

Gestures for Industry

Intuitive Human-Robot Communication from Human Observation

Brian Gleeson[†] Karon MacLean[†] Amir Haddadi[‡] Elizabeth Croft[‡] Javier Alcazar*
brian.gleeson@gmail.com maclean@cs.ubc.ca haddadi@ieee.org ecroft@mech.ubc.ca javier.alcazar@ieee.org

[†]University of British Columbia
Department of Computer Science
201-2366 Main Mall
Vancouver, BC V6T 1Z4, Canada

[‡]University of British Columbia
Department of Mechanical Engineering
6250 Applied Science Lane
Vancouver, BC V6T 1Z4 Canada

*General Motors
Research and Development Center
30500 Mound Road
Warren, Michigan, 48090, USA

Abstract—Human-robot collaborative work has the potential to advance quality, efficiency and safety in manufacturing. In this paper we present a gestural communication lexicon for human-robot collaboration in industrial assembly tasks and establish methodology for producing such a lexicon. Our user experiments are grounded in a study of industry needs, providing potential real-world applicability to our results. Actions required for industrial assembly tasks are abstracted into three classes: part acquisition, part manipulation, and part operations. We analyzed the communication between human pairs performing these subtasks and derived a set of communication terms and gestures. We found that participant-provided gestures are intuitive and well suited to robotic implementation, but that interpretation is highly dependent on task context. We then implemented these gestures on a robot arm in a human-robot interaction context, and found the gestures to be easily interpreted by observers. We found that observation of human-human interaction can be effective in determining *what* should be communicated in a given human-robot task, *how* communication gestures should be executed, and priorities for robotic system implementation based on frequency of use.

Index Terms—Human-Robot Communication, Collaborative Robotics, Gesture, Industrial Assembly

I. INTRODUCTION AND BACKGROUND

Over the past thirty years, robotic technology has been integrated into the manufacturing industry for the purpose of improving efficiency and reducing worker ergonomic stress and workload, but an evolving industry now requires more effective communication and collaboration between human and robot [1]. Existing robots that directly execute manufacturing operations, such as welding and painting, require no direct human guidance and are completely isolated from humans. Direct interaction between humans and robots is generally limited to physical guidance of assistive tools such as lift-assist devices.

The broad goal of our research is to develop methods of human-robot interaction (HRI) that will facilitate close cooperation between humans and robots on industrial tasks. Such a human-robot partnership would advance industrial productivity and quality by utilizing the different strengths of each partner: the dexterity, perception, and intelligence of human workers and the precision, repeatability, and

procedural knowledge of a “robotic assistant” (RA). An RA could also reduce ergonomic strain by executing repetitive or strenuous tasks.

Robotic technology is approaching the point where an RA and a human could physically interact in a safe and effective manner [2]. Promising avenues for RA hardware include lightweight robots designed specifically for safety [3] and existing industrial robot platforms augmented with improved sensing and control [4]. With the advancing physical capabilities of potential robots comes a need for improved HRI to facilitate human-robot collaborative work. Effective partnership means the development of intuitive robot behaviors and communication methods [5]. The growing potential for effective human-robot partnerships has inspired research into hardware and HRI, ranging from HRI for human-robot teams in space [6], to “LOCOTOT”¹, a large European project started in 2010 to develop customized robot co-workers for vehicle assembly lines.

In this paper, we look specifically at HRI for industrial human-robot teams. Our focus is explicit and intuitive gestural communication to facilitate assembly task execution, smooth flow of task variations, and resolution of assembly errors. To ensure the applicability of our results, we ground our research through interviews with industry experts and a characterization of a real-world assembly task. To produce HRI methods that are both natural and relevant to the desired task, we conducted a detailed analysis of human-human teams executing an assembly task. This analysis produced a lexicon of gestural communication, providing a template of *what* terms needs to be communicated and *how* these terms are expressed through gesture, and the frequency of each term-gesture pair. Finally, to verify our results, we implemented gestural communication on a robot arm in collaboration with a human and evaluated the efficacy of the communication. This study validated our gestural lexicon and gave insight into the application of human-human communication to human-robot tasks.

Our methodology builds upon previous investigations of gestural communication. Experiments where humans demonstrate gestures have been used to great advantage by

¹ LOW COst roBOT co-worker, <http://www.locobot.eu/>

other researchers and have produced valuable user-generated gesture sets, e.g., [7], [8]. In these previous studies, the experimenters knew *a priori* what terms needed to be communicated and provided participants with a list of these terms. Our study, however, did not start from an existing list of terms but from a set of challenge tasks, derived from industry examples. Following the methods used in a study of touch-screen interaction, [9], we presented participants with tasks to complete and we observed the resulting communication, producing two important results: a list of terms needed to accomplish the tasks and the lexicon of gestures used to communicate these terms. Because we desire communication methods for two-way HRI, it is important that our gestures can be both used and understood by humans. We conducted our experiments using pairs of participants, obliging participants to communicate in a way that could be understood by their partners. In this, we were inspired by an earlier study using human-human pairs in a similar, but less formal, methodology, [10].

