

Alignment to the Actions of a Robot

Anna-Lisa Vollmer, Katharina J. Rohlfing, Britta Wrede, Angelo Cangelosi

▶ **To cite this version:**

Anna-Lisa Vollmer, Katharina J. Rohlfing, Britta Wrede, Angelo Cangelosi. Alignment to the Actions of a Robot. International Journal of Social Robotics, 2015, 10.1007/s12369-014-0252-0 . hal-01249226

HAL Id: hal-01249226
https://inria.hal.science/hal-01249226

Submitted on 30 Dec 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

International Journal of Social Robotics manuscript No.
(will be inserted by the editor)

Alignment to the actions of a robot

Anna-Lisa Vollmer · Katharina J. Rohlfing · Britta Wrede · Angelo Cangelosi

Received: date / Accepted: date

Abstract Alignment is a phenomenon observed in human conversation: Dialog partners' behavior converges in many respects. Such alignment has been proposed to be automatic and the basis for communicating successfully. Recent research on human-computer dialog promotes a mediated communicative design account of alignment according to which the extent of alignment is influenced by interlocutors' beliefs about each other. Our work aims at adding to these findings in two ways. a) Our work investigates alignment of manual actions, instead of lexical choice. b) Participants interact with the iCub humanoid robot, instead of an artificial computer dialog system. Our results confirm that alignment also takes place in the domain of actions. We were not able to replicate the results of the original study in general in this setting, but in accordance with its findings, participants with a high questionnaire score for emotional stability and participants who are familiar with robots align their actions more to a robot they believe to be basic than to one they believe to be advanced. Regarding alignment over the course of an interaction, the extent of alignment seems to remain constant, when participants believe the robot to be advanced, but it increases over time, when participants believe the robot to be a basic version.

Keywords Alignment · Human-Robot Interaction · Action Understanding · Robot Social Learning

A.-L. Vollmer · A. Cangelosi
Centre for Robotics and Neural Systems, Plymouth University, Plymouth, PL4 8AA, UK
Tel.: +44-1752-584908
E-mail: anna-lisa.vollmer@plymouth.ac.uk

K. J. Rohlfing · B. Wrede
CITEC Center of Excellence Cognitive Interaction Technology, Bielefeld University, 33615 Bielefeld, Germany

1 Introduction

A long-term research goal in the field of social robotics is to develop robots that are capable of interacting and learning in natural social interaction with a human user (while not relying on pre-programmed interaction patterns). To achieve this goal, we subscribe to the view that *alignment* is beneficial in human communication [26] and transfer it to human-robot interaction.
Human interlocutors in a dialog adapt and converge on many aspects of linguistic behavior. They for example tend to choose the same words, adapt their speech rates and use the same syntactic structures. Such convergence or *alignment* phenomena have been found not only in spoken dialog, but also in written dialog, and non-verbal actions (cf. [9, 16, 19], 2.1). Alignment has been proposed to be the basis for communicative success [25]. In their *interactive alignment model*, Pickering and Garrod proposed that alignment processes are automatic and can be explained by mutual priming of interlocutors' internal representations [26]. Opposed to Pickering and Garrod's account of alignment, a mediated *communicative design* account has been proposed according to which interlocutors' extent of alignment is mediated by their judgment and beliefs about each other [4, 10, 23]: When interacting with a computer system, participants aligned their word choice more to a computer they believed to be basic than to one they believed to be advanced, irrespective of the feedback it gave. In a remote human-robot interaction study, we test the latter account for the alignment of actions. In a remote human-robot interaction game, participants take turns with the robot to demonstrate actions and match the partner's demonstrations to textual action descriptions. We examine to which extent participants who believe to be playing with a basic iCub version or

an advanced iCub version align the way they demonstrate actions to the way their robot partner has previously demonstrated them.
In the next section (Section 2), we will give an introduction to the topic of alignment, present the motivation for this work and describe the study serving as a model for the conducted experiment and analyses. We lay particular focus on the extensive experimental design (Section 3), as we put much effort in replicating the original study as exact as possible. In Sections 4 and 5, we present the results of two analyses. The analysis presented in Section 4 is based on a conference paper by Vollmer et al. [33]. It focuses on the questions "Do untrained users align their actions to a robot in order to successfully communicate?" and "Do they adapt the way they demonstrate actions to the robots capabilities and understanding?". A novel analysis on time presented in Section 5 is concerned with the question "Does alignment change over the time-course of an interaction?". We end with a conclusion in Section 7.

2 Background

2.1 Alignment

Alignment phenomena have been labeled with a great number of different terms (e.g. coordination, convergence, congruence, matching, mimicry) whose definitions are partly fuzzy and overlapping. The terms congruence, matching, and mimicry, however, seem to mostly refer to the alignment of non-verbal behavior [8, 20]. In this work, we use the word *alignment* to describe all relevant phenomena irrespective of the modality they occur in.

2.1.1 Verbal Behavior

In dialog, interlocutors align for example accents, speech rates, phonetic realizations, syntactic structure, speech rhythms, and choice of words. Alignment of choice of words is also known as lexical entrainment [14].

> For instance, when two people (interlocutor A and interlocutor B) have a conversation and interlocutor A is pointing out B's new "jeans", her partner, interlocutor B, will more likely use the same term, "jeans", to again refer to this garment, instead of for example using the word "trousers" or "pants" to refer to it.

Lexical entrainment has been found to be partner-specific [6].

> In our previous example, thus, if a third person (interlocutor C) would join the conversation, A and B would not have the same *conceptual pact* with C as they have not grounded this particular reference with C. Therefore, they might propose the same term "jeans" when addressing C, but they will accommodate to the new addressee.

Non-verbal Behavior

With regard to non-verbal behavior, interlocutors align their gestures, co-speech gestures, posture, timing of posture changes, body movements (e.g. foot shaking), mannerism, and facial expressions. For example, if one interlocutor frequently nods her head during the course of the conversation, her partner will be likely to also nod her head more frequently. Research revealed correspondence of gestures in handshape, motion, location, representation, technique, handedness, and palm orientation [2, 18].

2.2 Motivation - Why Study Alignment in Human-Robot Interaction?

There are two directions to consider, 1) the alignment from robot to human and 2) the direction from human to robot.

