The Role of the Timing between Multimodal Robot Behaviors for Joint Action

Kerstin Fischer
Institute for Design and Communication
University of Southern Denmark
DK-6400 Sonderborg
kerstin@sdu.dk

Lars Christian Jensen
Institute for Design and Communication
University of Southern Denmark
DK-6400 Sonderborg
larscj@sdu.dk

Stefan-Daniel Suvei
The Maersk Mc-Kinney Moller Institute
University of Southern Denmark
DK-5230 Odense M
stdasu@mmmi.sdu.dk

Leon Bodenhagen
The Maersk Mc-Kinney Moller Institute
University of Southern Denmark
DK-5230 Odense M
lebo@mmmi.sdu.dk

Introduction

That timing in HRI is important is clear ever since Suchman (1987) demonstrated the crucial role of timing for all interactions with technology. She showed that if system response is delayed, users, for whom the system behavior is not transparent, consider the lack of timely response as failure and initiate a new action, or the whole process may result in error altogether. Thus, users expect a timely response to their actions as a precondition for joint action.

In social robotics, much work concerns the timing of the robot's behavior with respect to the human's behavior (e.g. concerning gaze, cf. Mutlu et al. 2012, Fischer et al. 2013), yet the synchronization of robot behaviors such as movement of the body, speech and arm movement, for instance, has rarely been addressed.

To investigate the role of timing of multimodal robot behavior, we carried out an experiment in which the robot either employed its multimodal behaviors sequentially or synchronized and analyzed the effects of the timing on joint action.

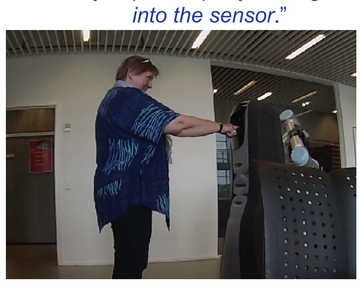
Results

The timing of individual actions plays a crucial role in how people understand how they should interact with the robot. The lack of synchronization of robot speech, movement and action in the first condition lead to confusion and insecurity on the part of the users.

Condition 1 - Sequential Actions:

Excerpt 1:

Robot says: "*please put your finger into the sensor.*"



"*here?*"
"*like that?*"

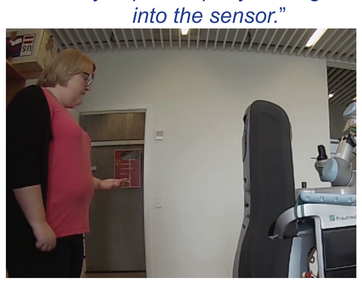
Robot extends arm with sensor.

"*no?*"
"*it's not the right one.*"

"*here?*"

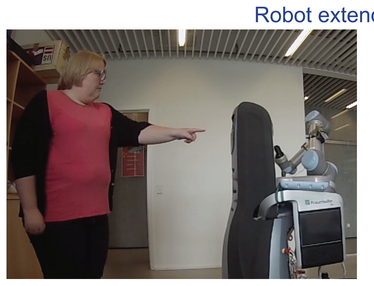
Excerpt 2:

Robot says: "*please put your finger into the sensor.*"

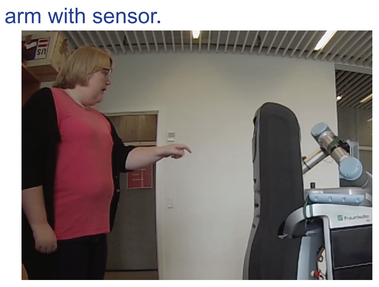


The participant takes out her and looks at it.

Robot extends arm with sensor.



The participant takes out her hand and ooks at it.



"*Aah!*" (participant redirects her finger at the arm being extended)

Excerpt 3:

Robot says: "please put your finger into the sensor."

Robot extends arm with sensor.

The participant looks searchingly to both sides of the robot.

"*Aah!*" (participant smiles while looking at arm being extended)

Condition 2 - Simultaneous Actions:

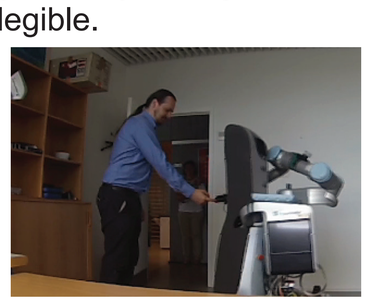
When the robot uses all of its modalities simultaneously, participants are much better at predicting the joint action.

