

A Human-Centered Approach to Robot Gesture Based Communication within Collaborative Working Processes

Tobias Ende, Sami Haddadin, Sven Parusel, Tilo Wüsthoff, Marc Hassenzahl, and Alin Albu-Schäffer

Abstract—The increasing ability of industrial robots to perform complex tasks in collaboration with humans requires more capable ways of communication and interaction. Traditional systems use separate interfaces such as touchscreens or control panels in order to operate the robot, or to communicate its state and prospective actions to the user. Transferring human communication, such as gestures to technical non-humanoid robots, creates various opportunities for more intuitive human-robot-interaction. Interaction shall no longer require a separate interface such as a control panel. Instead, it should take place directly between human and robot. To explore intuitive interaction, we identified gestures that are relevant for co-working tasks from human observations. Based on a decomposition approach we transferred them to robotic systems of increasing abstraction and experimentally evaluated how well these gestures are recognized by humans. We created a human-robot interaction use-case in order to perform the task of handling dangerous liquid. Results indicate that several gestures are well perceived when displayed with context information regarding the task.

I. INTRODUCTION

Recently, some robotic systems gained sufficient control capabilities to perform complex human-robot interaction tasks that heavily involve the mutual exchange of physical forces [1]. Some of these novel light-weight devices were already introduced to the industrial market, i.e. as state-of-the-art technology. This made it possible to automate difficult assembly tasks, which were still performed manually up to now. Classical industrial robots are unable to solve these tasks, such as rear axle assembly for automobiles, as they require sensible manipulation capabilities to prevent damages to the components during low tolerance assembly processes.

Apart from these automation oriented solutions involving robots capable of interaction with unknown environments, such novel work stations also aim to eliminate the need for classical safety barriers such as fences or light barriers. It is expected that these work stations can help realize applications where direct physical cooperation between human and robot is desired. Various safety investigations and high-performance safe interaction control schemes were developed [2]. However, despite the fact that these investigations and powerful methods enable robots to safely perform interaction tasks, their continuously increasing repertoire of (interaction & manipulation) behaviors makes it almost impossible to appropriately interpret the particular goal and state of the

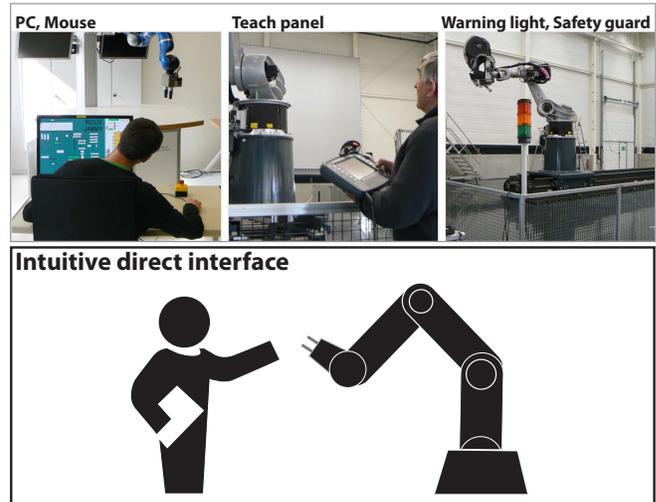


Fig. 1. Until now, PC workstation (upper left) and teaching panel (upper middle) are used to control the robot. Warning lights and safety fences (upper right) are installed for safety. Our long term goal is to eliminate physical safety barriers entirely in order to interact with robots in an intuitive direct manner (lower).

robot visually alone. In other words, the human observes the robot, but the robot's pure task motion does not provide sufficient information to the human to fully understand the overall task state.

Industrial systems would commonly resolve this issue by using external safety guards, touchscreens, control panels, or warning lights, see Fig. 14 (upper). They are used to operate the robot, or to communicate its state and/or prospective actions to the user. [3] proposed an interaction concept, which incorporates a touchscreen displaying a robotic system in a modeled scene of the real-world. The operator can then control the robot by “drawing commands into the virtual scene”. This technique is described as “drawing on the world” and reflects the desire to interact directly with the object to be controlled more easily and intuitively. However, the aforementioned solutions bear a significant drawback: *distraction from the actual interaction task*. This makes it necessary to provide additional interface capabilities to the robot to allow the human to intuitively interpret the robotic co-worker's “mode” or the current task situation. Enriching the direct interaction capabilities of robotic systems is, apart from interaction control, mainly pursued in the fields of speech interface and human gesture recognition. However, these require additional sensors and integration

T. Ende, S. Haddadin, S. Parusel, T. Wüsthoff, and A. Albu-Schäffer are with the Institute of Robotics and Mechatronics, DLR - German Aerospace Center, Wessling, Germany, contact: sami.haddadin@dlr.de

M. Hassenzahl is with the Folkwang University of the Arts, Essen, Germany

efforts. Furthermore, they still suffer robustness problems in real-world settings. The approach taken in this paper is to enable the robot to perform explicit interaction motions, i.e. robot gestures, to communicate directly with humans during work processes. Furthermore, our long term goal is to eliminate separate interfaces entirely in order to make it possible to interact with robots in a direct manner, see Fig. 14 (lower). [4] examined direct communication via some “cooperative gestures” using a humanoid robot and found encouraging results for the recognition of robotic gestures by humans, depending on certain conditions regarding their execution speed and direction. In this paper, we focus on the possibilities of displaying gestures that are observed in human-human-interaction during collaborative assembly scenarios by robotic systems of different anthropomorphism. This should help us understand, which information can be used in single arm manipulator (SAM) settings, common in industrial settings or professional service scenarios. This work searches for relevant gestures that can be executed not only by humanoid robots, but also by common setups in real-world applications¹.