1. Considering the alignment from robot to human would presumably yield the same benefits for the robot as for the aligning interlocutor in human-human interaction. An interlocutor aligning with her interaction partner results in positive affect towards the aligning party (cf. [3]). For instance, van Baaren et al. revealed that customers who were verbally mimicked by their waitresses (the waitresses repeated the customers order to them exactly) left larger tips [30]. Therefore, recently, researchers have aimed at endowing robots, artificial agents, and computers with the capability to align to their human interlocutors (linguistic alignment [7], emotional alignment [12], alignment of gestures [19]).
2. So far, related work investigating the other direction, namely how humans align to robots, has revealed alignment in speech [17] and qualitatively in gestures ([5], children adapting their behavior to a robot learner's behavior [22]). Alignment toward computers or artificial dialog systems has been identified as similar and even stronger alignment than toward human interlocutors [4, 15, 29].

How could a robot benefit from a human user aligning to it (2)?
Human alignment in human-robot *dialog* clearly appears to be beneficial.

> Consider an example of alignment of choice of words in a conversation with a robotic or artificial interlocutor that has a limited vocabulary: In an office building, the robot asks about the directions to a certain office, "Do I have to continue down this corridor and then turn left?" The human interlocutor now could answer: "No, you need to take the hallway to your right and then turn left." The robot does not know the word 'hallway' or what it means, so it has to ask a follow-up question and communicative success is delayed. Or the human interlocutor aligns to the robot and uses the same word the robot used: "No, you need to take the corridor to your right and then turn left." In this case, the robot knows and understands the word 'corridor' and communicative success is reached right away.

Similarly, the alignment of *actions* in human-robot interaction could increase the smoothness of cooperative interaction of a robot with a human user. It might also facilitate robot social learning of actions from a tutor's demonstration, especially when the user/tutor is naive to the internal functioning and learning mechanisms of the robot. Aligning to the actions of a robot would entail picking up what is already known and understood by the robot. To give a specific example, demonstrating an action "like the robot" by for instance restricting one's movements to a certain range (if the possible range of motion of the robot is limited) or grasping in a certain manner (for example with only two fingers, if the robot is equipped with only two finger-like manipulators on its end effector or hand) could facilitate mapping of the user's movements onto the body of the robot despite potential differences in human and robot embodiment (known as the correspondence problem). Thus, alignment might make understanding and learning easier by naturally lowering the complexity of communication.

2.3 The Mediated Communicative Design Account of Alignment

To investigate the influence of beliefs on people's alignment of lexical choice toward a computer, Pearson et al. conducted a study in which participants played a picture labeling and matching game with a computer (i.e. an artificial dialog system) [4, 23]. This section gives a description of the original study by Pearson et al. At the beginning of the study, the start-up screen of the system was displayed. It suggested the system to be either basic (i.e. old-fashioned and unsophisticated) or advanced (i.e. new and sophisticated). In the game, two pictures were presented side-by-side. One of them received a textual label from the system which was displayed to the participant whose task then was to match this label to the one of the two pictures it corresponded to. Next, two new pictures appeared and the participant had to label one of them via keyboard. The system then apparently matched this label to the respective picture, but the responses of the system were actually pre-programmed. Pictures which had to be labeled in experimental trials (experimental pictures) had two different labels of which both were acceptable for what was on the picture, but one was common and the other extremely uncommon. The system systematically gave the uncommon label for these pictures and at a later trial, the participants had to label the same picture again themselves. Pearson et al.'s findings indicate that the participants repeated the uncommon label more often (i.e. aligned their lexical choices more) when they believed to be playing with a basic computer than when they believed to be playing with an advanced computer, even though the performance of the system, i.e. the labeling and matching by the system, was equally good in both cases. This is in line with previous findings suggesting that people modify their behavior according to their interaction partners' understanding and needs (in human-robot interaction and adult-child interaction [13, 31, 32]).
Our work aims at adding to these findings in two ways. a) It investigates alignment of manual actions, instead of lexical choice. b) The participants interact with the iCub humanoid robot, instead of an artificial computer dialog system. Accordingly, in our work, the participants played an action demonstration and matching game, instead of a picture naming and matching game. How our game was designed and how it relates to the game in the original experiment is described in the next section.

3 Methods

The action demonstration and matching game in our experiment was designed based on Pearson et al.'s picture naming and matching game [23], see Section 2.3. All design choices of the study presented in this work are made to be equivalent to the design of the original study. A picture in their setup here corresponds to a text box containing the description of an action. Equivalently, the naming of a picture thus corresponds to the demonstration of the respective action described in the text box. As equivalent to the original setup which was a turn-based game on a computer, we chose to tell the participants that they would play a game with a basic or advanced robot in a different room over a network connection. The study investigates whether the

participants adapt their action demonstrations to the way the robot demonstrates the same action. Thus, the action demonstrations of the robot ideally should not introduce any variability in the interaction. Therefore, because some of the experimental actions were due to technical limitations difficult for the robot to perform (e.g. grasping and nesting of differently sized cups) and because the setup in the original study involved indirect communication, the responses of the robot were really pre-recorded videos of action demonstrations and the choice made by the robot for matching turns (i.e. the feedback of the robot) was scripted to choose always the correct answer in both conditions. On experimental trials, the robot demonstrated an action that had one common way and one uncommon way of execution. After two filler trials, the participants were asked to demonstrate the same action the robot had previously shown to them. We were interested in whether the participants would demonstrate the action in the same way the robot had demonstrated it in (i.e. whether they would align). Whereas the evaluation of Pearson et al.'s study by checking if two words were equal was straightforward, we had to compare the action demonstration of a robot with the action demonstration of a human participant. Laying importance on human perception, in our study, we obtained a score for alignment with five raters who individually and independently rated the similarity of the two action demonstrations shown in the videos on a scale from 1 [very similar] to 7 [not similar at all].

Analogously to Pearson et al.'s work, with beliefs about an interaction partner's capabilities influencing the participants' behavior, we expected alignment stronger in the basic robot condition than in the advanced robot condition.

In all pretests and the experiment, the participants were English native speakers to avoid language based differences in action understanding and executions, and members of the community of Plymouth University. All the participants responded to advertisement on campus. No participant took part in more than one pretest or experiment. The experiment used 31 participants (13 f, 18 m) aged 18–50 years ($M = 23.45, SD = 7.74$) who were paid to participate.

The robot platform with which the experiment was carried out was the iCub humanoid robot [21]. For the experiment, only the upper body of the robot was used (head, eyes, torso, arms and hands).

In the following, we will describe in detail how we designed and obtained counterparts of individual aspects of Pearson et al.'s original study.

3.1 Experimental Actions

The experimental pictures in the original experiment by Pearson et al. consisted of pictures with two different labels of which one was common and the other extremely uncommon (cf. Section 2.3). Pearson et al. obtained the pictures and their different labels from an existing database [28]. They conducted two pretests to find those pictures with two acceptable labels of which one was common and the other uncommon.

