showed more structure in the form of pauses and more direct movements. Rohlfing, Fritsch, Wrede and Jungmann found parameters, which were objective and automatically computable on object demonstration video data (Rohlfing et al., 2006). Rohlfing et al. found that movement in child-directed interaction is more round and slower than in adult-directed interaction, and Vollmer et al. extended their measures (Vollmer et al., 2009a), cf. 3.2.4. Motionese was found to also be preferred in comparison to adult-directed action by infants (Brand and Shallcross, 2008) and was argued to assist infant's action learning, just like infants' language learning benefits from motherese (Brand et al., 2002; Iverson et al., 1999). Brand et al. suggested that motionese helps infants understand the structure and goal of an action and maintains their attention.

Basically research revealed equivalent findings for motionese and motherese, but in a different modality, suggesting that the concept of modifications in infant-directed interaction seems to hold across modalities (Masataka, 1996). In fact, it has been shown

that synchronous verbal labeling of objects and object movement facilitates children's associative learning of the two (Gogate and Bahrick, 1998; Gogate et al., 2000).

### 3.2 Operationalizing Motionese

For further investigation of tutoring behavior on the basis of experimental data and to make the behavior modifications in tutoring children revealed in recent research available for the online use of robots, objective measures are necessary to measure the tutor's behavior and benefit from it. Because the movement modifications are more directly linked to learning manipulative actions, this work focusses on analyzing modifications in motion cues: Motionese. Additionally, the use of speech recognizers is—due to the complexity of natural interaction—unfeasible in the current state of development.

Analyses were carried out on the video data of different studies. They focused on investigating tutoring behavior modifications toward the learner in motion, gesture, and eye gaze. To computationally assess these differences in human interactional behavior, is not a trivial straight-forward process. It is a combination of qualitative and quantitative techniques, manual and automatic working steps, iteratively leading to objective measurements applicable to the whole data set. Because of the high variability of the human conduct, data mining and statistical learning algorithms are unable to obtain relevant results. Video data used in the analyses includes two camera views. Each tutoring interaction was thus examined on the basis of a frontal view on the tutor and the demonstration he/she presented and a frontal view on the learner observing the demonstration. One possible step of analysis consists in a qualitative data analysis on few video sequences of individual subjects. Observations and relevant features obtained in this qualitative step can be applied to identify features and acquire a set of manual and semi-automatic annotations on the pairs of videos for each subject of the corpus in selected tasks and conditions. From these and also other additional annotations, visualizations were developed as a first step to quantification. The visualizations combine different modalities and aid again qualitative analysis as well as the reformulation of the initial observations into concrete hypotheses computationally assessable on measures calculated on the existing set of annotations. The objective measures are partly based on existing analyses of motionese (Brand et al., 2002, 2007; Rohlfing et al., 2006) and aim at calculating the visually observed modifications.

#### 3.2.1 Methodology of Qualitative Data Analysis

As a means to analyze naturalistic interaction and to deal with the difficulties of conducting fine-grained analyses, the manual, qualitative analysis of video-taped natural

### 3. TUTORING BEHAVIOR

---

interactions is based on Ethnomethodological Conversation Analysis (EM/CA) (Goodwin, 1979; Mondada, 2006; Pitsch, 2006). The analysis, which was carried out in collaboration with Karola Pitsch (Bielefeld University, Germany), aims at understanding the sequential organization and the problems the participants are solving in their interaction. EM/CA provides a methodology for fine-grained analysis of video-taped interaction data. Michael Forrester (Forrester, 1999) stated in his talk at the Symposium on Asymmetric interactions at the Center of Excellence Cognitive Interaction Technology in Bielefeld that “In EM/CA, the focus is always on participants’ sense-making practices—sometimes termed ‘members methods’, meaning everyday methods that people use to make sense” and to achieve mutual understanding in social interaction. EM/CA follows Garfinkels ideal of the unmotivated examination of data (i.e., it aims at revealing analytical categories from the data themselves without formulating a pre-existing analytical interest) (Garfinkel, 1967). The qualitative analysis here is also applied at a later stage, when hypotheses or annotations already exist and is then restricted to certain modalities and features. The procedure is strictly empirical and qualitative. Analysis begins with a single case analysis on the video data (i.e., repeated inspection of video-taped data), transcribing the interaction in order to observe temporal patterns and relationships of the events of all interaction partners. The qualitative analysis results in observations or initial hypotheses, at this stage formulated in a general way, and a set of relevant features to be manually or computationally assessed on the whole corpus of video data. Note that observations of one single case (which might have been considered because it is particularly interesting) do not have to scale to the whole set of data and might even be misleading when trying to find measures to compute on a large set of data.

#### 3.2.2 Annotations

Systematic annotations of the corpus have been conducted. All features were annotated in a manner that they could also be algorithmically computed. The demonstrator’s hand motions were annotated using a semiautomatic hand tracker system allowing for manual adjustment in case of tracking deviation. The two-dimensional motion tracker is based on an Optical Flow based algorithm (Lucas et al., 1981) and was implemented as a plugin for the graphical plugin shell *iceWing* (Lomker et al., 2006). The generated output text file contains a time-stamped list of two-dimensional coordinates of the tracked hands, defining their position in the video frame (e.g., based on the standard video format *576p25* (720 by 576 pixels with a frame rate of 25 hertz)).

Additionally, several annotators systematically and objectively annotated the following features. The annotations are independent of content and theory and allow to attach precise timestamps to interactional events, see (Vollmer et al., 2010). Annotators used



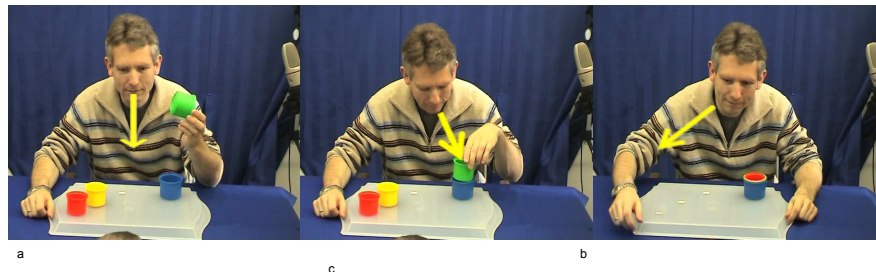
time-based annotation tools (ELAN and Interact, (Brugman and Russel, 2004; Mangold, 2006)) and verified each others work.

For the learner:

- Gaze: moving gaze toward a position. Possible eye gaze directions: all objects in scene, interaction partners face and hands, the experimenter. No annotation in case of occasional occlusion of infants face.
- Speech
- Pointing and reaching gestures: marked in three phases: preparation phase, peak phase, retraction phase.
- Smiles

For the adult:

- Gaze: eye gaze directions annotated with the program Interact (Mangold (Mangold, 2006)). Three categories of eye gaze directions were distinguished: looking at the interaction partner, looking at the object, and looking elsewhere (Figure 3.1).



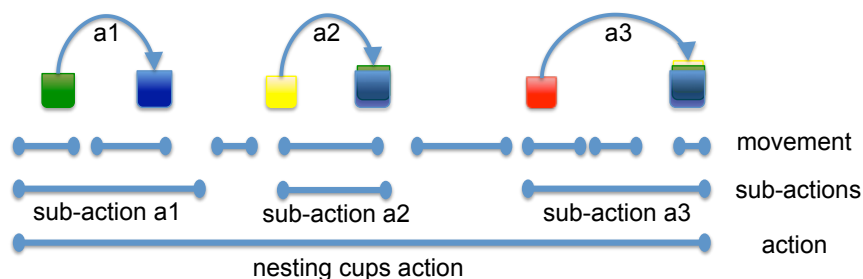
**Figure 3.1: Three different eye gaze directions** - Gaze to interaction partner (a), gaze to object (b), gaze elsewhere (c)

- Speech
- Action: the beginning and ending of actions and sub-actions were annotated as in (Vollmer et al., 2009a). *Sub-actions* correspond to the transportation of the objects involved or the action segments until completing sub-goals in a divided task. The *action* comprises all sub-actions, thus starting at the beginning of the first sub-action until the end of the last sub-action. In the example of a cup-nesting task, the sub-actions correspond to the transportation of the cups: a1, a2, a3, see Figure 3.2. One sub-action thus began, right when the tutor lifted one cup and ended, when releasing it into the blue cup, which is the end position.

### 3. TUTORING BEHAVIOR

---

Accordingly, the action was annotated as the whole process of transporting all objects to their goal positions.



**Figure 3.2: Segmentation of the cup-nesting action** - The action was divided into three sub-actions, each corresponding to the transportation of a cup.

The most detailed annotations mainly of the children’s gaze, speech, gestures and facial expressions, were structured and standardized using a set of conventions developed beforehand (see Appendix A.1). To integrate the different XML and text-based data structures for subsequent analysis, timestamps and annotation values are parsed from the transcripts and loaded into MATLAB (MATLAB version 7.10.0 (R2010a), The MathWorks Inc., Natick, Massachusetts) for further processing (i.e., visualization and computational investigation of the features and relevant aspects obtained in the qualitative analysis).

#### 3.2.3 Visualizations

An important part of quantification of the qualitative findings consists of visualizing the annotated data set in ways which combine different types of information in one representation and thus unveiling rather hidden interrelations which are hardly perceivable only considering the video data. These visualizations then help on the one hand to quickly be able to refine the qualitative observations on a few videos now on the whole data set and on the other hand to derive further systematic hypotheses which are easily computationally testable. Three ways of visualizing the data are presented. The first one shows the hand trajectories of one tutor plotted on top of a video still frame of the respective video and highlighting the transportations of the cups in color, hence revealing the shape of the tutor’s hand movements and considering the setting at the same time. The second way of visualization shows how the shapes of the three movements of transporting the cups differ in adult-child and adult-adult interaction by showing normalized versions of the sub-action trajectories. The third visualization links the tutor’s hand trajectories with the tutor’s gazing direction and at the same time shows the learner’s gaze at each time step. Beforehand, the annotations were

parsed and loaded into MATLAB, some were in a first step divided into more abstract categories created employing knowledge obtained in the qualitative analyses (Section 3.2.1).

### Visualizing Hand Trajectories

The hand trajectories obtained by the half-automatic hand tracking tool are drawn on a video frame showing a frontal view of the tutor with the object in the home position, meaning the position from which it is taken to be transported to its goal position. Employing the annotations of the structure of the action, the intervals corresponding to each sub-action are used to highlight those parts of the trajectory, which depict the hand motions when transporting the object. In the example image Figure 3.3 of a cup-nesting task, the transportation of each cup is colored in the respective cup color. This visualization enables us to directly view the trajectory shape and to compare the movements of transporting movements of different objects in a task with repeating structure.



**Figure 3.3:** Example for visualization of individual hand trajectories - Green/yellow/red trajectories mark the actions of stacking the cup of the corresponding color into the blue one. Thin lines represent movements without cups.

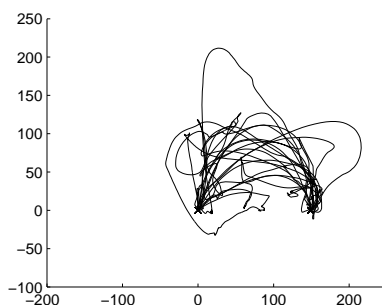
### Visualizing Hand Trajectories across all Video Data

Again the hand trajectory data are taken and cut according to the sub-actions of transporting the object or objects. For all tutors of the data set, the trajectory part corresponding to the first sub-action in the adult-child interaction is taken and transformed to have the same starting and end point. Then the sub-action trajectories are plotted in one coordinate system. The same is done for the other two sub-actions and for all sub-actions of the adult-adult interaction in separate images. The purpose of these pictures is to show overall differences in the shape of the sub-action trajectories

### 3. TUTORING BEHAVIOR

---

for the adult-child and adult-adult interactions, see Section 6.1.3. Figure 3.4 shows an example of adult-child trajectories for the first sub-action (i.e., the transportation of the first cup to the goal cup, in a nesting cups task).



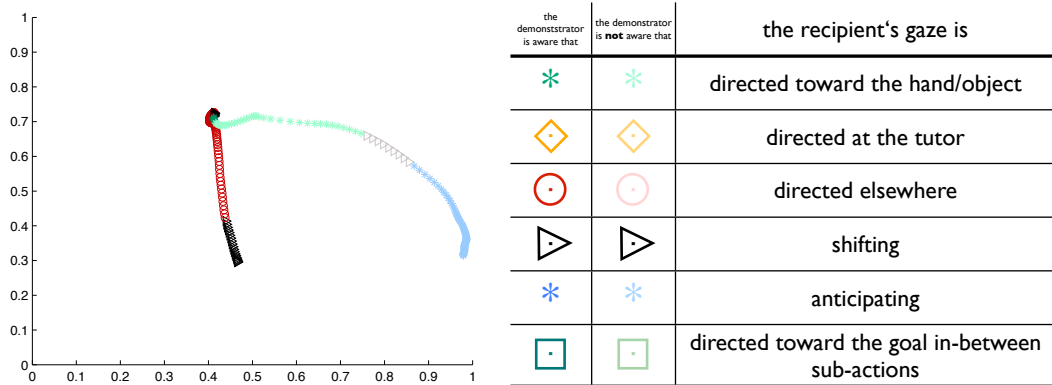
**Figure 3.4:** Example for visualization of hand trajectories across all tutor’s - The normalized trajectories for the first sub-action of the nesting cups task in adult-child interactions.

#### Visualizing Hand Trajectories, Tutor’s and Learner’s Gaze

The following visualization enables us to directly detect those moments, in which the learner is attentive to the action presentation and where he/she is not (marked in green vs. red), the child’s anticipating gaze defined in Section 6.1.5 (blue), whether the tutor is aware of the learners state of attention or not (dark vs. light) and how those instances link to the hand motion. This type of picture enables us to see series of subsequent changes in the participants’ orientation and the precise moments at which they occur in relation to each other and to the concrete shape of the hand trajectory, see Figure 3.5. The visualization thus can support qualitative analyses, but with the fusion of multiple modalities, it is also able to facilitate the generation of hypotheses concerning the interrelationship of the shown cues.

#### 3.2.4 Quantitative Measures for Motionese

Motionese parameters were defined, which serve as a measure for motionese behavior and more abstract features were computed from the annotated data. The motionese measures described here in detail are used for the analyses presented in Chapters 4, 5, 6, and 7. They are partly developed after (Brand et al., 2002, 2007; Rohlfing et al., 2006) and thus are based on previous research. The measures mainly deal with the annotated hand trajectories of the tutor and the tutor’s gaze, but also with the learner’s actions. Brand, Baldwin, and Ashburn measured motionese by means of rating the



**Figure 3.5: Example for visualization of hand trajectories and learner's gaze -** The trajectory for the second sub-action of the nesting cups task for one subject. Color codes for the learner's gaze.

infant-directed demonstrations in eight categories: interactiveness, enthusiasm, proximity to partner, range of motion, repetitiveness, simplicity, rate, and punctuation (Brand et al., 2002). They found that demonstrations to infants were higher in interactiveness, enthusiasm, proximity to partner, range of motion, repetitiveness, and simplicity. Rohlfing, Fritsch, Wrede and Jungmann derived from the manually coded measures proposed by Brand et al. a set of objective criteria for the movement modifications computable on the video streams using an automatic three-dimensional hand tracking system (Rohlfing et al., 2006). They segmented the hand trajectory of the demonstrations into actions and pauses and developed the motion parameters: pace, roundness, velocity, and acceleration. Rohlfing et al. did not find a significant effect for velocity for the three-dimensional posture tracking data. Their two-dimensional hand tracking data showed the statistically significant trend that hand movement in adult-adult interaction (AAI) is faster than in adult-child interaction (ACI). For pace, the authors found nearly significant differences comparing ACI and AAI. Their results suggest that pace values in ACI are lower than in AAI. They also found that hand movement is significantly rounder in AAI compared to ACI. The following measures for detecting motionese by means of the tutor's hand trajectories are based on Rohlfing et al.'s motion parameters:

Given:

- frame  $t \in \{1, \dots, d\}$ ,  $d \in \mathbb{N}$  last frame of action,
- duration  $\tau$  of one frame in seconds,
- action  $A = [1, d]$ ,  $d \in \mathbb{N}$ ,

### 3. TUTORING BEHAVIOR

---

- image coordinate  $x_t = (x_{t,1}, x_{t,2})$  of the hand executing the task in frame  $t$ ,  
 $x_{t,1} \in \{1, \dots, 720\}, x_{t,2} \in \{1, \dots, 576\}$ ,
- sub-actions  $a_1, a_2, a_3$  with  $a_i = [d_{i,1}, d_{i,2}], d_{i,j} \in \mathbb{N}, d_{i,1} < d_{i,2}, d_{1,j} < d_{2,j} < d_{3,j}$ ,
- movement threshold  $s \in \mathbb{R}, s > 0$ .

In a first step, the action was frame-wise automatically divided into movements and motion pauses, see Figure 3.2. For this the *velocity* of the tutor's hand from frame  $t$  to  $t + 1$  was computed as an approximation of the derivative of the two-dimensional hand coordinates of the hand which performed the action,

$$v_t = \frac{x_{t+1} - x_t}{\tau}. \quad (3.1)$$

Thus, a movement  $M_i$  was defined as a sequence of three or more consecutive frames  $t \in A$  with velocities  $\|v_t\| > s$ ,

$$\begin{aligned} \mathbf{movements}(A) = \mathbf{M} = \{M_i = [t, t + m_i] \mid t \in A, t + m_i \leq d, \\ v_j > s, j \in \{t, \dots, t + m_i - 1\}, m_i > 2\}. \end{aligned} \quad (3.2)$$

Analogously, a pause  $P_i$  was defined as a sequence of three or more consecutive frames  $j \in A$  with velocities  $v_j \leq s$ ,

$$\begin{aligned} \mathbf{pauses}(A) = \mathbf{P} = \{P_i = [t, t + p_i] \mid t \in A, t + p_i \leq d, \\ v_j \leq s, j \in \{t, \dots, t + p_i - 1\}, p_i > 2\}. \end{aligned} \quad (3.3)$$

Hence,

$$A = [1, \dots, M_{i-2}, P_{i-1}, M_{i-1}, P_i, M_i, P_{i+1}, \dots, d], \quad M_t \in \mathbf{M}, P_t \in \mathbf{P}. \quad (3.4)$$

The mean velocity is computed for each movement  $M_i$  as

$$\mathbf{velocity}(M_i) = \frac{1}{m_i} \sum_{j=t}^{t+m_i-1} v_j, \quad (3.5)$$

and the mean velocity for action  $A$  as

$$\mathbf{velocity}(A) = \frac{1}{|\mathbf{M}|} \sum_{i=1}^{|\mathbf{M}|} \mathbf{velocity}(M_i) \quad (3.6)$$

Equivalently, *acceleration* was computed as an approximation of the second derivative,

$$a_t = \frac{v_{t+1} - v_t}{\tau}. \quad (3.7)$$

for movement  $M_i$  as

$$\mathbf{acceleration}(M_i) = \frac{1}{m_i} \sum_{j=t}^{t+m_i-1} a_j, \quad (3.8)$$

and for action  $A$  as

$$\mathbf{acceleration}(A) = \frac{1}{|\mathbf{M}|} \sum_{i=1}^{|\mathbf{M}|} \mathbf{acceleration}(M_i). \quad (3.9)$$

*Pace* was defined for each movement by dividing the duration of the movement by the duration of the preceding pause,

$$\mathbf{pace}(M_i) = \frac{m_i}{p_i}. \quad (3.10)$$

Thus,

$$\mathbf{pace}(A) = \frac{1}{|\mathbf{M}|} \sum_{i=1}^{|\mathbf{M}|} \mathbf{pace}(M_i). \quad (3.11)$$

*Roundness* of a movement was defined by covered motion path divided by the distance between motion on- and offset,

$$\mathbf{roundness}(M_i) = \frac{\sum_{j=t}^{t+m-1} \|x_{j+1} - x_j\|}{\|x_{t+m_i} - x_t\|}, \quad (3.12)$$

and

$$\mathbf{roundness}(A) = \frac{1}{|\mathbf{M}|} \sum_{i=1}^{|\mathbf{M}|} \mathbf{roundness}(M_i). \quad (3.13)$$

Thus, a higher value in roundness means rounder movements.

*Frequency of motion pauses* was defined as the number of motion pauses per minute. Therefore, the number of motion pauses was computed automatically using the above-mentioned segmentation into movements and pauses:

$$\mathbf{fmp}(A) = \frac{60}{\tau} \cdot \frac{|\mathbf{P}|}{d} \quad (3.14)$$

### 3. TUTORING BEHAVIOR

---

Further, the *average length of motion pauses* (in frames) was computed as

$$\mathbf{almp}(A) = \frac{1}{|\mathbf{P}|} \sum_{i=1}^{|\mathbf{P}|} p_i. \quad (3.15)$$

The *total length of motion pauses* was computed as the percentage of time of the action without movement,

$$\mathbf{tlmp}(A) = \frac{100}{d} \sum_{i=1}^{|\mathbf{P}|} p_i. \quad (3.16)$$

Additionally, the trajectory during the actual transportation of the cups, when performing the task, was investigated. For each video and setting, the exact video frames of the beginnings and ends of the transportation for each of the three cups were annotated by hand, see Figure 3.2. This makes it possible to define variables for each individual sub-action ( $a_1, a_2, a_3$ ) and also detect changes in the demonstrators behavior in the course of fulfilling the task.

*Sub-action specific velocity* was computed as the average velocity for sub-actions  $a_1, a_2$ , and  $a_3$ , each without distinguishing pauses and motions:

$$\mathbf{velocity}(a_i) = \frac{1}{d_{i,2} - d_{i,1}} \sum_{j=d_{i,1}}^{d_{i,2}-1} v_j \quad (3.17)$$

is the velocity for sub-action  $a_i$  in seconds.

*Sub-action specific acceleration* was computed analogously as the average acceleration for sub-actions  $a_1, a_2$ , and  $a_3$ ,

$$\mathbf{acceleration}(a_i) = \frac{1}{d_{i,2} - d_{i,1}} \sum_{j=d_{i,1}}^{d_{i,2}-1} a_j \quad (3.18)$$

*Range* was defined as the covered motion path divided by the distance between sub-action, on- and offset.

$$\mathbf{range}(a_i) = \frac{\sum_{j=1}^{d_{i2}-d_{i1}} \|x_{j+1} - x_j\|}{\|x_{d_{i2}} - x_{d_{i1}}\|} \quad (3.19)$$



*Action length* denoted the overall action length in seconds and was measured from the beginning of sub-action  $a_1$  to the end of sub-action  $a_3$ .

$$\mathbf{length}(A) = (d_{3,2} - d_{1,1}) \cdot \tau. \quad (3.20)$$

Concerning the tutor’s eye gaze, three parameters were defined to measure the “contingency” of the interaction. Contingency is a concept related to the one of synchrony in interaction. It was defined as being present when a temporal, probabilistic relationship exists between two events in the interaction (Gergely and Watson, 1997; Harrist and Waugh, 2002; Watson, 1985). J.S. Watson defines contingency as the human infant’s means for detecting socially responsive agents and therefore postulates the existence of an innate contingency detection module as one of the most fundamental innate modules (Watson, 1985). Contingency is argued to be a characteristic aspect of social interaction (Csibra and Gergely, 2005) and to play an important role in infant development (Gergely and Watson, 1997). ”The discovery that another agent’s gaze is a cue worthy of monitoring relies on the infant’s ability to detect the contingency structure in interactions with that agent” (Fasel et al., 2002).

Brand, Shallcross, Sabatos, and Massie measured interactiveness investigating the tutor’s eye gaze behavior (Brand et al., 2007). The variables related to eye gaze they measured were the number of eye gaze bouts to the learner’s face per minute, the percentage of the demonstration spent gazing at the learner, and the average length of bout. Brand et al. found that infants received significantly more eye-gaze bouts per minute (Brand et al., 2007), so the frequency of eye-gaze bouts to the interaction partner was significantly higher in ACI than in AAI. The total and average length of eye-gaze bouts to the interaction partner in their study was significantly greater in ACI than in AAI. Brand et al.’s measures were adapted to form parameters suited to measure the contingency of the interaction, see (Vollmer et al., 2009a). *Frequency of eye-gaze bouts* to interaction partner (i.e., eye gaze bouts per minute) was computed analogously to the computation of the frequency of motion pauses, but from the Interact annotations. Also, the *average length of eye-gaze bouts* to interaction partner and the *total length of eye-gaze bouts* to interaction partner as the percentage of time of the action spent gazing at the interaction partner were computed. Equivalent measures were calculated for the eye gaze on the demonstrated object. Namely, values for frequency of eye-gaze bouts to object, average length of eye-gaze bouts to object, and total length of eye-gaze bouts to object as the percentage of time of the action spent gazing at the object, were obtained.

Additionally to the measures computed on the tutor’s behavior, for some analyses the

### 3. TUTORING BEHAVIOR

---

annotations of the learner's eye gaze were divided into the more abstract direction categories also used for visualization: anticipating, interaction partner, moving gaze, and elsewhere. Also directly from the annotation values made available as Matlab variables, the duration, the number of certain annotations and the relationship (e.g., the distance) between annotations were computed.

## 4

# Analyzing Tutoring Behavior

As described in the previous chapter, in Section 3.1, adults not only adjust their speech (Fernald and Mazzie, 1991), but also their gesture (Iverson et al., 1999) and motion (Brand et al., 2002; Gogate et al., 2000), when interacting with children. It has been shown that children not only prefer (Brand and Shallcross, 2008), but also can benefit from these modifications (Masataka, 2003).











This benefit has attracted attention of research in developmental robotics. The objective here is that, if the interaction between a robot and its user could be designed based on the natural adult-child tutoring interaction, the robot—similar to the child—could obtain the more structured and enriched input and benefit from it in its learning process (Nagai and Rohlfing, 2007; Rohlfing et al., 2006; Wrede et al., 2009). This is particularly interesting for learning actions, since—without support and only by observation—it is difficult for a robot to decide what and when to imitate (Carpenter et al., 2005; Csibra and Gergely, 2005), see Section 2.1. With these problems in mind, it has been suggested that using modifications in tutors’ behavior, a robot could learn to detect the meaningful structure of the demonstrated action (Nagai and Rohlfing, 2007; Rohlfing et al., 2006).

In the current chapter, in Section 4.2, light is shed on the question if this principle of benefitting from the tutoring behavior modifications (“motionese”, see Section 3.1.2) is transferable to human-robot interaction. Are robots tutored like children? Also some more insights are gained on which features characterize motionese behavior in general and motionese behavior toward learners of different age—the latter is addressed in Section 4.3. Beforehand, the corpus on which the analyses are carried out is described in Section 4.1. Analyses and results have been reported in (Vollmer et al., 2009a), (Lohan et al., 2009), and (Vollmer et al., 2009b) in collaboration with Katrin Lohan (Bielefeld University, Germany).

## 4. ANALYZING TUTORING BEHAVIOR

---

Please show **how**...

...to switch on the light with the lamp.		...to ring the bell by hitting the button.	
...to stack the blocks onto the poles.		...to use the saltshaker.	
...to nest the cups. Please start with the one closest to you.		...to open the bag.	
...to stamp and make three stamps onto the marks.		...to put the rings into the box.	
...to open and close the shelf.		...to put the books into the box.	

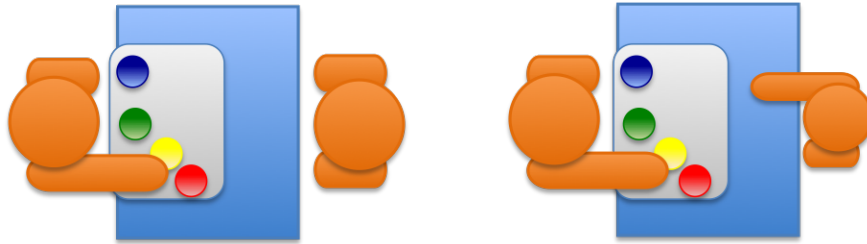
**Figure 4.1:** Objects and task instructions for the Motionese Corpus - Parents presented a set of ten manipulative tasks.

### 4.1 The Motionese Corpus

The Motionese Corpus is the main corpus on which the following analyses were carried out. The corpus comprises video recorded adult-child (ACI) and adult-adult interactions (AAI). Data was assembled by Katharina Rohlfing (Bielefeld University, Germany). In the semi-experimental setting, parents were asked to present a set of ten manipulative tasks both to their infant and to another adult. The objects, tasks and the corresponding instructions are presented in Figure 4.1. During the tasks, the tutor and the learner were facing each other, sitting across from each other at a table, see Figure 4.2. The situation was videotaped with three cameras: One recording the scene from above and the other two each focussing on one interaction partner. From these recordings only the video data of the camera filming the tutor and for the adult-child interactions additionally the camera filming the learner (i.e., the child) were digitalized. 67 families of which twelve were invited repeatedly in six months intervals participated in the study (see Table 4.1). Children were divided into three different age groups, defined through lexical development: preverbal children in group 1 (8–11 months), early lexical children in group 2 (12–24 months), and lexical children in group 3 (25–30

## 4.2 Motionese Compared to Modifications in Tutoring Robots

months). The age group 2 of early lexical children was again divided into two subgroups because around 18 months, there seems to be a drastic increase in the children’s vocabulary (“vocabulary spurt/burst” (Bates et al., 1988; Benedict, 1979; Goldfield and Reznick, 1990) from which children in age group 2b benefit, whereas in age group 2a, children usually only use one-word utterances: group 2a (12–17 months) and group 2b (18–24 months).



**Figure 4.2:** AAI and ACI setting - Setting in the adult-adult interaction condition (left) and the adult-child interaction condition (right).