The first step in this paper is to identify and gather gestures from human-human interaction, which are relevant for co-working tasks. Section II describes the steps we applied for observing humans interacting in different situations. Relevant gestures are isolated and presented in Sec III. In Sec IV we present a robot gesture lexicon containing the selected gestures from human observation. These were performed with a human arm, an anthropomorphic hand-arm system, and ultimately with a ceiling-mounted SAM. We analyze the recognition performance by humans via a set of experiments described in Sec. V. The use-case in Sec. VI shows that the selected gestures can provide valuable information to the human during a rather complex scenario, see Fig. 2.

II. HUMAN OBSERVATION

We intend to use robot gestures as a non-verbal interactive modality that acts as an integral part for a given cooperative task. Furthermore, our primary aim for HRI concepts is intuitiveness. We take gestures from cooperative human-human-interaction in co-working scenarios as our antetype model. Our analysis utilizes three types of observation: *observation of human gesturing*, *acting out*, and *observation of static poses* [6]. The results and selected gestures according to the criteria for cooperative gestures that we transferred to the robotic systems are given in Sec. III.

A. Observation of human gesturing

Interaction processes between colleagues, family members, friends, and unknown people, during everyday activities were observed and video-recorded. The filmed persons were told to execute complex human-human collaborative activities, such as modular construction element assembly, electrical repair, furniture assembly, cooking, and gardening, see Fig. 3.

¹A video showing several aspects presented in this paper can be found as a video attachment or at www.safe-robots.com/iros2011-process-gestures.html

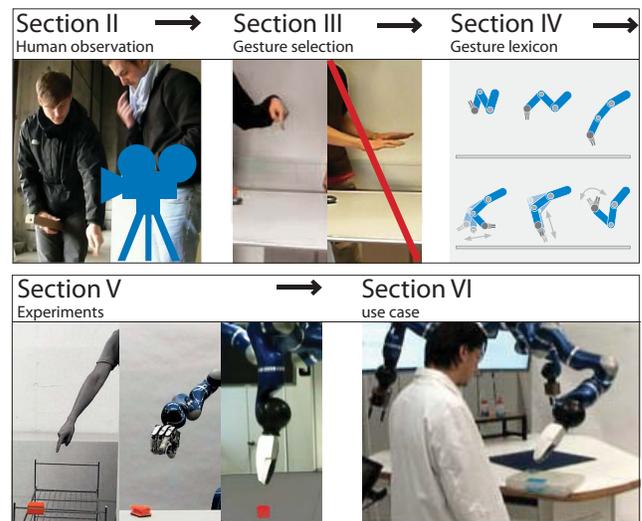


Fig. 2. Observation of interaction between humans (upper left), selection of hand/arm gestures, which are relevant for a single arm manipulator (upper middle), gesture lexicon (upper right), experiments including human arm, DLR mobile humanoid robot Justin [5] (lower left), Use case with the task of handling dangerous liquid (lower right).



Fig. 3. Observation of colleagues, family members, friends, and unknown people during everyday activities.

B. Acting out

In order to

- 1) provoke situations that were not observed during the aforementioned observation of natural gesturing, and
- 2) maximize the information flow (e.g. the gesturing of one participant) from one person to the other,

we performed experiments in which two participants were acting out an artificially constructed scenario without speech. In this cooperative assembly scenario, plastic cube structures were built according to a predefined plan, which was known only to one of the two persons, see Fig. 4. Initially, the informed person sat in front of the scenario according to the given plan, two distinct situations were provoked in case the second human enters the scene:

- The sitting person tells the arriving one to leave, or in case he does not comply shields the scenario²
- The sitting person was told to collaborate with the arriving one and guide the arriving person through the process of building the cube structure.

²The scenario incorporates “delicate” process (which could also be a non-interactive part of a larger collaborative assembly scenario only) that must not be disturbed by another person.



Fig. 4. Acting out: guiding a person's actions via gestures.

C. Observation of static poses

Besides active gestures during a cooperative assembly task, it is clear that some special system states, which are usually not part of observed human-human scenarios, should also be displayed by the robots's geometric configuration. This can already provide relevant information about the system state to the human without additional external communication devices. We also observed several "static gestures" of the robot (see Fig. 5), which correspond to the most important static system states **Sleep**, **System Standby**, and **Error**, see Fig. 10.



Fig. 5. Sleeping position (left), waiting position (middle), and error pose (right).