Analogously, our experiment required experimental actions with two different acceptable ways of execution of which one was common and the other uncommon. Thus, before the main experiment, we conducted three separate pretests to build a dataset of actions and their executions and to find appropriate experimental items. These are presented in the following section. Pretest 1 revealed actions with more than one way of execution that were compatible with the setup of the experiment. In pretest 2, we singled out those actions obtained from pretest 1, which had two equally acceptable ways of execution. In pretest 3, these actions were tested to determine which way of an execution of an action the participants would favor. From the pretests resulted the actions on which the experimental items were based. The iCub robot as well as the setup imposed constraints on the actions in the experiment. As the experiment was set in a sitting position at a table (see Section 3.3.1), the actions should only involve movement of the upper body which could be carried out on a table top in a sitting position. Additionally, the hardware of the robot only permitted actions with light objects, such as soft toys, and a low degree of finicky accuracy.

3.1.1 Pretests

1. Sitting at a table, 24 participants (9 f, 15 m) were asked to perform 34 actions (e.g. how to stamp). They were instructed to give only one unambiguous demonstration for each action description and perform the first one that came to their minds. The action demonstrations were video recorded. Of the actions which were demonstrated by the participants, we chose those actions which appeared to have more than one type of execution (24 actions) for the next pretest.
2. Another twelve participants (6 f, 6 m) were presented with video clips of an actor acting out the ways of execution for the 24 actions obtained from pretest 1 and rated the acceptability of each demonstration for the corresponding action description on a scale from 1 [completely unacceptable] to 7 [completely acceptable]. We then selected those actions

which had two different types of execution both with a rating of more than 5 (M=6.18, SD=0.61). We selected 16 actions.

3. In a forced-choice task, 14 participants (7 f, 7 m) named which one of the two demonstrations for each action they would perform. For ten actions one demonstration was favored by more than 80% of the participants (M=92.86%, SD=5.83%). Nine of these ten actions constitute the experimental items for the experiment. One item was excluded from the experiment because the robot was not able to perform either way of execution of the action as the objects involved were too small for grasping and the task was too intricate. Summarizing, the actions resulting from the pretests can be demonstrated – also by the robot – in two different ways which are equally acceptable for the task, but of which only one is common and favored and the other one is uncommon and unfavored (for an overview, see Table 1).

Experimental actions	common way	uncommon way
"how to ring a bell"	with the flat hand	with the thumb
"how to brush (the table)"	long strokes to the side	circular motion
"how to nest cups into each other"	smaller cup into next bigger one	biggest cup into goal cup, smallest into next bigger, then into goal cup
"how to roll dice"	shaking before rolling	direct rolling
"how to pet a dog"	long strokes along back	several pats on head
"how to cut fruit"	one cut	sawing movement
"how to point to the die"	straight	like a pistol
"how to use a salt-shaker"	several shakes	tapping with the index finger
"how to swim"	breast stroke	butterfly stroke

Table 1 Experimental actions with common and uncommon way of execution

3.2 Items

With the nine experimental items obtained from the three pretests described in Section 3.1.1, we constructed nine *experimental items* that consisted of the descriptions of a *prime action* and a *distractor action* in text boxes, together with a common and an uncommon way of execution for the prime action, and the descriptions of a *target action* (identical to the prime action) and another distractor action in text boxes. As most actions involved an object, the prime/target action and distractor actions had to involve the same object because the actions should not be recognized by object only.

For example, one item comprised

- the descriptions of a *prime action* and a *distractor action* in text boxes
  - the prime action "how to pet a dog"
  - with the distractor action "how to give a dog treats"
- a common and an uncommon way of execution for the prime action
  - the common way of execution: to pet the dog by doing several long strokes across its back
  - the uncommon way: giving the dog a couple of pats on its head
- the descriptions of a *target action* (identical to the prime action) and another distractor action in text boxes
  - the target action "how to pet a dog"
  - the distractor action "how to pull a dog's tail".

Actions that did not involve an object (e.g. "how to wave to someone") were paired with another action that did not involve any objects either.
Additionally, equivalently to Pearson et al.'s experiment, we constructed 22 pairs of *filler actions*. Thus, one action in each pair was executed, the other was a distractor action. Each of the executed actions, as for example "how to nod your head", "how to grasp an apple", "how to wave to someone", etc., was paired with a distractor action, again involving the same (or no) object just like the filler action with which it was paired. A *filler item* consisted of two different, independent pairs of filler actions in text form. One action in each of eleven of the 22 pairs of filler actions was demonstrated by the robot and equivalently one action in each of the other eleven pairs was demonstrated by each human participant.

3.2.1 Robot action demonstration

The experiment required videos of the iCub robot carrying out the nine actions for the nine experimental items in two different ways (one common, one uncommon) and the eleven actions of the filler items in a preferably natural manner. For this, we applied slow kinesthetic teaching and replaying of the movements taught using the Aquila framework [24]. This method was not suitable for actions involving large movements of both arms or large shoulder movements. For these actions, joint angles of separate discrete positions were recorded and then interpolated for subsequent replay.

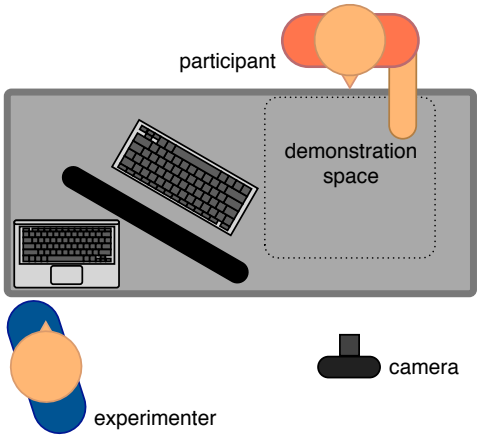


Fig. 1 Experimental setup

Additionally, the robot exhibited a simple gaze behavior: It gazed in direction of the camera at the beginning of each video and action execution, then it gazed down to the object or hand carrying out the action and at the end of the execution, it gazed back to the camera. A front view of the actions of the robot was video-recorded and the video clips then sped up partially to compensate for artifacts from the methods used and achieve more natural robot action executions which are more similar to the human action executions observed in pretest 1.