	Group 1	Group 2 Group 2a	Group 2b	Group 3
Development	prelexical	early lexical	early lexical	lexical
Age in Months	8–11	12–17	18–24	25–30
Mean Age in Months	10.25	15.02	20.89	26.91
Standard Deviation of Age	1.13	1.94	1.72	2.08
Number of Infants	18	15	16	18
Gender of Infants	10m, 8f	8m, 7f	9m, 7f	7m, 11f

**Table 4.1:** The subjects of the three different age groups of the Motionese Corpus.

## 4.2 Motionese Compared to Modifications in Tutoring Robots

The crucial characteristics that establish a natural tutoring situation are yet unknown. In the field of developmental robotics, it is often assumed that in human-robot interaction, robots—because of their immature cognitive capabilities—can trigger a tutoring behavior in their interaction partner similar to the one observed in adult-child interaction (Nagai et al., 2008). However, this assumption has barely been studied. Recently, a study by Herberg and his colleagues (Herberg et al., 2008) investigated the question

#### 4. ANALYZING TUTORING BEHAVIOR

---

whether people modify their actions for computers. They presented a picture of an interaction partner to the subjects, which varied depending on the condition: a child, an adult and a computer together with a monitor and a mounted camera on it in a second condition (Herberg et al., 2008). The authors found that subjects modified their actions when speaking to a computer. These modifications differed from how they interacted with a picture of a child or an adult. Herberg and his colleagues (Herberg et al., 2008) interpret the difference in terms of assigning—to the persons, but not to the computer—the capability of reasoning about goals. However, it is difficult to expect from a user to assign some capabilities just from viewing a picture. It has been shown that subjects, when asked to speak to an imaginary infant, were not able to produce speech that exhibits all the features that are characteristic for motherese as it is produced in real adult-infant interactions (Knoll and Scharrer, 2007). The results from Herberg et al. should thus be interpreted with caution. Also, interactions with a computer are differently processed by subjects than interactions with robots, especially with respect to the assignment of intentions. In an fMRI study Krach et al. (Krach et al., 2008) have shown that the brain area that is generally associated with theory-of-mind (thus, the reasoning about the other’s intentions) is significantly stronger activated when the subjects thought they were interacting with a humanoid robot than when they thought they were interacting with a computer. Another relevant concept that has to be considered is contingency, see Section 3.2.4. Contingency describes situations in which two agents socially interact with each other. As mentioned previously, Csibra and Gergely showed that contingency is a characteristic aspect of social interaction (Csibra and Gergely, 2005). In the study published by Herberg et al. there is no possible reactiveness in the interaction partner, so in (Vollmer et al., 2009a), it was argued that social interaction cannot take place. In this section, therefore results from real interactions are presented with an embodied simulated robot based on the assumption that real interaction is needed in order to coordinate the behavior with the partner and to open up for mutual influence (Fogel and Garvey, 2007). Only such a scenario can create an environment in which it is possible to find out about the crucial characteristics of a natural tutoring situation.

In the study presented, similar to Herberg et al. (Herberg et al., 2008), the question of whether people will modify their actions when interacting with a machine was pursued. In contrast to Herberg et al., who used a computer, here the interaction with a virtual robot was investigated. For this purpose, real interactions—and not just a picture of the partner as in the previous study—with the artificial system were analyzed and the results compared to the results obtained from real interactions with a child and an adult. For the analysis, a battery of measurements was applied allowing for a fine-grained analysis of performed motions and their changes in the interaction as it

unfolds, cf. Section 3.2.4. The tutors' motionese behavior was measured using movement and eye gaze parameters and compared to the tutoring behavior in adult-child and adult-adult interaction. In the following the results of the analysis as described in (Vollmer et al., 2009a) are presented.

### 4.2.1 Data

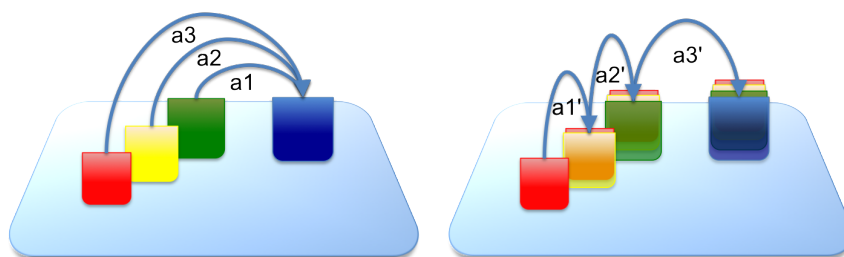
Data was obtained in two experiments. The data on adult-child interaction was obtained in the Motionese experiment, see Section 4.1, which is based on the same setting as in (Rohlfing et al., 2006) and (Nagai and Rohlfing, 2007). The data on human-robot interaction was obtained in a second experiment.

#### Motionese

The adult-child and adult-adult interaction data used for the analysis was taken from the Motionese Corpus. The cup nesting task was chosen to be analyzed because its repeating structure is assumed to provide most information about how the cognitive development of the learner is assisted by the tutor (Nagai, 2010). Only the age group of the preverbal children (8–11 months) was considered for the analysis because previous research already revealed differences in tutoring behavior between the adult-child interactions for this age group and adult-adult interactions (Rohlfing et al., 2006). From the 18 couples (36 subjects) and their children, a subgroup of eight parents (four fathers, four mothers) for the adult-child interaction condition and a subgroup of twelve parents (seven fathers, five mothers) for the adult-adult interaction condition were selected (see Table 4.2). The selection is based on comparability of behavior and sufficiency of video quality. The latter was essential for the annotations described in Section 3.2.2, which were carried out on the videos. The variability in the demonstration behavior arises from an alternate execution of the task. More specifically, the order in which the cups are nested can vary: The instruction contained the request to start the action with the cup closest to the participant's body, which means to sequentially pick up the green (a1), the yellow (a2), and the red cup (a3) and to place them subsequently into the blue one (Figure 4.3, left). However, some parents performed the action differently and placed the red cup into the yellow one (a1'), then the yellow cup containing the red one into the green cup (a2') and finally nested the green cup (containing the red and yellow one) into the blue cup (a3') (Figure 4.3, right). Those parents were selected, who executed the task in a comparable manner by executing the task in the first way (Figure 4.3, left).

#### 4. ANALYZING TUTORING BEHAVIOR

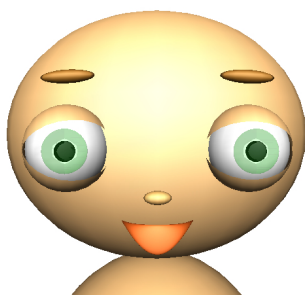
---



**Figure 4.3: Two ways of nesting cups** - First way of nesting the cups conforming to the instructions (left), second way of nesting the cups disregarding the instructions (right).

#### Robot-Directed Interaction Experiment

The adult-robot interaction (ARI) video data was acquired by Kerstin Fischer (University of Southern Denmark, Denmark) in a study with 31 adult subjects (17 male, 14 female) and a robot simulation, called Aka-chan, a Japanese name alluding to its baby-like face, on a computer screen, see Figure 4.4. The virtual robot was originally developed by Ogino et al. (Ogino et al., 2006). It was equipped with a saliency-based visual attention system originally proposed by Itti and Koch (Itti et al., 1998) and further adapted by Nagai and Rohlfsing (Nagai and Rohlfsing, 2007). Visual features, such as colors, intensity, orientations, flicker, and motions, are extracted to find locations in the scene, which stand out from their surroundings. These locations are called salient and the robot eyes will gaze to the most salient location the attention system finds.



**Figure 4.4: The robot simulation “Aka-chan”** - The virtual robot with its baby-like face developed by Ogino et al..

The study was designed for the data to be as similar as possible to the data of the Motionese corpus. For the setting, participants were seated at a table with the objects placed in front of them (see Figure 4.5). A computer monitor showing the Aka-chan robot was installed on the table at the side opposite to the subject’s seat. On the monitor, a camera was mounted for online calculation purposes of the saliency module.





**Figure 4.5:** ARI setting - Setting in the adult-robot interaction condition.

The objects the participants had to demonstrate were six of the same objects as in the Motionese corpus: *Lampe* (lamp), *Minihausen* (blocks on poles), *Becher* (cup nesting), *Klingel* (bell), *Salz* (saltshaker), and *Ringe* (rings), see Figure 4.1. For the analysis twelve participants (four male and eight female) who performed the nesting cups task in a comparable manner were selected, see Table 4.2.

	ACI Group 1	AAI	ARI
Child Development	prelexical		
Child Age in Months	8–11		
Number of Tutors	8	12	12
Gender of Tutors	4m, 4f	7m, 5f	4m, 8f

**Table 4.2:** The subjects considered in the analysis for adult-child (ACI), adult-adult (AAI), and adult-robot interaction (ARI).

### 4.2.2 Method

The goal of this analysis is to investigate tutoring behavior from two perspectives, motionese and contingency. For this reason, motionese and contingency features were analyzed. From the annotations described in Section 3.2.2, the data for the 2D hand trajectories were obtained, the action and sub-actions of the demonstration and the tutor’s eye gaze directions were coded.

### Hypothesis

The hypothesis stated that robots are tutored similar to children. Behavior modifications similar to motionese behavior should thus be measurable in the adult-robot interactions as well.

## 4. ANALYZING TUTORING BEHAVIOR

---

### Annotations

For all annotations, the video captured by a camera showing the front view on the demonstrator was used. It is best suited for action, movement, and gaze annotations, which are discussed in detail in Section 3.2.2 and again mentioned below.

*Action Segmentation:* For analyzing the data, the beginning and the end of the action of nesting the cups and additionally, the sub-actions (a1–a3) of grasping one cup until releasing it into the blue cup, which is the end position, (Figure 3.2) were marked in the video.

1. Action is defined as the whole process of transporting all objects to their goal positions.
2. Sub-action is defined as the process of transporting one object to its goal position.
3. Movement is defined as phases where the velocity of the hand is above a certain threshold. All other phases are defined as pauses (see Section 3.2.4).

*Hand Trajectories:* The videos of the two experiments were annotated via the semiautomatic hand tracker system mentioned in Section 3.2.2.

*Eye Gaze:* In annotating the eye gaze directions with the program Interact (Mangold, 2006), three categories of eye gaze directions are distinguished: looking at the interaction partner, looking at the object, and looking elsewhere (Figure 3.1).

### Measures

For quantifying *motionese* and *contingency*, 17 variables related to the two-dimensional hand trajectories derived from the videos and the eye gaze bout annotations produced with Interact were computed.

#### Motionese

Motionese was operationalized in terms of velocity, acceleration, pace, roundness, and motion pauses as defined in (Rohlfing et al., 2006). See Section 3.2.4 for details and formal descriptions. As already mentioned, Rohlfing and colleagues found that in adult-child interaction roundness is significantly lower than in adult-adult interaction and a trend for pace to be lower in adult-child interaction as well (Rohlfing et al., 2006). The authors found more pauses in adult-child interaction, but no significant difference in the tutor's hand movement velocity and acceleration. Here, the following measures were computed:

- velocity
- acceleration

## 4.2 Motionese Compared to Modifications in Tutoring Robots

---

- pace
- frequency of motion pauses
- average length of motion pauses
- total length of motion pauses

Additionally, the trajectory during the actual transportation of the cups, when performing the task, was investigated. The annotations of the sub-actions enable to define variables for each individual sub-action (a1, a2, a3) and also detect changes in the demonstrator's behavior in the course of fulfilling the task:

- sub-action specific velocity
- sub-action specific acceleration
- range
- action length

### Contingency

As described in Section 3.2.4 the contingency of the interactions was quantified in terms of variables related to eye gaze, as defined in (Brand et al., 2007) for measuring interactiveness.

- frequency of eye-gaze bouts to interaction partner / object
- average length of eye-gaze bout to interaction partner / object
- total length of eye-gaze bouts to interaction partner / object

Brand et al. found that infants received significantly more eye-gaze bouts per minute (Brand et al., 2007), so the frequency of eye-gaze bouts to the interaction partner was significantly higher in ACI than in AAI. The total and average length of eye-gaze bouts to the interaction partner in their study was significantly greater in ACI than in AAI.

### 4.2.3 Results

A multivariate ANOVA was run to test for differences of motionese and contingency in tutoring behavior in ACI, AAI, and ARI.

## 4. ANALYZING TUTORING BEHAVIOR

---

### Motionese

Differences were highly significant for all measures, see Table 4.3 for the results of the ANOVA as well as means and standard deviations. Tukey-HSD post-hoc comparisons of the three groups were carried out:

For *velocity*, the test revealed highly significant differences for ACI vs. AAI ( $p = 0.000$ ) and AAI vs. ARI ( $p = 0.000$ ), and a trend when testing ACI vs. ARI ( $p = 0.099$ ). These results show that in ARI hand movements seem to be slower than in ACI and hand movements in ACI are significantly slower than in AAI.

For the *sub-action specific velocity* measure, which only takes into account the hand movement during the transportation of the respective cup, the results were even more significant. For all pairs of conditions, significant differences for almost all three sub-actions were also found. These results clearly show that in AAI hand movements are very fast compared to ACI and ARI and additionally that hand movement is slowest in the ARI condition (ACI vs. AAI in a1:  $p = 0.000$ , in a2:  $p = 0.000$ , in a3:  $p = 0.002$ , ACI vs. ARI in a1:  $p = 0.156$ , in a2:  $p = 0.041$ , in a3:  $p = 0.029$ , AAI vs. ARI in a1:  $p = 0.000$ , in a2:  $p = 0.000$ , in a3:  $p = 0.000$ ). Also note that for all conditions the mean values increase for the consecutive sub-actions. This also holds for the variances (i.e., mean and variance for the velocity of hand movement in sub-action a3 are greatest). In the ARI, the rate in which the mean values increase is slowest.

The tests showed no significance for *acceleration* in ACI vs. AAI ( $p = 0.082$ ), but show a trend which is that acceleration of hand movement in ACI is smaller than in AAI. They show significant results for ACI vs. ARI ( $p = 0.047$ ) and AAI vs. ARI ( $p = 0.000$ ) conditions (i.e., in ARI, hand movement acceleration is significantly smaller than in AAI and ARI).

Viewing this measure again for only the transportation of the cups in the different sub-actions, the test results reveal significant differences and statistical trends for all pairs of conditions and almost all sub-actions. Results suggest that *sub-action specific acceleration* of hand movement is lower in ACI than in AAI. The mean values for each consecutive sub-action increase for both conditions, so that results for a2 revealed significance ( $p = 0.002$ ), whereas results for a1 ( $p = 0.058$ ) and a3 ( $p = 0.081$ ) show a trend. Also hand movement acceleration is highly significantly lower in ARI than in AAI ( $p = 0.000$  for a1, a2, and a3). For ACI vs. ARI results reveal significance only for a3 ( $p = 0.035$ ). Note again that for ARI mean values increase at a lower rate.

*Pace* results revealed significant differences for AAI vs. ARI ( $p = 0.003$ ) and a trend for ACI vs. AAI ( $p = 0.088$ ). The latter confirms the findings in (Rohlfing et al., 2006) that pace in AAI is higher than in ACI. The results indicate ARI having significantly slower pace than AAI and ACI having significantly slower pace than AAI. Note that the variance of pace in ARI is very small.

## 4.2 Motionese Compared to Modifications in Tutoring Robots

The results for the *roundness* measure show that movement is roundest in AAI compared to the other two conditions. Differences between ACI and AAI ( $p = 0.000$ ), and AAI and ARI ( $p = 0.000$ ) are highly significant which is again confirming previous findings (Rohlfing et al., 2006). No significance was found for ACI vs. ARI.

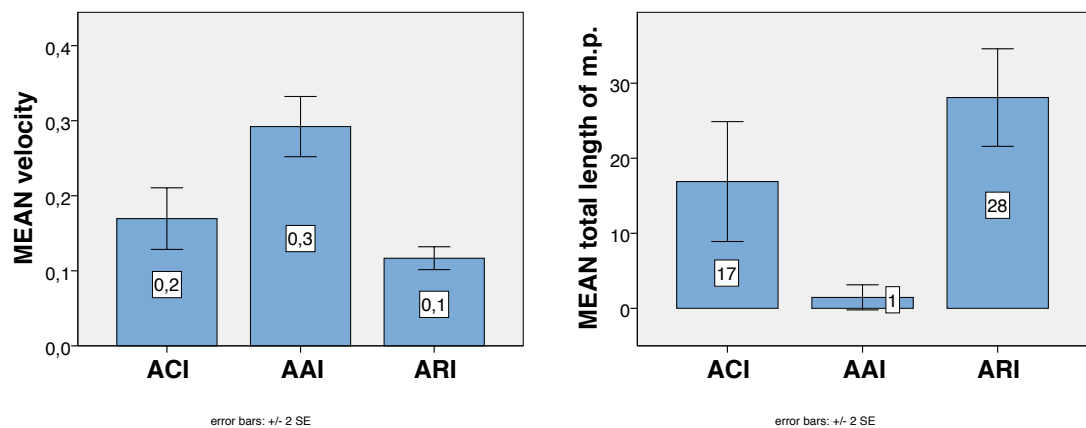
The *range* measure suggests that ARI exhibits the greatest range and for this reason most exaggerated movement for all sub-actions a1 to a3 and also that range is greater in ACI than in AAI. For ACI vs. AAI results revealed a trend for sub-action a2 ( $p = 0.086$ ). For ACI vs. ARI solely results for sub-action a1 showed significance ( $p = 0.012$ ); results for a2 and a3 did not. For AAI vs. ARI sub-actions a1 to a3 revealed significance (a1:  $p = 0.000$ , a2:  $p = 0.007$ , a3:  $p = 0.007$ ).

When analyzing motion pauses, tests revealed that in AAI the *total length of motion pauses* is significantly lower than in ACI ( $p = 0.002$ ) and ARI ( $p = 0.000$ ) and additionally that it is lower in ACI than in ARI ( $p = 0.029$ ).

In AAI the *frequency of motion pauses* is significantly lower than in ACI ( $p = 0.013$ ) and ARI ( $p = 0.001$ ). For ACI vs. ARI no significant differences were found.

The *average length of motion pauses* is significantly smaller in the AAI condition than in the ACI ( $p = 0.013$ ) and ARI ( $p = 0.000$ ) condition. For ACI vs. ARI test results also show significance ( $p = 0.019$ ). Values for ARI are greater than for ACI.

The overall *action length* is significantly greater in ARI than in ACI ( $p = 0.045$ ), where the action length is again significantly greater than in AAI ( $p = 0.009$ , AAI vs. ARI:  $p = 0.000$ ). Adults thus take more time, when demonstrating object functions to children compared to demonstrating them to adults, but they take even more time when demonstrating objects to a robot. Thus, in general, the movement in ARI appears to be even more accentuated than in ACI.



**Figure 4.6: Motionese results in bar charts** - Exemplarily, the values for velocity (left) and total length of motion pauses (right) are represented to illustrate the differences in motionese of the demonstrations for adult-child, adult-adult, and adult-robot interaction.

## 4. ANALYZING TUTORING BEHAVIOR

---

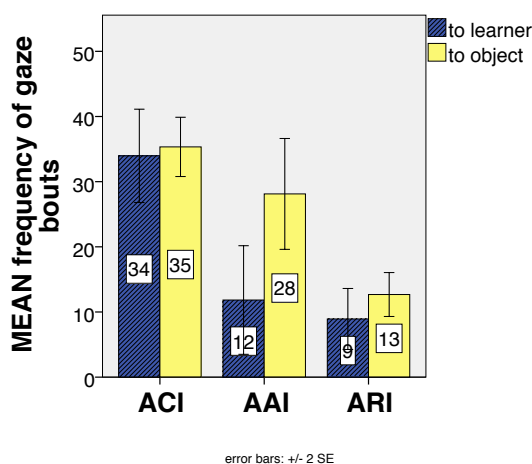
### Contingency

Most interestingly the results for eye gaze show a completely different picture. The contingency measures revealed for *total length of eye-gaze bouts to interaction partner* that in ACI significantly more time was spent gazing at the interaction partner than for AAI ( $p = 0.001$ ) and ARI ( $p = 0.002$ ). Differences between AAI and ARI are not significant.

For *frequency of eye-gaze bouts to interaction partner* the results showed significant differences for ACI vs. AAI ( $p = 0.001$ ) and ACI vs. ARI ( $p = 0.002$ ) again, but not for AAI vs. ARI. In ACI eye-gaze bouts to the interaction partner were most frequent. Testing the *average length of eye gaze bout to interaction partner*, on average significantly longer bouts in ACI than in AAI and ARI and a trend for AAI vs. ARI were found.

The same is true for eye-gaze to the object. For the measure *total length of eye-gaze bouts to object*. Values are significantly lower in ACI than in AAI ( $p = 0.001$ ) and ARI ( $p = 0.001$ ), where again differences between AAI and ARI did not exhibit significance. The results reveal that *frequency of eye-gaze bouts to object* is significantly lower in ARI than in ACI ( $p = 0.000$ ) and AAI ( $p = 0.003$ ). Differences in ACI and AAI were not significant.

*Average length of eye gaze bout to object* was significantly smaller for ACI than for ARI ( $p = 0.014$ ). Here, differences between ACI and AAI, and AAI and ARI were not significant.



**Figure 4.7: Contingency results in bar chart** - Exemplarily, the values for frequency of eye gaze bouts to learner and object are represented to illustrate the differences in contingency of the demonstrations for adult-child, adult-adult, and adult-robot interaction.

Variable	ACI		AAI		ARI		<i>F</i>	sig. <i>p</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
velocity	0.17	0.06	0.29	0.07	0.12	0.03	32.96	0.000
velocity a1	4.33	1.71	7.89	2.01	2.95	0.82	30.75	0.000
velocity a2	5.9	2.25	11.14	2.38	3.59	1.16	45.67	0.000
velocity a3	7.24	2.42	13.93	3.75	4.83	1.66	33.66	0.000
acceleration	0.06	0.03	0.08	0.03	0.03	0.01	14	0.000
acceleration a1	1.18	0.64	1.75	0.56	0.78	0.37	10.57	0.000
acceleration a2	1.58	1.06	2.93	0.84	0.84	0.34	22.8	0.000
acceleration a3	2.67	1.27	3.88	1.53	1.19	0.57	15.56	0.000
pace	17.68	32.78	45.02	35.59	4.25	1.98	6.92	0.003
roundness	2.87	2.49	7.26	2.71	1.74	0.30	23.07	0.000
total length m.p.	16.89	11.29	1.46	2.9	28.08	11.25	26.11	0.000
frequency m.p.	37.88	14.28	23.28	10.56	40.05	7.1	8.68	0.001
average length m.p.	5.92	3.68	0.58	1.14	11	5.39	22.02	0.000
range a1	2.54	1.07	1.76	0.42	4.09	1.52	13.69	0.000
range a2	1.69	0.41	1.33	0.18	1.81	0.44	5.76	0.008
range a3	1.45	0.25	1.24	0.18	1.64	0.4	5.55	0.009
action length	9.68	4.2	3.65	1.11	14.41	5.66	20.66	0.000
total length eye-gaze to l.	36.38	22.61	7.78	9.65	9.99	13.25	10.14	0.000
frequency eye-gaze to l.	33.96	10.13	11.84	14.43	8.93	8.11	13.16	0.000
average length eye-gaze to l.	0.94	0.39	0.22	0.29	0.45	0.41	9.47	0.001
total length eye-gaze to o.	59.48	23.17	90.74	11.31	88.87	14.13	10.91	0.000
frequency eye-gaze to o.	35.34	6.43	28.12	14.73	12.68	5.83	13.15	0.000
average length eye-gaze to o.	1.3	0.76	5.74	3.36	10.04	9.72	4.61	0.018

**Table 4.3:** Description of means and standard deviations for the groups. *F* and *p* values illustrate results of the ANOVA with hypothesis  $df = 2$  and error  $df = 29$  for all measures.

## 4. ANALYZING TUTORING BEHAVIOR

---

Similar results have been shown by Lohan, Vollmer, Fritsch, Rohlfing, and Wrede in (Lohan et al., 2009) for the Minihausen task (see Figure 4.1). The authors found significant differences for all three sub-actions for all pairs of conditions for the velocity measure, which is computed for each sub-action. The results clearly show that in AAI hand movements are faster than in ACI and ARI and additionally that hand movement is slowest in the ARI condition, supporting the previously described findings. Also note that for all conditions the mean values increase for the consecutive sub-actions: velocity in sub-action a1 < velocity in a2 < velocity in a3. In ARI, the rate in which the mean values increase is lowest and in AAI the rate is highest.

The range measure suggests that ARI exhibits the greatest range for each sub-action and therefore movement is most exaggerated. Also, range is greater in ACI than in AAI.

The results for eye gaze show here too a completely different picture. For total length of eye-gaze bouts to interaction partner they show that in ACI significantly more time was spent gazing at the interaction partner than in AAI and ARI. Differences between AAI and ARI are not significant.

For the measure total length of eye-gaze bouts to object, values are significantly lower in ACI than in AAI and ARI, where differences between AAI and ARI exhibit that values are significantly lower in ARI.

### 4.3 Motionese Toward Children of Different Age

This section presents work described in (Vollmer et al., 2009b). While it is already known that parents modify their demonstrations toward children (Brand et al., 2002, 2007), see Section 3.1, and that young infants aged six to eight months prefer ‘motionese’ (Brand et al., 2007), little is known about whether the modified behavior can also be found in interaction with older children. Here, therefore parental behavior toward children of three different age groups was investigated: parents of prelexical (8–11 months), early lexical (12–24 months) and advanced lexical (25–30 months) children.

#### 4.3.1 Data

The videos investigated in this analysis are part of the three different age groups of the Motionese corpus in the nesting cups task. Again they were selected based on task performance comparability. The subjects included are the following according to the learner’s age:



	Group 1	Group 2	Group 3
Development	prelexical	early lexical	lexical
Age in Months	8–11	12–24	25–30
Number of Parents	8	11	10
Gender of Parents	3m, 5f	6m, 5f	4m, 6f

**Table 4.4:** The subjects considered of the three different age groups.

### 4.3.2 Method

#### Hypothesis

This analysis has an exploratory character and therefore does not start out with a hypothesis, but with the questions, whether motionese can be found in the behavior of tutors interacting with children of different age and which motionese parameters change.

#### Annotation

For the analysis described in this section, the two-dimensional hand trajectory annotations and the annotated division of the action into three sub-actions were utilized (Figure 3.2).

#### Measures

The focus of this investigation lies on the following features computed on the annotated data (see Section 3.2.2) because they exhibited significant differences between the groups of participants shown in Section 4.2:

- Range
- Pace
- Total length of motion pauses
- Total length of teacher’s eye-gaze bouts to learner

### 4.3.3 Results

A repeated measures ANOVA with interaction condition (adult-child interaction (ACI), adult-adult interaction (AAI)) as intersubjective and infants’ age as intrasubjective factors revealed significant main effects for the interaction condition for all measures

## 4. ANALYZING TUTORING BEHAVIOR

---

( $p < 0.05$ ). Paired T-tests were computed for the three age groups separately to compute the measures of movement and eye gaze in ACI and AAI conditions. For the range measure, only in group 1 differences between the conditions were significant for sub-action 3 (ACI:  $M = 1.38$ ,  $SD = 0.16$ , AAI:  $M = 1.22$ ,  $SD = 0.14$ ,  $t(7) = 2.55$ ,  $p = 0.038$ ) and marginally significant for sub-action 2 (ACI:  $M = 1.7$ ,  $SD = 0.42$ , AAI:  $M = 1.35$ ,  $SD = 0.18$ ,  $t(7) = 2.15$ ,  $p = 0.069$ ). This suggests that the modified range of hand movements is present only in demonstrations toward pre-lexical infants. A reason for this could be that younger infants need gestures to attract their attention. The pace measure shows significance for groups 1 (ACI:  $M = 10.45$ ,  $SD = 10.54$ , AAI:  $M = 64.96$ ,  $SD = 30.27$ ,  $t(7) = -4.95$ ,  $p = 0.002$ ) and 3 (ACI:  $M = 13.9$ ,  $SD = 19.08$ , AAI:  $M = 49.61$ ,  $SD = 35.46$ ,  $t(9) = -2.82$ ,  $p = 0.02$ ), which suggests that pace in interactions with infants of all three age groups remains higher than in the AA condition. For motion pauses, significant differences for age groups 2 (ACI:  $M = 15$ ,  $SD = 14.8$ , AAI:  $M = 2.46$ ,  $SD = 4.28$ ,  $t(10) = 2.79$ ,  $p = 0.019$ ) and 3 (ACI:  $M = 12.16$ ,  $SD = 7.3$ , AAI:  $M = 2.96$ ,  $SD = 5.66$ ,  $t(9) = 4.55$ ,  $p = 0.001$ ) and a trend for group 1 (ACI:  $M = 19.75$ ,  $SD = 16.48$ , AAI:  $M = 1.8$ ,  $SD = 3.48$ ,  $t(7) = 3.2$ ,  $p = 0.015$ ) were found. Pauses structuring the shown action seem to be used over all age groups. For the eye gaze measure, a decrease in significance could be found over the children's age: In the AC condition, the learner was gazed at significantly longer in groups 1 (ACI:  $M = 35.34$ ,  $SD = 21.33$ , AAI:  $M = 5.74$ ,  $SD = 9.89$ ,  $t(7) = 3.96$ ,  $p = 0.005$ ), 2 (ACI:  $M = 30.69$ ,  $SD = 22.8$ , AAI:  $M = 6.55$ ,  $SD = 7.73$ ,  $t(10) = 3.61$ ,  $p = 0.005$ ) and 3 (ACI:  $M = 21.98$ ,  $SD = 11.72$ , AAI:  $M = 11.34$ ,  $SD = 13.28$ ,  $t(9) = 2.34$ ,  $p = 0.044$ ) and objects were gazed at significantly less in groups 1 (ACI:  $M = 64.54$ ,  $SD = 21.29$ , AAI:  $M = 94.26$ ,  $SD = 9.89$ ,  $t(7) = -3.98$ ,  $p = 0.005$ ) and 2 (ACI:  $M = 69.26$ ,  $SD = 22.76$ , AAI:  $M = 93.45$ ,  $SD = 7.73$ ,  $t(10) = -3.62$ ,  $p = 0.005$ ) suggesting that the young infants' attention is more often checked on.

### 4.4 Discussion

In sum, the results show a differentiated picture for modifications in human-robot interaction. On the one hand, the initial hypothesis is confirmed: A robot seems to receive even more strongly accentuated input than an infant: almost all hand movement-related variables, when pooled over the whole action sequence, showed a significant difference, or at least a trend, between the three conditions with a clear ordering (AAI < ACI < ARI). ARI movements can thus be characterized as slower (velocity, acceleration, and pace), more exaggerated (range), and less round (roundness) than AAI movements. In contrast to ACI, where the tutoring behavior seems to bear lots of variability, in the ARI, more stability could be observed. This suggests that ARI allows controlling the parameters of the learner and is thus a promising method for studying tutoring

behavior. On the other hand, contrary to the initial hypothesis, the contingency measurements show less contingent eye gazing behavior in ARI than in ACI (frequency and length of eye-gaze bouts to interaction partner).

These results raise an interesting question: Why is the behavior of the tutors in the ARI condition less contingent than in the ACI condition? As contingency is a bi-directional phenomenon, it is likely to be related to the robot's feedback behavior. Indeed, while the frequency of motion pauses is similar in ARI and ACI, the length of motion pauses is significantly longer in ARI than in AAI and ACI indicating that the tutor is waiting—possibly in vain—for a sign of understanding from the robot. The lower amount of eye-gaze bouts to the interaction partner in ARI as opposed to ACI could be interpreted similarly: as the tutor does not receive the expected feedback of understanding from the robot, he/she does not search for eye-contact with the robot. The question of why movement in ARI is less variable than in ACI can also be answered in the same line of argument: There is no learner behavior that could directly influence the tutor's demonstration. This suggests that the variability in natural adult-child tutoring interactions is caused by the learner's feedback, which shapes the tutor's action presentation online.