Excerpt 4:

The participant realizes what he is supposed to do already halfway through the robot's approach. Thus, the multimodal behavior is more legible.

The robot starts to drive, says: "*please put your finger into the sensor,*" and extends it's arm with the sensor.

The participant holds up his finger as soon as the robot has made its verbal request.



The participant puts his finger into the sensor the moment the robot is close enough.

The results from the qualitative analysis are supported by the quantitative analysis. The analysis shows that people interacting with the robot using inappropriate timing show significantly more signs of confusion and insecurity ($p<.05$, Fisher's Exact Test).

Method

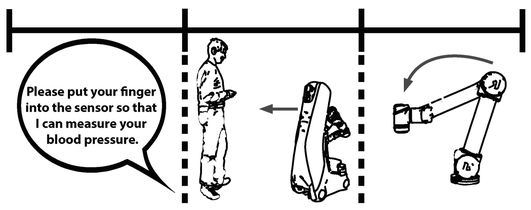
Experimental Conditions
Between-subject design

Condition 1:

1. Robot says: "Please put your finger into the sensor so that I can measure your blood pressure"
2. then drives up to the participant
3. and extends arm

Condition 2:

- speech, movement and gesture at the same time



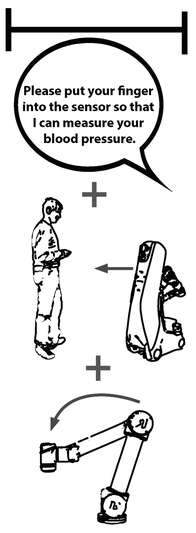


Figure 1 (top): Sequential performance of actions: when one action ends, the next one commences.

Figure 2 (left): Simultaneous performance of actions: the three modalities are performed at the same time.

Robot
Care-O-bot 3 with an ostensible blood pressure measuring device.

Participants

- university staff/students and people from the general public (24 m, 12 f)
- mean age 30.7 (SD=9.9)

Procedure

1. Participants signed a consent form
2. met the Care-O-bot in the lab
3. then filled out a survey
4. were shown around our department and interacted with other robots
5. they were then taken back to the Care-O-bot where the actual experiment took place
6. participants filled out a survey

Analysis

The focus of the current investigation is on participants' behavioral responses to the robot's actions and is analyzed by using multimodal conversation analysis of video recordings. The analysis is supplemented with a quantitative analysis, which was done by counting the visible signs of confusion and insecurity shown by the participants (e.g. stepping back, looking at the robot searchingly, hesitating, or asking for help).

Conclusion

The results show that the timing of robot multimodal actions play a crucial role even for a situation that requires as limited joint action as the one under consideration. In particular, we found that

- people assume that they should be able to carry out an instruction in the moment the instruction is uttered; thus speech needs to be carefully coordinated with the moment all preconditions for the human partner to carry out the required action are fulfilled;
- people process the robot's behavior incrementally and on the basis of partial information and start predicting the robot's actions on the basis of what is available at each given moment. This can lead to inappropriate proactive behavior if the robot's individual actions are not sufficiently legible.

This has consequences for the legibility of robot action since users don't wait for the whole action before making their predictions and acting proactively.
The design of legible robot behavior thus needs to take the timing between actions and processing in time into account if human-robot joint action is supposed to be successful.

Acknowledgements

We wish to thank Sibel S. Isikli-Kristiansen and Kristina Holz for their help during the experiments. Furthermore, we gratefully acknowledge the support of the SPIR project patient@home.

References

Fischer, K., Lohan, K., Nehaniv, C. & Lehmann, H. (2013). Effects of Different Types of Robot Feedback. International Conference on Social Robotics'13, Bristol, UK.

Mutlu, B., Kanda, T., Forlizzi, J., Hodgins, J., & Ishiguro, H. (2012). Conversational Gaze Mechanisms for Humanlike Robots. ACM Transactions on Interactive Intelligent Systems 1, 2, art. 12.

Suchman, L. (1987): Plans and Situated Actions. Cambridge: Cambridge University Press.

UNIVERSITY OF SOUTHERN DENMARK.DK