3.3 Procedure

3.3.1 Setup and Course of the Experiment

The participants sat at a table with a monitor on the right in front of them and a camera video-recording their action demonstrations from the front (see Fig. 1). The experimenter was seated behind the monitor so that her face was hidden from the participant. She had a laptop to observe the course of the experiment and to know which objects to get. The participant monitor showed a purpose-built Skype-like application with a big window in which the video of the robot was shown and a smaller window in which the participant's video was shown. The participant window overlaid the lower right corner of the bigger robot window (see screenshots in Fig. 2) so that no parts of the actions of the robot were occluded. While the windows did not show an action demonstration, they displayed "no video input" to give the impression that the network video connection was running even though no video was shown. Below the windows two text boxes were displayed side by side which contained a pair of action descriptions as part of an experimental or a filler item. The participant and the robot took turns in demonstrating an action and matching it to one of two action descriptions in text boxes. The videos of the robot and the participant were only displayed during the action demonstrations and thus robot and participant videos were never shown at the same time. The robot began with an action demonstration turn which was followed by a participant matching turn (see Fig. 2). Then succeeded the participant's demonstrating turn with a subsequent robot matching turn and so on. From the participant's view, the game began with two text boxes appearing beneath the video-windows reading "no video input". The demonstration turn of the robot was indicated by the token "A" in the upper left corner of the screen. Then (after a delay of $10s$ for experimental items and a randomly chosen delay of $2s$ to $10s$ for filler items), a countdown from 3 to 1 initiated the transmission of the robot video showing its action demonstration in the big window. After the robot demonstration, the window went back to the initial "no video input" display and the black border around the text boxes changed to a light gray, prompting the participant to match the robot demonstration to one of the action descriptions. To match an action description the participants first chose an action with the arrow keys of the keyboard. Their choice was visualized with a yellow border around the respective text box. Then, they confirmed their choice by pressing the enter key. A red border around the respective text box appeared. All keys of the keyboard other than the ones used at a certain point in time were disabled. In the next step, the participant's demonstration turn began. The token "B" appeared in the corner of the screen and two new text boxes with black borders appeared. One of them was subsequently (after $2s$) highlighted with a yellow border indicating that the contained action should be demonstrated. If the action involved an object, the experimenter gave it to the participant and when she was ready to demonstrate the action, she pressed the space bar to start the transmission of the video. Again a countdown was shown, here, in the small participant window, before the video of the webcam was visible. After the action demonstration, the participant pressed the space bar again to stop the video and the window reverted to the "no video input" display. The space bar signal was used not only to show the videos, but also to store them for later analysis. After a short delay (randomly chosen between 2s to 10s), a red border appeared around the text box that indicated the target action, to give the impression that the robot had correctly matched the action demonstration. The game continued with another robot demonstration and participant matching turn and so on until all items had been demonstrated.

The game began with two filler items before the first experimental item was presented, namely "how to hammer" and "how to grasp an apple". This number is less then the number of initial filler items in the original study by Pearson et al., but compared to their items our items are much greater in terms of effort and time. Nevertheless, the two filler items ensured that the participants had understood the instructions. Each experimental item was presented as an *experimental trial* consisting of a robot demonstration turn and participant matching turn for the prime action of the experimental item, followed by two *filler trials*: the participant demonstration and corresponding robot matching turn for a pair of actions of a filler item and robot demonstration and corresponding participant matching turn for the other pair of filler actions of the filler item. Then the target action (same as prime action) was demonstrated by the participant and subsequently matched by the robot. Thus, there was always one participant demonstration turn and one participant matching turn in-between the matching and naming of an experimental action. All nine experimental trials were presented in succession. At the end of the experiment, the participants filled out a questionnaire, which included questions about how much experience they had with robots, if they had known the iCub robot before participating, which other robots they knew, what they thought the study was investigating and whether they believed there was a real robot on the other side. It concluded with a small personality questionnaire based on Rammstedt and John's 10-question Big Five Inventory [27], which comprises two questions for each factor of the Big Five personality traits [11], (for instance see Table 2 for the questions on emotional stability).

Randomizations

The participants were randomly assigned to the basic robot or the advanced robot condition. Just like in the original study, we constructed two lists of experimental items. In the one list, four (in the other list the remaining five) experimental actions were presented in the uncommon way and five (resp. four) were presented in the common way. Thus, one version of each item appeared in each list. For the filler items we defined a fixed randomized order. The order of the experimental items was randomized individually for each participant.

The action text boxes containing a matching and a distractor action were always displayed side-by-side (see Figure 2). The position of the text box containing the matching action was randomized to be displayed on the right or left side for half of the participant matching turns in order to avoid bias. Accordingly, the text box containing the target action appeared as well on the right or left for half of the participant demonstration turns.

Instructions

The experimenter gave verbal instructions to the participants. These comprised two parts. In the first part, the experimenter explained the game and setup. The participants were told that iCub was in a different room and that communication took place over a network connection. They were instructed that they would be referred to as player B and the robot as player A. Then the network application and the course of the game was shown and explained to them in detail including the participant and robot turns and controls for matching an action and starting the video for demonstration. They were told that on the robot screen, their action demonstration was shown in the big window and the robot demonstration in the small window. The second part consisted of an introduction to the iCub robot according to the condition the respective participant belonged to: *basic* or *advanced*. In the basic condition the experimenter stated that the iCub they would play with was a very basic version and explained that this meant it was restricted in hardware and software and was not equipped with the newest sensors. Additionally, it had not received much training in action recognition or execution. In the advanced condition, the iCub was presented as an advanced version which received all updates and was equipped with the newest sensors. The experimenter told the participants in this condition that the iCub had received much training in action recognition and execution.

Scoring

We obtained scores for the degree of alignment with five raters who individually and independently rated the similarity of the initial action demonstration by the robot and the subsequent action demonstration by the human participant on a scale from 1 [very similar] to 7 [not similar at all].
Five raters (3 f, 2 m) were consecutively presented with pairs of videos of all experimental trials. Each pair of videos consisted of the robot demonstration of the experimental action in one particular experimental trial and the respective demonstration of the same action by the participant in that trial. The order in which the pairs of videos were presented was randomized for each rater individually. The similarity of the two action demonstrations shown in the pairs of videos was rated independently on a 7-point Likert scale from 1

Instructions: How well do the following statements descibe your personality?

