hHRI 2013 Papers

Reviews of submission #135: "Gestures for Industry: Intuitive Human-Robot

Communication from Human Observation"

------------------------ Submission 135, Review 4 ------------------------

Reviewer: primary

Expertise

4 (Expert)

Summary

The paper describes a gesture set for human-robot collaboration on

assembly tasks designed by watching two people work together to solve the

tasks,

Meta Review

Asking people to indicate what gestures they would use to control a

computer or robot (or to give information to another person) can be an

effective method for identifying gestures that can be easily understood

by other people without significant amounts of training. As noted by the

reviewers, this type of methodology has been used before in HCI (Jacob O.

Wobbrock, Meredith Ringel Morris, and Andrew D. Wilson, "User-Defined

Gestures for Surface Computing," CHI 2009, Boston, MA, April 4-9, 2009)

and HRI (Mark Micire, Munjal Desai, Amanda Courtemanche, Katherine M.

Tsui, and Holly A. Yanco, Analysis of Natural Gestures for Controlling

Robot Teams on Multi-touch Tabletop Surfaces. ACM International

Conference on Interactive Tabletops and Surfaces, Banff, Alberta,

November 23-25, 2009). The authors should cite these papers (and read

them to see the methodologies used).

The paper presents interesting results, although there are several flaws,

as noted by all three reviewers. The authors should address how they

intend to revise the paper based upon the reviewer feedback in their

rebuttal.

It would also be interesting to see all of the gestures that were

performed, as well as their frequency counts. It is understandable that

one of a kind gestures would be omitted from the final set, but they

should not be omitted from the results in the paper.

The authors state that the gestures "generally employed only a single

hand and arm, even though single-limb communication was never a

constraint." However, the experimental set up in Figure 2 shows two

people sitting side by side, instead of across a table. It is highly

likely that two handed gestures would have been seen if the people were

across from one another -- the side by side seating likely created a bias

towards only using the arm next to the other person. Running another set

of experiments in the condition with people seated across from one

another would allow this hypothesis to be tested. (However, it would

need to be done for another paper, as the conference does not allow for

additional experiments to be conducted after the submission deadline and

added.)

Overall Rating

3 (Borderline: Overall I would not argue for accepting this paper.)

------------------------ Submission 135, Review 1 ------------------------

Reviewer: external

Expertise

4 (Expert)

Summary

The authors present a methodology for designing gestures to communicate

with robotic assistants that perform tasks in an industrial process. The

methodology consists in observing human-human communication via gestures

in such industrial processes, and construct a set of representative

gestures for the main tasks. The authors validate their results with a

follow-up study in which human participants gave feedback on such

gestures performed by a robotic assistant.

Strengths and Weaknesses

The article is clearly written and easy to follow, I would like to see it

in the conference. The methodology is very easy to replicate and I know

from past experience it leads to good gesture results, reflective of user

behavior. The idea is not new for HCI (see comment in the Related Work

section of this review) but, but its application to HRI is definitely

new and useful.

Soundness

Yes

Related Work

I want to point the authors to an existing methodology from HCI which

elicites gestures from users in response to showing the gesture effect

[\*] and which has found great success up to date. A comparison between

the authors' methodology for HRI and the one from Wobbrock et al. [\*] for

HCI, could be valuable to the HCI readers of the paper.

[\*] Jacob O. Wobbrock, Meredith Ringel Morris, and Andrew D. Wilson.

2009. User-defined gestures for surface computing. In Proceedings of the

27th international conference on Human factors in computing systems (CHI

'09). ACM, New York, NY, USA, 1083-1092. DOI=10.1145/1518701.1518866

Presentation

Yes

Suggestions

See the suggestion for the Related Work section of this review.

Overall Rating

4 (Probably accept: I would argue for accepting this paper.)

------------------------ Submission 135, Review 2 ------------------------

Reviewer: external

Expertise

3 (Knowledgeable)

Summary

This paper investigates potential gestures to be used by robots when they

act as direct interaction assistants to humans in industrial situations.

The paper first presents a task analysis that informed the design of

human-human partner assembly related tasks. The human-human study informs

the gestures to be implemented on the industrial robot, and a video based

evaluation for interpreting the robot gestures for the same assembly

tasks with a human partner is provided. The paper presents a number of

design conclusions along with a lexicon of communication of verbal terms

and gestures.

Strengths and Weaknesses

This is a very nice paper! It has just about everything I would like to

see in a paper. It has good foundational research, a sound problem

statement, good design conclusions, etc. The authors have done their

homework on understanding fundamental real-world tasks, understanding how

humans would do the task, and conducted a preliminary evaluation based on

the robot implementation. The results are novel and original. The domain

application is one not often seen in HRI - industrial applications.

Soundness

The ideas, task analysis and task identification are sound. The

evaluations have some minor issues, but there are very few perfect

papers.

Specific issues:

1. It is not clear until the second paragraph of section 3.2 that the

participants' faces are not covered, as is implied in the first paragraph

of Section 3.1. In printed version it is hard to discern if the

participants can see at all and the sentence in 3.1 can be interpreted to

mean that the participants are blind. There are, however, statements

prior to section 3.2 that imply the participants can see. This is a

confusing aspect until the dark glasses are mentioned. This point should

be made clear at the beginning.