The observations and considerations presented up to now yielded a set of observed static postures and active gesticulation that are not categorized or arranged in any meaningful order. In order to eliminate the ones that are not suitable to be translated into robot gestures, we developed several criteria based on gesture classifications elaborated in the following section.

After a short survey on these aspects, we describe the gesture repertoire obtained from the observation process analysis that were transferred to the robotic system.

III. THEORY OF WORK PROCESS AND GESTURING

According to [7], a task generally consists of actions, operations, and motions/gestures that are arranged in a hierarchical order, see Figure 6. A task is composed of several actions, which have distinct goals (e.g. Human H1 and human H2 cooperatively assembling part A). A potentially complex structure of operations and gestures represents one action in a more detailed manner. In particular, it reflects the conditional substructure of an action (operation: H2 assembles part A if human H1 gives it to H2, gesture: H1 shows "stop!" gesture if H2 does something wrong.). Their execution is then a simple chain of atomic motions. The distinction between operations and gestures is not necessarily unique, as several motions can be assigned to both (e.g. handing

over, or receiving an object). Gestures are operations, which are composed of motions in a chronological order.

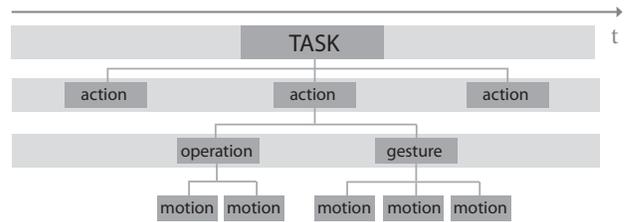


Fig. 6. Extended macro structure of working process, similar to [8].

According to [9], there exists a hierarchy of gestures called "**Kendon's Kontinuum**" [10], where the presence/fraction of language decreases gradually. The robotic system will deal with emblematic gestures (symbols) which are characterized by a clear association of their form to specific meanings without the aid of speech. Furthermore, the selected gestures are easily and intuitively interpretable. As a result, these gestures deemed to be most relevant.

In this paper we are particularly interested in hand-arm motions for active gestures. A well known classification of hand gestures was developed in [11]. In this work, hand and arm motions are split into five categories.

- **Symbolic gestures:** gestures that are assigned with firm meanings in our cultural environment and which are frequently used in everyday experience.
- **Interactional gestures:** gestures used to regulate interaction with a partner, i.e. to initiate, maintain, invite, synchronize, organize, or terminate a particular interactive/cooperative behaviour.
- **Pointing gestures:** gestures which refer to objects or humans in the environment.
- **Side effects of expressive behavior:** while communicating with others, motions of hands, arms, and face occur as part of the overall communicative behavior but without any specific interactive, communicative, symbolic, or referential roles.
- **Body/manipulator motions:** gestures which are neither communicative nor socially interactive, but are rather instances of effects of human motion. They may be salient but are not movements that are primarily employed to communicate or engage a partner in interaction.

The following **criteria** were the basis for our selection of observed gestures.

- The gesture has to be conceivable without specific knowledge as e.g. sign language.
- The gesture does not involve other parts of the body than the arm and hand.
- Regarding the hand, the gesture does not involve other fingers than the index finger.

The last two categories from [11] are not important for our consideration, since we aim to transfer the gestures to a single manipulator, which lacks a well defined body. Furthermore, we selected only gestures that can be executed at the height between head and pelvis.

The identification of gestures followed our derived criteria for co-working gestures and general properties for gestures (see Fig. 8). The video of the interaction processes was analysed by three study participants. All participants obtained similar results.

1) **Isolated gestures of human observation:** In this experiment four students were observed while assembling an Ikea table. During the 2-minute film sequence, 10 different hand/arm gestures were executed by the participants. In 9 out of 11 interactions from one participant to the other, a meaningful hand/arm gesture was used to support the spoken words. In one of the 11 interactions, a single hand/arm gesture was used for this particular example without speech, see Fig. 7, where the observed interaction structure is displayed.

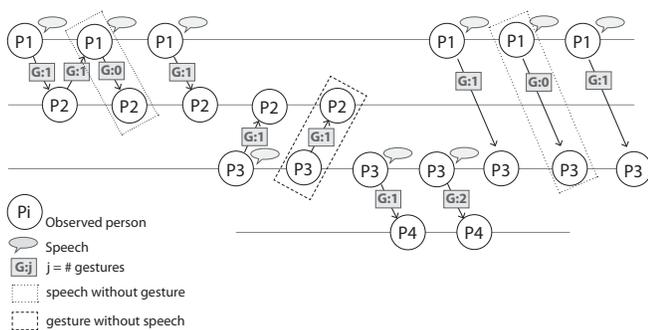


Fig. 7. Observed interaction structure with speech and executed gestures in the cooperative assembly task “Ikea table with 4 participants”.

Isolated hand/arm gestures according to the aforementioned criteria from human observation are:

- “Come here!”, “Come closer!”, “This one!”, “From here, to there!”, “Display object”, “To give something”, “Give it to me!”, “Go away!”, “No!, Caution!”