The key contributions of our work include the following:

- Characterization of industry needs and a useful abstraction of an industrial assembly task
- An observationally-derived lexicon of communication terms and gestures for human-robot collaborative assembly
- Validation and evaluation of this lexicon through human-robot interaction with implications for communication design

II. CHARACTERIZATION OF A REAL-WORLD ASSEMBLY TASK

In this work, we follow a use-case approach to ground and focus our efforts in a real, high-value manufacturing task while aiming for broad applicability, beginning with a study of the needs of workers in advanced manufacturing plants.

A. Methods

Our understanding of human-robot interaction in manufacturing assembly was developed through two visits to automotive assembly plants and interviews with an industry partner and subject matter experts, including manufacturing engineers, researchers, and manufacturing floor managers. We identified key requirements and constraints for a manufacturing RA system and identified a representative assembly task to use as a basis for task modeling. While our information comes from the automotive industry, many of the perspectives presented here are applicable to a range of industrial assembly applications.

We reviewed process documents related to the chosen task (vehicle door assembly), classifying and coding worker actions into generalized sub-task categories. For one view of the impact of the various sub-tasks, we tallied the occurrences of each sub-task. The most frequent subtasks were then abstracted into simple actions that could be implemented in laboratory work.

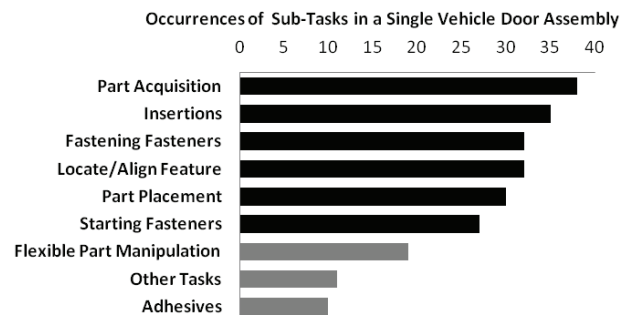


Fig. 1. Classification of operator actions in a representative assembly task. Black bars indicate sub-tasks which were abstracted in our experiments.

B. A Representative Task with Potential for Robotic Assistance

Industry experts involved in our research suggested that we focus on HRI methods for manufacturing assembly, where an RA could add significant value. They stated that while many assembly tasks are currently impractical to fully automate, an RA could be of great value to human workers in these situations. An effective RA would advance process quality and reliability and also reduce the ergonomic stresses on the human worker. They specifically suggested the exemplar use case of assembling a vehicle door. Representative, complex, dynamically changeable and labor-intensive, it is an appropriate place to implement and evaluate an RA.

C. Abstraction of the Assembly Task

The results of our task analysis and classification are shown in Fig. 1. It was possible to classify the various sub-tasks required to assemble a vehicle door into a relative small number of categories, with part acquisition (selecting and fetching parts) and part insertions (peg-in-hole type tasks) being the most common. We further generalized these sub-tasks into a set of three abstract actions that capture the essential features of the sub-tasks:

Part Acquisition is the most common sub-task executed by workers in our example case. This includes the selection of appropriate parts, locating these parts in the supply area, acquiring the parts and bringing them to the work area. The experts we interviewed identified part acquisition as key for RA intervention. In assemblies with multiple variations and options, remembering correct part combinations and quickly locating parts in the supply is challenging for human workers. Additionally, the physical acquisition and transport of parts requires a significant amount of worker time and is a potential source of unwanted ergonomic strain.

Part Manipulation encompasses several important sub-tasks, including part insertions, feature location and alignment, general part placement and starting (threading) fasteners. In all of these operations, the worker is required to move a part into a specific location and orientation on the assembly. In most cases, the industry experts felt that part manipulation is best done by human workers. Manipulation tasks often require

qualities that are lacking in current robotic systems: a high degree of manual dexterity, acute visual and tactile perception, and the ability to reach into tight spaces.

Part Operations include all actions done to a part once it has been placed: for example, securing fasteners (tightening bolts, snapping pres-fit fasteners) and seating electrical connectors. Some part operations, such as tightening bolts, may be best accomplished by an RA, ensuring precision torque and avoiding ergonomic stress on the human, while others, e.g. manipulating electrical connectors, require the dexterity of human workers.

With these three abstracted action categories, we capture the essential elements of the most common assembly sub-tasks. These categories form a set of simple, abstract actions that are appropriate for laboratory work while being firmly grounded in industry practice and needs.

D. Communication Constraints

The industry experts we interviewed stressed the importance of intuitive, non-verbal methods for communicating with an RA. In noisy industrial conditions, reliable verbal communication is difficult for humans, and impractical for machines. We therefore focus our efforts on gestural expression as a means of communication. Gestures are a powerful, non-verbal means of natural and intuitive communication [11].