I see myself as someone who ...	Disagree strongly	Disagree a little	Neither agree nor disagree	Agree a little	Agree strongly
... is rather reserved, cautious	(1)	(2)	(3)	(4)	(5)
... gets nervous easily, is insecure	(1)	(2)	(3)	(4)	(5)

Table 2 Questions on emotional stability in the Big Five Inventory of the questionnaire. Scoring the scales: Emotional Stability: 1, 2R (R = item is reversed-scored). Inventory based on [27]

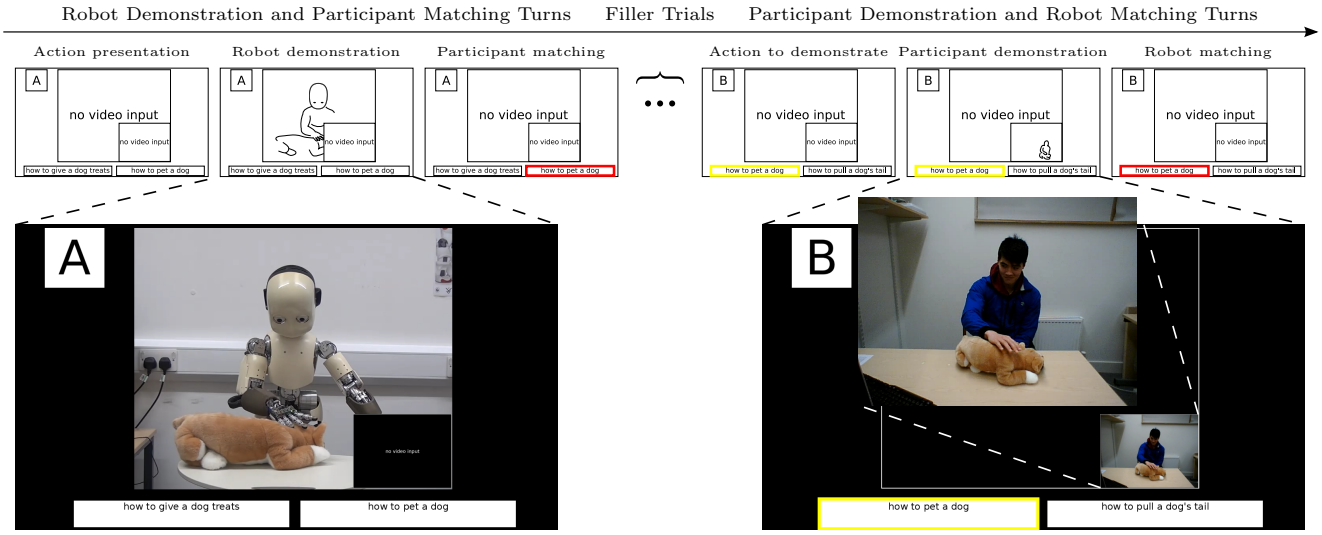


Fig. 2 Schematic view of the course of the experiment showing robot demonstration / participant matching turns and subsequent participant demonstration / robot matching turns with action demonstration screenshots

[very similar] to 7 [not similar at all]. For the ratings, handedness was not taken into account as a factor for similarity. Raters were told to consider the ways in which each action could be carried out, but were not instructed which factors to take into account. An interrater reliability analysis was performed to determine consistency among the raters. We computed the absolute agreement for the average scores of the raters via a two-way random effects model of intraclass correlation. It revealed a strong agreement: $ICC = 0.88$ with $p < 0.001$, 95% $CI(0.86, 0.90)$. The mean of the ratings was taken as a measure for the extent of alignment of the participants to the robot demonstrations. For the analysis we considered only eight of the experimental items excluding one experimental item, "how to swim" (common way: breast stroke, uncommon way: butterfly stroke), because during the course of the study, it became evident that most participants were unfamiliar with the butterfly stroke and did not feel confident enough to perform it. For each participant, the ratings for all common experimental items (i.e. actions which were presented by the robot in a common way) and respectively also the ratings for all uncommon experimental items (i.e. actions which were presented by the robot in an uncommon way) were averaged.

4 Analysis 1

4.1 Results

In line with Pearson et al.'s findings [23], we hypothesized that the participants adapt the way they demonstrate actions to the capabilities of the robot. In particular, the participants should align their action demonstrations more to the robots actions (i.e. presenting the uncommon way of execution more often) when they believe the robot to be basic than when they believe it to be advanced.

A two-way mixed ANOVA with robot demonstration condition (common, uncommon) as within-subject factor and the belief about robot capabilities (basic, advanced) as between-subjects factor for the remaining eight experimental items revealed a statistically significant difference in the extent of alignment based on the way the robot demonstrated an action (common: $M = 2.39$, $SD = 0.6$, uncommon: $M = 3.68$, $SD = 1.04$) , $F(1, 29) = 52.72$, $p < 0.001$; *Wilks'* $\Lambda = 0.355$, *partial* $\eta^2 = 0.65$. The participants rarely demonstrated an action in a different way, when the robot previously demonstrated the action in the common way (i.e. after a common prime). We expected this main effect,

Participant subsample	Within-subject Robot action execution	Between-subject Robot condition	M	SD	N
Higher emotional stability	common	basic	2.52	0.57	10
		advanced	2.39	0.5	9
	uncommon	basic	3.45	0.83	10
		advanced	4.04	0.91	9
More experience with robots	common	basic	2.44	0.72	6
		advanced	2.07	0.65	5
	uncommon	basic	3.05	0.68	6
		advanced	3.86	1.1	5
More robots known	common	basic	2.11	0.63	7
		advanced	2.14	0.47	11
	uncommon	basic	3.13	1.01	7
		advanced	3.96	0.99	11
Knew iCub	common	basic	2.26	0.68	6
		advanced	2.14	0.54	7
	uncommon	basic	3.17	0.86	6
		advanced	3.89	1.15	7

Table 3 Descriptive statistics of the alignment scores for the different subsamples of participants (low mean values indicate higher alignment)

as the common way of demonstrating an action was chosen because it is preferred over the uncommon way by over 80% of people. Still, we observed that the participants aligned their actions also to the uncommon way of execution by the robot, which is reflected by a relatively high mean score of 3.68 of 7 for the uncommon actions. We were not able to find any significant main effect for a participant's prior beliefs about the capabilities of the robot or any interaction effect for within- and between-subject factors. However, including the participants' personality traits and their familiarity with robots in the analysis showed significant interaction effects. We split the participants into two groups at the median of factors assessed in the questionnaire. We obtained different splits of the participants for the Big Five personality traits (openness, conscientiousness, extraversion, agreeableness, neuroticism) according to the Five Factor model [11], for the participants' experience with robots, for the number of robots they knew, and for if they knew iCub before the experiment. Separate two-way mixed ANOVAs for each subset revealed a significant interaction effect for the subsamples of participants with a high score in emotional stability ($Median = 2.5$, on a scale from 1 [lowest value] to 5 [highest value], observed range: 1.5–4.5) and the participants that had some experience with robots ($Median = 1$, on a scale from 1 [no experience] to 5 [much experience], observed range: 1–5). They revealed a trend for the participants that knew more robots ($Median = 5$, observed range: 0–13), and the participants that knew iCub before (for an overview of descriptive statistics and ANOVA results, see Tables 3 and 4). Follow-up one-tailed independent t-tests were conducted to test our hypothesis that scores in the basic condition are lower than in the advanced condition for uncommon robot demonstrations (see Table 5). Results suggest that the participants in these subsamples were more likely to align their actions to the uncommon way the robot previously demonstrated, when they believed the robot to be basic, then when they believed it to be advanced.