2) **Isolated hand/arm gestures from acting out are:**

- “Stop!”, “Higher!”, “Turn!”, “Get out of the way!”, “Lower!”, “Underneath!”, “Slow down!”, “No idea!”, “Carry on!”

Generally, the described observations show that interaction between people during a collaborative task is strongly accompanied by various gestures. Almost every observed interaction consists of spoken words combined with particular gestures that are used to emphasize the commands, or to refer to objects or points in the surrounding. Only a few gestures are used without speech (see Fig. 7). This observation shows that the modality of gestures is strongly present in our everyday communication within collaborative tasks. They support verbal information or even entirely replace spoken commands. Since we aim at gestures for every phase of a specific task, we arranged the isolated gestures corresponding to the structure of interactional gestures. That shows that gestures are used, sometimes exclusively, to initiate, maintain, invite, synchronize, organize or terminate a particular interactive, cooperative behaviour [11]. The order of isolated

gestures is illustrated in Fig. 15, where they are used for evaluating their respective recognition rate. To design the isolated gestures for recognition analysis, they followed the specific rules according to Efron, which are e.g. described in [12]. Gestures are subject to following general properties.

- The size of the radius of movement and from which part of the arm the movement originates.
- The form of the movement, whether it is sinusoidal, elliptical, angular or straight.
- The plane of the movement, whether sideways or transverse, towards or away from the auditor.
- The body parts involved in the gesticulation (hand/arm gestures), digital gestures (considering finger position), the laterality of the gestures (only one hand is employed), and
- the speed of the gesture - whether it has abrupt transitions from one speed of movement to another.

When performing pointing gestures it has to be taken into account that the finger, hand and arm are aligned in such a way that the motion is target-oriented with a straight dynamic of motion and with a pause at the maximum distance [13]. To distinguish gestures accurately each one needs to be preceded and followed by a pause. Furthermore, according to [4], every gesture consists of the following three temporal phases (decomposition), see Fig. 8.

- preparatory phase
- execution phase (i.e. stroke), potentially containing
 - repeating sequences (i.e. wave)
 - or hold (i.e. stop gesture, pointing gesture)
- retraction phase

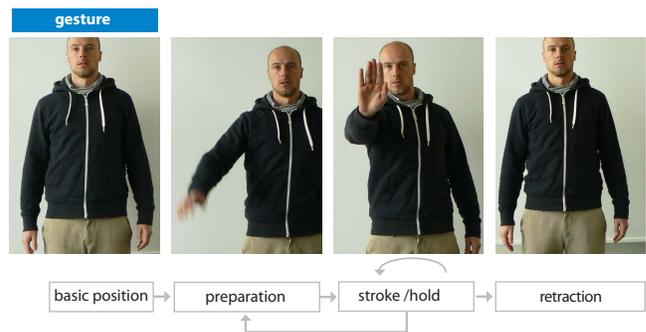


Fig. 8. Structure of gestures with four phases: basic position, preparation, stroke/hold and retraction. A repetitive behavior can be observed either from stroke/hold back to preparation phase or only within the stroke/hold phase.

With respect to the specific rules of performing gestures, we implemented the 20 isolated gestures from human observation into the anthropomorphic robot as well as the SAM (described in section V), in order to obtain a lexicon of gestures.

IV. LEXICON OF ROBOT GESTURES

The potential use of gestures is strongly coupled to the particular state the system occupies, see Fig. 12. Static gestures should be associated to the “static” part of the generalized automaton. In order to derive appropriate configurations,

representing the aforementioned static system states, we used images derived from different static poses of the human body, which intuitively represent those states. As already mentioned, Fig. 5 depicts configurations of humans that correspond to these three states. The respective “translated” robot-arm poses are displayed in Fig. 9.

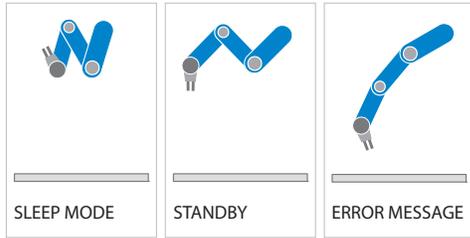


Fig. 9. Sleep mode (left), Standby position (middle), Error message (right)

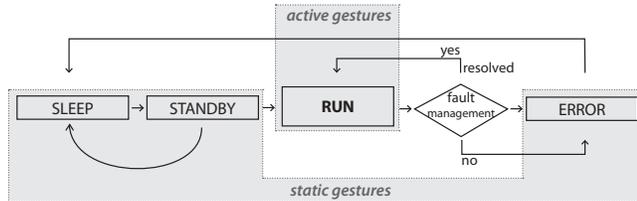


Fig. 10. Generic state machine incorporating static and active gestures.