2. Most evaluations that use coding involve multiple coders to ensure

proper coding. This is a limitation of this paper, but I would not

consider it to be significant.

3. The second evaluation of the human-robot team also has a couple of

limitations. The first limitation is that the participants were the same

as those who participated in the human-human study. There is some

justification for using the same participants, but it would be better to

use new participants. This is a minor limitation in this reviewer's mind.

4. A more significant limitation of the human-robot evaluation is that

the participants watch a video recording of the team and interpreted the

gestures. This evaluation would be significantly stronger if the

participants had completed the tasks with the robot and objective metrics

were collected regarding the participants' real-time interpretation of

the robot gestures. I am willing to overlook this weakness because this

paper offers so much!

Related Work

The literature review is adequate. One point that the authors do not

include in their paper is the impact of working/training with the robot

on a daily basis. There is literature related to training and repeated

interaction that can be used to justify the not so perfect implementation

of gestures on the robot. Presumably, with repeated interaction over a

long period of time, the human would learn to better interpret the

ambiguous gestures given the task context.

Presentation

Absolutely.

The only thing would be to make figures 8 and 9 larger, but that is not

very easy.

Suggestions

None. This paper does a nice job of cover a broad set of data collection,

motivating the problem, explaining results, and providing design

guidance.

Overall Rating

5 (Definite accept: I would argue strongly for accepting this paper.)

------------------------ Submission 135, Review 3 ------------------------

Reviewer: external

Expertise

4 (Expert)

Summary

The paper investigates robot assistants in collaborative work

environments. Specifically, it looks into gestural communication between

human workers and robots in assembly tasks. The paper is grouped into

three experiments. First, the authors look at existing work assembly

documents and coded for occurrences of sub-tasks. Then, using the top

two categories of occurrences, the authors create situations in which two

humans are required to cooperate in an two-dimensional assembly task.

From this data, they coded and generated a classification of the

gestures. They then showed these earlier participants video of a 6 DOF

robot performing similar tasks and had participants rated the movement

based on three questions.

Strengths and Weaknesses

The work is in the area of human-generated gestures sets and is therefore

a well explored area of research. I believe that the paper is relevant

and if executed correctly. Unfortunately, the authors have given the

reader few ways to validate or comment constructively on the results.

The weaknesses of the paper are significant. First, I would suggest that

the authors look closely at Cohen’s Kappa published by Jacob Cohen in

the journal Educational and Psychological Measurement in 1960. Any time

a data set is analyzed and coded, it is vital that the correctness of

this classification be considered. In this paper, there is no indication

that the coding scheme was tested for inter-rater agreement. This is

important since the experiment and results should be reproducible.

This problem is further compounded by the fact that coding happens in the

first two experiments and, therefore, the bias has the potential to be

summative and dramatically effect the third experiment.

By the third experiment, I was having a hard time parsing the results and

discussion. The task resulted in three questions being asked – 1) What

should the robot/human do? 2) How easy was it or you to understand the

this gesture? 3) How natural was the gesture? As far as I can tell, the

authors do not fully provide the data/results to these questions. They

“classified the participants interpretations as correct, partially

correct, or incorrect.” I am unsure how the authors can take a

questions like #2 (how easy was it to understand) and rate it as correct

or incorrect. If #2 was not included in the correctness measure, I am

unclear about where this question is discussed in the results. The

analysis by the authors related to the open-ended question #1 would

require some type of coding and cross validation. (Which like the other

two experiments is not provided.)

The reader is left to assume that Figure 9 represents the arithmetic mean

(and not the mean through some other probability distribution) and the

chart is left without standard deviation bars, making it relatively

meaningless for analysis. Since there is no tabular data in the text,

the reader is provided no way to independently validate the results.

As a result, the results and discussion points can only be considered

qualitative commentary to an experiment that cannot be independently

verified.

Soundness

The methods are relatively sound from a experimental design standpoint.

The authors should provide a lot more data on the participants

demographics, experimental procedures. I also do not support the

author’s justification for re-using the same participants for the third

study. The authors attempt to justify this by the need for context, but

I believe that this need could have been mitigated through descriptions

of context to the later study and therefore eliminate bias that would be

carried forward from earlier experiments.

The soundness is in serious question since details were missed in the

analysis that jeopardize the later two studies. Since the experiments

build on one another, it is vital that there are no analysis errors in

the earlier stages. Since the authors did not perform a kappa and have

not provided any background on the coding set, we have no way of

confirming or disputing the results for the first and second experiments.

We just have to trust that they coded the trials without bias. This is

further compounded by the lack of statistical significance in the

follow-on study. If there is a signal in the noise, the reader is left

with no way to recreate the experiments or evaluate the author’s

approach themselves.

Related Work

The authors seem to concentrate their prior work on robotics. I would

strongly suggest that they look at the user generated gesture set

research in the human computer interaction literature. This will help in

the experimental design and analysis portions of the research.

Presentation

It is well organized and despite lacking technical detail and

correctness, it is clearly presented.

Suggestions

I would politely request that the authors look at the other user

generated gesture literature and re-formulate their experiments. The

research is on the right track, but it is vital as scientists that we are

able to look at the experimental design, data, results, and discussion

with a critical eye and make informed analysis of the research. As it

stands now, the reader is left trusting that the authors did everything

correctly and without bias.

Overall Rating

2 (Probably reject: I would argue for rejecting this paper.)