According to the post-experimental questionnaire, no participants had suspicions about the nature of the interaction or knew what the study was investigating. All participants believed they were interacting with the robot in real-time.

Participant subsample	t	df	p
higher emotional stability	-1.48	17	0.08
more experience with robots	-1.5	9	0.08
more robots known	-1.72	16	0.05
knew iCub	-1.26	11	0.12

Table 5 Results from independent samples t-tests for subsamples of participants; one-tailed p-values are reported

4.2 Discussion

The results of this analysis in general reveal the occurrence of the phenomenon of alignment in the domain of actions in human-robot interaction. People adapt their action demonstrations to the way their robot partner

Participant subsample	Wilks' Lambda	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
higher emotional stability	0.79	4.53	1	17	0.05	0.21
more experience with robots	0.38	14.65	1	9	0.00	0.62
more robots known	0.8	3.94	1	16	0.06	0.2
knew iCub	0.78	3.17	1	11	0.10	0.22

Table 4 Interaction effects from two-way mixed ANOVAs for subsamples of participants

has previously shown them. We did not find an overall influence of people's beliefs about their robot partner on the extent of alignment to the actions of the robot in this setup. However, our results suggest that the participants' personality as well as their familiarity with their interaction partner affect how their beliefs influence the alignment to their partner's actions. For the field of human-robot interaction, the sole occurrence of alignment implies that in principle for action understanding, robots can benefit from actions being shown to them in a way similar to the ones they already know and similar to the way they execute them.

We were not able to replicate Branigan et al.'s results in general for the alignment of actions in human-robot interaction in this setup. Their findings were confirmed only to some extent. It is arguable that the fact that we could not find overall differences might be due to the experimental design: Even though we carefully constructed the experimental items, certain distractor actions paired with a target action in the participant demonstration turns may have lead the participants to demonstrate the action in a way that was as distinct from the distractor as possible. In the "how to pet a dog" item for instance, the relevant distractor action was "how to pull a dog's tail" leading the participants in all conditions to rather perform the uncommon way of execution (several pats on head), as it was spatially farther from the dog's tail than the common way (long strokes along back). This might have overridden any tendency to align. For the experiment conducted by Branigan et al. the same issue holds true. In their experiment, the participants might have given a label naming the target picture which is as distant as possible from the word they would use for the distractor picture it was paired with (considering e.g. length of word, initial letter or number of letters shared) to avoid an accidental confusion of words by the partner. As a possible example, the participant might prefer the label "building" over the label "house" to avoid confusion with the distractor "mouse". In the future, this issue should be addressed and the experimental design improved in this respect.

One possible explanation for our findings could be that people familiar with robots may have an idea about the potential action recognition mechanisms of the robot. For that reason they might have paid closer attention to the hands and object movements, as they were aware that the robot was likely to track any of them. In the basic robot condition, they then rather copied the movement the robot had demonstrated and already knew. Another explanation could be that higher emotional stability and greater familiarity with robots cause a positive attitude and less tension and stress for the participant in the interaction with the robot. Thus, these factors enable the participant to be less focused on herself and be more responsive to the interaction partner and its needs in the communicative situation and equally to consider how to achieve successful communication. Therefore, the participants with high scores on emotional stability and the participants with greater familiarity with robots align more strongly to their basic interaction partner they believe to be untrained and less capable of recognizing the actions they show. More feedback from the robot which gives them cues about the understanding of the robot and possibly emphasizes the condition of the capabilities of the robot might promote and boost the participants' reactiveness and adaptivity. Branigan et al. suggested that the beliefs about a computer as interlocutor may be resistant to dynamic updates and changes induced by the feedback from the computer [4]. The system both in their basic and advanced condition gave always the correct response, so an assumed basic system gave an incongruous feedback. Still a system that was believed to be basic was treated differently than one that was believed to be advanced. In the current study, the feedback by the robot also consisted of the answers of the robot in the robot matching trials which always indicated perfect understanding and additionally, beliefs could possibly have been updated when seeing the robot perform the actions in a sophisticated manner.

It might thus not be sufficient to tell the participants about the capabilities of the robot before the experiment in order to prime their beliefs. A number of variables might have an impact on how the robot is perceived. We propose to investigate the influence of the following manipulations:

- the action performance by the robot
- the matching answers of the robot
- the time it takes the robot to perform the action

– the time it takes the robot to give an answer on matching turns
– the appearance of the robot

Investigating this aspect more closely for a robot as interaction partner will be subject of following research.

5 Analysis 2

Another interesting factor to look at is time. Time might change participants' alignment behavior. Participants might need time to feel at ease in the interaction, get the hang of the game, form a strategy in which way to demonstrate the actions and get accustomed to the ways of the robot. Branigan et al. showed that beliefs also affect the time-course of alignment [4]. They investigated time within one trial and found reduced alignment for the participants who believed to be playing with a human partner when they had to produce the label some turns after their partner opposed to when they had to produce it right away. When the participants believed they were playing with a computer partner, there was no decay of alignment over time. However, Branigan et al. did not consider time over the course of an interaction comprising several trials. In the setting of this work, in a next step, we would like to address the question of how the extent of alignment changes over the course of an interaction.

5.1 Additional Methods

Instead of averaging the obtained ratings over all uncommon/common experimental items as done for Anaysis 1, the experimental items were divided in half according to their chronological order for each participant. The ratings for the first half of experimental items and the ratings for the second half of items were averaged separately for each participant.