III. UNDERSTANDING NATURAL GESTURES

In order to design naturalistic and intuitive HRI, we conducted a human observation experiment to determine what *terms* need to be communicated and the *gestures* most naturally used to communicate these terms.

A. Methods

We recruited sixteen participants (8 male, 8 female; average age 26.2; most from an engineering background) to work in pairs on two simulated assembly tasks. To stimulate communication between participants, we imposed task-by-task variations in which one participant was instructed to slightly alter the task execution (see III.A.2). Communication was restricted to either verbal or gestural communication in two separate test conditions. In both conditions, participants were permitted to communicate through touch. Participants covered their face and eyes, (Fig. 2), because industrial technology for communication through facial expression and eye gaze is not yet available. Communication between participants was observed during experiments and video recorded.

1) Experimental Tasks

Based on the analysis and abstraction described in Section II, we designed two experimental tasks to capture key elements of industrial assembly tasks: a part assembly task and a fastener insertion task, depicted in Fig. 3. These experimental tasks were designed using input from industry experts to simulate real-world tasks in which an RA could be beneficial.

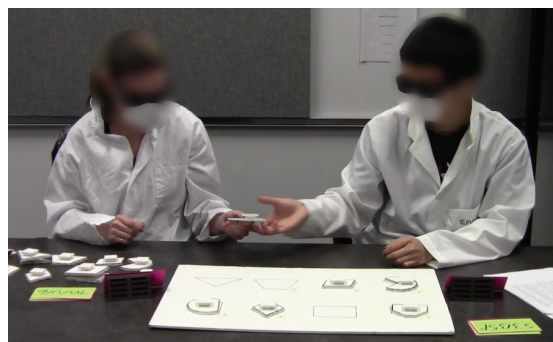


Fig. 2. Human-human experiment setup.

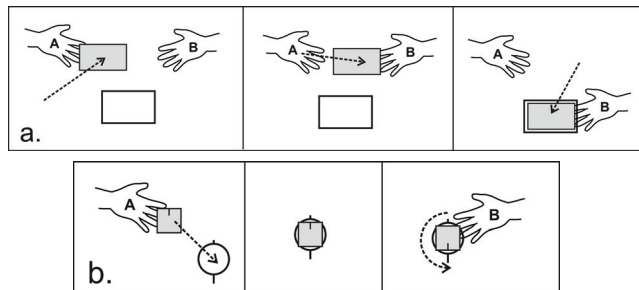


Fig. 3. Tasks in human-human experiment.
a. Part Assembly b. Fastener Insertion

a) Part Assembly Task: RA in Support Role

In an envisioned human-robot part assembly task, the RA provides correctly sequenced parts to a human worker, while the human worker performs the assembly. This is an example of an RA in a support role, in which the human worker executes all assembly tasks while the RA provides support to optimize the worker’s performance. In our experiment, Person A (playing the role of the RA) selected parts from a supply and handed them to Person B. Person B (acting as the worker) placed the parts on a simulated assembly. These roles simulated the imagined RA-human division of labor and also forced the participants to communicate and work as a team. This task was designed to focus on the key assembly actions of *Part Acquisition* and *Part Manipulation*.

b) Fastener Insertion Task: RA in Collaborative Role

In an envisioned fastener insertion task, the worker threads fasteners while the RA tightens them. This is an example of an RA in a collaborative role, in which the worker and RA work together on the same assembly, each executing actions suited to their abilities. In our experiment, Person A (playing the worker) acquired simulated bolts from a supply and placed the bolts on a simulated assembly in an ‘unscrewed’ state. Person B (playing the RA) performed a tightening action on each bolt as they were placed on the assembly. Again, these roles simulated the imagined RA-human division of labor and also forced the participants to communicate and work as a team. This task contained elements of *Part Acquisition* and *Part Manipulation*, but was designed to focus on the key assembly action of *Part Operations*, in this case the tightening and loosening of screws.

We simulated the parts as two-dimensional shapes and used drawn outlines on a board to represent assembly locations, as shown in Fig. 2. The simulated bolts had orientation marks to indicate the screwed/unscrewed states (Fig. 3b).

2) Task Variations

We used task variations to stimulate communication between participants. We derived these variations from the abstracted action categories described in Section II. Participants executed five different types of task variations: omit a part, add an additional part, change the orientation of a part (or screw/unscrew a bolt), use a different part in place of a standard part, and swap the locations of two parts.

Instructions were given randomly and privately to only one pair member during a given task iteration, but in every case, the task variations required some action from both participants. This forced the participant with knowledge of the variation to communicate with her partner to execute the special instructions. Instructions were delivered on cards containing both graphic and text instructions.

Task variations appeared randomly either before or after the execution of each task. Instructions received before task execution simulated a variation in an assembly task, e.g., the assembly of a car with a particular set of options. Instructions received after task execution simulated an error correction.

An experimental block consisted of six iterations of a given task, comprising one iteration for each of the five task variations and one iteration without any task variation. Each pair of participants completed four experimental blocks, for a full-factorial exploration of task (Part Assembly or Fastener Insertion) and communication condition (Verbal or Gestural). The order of test blocks was balanced between participants.