These results have important consequences for human-robot interaction in developmental robotics. They indicate that the behavior of the robot shapes the behavior of the tutor. Although all tutors showed strong modifications in their movement behavior toward a robot, thus stressing important aspects of the demonstrated action, they did not increase their contingency behavior, as other tutors would do in interactions with infants. Even though the purely reactive behavior of the robot in the study does induce parent-like teaching (as indicated in a qualitative study by Nagai et al. (Nagai et al., 2008)), it does not seem to be sufficient to produce a contingent interaction. As studies show, contingent behavior is an important feature for learning in human development (Gergely and Watson, 1997). Thus, in order for robots to be able to learn from a human tutor, they should have the capability to engage in a contingent interaction.

The findings of the second analysis of tutoring behavior toward children of different age suggest that actions chosen to attract attention (range) can primarily be found in interaction with younger infants, whose attention needs more guidance. Whereas interactions with older children seem to differ due to either the increase of children's attention abilities or that parents use other means to attract their attention (e.g., speech). In contrast, parameters that appear to be more in charge of structuring the action (motion pauses) seem to persist over the children's age and their verbal capabilities. Hence, the results support the hypothesis of learner feedback influencing the tutor's demonstration. The children's feedback according to their understanding of the

#### **4. ANALYZING TUTORING BEHAVIOR**

---

demonstrated action and depending on their age and capabilities seems to prompt the differences in tutoring behavior.

## 5

# Analyzing Learner Behavior

In the previous chapter, the tutor’s behavior has been analyzed. Results suggest that the feedback of the learner is important for creating a natural contingent tutoring interaction and that it could shape the tutor’s behavior—a resource which is highly valuable if we aim at enabling robot systems to learn within and from social interaction. But what kind of feedback should a robot produce in a tutoring situation and at which time? Robots provided with appropriate feedback strategies in tutoring interactions could elicit and benefit from behavior modifications like motionese behavior and learn in social interaction.

As drawn from these results, we focus on adult-child interaction to investigate what kind of feedback children contribute to a tutoring interaction with their parents because this could serve as inspiration for developing a robot’s feedback behavior. However, existing feedback models provided in social robotics and artificial agents mostly operate on the level of context-independent rules attempting for smooth turn-taking (Wrede et al., 2010), and do not address the issue of displaying “understanding” of an action as it is crucial in a tutoring/learning scenario.

The analysis presented in the current chapter concerns what kind of feedback the learner in a tutoring interaction gives and exactly how the learner’s understanding is signaled. The ways in which parents demonstrate actions to their infants are commonly related to children’s cognitive abilities (Vollmer et al., 2009b), see Section 4.3. Therefore, building on the previous analyses on the tutor’s behavior (Chapter 4), in this analysis, which was published in (Vollmer et al., 2010), the feedback provided by infants of different age groups—pre-lexical (8–11 months), early lexical (12–23 months), lexical (24–30 months)—to a parent’s action presentation is investigated.

The motivation here is that some insights can be gained into how people adapt their interaction to cognitive abilities of their partner and what feedback the partner makes use of. The expectation was that infants—due to their different levels of (cognitive, verbal, motoric) development—might produce different kinds of feedback which display

## 5. ANALYZING LEARNER BEHAVIOR

---

their current understanding of the demonstrated action. According to the Denver Developmental Scale (Frankenburg and Dodds, 1967), a screening test for cognitive and behavioral problems in preschool children, during normal development different behavior can be observed depending on the child’s age:

- 8 to 11 months: The child looks at a face, smiles back, smiles spontaneously and reaches for objects beyond its reach. It follows with the eyes 180 degrees, reacts to a bell, turns toward speech, begins to utter the words “mom” and “dad” undirectedly and can sit without help.
- 12 to 23 months: A child reveals wishes, begins to say “mom” and “dad” directedly, begins to combine words and pours raisins out of a jar as demonstrated.
- 24 to 30 months: A child uses syntactic constructions and says first name and last name, it easily accepts to be separated from its mother. Note that children begin to recognize colors only later, at the age of 30 to 36 months.

### 5.1 Feedback: Children’s Contribution to Tutoring Interactions

The *learner’s* contribution to the modified tutoring behavior and the learning process has received only little attention in research so far.

From an interactional perspective, the learner’s feedback is important as it provides information about the learner’s current understanding, which in turn enables the tutor to adjust his/her presentation accordingly (Estigarribia and Clark, 2007). It has been documented that, once the infant’s communication tends to break down, caregivers sensitively adjust subsequent messages (Zukow-Goldring, 1996).

Interactional research has revealed to which extent in authentic social interaction, the co-participants’ actions are closely related to each other and contingently respond to and build upon each other in a fine-grained interactional loop (Estigarribia and Clark, 2007; Sacks, 1992). In this line, the recipient’s verbal “back-channeling” behavior has become an important research topic and, as a multimodal account, it has been shown how some speaker’s talk step by step emerges with regard to the recipients’ changing foci of attention (Goodwin, 1979).

In the current chapter, patterns and features in demonstration and feedback will in a first step be ascertained by means of hypotheses acquired from qualitative investigation derived from Conversation Analysis, see Section 3.2.1. These patterns will then be found with quantitative measurements computationally from annotations and features in motion and on a verbal level.

### 5.1.1 Data

For the analysis presented here, data is again taken from the Motionese corpus, described in Section 4.1. The focus lies on parent-infant-interaction and on the task of nesting differently sized cups, see Figure 4.1. The two main ways of task performance (see Figure 4.3) were included in the analyses.

See Table 5.1 for an overview of the subjects that were considered.

	Group 1	Group 2	Group 3	
		Group 2a	Group 2b	
Development	prelexical	early lexical	early lexical	lexical
Age in months	8–11	12–17	18–24	25–30
Number of Parents	22	11	13	18
Gender of Parents	10m, 12f	6m, 4f	6m, 7f	9m, 9f

**Table 5.1:** The subjects of the three different age groups.

### 5.1.2 Method

As human interactional behavior in natural interaction is highly complex and variable, it has been drawn upon a combined qualitative and quantitative, manual and computational analysis in cooperation with Karola Pitsch (Bielefeld University, Germany) to investigate the infants’ feedback, see Section 3.2. To illustrate the procedure of the qualitative conversation analytic part of the analysis, for the first age group of prelexical infants a transcript is presented and described in length. This degree of detail will be excluded for the remaining age groups, but can be found in (Vollmer et al., 2009a). For the computation, the annotations of learner behavior are mainly used, but also the annotations of the tutor’s verbal utterances as well as the annotations of the division of the action into sub-actions are partly used, cf. Section 3.2.2.

### Hypothesis

For the current analysis the hypothesis was that children give different kinds of feedback depending on their age and abilities and showing their understanding of the demonstrated action.

### 5.1.3 Group 1: Prelexical Infants (8 to 11 months)

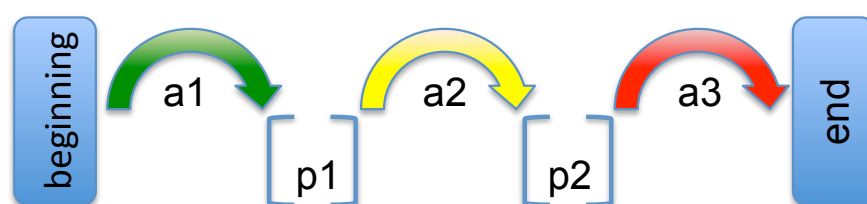
First, the basic structure of the parents’ action presentation is presented, which serves as a baseline with regard to which the infants’ feedback is located.

## 5. ANALYZING LEARNER BEHAVIOR

---

### Parents' Action Presentation

In their basic version, the parents' presentation of how to nest the differently sized cups consists of (i) marking the beginning, (ii) the three movements of transporting cups (a1, a2, a3) separated by short pauses (p1, p2) and (iii) marking the end of the action, see Figure 5.1. When carrying out these action demonstrations, parents use both verbal language and bodily actions, such as gesture, gaze, facial expressions, manipulation of objects etc.



**Figure 5.1: Course of the demonstration** - a1, a2, and a3 mark the transportations of the three cups. p1 and p2 mark the pauses in between cup transports. Figure adapted from (Vollmer et al., 2010).

### Qualitative Analysis

Parents presenting the action to their prelexical infants can be seen to mainly deal with the problem of helping the infant visually orient to relevant features of the scene: Infants appear to often look “somewhere” (i.e., unmotivated with regard to the task), and parents explicitly call for the infant’s attention either verbally (name + look here) or by extended hand/arm movements (Vollmer et al., 2009a). If these interactions are considered more closely with regard to the infant’s feedback, the investigation reveals that the infants respond to these cues offered by the parents by orienting their gaze to specific places at specific moments in time. The following interaction fragment, presented in (Vollmer et al., 2009a), exemplifies the conversation analytic transcription, which here constitutes the qualitative analysis. It shows such typical “attention grabbing”-patterns. The authors describe the transcript:

(i) Before the adult tutor (T) begins to demonstrate the action, the infant learner (L) gazes to the experimenter (Figure 5.2, img.1). When the tutor first starts to move his left hand to take the blue cup (Figure 5.2, img.2), the learner instantly orients his attention to the relevant hand carrying the cup (Figure 5.2, img.3).

(ii) After that, the tutor lets go of the blue cup again and picks up the green cup instead and utters “LOOK” (Figure 5.2, l.01, img.4) and with that another time re-orientes the learner’s attention to the relevant, green cup (Figure 5.2, img.5). The learner follows the trajectory of the cup with a short delay until the tutor lets it fall into the blue goal



## 5.1 Feedback: Children's Contribution to Tutoring Interactions

---

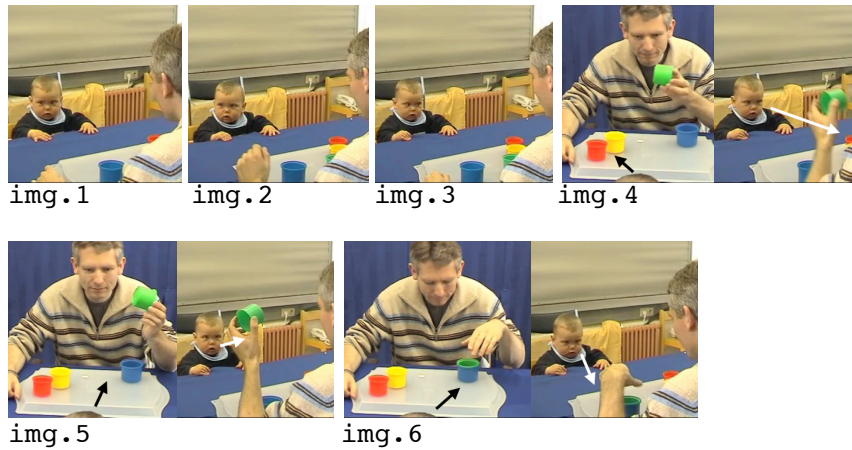
cup (Figure 5.2, img.6).

(iii) After the first cup transport, the tutor makes a pause of 1.0 second in motion and

```

01 T:                                     |↑GUCK mal;
T-act:  _____|LH↑↓|_____g grab|g lift|hold .
L-gaz:  @∅ |  _____|@cups |  _____|@g |>>>|@g .
          *1  *2      *3          *4          *5

02 T:  ERST nehmen wir den GRÜ:|Nen; (1.0) |
T-act:  . _____|g place
L-gaz:  . _____|
                                     *6
    
```



**Figure 5.2: Example fragment group 1 - 1.** Figure adapted from (Vollmer et al., 2009a)

speech (Figure 5.3, l.01). The infant's gaze shifts off to the side (Figure 5.3, img.7). The tutor begins the second sub-action by grasping and lifting the yellow cup.

## 5. ANALYZING LEARNER BEHAVIOR


---

He then shakes it and calls “HE:LLO <name> LOOK here”, which once more, prompts the learner to orient to the relevant cup (Figure 5.3, img.8).


Vollmer et al. identify the learner in this example to be a silent observer, who does not verbalize and whose other body movements—except the gaze behavior—seem to “freeze” during the action presentation.

03	T:	(0.5)	DA:NN,	(0.8)	↑HA:LLO	↑RAS	MUS;	
	T-act:	y grab	y lift	y shake	.			
	L-gaz:	>>>>>	@∅					>>>>
				*7				

04	T:	HIERher gucken;	(0.2)	DANN	den	GELBEN;	
	T-act:	.	y lift	y place			
	L-gaz:	@y	@b				
		*8					



img.7



img.8

**Figure 5.3: Example fragment group 1 - 2.** Figure adapted from (Vollmer et al., 2009a)

These observations suggest that, in age group 1, the infants’ feedback primarily consists of gazing behavior. As the analysis reveals, it matters to the tutor that the infant gazes at the appropriate place at a given moment. Its exact timing in relationship to the adult’s actions thus is important. The learner’s verbal utterances and other bodily behavior, however, seem to play only a marginal role.

### Quantitative Analysis

As a very first step, to underline the importance of gaze as feedback, the number of subjects, who give other active feedback in terms of verbalization, pointing or reaching gestures and smiling during the demonstrated action (for the definition of action, see Figure 3.2) were counted and the result shows that only three of 21 subjects verbalized, two pointed or reached for the object and three smiled in this age group.

Because gaze appears to be the most important type of feedback continuously given for this age group, this feature was explored in more detail, especially its precise timing with regard to the adult’s actions, in order to verify, whether the patterns revealed in the qualitative analysis could be found.

## 5.1 Feedback: Children’s Contribution to Tutoring Interactions

---

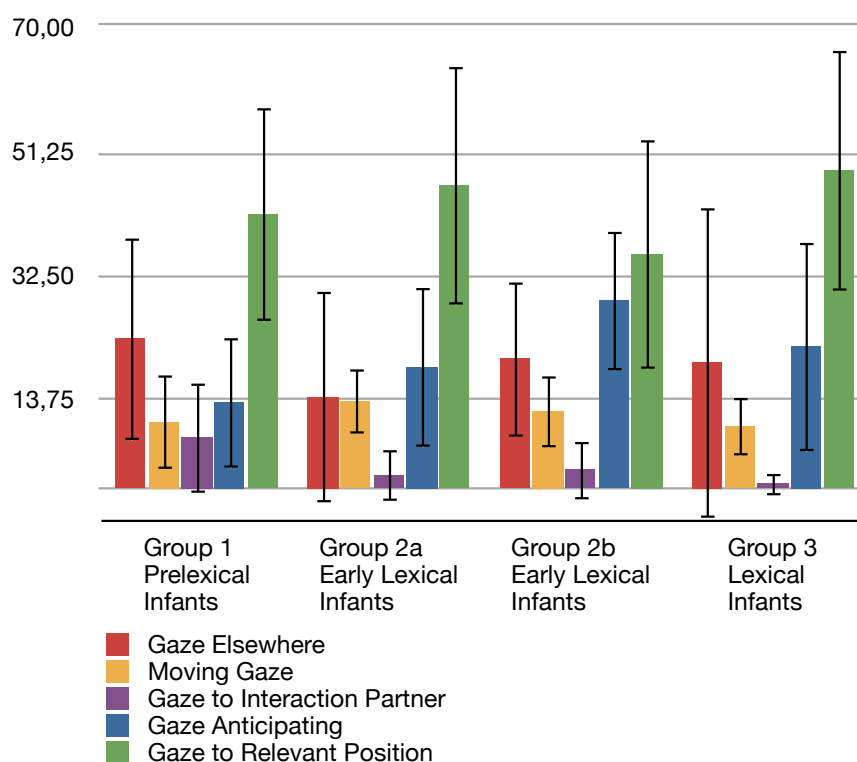
For the calculation, the infants’ gazing directions were classified into gaze to relevant position, anticipating gaze, gaze to interaction partner, moving gaze and gaze elsewhere. The following features were used in the investigation:

- *Eyegaze to Relevant Position*: Defined as the percentage of time of the demonstrated action looking to the right position, which means to the relevant object or hand. In the subactions a1, a2, and a3, this is always the cup which is being transported. During the time intervals in between subactions, when no cup is transported, but the hand reaches for the next cup, p1 and p2, the right position is considered the hand performing the next action.
- *Eyegaze Anticipating*: Percentage of time of the demonstrated action spent anticipating, that means looking at the goal position of the cup or hand. In subactions a1, a2, and a3, this is the cup, into which the cup, which is currently transported, will be stacked. In p1 and p2 this is one of the remaining cups, which could be transported next and to which the hand is being moved.
- *Eyegaze to Interaction Partner*: Percentage of time of the demonstrated action spent gazing at the interaction partner. At all time of the demonstration this is the case, when the child is looking at the face of the parent.
- *Moving Eyegaze*: Percentage of time of the demonstration, when eye gaze is shifting or in the process of moving.
- *Eyegaze Elsewhere*: Percentage of time of the demonstration spent gazing anywhere else than the directions above.

Figure 5.4 shows the results of the gaze features for all age groups. When assessing how much children in group 1 anticipate future actions with their gaze, the mean percentage of time a child in this age group anticipates a goal by shifting the eye gaze early in direction of the goal position was measured. The results reveal that the percentage of *Eyegaze Anticipating* a next action averages only 13.21% for group 1, whereas they *gaze Elsewhere* 22.83% of the demonstration. To measure the amount of attention grabbing patterns, first, the parents’ utterances annotated in the praat textgrids for the term “guck mal”, which is German for “look” were parsed and then, the focus lied on the gazing direction of the child right at the beginning of the utterance of the signal using the time stamp of the utterance obtained from the textgrid. The computation shows that 13 of 23 times the term was uttered, the child looked to a position which was not relevant at that particular moment. Additionally, the question was posed, where children who did not look to a relevant position before the attention grabbing pattern would look after the term was uttered by their parents. Out of the 13 children who did not gaze to the relevant direction, nine shifted their gaze either to the objects (5)

## 5. ANALYZING LEARNER BEHAVIOR

or the parent’s face (4) within two seconds after the attention grabber. The rest of the children all except for one, shifted gaze to a more relevant position, such as the hand of the parent or the plate supporting the objects. These findings suggest that “guck mal” is often used as an attention getter in this age group which seems to effectively orient the children’s attention toward a relevant position.



**Figure 5.4: Child gaze** - Graphic depicting the percentages of the demonstration the child gazes in the different directions with standard deviations. Figure adapted from (Vollmer et al., 2010)

### 5.1.4 Group 2: Early Lexical Infants (12 to 24 months)

#### Qualitative Analysis

For the interaction with early lexical infants, (a) some children continue to exhibit the “observer feedback” revealed for group 1, while (b) other infants begin to respond differently to the actions presented. In the example fragment of the qualitative analysis for this age group, Vollmer et al. attend to the new features and issues exhibited by group (b).

The qualitative conversation analytic transcripts are omitted and instead summarized

## 5.1 Feedback: Children’s Contribution to Tutoring Interactions

---

from this point on. Refer to (Vollmer et al., 2009a) for the full fragments including visualizations.

The qualitative analysis revealed: (i) When new objects are placed on the table, infants begin to tend to them by themselves and claim physical access. (ii) The infant’s initial proactive reaction toward the objects creates a different starting situation, in which the parent’s presentation takes place. As the infant is already oriented to the relevant object, the adult’s “attention grabbing” actions as observed in group 1 would not be required and the infant’s gaze direction does not change. However, surprisingly parents still use the same communicational devices—such as “LOOK” at the onset of their action presentation. Thus, “attention getters”—although produced by the parents in both groups—now begin to change their interactional functions: They assume the role of “structuring signals” which mark the beginning of the action demonstration.

(iii) While in group 1, the infant’s gaze continuously follows the parent’s action presentation, in group 2, results show infants anticipate next actions in the series of sub-actions. In the fragment presented in (Vollmer et al., 2009a) for this age group, the learner directs her gaze already to the third (red) cup while the tutor still finishes dropping the yellow cup into the blue one.

(iv) Not only does the infant’s anticipating gaze display (both to the tutor and the researcher) an understanding of the action and its serial character, but also do other forms of feedback provide further insights into the learner’s cognitive processing capabilities. The qualitative analysis further shows that the learner requests the cups at the onset of the demonstration verbally and by reaching for them, then she rests her—still extended arm—on the table. Her arm “freezes” in this posture during the entire action presentation and the learner again reaches toward the cup and verbally calls for the tutor’s attention the instant the last (red) cup falls into the blue one. Thus, the learner also displays an understanding of the expectable end of the demonstration and that the object can again be “requested” by her.

### Quantitative Analysis

Quantitative Analysis shows that in this second group, eleven of 23 children verbalize during the demonstrations, eight of them point or reach and five of them smile. This suggests a much more active feedback behavior in speech and movement.

Compared to group 1, the infants of subgroup 2b anticipate significantly longer (Mann-Whitney U test (data not normally distributed),  $U = 36$ ,  $Z = -3.368$ ,  $p = 0.001$ ,  $r = 0.59$ ).

The findings of the qualitative analysis suggest that infants in this age group should more often look at the right cup before the parent utters “guck mal” than gaze at irrelevant positions because this would confirm the change of use of the term toward

## 5. ANALYZING LEARNER BEHAVIOR

---

a structuring signal. Indeed, this is the case for seven out of nine times the term was uttered.

### 5.1.5 Group 3: Lexical Infants (25 to 30 months)

#### Qualitative Analysis

In group 3 of lexical infants, the qualitative analysis revealed that some parents begin to redefine the task of mere action presentation by more actively requesting the infant's feedback (e.g., through tag-questions, delaying actions or asking "do you know which color this is?"). In addition to the still remaining "observer feedback" from group 1, the infants' feedback thus becomes more elaborated. While appropriate gaze behavior remains an important feature, examples show that infants not only display their understanding of sub-actions, relevant next actions and the action as a whole, but also begin to translate this understanding into suggestions/instructions for the demonstrating adult located at precise moments in time. In an example fragment Vollmer et al. (Vollmer et al., 2010) find that the learner observes the first sub-action, subsequently points to the next cup, which should be transported, and then points to the respective goal cup and with that anticipates and even directs the tutor's next action. At the end of the complete nesting action (i.e., all cups are nested), the learner's pointing gesture changes to an open hand reach. When providing such "action guides" to the demonstrator, the learner's timing of her own actions in relation to the adult's presentation appears to be very systematic.

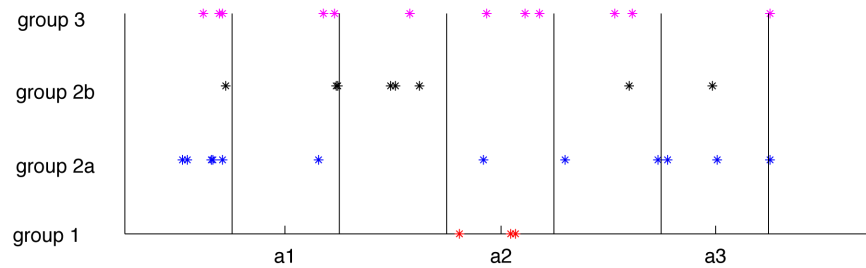
#### Quantitative Analysis

To substantiate the advanced systematicness of the infant's feedback in this group, a measure which shows that the infant's feedback follows the structure of the action has to be found. For each kind of feedback, considered for the other age groups, the time stamp of the beginning of the respective feedback intervals were taken and the distance to the nearest action boundary from the ELAN files as described in Section 3.2.2, see Figure 3.2, was computed. Unfortunately, no meaningful results could be found. Even when scaling the action parts to all have unity length and visualizing the beginnings of feedback intervals, no regularities can be seen, see Figure 5.5.

## 5.2 Discussion

As the hypothesis stated, the analysis has shown that infants indeed provide different kinds of feedback in the three age groups. Also, close inspection has revealed that the infants' feedback operates on two levels: as continuous involvement (e.g., through gaze)

and at specific places within the structure of the interaction (e.g., through pointing gestures at objects).



**Figure 5.5: Pointing** - Interval starts of pointing and reaching gestures for the different age groups. Figure adapted from (Vollmer et al., 2010)

Even though the age distinctions between group 1 through 3 are small, the results reveal noticeable differences in feedback: In group 1 feedback consists primarily of gazing behavior displaying the infant’s state of attention. In group 2, children begin to anticipate next actions with the direction of gaze and use more gestures and other modalities as feedback with which they provide the parent with information about the understanding of the presented action. This becomes even more evident in the feedback of the children in group 3, who give feedback much more systematically according to the structure of the action. Thus, feedback has to be considered in relation to the interaction partner’s current actions. In the presented analysis, a first attempt to investigate such links between the infants’ feedback and the parents’ presentations has been undertaken. The analysis has revealed two central interactional patterns which take this interrelationship into account: (1) Considering the precise timing of the infant’s gaze in relation to the adult’s hand movements, results showed that the infant’s gaze *follows* current actions or *anticipates* the next relevant action. The latter is mostly the case for the children of the early lexical and lexical groups 2 and 3. (2) Considering the precise timing of the infant’s gaze in relation to the adult’s verbal utterance “look”/“guck mal”, its function seems to change with the infant’s age: While it serves to grab the child’s attention in group 1, it becomes a structuring signal that marks important points of the demonstration to the children in group 2 and 3.

When trying to bring the structure of the action and the children’s feedback closer together taking objective action and sub-action boundaries, however, the attempts fail due to the variability of human interactional conduct. While the moments in time at which an infant provides feedback are highly systematic for the child in each single case, once these are tried to be detected over the corpus, problems arise. From this it can be concluded that more advanced methods and more precise patterns of features drawn from concrete hypotheses generated by qualitative analyses are required to link the infant’s feedback to the adult’s actions—given the complexity and variability of

## 5. ANALYZING LEARNER BEHAVIOR

---

human social conduct.

From this work, the following implications can be derived for the development of robotic systems that should learn from a tutor in social interaction: The feedback a robot should give should be twofold. It should provide a continuous part and a part transmitted at specific moments in time making use of multimodal conduct and thus, making it possible for the robot to influence the presenter's actions.



## 6

# The Interactional Account of Motionese

### 6.1 On the Loop of the Tutor’s Action Modifications and the Learner’s Gaze

In the previous chapters, in a first step, the tutors’ action presentations to different interaction partners (i.e., infants, adults, and robots) have been investigated and compared and the differences and modifications presented (Chapter 4). In a second step, the learner’s feedback has as well been investigated revealing that it signals the learner’s understanding of the presented action (Chapter 5). These analyses of tutor and learner have not yet considered how the participants’ actions interact and relate to each other, but reveal the necessity to do so. If interactional patterns can be identified, this could aid the interpretation of the tutor’s and learner’s actions in the tutoring interaction and shed light on the question of how to further transfer the results from adult-child interaction to human-robot interaction. This will be addressed in the current chapter, where an interactional point of view is taken to examine social learning and more specifically the reasons and consequences of the tutor’s “motionese” behavior. The aims here are to support previous postulations that the motionese behavior in tutoring is caused by the learner’s feedback behavior, to find out how the feedback is consequential for the action-presentation, and to find specific behavioral patterns as sources of the variability in adult-child interaction looking more closely inside the interaction between tutor and learner.

In research on adult-child interaction, only seldom an interactional perspective is assumed. Zukow-Goldring and colleagues for example show that tutors aim at guiding the learners’ attention to relevant aspects of the shown action (Zukow-Goldring, 1997; Zukow-Goldring and Arbib, 2007). Estigarribia and Clark revealed sequence structure

## 6. THE INTERACTIONAL ACCOUNT OF MOTIONESE

---

with a pattern of subsequent interactional moves of both tutor and learner (Estigarribia and Clark, 2007): At the beginning the tutor organizes the infant’s attention to the object, then when the infant’s gaze is oriented to the object, the tutor proceeds with the introduction of new information about the shown object and further tries to maintain the infants attention. Opposed to Estigarribia and Clark, who coded the learner’s gaze in three categories: the tutor, the object, and elsewhere, and considered a fixed set of gestures and verbal attention getters the tutor used, the current analysis presented in this chapter extends their approach by considering the full interaction on a micro-level, pointing out that gestures emerge in relation to precise moments of shifting gaze and, in this line, arguing that the task of “orienting the co-participant” is—at points—part of the action presentation itself.

For the analysis, which was carried out and published in collaboration with Karola Pitsch (Bielefeld University, Germany) (Pitsch et al., 2009, 2011), presented in this chapter, the principles of “co-construction” and “mutual monitoring” established in Conversation Analysis (Section 3.2.1), are used. These principles propose that actions within social interaction are conceived as being a co-construction (a joint accomplishment) of all participants and that participants of the interaction constantly monitor each other, interpret the others’ actions and display their online analysis with their actions, which successively shape the others’ actions while they are created, see (Goodwin, 1979; Mondada, 2006).

Thus, here, it is argued that the actions of tutor and learner are closely interwoven. The learner’s online feedback during the tutor’s action demonstration shape this presentation while it is created. In turn, the tutor’s presentation influences and prompts the learner’s actions.

More specifically, in this chapter, the hand trajectories of the tutors in the Motionese corpus described in Section 4.1 and their observed modifications leading to the high variability of tutors’ hand motion in adult-child interaction will be inspected with qualitative and quantitative means (Section 3.2) guided by the questions of how they are generated in the interaction and which functions they might have.

### 6.1.1 Data

For the analysis presented in this chapter, the focus lies on the cup-nesting task of group 1 of prelexical infants (8–11 months) of the Motionese corpus, see Section 4.1.

### 6.1.2 Method

To examine “motionese” behavior in adult-child interaction, highly variable human behavior has to be investigated on a set of data. Therefore, as described in Section 3.2, this analysis combines qualitative and quantitative as well as manual and automatic

## 6.1 On the Loop of the Tutor’s Action Modifications and the Learner’s Gaze

---

methods. First, a qualitative analysis is carried out manually on single videos. Second the obtained findings have to be described systematically and formalized as means to analyze the set of data computationally.

	Group 1
Development	prelexical
Age in months	8–11
Number of Parents	22
Gender of Parents	10m, 12f

**Table 6.1:** The subjects of age group 1.

### Hypothesis

In the analysis of this chapter, it was hypothesized that the motionese behavior in tutoring and its variability is caused by the learner’s feedback behavior and that the tutor’s demonstration and learner’s actions are closely interwoven. More specifically it was hypothesized that the tutor’s hand trajectories during the action demonstration and the learner’s gaze behavior mutually influence each other.

### Conversation Analysis

In a first step, a manual qualitative micro-analysis of interactions between an adult tutor and his/her infant, is carried out. See Section 6.1.4 for the results of the analysis and Section 3.2.1 for a description of the analytic procedure, for which Ethnomethodological Conversation Analysis (EM/CA) is used as analytical framework.

This analysis requires EM/CA to start with a specific research assignment: to investigate the reasons and effects of the variability in the tutor’s manipulative actions previously argued to be a manifestation of the interplay between the tutor’s action demonstration and the learner’s feedback, see Chapter 4. For that reason EM/CA does not consider the full amount of multimodality available in the interactions in the qualitative analyses, but focusses on the tutor’s hand movements and the way they are created during the interaction.