5.2 Results

We performed a two-way mixed ANOVA with chronological order (first, last) and robot demonstration condition (common, uncommon) as within-subject factors and the belief about robot capabilities (basic, advanced) as between-subjects factor. Additionally to the statistically significant difference in the extent of alignment based on the way the robot demonstrated an action we already reported in Analysis 1, it revealed a significant interaction effect of the chronological order and the belief about robot capabilities, $F(1, 28) = 6.62$, $p < 0.05$; *Wilks'* $\Lambda = 0.81$, *partial* $\eta^2 = 0.19$. As planned contrasts, two paired-samples t-tests were conducted to compare the extent of alignment during the first half of experimental items and the second half of experimental items in the basic and advanced robot condition separately. In the advanced robot condition, there was no significant difference of alignment between the first ($M = 3.02$, $SD = 0.65$) and second ($M = 3.02$, $SD = 1.1$) half of items, $t(14) = 0.01$, $p = 1$. In the basic robot condition, results revealed a significant difference between first ($M = 3.38$, $SD = 0.93$) and second ($M = 2.7$, $SD = 1.01$) half of items, $t(15) = 2.43$, $p < 0.05$, indicating that in this condition, the participants aligned significantly stronger in the second half of the interaction (see Fig. 3).

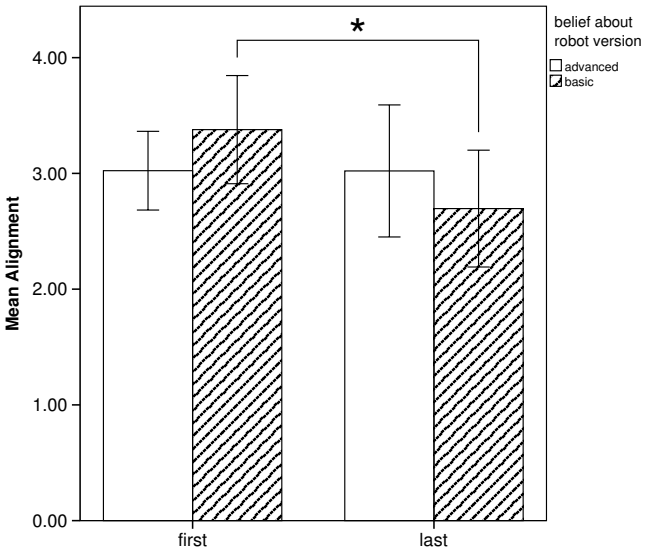


Fig. 3 Mean alignment score (low scores correspond to less differences and thus a high extent of alignment) for chronologically first and last half of experimental items in both belief conditions: advanced robot version (white), basic robot version (hatched); error bars represent standard errors

5.3 Discussion

With this analysis about the change of alignment over the course of an interaction, we found that when the participants, who believed they were interacting with a basic version of the iCub, had interacted with the robot for a while, their alignment to the actions of the robot increased. The participants, who believed they were interacting with an advanced version of the iCub, on the other hand, seem to align to the same extent over the course of the interaction. We believe that this is an interesting result which is worth investigating further. When observing an interaction partner's action demonstrations, there are many aspects to attend to. One

possible explanation for our findings therefore could be that the participants in the basic robot condition first have to find out what is basic about the robot they are interacting with. When participants are playing the game with a robot for the first time, in the first trial they might only pay attention to what the robot looks like, if it can perform the action at all and also as a first priority they fulfill the tasks they were told to do: a) In the participant matching turns, pick which action the robot is demonstrating (there is no need to pay close attention to the exact demonstration the robot is performing, as long as they can distinguish the target action from the distractor action). b) In the participant demonstration turn, demonstrate the movement. Once the participants have done a few trials and get accustomed to the game and know that matching the action correctly is never difficult for them, they can pay closer attention to what the robot might understand and what it could need. They can attend to the actions of the robot and do what they believe leads to success in the communication, for example adapt their demonstrations according to their beliefs about their partner's understanding. Our results thus could suggest familiarity to play a role in enabling beliefs to affect alignment. In accordance, Baddoura et al. found that the experience of familiarity ("the familiar") of an interaction is associated to a pleasant and safe state and correlates with the perception of a robot as a machine with human-like aspects [1]. A future experiment on long-term human-robot interaction with a larger number of experimental items or two or three separate sessions could be conducted to investigate the temporal increase of familiarity and its effect on alignment.

6 Challenges and Limitations

In this work, we presented a thorough replication of an existing study, which we extended to the domain of actions in a human-robot interaction setup. In contrast to the original study by Pearson et al., our study for investigating the alignment of actions encompassed a setup, where the interaction partners had to see their respective demonstrations, but should not communicate at any other time than on the turns in which they demonstrate and match the actions. We solved this issue by implementing a purpose-built Skype-like application with which the participants were able to communicate with the robot via a network connection. In the application, which was implemented as a Linux shell script using the window manager fvwm, video playbacks of the robot demonstrations were followed by the participants' matching responses, which were entered using key presses on the keyboard (left and right arrow keys for selecting a response and the enter key for confirmation), and the display and recording of the participants' demonstrations, which were also triggered by key presses. The application was implemented in a way that disabled all unused keys and prevented any errors in the flow of the game caused by the user. With this interface design, we restricted the interaction considerably, but still aimed at keeping it natural while closely matching Pearson et al's design.

Another challenge we had to solve was that the robot has technical limitations which make it difficult for it to perform certain actions as for instance to grasp an object. These limitations in an interaction with the user would have caused some action demonstrations to fail and the demonstrations of the same action to be incomparable across participants. We therefore recorded videos of the robot demonstrating the actions and replayed them in the application.

Additionally, we could not rely on an existing dataset of actions with different ways of execution, so we had to create a new dataset to find appropriate experimental actions.

The main limitation of the presented study is that a video display of a robot presenting actions constitutes a much richer source of information than the simple display of labels in the original work. The participants might be influenced by the appearance of the robot, as well as the way in which it moves and performs the actions. While the presented work is not able to provide a comprehensive understanding of alignment of actions in human-robot interaction, it still gives meaningful insights on interaction dynamics for future research on this topic.

7 Conclusion

In our experiment, we found that people seem to also align their *non-verbal actions* to the ones of a *robot* interaction partner. They aligned even with very uncommon action executions. In particular, people who are more emotionally stable as well as people who are familiar with robots seem to align their actions to a greater extent when they believe the robot to be basic and less capable than when they believe it to be advanced and more capable. Thus, in these subsamples of participants, we were able to verify our hypothesis and replicate the results of Branigan et al. suggesting that also in the domain of actions, alignment can be mediated by beliefs about the interaction partner.

Regarding the effect of time on alignment, the extent of alignment for people interacting with an interaction partner which is incapable in their eyes increases over

time. Research on alignment in human-robot interaction should thus not only study short-term interaction. Time and familiarization seem to play a decisive role in the process. Future research should focus on identifying enabling and promotive factors of alignment. We believe that a thorough understanding of alignment toward robot interaction partners would enable us to develop robots that can elicit alignment phenomena in interaction and benefit from a reduced interactional complexity for action learning and understanding.