- **Sleep** is a distinct posture that indicates the robot’s sleep mode. This implies also that the robot is deliberately put into this particular position by a previous user. In order to proceed operation, the robot waits for an active human action to be started (boot up). Then, the robot moves autonomously to the system standby position and awaits for new operation commands.
- **System standby** posture indicates that the robot completed a preceding task and returned to the system standby position afterwards. The robot can then be entrusted with new tasks to be carried out. If no new commands arrive for some time, the robot returns to sleep mode, see Fig. 10.
- **Error message** is a posture indicating malfunctions that occurred during the execution phase. The robot remains in an arbitrary configuration, which is neither one of the two previous modes as the system stopped due to the system error. Therefore, an arbitrary position in the workspace indicates that a severe fault during task execution occurred, needs to be eliminated, and confirmed for continuing operation.

During a running task the robot needs to communicate via various active visual gestures. Active gestures are part of the nominal task program in the generic “RUN” state, see Fig. 10. This can provide valuable information to the human about possible risks, faults of the user, or requests. In the sense of the process, possible classes of such gestures are

- action initiating gestures,
- action terminating gestures, and

- synchronization gestures.

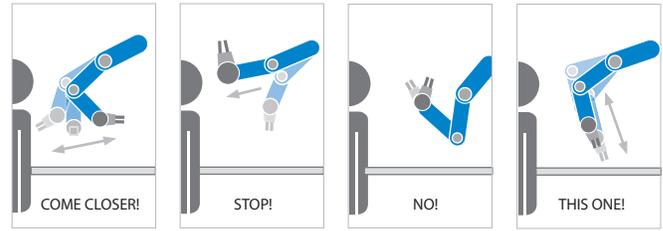


Fig. 11. Active gestures: Come closer!, Stop!, No!, This one!

Some examples of the robot designed gestures are depicted in Fig. 11. Fig. 12 depicts the overall integration into the task state machine that runs our robots. The nominal task consists of parallel and hierarchically structured hybrid action patterns, which define the overall nominal robot behavior. The gesture machine is activated in case a fault (of human behavior, not a system fault) occurs, or the human intends to interact. This causes a task focus shift from the nominal hybrid state machine *task* to *gesture machine*. The gesture machine selects the gesture according to the active context and executes the motion based on the gesture database. Each gesture G_i reflects the human gesture structure outlined in Sec. III.

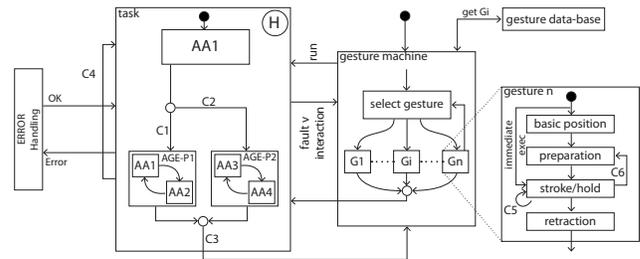


Fig. 12. Task machine with gesture lexicon. See Sec. III for comparison with gesture theory.

For analyzing the recognizability of active gestures performed by a SAM, we conducted experiments with a human, a humanoid system, and a SAM itself.

V. GESTURE EXPERIMENTS AND SELECTION

One goal of this work is to explore the ability of humans to understand gestures performed by robots. We conducted three experiments for analyzing active gestures executed by a human, a humanoid system, and a 7 DoF SAM equipped with a two jaw gripper. First, we performed and filmed the isolated gestures from Sec. III with the three subjects. The gestures are performed without speech by a human facing a camera that is positioned at eye level in the first case. Only the human/robot arm are shown in order to avoid any unintentional influence of the results by perceiving the trunk. A red cube and a yellow sponge serve as targets for pointing gestures. Secondly, we conducted an online survey in which the three videos were presented to study participants in order to analyze the recognition rate of the gestures. Male and

female participants of ages 20 – 55 were recruited from a pool of 400 people for each video, not all of whom had a technical background and experience in robotics. Thirdly, based on the outcome of the experiments we selected a set of gestures for incorporation into the use-case. The experiment video shown to participants contains all gestures, which were separated by three second pauses.

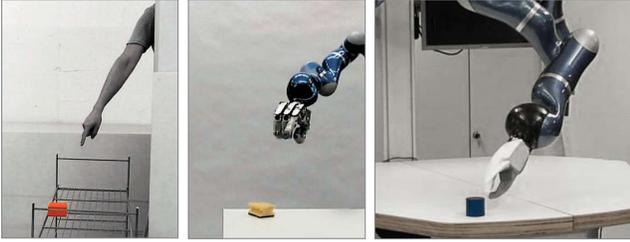


Fig. 13. Experimental setup (pointing gesture) with human, anthropomorphic hand-arm system, and single arm manipulator.

A. Survey one: human

In our first survey we analyzed whether the gestures identified from observation of natural scenes, can be clearly understood by test persons when performed by a human without task-related information. This video was shown to 82 test-persons via online survey. The participants were asked to watch the video and to identify the meaning of each presented gesture. Their interpretation was then compared with the true meanings of the gestures. The recognition rate is depicted in Fig. 15.

In the following two surveys we performed the designed gestures with an anthropomorphic hand-arm-system (four-finger-hand and a 7 DoF robot arm) and a SAM. The aim was to elaborate how well these gestures can be identified when being alienated.