### Annotation

In a second step, systematic annotations are used for computational investigation of the data on corpus level, see Section 3.2.2. For the tutor, the annotations of the

## 6. THE INTERACTIONAL ACCOUNT OF MOTIONESE

---

hand trajectories during the action demonstration and the gaze directions were used. On the learner's side only the gaze directions were employed because previous results (presented in Chapter 5) revealed that in the age group of prelexical infants, gaze is the learner's main channel of feedback.

### Systematization and Quantification

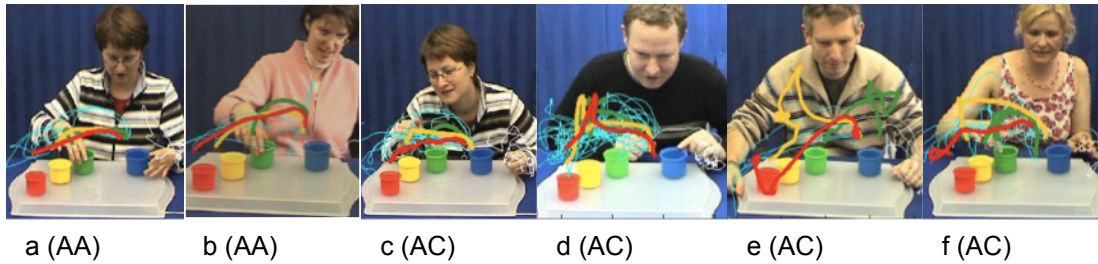
In a third step, the observations obtained in the qualitative analysis were adapted to find measures and approaches, which can assess the manual analyses with computational methods on the annotated data. For this, the timestamps and annotation values are parsed and loaded into MATLAB for further processing (i.e., for visualization, algorithmic systematization and quantification (see Section 6.1.5)).

#### 6.1.3 Starting Point: Variability of Hand Trajectories

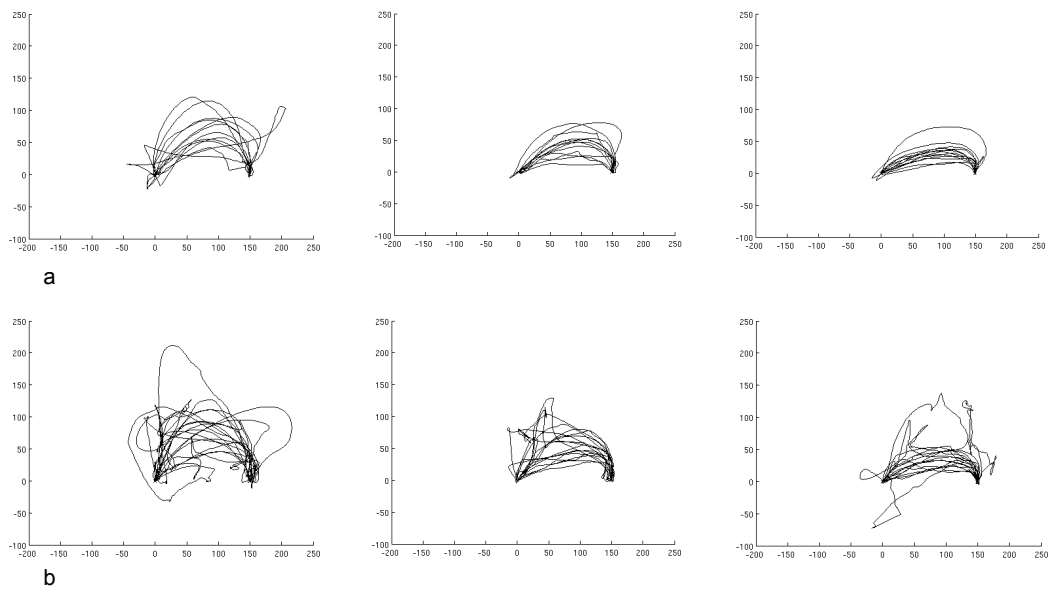
To better understand the variability in the tutors' hand motions when presenting a task to the infants, the tracked hand trajectories were visualized by plotting them over the corresponding video-frames, see Section 3.2.3. In the visualizations obtained (Figure 6.1) modifications in the hand trajectories can be observed, in which the considered trajectories differ from those suggested as typical for AAI and ACI: Rohlfing et al. depicted an ideal sub-action hand trajectory of an adult-child interaction, when defining their motion parameter roundness (Rohlfing et al., 2006). Parents are assumed to lift the cup straight up, pause in the air and then in a second motion transport the cup in a straight line to its goal position. This square motion is opposed to the observed smooth and round movement without pauses in adult-adult interaction. The presented visualizations (Figure 6.1) show that this ideal shape seems not to be observable in the adult-child interactions, but instead that there is a lot of individual variability involved in the movement execution. Results show: (i) cases, in which the tutor's sub-action hand trajectories are flat without particularly marked points (Figure 6.1 a, c); (ii) cases, in which the trajectories are more pronounced with a small peak toward the end of the sub-actions (Figure 6.1 b); (iii) cases, in which the presenter's hand performs a modification at the onset (Figure 6.1 d, e); (iv) combinations of these trajectory types: particularly cases, in which the first two nesting actions (green, yellow) show a high/pronounced shape, while the third action (red) is performed in a rather flat manner (Figure 6.1 e, f).

Considering these instances for presentations from parents toward their prelexical infants aged 8 to 11 months (group 1), the presenters' hand trajectories appear to have a relatively homogenous parabolic shape in the adult-adult interaction (Figure 6.2a). The trajectories in the adult-child interaction exhibit more variation (Figure 6.2b): higher arches and various modulations, particularly in the first sub-action.

## 6.1 On the Loop of the Tutor's Action Modifications and the Learner's Gaze



**Figure 6.1: Example trajectories** - Individual hand trajectories in Adult-Adult-Interaction (AA) and Adult-Child-Interaction (AC). Green/yellow/red trajectories mark the actions of nesting the cup of the corresponding color into the blue one; thin lines represent movements without cup.



**Figure 6.2: Normalized trajectories** - Normalized hand trajectories of groups of participants in Adult-Adult and Adult-Child Interaction. a: first (a1), second (a2), third (a3) sub-action in Adult-Adult Interaction for age group 1. b: first (a1), second (a2), third (a3) sub-action in Adult-Child Interaction for age group 1.

## 6. THE INTERACTIONAL ACCOUNT OF MOTIONESE

---

The aim of the following analysis is to find possible reasons for the observed modifications and the high between- and inner-subject variability of hand trajectories and to reveal the functions these modifications have in the tutoring interaction—for the participants, but also for the performance of the action.

### 6.1.4 Empirical Observations on the Interplay between the Tutor’s Hand Motions and the Learner’s Gaze

In a first analytic step, a manual qualitative analysis of the videotaped data is carried out using Conversation Analysis to understand the interactional organization and the problems that the participants are solving in their interaction. Here this analysis, which was published in (Pitsch et al., 2011), is summarized, refer to (Pitsch et al., 2011) for a detailed discussion. To be able to learn and reproduce an action, the learner has to have attended to the action demonstration. The infants do not always attend to the relevant aspects of an action by themselves. One main problem the tutor needs to solve is thus to orient the infant’s attention to the relevant aspects. In the qualitative analysis, Pitsch, Vollmer et al. introduced new revelations about the variability of action presentations in tutoring situations. They identified the sources and effects of the tutor’s “motionese” behavior in the interaction:

1. In tasks, in which the demonstration makes it necessary for the tutor to focus on an object, as for example in the cup nesting task, he/she is confronted with the issue of “dual orientation” between the object and the learner. A tutor mainly looking at the object (*task-oriented*) during the action demonstration, is not able to pay attention to what the learner does or observe the learner’s current state of attention and is thus not able to coordinate his/her actions with those of the learner. A tutor, who mainly monitors the learner (*recipient-oriented*), is able to adjust his/her activities according to the learner’s needs and thus is able to adjust his/her hand motions during the task to guide and to re-orient the learner’s attention. For this, one example fragment of an adult-child interaction was investigated. In this fragment, the mother gazes at her son only outside the transporting actions of the cups and performs the three cup transport trajectories with low arches. The child at most observes a small fragment of the action presentation, but by chance gazes to a relevant position, when his mother is checking on his attention giving her the impression that the child witnessed the whole presentation.
2. The shape of the action presentation of a recipient-oriented tutor has been shown to be involved in an interactional loop with the learner’s gaze: they shape each other. An example fragment of a father and his son was considered. The infant initially looks to the tray on which the cups are placed, but with a short delay

## 6.1 On the Loop of the Tutor's Action Modifications and the Learner's Gaze

---

after the father starts moving his hand to transport the first cup, the child shifts his gaze. The father monitoring his child stops his motion in the air and waits for his son to reach the cup with his gaze. Only then, the father starts to verbally comment on his actions.

3. Concerning the function of the tutor's movement modifications in the interactional loop, these seem to serve as orienting devices for guiding the infant's attention. Especially the upward movements of the tutor's hand carrying the cup (these can be seen as high arches in the printed trajectories) have been found to be used to attract the infant's attention to the movement. The analysis proceeds with the same example. During the next cup transport, at the beginning the child is gazing to the opposite direction. The father initiates a repair activity: He again stops the motion, then shakes the cup and also verbally calls for the child's attention.
4. Recipient-oriented tutors can infer the learners' understanding of the action by observing the learners' gaze. Additionally, for the case of the infant's task-anticipating gaze after the 2nd nesting action (i.e., the child in the investigated example fragment gazes at the last remaining cup initiating the tutor's next action, while the tutor's hand is still placed on the table where it was set after the second cup transport) three different interpretations of the infant's action have been reported:
  - Some tutors were reported to react by performing the 3rd nesting action with a flat hand trajectory, which does not contain any ostensive signals, and thus seem to treat the infant's gaze as revealing his/her knowledge about relevant next steps in the action.
  - Other tutors treat the infant's anticipating gaze as a lack of attention toward the ongoing action presentation and try to repair it with pronounced hand motions for example.
  - Yet other tutors do not show any reaction toward the infant's anticipating gaze behavior.

### 6.1.5 Systematization: From Empirical Observations to Formal Description

If we want to confirm these findings computationally on a broader base of data and finally, use them as inspiration for the design of human-robot-interaction, ways of transferring the results from the manual, qualitative analysis (Section 6.1.4) toward a more formalized description have to be developed. The first aim is to investigate whether

## 6. THE INTERACTIONAL ACCOUNT OF MOTIONESE

---

the qualitative observations derived from a few examples generalize or if these cases are only exceptions.

### Motionese Behavior: Task-oriented vs. Recipient-oriented

The qualitative analysis has revealed a difference in the ways in which tutors handle the “dual orientation” between the learner and the objects involved in the task (recipient-oriented vs. task-oriented). This has an impact on the tutor’s ability to micro-coordinate his/her actions with those of the learner and is reflected in more pronounced (recipient-oriented) vs. rather flat (task-oriented) hand motions. From this, it is hypothesized that—considering the corpus—tutors who orient to the recipient, would perform more motionese features than those who orient toward the task. To describe this phenomenon formally a sequence of two steps is undertaken:

1. Depending on the tutor’s gaze behavior (gaze at the infant for recipient-oriented, gaze at the objects for task-oriented) the data is separated into two classes. A sub-action (a1, a2, a3) is defined as belonging to the task-oriented category if the tutor gazes max. 25% of the time at the learner (the 25% threshold is inspired by the adult’s gazing patterns and is a simplification of the phenomenon described in Section 6.1.4). All other sub-actions, for which the tutor is gazing for more than 25% of the duration of the cup transportation at the learner, fall in the category of recipient-oriented sub-actions. This way, the sub-actions of the demonstrations by 18 parents (9m, 8f) were automatically divided into being task-oriented (20 sub-actions) and recipient-oriented (31 sub-actions).
2. For these two classes, the tutor’s motionese features were calculated using the values of the tutor’s hand trajectories and the annotation of the action structure intervals (a1, a2, a3) and applying the measures suggested in (Vollmer et al., 2009a) (action length, velocity, acceleration, range, total/average length of motion pause and pace).

Analysis reveals a significantly stronger motionese behavior in the recipient-oriented (r-o) compared to the task-oriented sub-actions (t-o) (independent sample t-test for all measures):

- The recipient-oriented sub-actions are longer (*action length*, **r-o**:  $M = 3.25$ ,  $SD = 2.06$ , **t-o**:  $M = 1.13$ ,  $SD = 0.35$ ,  $t(33) = -5.62$ ,  $p = 0.000$ ),
- performed at a lower speed (*velocity*, **r-o**:  $M = 0.09$ ,  $SD = 0.05$ , **t-o**:  $M = 0.15$ ,  $SD = 0.07$ ,  $t(49) = 3.63$ ,  $p = 0.001$ , *acceleration*, **r-o**:  $M = 1.08$ ,  $SD = 0.78$ , **t-o**:  $M = 2.14$ ,  $SD = 1.33$ ,  $t(27) = 3.2$ ,  $p = 0.003$ , and *pace*, **r-o**:  $M = 8.77$ ,  $SD = 11.18$ , **t-o**:  $M = 19.16$ ,  $SD = 12.91$ ,  $t(43) = 2.87$ ,  $p = 0.006$ ),



## 6.1 On the Loop of the Tutor’s Action Modifications and the Learner’s Gaze

---

- exhibit more *range* (**r-o**:  $M = 3.21$ ,  $SD = 1.72$ , **t-o**:  $M = 2.15$ ,  $SD = 0.91$ ,  $t(48) = -2.85$ ,  $p = 0.006$ )
- and longer *motion pauses* (*total* (**r-o**:  $M = 6.03$ ,  $SD = 10$ , **t-o**:  $M = 0.13$ ,  $SD = 0.57$ ,  $t(26) = -3.1$ ,  $p = 0.005$ ) and *average length of motion pauses* (**r-o**:  $M = 6.27$ ,  $SD = 8.83$ , **t-o**:  $M = 0.11$ ,  $SD = 0.47$ ,  $t(26) = 3.62$ ,  $p = 0.001$ )).

This shows that, also on the corpus level and with a formalized description, the tutor’s motionese conduct is linked to the concept of recipient design (Sacks et al., 1974) in the concrete interaction. Not only the mere physical presence of an infant (as opposed to an adult) plays a role for the tutor’s motionese behavior, but the tutor’s local monitoring and online analysis of the recipient’s actions and the recipient’s feedback are the basic condition for the observed conduct, as this differentiation within the ACI-condition shows.

### Tutor’s Hand Motions as Orienting Device

For the case of recipient-oriented tutoring, the qualitative analysis has revealed an interactional loop between the tutor’s hand motions and the learner’s gaze. More precisely, it has been shown that the tutor’s high, upward hand motions function as orienting devices for attracting and guiding the learner’s attention (Section 6.1.4). How can these orienting devices be described in a systematic and formalized way? Such description involves not only processing the simultaneous occurrence of events in one participant (as in the previous paragraph), but requires to render sequential structures of interactional coordination between two participants. To describe such interaction patterns in a formalized way, the following steps are undertaken to build a classifier operating on the annotated data:

1. For the cases of recipient-oriented tutoring (i.e. where the tutor is aware of the infant’s actions), identify the beginning of a sub-action.
2. Identify the infant’s orientation (i.e., gaze direction), which can be classified in two sub-groups:
  - (a) infant is attentive and gazes at the cups or the tutor vs.
  - (b) infant is gazing elsewhere.
3. Investigate the tutor’s hand motion during the nesting action: Is he/she performing a high hand trajectory, which—as hypothesized from the qualitative analysis—is supposed to attract the infant’s gaze? For analysis the issue arises how to best define a trajectory as being a “high” one? Here, the height of a

## 6. THE INTERACTIONAL ACCOUNT OF MOTIONESE

---

trajectory peak is considered relative to the height of the rest of the motion trajectory and a “high” trajectory is defined to lie above a threshold calculated by adding the standard deviation of the trajectory height of the three sub-actions (a1, a2, a3) to the mean trajectory height.

4. Analyze the infant’s reaction once the tutor’s hand motion has reached the defined threshold: Does the infant follow the tutor’s hand or not?

Results from classification: Two different types of orienting devices were identified. The first type engages the learner to follow the transported cup with his/her gaze, even though the learner might already be looking to a relevant position at the beginning of the sub-action (2a) and the second type at the same time constitutes a repair of the learner’s attention, which, at the beginning of the sub-action, is not directed toward a relevant position, but through the orienting device is redirected to a relevant position (2b). When the child’s gaze is still inattentive, once the threshold has been reached, the learner’s gaze has to start orienting toward a relevant position (the transported cup), while the tutor’s hand still moves above the threshold, for the orienting device to be considered a successful repair activity. The classifier found:

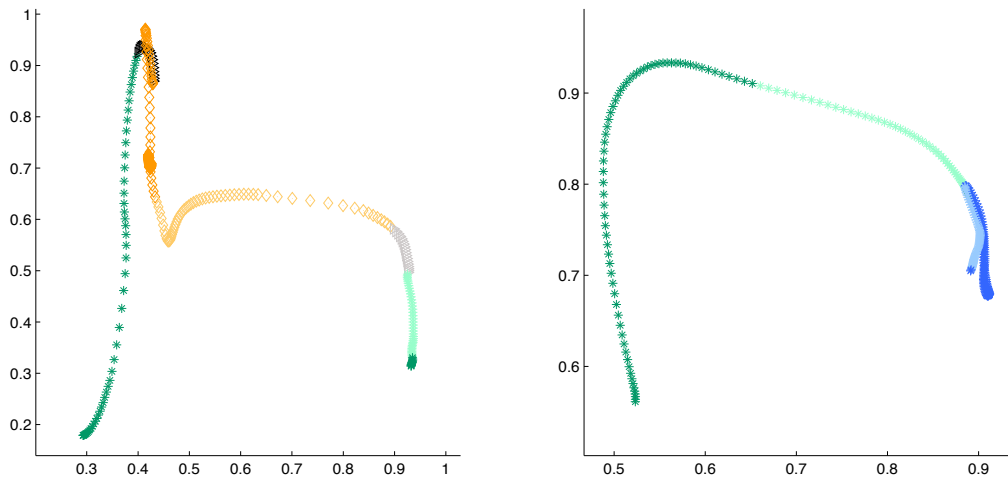
- Out of the analyzed 31 sub-actions in the recipient-oriented group 24 trajectories have been considered as being “high”.
- Ten sub-actions were identified as the first type of orienting device (2a).
- 14 sub-actions were identified as the second type of orienting device/ repair activity (2b), of which seven were successful.

To verify the choice of features and the definition of an orienting device/repair activity, these results were compared to independent qualitative-manual analysis of the same sub-actions by EM/CA methodology as presented in Section 6.1.4. Results were consistent. For 2b, the manual analysis reported only one additional sub-action with orienting device/repair activity, which the computational analysis did not find because the learner in this case did not change his/her gaze direction to the transported cup, but the tutor’s face, which the tutor considered to be relevant. Thus, a corpus query is derived that enables—for further analytical purposes—to build sub-corpora with very well defined interactional patterns.

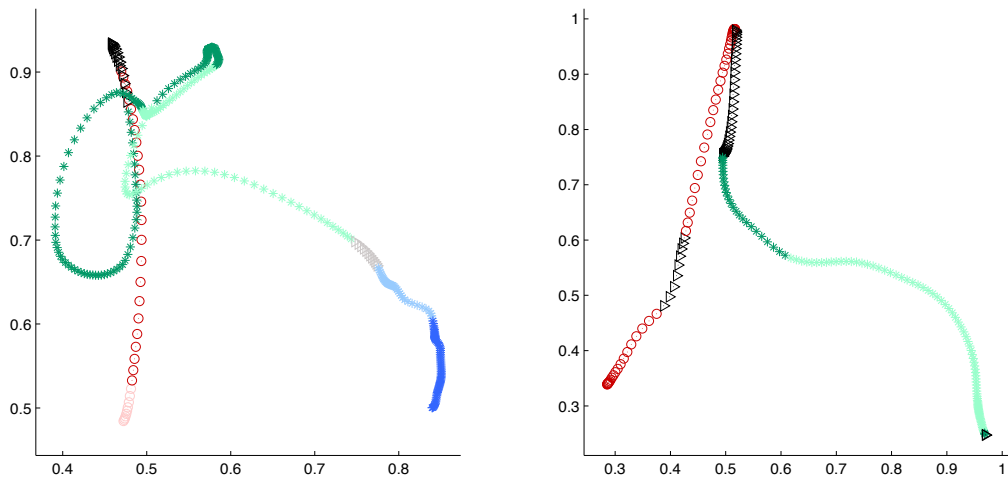
For 2b, the delay until the children reached a relevant point with their gaze was measured. The time it took for the orienting device to work and change the child’s gaze, ranged from 0.52 to 1.25 seconds ( $M = 0.83, SD = 0.28$ ). This has implications for the design of human-robot interaction, where a robot also should change its gaze and follow the cup transport in this time frame.

## 6.1 On the Loop of the Tutor's Action Modifications and the Learner's Gaze

---



**Figure 6.3: Example trajectories including orienting devices** - Tutor's high hand motions as orienting device for the infant. The trajectories are normalized using the maximum span of movement of the full demonstration in height and width (i.e., the axis-aligned minimum bounding box). (For color code, see Figure 3.5.)



**Figure 6.4: Example trajectories including orienting devices functioning as repair activity.** - Tutor's high hand motions as orienting device for the infant. The trajectories are normalized to the maximum height and width of the full demonstration. (For color code, see Figure 3.5.)

## 6. THE INTERACTIONAL ACCOUNT OF MOTIONESE

---

### **Anticipating Next Actions and its Impact on the Tutor's Action Presentation**

Qualitative analysis has shown that some infants anticipate the next relevant action during the tutor's demonstration through their gaze behavior and that tutors treat this anticipating gaze behavior differently: as displaying understanding of the action, to which they respond with a flat third nesting action; as displaying lack of attention to the ongoing action and in need of repair, to which they respond either online with an action modulation or subsequently with a higher next (third) nesting action; or they do not react to it. To describe this interaction pattern in a formalized way, the following steps need to be undertaken:

1. For the cases of recipient-oriented tutoring (i.e., where the tutor is aware of the infant's actions), identify the moments at which the infant's gaze anticipates the tutor's next action.
2. Investigate the tutor's reaction to the infant's anticipating gaze and classify it into the three interpretations
  - (a) understanding of action,
  - (b) repair, and
  - (c) indifferent,

each of which being related to a certain observable action of the tutor.

In doing this, two issues arise which are of general interest when transforming empirical observations toward formalization: First, considering the timely, emergent nature of natural interaction, to describe phenomena such as "repair" requires to reconstruct a previous action post hoc as being a repair. In the moment when that action is being produced, it is only a potential repairable. Second, the qualitative analysis has suggested the idea of "anticipating" a relevant next action. However, to find these instances on our corpus requires a new level of descriptive precision. In addition to the qualitative analysis, this formal approach points us to a set of two different types of "anticipation" in the data:

- The learner's gaze starts from the tutor's hand and suggests the next relevant action, as described in Section 6.1.4.
- The tutor's hand starts moving before the learner's gaze follows and eventually passes the tutor's hand directed to the target position.

## 6.1 On the Loop of the Tutor’s Action Modifications and the Learner’s Gaze

---

To define “anticipation”, the following steps have been carried out: An infant’s gaze interval was defined to be “anticipating gaze” if the child gazes to the next relevant object and has gazed at a relevant position before. The gazing directions were taken from the annotations. It was in a first step abstracted from the original notations to form groups of gaze toward positions related to the tutoring situation and task: For the first sub-action, when the green cup is transported, the relevant positions are the green cup and the tutor’s hand that transports it. At the beginning of the sub-action the child cannot anticipate because he/she has not seen the tutor perform the action, yet. However, when the child follows the transport of the cup and anticipates its goal position, this stretch of the infant’s gaze is considered to be anticipating. The following rules were applied:

relevant a1 = {green cup, parent’s relevant hand a1}  
anticipating a1 = {}  
= {blue cup} if child gaze was relevant before in a1

In the following pause (p1), the hand grasping the cup is the same hand, which transports the second cup. Only this hand is considered to be a relevant gazing target. Anticipating gaze can only take place, when the child is gazing toward the next relevant cup, which in this case, could be the yellow or red cup.

relevant p1 = {parent’s relevant hand a2}  
anticipating p1 = {yellow cup, red cup}

For the transport of the second cup, the relevant position is again the transported (here: yellow) cup and the hand transporting it. Anticipation can occur as in sub-action a1—the child gazes to the relevant position, before gazing to the target position—but also, when the child anticipated the next relevant cup in the pause beforehand (i.e., in p1).

relevant a2 = {yellow cup, parent’s relevant hand a2}  
anticipating a2 = {}  
= {blue cup} if child gaze was relevant before in a2  
= {blue cup} if child gaze was anticipating in p1

Analogously, the classes are defined in the second pause (p2) and the third cup transport (a3):

relevant p2 = {parent’s relevant hand a3}  
anticipating p2 = {red cup}

## 6. THE INTERACTIONAL ACCOUNT OF MOTIONESE

---

relevant a3 = {red cup, parent's relevant hand a3}  
anticipating a2 = {}  
= {blue cup} if child gaze was relevant before in a3  
= {blue cup} if child gaze was anticipating in p2

Quantification according to these rules suggests that

- infants anticipate the next action or pursuit of the hand trajectory in 13 out of the 31 sub-actions when the cup is being transported. In three of these sub-actions, the tutor does not look at the infant while he/she is anticipating and thus might not be aware of it.
- Ten cases were found in which the infants anticipate during the nesting pauses when the tutor's hand is being brought back to grab the next object.

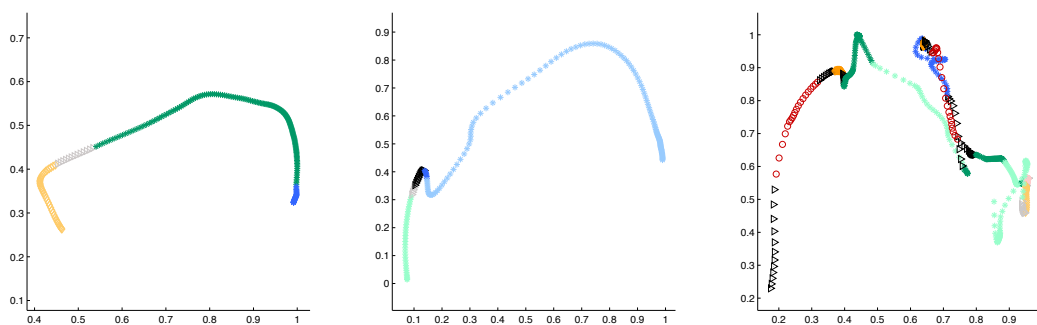
For investigating the tutor's reactions to the infant's anticipating gaze behavior, only those sub-actions in which the tutor actually sees the infant anticipating were considered, so that it was started from a sub-corpus of ten sub-actions and ten nesting pauses, in which an infant shows anticipating gaze behavior. Considering this phenomenon in closer detail, our automated analysis detects the followings groups:

1. Action-final anticipation (Figure 6.5-1): One type of the infant's "anticipation" occurs in the moment when the adult's hand hovers right above the big blue cup just before he/she drops the smaller cup into the big one. These forms of anticipation are very short, barely visible in the video-data, are to some extent an artifact of the annotation and finalize only the almost finished action, so that they have a very limited possibility of exhibiting the infant's understanding of some action. It is not clear whether tutors would realize these forms of "anticipation". In our corpus, three of the ten sub-actions and eight of the ten pauses belong to this type. All other anticipations start during the tutor's sub-actions (seven cases by seven children: four in a1, two in a2, one in a3) or pauses (two cases by two children: one in p1, one in p2).
2. Anticipation during sub-action (Figure 6.5-3): Six cases have been found, in which the infant anticipates the goal of the sub-action. In two of these cases, the tutor reacts with a flat trajectory in the next sub-action, thus displaying his/her interpretation of the infant's gaze as understanding of the action. In four of these cases, the tutor reacts with a higher trajectory in the next sub-action, thus showing his/her understanding of the infant's conduct as lack of attention.

## 6.1 On the Loop of the Tutor’s Action Modifications and the Learner’s Gaze

3. Anticipation during sub-action treated as being in need of immediate repair (Figure 6.5-2): In one case (a3) of our corpus cases, the tutor reacts upon the infant’s “anticipation” on-line with an elevated hand motion to re-orient the infant’s gaze, to which the infant indeed responds by following the tutor’s orienting device. This shows that the tutors do not treat the infants’ gaze as displaying their knowledge about an action, but rather as a lack of orientation.
4. Anticipation during nesting pauses (i.e., when the tutor’s empty hand travels back to grab the next cup): In two cases, the infant anticipates during the nesting pause, to which—in both cases—the tutor reacts with a flat next nesting sub-action and thus shows his/her interpretation of the infant’s gaze behavior as displaying their knowledge of the current action.

This suggests that tutors—across the corpus—are indeed sensitive to the infant’s gaze display and use it as indication either of their current state of understanding, but also as lack of attention. While the data sample is too small for statistical comparison, a small tendency can be observed: The infant’s anticipating gaze during the nesting pauses seems to be interpreted as the infant displaying his/her understanding of the current action. This has been identified through the tutor’s flat hand motion during the next action. However, The infant’s anticipating gaze during the sub-action is either considered as knowledge display or—more often—as lack of attention, which is repaired on-line (with an orienting device, see Section 6.1.4) or in the next sub-action with a high trajectory. These different options of subsequent reactions toward the infant’s conduct could be induced by the preverbal children’s cognitive abilities and the tutor’s expectations in the infant’s actions.



**Figure 6.5: Example trajectories: Anticipating gaze** - Patterns of infant’s anticipating gaze behavior. The trajectories are normalized to the maximum height and width of the full demonstration. (For blue parts and color code, see Figure 3.5.)

## 6. THE INTERACTIONAL ACCOUNT OF MOTIONESE

---

A first step toward investigating the tutors' reactions in interaction with children of different age has been undertaken in (Vollmer et al., 2010), see Chapter 5: Comparing the infant's anticipating behavior across the corpus, results revealed for the age group 8 to 11 months that infants spent on average 13.2% of the action presentation anticipating, and this increases in the older age groups: for the 12 to 17 months-olds, infants spent about 18.5% anticipating, for the 18 to 24 month-olds 28.6% and for the 25 to 30 month-olds 21.7%. Thus, effects of cognitive ability and interest/novelty seem to play a role.