3) Data Analysis

We utilized an open coding approach to distill communication *terms* and *gestures* from the participant interactions. We use *terms* to mean *what* idea was being communicated and *gesture* to mean *how*, physically, the participant represented the given term.

A coder (one of the authors) first reviewed video of all verbal test conditions. Extraction of communication terms from the verbal record was straightforward; we found that the participants said, in explicit language, what they were trying to communicate. We formed a list of terms by collecting and coding the participants' statements, e.g., statements such as "Put the rectangle in spot #2" or "Place that in the top-left corner" were collected with other semantically similar expressions to form the term "Place Part". Similarly, we produced a list of gestures (the physical actions, not their meanings) used by participants under the gestural communication test condition.

The coder then reviewed the gestural communication to connect gestures to terms, that is, to map physical actions onto meaning. The result was a transcript of the gestural communication, written in codes of gesture and term.

This process of analyzing first verbal and then gestural communication minimizes subjectivity and made incorrect

coding rare, although it was still possible for the coder to miss or fail to observe some gestures. Failing to observe a gesture would affect the reported usage frequencies, but would not alter our lexicon. To validate our results, a second coder analyzed a random subset (25%) of the gestural video data. Inter-coder reliability was good (Cohen's Kappa for terms and gestures = 0.76 and 0.71, respectively). If we exclude missed gestures and consider only instances where coders disagreed, the inter-coder reliability is excellent (0.97 and 0.96, respectively).

B. Results

We coded 276 communication events, which we reduced to nine basic gestures (depicted in Fig. 4) and eleven different terms. Fig. 5 summarizes our results as a graphical lexicon, listing terms, gestures, and their relationships; line widths indicate the strength of association between terms and gestures, as measured by the relative frequency with which each gesture was used to represent each term. While the term usage frequencies shown in Fig. 5 are influenced by the design of the experimental tasks, these data give a general sense of which communication terms are likely to be used most often in interactions. Note that the usage frequency is not necessarily a measure of importance. The measure of gesture importance remains an open question for this work.

While we coded and analyzed all gestures and terms, we show only those that were observed more than twice. Infrequently used communication terms include 'Clear Workspace', 'Don't Place Part', 'You' and 'Me'.

An important characteristic of the observed communication is that it generally employed only a single hand and arm, even though this was never a constraint. In a few cases, participants attempted to use eye contact (through head movement) as part of their communication (despite the dark glasses), but this was rare; in only 7.2% of cases did the communicator attempt eye contact, and eye contact was reciprocated in only 2.9% of communication events. In addition, touch-based communication was almost never used, despite participants being explicitly instructed that they could use touch. Touch was used in only two instances, where one participant tapped his partner on the shoulder to get her attention.

In addition to communication about task execution and part manipulation, we also observed a more subtle form of communication used to control the flow of the task. Participants used flow-control communication to regulate turn-taking, to make general statements and inquires regarding the state of the task, and to demand or pass control. This communication often involved eye contact and gestures of the head, face and shoulders. Our analysis of this flow-control communication is ongoing and beyond the scope of this paper. We will present our detailed findings in a future work.

C. Discussion

1) Gestures are Physically Simple

The gestures we observed are all physically simple. With the exception of the two-handed 'swap gesture,' all gestures

required only one hand to execute. None of the gestures used by participants to communicate about part manipulation involved non-hand body language, except for the few instances when participants attempted eye contact along with a hand gesture. While many of the gestures involved articulating the fingers, only the ‘indicator position’ gesture (Fig. 4, center) relied on a specific finger pose. All other gestures were executed with variations or simplification of finger position. For example, a ‘Point at Part’ gesture could be executed with a single finger, with all fingers held together, or with a tool or part held in the hand.

The physical simplicity of these gestures makes them particularly suitable for implementation on an RA. Because only a single arm is required for gesturing, communication could be achieved with existing “one-arm” industrial robots. Furthermore, the lack of facial expression or subtle body language greatly simplifies the task of recognizing the gestures of the human operator. While our experimental tasks mimicked the complexity of real-world assembly tasks, we acknowledge that more complex human-robot tasks may require more complex and subtle gestures.