B. Survey two: anthropomorphic hand-arm-system (DLR mobile humanoid robot Justin)

The second survey was conducted in the same manner as the first survey, with 47 participants. The main aspects to be considered for this experiment are:

- The gestures are programmed with the help of a motion capture suit. The transfer of the human kinematics to the robot behavior is done with the algorithms presented in

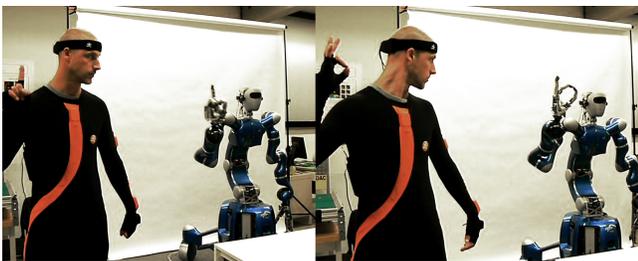


Fig. 14. Humanoid mobile system Justin performing gestures. The motion capture suit was used for transferring human gestures to the robot via the algorithms developed in [14].

[14]. For repeatability the gestures are executed by the same person who performed the initial human gestures used in the first survey.

- The arm is mounted to the anthropomorphic robot Justin and complies with the human model in terms of height and orientation.
- The robot hand shows an anthropomorphically plausible character.

C. Survey three: single arm manipulator (SAM) (DLR LWR III equipped with an industrial two jaw gripper)

The third survey was conducted in the same manner as the first survey, with 40 participants. The main aspects to be considered for this experiment are:

- The position of the arm does not comply with the human model, as it is mounted to the ceiling.
- The limited anthropomorphic character is mainly due to the mounted industrial gripper.
- The overall kinematics is very different from the human body and therefore, performs very unusual movements as compared to a human.
- Gestures are programmed via manual guidance and are therefore prone to unintended changes in accuracy, speed, and acceleration. This may lead to different appearance and characteristics of the gesture in general. We consider this aspect of shortcoming precision important, as it is immanent to the process of manual teaching.

D. Results

The identification rate of the three surveys is depicted in Fig. 15. 20 gestures were tested with the human arm and the anthropomorphic system. Six out of the 20 gestures, showed low identification rates with human arm and with the anthropomorphic system, were not tested on the single arm manipulator. In the experiment, 11 gestures performed by a human arm without contextual information had a recognition rate $\geq 80\%$. A similar result applies to the anthropomorphic hand-arm-system: 9 gestures have a recognition rate $\geq 80\%$. In contrast, only 4 gestures performed by the SAM system have a recognition rate $\geq 80\%$. Gestures that performed well on the SAM are the two referencing gestures “this one!” with 84% and “from here to there!” with 90%. These gestures are also characterized as pointing gestures that use the index finger and a target that represents context information. The terminating gestures also performed well, with 84% for “go away!” and 92% for “stop!”. Gestures that performed significantly worse on the SAM than on the other two systems (human and humanoid) are the three initiating gestures “wave ones hand!”, “Come here!” and “Come closer!”. Generally, terminating and referencing gestures performed well in all three surveys, while the synchronizing gestures showed low recognition rates.

In the next section we describe a use-case in which a three-robot system performs an assembly task, during which one of the robots is able to respond to the human activity by executing the according context based gestures. The applied

Command	Identification rate		SAM	
	Human	Humanoid robot		
Wave one's hand!	80 %	74 %	30 %	initiate
Come here!	97 %	96 %	47 %	
Come closer!	98 %	98 %	26 %	
This one!	93 %	94 %	84 %	reference
From here, to there!	88 %	100 %	90 %	
Display object!	80 %	70 %	63 %	
To give something!	28 %	52 %	71 %	synchronize
Give it to me!	23 %	80 %	39 %	
Carry on!	30 %	26 %		
Higher!	60 %	43 %		synchronize
Turn!	72 %	39 %		
Lower!	43 %	09 %		
Underneath!	13 %	22 %		synchronize
Slow down!	97 %	72 %	04 %	
Get out of the way!	40 %	26 %		
No idea!	72 %	83 %	17 %	terminate
Go away!	93 %	48 %	84 %	
No!	98 %	85 %	58 %	
Caution!	97 %	94 %	68 %	terminate
Stop!	93 %	98 %	92 %	

t

- very well identified
- well identified
- not identified

Fig. 15. Experimental results including recognition rate of three subjects: human, anthropomorphic robot, single arm manipulator

gestures in the use-case were selected according to the results of the aforementioned experiments.