These results are interesting to compare to Falck-Ytter and colleagues' report (Falck-Ytter et al., 2006). Falck-Ytter and colleagues suggest that infants were able to anticipate the goal of a presenter's reaching actions at the age of 14 months (Gredebäck et al., 2009), but they were not able to do so at ten months. On the contrary, our data suggests that infants already begin to anticipate next actions at the age of eight months (youngest age in our corpus, cf. (Vollmer et al., 2010)). These differences might be linked to methodological issues: Using eye-tracking methods, Falck-Ytter and colleagues define "anticipation" as a learner's gaze toward the target position, which arrives there at least 200 ms before the tutor's hand (Falck-Ytter et al., 2006). On the one hand, this level of precision cannot reliably be used with our annotation methods. When focusing on the instances of anticipation, for which the learner does not change the gaze direction again before the sub-action is completed, our annotations dividing the action into different sub-actions (a1, a2, a3) do not permit us to measure the offset of child's gaze arrival at target and tutor's hand arrival because the end of a sub-action is defined as the point in time, when the tutor releases the cup and not when the transported cup has reached the end position inside the goal cup and analogously in the pauses, when the tutor lifts the next cup, but not, when it is grasped. On the other hand, additionally, while the infants in Falck-Ytter et al.'s experiment were only confronted with a systematically moving hand (while all other parts of the tutor were hidden behind a shield), the infants in our study were immersed in a dynamic interactional process, in which the tutor was able to adjust their conduct to the participant's needs and in which the learner had access to the full range of the tutor's communicational resources (talk, gaze, head orientation etc.). This highlights the role of social cues in tutoring and points to a crucial difference of participants' skills exhibited in real world settings vs. under highly controlled lab conditions, which focus on one particular aspect of interactional conduct and neglect their interplay with other factors.

### **Summary: Systematization and Quantification across the Corpus**

The hypothesis drawn from the qualitative analysis that recipient-oriented tutors would produce more motionese features in their action presentations than task-oriented tutors



could be verified. Recipient-oriented tutors produce significantly longer action trajectories, with lower speed, more range, and longer motion pauses. Further investigation into the tutor's hand motions, in particular: the observation of high trajectories as orienting devices for organizing the infant's attention, has revealed the following: For recipient-oriented tutoring: From 31 sub-actions 24 cases with high trajectories (identified as prototypical for orienting attention) have been found. From these, one pattern has been found (ten cases), in which the infant is attentive at the beginning (i.e., orients to the cups or the tutor), then the tutor produces a high action trajectory and the infant's gaze follows. Another pattern has been found (seven cases), in which, at the beginning, the infant is gazing elsewhere, then the tutor produces a high trajectory, which attracts the infant's attention and re-orientes to follow the tutor's hand. However, in seven cases, this repair initiation does not work: the infant does not re-orient. Starting from the observation of the infant anticipating the next action in the nesting cups scenario, a systematic description of the concept "anticipation" has been developed and with this revealed a set of different types of anticipation in the data: Action-final anticipation, anticipation during sub-action, and anticipation during nesting pause. Investigating the tutor's reactions to this has revealed their close sensitivity to the infant's gaze behavior and their interpretation of the infant's anticipating gaze either as display of knowledge or as lack of attention. For this, there seems to be a tendency (on a data-set too small for statistical analysis) that anticipation during pauses is likely to be treated as display of action understanding, whereas anticipation during sub-actions provokes both forms of interpretation with the "lack of attention" occurring slightly more often.

## 6.2 Discussion

In this chapter, an interactional account of motionese has been suggested and its causes and effects in the concrete interaction between tutor and learner have been investigated. The analysis has supported the hypothesis by revealing that the tutor's presentation is interleaved with the learner's conduct on a micro-level: a direct relationship exists between the ways in which parents modify their actions directly with regard to the child's focus of attention. Action modification and the recipient's gaze can be seen to have a reciprocal sequential relationship and constitute a constant loop of mutual adjustments. In this loop, a set of interaction patterns have been revealed:

1. In tasks, in which the demonstration makes it necessary for the tutor to focus on an object, as for example in the cup nesting task, he/she is confronted with the issue of "dual orientation" between the object and the learner. Quantitative analysis has revealed that 20 of 51 sub-actions were classified as being task-oriented (i.e., the tutor mainly focussed on the object and did not monitor the learner), and the remaining 31 sub-actions as being recipient-oriented (i.e., the

## 6. THE INTERACTIONAL ACCOUNT OF MOTIONESE

---

tutor primarily payed attention to the learner). In comparison, the latter class showed stronger motionese behavior modifications than the task-oriented class. Thus, the computational investigation supports the qualitative analysis in that the learner's feedback shapes the tutor's action demonstration.

2. The form of the action presentation of a recipient-oriented tutor and the learner's gaze have been shown to shape each other.
3. The tutor's movement modifications in the interactional loop seem to serve as orienting devices for guiding the infant's attention. Especially the upward movements of the tutor's hand transporting the cup have been found to be used to attract the infant's attention to the movement. Quantification has revealed that high arches are a frequent means of orienting or reorienting the learner's attention, and that half of the tutors' attempts to reorient the learner's attention were successful.
4. Recipient-oriented tutors can infer the learners' understanding of the action by observing the learners' gaze (e.g., anticipating gaze). Possible reactions on the learner's anticipation from the tutor include a flat next nesting action without ostensive signals or attention getters conveying that the tutor treats the anticipation as display of the learner's correct understanding, reorienting the infant's attention and thus, treating the infant's gaze as incorrect, and not showing any reaction. A formal definition of anticipating gaze was developed, identifying action-final anticipation, anticipation during sub-actions, and anticipation during nesting pauses from the child's gaze directions. For this, there seems to be a tendency (on a data-set too small for statistical analysis) that anticipation during pauses has good chances to be treated as conveying action understanding, whereas anticipation during sub-actions is treated as conveying understanding, but as well as lack of attention, which is found slightly more often.

The presented analysis proposes an interactional perspective on social learning and also yields implications for research on tutoring in adult-child interaction as well as social robotics.

In eye-tracking studies, Gredebäck and colleagues investigated the understanding of different manual action conditions: reaching gestures, transporting actions, and stylized moving fists, in children of different age (Gredebäck et al., 2009). They found that 14 month-old infants anticipate goals of reaching movements, but only followed hand movements in transporting actions or moving fists with their gaze. Ten months-olds in their study only followed the movements in all conditions reactively suggesting that anticipating gaze emerges around the age of 14 months. Opposed to the findings presented by Gredebäck et al. the results of the analysis presented in this chapter have

shown that infants already anticipate next relevant steps of a demonstrated action with their gaze as early as eight months of age (youngest age in our corpus, cf. (Vollmer et al., 2010)). The main difference of the two studies is that in the adult-child interactions of the Motionese corpus analyzed in this chapter, the learners are part of an interaction with the tutor, who teaches with motionese behavior and reacts and modifies his/her presentation according to the learner's needs. The infants here also have access to information across all modalities and communicational channels. In the study conducted by Gredebäck et al., infants only were presented with videos of actors performing the movement of each condition. This suggests a higher level of skill and a better performance of participants in natural interactions compared to participants' performance in studies conducted with lab conditions that focus on only one aspect of communication. It additionally highlights the importance of interaction, multimodal communication and social cues for tutoring.

Concerning robotic learning, the presented findings imply that when the learner only observes the tutor's action demonstration, this might not be enough to understand the action. A robot system, which is supposed to learn new skills in social interaction with a human tutor, should rather be involved in a situated interaction with the tutor, influencing and shaping the tutor's ongoing presentation (shape of the trajectories, speed of the demonstration and so on) online with its feedback. Through its feedback (e.g., its gaze and other features), the robot could communicate information about its cognitive state, meaning for example, which parts of the presented actions are known or already understood and which parts are new or where there are uncertainties or incorrect assumptions.

A robotic system being aware of the interactional consequences of its own actions would have a powerful tool to actively influence and shape the action demonstration to its own advantage.

## 6. THE INTERACTIONAL ACCOUNT OF MOTIONESE

---

# 7

## A Human-Robot Interaction Study of Feedback in an Imitation Learning Scenario

In the previous analyses, see Chapters 4, 5, and 6, it has been shown that feedback is important to create a natural tutoring interaction. Types of feedback, a robot should give, namely continuous and at specific places in time, have been proposed and it has been argued that with its feedback, a robot could possibly shape the tutor's action presentation according to its benefits. These findings lead to the design of a human-robot interaction (HRI) study evaluating the impact of online feedback behavior during action demonstration and investigating the consequences of reproduction behavior (imitation/emulation) on the demonstration.

The study will be presented and discussed in detail in the current chapter. In a first section, the study and its underlying research questions are motivated (Section 7.1), then, design and realization are reported in a second section (Section 7.2), the methods and results of the data analysis are presented in a third section (Section 7.3) and finally discussed in Section 7.4.

### 7.1 Motivation to Investigate Online and Turn-based Feedback in a Demonstration-Action Loop

In the previous chapters (Chapters 4, 5, 6), it has been argued that a natural tutoring situation could only be created when the learner's feedback is involved. Information should flow bidirectionally and not only from tutor to learner, to enable online analysis and mutual monitoring. These dynamic processes shape the interaction and each

## 7. A HUMAN-ROBOT INTERACTION STUDY OF FEEDBACK IN AN IMITATION LEARNING SCENARIO

---

participant’s individual conduct.

De Jaegher, Di Paolo and Gallagher state that social interaction dynamics could constitute social cognition emphasizing the importance of interactive processes in the ability to understand others and act appropriately toward them (De Jaegher et al., 2010). Moreover, Wrede, Rohlfing, Hanheide and Sagerer argue that “learning necessarily needs to be embedded in an interactive situation”, (Wrede et al., 2009).

In most current robotic systems there is no interaction taking place and information only flows unidirectionally from tutor to learner: The tutor has knowledge about the action and demonstrates the action to the learner. The learner observes the action and then replicates it. With that the question of what to imitate is considered to be solved. Some recent approaches attempt to benefit from interaction with the human tutor (Breazeal et al., 2006; Nicolescu and Matarić, 2005), but interaction here is so far very basic, restricted, and follows predefined rules, as for example Nicolescu and Matarić in (Nicolescu and Matarić, 2005). The authors focus on imitation as trajectory following by a mobile robot, which has a small set of given behaviors like to go to, to track, to pick up and to drop colored boxes and follows the human tutor through the task. The tutor knows the robot’s behaviors and the sensors the robot uses. The interaction takes place, when the robot is executing the task. It—step by step—executes the sequence of behaviors previously acquired from the tutor’s demonstrations. The tutor can provide corrective hints for each step to the learner, which are uttered as pre-defined spoken commands, one set of commands to add a step, which the robot had missed and another set to delete a step from the sequence, when an irrelevant step was learned. During the demonstrations, the robot was set to “recording mode” and it repeated the learned sequence when it was set to “play”. The authors called this action-based interaction.

In another example, Lockerd et al. claim to teach their robot ‘Leo’ through a natural dialog in social interaction, but use a set of predefined social cues, which the tutor knows beforehand (Lockerd and Breazeal, 2004). Leo also gives feedback communicating difficulties and problems and eliciting help from the teacher. For example, when the robot perks its ears and leans forward after execution of the task, the tutor is expected to give binary feedback: either a “Not quite..” for unsuccessful execution, followed by a further task demonstration or a “Good” for successful task execution.

Current approaches do not allow for bidirectional dynamic interaction with continuous involvement of both tutor and learner. And this situation has hardly been studied, especially not with non-expert users. There are several open questions: What determines a tutor’s demonstration and more specifically, does the tutor’s behavior or behavior modification change depending on the robot’s feedback? How do unexperienced users react if a robot replicates their movements? How is the robot corrected, when it does not do it right? Are there robust social cues, which the robot could use?

In this study, which aims at investigating these questions, the robot’s feedback should

consist of an *online feedback* during the demonstration, which—as drawn from the previous findings in Chapter 5—should be continuous as the robot’s gaze on the one hand and on the other hand at specific moments in time like (sub-)action boundaries, and a *turn-based feedback* revealing more explicitly the learner’s understanding of the task by reproducing it.

The analyses presented so far have only looked at the demonstration of actions, but they did not consider a demonstration-action loop as in an imitation learning scenario, in which the learner has the chance to replicate the shown action. The replication of the action the tutor has presented—additional to the online feedback during the demonstration (e.g., continuous eye gazing)—directly following the demonstration would provide a concrete and very explicit feedback to the tutor of what the learner has understood of the action. The feedback would give the tutor the opportunity to repeat the demonstration in an adjusted manner (e.g., highlighting what has not yet been understood, emphasizing crucial aspects and removing potential ambiguities).

Section 2.3 brought to light that in robotic approaches to imitation learning, the term “imitation” is generally used for the replication of movements and the concept is not considered in detail and suffers from a lack of further differentiation to related concepts presented in Section 2.2. The two main ways of replicating movement identified in human children and also apes are imitation and emulation and have never been studied in the context of imitation learning in a human-robot interaction study before. How should a robot reproduce shown actions and what is the reaction to this type of feedback? To study how the participants reacted to the turn-based feedback, the robot was controlled to either reproduce the action by emulating or imitating it. The tasks were intended to either yield emulation as correct reproduction behavior and imitation as rather incorrect one or the other way around, see Figures 7.2 and 7.8. Since Imitation and emulation had not been studied before in an imitation learning scenario, the study has a partly exploratory character and two additional questions arose: Does the tutor notice the difference in the robot’s reproduction behavior (e.g., imitation or emulation)? Can measurable repair and revision behavior be observed, when the robot reproduces movements?

## 7.2 Design and Realization

This section concerns how the presented study was designed and gives information about how it was conducted, both in collaboration with Manuel Mühlig (Honda Research Institute Europe, Offenbach/Main, Germany). Setting and subjects are presented as well as experimental conditions, hypotheses and technical details.

## 7. A HUMAN-ROBOT INTERACTION STUDY OF FEEDBACK IN AN IMITATION LEARNING SCENARIO

---

### 7.2.1 Subjects

59 subjects (28 m, 31 f) participated in the study. One subject was excluded from all analyses because she neglected the task instructions. The subjects were right-handed to avoid side differences in action presentation, they were German native speakers to avoid language-based differences in action presentation, and they did not have any experience with robots (The majority of subjects had some experience working with computers,  $M = 3.42$ ,  $SD = 1.06$  on a scale of 1 [no experience] to 5 [very much experience], but subjects indicated that they had minimal to no experience interacting with robots.  $M = 1.24$ ,  $SD = 0.5$  on the same scale.). The study was gender-balanced and subjects were equally distributed across four age groups (20–30 years, 30–40 years, 40–50 years and above 50 years). Additionally, equal gender balanced numbers of subjects from each age group were randomly assigned to three robot online feedback behavior conditions. Please see Table 7.1 for a clear visualization.

Age groups	Group 1	Group 2	Group 3	Group 4
Age	20–30 years	30–40 years	40–50 years	50+ years
Min	19	30	40	51
Max	29	39	49	66
Mean	24.46	33.33	44.8	58.25
Gender	7 m, 6 f	7 m, 8 f	7 m, 8 f	7 m, 9 f
Robot online feedback				
Social gaze	2 m, 2 f	2 m, 3 f	2 m, 2 f	3 m, 3 f
Random gaze	2 m, 2 f	2 m, 3 f	3 m, 3 f	2 m, 3 f
Static gaze	3 m, 2 f	3 m, 2 f	2 m, 3 f	2 m, 3 f

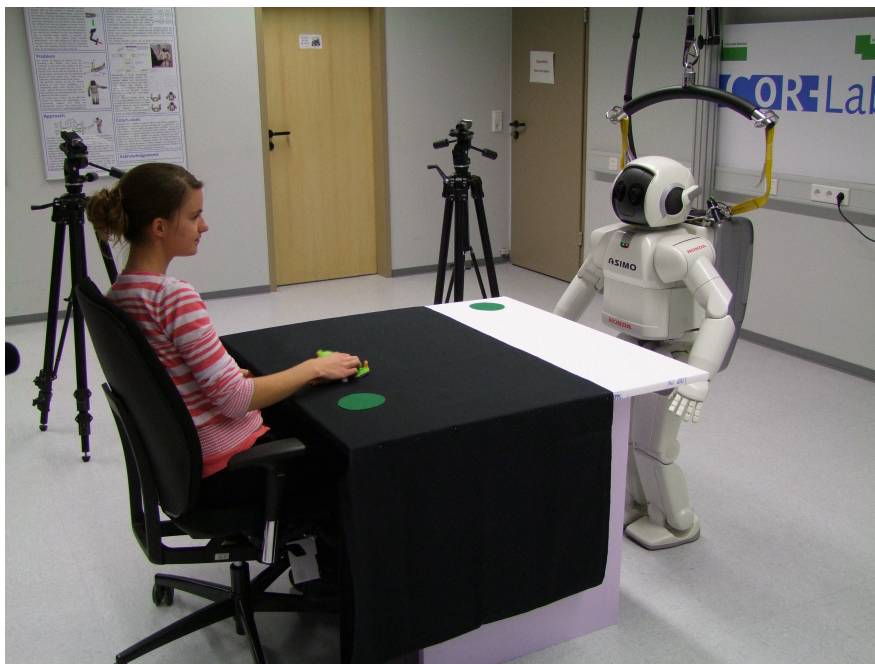
**Table 7.1:** The subjects of the different age groups and assigned robot online feedback condition.

### 7.2.2 Setting and Experimental Conditions

The robot (a full-size humanoid robot) stood started up at a fixed position at a table when the subject entered the laboratory. The interaction took place at this table with the subject seated opposite to the robot (see Figure 7.1). The experimenter gave an introduction to the general course of action and explained the task guiding the participant through the interaction step by step with an example using a rubber duck. The participants had to present eight different object manipulation actions to the robot. These actions fell into two categories: *manner-crucial*, and *goal-crucial*



actions. In manner-crucial actions, the manner and path is most important about the action. As for example for the task to show how to clean a window with a sponge, the movements are important and not where the sponge is set down. For goal-crucial actions in contrast, the goal position of the object is important and not so much how it got there. For example when a phone is hung up, it is important that the handset is properly put on the hook, but it does not matter if it reaches this position in a curved or straight movement. An overview of the objects and task instructions is given in Figure 7.2. After one demonstration of an action, the robot gave *turn-based feedback* by reproducing the action. The action could then be demonstrated again and the robot reproduced the action again forming a loop, which repeated until the participant decided against it. One interaction with one of the eight objects thus was composed of several steps depicted in Figure 7.3.











**Figure 7.1: HRI setting** - Setting of the human-robot interaction

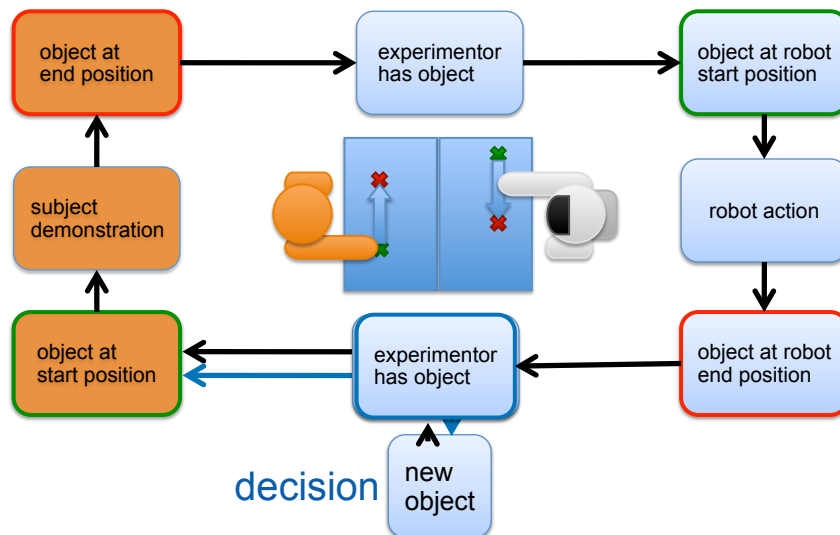
For each participant four actions (of which two were manner-crucial and two were goal-crucial) were randomly chosen to be imitated by the robot (cf. Section 2.2) (i.e., it reproduced the trajectory of the object as exact as possible). (For example sequences see Figures 7.4 and 7.5.) The other four actions (of which analogously two were manner-crucial and two were goal-crucial) were emulated (again cf. Section 2.2) (i.e., the robot reproduced the end state only with a straight, goal-directed movement). (For example sequences see Figures 7.6 and 7.7.) Together with the action conditions, the replication behaviors form a two-by-two design (see Figure 7.8).

## 7. A HUMAN-ROBOT INTERACTION STUDY OF FEEDBACK IN AN IMITATION LEARNING SCENARIO

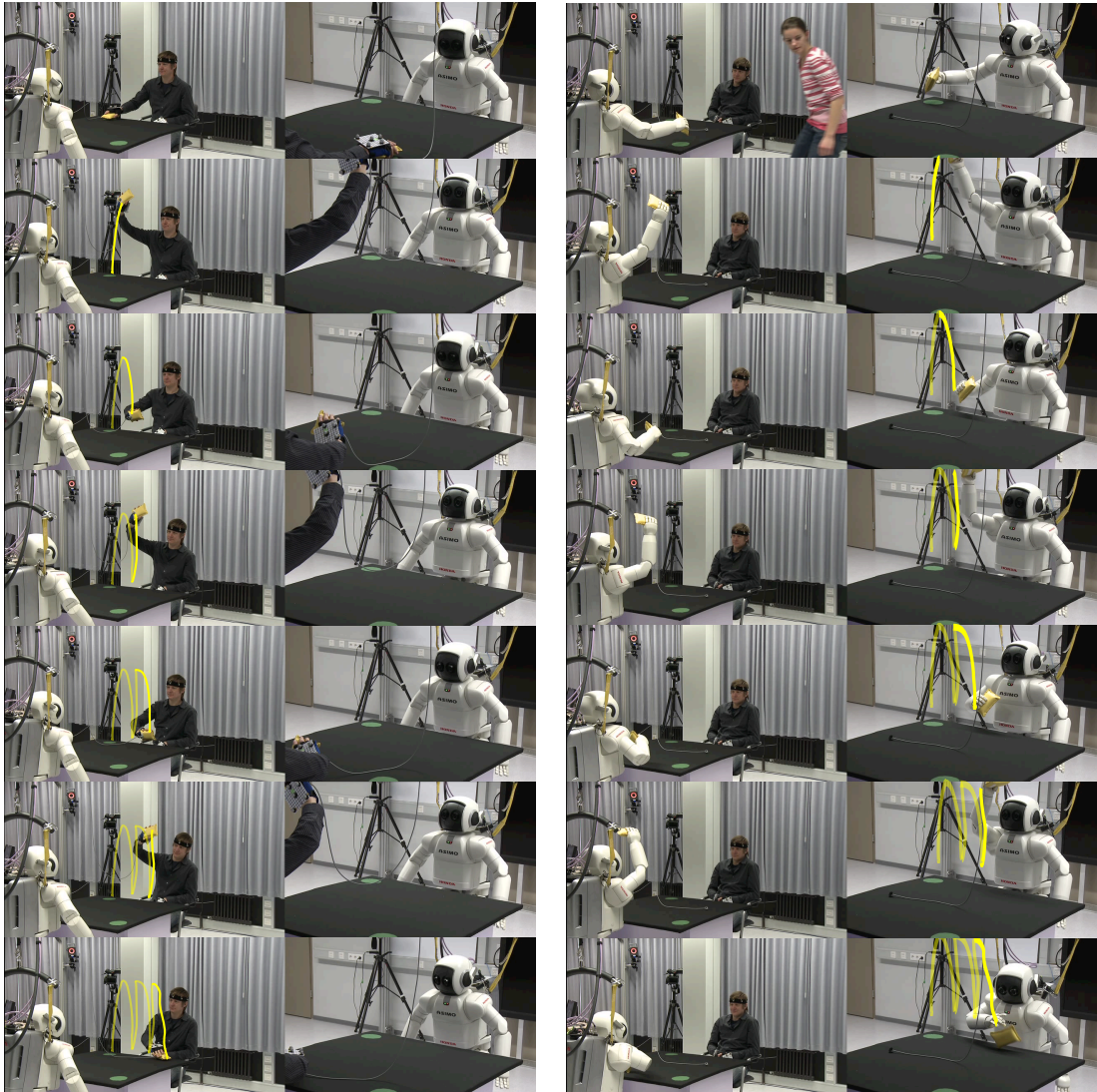
Please show the robot **how...**

Goal-crucial actions	Manner-crucial actions
...the airplane flies to the airport. 	...the airplane does a loop. 
...to hang up the phone. 	...to clean a window with a sponge. 
...the dog walks to the bowl. 	...the frog jumps. 
...the elevator moves down. 	...a feather falls. 

**Figure 7.2: Objects and task instructions** - Objects and tasks were divided into goal-crucial actions and manner-crucial actions.



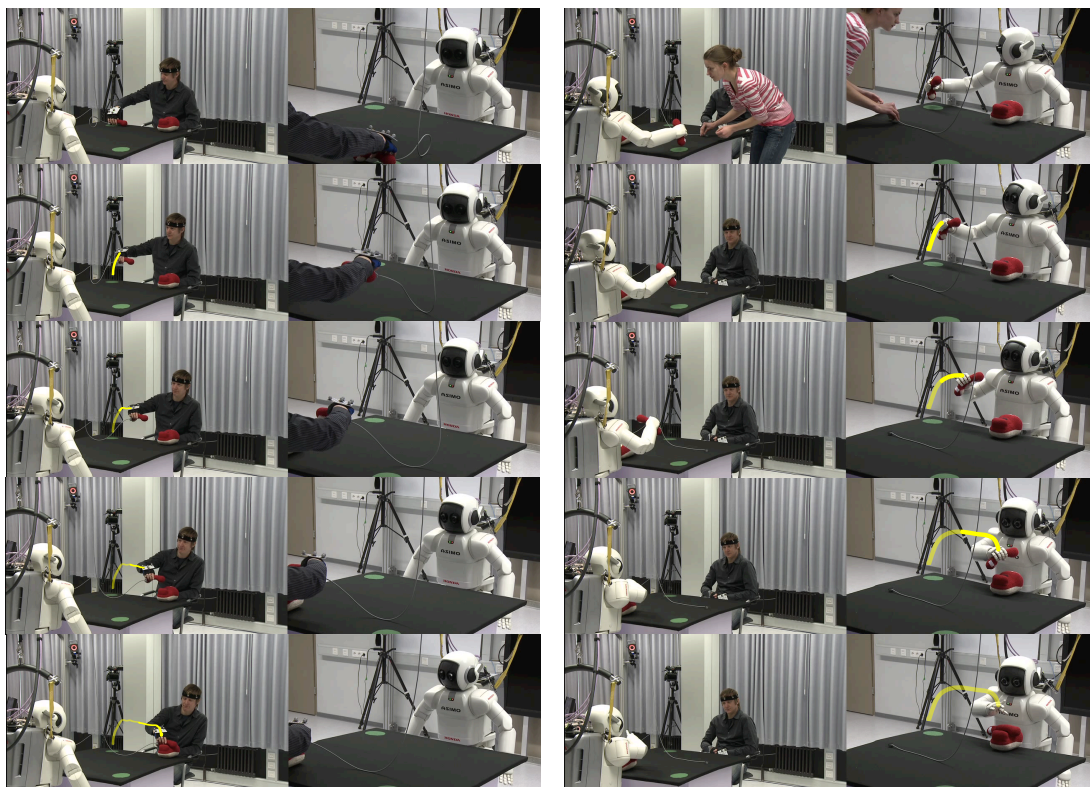
**Figure 7.3: Interaction loop** - The course of the interactional loop for one task beginning with the introduction of the new object and following the black arrows.



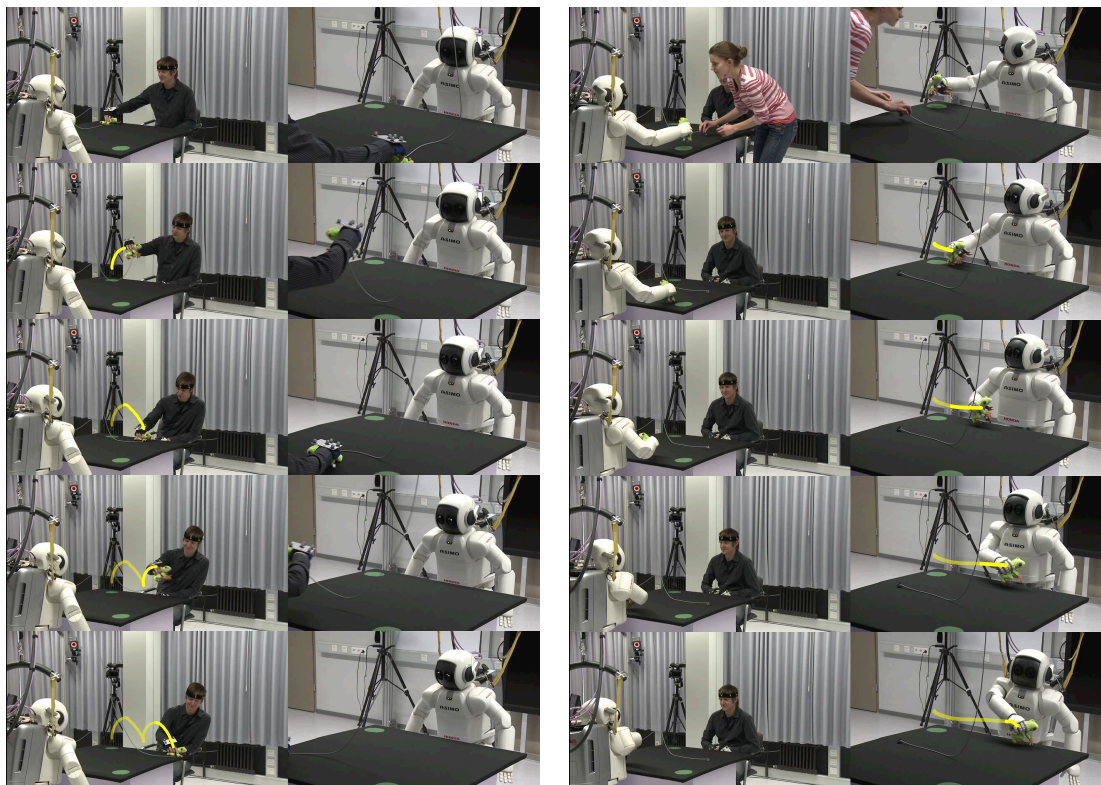
**Figure 7.4:** Example sequence of imitation of a manner-crucial action - The robot imitates (right) the subject's demonstration of the manner-crucial action of cleaning a window with the sponge (left).