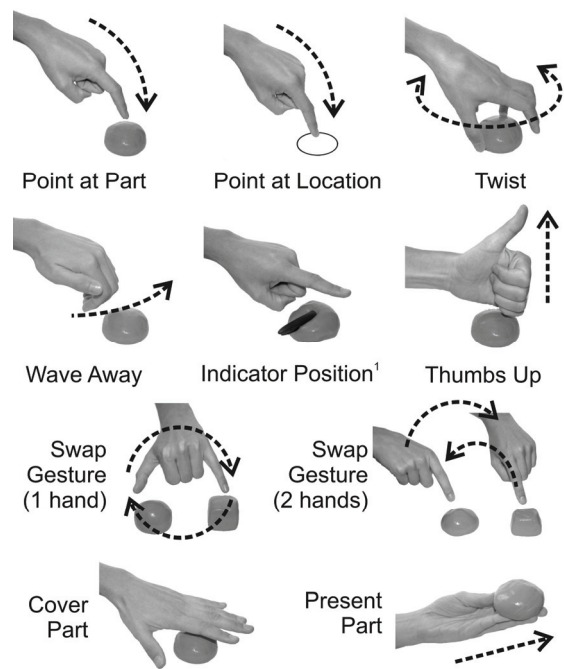
2) A Simple Gesture/Term Set Communicates Diverse Ideas

A notable feature of the gesture lexicon depicted in Fig. 5 is the relatively small number of terms, and the even smaller number of gestures. Each of the 96 task iterations observed in the study required some unique communication, and communication schemes were independently developed by 8 teams for two different tasks. Still, participants settled on the same basic gestures, mapped to the same terms, to meet their diverse communication needs. This suggests that these gestures have a certain broadly understood interpretation and that people are not only able, but tend naturally, to communicate a wide range of ideas with a simple set of terms and gestures.

This result is encouraging for HRI design, as an RA need only be capable of executing and recognizing a small set of gestures. This lessens the burden on those designing and implementing the gestures and makes the task of visual gesture recognition somewhat simpler.

3) Two Basic Gesture Structures

In the observed human-human interactions, we observed two basic ‘sentence structures’ for gestural communication: communications that specified both an object and an action, and those which specified only an object and relied on an implied understanding of the desired action. Examples of the object-action structure include two-gesture combinations (e.g., pointing to a part, and make the ‘thumbs up’ gesture, meaning, ‘this part here: unscrew it’) and single gestures that capture both the object and the action (e.g., the ‘twist’ gesture, when executed over the desired part. The gesture indicates the action, and the location of the gesture indicates the object). An example of the object-(implied action) form is a simple pointing gesture meaning ‘unscrew this part’. Here, the communication specifies only the part, but it was generally possible for the participant to infer what action she should do.



¹Finger used to indicate the desired position of a salient feature on the part.
 Fig. 4. Depiction of gestures used in our lexicon. Typical hand positions shown, but variations were observed.

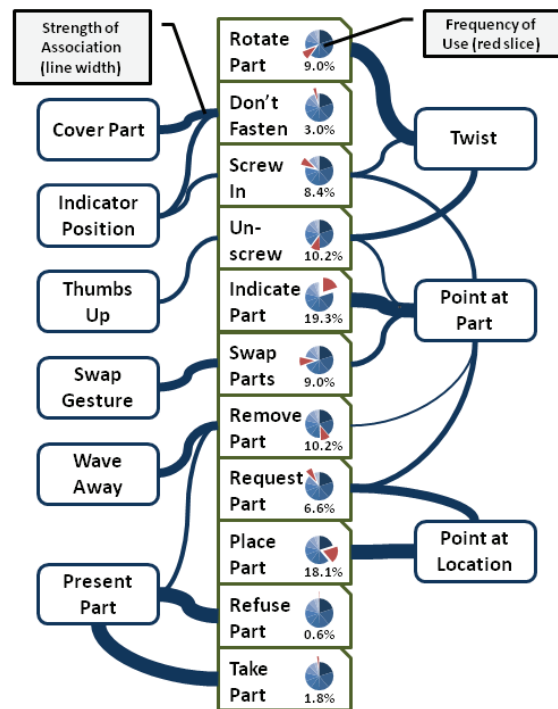


Fig. 5. A lexicon of communication terms and gestures.

This result has mixed implications for HRI. The object-action form of communication is well structured for machine understanding, and contains all of the information necessary to act. The object-(implied action) form is less tractable, and would require the RA to estimate the intent of the human operator based on ambiguous communication plus the state of the task. While an interaction designer could stipulate that

operators use the object-action form, the most natural HRI allow for either communication structure.

4) *Interpretation of Gestures is highly Context Sensitive*

As can be seen in Fig. 5, many of the gestures are associated with more than one communication term. In the most extreme case, participants used ‘Point to Part’ to communicate 6 distinct terms, with the intended meaning determined by the task context. In each case, the target of the communication would have to fully understand the context of the task and the intentions of her partner in order to correctly interpret the gesture. Our findings on the importance of context in gesture interpretation are in accordance with prior work, for example, [10].

The context dependence of gesture interpretation provides a challenge for HRI. It requires the RA to fully understand the state of the task, the array of possible actions at any given time, and maintain a model of the human operator’s intentions.

5) *Gestures are Highly Intuitive and Broadly Applicable*

One of the primary goals of our research was to generate a lexicon of gestures that is both intuitive and useful beyond the bounds of the experimental task. In that goal, the experiment appears to have succeeded. Of the 9 gestures in our lexicon, only one was symbolic and specific to the experimental task: the ‘Indicator Position’ gesture. It is significant that participants selected a salient feature on a part and formed a symbol based on this feature, indicating that such task-specific symbols could be useful in other tasks, even though such symbolic gestures are not inherently intuitive.