VI. USE-CASE

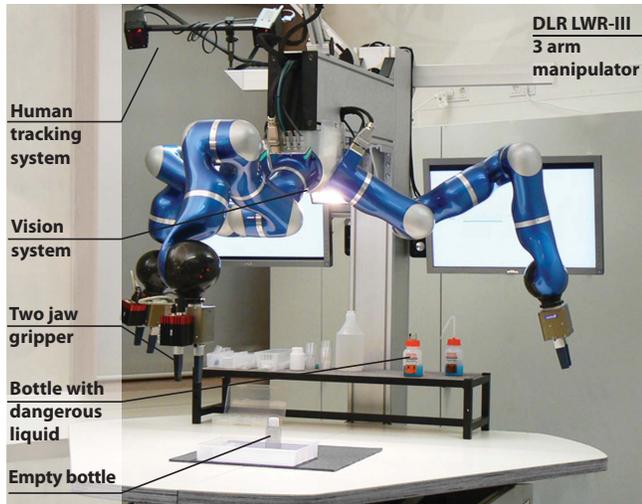


Fig. 16. Use-case setup with DLR LWR-III 3 arm manipulator

The main objective of the use-case task is for a human to hand over a closed, empty bottle to the robot system consisting of three ceiling mounted robots. The bottle is then filled with a potentially dangerous liquid. Finally, the human shall receive the bottle from the system. During the execution phase (filling the bottle) the human is considered as a disturbance to the process. In our use-case two robots are assigned to perform a “dangerous” task, whereas the third one is mainly responsible for the interaction with the

human by means of haptic and gestural communication. In the interest of readability we will name the robots Jimmy, Eric and Eddie. Eric and Eddie handle the bottle task, while Jimmy interacts with the human. The operator is equipped with an ART passive marker tracking device attached to his wrist. Initially, the “dangerous” bottle is placed in the back and the human is assumed to be absent. The setup is depicted in Figure 16. Overall, the implemented use-case consists of the following steps (each number refers to the respective picture in Figure 17).

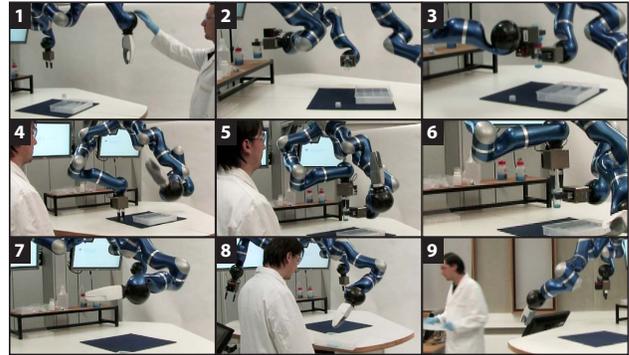


Fig. 17. Human-robot collaborative task

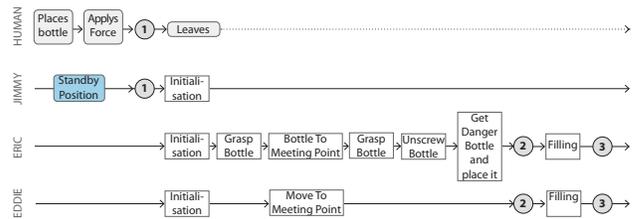


Fig. 18. Structure of the task, grasping a bottle and filling it with potentially dangerous liquid.

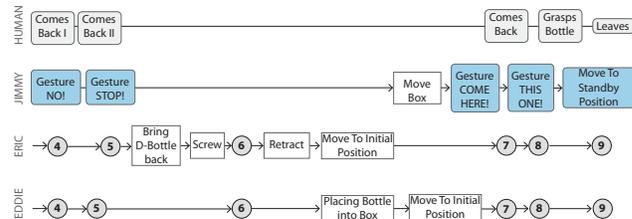


Fig. 19. Structure of the task filling a bottle with potentially dangerous liquid and delivering it to the human.

- 1) The robots rest in **system standby** position to indicate that they can be entrusted with new tasks. The human enters the workspace, places the bottle to be filled, and activates the process by applying a force to the robot in a predefined direction. The human may leave now and carry out another task. Next, the vision system recognizes the bottle. Meanwhile, all manipulators

initialize. Eric lifts the bottle and Eddie grasps it below the cap to prepare the unscrewing of the cap.

- 2) Eric and Eddie unscrew the bottle cap. Eric then retrieves the “danger bottle”.
- 3) Eddie and Eric cooperatively fill the bottle.
- 4) **Human enters the sensing range of the system:** If the human enters the workspace, while the robots have not yet finished the potentially dangerous part of the task, Jimmy can warn the human by executing the active gesture “no!” in direction of the observer as soon as he gets close to the dangerous liquid. This indicates that the human should not come closer until the dangerous part is finished.
- 5) **Human enters the workspace:** If the human gets even closer and disregards the warning gesture “no!”, the stronger gesture “stop!” is activated.
- 6) If the human retracts, Eric places the dangerous bottle back into the shelf, Eddie and Jimmy screw back the cap and place it back into the box again.
- 7) **The robot indicates the human to come closer:** To arouse human attention, Jimmy executes the “come closer!” gesture, indicating that the dangerous part of the task is finished and the human may now safely enter the workspace to pick up the filled bottle.
- 8) **The robot refers to the respective object:** If the human comes closer, the robot points to the respective object with the pointing gesture “this one!”.
- 9) **The human leaves the workspace and all robots return to system standby position** The robot can now be entrusted with new tasks to be carried out.