## 7. A HUMAN-ROBOT INTERACTION STUDY OF FEEDBACK IN AN IMITATION LEARNING SCENARIO

---



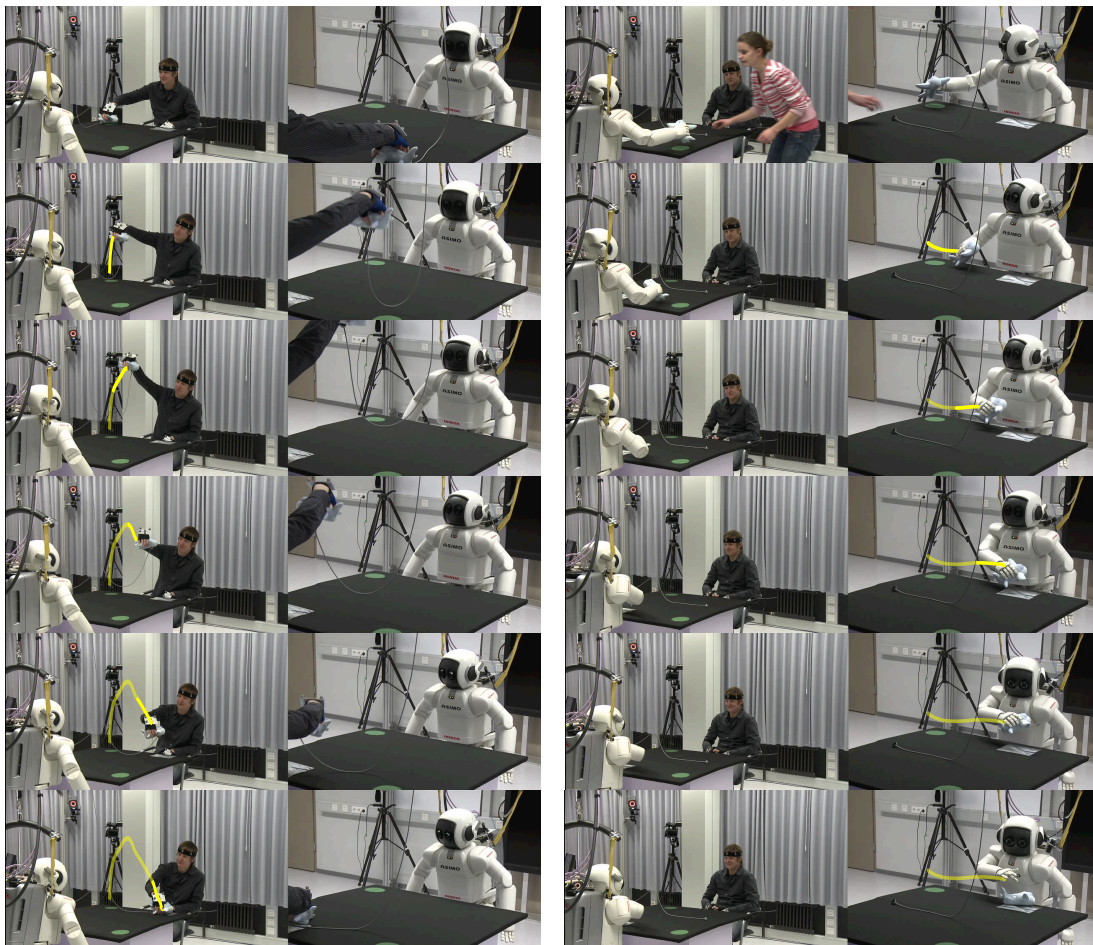
**Figure 7.5:** Example sequence of imitation of a goal-crucial action - The robot imitates (right) the subject's demonstration of the goal-crucial action of hanging up the phone (left).



**Figure 7.6:** Example sequence of emulation of a manner-crucial action - The robot emulates (right) the subject's demonstration of the manner-crucial action of how the frog jumps (left).

## 7. A HUMAN-ROBOT INTERACTION STUDY OF FEEDBACK IN AN IMITATION LEARNING SCENARIO

---



**Figure 7.7:** Example sequence of emulation of a goal-crucial action - The robot emulates (right) the subject's demonstration of the goal-crucial action of how the airplane flies to the airport (left).

		Feedback	
		Imitation	Emulation
Event knowledge	Reproduction behavior Action property		
	Manner-crucial	✓	✗
	Goal-crucial	-	✓

**Figure 7.8: Conditions** - The two-by-two design of the action properties and reproduction conditions. Check marks, minus, and crosses indicate the degree of correctness of the conditions. For explanation see Section 7.2.3.

Additionally, each participant was presented with one of three robot *online feedback* behaviors in terms of three robot eye gaze behaviors: *social gaze*, *random gaze*, and *static gaze*.

The robot's gaze was initially pointed at a fixed scene position (i.e., a point between the face of the tutor and the table).

**Social gaze** This robot gaze behavior was designed to reflect the learner's behavior observed in adult-child tutoring interactions. The findings of the analyses presented in Chapters 5 and 6 were incorporated in the design. The robot either exhibited attentive gaze following the object movements or anticipating expected end positions of the transported object.

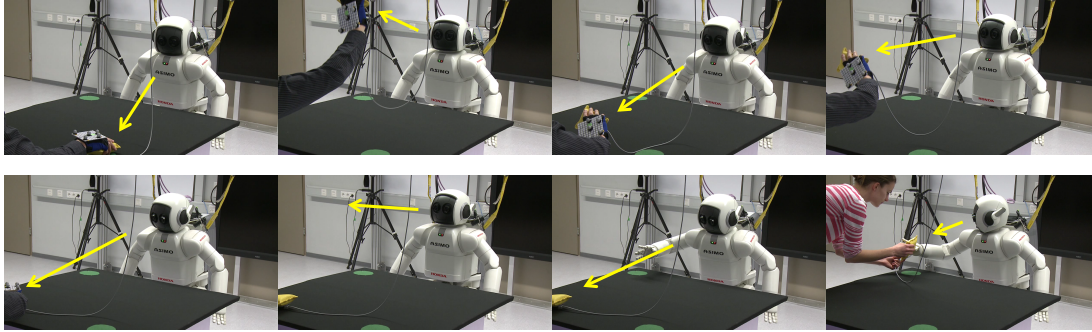
When the object was set down at the start position, the robot shifted its gaze toward the object. When the object was moving, thus, during the action demonstration, the robot gave continuous *online feedback* by following the object with its gaze depending on the turn-based feedback condition:

- **Imitation:** The robot followed the object with its gaze, until the subject had finished the action demonstration, see Figure 7.9 for an example sequence of eye gaze in this condition and see Section 7.2.4 for definitions of movement start and end.
- **Emulation:** The robot followed the object with its gaze for two seconds and then switched its gazing direction toward a predefined end position,

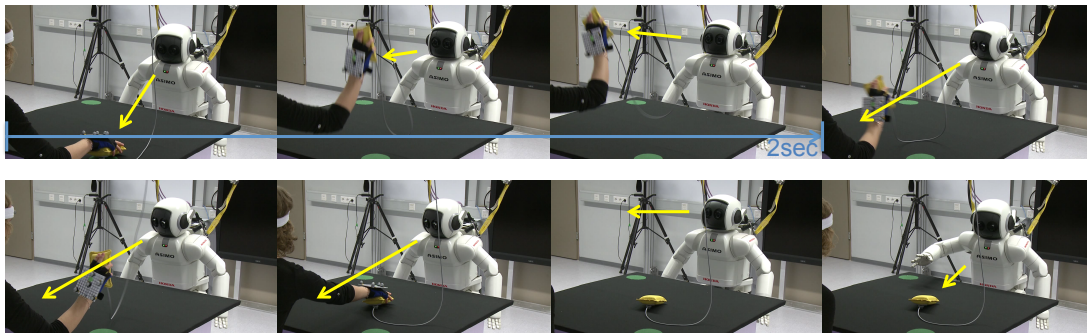
## 7. A HUMAN-ROBOT INTERACTION STUDY OF FEEDBACK IN AN IMITATION LEARNING SCENARIO

---

anticipating where the object should be set down. For an example sequence of eye gaze in this condition, see Figure 7.10.



**Figure 7.9: Social gaze imitation** - Window cleaning demonstration sequence with social gaze in imitation condition. Yellow arrows mark robot gaze direction.



**Figure 7.10: Social gaze emulation** - Window cleaning demonstration sequence with social gaze in emulation condition (i.e., anticipating gaze after two seconds). Yellow arrows mark robot gaze direction.

At the specific point in time, right after the task demonstration was complete, the robot again gave feedback in a sequence of actions. It gazed at the tutor's face and then to the object, while reaching out its right arm in direction of the object. After that, the robot followed the object, until it was placed into the robot's hand.

The social gaze condition additionally included a behavior after the robot replicated the action. While setting down the object on the table after the action replication was complete, the robot gazed at the object and after that at the tutor encouraging the tutor to give feedback to the shown replication.

**Random gaze** Here the robot's gaze had five directions between which it alternated beginning when the object was set down at the start position. For an example



sequence of eye gaze in this condition, see Figure 7.11. The duration of the gaze intervals and number of occurrences of a specific direction were designed to follow random distributions modeled after 12 to 24 months old children’s gaze directions during action demonstrations in parent-infant interactions. The intervals and gaze directions were investigated and corresponding statistics calculated on the Motionese corpus data, see Section 4.1. The fix points of the children’s gaze behavior was divided into four classes, of which only three were considered and their likelihood was calculated.

1. Gaze to object: 88.41%

To cover all relevant positions of the tutoring situation and task, this figure was divided into three equally distributed classes for the robot:

- Object: 29.47%
- Start position: 29.47%
- End position: 29.47%

2. Gaze to tutor’s face: 10.87%
3. Gaze to tutor’s stationary hand: 0.72%
4. Gaze elsewhere: not considered

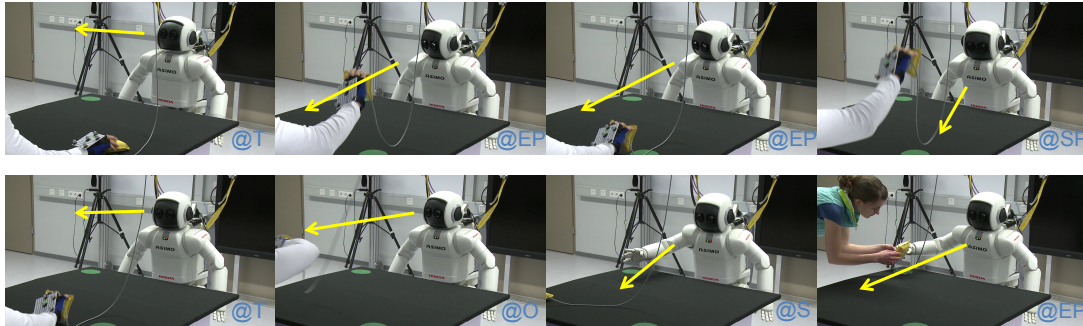
For the duration of gaze intervals to each of the three gazing directions, lognormal distributions were fit to the histograms of the data obtained from the corpus to serve as probability distributions for the modeled random gaze behavior.

1. Gaze to object (equal for all sub-classes):  $\mu = -0.246$ ,  $\sigma = 0.926$
2. Gaze to parent’s face:  $\mu = -0.586$ ,  $\sigma = 0.772$
3. Gaze to tutor’s stationary hand:  $\mu = -0.455$ ,  $\sigma = 0.711$

The fourth class of all gaze elsewhere than to the object, the parent or the stationary hand was not taken into account because the random gaze condition aimed at controlling the timing of gaze to relevant positions, but was not designed to include gaze to positions entirely irrelevant to the task, which independent of the timing of gaze trigger attention getters at any given moment. After the demonstration, the robot gazes to the fixed scene position between the table and the tutor’s face and lifts its arm to reach for the object. Concerning the end of the robot’s replication, in this gazing condition the robot gazes to the fixed scene position as well when releasing the object.

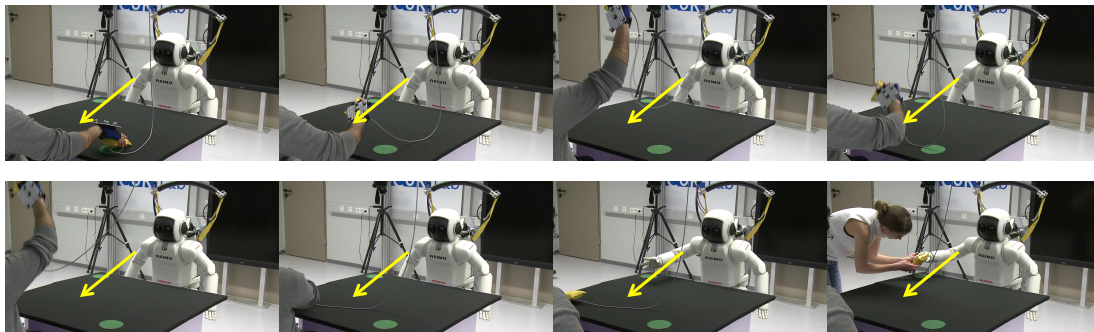
**Static gaze** In the static gazing condition, the robot maintained the fixed scene gazing direction at all times. For an example sequence of eye gaze in the static gaze condition, see Figure 7.12. This direction was chosen between the face of the

## 7. A HUMAN-ROBOT INTERACTION STUDY OF FEEDBACK IN AN IMITATION LEARNING SCENARIO



**Figure 7.11: Random gaze** - Window cleaning demonstration sequence with random gaze. Yellow arrows mark robot gaze direction additionally indicated at bottom right: @*T* gaze at tutor, @*EP* gaze at object end position, @*SP* gaze at object start position, @*O* gaze at object, @*S* gaze at fixed scene position.

tutor and the height of the starting point of the task, such that the tutor had the impression, the robot had witnessed the demonstration. After the action demonstration, the robot's gaze remained unchanged as it reached for the object. Concerning the end of the robot's replication, in this gazing condition the robot also gazes to the fixed scene position when releasing the object.



**Figure 7.12: Static gaze** - Window cleaning demonstration sequence with static gaze toward the scene. Yellow arrows mark robot gaze direction.

For the time, during which the robot replicated the movement, it was technically not possible to control the robot's head movements because this would have restricted the robot's Whole Body Motion Controller (Gienger et al., 2005) to an unsustainable degree (i.e., the robot would not have been able to perform the action as desired). Task order and action belonging to reproduction condition were randomized within the above constraints. After the eight tasks had been completed, the participants filled out a questionnaire and were interviewed. For the questionnaire and interview forms, see Appendix C.1 and D.1.

### 7.2.3 Hypotheses

Hypotheses were formed corresponding to the conditions described in Figure 7.8 prior to the study.

For the **action properties** and **turn-based feedback** (imitation/emulation):

In the condition of the robot *imitating a manner-crucial action* with an object, the tutor should treat the robot's action replication as being correct because the important aspect of the action—in this case the manner or motion path with which the action was performed—is reproduced. On that account the tutor might assume that the robot already knows the object and thus he/she does not repeat the demonstration.

Similarly, if the robot *reproduces a goal-crucial action by emulating* it and thus replicates the end-state (i.e., the result and goal) (which here are equivalent, cf. Section 2.2), and ignoring additional incidental or unnecessary parts of the movement, the tutor should also treat the action as being correctly replicated by the robot. Hence, the hypotheses for this condition are also the same.

The case of *replicating a manner-crucial action with an object by emulating* it, should be considered as being incorrect by the tutor because the robot only replicates the end position of the movement without regarding, what is essential to the presented task. The hypotheses for this condition are contrary to the ones in the previous conditions. Here the tutor should rather assume that the object is not known to the robot and thus repeats the demonstration.

The hypotheses for the remaining condition of the robot *imitating a goal-crucial action* are analogous. The tutor here does not consider the robot's replication of the action as correct because the robot does not distinguish between the important parts of the demonstrated action and precisely those movements, which are incidental or unnecessary. This hypothesis implies that the tutors' demonstrations include these incidental or unnecessary parts. This might not be the case for all demonstrations and therefore, this condition is considered as incorrect, but it might not be considered as incorrect as the case when a manner-crucial action is emulated. (For this reason in Figure 7.8 this condition is marked with a red minus.)

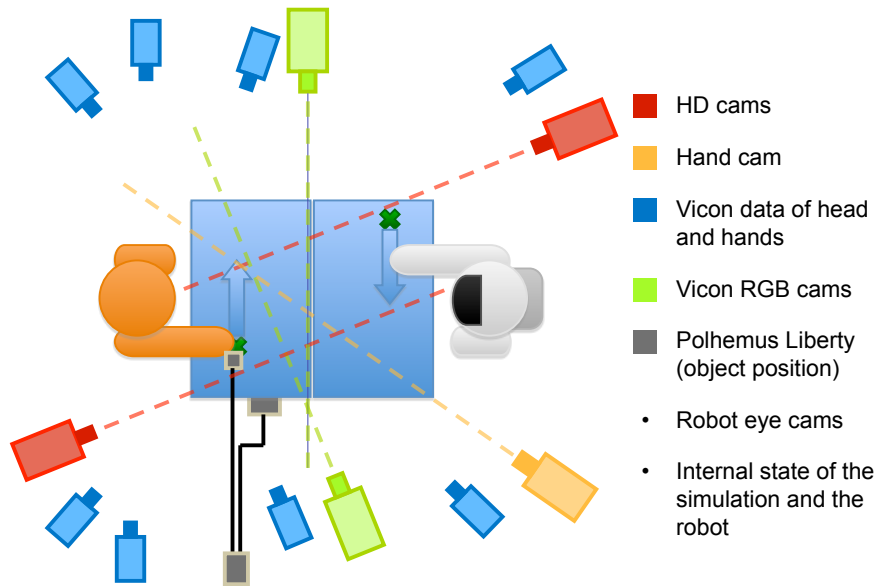
For the **online feedback** behavior, the different gaze behaviors also yield different hypotheses. The *social gaze* behavior was hypothesized to elicit stronger motionese behavior in the tutor's demonstrations as the other two conditions. In interactions with *random robot gaze* behavior the tutor's demonstration should include more disturbances or attention getters than in the other two conditions. Both, the social and the random gaze behavior, should provoke a different form of tutoring behavior than what can be observed in the *static gaze* condition.

## 7. A HUMAN-ROBOT INTERACTION STUDY OF FEEDBACK IN AN IMITATION LEARNING SCENARIO

---

### 7.2.4 Technical Realization

The current chapter includes an illustration of the data, which has been collected during the study, and a brief description of the technical design of the study. An overview of the collected data is given in Figure 7.13. The following data were recorded:



**Figure 7.13: Technical setup** - The technical realization involved several cameras and tracking systems

- Two videos were recorded of each interaction by two **high-definition cameras** at 50 frames per second. One of them focussed on the tutor and one on the robot and both recorded a frontal view on the subjects.
- a small **hand camera** filmed the interaction from the side of the table at 25 frames per second. This video was used for interview purposes.
- eight **Vicon cameras** registered the three-dimensional positions of the tutor's head and hands at  $100Hz$ . For this purpose the participants wore a head band and gloves equipped with sets of Vicon markers.
- Additionally, the scene was recorded by two **Vicon RGB cameras** at 25 frames per second. The two resulting videos are synchronous with the Vicon three-dimensional tracking data and were intended to be utilized for visualization purposes.
- The position of the objects in three-dimensional coordinates were recorded at  $40Hz$  by a Polhemus Liberty System, a magnetic field based tracking system.

A marker, which was linked to the tracking device via cable, was attached to each object. This was only needed during action demonstrations by the tutor for online usage (e.g., for the eye gaze direction).

- The robot eye cameras both recorded the scene at eight frames per second mainly for qualitative analysis.
- The internal state of the simulation and the robot were logged at  $40Hz$ .

The robot control architecture was developed by Manuel Mühlig and Michael Gienger (see (Mühlig et al., 2010)).

### Collecting Data from Sensors

The information from the Vicon system, the Polhemus Liberty system, the robot, and predefined static elements are subsumed in a “Persistent Object Memory” (POM), which kept track of the object in the scene and also handle temporary occlusions (e.g., of the subject’s hands) (Mühlig et al., 2010).

### Structuring the Interaction

Based on the POM, a state machine was used to structure the interaction. The whole scenario was automated including the segmentation of the movements.

### Automatic Movement Segmentation

The beginning and end of the movement presented by the tutor, which the robot should reproduce, were segmented from the movement stream automatically. For the two events, the following preconditions had to be met:

#### Movement start

- Right or left hand is near the object
- Object has a minimum speed
- Object leaves the start position

#### Movement end

- Object lies on the table
- Object does not move anymore
- No hand is near the object

## 7. A HUMAN-ROBOT INTERACTION STUDY OF FEEDBACK IN AN IMITATION LEARNING SCENARIO

---

### 7.3 Data Analysis

59 subjects participated in the study in 13 days. Each run took approximately 45 minutes. In sum around 200 hours of video material were recorded. All capturing devices presented in the previous section together produced about 1.6 TB of data. Moreover one folder full of questionnaires and interview notes was filled during the study.

#### 7.3.1 Methods

In a first part, this section shows the variables considered in the analysis. In a second part, the steps of preprocessing the data necessary for assessing the dependent variables are described and the last part is concerned with the actual measures used to assess the variables.

##### **Independent Variables:**

The independent variables are available at the beginning of the study and controlled by the experimental design. The following independent variables are involved in the study:

- The robots replication behavior: imitation/emulation
- The robot's gazing behavior during the action demonstration
  - Designed social gaze (during demonstration tied to imitation behavior)
    - \* Imitation: following object with gaze
    - \* Emulation: anticipating gaze
  - Random gaze
  - Static gaze
- Object movement properties: goal- or manner-crucial
- Experience with robots before study

##### **Dependent Variables:**

The dependent variables are created during the study and constitute what is measured and observed as output of the study. The study comprises the following dependent variables:

- Number of demonstrations

- Motionese behavior (velocity, pace, roundness etc.)
- Use of Attention Getters
- Subjects experience from
  - Questionnaire
  - Interview

Once the data were collected, they were synchronized and the videos were cut automatically. Several features, which will be presented in the next paragraph, were calculated on the trajectory data and the results were evaluated statistically.

### Motionese Measures

To be able to assess the behavior modifications in the tutors' demonstrations, the quantitative measures for motionese as described in Section 3.2.4 were utilized and the number of times each action was demonstrated to the robot was counted. For the trajectory data, the tracked object positions obtained via the Polhemus Liberty system were utilized. Because the trajectory data for the current analysis are thus three-dimensional, in addition five measures were defined to analyze the movement in its third dimension. The *depth*, *height*, and *width* of the movement of the demonstration were calculated as the minimum distance the object had to the robot, the maximum distance of the object to the table, and the maximum span of the movement from the tutor's right to left. The *area* of the movement was defined as the overall area of the convex hull of the movement in two dimensions. Analogously the *volume* of the movement was measured as the volume of the convex hull of the movement in three dimensions during the demonstration. All measures were computed for all objects and averaged over the four manner-crucial actions on the one hand and the four goal-crucial actions on the other hand.

### 7.3.2 Results

Results revealed differences in the participants' action demonstrations with respect to three factors: the participants *event knowledge*, the robots *turn-based feedback* and the robots *online feedback*.

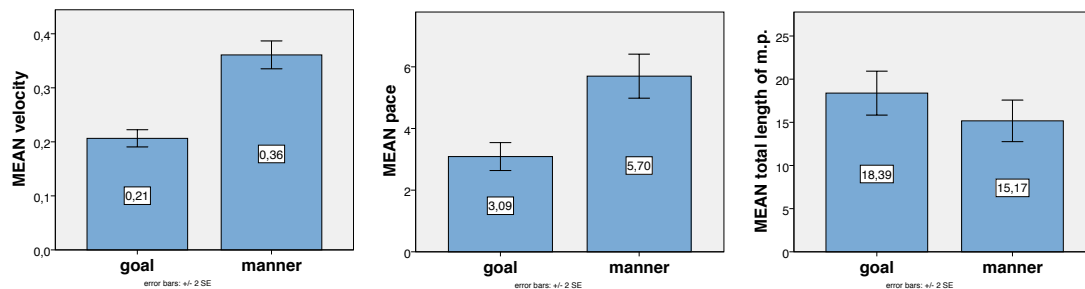
#### Event Knowledge

The participants' event knowledge was assessed using a one-way analysis of variance (one-way ANOVA) on the motionese measures calculated on the first demonstration of each object and averaged over the four manner-crucial actions on the one hand and the

## 7. A HUMAN-ROBOT INTERACTION STUDY OF FEEDBACK IN AN IMITATION LEARNING SCENARIO

four goal-crucial actions on the other hand. The object movement property (manner- or goal-crucial) was considered an inner-subject factor in the repeated measures design. Only the tutors' first demonstrations of each object were included in this analysis because they were not influenced by the robot's turn-based feedback, yet. Several significant differences have been found. (Results without significance are omitted due to the number of variables.)

The *action length* differed significantly across the two groups: manner ( $M = 9.81$ ,  $SD = 3.9$ ) and goal ( $M = 6.58$ ,  $SD = 2.24$ ),  $F(1, 53) = 71.65$ ,  $p = 0.000$ . Thus, the participants demonstrated manner-crucial actions longer than goal-crucial actions. In addition the same was revealed concerning the speed of the demonstrations. Manner-crucial actions ( $M = 0.36$ ,  $SD = 0.1$ ) were carried out faster than goal-crucial actions ( $M = 0.21$ ,  $SD = 0.06$ ), *velocity*:  $F(1, 53) = 259.19$ ,  $p = 0.000$ , with higher *acceleration* (manner-crucial:  $M = 2.01$ ,  $SD = 0.7$ , goal-crucial:  $M = 1.1$ ,  $SD = 0.38$ ),  $F(1, 53) = 198.98$ ,  $p = 0.000$ , and with higher *pace* (manner-crucial:  $M = 5.7$ ,  $SD = 2.62$ , goal-crucial:  $M = 3.09$ ,  $SD = 1.66$ ),  $F(1, 53) = 47.28$ ,  $p = 0.000$ , see Figure 7.14.



**Figure 7.14: Event knowledge results for speed in bar charts** - Exemplarily, the values for velocity (left) and pace (middle), and total length of motion pauses are represented to illustrate the differences in speed of the demonstrations for the different object movement properties: goal- and manner-crucial.

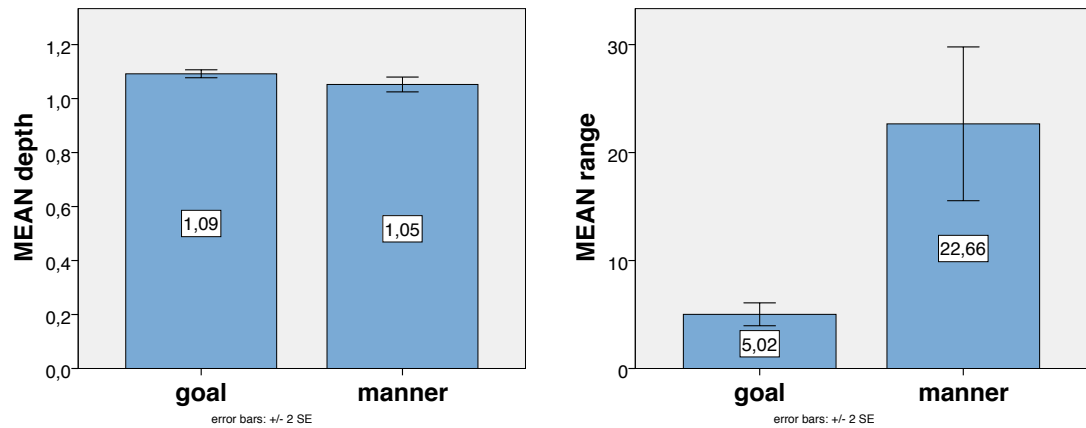
For the *total length of motion pauses*, the manner-crucial actions ( $M = 15.17$ ,  $SD = 9.03$ ) were demonstrated with less pauses than the goal-crucial actions ( $M = 18.39$ ,  $SD = 9.52$ ),  $F(1, 53) = 5.85$ ,  $p = 0.019$ .

The manner-crucial actions ( $M = 3.46$ ,  $SD = 1.39$ ) were also carried out with less *roundness* than the goal-crucial actions ( $M = 1.43$ ,  $SD = 0.43$ ),  $F(1, 53) = 102.28$ ,  $p = 0.000$ .

For *range* the results show that participants demonstrated manner-crucial actions ( $M = 22.66$ ,  $SD = 26.64$ ) with a higher range than goal-crucial actions ( $M = 5.02$ ,  $SD = 3.96$ ),  $F(1, 53) = 26.01$ ,  $p = 0.000$ . They furthermore demonstrated manner-crucial actions ( $M = 0.55$ ,  $SD = 0.08$ ) with wider movement than goal-crucial actions ( $M =$



0.47,  $SD = 0.05$ ), *width*:  $F(1, 53) = 43.49$ ,  $p = 0.000$ , manner-crucial actions ( $M = 0.49$ ,  $SD = 0.08$ ) with higher movement than goal-crucial actions ( $M = 0.32$ ,  $SD = 0.09$ ), *height*:  $F(1, 53) = 205.66$ ,  $p = 0.000$  and manner-crucial actions ( $M = 1.05$ ,  $SD = 0.1$ ) with closer proximity to the robot than goal-crucial actions ( $M = 1.09$ ,  $SD = 0.06$ ), *depth*:  $F(1, 53) = 10.45$ ,  $p = 0.002$ , see Figure 7.15. The same seems to be the case for *area* and *volume* of the demonstrations. The area and volume of the demonstrations of manner-crucial actions ( $M = 0.19$ ,  $SD = 0.05$ ,  $M = 0.012$ ,  $SD = 0.006$ , respectively) were greater than those of the goal-crucial actions ( $M = 0.07$ ,  $SD = 0.02$ ,  $M = 0.002$ ,  $SD = 0.001$ , respectively), area:  $F(1, 53) = 369.66$ ,  $p = 0.000$ , volume:  $F(1, 53) = 140.29$ ,  $p = 0.000$ .



**Figure 7.15: Event knowledge results for range in bar charts** - Exemplarily, the values for depth (i.e., smallest number represents closest proximity to robot) (left) and range (right) are represented to illustrate the differences in range of the demonstrations for the different object movement properties: goal- and manner-crucial.

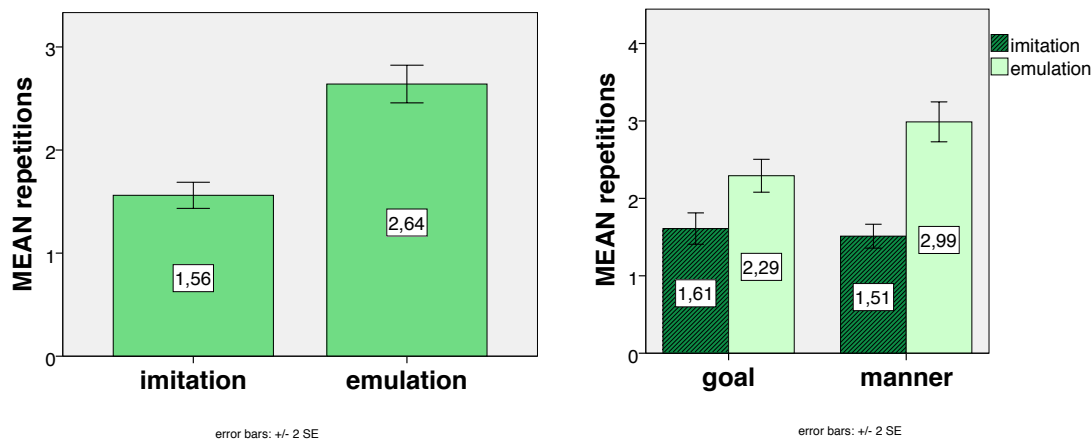
### Turn-based Feedback

A one-way within subjects (or repeated measures) ANOVA was conducted to compare the effect of robot's turn-based feedback behavior on the number of the tutor's demonstrations in imitation and emulation conditions.