The remaining 8 gestures are far more intuitive. The gestures ‘Twist’, ‘Wave Away’, and ‘Thumbs Up’ are easy to interpret because they indicate the desired motion of the part. The gestures ‘Cover Part’ and ‘Present Part’ almost force the action they are trying to communicate, preventing the partner from acting, and thrusting a part directly at the target, respectively. The remaining gestures are all built around generic pointing gestures.

In addition to being intuitive, the gestures in our lexicon are applicable beyond the specific tasks used in our experiment. Our experiment was designed to address the fundamental actions used in industrial assembly tasks. The gestures resulting from our study could be used for many real-world tasks composed of these fundamental assembly actions. An exception might be the gestures relating to part operations; our experiment only involved part operations relevant to threaded fasteners. While the threading of fasteners is common, other part operations, such as press-fitting snap fasteners or making electrical connections, would likely inspire different gestures specific to those operations.

However, the context-sensitivity of gestural communication does limit the applicability of our lexicon. Because the interpretation of gestures is dependent on task specifics, it will be necessary to examine each new task and form a new understanding of the relationship between task state, gesture, and the intended meaning.

IV. VALIDATION

The final stage of our exploration was a small, focused study to verify our lexicon, to show that the gestures could be correctly interpreted in context, and to show that gestures inspired by humans could be successfully translated to a robot.

A. Methods

In this verification phase of our experiment, 12 subjects (6 female) were asked to interpret and evaluate robot and human gestures seen in video recordings of a human-robot team working together on a simplified version of the part assembly task used in the human-human study, as shown in Fig. 6. As discussed above, the meaning of a given gesture is highly sensitive to context and the task. Therefore, we showed the gestures in context, as part of a task, and we only recruited participants who were familiar with the task through their participation in the earlier human-human study.

1) *Physical Setup*

Because we intend our results to be applicable to industrial applications, we limited our robot to conditions that could be expected in a near-term industrial environment, i.e., non-anthropomorphic, with limited degrees of freedom, and without articulated fingers. We implemented our gestures on a 7-DOF Barrett WAM arm (Fig. 6), but restricted ourselves to only the first 6-DOFs (the second wrist rotation was inactive). Recognizing that the end effector of an industrial robot assistant could take many forms, such as a 2-fingered gripper or a tool (e.g. a torque wrench), we chose not to rely on any expressive features of the end effector. Instead, we placed a non-actuated stuffed glove on the end of the robot, proving a compliant surface for manipulating the parts and making the end effector highly visible in the video recordings. The robot manipulated the experimental parts using simple Velcro pads on the robot’s hand, on the parts, and on the assembly board.

2) *Gesture Implementation*

This verification study tested all term and gesture combinations from the Part Assembly Task with one exception: the two-handed swap gesture could not be implemented on the one-armed robot.

Robot gestures were implemented by recording manually generated motions and tuning the resultant trajectories. For each gesture, as well as the movements necessary for the manipulation of the parts used in the task, we manually moved the passive robot arm through the desired motions and recorded the joint trajectories. We then manipulated these



Fig. 6. Human-robot experiment setup.

trajectories in MATLAB, speeding them up to a natural pace, passing them through a low-pass filter to smooth the motions, and removing any unintended pauses. Gestures and motions were made to resemble, as closely as possible, the human gestures recorded during our earlier study.

We note that other researchers have developed more complex methods of producing human-inspired robot gestures, such as motion-capture of human actors [12] or automated imitation of human behavior [13]. While such methods have been shown to be effective, we aimed to establish a gesture production method that is easier to implement in industry. Ideally, such a method would not require any specialized equipment or expertise, and would be fast and simple.

3) Experimental Task

Each participant watched video recordings of a human and a robot executing a 3-part version of the part assembly task (see Section III.A.1). At some point in each recording, either the human or the robot used a gesture to issue a command, after which the video would end. After each gesture, the participants answered three questions: 1. *What should the robot/human do* (in response to the gesture)? 2. *How easy was it for you to understand this gesture* (on a scale of 1-7)? 3. *How natural was this gesture* (on a scale of 1-7)? Participants were also given the opportunity to provide comments.

An experimental block comprised one iteration of each gesture, executed by either the robot or the human, randomly ordered. Each participant viewed one block of robot gestures and one block of human gestures, with the order balanced between participants.

4) Data Analysis

We classified the responses to our first survey question (*What should the robot/human do?*) as Correct, Partially Correct, or Incorrect. Responses were classified as Partially Correct when the participant gave the correct interpretation of the gesture along with some qualification, uncertainty, or alternate response. For example, when the Wave Away gesture was used to indicate that the circle piece should be removed, one participant responded, “Remove the circle from the board. (or maybe remove all the parts?)”. Our classification of all responses was verified by a second coder, with excellent agreement (Cohen’s Kappa = 0.91).