The overall “screenplay” of the use-case is depicted in Fig. 18 and 19, where the particular actions of the four entities are displayed and the gestures executed by Jimmy are labeled. The full implementation of the use-case can be viewed as a video attachment or at www.safe-robots.com/iros2011-process-gestures.html.

VII. CONCLUSIONS

In this paper we investigated the usability of robot gestures as an interaction modality for human-robot collaborative assembly. Based on a view on the theory of working processes and human gesturing, we elaborated various insights into human-human interaction during collaborative assembly processes. Generally, gestures were observed to be an integral part of this interaction type. Furthermore, gesture execution can be decomposed into four atomic units, which then can be easily transferred into our robot control systems. Based on an experimental analysis we transferred a pre-selected ensemble of gestures to robotic systems of increasing abstraction. We isolated suitable ones, which could even be used by a single arm in a rather complex use-case. An important result is that gestures, as e.g. “come here!” and “come closer!”, which arm motions are strongly related to the particular torso position, are not well conceived for a SAM. This can be concluded since the human’s and anthropomorphic hand-arm-system’s gestures were still well recognized, though their torso was concealed. Synchronizing gestures, as e.g.

“higher” or “underneath” showed poor results for all systems, which is probably due to the lack of contextual information in the experiment.

The set of selected gestures, which we consider as suitable for collaborative applications were either of referencing type, as e.g. “This one!” and “From here, to there!”, or terminating gestures as “Stop” or “No”. These were shown to provide valuable information to the human user for the designed use-case. Furthermore, we also selected some of the poorly recognized gestures such as “come here”, since the context information is well defined for the use-case. This turned out to be well perceived by the test persons.

VIII. ACKNOWLEDGMENTS

We would like to thank Hannes Widmoser, Chris Schindlbeck, and Christian Ott for their valuable help and support. This work has been partially funded by the European Commission’s Sixth Framework Programme as part of the project SAPHARI under grant no. 287513.

REFERENCES

- [1] S. Haddadin, M. Suppa, S. Fuchs, T. Bodenmüller, A. Albu-Schäffer, and G. Hirzinger, *The 14th International Symposium ISRR2009: Towards the Robotic Co-Worker*, C. Pradalier, R. Siegwart, and G. Hirzinger, Eds., Springer-Verlag, Berlin-Heidelberg, 2011.
- [2] S. Haddadin, A. Albu-Schäffer, A. D. Luca, and G. Hirzinger, “Collision detection & reaction: A contribution to safe physical human-robot interaction,” in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS2008)*, Nice, France, 2008, pp. 3356–3363.
- [3] A. Correa, R. Walter, L. Fletcher, J. Glass, S. Teller, and R. Davis, “Multimodal interaction with an autonomous forklift,” in *International Conference on Human-Robot Interaction (HRI)*, 2010 5th ACM/IEEE, Massachusetts Institute of Technology Cambridge, MA, USA, 2010.
- [4] D.Riek, A.Pipe, M.Fraser, and P.Robinson, “Cooperative gestures: Effective signaling for humanoid robots,” University of Chambridge, UK, 2010.
- [5] C. Borst, C. Ott, T. Wimböck, B. Brunner, F. Zacharias, B. Bäuml, U. Hillenbrand, S. Haddadin, A. Albu-Schäffer, and G. Hirzinger, “A humanoid upper body system for two-handed manipulation,” in *IEEE Int. Conf. on Robotics and Automation (ICRA2007)*, Rome, Italy, 2007, pp. 2766–2767.
- [6] N. Mack, C. Woodsong, K. M. MacQueen, G. Guest, and E. Namey, *Qualitative Research Methods: A Data Collector’s Field Guide*, North Carolina, USA, 2005.
- [7] W.Nerdinger, G.Blickle, and N.Schaper, *Theoretische Modelle des Arbeitshandeln*, Heidelberg, Germany, 2008.
- [8] U. Clement and M. Lacher, *Produktionssysteme und Kompetenzerwerb*, Stuttgart, Germany, 2006.
- [9] D. McNeill, *Hand and Mind*, Chicago, USA, 1992.
- [10] A. Kendon, *Signlanguages of Aboriginal Australia*, Cambridge University Press, UK, 1988.
- [11] C.Nehaniv, K.Dautenhahn, J.Kubacki, M.Haegele, C.Parlitz, and R.Alami, “A methodological approach relating the classification of gesture to identification of human intent in the context of human-robot interaction,” in *IEEE Int. Symposium on Robot and Human Interactive Communication (RO-MAN)*, 2005.
- [12] A. Kendon, *Gesture*, Cambridge University Press, UK, 2004.
- [13] A. Kendon, *An Agenda for Gesture Studies*, Cambridge, UK, 1996.
- [14] D. Lee, C. Ott, and Y. Nakamura, “Mimetic communication model with compliant physical contact in human-humanoid interaction,” *Int. J. of Robotics Research*, vol. 29, no. 13, pp. 1684–1704, 2010.