Results revealed a significant effect of robot's turn-based feedback behavior on the number of times the tutor repeated the demonstration, Wilks' Lambda,  $\Lambda = 0.22$ ,  $F(1, 40) = 140.93$ ,  $p = 0.000$ . The participants thus repeated the demonstration more often, when the robot emulated ( $M = 2.64$ ,  $SD = 0.75$ ) than when it imitated ( $M = 1.56$ ,  $SD = 0.57$ ) the action, see Figure 7.16. The highest number of demonstrations was carried out, when a manner-oriented action was presented, which the robot

## 7. A HUMAN-ROBOT INTERACTION STUDY OF FEEDBACK IN AN IMITATION LEARNING SCENARIO

emulated ( $M = 2.99$ ,  $SD = 0.83$ ), see Figure 7.16. To protect against violating the assumption of normality, variables additionally were transformed and provided results of the same significance.



**Figure 7.16: Turn-based feedback results in bar charts** - The number of repetitions is shown depending on replication behavior: imitation and emulation (left) and replication and object movement property: goal- and manner-crucial (right).

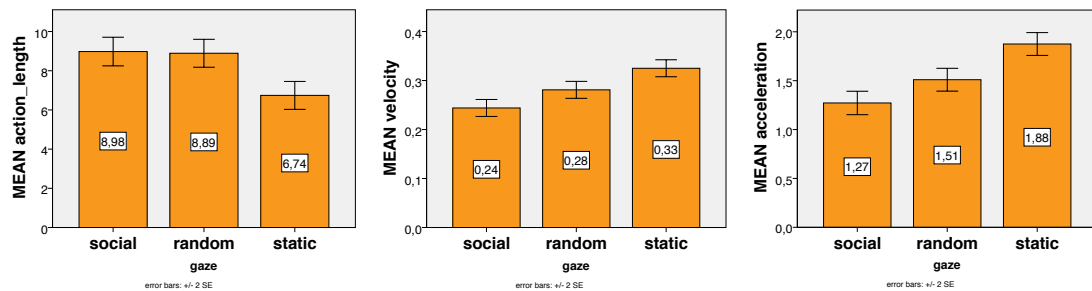
### Online Feedback

The effect of the online feedback behavior on the motionese features of the demonstration the tutor carried out, was considered using a one-way between subjects ANOVA in the social gaze, random gaze, and static gaze conditions. Here, also, only the first demonstrations of the actions were considered. (Results without significance are omitted due to the number of variables.)

There was a significant effect of robot's online feedback on the *velocity* and *acceleration* of the presentation, velocity:  $F(2, 53) = 7.302$ ,  $p = 0.002$  and acceleration:  $F(2, 53) = 8.824$ ,  $p = 0.000$ , see Figure 7.17. A Scheffé test was used to make post hoc comparisons between conditions. It uncovered that participants in the social gaze condition (velocity:  $M = 0.24$ ,  $SD = 0.08$ , acceleration:  $M = 1.27$ ,  $SD = 0.53$ ) demonstrated significantly slower than in the static gaze condition (velocity:  $M = 0.32$ ,  $SD = 0.07$ , acceleration:  $M = 1.87$ ,  $SD = 0.53$ ), velocity:  $p = 0.002$  and acceleration:  $p = 0.001$ . For velocity the comparison between the other groups did not reveal any significant results, but for acceleration the test uncovered that participants in the random gazing condition ( $M = 1.51$ ,  $SD = 0.39$ ) also demonstrated with a lower acceleration than participants with static robot gaze ( $M = 1.87$ ,  $SD = 0.53$ ),  $p = 0.046$ . Likewise, there was a significant effect of robot's online feedback on the *action length* of

the presentation,  $F(2, 53) = 4.18$ ,  $p = 0.021$ , see Figure 7.17. Again a Scheffé test was used to make post hoc comparisons between conditions. It revealed that participants in the social gaze condition ( $M = 8.98$ ,  $SD = 2.98$ ) demonstrated significantly longer than in the static gaze condition ( $M = 6.74$ ,  $SD = 2.23$ ),  $p = 0.049$ .

Additionally to the computational evaluation, the data was also analyzed qualitatively. The qualitative findings suggest that participants attended to the robot's gaze following in the social gazing condition. They delayed the upward movement of the object and adjusted the presentation to the robot's gazing speed. It seems that this was mostly done when the current movement was important for the task. Unimportant parts of an action, like the lifting of the elevator, were carried out faster without waiting on the robots gaze. When showing a goal-crucial action, which the robot emulated, the robot's anticipating gaze was nearly never noticed. When showing a manner-crucial action, the participants treated the anticipating gaze in the robot's emulation condition as lack of attention and tried to repair it by applying attention getting devices (e.g., pausing the motion).



**Figure 7.17: Online feedback results in bar charts** - Exemplarily, the values for action length (left), velocity (middle), and acceleration (right) are represented to illustrate the differences in speed of the demonstrations for the different gaze directions: social, random, and static gaze.

Also, the participants clearly assumed a connection between this lack of attention and the following incorrect replication of the movement. They adjusted their following action demonstrations, presenting the action slower and simpler (e.g., by using straight movements without curves and suppressing unnecessary movements like letting the dog eat out of the bowl) or using the strategy to hold the object so that the marker, which was attached to the object, was visible to the robot for example. Also during the interview one participant uttered that the robot could not do it if he did not look.

## 7. A HUMAN-ROBOT INTERACTION STUDY OF FEEDBACK IN AN IMITATION LEARNING SCENARIO

---

### 7.4 Discussion

Results show that participants' action demonstrations were determined by three factors: the participant's event knowledge, the robot's turn-based feedback and the robot's online feedback.

**Event Knowledge** Participants' event knowledge significantly influenced action demonstrations. Manner crucial actions were demonstrated significantly longer (action length), faster (movement velocity, acceleration, and motion pauses), rounder, as well as with significantly more range (range of motion, width, height, depth, area, and volume).

**Turn-based Feedback** The robot's turn-based feedback determines how often a subject demonstrates the action. Emulated actions were shown significantly more often than imitated ones. Manner-oriented actions, which were emulated by the robot, were demonstrated most often. These results partly confirm our initial hypotheses. Interestingly, contrary to the hypotheses, the robot's way of replication for imitated goal-crucial actions was considered more correct than for emulated goal-crucial actions. On the one hand, the fact that imitation of a goal-crucial action was not considered incorrect, could be explained by the omission of unnecessary and incidental movements during the action demonstrations. On the other hand—even though not significantly—the imitated goal-crucial actions were repeated slightly less often than the emulated goal-crucial actions. This could indicate that participants pay close attention to the details of goal-crucial actions. Also, the actions, which were demonstrated to the robot might not be exclusively goal-crucial, but could also involve a certain manner, which was reproduced by the robot in the imitation, but not in the emulation condition. For example participants paid close attention on how the airplane landed, so that imaginary passengers were not injured. It was also hypothesized that emulated manner-crucial actions were considered as incorrectly reproduced, which was supported by the statistical results.

**Online Feedback** The robot's gaze behavior had an impact on the tutor's action demonstration. This confirms the initial hypothesis on the gaze conditions. When the robot was in the social gaze condition, the participants demonstrated the actions significantly slower (action length, velocity, acceleration) compared to the static gaze condition.

Also participants noticed in the manner-crucial action condition that the robot's anticipating gaze was coupled with the following turn-based incorrect reproduction (emulation) of the action.

Thus the results reveal that the robot's feedback shapes the interaction. Concerning the turn-based feedback, the findings imply that imitation seems to be considered more correct than emulation, independent of the nature of the action. The action was repeated significantly most often, when a manner-crucial action was emulated. Koterba and Iverson investigated the effect of motionese on infant object manipulation (Koterba and Iverson, 2009). They conducted an exploratory study with 8- to 10-month-old infants and presented movements with different numbers of enhanced action parameters (demonstrations with varying amplitude (low amplitude and high amplitude) and varying number of repetitions (low repetition: one time, and high repetition: four times)) to them. Koterba and Iverson found that infants engaged in bangs and shakes of the object for a significantly longer time, when they saw a high number of repetitions, and argued that this might display the infants' efforts to imitate the demonstrated action and that infants focussed more on the tutor's movements during demonstration. In the low-repetition condition infants focussed on and explored the object more. This is in agreement with our findings of the human-robot interaction study. Tutors repeated a manner-crucial action, which the robot emulated, most often relative to all other conditions presumably with the assumption that the robot would finally correctly replicate the action by imitating it.

Concerning the online feedback, the results reveal significant differences in speed only between the social and the static gaze conditions. This supports the findings of the qualitative analysis, which showed that the participants adjusted their demonstration speed to the robot's gaze speed. The connection of the robot's anticipating gaze to the failed reproduction of the action shows that the robot's gaze is interpreted as being intentional and as reacting upon their actions. The participants monitor the robot's gaze and attention and even try to repair it. They clearly use the robot's gaze as an indicator of what the robot has understood of the action. This suggests that the robot's gaze is a powerful instrument for the robot to shape the action demonstration according to its benefits.

#### 7.4.1 Feedback Strategies

The following ideas on how a robot could employ its gazing behavior to improve learning of actions from a human tutor were developed in close cooperation with Manuel Mühlig (Honda Research Institute Europe, Offenbach/Main, Germany). For the goal of realizing this capability, it is necessary that the robot has a perceivable "gaze direction", for example a controllable head with one or more cameras. Additionally, the robot should exhibit a perceivable gazing strategy, such as an attentive gazing behavior, where the robot tracks the tutor's hand or an object, which is involved in the

## 7. A HUMAN-ROBOT INTERACTION STUDY OF FEEDBACK IN AN IMITATION LEARNING SCENARIO

---

demonstration. Even further, a more complex social gazing is possible, which leads to a higher chance that the tutor anthropomorphizes the robot. During the tutor's action demonstration the robot could generate hypotheses about the degree of importance of parts of the shown movement. The system could then employ different strategies for hypotheses generation. For example, when the tutor repeatedly shows a movement, the robot compares the respective demonstrations using a method as for example Dynamic Time Warping (Mühlig et al., 2009). This leads to a degree of similarity for specific parts of the compared movements. Movement parts exhibiting a high variance over demonstrations can be hypothesized to be unimportant, whereas parts with low variance are important for the execution of the demonstrated action and can thus not be modified much, when the robot carries out the movement. Another strategy could be for example to measure the speed of the demonstration and classify fast parts as unimportant and parts carried out more slowly as important.

Once the robot has generated a hypothesis on the importance of a certain movement part, there are several possible ways of determining the validity of the hypothesis. In the following, some strategies are described exemplarily. To detect the onset of a demonstrated task, the robot systematically directs its gaze toward points with only low saliency instead of the point, where it hypothesizes the action to take place. The tutor will then change his/her behavior by for instance pausing the movement or shaking/waving his/her hand or the object with which the action is executed trying to reorient the robots attention or shifting his/her own gaze toward the important point, the robot should attend to.

To test a generated hypothesis about the degree of importance of a certain movement part, the robot could actively look away from the object toward an estimated goal position (from increasing the speed of tracking to anticipating the goal position) to test if the focus of the demonstrated movement is more on the goal or on how it is reached. Both cases are distinguishable based on the behavior of the tutor. The robot can recognize this behavior: if the tutor does not react and the estimated goal position is the correct one, the path and manner are not an important part of the action.

Another strategy is that the robot actively reduces its object tracking speed to evaluate if certain parts of a demonstrated movement are task-relevant. It is likely that in this case the tutor would reduce its demonstration speed, which is a hint for importance to the robot.

Also during the demonstration, the robot could actively look back to the starting point again if—based on its experience—it is not clear if the current demonstration is important or related to previously shown demonstrations. Under the assumption that the tutor stops his/her demonstration and restarts it, when he/she recognized the robot's behavior, it is possible for the robot to recognize actual importance of the currently shown action.

## 8

# Conclusion

The goal of the thesis is to contribute answers to the issue of enabling robots to learn new manipulative actions in social interaction. It started out with a set of research questions revolving around the issue. The first one arose from the idea of letting robots learn the way infants learn, assuming that infants' acquisition of new skills is supported by their social environment, especially in interaction with caregivers, and is concerned with what constitutes a natural tutoring interaction. The second question addresses a methodical issue and requests new approaches to computationally analyze human behavior in naturalistic interaction. The third question deals with the development of technical systems, which are able to learn in social interaction with a human tutor. It focusses on imitation learning as the most important field of social learning addressed in robotics and the open question of what to imitate.

### **What Constitutes a Natural Tutoring Interaction?**

This question was investigated throughout all analyses presented. The analysis described in Chapter 4 reveals that the robot's feedback is important for the robot to be recognized as a full interaction partner and it was shown that the tutor's behavior is modified according to the learner's capabilities, understanding, and needs. In the detailed analyses in Chapter 6, the importance of the coordination and interplay of the tutor's and learner's actions for natural tutoring interactions becomes evident. A natural tutoring interaction is bidirectional and allows for mutual online analysis.

For human-robot interaction the findings imply that a robot should give feedback, with which it signals its understanding of the current action demonstration. It is argued that a tutoring situation such as for imitation learning should necessarily be bidirectional and interactive to a high degree in order to enable robots to learn manipulative actions from human demonstrations.

## 8. CONCLUSION

---

### **How Can Human Behavior in Naturalistic Interaction Be Analyzed?**

Human behavior in naturalistic interactions is very complex and variable. Therefore, methods were needed to analyze this behavior computationally on a corpus of data. As described in Chapter 3, the methods utilized for the analyses are interdisciplinary and involve initial manual qualitative analyses based on conversation analysis and automatic quantitative analyses including formal descriptions, visualizations and statistical verifications. The developed methods are valuable for research in behavior analysis as they provide means to investigate human behavior in naturalistic interaction.

### **How Could a Robot Know What is Important about a Shown Action and What to Imitate?**

With feedback a robot could signal its understanding of a demonstrated action to the human tutor or even actively trigger tutoring behavior and modifications, which are beneficial for its learning processes. The discussion of the results obtained in Chapter 7 suggests concrete strategies for this achievement. The robot is proposed to employ its gazing behavior during the tutor's action demonstration to draw information from the tutor's reactions to form and corroborate hypotheses about the degree of importance of parts of the action.

### **Discussion and Future Perspectives**

The thesis has cast light on the issue of letting robots learn in social interaction. By investigating tutoring interactions of adults and their children, this work has shown that the children learners—through their feedback—actively influence and shape the tutoring interaction. In a human-robot interaction study it has been demonstrated that the robot's feedback can also shape the tutoring interaction. The feedback implemented was designed to be perceived to reflect the robot's understanding of the demonstrated action. The robot's understanding relies on the technical implementation of learning mechanisms. In the current study, the robot thus pretended to have prior knowledge and understanding of the action, even though there was no learning of the action involved. Future research should involve robotic systems equipped with learning mechanisms, which online and incrementally could build representations of a shown action and adapt them by observing the tutor's demonstration and using the tutor's social signals. The robot could generate hypotheses about what to imitate and give feedback controlled by a feedback module during the action demonstration. By that means the robot could elicit changes in the tutor's action presentation, with which in turn the robot could support or falsify its hypotheses. The strategies proposed in Section 7.4 could be tested with such a system in further studies with human tutors. Moreover, the coupled system with its combined feedback and learning mechanisms



---

could be evaluated by testing it against a system with the learning mechanism alone, where it is expected that the coupled system should be able to outperform the simpler observer-only system in speed and accuracy.

Further implications for research in human-robot interaction and imitation learning can be derived from the fact that robots shape the tutors demonstrations with their actions. The interactional aspect of social learning has to shift to the focus of efforts in robotics research on this topic. Tutoring does not take place in uni-directional information flow. The tutor has knowledge about the task he/she should teach. Tutoring is not about bringing this knowledge to the learner's head, but the learner also has prior knowledge, experience and capabilities, which together with the tutors ideas form ever new concepts through both co-participants actions shaping the interaction while it is being created. Only through consideration of the dynamical processes of interaction can natural tutoring situations be realized in human-robot interaction and robots infer the goals of actions. This route might bring research closer to the comprehension of the development of social cognition and is worth following.

## 8. CONCLUSION

---

# References

- B. D. Argall, S. Chernova, M. Veloso, and B. Browning. A survey of robot learning from demonstration. *Robotics and Autonomous Systems*, 57(5):469–483, 2009. ISSN 09218890. doi: 10.1016/j.robot.2008.10.024. 12, 13
- C. G. Atkeson and S. Schaal. Robot learning from demonstration. In *Proc 14th International Conference on Machine Learning*, pages 12–20. Morgan Kaufmann, 1997. ISBN 1558604863. 13
- E. Bates, I. Bretherton, and L. Snyder. *From first words to grammar: Individual differences and dissociable mechanisms*. Cambridge University Press, 1988. 31
- P. J. Bauer and E. E. Kleinknecht. To ape or to emulate? Young childrens use of both strategies in a single study. *Developmental Science*, 5(1):18–20, 2002. ISSN 1363755X. doi: 10.1111/1467-7687.00197. 12, 14
- H. Bekkering, A. Wohlschlagel, and M. Gattis. Imitation of gestures in children is goal-directed. *Quarterly Journal of Experimental Psychology*, 53(1):153–164, 2000. ISSN 14640740. doi: 10.1080/027249800390718. 12
- H. Benedict. Early lexical development: comprehension and production. *Journal of Child Language*, 6(2):183–200, 1979. 31
- A. Billard and M. J. Matarić. Learning human arm movements by imitation: Evaluation of a biologically inspired connectionist architecture. *Robotics and Autonomous Systems*, 37(2-3):145–160, 2001. ISSN 09218890. doi: 10.1016/S0921-8890(01)00155-5. 13
- A. G. Billard, S. Calinon, and F. Guenter. Discriminative and adaptive imitation in uni-manual and bi-manual tasks. *Robotics and Autonomous Systems*, 54(5):370–384, 2006. ISSN 09218890. doi: 10.1016/j.robot.2006.01.007. 13
- R. J. Brand and W. L. Shallcross. Infants prefer motionese to adult-directed action. *Developmental Science*, 11(6):853–861, 2008. ISSN 1467-7687. 16, 29
- R. J. Brand, D. A. Baldwin, and L. A. Ashburn. Evidence for motionese: modifications in mothers infant-directed action. *Developmental Science*, 5(1):72–83, 2002. ISSN 1467-7687. 16, 17, 22, 23, 29, 42
- R. J. Brand, W. L. Shallcross, M. G. Sabatos, and K. P. Massie. Fine-Grained Analysis of Motionese: Eye Gaze, Object Exchanges, and Action Units in Infant-Versus Adult-Directed Action. *Infancy*, 11(2):203–214, 2007. ISSN 1532-7078. 17, 22, 27, 37, 42
- C. Breazeal and B. Scassellati. Challenges in Building Robots That Imitate People. In K. Dautenhahn, editor, *Imitation in animals and artifacts*, pages 363–390. MIT Press, 2002. ISBN 0262042037. 8
- C. Breazeal, A. Brooks, J. Gray, G. Hoffman, C. Kidd, H. Lee, J. Lieberman, and A. Lockerd. Tutelage and Collaboration for Humanoid Robots. *International Journal of Humanoid Robotics*, 1(2):315–348, 2004. 14
- C. Breazeal, M. Berlin, A. Brooks, J. Gray, and A. L. Thomaz. Using perspective taking to learn from ambiguous demonstrations. *Robotics and Autonomous Systems*, 54(5):385–393, 2006. ISSN 09218890. doi: 10.1016/j.robot.2006.02.004. 80
- A. Brugger, L. A. Lariviere, D. L. Mumme, and E. W. Bushnell. Doing the right thing: infants’ selection of actions to imitate from observed event sequences. *Child Development*, 78(3):806–824, 2007. 9, 14
- H. Brugman and A. Russel. Annotating multimedia/multimodal resources with ELAN. In M. Lino, M. Xavier, F. Ferreira, R. Costa, and R. Silva, editors, *Proceedings of the Fourth International Conference on Language Resources and Evaluation*, pages 2065–2068. ELRA, 2004. 19
- D. Buttelmann, M. Carpenter, J. Call, and M. Tomasello. Enculturated chimpanzees imitate rationally. *Developmental Science*, 10(4):F31–F38, 2007. 11
- S. Calinon and A. Billard. Incremental learning of gestures by imitation in a humanoid robot. *Proceeding of the ACM/IEEE International Conference on Human-Robot Interaction*, page 255, 2007. doi: 10.1145/1228716.1228751. 14
- S. Calinon, F. Guenter, and A. Billard. Goal-Directed Imitation in a Humanoid Robot. In *IEEE International Conference on Robotics and Automation*, volume 1, pages 299–304. IEEE, 2005. ISBN 078038914X. doi: 10.1109/ROBOT.2005.1570135. 13
- J. Call and M. Carpenter. Three Sources of Information in Social Learning. In *Imitation in animals and artifacts*, pages 211–228. The MIT Press, 2002. ISBN 0262042037. 9, 10, 12
- J. Call and M. Tomasello. The Social Learning of Tool Use by Orangutans (*Pongo pygmaeus*). *Human Evolution*, 9(4):297–313, 1994. ISSN 03939375. doi: 10.1007/BF02435516. 11
- M. Carpenter and J. Call. The question of ‘what to imitate’: inferring goals and intentions from demonstrations. In C. L. Nehaniv and K. Dautenhahn, editors, *Imitation and Social Learning in Robots Humans and Animals*, chapter 7, pages 135–151. Cambridge University Press, 2007. 7
- M. Carpenter, J. Call, and M. Tomasello. Twelve-and 18-month-olds copy actions in terms of goals. *Developmental Science*, 8(1):F13–F20, 2005. ISSN 1467-7687. 29
- S. C. F. Chong, J. F. Werker, J. A. Russell, and J. M. Carroll. Three Facial Expressions Mothers Direct to Their Infants. *Infant and Child Development*, 232(3):211–232, 2003. ISSN 15227227. doi: 10.1002/icd. 16
- R. Cooper. The development of infants’ preference for motherese. *Infant Behavior and Development*, 20(4):477–488, 1997. ISSN 01636383. doi: 10.1016/S0163-6383(97)90037-0. 16

## REFERENCES

---

- G. Csibra and G. Gergely. The teleological origins of mentalistic action explanations: A developmental hypothesis. *Developmental Science*, 1(2):255–259, 1998. ISSN 1363755X. doi: 10.1111/1467-7687.00039. 12
- G. Csibra and G. Gergely. Social learning and social cognition: The case for pedagogy. *Processes of change in brain and cognitive development. Attention and performance*, 21, 2005. 27, 29, 32
- K. Dautenhahn. Getting to know each other Artificial social intelligence for autonomous robots. *Robotics and Autonomous Systems*, 16(2-4):333–356, 1995. ISSN 09218890. doi: 10.1016/0921-8890(95)00054-2. 13
- H. De Jaegher, E. Di Paolo, and S. Gallagher. Can social interaction constitute social cognition? *Trends in Cognitive Sciences*, 14(10):441–447, 2010. 80
- J. Decety. Do imagined and executed actions share the same neural substrate? *Brain Research*, 3(2):87–93, 1996. 11
- J. Decety, D. Perani, M. Jeannerod, V. Bettinardi, B. Tadary, R. Woods, J. C. Mazziotta, and F. Fazio. Mapping motor representations with positron emission tomography. *Nature*, 371(6498):600–602, 1994. ISSN 00280836. doi: 10.1038/371600a0. 11
- Y. Demiris and G. Hayes. Imitation as a dual-route process featuring predictive and learning components : a biologically-plausible computational model. *Electronic Engineering*, 2002:327–361, 2002. 13
- G. Di Pellegrino, L. Fadiga, L. Fogassi, V. Gallese, and G. Rizzolatti. Understanding motor events: a neurophysiological study. *Experimental Brain Research*, 91(1):176–180, 1992. 11
- B. Estigarribia and E. V. Clark. Getting and maintaining attention in talk to young children. *Journal of Child Language*, 34(4):799–814, 2007. 48, 60
- L. Fadiga, L. Fogassi, G. Pavesi, and G. Rizzolatti. Motor facilitation during action observation: a magnetic stimulation study. *Journal of Neurophysiology*, 73(6):2608–2611, 1995. 11
- T. Falck-Ytter, G. Gredebäck, and C. Von Hofsten. Infants predict other people’s action goals. *Nature Neuroscience*, 9(7):878–879, 2006. doi: 10.1038/nn1729. 74
- I. Fasel, G. O. Deak, J. Triesch, and J. Movellan. Combining embodied models and empirical research for understanding the development of shared attention. *Proceedings 2nd International Conference on Development and Learning ICDL 2002*, pages 21–27, 2002. doi: 10.1109/DEVLRN.2002.1011724. 27
- C. A. Ferguson. Baby Talk in Six Languages. *Communication*, 66(6):103–114, 1964. 15
- A. Fernald. The perceptual and affective salience of mothers’ speech to infants. In L. Feagans, D. Garvey, and R. Golinkoff, editors, *The Origins and Growth of Communication*, pages 5–29. Ablex, 1984. 16
- A. Fernald. Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development*, 8(2):181–195, 1985. ISSN 01636383. doi: 10.1016/S0163-6383(85)80005-9. 15, 16
- A. Fernald and P. Kuhl. Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development*, 10(3):279–293, 1987. ISSN 01636383. doi: 10.1016/0163-6383(87)90017-8. 16
- A. Fernald and C. Mazzie. Prosody and focus in speech to infants and adults. *Developmental Psychology*, 27(2):209–221, 1991. ISSN 0012-1649. 29
- A. Fernald and T. Simon. Expanded intonation contours in mothers’ speech to newborns. *Developmental Psychology*, 20(1):104–113, 1984. ISSN 19390599. doi: 10.1037/0012-1649.20.1.104. 15, 16
- A. Fernald, T. Taeschner, J. Dunn, M. Papousek, B. De Boysson-Bardies, and I. Fukui. A cross-language study of prosodic modifications in mothers’ and fathers’ speech to preverbal infants. *Journal of Child Language*, 16(3):477–501, 1989. 16
- L. Fogassi, V. Gallese, L. Fadiga, and G. Rizzolatti. Neurons responding to the sight of goal-directed hand/arm actions in the parietal area PF (7b) of the macaque monkey. *Society of Neuroscience Abstracts*, 24(257.255), 1998. 11
- A. Fogel and A. Garvey. Alive communication. *Infant Behavior and Development*, 30(2):251–257, 2007. ISSN 0163-6383. 32
- M. A. Forrester. Conversation analysis: A reflexive methodology for critical psychology. *Annual Review of Critical Psychology*, 1:34–49, 1999. 18
- W. K. Frankenburg and J. B. Dodds. The Denver Developmental Screening Test\*. *The Journal of Pediatrics*, 71(2):181–191, 1967. 48
- V. Gallese, L. Fadiga, L. Fogassi, and G. Rizzolatti. Action recognition in the premotor cortex. *Brain*, 119(2):593–609, 1996. ISSN 00068950. doi: 10.1093/brain/119.2.593. 11
- H. Garfinkel. *Studies in Ethnomethodology*, volume 33 of *Biblioteka socjologiczna*. Prentice-Hall, 1967. ISBN 0745600050. doi: 10.2307/2092244. 18
- O. K. Garnica. Some prosodic and paralinguistic features of speech to young children. In *Talking to children*, pages 63–88. CUP, 1977. 15
- G. Gergely and G. Csibra. Teleological reasoning in infancy: the naive theory of rational action. *Trends in Cognitive Sciences*, 7(7):287–292, 2003. ISSN 13646613. doi: 10.1016/S1364-6613(03)00128-1. 12
- G. Gergely and J. S. Watson. Early Socio-Emotional Development : Contingency Perception and the Social-Biofeedback Model. *Early Social Cognition Understanding Others in the First Months of Life*, 34(4):101–136, 1997. doi: 10.1111/j.1365-2214.1992.tb00355.x. 27, 45
- G. Gergely, H. Bekkering, and I. Kiraly. Rational imitation in preverbal infants. *Nature*, 415(6873):755, 2002. 12, 14
- M. Gienger, H. Janssen, and C. Goerick. Task-oriented whole body motion for humanoid robots. In *IEEERAS International Conference on Humanoid Robots*, pages 238–244. IEEE, 2005. ISBN 0780393201. doi: 10.1109/ICHR.2005.1573574. 92