B. Results

The results of our human-robot gesture recognition study are shown in Fig. 7 and Fig. 8. Fig. 7 displays the number of participants who identified each gesture correctly. Fig. 8 shows the mean ratings of naturalness and ease of interpretation for each gesture, on scale of 1-7 with 7 being the most easy/natural.

C. Discussion

1) Robot Gestures were interpreted with High Accuracy and evaluated as Natural and Easy to Interpret

The robot gestures proved to be effective, with participants correctly interpreting most robot gestures and rating them as both easy to interpret and natural. This result indicates that the

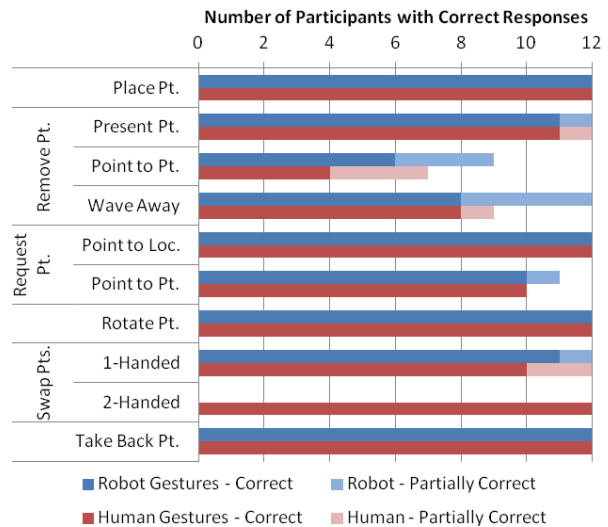


Fig. 7. Correct gesture interpretation for gestures rendered by a human (red) and a robot (blue). Recognition rates are generally high for all gestures.

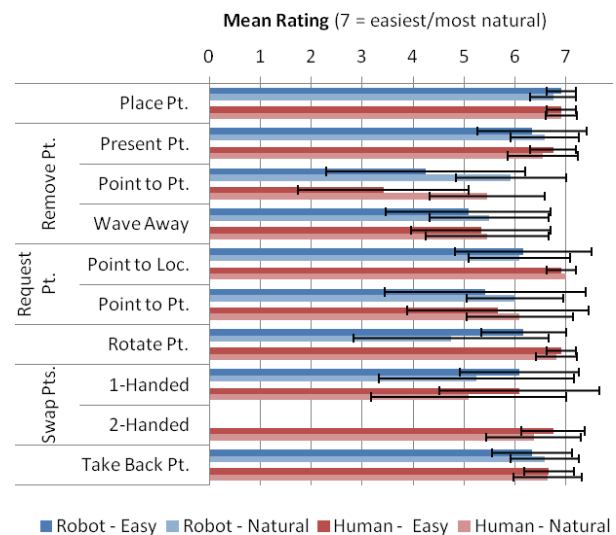


Fig. 8. Ease of interpretation and naturalness for human and robot gestures. Black error bars show standard deviations.

lexicon of gestures and terms produced in our human-human study can be directly applied to human-robot communication.

Also notable is the similarity in communication performance of the human and robot gestures. In all metrics, the performance rating of the robot gestures was statistically indistinguishable from that of the human gestures: In all cases, $p > 0.05$. While this does not mean that the human and robot gestures performed identically, the close similarity in performance leads us to conclude that the gestures were effectively rendered by the robot, despite the limitations of the hardware. We note, however, that ratings may have been skewed by participants’ expectations; participants’ comments indicated that they may have had lower standards for the performance of the robot, for example, “This gesture somehow seemed more natural and clear than the human

equivalent,” and “The gesture is artificial for a human but natural for a robot.”

The exceptions to these generally positive results are two of the gestures used to communicate ‘Remove Part’: the ‘Point to Part’ gesture and the ‘Wave Away’ gesture, and to a lesser extent, the use of ‘Point to Part’ to request a part. It is notable, however, that the performance of these gestures-term pairings was similarly low for both the robot and the human implementations, suggesting that the problem was with the communication form itself, not the robotic implementation of the gestures.

2) Subject-Action Gestures are more Effective than Subject-(Implied Action) Gestures

Those gestures that were not accurately interpreted all left some part of the intended action unspecified, generally taking the subject-(implied action) form. In many cases, participants correctly interpreted the implied portion of the communication, but were unconfident in their interpretations, as in the following example: “The gesture was very natural for indicating “this piece”, but it wasn’t really clear what the robot wanted the human to do about that piece. Because of the task context, I figure it wants another square. But the gesture itself doesn’t capture that without the context.” This and other comments confirm our earlier conclusion regarding the importance of task context in gesture interpretation, and indicate that it may be difficult to infer meaning from the task context in some cases. Therefore, we conclude that in cases where the task context allows for some ambiguity in interpretation, fully specified subject-action gestures should be used, even though such gestures may be more time consuming.