Acknowledgements This research has been supported by the EU project RobotDoC (235065) from the FP7 Marie Curie Actions ITN.

References

1. Baddoura R, Venture G, Matsukata R (2012) The familiar as a key-concept in regulating the social and affective dimensions of hri. In: ACM/IEEE 7th International Conference on Human-Robot Interaction
2. Bergmann K, Kopp S (2012) Gestural alignment in natural dialogue. In: Proceedings of the 34th Annual Conference of the Cognitive Science Society (CogSci)
3. Branigan HP, Pickering MJ, Pearson J, McLean JF (2010) Linguistic alignment between people and computers. Journal of Pragmatics 42(9):2355–2368
4. Branigan HP, Pickering MJ, Pearson J, McLean JF, Brown A (2011) The role of beliefs in lexical alignment: Evidence from dialogs with humans and computers. Cognition 121(1):41–57
5. Breazeal C (2002) Regulation and entrainment in human-robot interaction. The International Journal of Robotics Research 21(10-11):883–902
6. Brennan SE, Clark HH (1996) Conceptual pacts and lexical choice in conversation. Journal of Experimental Psychology: Learning, Memory, and Cognition 22(6):1482
7. Buschmeier H, Bergmann K, Kopp S (2009) An alignment-capable microplanner for natural language generation. In: Proceedings of the 12th European Workshop on Natural Language Generation, Association for Computational Linguistics, pp 82–89
8. Chartrand TL, Bargh JA (1999) The chameleon effect: The perception–behavior link and social interaction. Journal of Personality and Social Psychology 76(6):893
9. Chartrand TL, Maddux WW, Lakin JL (2005) Beyond the perception-behavior link: The ubiquitous utility and motivational moderators of nonconscious mimicry. The New Unconscious pp 334–361
10. Clark HH, Schaefer EF (1987) Concealing one's meaning from overhearers. Journal of Memory and Language 26(2):209–225
11. Costa Jr P, McCrae RR (1992) Neo personality inventory–revised (neo-pi-r) and neo five-factor inventory (neo-ffi) professional manual. Odessa, FL: Psychological Assessment Resources
12. Damm O, Malchus K, Hegel F, Jaecks P, Stenneken P, Wrede B, Hielscher-Fastabend M (2011) A computational model of emotional alignment. In: 5th Workshop on Emotion and Computing, German Conference on Artificial Intelligence (KI)
13. Fischer K, Foth K, Rohlfing KJ, Wrede B (2011) Mindful tutors: Linguistic choice and action demonstration in speech to infants and a simulated robot. Interaction Studies 12(1):134–161
14. Garrod S, Anderson A (1987) Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. Cognition 27(2):181–218
15. Gustafson J, Larsson A, Carlson R, Hellman K (1997) How do system questions influence lexical choices in user answers. In: Proceedings of EUROSPEECH, Citeseer, vol 97, pp 2275–2278
16. Hartsuiker RJ, Bernolet S, Schoonbaert S, Speybroeck S, Vanderelst D (2008) Syntactic priming persists while the lexical boost decays: Evidence from written and spoken dialogue. Journal of Memory and Language 58(2):214–238
17. Iio T, Shiomi M, Shinozawa K, Miyashita T, Akimoto T, Hagita N (2009) Lexical entrainment in human-robot interaction: can robots entrain human vocabulary? In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, pp 3727–3734
18. Kimbara I (2006) On gestural mimicry. Gesture 6(1):39–61
19. Kopp S (2010) Social resonance and embodied coordination in face-to-face conversation with artificial interlocutors. Speech Communication 52(6):587–597
20. Lakin JL, Chartrand TL (2003) Using nonconscious behavioral mimicry to create affiliation and rapport. Psychological Science 14(4):334–339
21. Metta G, Sandini G, Vernon D, Natale L, Nori F (2008) The icub humanoid robot: an open platform for research in embodied cognition. In: Proceedings of the 8th Workshop on Performance Metrics for Intelligent Systems, ACM, pp 50–56
22. Nalin M, Baroni I, Kruijff-Korbayová I, Canamero L, Lewis M, Beck A, Cuayáhuitl H, Sanna A (2012) Children's adaptation in multi-session interaction

with a humanoid robot. In: IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), IEEE, pp 351–357
23. Pearson J, Hu J, Branigan HP, Pickering MJ, Nass CI (2006) Adaptive language behavior in hci: how expectations and beliefs about a system affect users' word choice. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, pp 1177–1180
24. Peniak M, Morse A, Larcombe C, Ramirez-Contla S, Cangelosi A (2011) Aquila: An open-source gpu-accelerated toolkit for cognitive robotics research. In: International Joint Conference on Neural Networks (IJCNN)
25. Pickering MJ, Garrod S (2006) Alignment as the basis for successful communication. Research on Language and Computation 4(2-3):203–228
26. Pickering MJ, Garrod S, et al (2004) Toward a mechanistic psychology of dialogue. Behavioral and Brain Sciences 27(2):169–189
27. Rammstedt B, John OP (2007) Measuring personality in one minute or less: A 10-item short version of the big five inventory in english and german. Journal of Research in Personality 41(1):203–212
28. Snodgrass JG, Vanderwart M (1980) A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. Journal of experimental psychology: Human learning and memory 6(2):174
29. Stoyanchev S, Stent A (2009) Lexical and syntactic priming and their impact in deployed spoken dialog systems. In: Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Short Papers, Association for Computational Linguistics, pp 189–192
30. Van Baaren RB, Holland RW, Steenaert B, van Knippenberg A (2003) Mimicry for money: Behavioral consequences of imitation. Journal of Experimental Social Psychology 39(4):393–398
31. Vollmer AL, Lohan KS, Fischer K, Nagai Y, Pitsch K, Fritsch J, Rohlfing KJ, Wrede B (2009) People modify their tutoring behavior in robot-directed interaction for action learning. In: IEEE 8th International Conference on Development and Learning (ICDL)
32. Vollmer AL, Lohan KS, Fritsch J, Wrede B, Rohlfing KJ (2009) Which motionese parameters change with childrens age? Poster presented at the biennial meeting of the Cognitive Development Society (CDS), San Antonio, Texas, USA
33. Vollmer AL, Wrede B, Rohlfing K, Cangelosi A (2013) Do beliefs about a robots capabilities influence alignment to its actions? In: IEEE 3rd Joint International Conference on Development and Learning and Epigenetic Robotics