## REFERENCES

- M. Gienger, M. Mühlig, and J. J. Steil. Robot with automatic selection of task-specific representations for imitation learning, 2010. 13
- E. Gleitman, L., & Wanner. Richly specified input to language learning. *Adaptive control of ill-defined systems*, 1984. 16
- L. J. Gogate and L. E. Bahrick. Intersensory redundancy facilitates learning of arbitrary relations between vowel sounds and objects in seven-month-old infants. *Journal of Experimental Child Psychology*, 69(2):133–149, 1998. 17
- L. J. Gogate, L. E. Bahrick, and J. D. Watson. A study of multimodal motherese: the role of temporal synchrony between verbal labels and gestures. *Child Development*, 71(4):878–894, 2000. 17, 29
- B. A. Goldfield and J. S. Reznick. Early lexical acquisition: rate, content and the vocabulary spurt. *Journal of Child Language*, 17:171–183, 1990. 31
- H. Goodluck. *Language acquisition: a linguistic introduction*. Blackwell, 1991. ISBN 0631173854. 16
- C. Goodwin. The Interactive Construction of a Sentence in Natural Conversation. *Everyday language: Studies in ethnomethodology*, pages 97–121, 1979. 18, 48, 60
- G. Gredebäck, D. Stasiewicz, T. Falck-Ytter, C. Von Hofsten, and K. Rosander. Action type and goal type modulate goal-directed gaze shifts in 14-month-old infants. *Developmental Psychology*, 45(4):1190–1194, 2009. 74, 76
- A. Harrist and R. M. Waugh. Dyadic synchrony: Its structure and function in children’s development. *Developmental Review*, 22(4):555–592, 2002. ISSN 02732297. doi: 10.1016/S0273-2297(02)00500-2. 27
- G. Hayes and J. Demiris. A robot controller using learning by imitation. In *Proceedings of the 2nd International Symposium on Intelligent Robotic Systems*, pages 198–204. Citeseer, 1994. 13
- J. S. Herberg, M. M. Saylor, P. Ratanaswasd, D. T. Levin, and D. M. Wilkes. Audience-Contingent Variation in Action Demonstrations for Humans and Computers. *Cognitive Science*, 32(6):1003–1020, 2008. ISSN 1551-6709. 31, 32
- A. J. Ijspeert, J. Nakanishi, and S. Schaal. Movement imitation with nonlinear dynamical systems in humanoid robots. *Proceedings 2002 IEEE International Conference on Robotics and Automation Cat No02CH37292*, 2(May):1398–1403, 2002. doi: 10.1109/ROBOT.2002.1014739. 13
- L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1254–1259, 1998. ISSN 14631318. doi: 10.1109/34.730558. 34
- J. M. Iverson, O. Capirci, E. Longobardi, and M. Cristina Caselli. Gesturing in mother-child interactions\*. *Cognitive Development*, 14(1):57–75, 1999. ISSN 0885-2014. 16, 29
- M. Knoll and L. Scharrer. Acoustic and affective comparisons of natural and imaginary infant-, foreigner-and adult-directed speech. In *Eighth Annual Conference of the International Speech Communication Association*, 2007. 32
- E. A. Koterba and J. M. Iverson. Investigating motionese: The effect of infant-directed action on infants’ attention and object exploration. *Infant behavior development*, 32(4):437–444, 2009. 103
- S. Krach, F. Hegel, B. Wrede, G. Sagerer, F. Binkofski, and T. Kircher. Can machines think? Interaction and perspective taking with robots investigated via fMRI. *PLoS one*, 3(7):e2597, 2008. 32
- A. Lockerd and C. Breazeal. Tutelage and socially guided robot learning. *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems IROS IEEE Cat No04CH37566*, 4:3475–3480, 2004. doi: 10.1109/IROS.2004.1389954. 80
- K. S. Lohan, A. L. Vollmer, J. Fritsch, K. Rohlfing, and B. Wrede. Which ostensive stimuli can be used for a robot to detect and maintain tutoring situations? In *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*, pages 1–6. IEEE, 2009. 29, 42
- F. Lomker, S. Wrede, M. Hanheide, and J. Fritsch. Building modular vision systems with a graphical plugin environment. *Proceedings of the IEEE International Conference on Computer Vision Systems*, 2006. 18
- M. Lopes and J. Santos-Victor. Visual learning by imitation with motor representations. *IEEE transactions on systems man and cybernetics Part B Cybernetics a publication of the IEEE Systems Man and Cybernetics Society*, 35(3):438–449, 2005. 13
- B. D. Lucas, T. Kanade, and Others. An iterative image registration technique with an application to stereo vision. In *International joint conference on artificial intelligence*, volume 3, pages 674–679. Citeseer, 1981. 18
- P. Mangold. Getting better results in less time: When using audio/video recordings in research applications make sense. *3rd Congress of the European Society on Family Relations, Darmstadt, Germany*, 2006. 19, 36
- N. Masataka. Perception of motherese in a signed language by 6-month-old deaf infants. *Developmental Psychology*, 32(5):874–879, 1996. ISSN 00121649. doi: 10.1037/0012-1649.32.5.874. 16
- N. Masataka. *The onset of language*. Cambridge Univ Pr, 2003. ISBN 0521593964. 15, 16, 29
- A. N. Meltzoff and M. K. Moore. Newborn infants imitate adult facial gestures. *Child Development*, 54(3):702–709, 1983. 11
- L. Mondada. Participants’ online analysis and multimodal practices: projecting the end of the turn and the closing of the sequence. *Discourse Studies*, 8(1):117–129, 2006. ISSN 14614456. doi: 10.1177/1461445606059561. 18, 60
- M. Mühlig, M. Gienger, S. Hellbach, J. J. Steil, and C. Gericke. Task-level imitation learning using variance-based movement optimization. *Proceedings of the IEEE International Conference on Robotics and Automation (2009)*, pages 1177–1184, 2009. ISSN 10504729. doi: 10.1109/ROBOT.2009.5152439. 13, 14, 104
- M. Mühlig, M. Gienger, and J. J. Steil. Human-Robot Interaction for Learning and Adaptation of Object Movements. *Memory*, pages 4901–4907, 2010. 95

## REFERENCES

---

- Y. Nagai. How a robot's attention shapes the way people teach. In B. Johansson, E. Sahin, and C. Balkenius, editors, *Proceedings of the 10th International Conference on Epigenetic Robotics*, pages 81–88, 2010. 33
- Y. Nagai and K. J. Rohlfing. Can Motionese Tell Infants and Robots. What to imitate? In *Proceedings of the 4th International Symposium on Imitation in Animals and Artifacts*, pages 299–306. Citeseer, 2007. 9, 14, 29, 33, 34
- Y. Nagai, C. Muhl, and K. J. Rohlfing. Toward designing a robot that learns actions from parental demonstrations. In *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*, pages 3545–3550. IEEE, 2008. 14, 31, 45
- K. Nagell, R. S. Olguin, and M. Tomasello. Processes of social learning in the tool use of chimpanzees (*Pan troglodytes*) and human children (*Homo sapiens*). *Journal of Comparative Psychology*, 107(2):174–186, 1993. 11
- C. L. Nehaniv and K. Dautenhahn. Of hummingbirds and helicopters: An algebraic framework for interdisciplinary studies of imitation and its applications. *World Scientific Series in Robotics and Intelligent Systems*, 24:136–161, 2000. 8
- E. L. Newport. Motherese: The Speech of Mothers to Young Children, 1975. 15
- M. N. Nicolescu and M. J. Mataric. Experience-Based Representation Construction: Learning from Human and Robot Teachers. In *Proceedings of the IEEERSJ International Conference on Intelligent Robots and Systems*, volume 2, pages 740–745. IEEE, 2001. doi: 10.1109/IROS.2001.976257. 13
- M. N. Nicolescu and M. J. Matarić. Task Learning Through Imitation and Human-Robot Interaction. *Learning*, pages 407–424, 2005. 13, 80
- M. Ogino, A. Watanabe, and M. Asada. Mapping from Facial Expression to Internal State based on Intuitive Parenting. *Life*, 11:31–62, 2006. 34
- M. Pardowitz, S. Knoop, R. Dillmann, and R. D. Zöllner. Incremental learning of tasks from user demonstrations, past experiences, and vocal comments. *IEEE transactions on systems man and cybernetics Part B Cybernetics a publication of the IEEE Systems Man and Cybernetics Society*, 37(2):322–332, 2007. 13
- K. Pitsch. *Sprache , Körper , Intermediäre Objekte : Zur Multimodalität der Interaktion im bilingualen Geschichtsunterricht*. PhD thesis, Universität Bielefeld, 2006. 18
- K. Pitsch, A. L. Vollmer, J. Fritsch, B. Wrede, K. Rohlfing, and G. Sagerer. On the loop of action modification and the recipient's gaze in adult-child interaction. *Gesture and Speech in Interaction GESPIN*, 2009. 60
- K. Pitsch, A.-L. Vollmer, J. Fritsch, K. J. Rohlfing, and B. Wrede. Tutoring in Adult-Child-Interaction: On the Loop of Action Modification and the Recipients Gaze. *Interaction Studies*, 2011. 60, 64
- P. K. Pook and D. H. Ballard. Recognizing teleoperated manipulations. *1993 Proceedings IEEE International Conference on Robotics and Automation*, pages 578–585, 1993. doi: 10.1109/ROBOT.1993.291896. 13
- G. Rizzolatti and M. A. Arbib. Language within our grasp. *Trends in Neurosciences*, 21(5):188–194, 1998. 11
- G. Rizzolatti, L. Fadiga, V. Gallese, and L. Fogassi. Premotor cortex and the recognition of motor actions. *Brain Research*, 3(2):131–141, 1996. 11
- K. J. Rohlfing, J. Fritsch, B. Wrede, and T. Jungmann. How can multimodal cues from child-directed interaction reduce learning complexity in robots? *Advanced Robotics*, 20(10):1183–1199, 2006. ISSN 0169-1864. 16, 17, 22, 23, 29, 33, 36, 38, 39, 62
- H. Sacks. *Lectures on Conversation*, volume 2. Blackwell, 1992. ISBN 1557867054. 48
- H. Sacks, E. A. Schegloff, and G. Jefferson. A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4):696–735, 1974. ISSN 00978507. 67
- J. Saunders, C. L. Nehaniv, and K. Dautenhahn. Teaching robots by moulding behavior and scaffolding the environment. *Proceeding of the 1st ACM SIGCHISIGART conference on Humanrobot interaction HRI 06*, 2006:118–125, 2006. doi: 10.1145/1121241.1121263. 13
- S. Schaal. Is imitation learning the route to humanoid robots? *Trends in Cognitive Sciences*, 3(6):233–242, 1999. ISSN 1879307X. doi: 10.1016/S1364-6613(99)01327-3. 11, 12
- B. Shute and K. Whezldall. The incidence of raised average pitch and increased pitch variability in British'motherese'speech and the influence of maternal occupation and discourse form. *First Language*, 15(43):35–55, 1995. 16
- C. E. Snow. The development of conversation between mothers and babies. *Journal of Child Language*, 4(01):1–22, 1977. 16
- D. Stern. Mother and infant at play: The dyadic interaction involving facial, vocal, and gaze behaviors. In M. Lewis and L. Rosenblum, editors, *The Effect of the Infant on its Caregiver*, pages 187–213. Wiley-Interscience, 1974. 16
- D. N. Stern, S. Spieker, and K. MacKain. Intonation contours as signals in maternal speech to prelinguistic infants. *Developmental Psychology*, 18(5):727–735, 1982. ISSN 00121649. doi: 10.1037/0012-1649.18.5.727. 16
- C. Tennie, J. Call, and M. Tomasello. Push or Pull: Imitation vs. Emulation in Great Apes and Human Children. *Ethology*, 112(12):1159–1169, 2006. ISSN 01791613. doi: 10.1111/j.1439-0310.2006.01269.x. 9, 11, 14
- W. H. Thorpe. *Learning and Instinct in Animals*. Methuen & Co. Ltd, London, 1956. 9
- M. Tomasello. Cultural transmission in the tool use and communicatory signalling of chimpanzees. In S. Parker and K. Gibson, editors, *Language and Intelligence in Monkeys and Apes Comparative Developmental Perspectives*, pages 274–311. Cambridge University Press, 1990. 9
- M. Tomasello. The cultural origins of human cognition. *book*, 13(3):1–254, 1999. ISSN 10420533. doi: 10.1002/ajhb.1073. 7, 9

## REFERENCES

- M. Tomasello, M. Davis-Dasilva, L. Camak, and K. A. Bard. Observational learning of tool-use by young chimpanzees. *Human Evolution*, 2(2):175–183, 1987. ISSN 03939375. doi: 10.1007/BF02436405. 9
- M. Tomasello, A. C. Kruger, and H. H. Ratner. Cultural learning. *Behavioral and Brain Sciences*, 16(3):495–552, 1993. ISSN 0140525X. 9, 11
- M. Tomasello, M. Carpenter, J. Call, T. Behne, and H. Moll. Understanding and sharing intentions: the origins of cultural cognition. *Behavioral and Brain Sciences*, 28(5):675–691; discussion 691–735, 2005. 7, 11
- A. L. Vollmer, K. S. Lohan, K. Fischer, Y. Nagai, K. Pitsch, J. Fritsch, K. J. Rohlfing, and B. Wrede. People modify their tutoring behavior in robot-directed interaction for action learning. *Development and Learning (ICDL), 2009 IEEE 9th International Conference on*, 2009a. 16, 19, 27, 29, 32, 33, 49, 50, 51, 52, 55, 66
- A.-L. Vollmer, K. S. Lohan, J. Fritsch, B. Wrede, and K. J. Rohlfing. Which Motionese Parameters Change with Children’s Age? *The 2009 Cognitive Development Society’s Biennial Meeting*, 2009b. 29, 42, 47
- A. L. Vollmer, K. Pitsch, K. S. Lohan, J. Fritsch, K. J. Rohlfing, and B. Wrede. Developing feedback: how children of different age contribute to a tutoring interaction with adults. In *Development and Learning (ICDL), 2010 IEEE 9th International Conference on*, pages 76–81. IEEE, 2010. 18, 47, 50, 54, 56, 57, 74, 77
- J. S. Watson. Contingency perception in early social development. *Social perception in infants*, pages 157–176, 1985. 27
- A. Whiten and R. Ham. On the nature and evolution of imitation in the animal kingdom: Reappraisal of a century of research. *Advances in the Study of Behavior*, 21:239–283, 1992. ISSN 00653454. doi: 10.1016/S0065-3454(08)60146-1. 9
- B. Wrede, K. Rohlfing, M. Hanheide, and G. Sagerer. Towards learning by interacting. *Creating Brain-Like Intelligence*, pages 139–150, 2009. 29, 80
- B. Wrede, S. Kopp, K. Rohlfing, M. Lohse, and C. Muhl. Appropriate feedback in asymmetric interactions. *Journal of Pragmatics*, 42(9):2369–2384, 2010. ISSN 03782166. doi: 10.1016/j.pragma.2010.01.003. 47
- R. Zangl and D. L. Mills. Increased brain activity to infant-directed speech in 6- and 13-month-old infants. *Infancy*, 11(1):31–62, 2007. 16
- P. Zukow-Goldring. Sensitive Caregiving Fosters the Comprehension of Speech: When Gestures Speak Louder than Words. *Early Development and Parenting*, 5(4): 195–211, 1996. ISSN 10573593. doi: 10.1002/(SICI)1099-0917(199612)5:4(195::AID-EDP133)3.0.CO;2-H. 48
- P. Zukow-Goldring. A social ecological realist approach to the emergence of the lexicon: Educating attention to amodal invariants in gesture and speech. In C. Dent-Read and P. Zukow-Goldring, editors, *Evolving explanations of development Ecological approaches to organismenvironment systems*, pages 199–250. American Psychological Association, 1997. ISBN 1557983828. 59
- P. Zukow-Goldring and M. Arbib. Affordances, effectivities, and assisted imitation: Caregivers and the directing of attention. *Neurocomputing*, 70(13-15):2181–2193, 2007. ISSN 09252312. doi: 10.1016/j.neucom.2006.02.029. 59

## REFERENCES

---



## Appendix A

# Conventions for Transcription

**Table A.1:** Conventions for transcription used for manual annotations of video data.

	<b>Allgemeine Konventionen</b>	<b>general conventions of locations</b>
TS	vom Kind aus links	left from the child's point of view
VS	vom Kind aus rechts	right from the child's point of view
rH/lH	rechte Hand vom jeweiligen Akteur aus gesehen (linke Hand)	right hand/left hand of the actor
MH	Minihausen	the blocks on poles toy
T	Tasche	bag
B	Becher	cup
M	Mutter	mother
F	Vater	Father
C	Kind	Child
VL	Versuchsleiter	experimenter
DB	Klingel	bell
L	Lampe	lamp
S	Sitz	seat
t	Tisch	table
P	Platte	tablet
stempel	Stempel	stamp
c	Decke	ceiling
cam	Kamera	camera
HP	Home position	home position
H	Kopf	Head

Continued on next page

## A. CONVENTIONS FOR TRANSCRIPTION

Table A.1 – continued from previous page

	<b>Allgemeine Konventionen</b>	<b>general conventions of locations</b>
FrH/FIH	rechte/linke Hand des Vaters	father's right/left hand
MrH/MIH	rechte/linke Hand der Mutter	mother's right/left hand
FrA/FIA	rechter/linker Arm des Vaters	father's right/left arm
MrA/MIA	rechter/linker Arm der Mutter	mother's right/left arm
bot	Flasche	bottle
tafel	Tafel	chalk board
	<b>verbal (ver)</b>	
<lacht>	GAT-Konventionen lachen	laugh
<Laut>	laut	loud
<räuspern>	räuspern	cough
<XXX>	nicht verstanden	not understood
	<b>Blickrichtung (gaz)</b>	<b>gaze direction</b>
@X	Blick auf etwas gerichtet	gaze fixed on something
@CrH+Kx	Blick auf rechte Hand des Kindes mit Klotz x	gaze to the child's right hand with the block
@sx	Blick auf Säule x	gaze to pole x
@Bx/By	Blick auf Bx oder auf By (erstes Priorität)	gaze to Bx or By (first direction more likely)
@Bx/By/Bz	Blick geht auf Bx oder By oder Bz (nach Prioritäten)	gaze to Bx, By, or Bz (likelihood decreases from first to last)
@Bx/C	Blick geht zu Bx oder Kind (nach Prioritäten)	gaze to Bx or child (first direction more likely)
@XXX	Blickrichtung nicht erkennbar (aber sichtbar)	gaze direction not recognizable (but visible)
	<b>Blickbewegung (gaz)</b>	<b>gaze movement</b>
~X	Blick bewegt sich	shifting gaze
~down	Blick geht runter	gaze moves down
~up	Blick geht rauf	gaze moves up
0	Blick ins off	gaze to off
~back	nach hinten schauen	gaze behind
Continued on next page		

**Table A.1 – continued from previous page**

	<b>Allgemeine Konventionen</b>	<b>general conventions of locations</b>
~Bx/By	Blick geht zu Bx oder By (erstes Priorität)	gaze moves to Bx or By (first direction more likely)
~Bx/By/Bz	Blick geht zu Bx oder By oder Bz (nach Prioritäten)	gaze moves to Bx, By, or Bz (likelihood decreases from first to last)
~Bx/C	Blick geht zu Bx oder Kind (nach Prioritäten)	gaze moves to Bx or child (first direction more likely)
~XXX	Blickrichtung nicht erkennbar (aber sichtbar)	gaze direction not recognizable (but visible)
	<b>Mimik (fac)</b>	<b>facial expression</b>
smile	lächeln	
laughing	lachen	
konzentrierter Blick		concentrated look
entspannter Blick		relaxed look
interessierter Blick		interested look
desinteressierter Blick		uninterested look
unzufriedener Blick		discontent look
staunender Blick		astonished look
überraschter Blick		surprised look
Lippen zusammenpressen		pressing lips together
Lippen spitzen		puckered lips
stirnrunzeln		frown
Augenbrauen hochziehen		raising eyebrows
Zunge rausstrecken		sticking tong out
fragender Blick		questioning look
angestregter Blick		stressed look
gelangweilter Blick		bored look
zwinkern		wink
ernster Blick		stern look
Gesicht zusammenkneifen	squinted face	

Continued on next page

## A. CONVENTIONS FOR TRANSCRIPTION

Table A.1 – continued from previous page

	<b>Allgemeine Konventionen</b>	<b>general conventions of locations</b>
	<b>Hand- /Armbewegungen (act)</b>	<b>hand /arm movements</b>
~	Handbewegung	moving hand
HP	home position	home position
P	Hand an der Platte	hand at tablet
X	Hand an Objekt X	hand at object X
holding pos	Haltestellung der Hand in der Luft	hand holds still in the air
~up	Hand geht rauf	hand moves up
~down	Hand geht runter	hand moves down
~holding pos	Bewegung in eine Haltestellung der Hand (in der Luft)	moving the hand to a position, where hand is held still in the air
~HP	Bewegung in die home position	movement to home position
~H	Bewegung zum Kopf	movement to head
~t	Bewegung zum Tisch	movement to table
~P	Bewegung zur Platte	movement to tablet
~Bx	Bewegung zu Becher x	movement to cup x
~S	Bewegung zum Sitz/Schoß	movement to lap
X>Y	X bewegt sich zu Y	X is moving to Y
lH~H	linke Hand bewegt sich zum Kopf	left hand is moving to head
rH>lH	rechte Hand zur linken Hand	right hand to left hand
P>X	Platte zum Interaktionspartner (X=C,F,M,VL)	tablet to interaction partner (X=C,F,M,VL)
clap	klatschen	clap
grab X	etwas greifen	grab something
lift X	etwas heben	lift something
hold X	etwas in der Luft halten	hold something
place X	etwas plazieren	place something
drop X	etwas fallenlassen	drop something
X~t	etwas wird auf dem Tisch verschoben	something is moved on the table
clench F	Faust ballen	clench fist
-clench F	Faust öffnen	open fist
move t	Tischdecke verschieben	move the tablecloth
Continued on next page		

Table A.1 – continued from previous page

	<b>Allgemeine Konventionen</b>	<b>general conventions of locations</b>
rotate X shake X ~body ~T-Shirt x>down ~Ball Bx~t  Bx>down  Bx+Bz/~Bx+Bz	etwas drehen etwas schütteln Hände am Körper am T-Shirt ziehen Objekt runter schmeißen Bewegung zum Ball Becher x wird auf dem Tisch verschoben Becher x wird runter geschmissen Hand liegt auf (bewegt sich zu) zwei Bechern gleichzeitig  <b>Gesten (in der Zeile act annotiert)</b>	rotate something shake something hands to body pull shirt throw object down movement to ball cup x is moved on table  cup x is thrown down  hand rests on (moves towards) two cups at the same time  <b>gestures</b>
prep peak retr	<b>Körperpositur (pos)</b>	<b>body posture</b>
zurück, vor, zurück-TS, zurück-VS, frontal, vor-frontal, vor-TS, vor-VS, runter, vor-runter, zurück-runter, hoch, vor-hoch, zurück-hoch, umdrehen, aufstehen	<b>Mot Becher</b>	<b>nesting cups specific</b>
Bb By Bg Br	Becher blau Becher gelb Becher grün Becher rot	blue cup yellow cup green cup red cup
	<b>Mot Minihausen</b>	<b>blocks on pole specific</b>
Kbk	Klotz blau klein	small blue block
Continued on next page		

## A. CONVENTIONS FOR TRANSCRIPTION

---

Table A.1 – continued from previous page

	<b>Allgemeine Konventionen</b>	<b>general conventions of locations</b>
Kyk	Klotz gelb klein	small yellow block
Kgd	Klotz grün dunkel	dark green block
Ko	Klotz beige	beige block
Kr1	Klotz rot 1	red block 1
Kr2	Klotz rot 2	red block 2
s1	von F/M aus gesehen die linke Seite	left side from F/M's point of view
s2	mittlere Säule	pole in the middle
s3	rechte Säule	right pole
Kx>sy	Klotz x zur Säule y	block x on pole y

## Appendix B

# Table Overviewing the Results of the Analysis of Tutoring Behavior

Compared to AAI, ACI shows	Compared to ACI, ARI shows	Compared to AAI, ARI shows
slower hand movement	slower hand movement	slower hand movement
lower hand movement acceleration	lower hand movement acceleration	lower hand movement acceleration
smaller pace		smaller pace
less round movement		less round movement
greater range and therewith more exaggerated movement	greater range and therewith more exaggerated movement in the first sub-action	greater range and therewith more exaggerated movement
greater total length of motion pauses	greater total length of motion pauses	greater total length of motion pauses
higher frequency of motion pauses		higher frequency of motion pauses
greater average length of motion pauses	greater average length of motion pauses	greater average length of motion pauses
longer action	longer action	longer action
more time spent gazing at the learner	less time spent gazing at the learner	
more frequent eye-gaze bouts to the learner	less frequent eye-gaze bouts to the learner	
on average longer eye-gaze bouts to the learner	on average shorter eye-gaze bouts to the learner	on average shorter eye-gaze bouts to the learner
less time spent gazing at the object	more time spent gazing at the object	
	lower frequency of eye-gaze bouts to object	lower frequency of eye-gaze bouts to object
	greater average length of eye-gaze bout to object	


**Table B.1:** This table shows a short summary of the results of Section 4.2.




## Appendix C

# Questionnaire for Human-Robot Interaction Study

## C. QUESTIONNAIRE FOR HUMAN-ROBOT INTERACTION STUDY

 **Universität Bielefeld** Research Institute for Cognition and Robotics – CoR-Lab

 **CORE Lab**

**Fragebogen Versuch IS-1**

1	2
---	---

**Code**

1. Wie alt sind Sie? (How old are you?) \_\_\_\_\_ Jahre (years)

2. Sie sind (You are):  männlich (male)  weiblich (female)

3. Ist Deutsch Ihre einzige Muttersprache? (Is German your only first language?)

ja (yes)  nein (no)

Wenn nein, welche ist / sind Ihre weitere(n) Muttersprache(n)?  
(If no, which is/are your further first language(s)?)

\_\_\_\_\_

4. Welcher Tätigkeit gehen Sie derzeit nach?  
(What is your current occupation?)

Student (student)  Arbeitnehmer (employee)  selbstständig (self-employed)

Sonstiges (other): \_\_\_\_\_

Wenn Sie Student sind, seit wie vielen Semestern sind Sie bereits eingeschrieben?  
(If you are a student, for how many semesters already?)

\_\_\_\_\_ Semester (semesters)

In welchem Bereich studieren/arbeiten Sie?  
(What is your area of study/work?)

\_\_\_\_\_

5. Welchen Abschluss haben Sie? (Which is your highest degree?)

Hauptschulabschluss	<input type="radio"/>	Diplom (FH)	<input type="radio"/>
Mittlere Reife	<input type="radio"/>	Magister	<input type="radio"/>
Fachhochschulreife	<input type="radio"/>	Diplom	<input type="radio"/>
Abitur	<input type="radio"/>	Master	<input type="radio"/>
Bachelor	<input type="radio"/>	Doktor	<input type="radio"/>

1

**Figure C.1: Questionnaire** - Questionnaire form for the human-robot interaction study presented in 7.

6. Wie viel Erfahrung haben Sie im Umgang mit Computern?  
(What is your experience with computers?)

<p><b>Ich habe gar keine Erfahrung mit Computern.</b> (no experience)</p>	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	<p><b>Ich habe sehr viel Erfahrung mit Computern.</b> (very much experience)</p>
---	---	--

7. Kreuzen Sie bitte die Roboter an, die Sie kennen!  
(Please check the robots you know!)

- |   |  |
|---|--|
| <input type="radio"/> Aibo                                | <input type="radio"/> Serviceroboter (service robot)     |
| <input type="radio"/> Nao                                 | <input type="radio"/> Marserkundungsroboter (Mars rover) |
| <input type="radio"/> Kismet                              | <input type="radio"/> ASIMO                              |
| <input type="radio"/> iCub                                | <input type="radio"/> BIRON                              |
| <input type="radio"/> BARTHOC                             | <input type="radio"/> Paro                               |
| <input type="radio"/> Fussballroboter (soccer robot)      | <input type="radio"/> R2D2                               |
| <input type="radio"/> Lego Mindstorms                     | <input type="radio"/> Roomba (vacuum cleaning robot)     |
| <input type="radio"/> Industrieroboter (industrial robot) | <input type="radio"/> Wall-E                             |
| <input type="radio"/> Pleo                                |  |
- Sonstige (other): \_\_\_\_\_

8. Wie viel Erfahrung haben Sie im Umgang mit Robotern wie den eben genannten?  
(How much experience do you have with robots like the ones mentioned above)

<p><b>Ich habe gar keine Erfahrung mit Robotern.</b> (no experience)</p>	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	<p><b>Ich habe sehr viel Erfahrung mit Robotern.</b> (very much experience)</p>
--	---	---

9. Inwieweit treffen die folgenden Aussagen auf Sie zu?  
In how far do the following statements apply to you?  
*Bitte kreuzen Sie auf der Skala die Antwort an, die am ehesten Ihrer Einschätzung entspricht!*  
*Bitte in jeder Zeile ein Kästchen ankreuzen!*  
*(Please check the answer on the scale, which is closest to your estimation!*  
*Please only tick one box in each row!)*

Figure C.1: Questionnaire continued - Questionnaire form for the human-robot interaction study presented in 7.

## C. QUESTIONNAIRE FOR HUMAN-ROBOT INTERACTION STUDY

Ich (I)...	trifft überhaupt nicht zu (does not apply at all)	trifft eher nicht zu (does rather not apply)	weder noch (neither)	eher zutreffend (rather applies)	trifft voll und ganz zu (fully applies)
...bin eher zurückhaltend, reserviert. (am rather quiet, reserved.)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...schenke anderen leicht Vertrauen, glaube an das Gute im Menschen. (easily trust people, believe in the good in man.)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...bin bequem, neige zur Faulheit. (am easygoing, tend toward laziness.)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...bin entspannt, lasse mich durch Stress nicht aus der Ruhe bringen. (am relaxed, do not get stressed out easily.)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...habe nur wenig künstlerisches Interesse. (do not have much interest in art.)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...gehe aus mir heraus, bin gesellig. (feel comfortable around people.)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...neige dazu, andere zu kritisieren. (tend to criticize others.)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...erledige Aufgaben gründlich. (do chores thoroughly.)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...werde leicht nervös und unsicher. (easily get nervous and insecure)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...habe eine aktive Vorstellungskraft, bin phantasievoll. (possess an active imagination, am imaginative)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**Figure C.1: Questionnaire continued** - Questionnaire form for the human-robot interaction study presented in 7.

## Appendix D

# Interview for Human-Robot Interaction Study

## D. INTERVIEW FOR HUMAN-ROBOT INTERACTION STUDY

---

Versuch IS-1  
[www.cor-lab.de](http://www.cor-lab.de)

1	2
---	---

Code

### Interview – Guided questions

Sie haben dem Roboter einige Objekte gezeigt und wie man sie bewegt/benutzt. Ich würde gerne über Ihre Eindrücke und Erfahrungen im Umgang mit dem Roboter sprechen.  
(You showed several objects to the robot and how to move/use them. I would like to talk with you about your impressions and experiences with the robot.)

- 1. Wenn Sie über Ihre speziellen Eindrücke nachdenken – wie war die Interaktion mit dem Roboter?**  
(When you think about your particular impressions – how was the interaction with the robot?)
- 2. Was ist Ihnen besonders aufgefallen? positiv oder negativ? Gab es besondere Situationen?**  
(What did you notice? positively or negatively? Were there any special situations?)
- 3. Worauf haben Sie bei der Interaktion geachtet? Hatten Sie eine Strategie?**  
(What did you pay attention to during the interaction? Did you have a strategy?)
- 4. Wohin hat der Roboter geguckt?**  
(Where did the robot look?)

**Figure D.1:** Interview - Guided interview form for the human-robot interaction study presented in 7.

---

Versuch IS-1  
[www.cor-lab.de](http://www.cor-lab.de)

1	2
---	---

Code

5. **Was hat er verstanden, von dem, was Sie ihm vorgemacht haben?**  
(What did it understand of what you demonstrated to it?)
  
6. **Hat es lang gedauert bis er verstanden hat, was Sie ihm zeigen wollten?**  
(Did it take long until it understood what you wanted to show it?)
  
7. **Was von dem, was Sie ihm vorgemacht haben, hat der Roboter nachgemacht?**  
(What of your demonstrations did the robot reproduce?)
  
8. **Haben Sie evtl. weitere Anmerkungen oder Kommentare zur Studie oder dem Roboter?**  
(Do you maybe have any other remarks or comments about the study or the robot?)
  
9. **Hat der Versuch Spaß gemacht?**  
(Did you enjoy the study?)

**Figure D.1: Interview continued** - Guided interview form for the human-robot interaction study presented in 7.