3) Human-Human Gesture Frequency is an Indication of Gesture Efficacy

The results of our validation study suggest that the potential value of a given gesture-term pair can be predicted by the popularity of the communication pattern in human-human trials. In general, the most misunderstood gesture-term pairs and those with the lowest ‘easiness’ and ‘naturalness’ ratings were those used least frequently in the human-human study. For example, ‘Point to Part’ had the lowest relative usage (12%) of all gestures used to communicate the term ‘Remove Part’, and this gesture-term pair was also the most frequently misunderstood in our human-robot study and received the lowest ratings from participants. Our results suggest that gesture-term pairing frequencies in human-human studies can be valuable guides to determine which gestures should be selected to communicate a given term in human-robot systems.

V. CONCLUSIONS AND FUTURE WORK

In this study we produced a gestural lexicon from the observation of human interactions and validated the lexicon in human-robot trials. Our study is based upon an analysis of real-world needs, ensuring the relevance of our results. The methods we use here could be reproduced to design gestural HRI for many different tasks.

Our ongoing work focuses on understanding the communication used to regulate task flow and turn-taking and on the implementation of this communication on both anthropomorphic and non-anthropomorphic robots. With the addition of flow-control gestures, we plan to implement a complete HRI system for industrial assembly, in conjunction with collaborators currently working on robotic hardware, vision, control and safety.

ACKNOWLEDGMENTS

Our thanks go to Chris Parker and the GM staff, including Roland Menassa, James Wells, Irene O’Hara, Ian Soutter, Ivan Novak, Rebecca Estoesta, and Steve Quinn.

REFERENCES

- [1] C. Heyer, “Human-robot interaction and future industrial robotics applications,” in *Int’l Conf. on Intelligent Robots and Systems*, 2010, pp. 4749–4754.
- [2] S. Haddadin, M. Suppa, S. Fuchs, T. Bodenm, A. Albu-Schaeffer, Hirzinger, and Gerd, “Towards the Robotic Co-Worker,” in *Robotics Research*, C. Pradalier, R. Siegwart, and G. Hirzinger, Eds. 2011, pp. 261–282.
- [3] S. Haddadin, A. Albu-Schaeffer, A. De Luca, and G. Hirzinger, “Collision detection and reaction: A contribution to safe physical Human-Robot Interaction,” in *Int’l Conf. on Intelligent Robots and Systems*, 2008, pp. 3356–3363.
- [4] M. Giuliani, C. Lenz, T. Müller, M. Rickert, and A. Knoll, “Design Principles for Safety in Human-Robot Interaction,” *Int’l J. of Social Robotics*, vol. 2, no. 3, pp. 253–274, Mar. 2010.
- [5] M. A. Goodrich and A. C. Schultz, “Human-Robot Interaction: A Survey,” *Foundations and Trends in Human-Computer Interaction*, vol. 1, no. 3, pp. 203–275, Feb. 2007.
- [6] T. Fong, J. Scholtz, J. A. Shah, L. Fluckiger, C. Kunz, D. Lees, J. Schreiner, M. Siegel, L. M. Hiatt, I. Nourbakhsh, R. Simmons, B. Antonishek, M. Bugajska, R. Ambrose, R. Burrige, A. Schultz, and J. G. Trafton, “A Preliminary Study of Peer-to-Peer Human-Robot Interaction,” in *Int’l Conf. on Systems, Man and Cybernetics*, 2006, pp. 3198–3203.
- [7] J. O. Wobbrock, M. R. Morris, and A. D. Wilson, “User-defined gestures for surface computing,” in *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09*, 2009, p. 1083.
- [8] S. Yohanan and K. E. MacLean, “The Role of Affective Touch in Human-Robot Interaction: Human Intent and Expectations in Touching the Haptic Creature,” *Int’l J. of Social Robotics*, vol. 4, no. 2, pp. 163–180, Dec. 2011.
- [9] M. Micire, M. Desai, A. Courtemanche, K. M. Tsui, and H. A. Yanco, “Analysis of natural gestures for controlling robot teams on multi-touch tabletop surfaces,” in *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces - ITS ’09*, 2009, p. 41.
- [10] T. Ende, S. Haddadin, S. Parusel, T. Wusthoff, M. Hassenzahl, and A. Albu-Schaeffer, “A human-centered approach to robot gesture based communication within collaborative working processes,” in *Int’l Conf. on Intelligent Robots and Systems*, 2011, pp. 3367–3374.
- [11] R. Hari and M. V. Kujala, “Brain Basis of Human Social Interaction: From Concepts to Brain Imaging,” *Physiological Reviews*, vol. 89, no. 2, pp. 453–479, 2009.
- [12] D. Matsui, T. Minato, K. F. MacDorman, and H. Ishiguro, “Generating natural motion in an android by mapping human motion,” in *Int’l Conf. on Intelligent Robots and Systems*, 2005, pp. 3301–3308.
- [13] M. J. Mataric, “Getting humanoids to move and imitate,” *IEEE Intelligent Systems*, vol. 15, no. 4, pp. 18–24, Jul-2000.