**2011 RO-MAN
20th IEEE International Symposium on Robot and
Human Interactive Communication,
July 31 - August 3, 2011, Atlanta, GA, USA**

A Friendly Gesture: Investigating the Effect of Multimodal Robot Behavior in Human-Robot Interaction

Maha Salem, Katharina Rohlfing, Stefan Kopp, Frank Joublin

Abstract— **Gesture is an important feature of social interaction, frequently used by human speakers to illustrate what speech alone cannot provide, e.g. to convey referential, spatial or iconic information. Accordingly, humanoid robots that are intended to engage in natural human-robot interaction should produce speech-accompanying gestures for comprehensible and believable behavior. But how does a robot's non-verbal behavior influence human evaluation of communication quality and the robot itself? To address this research question we conducted two experimental studies. Using the Honda humanoid robot we investigated how humans perceive various gestural patterns performed by the robot as they interact in a situational context. Our findings suggest that the robot is evaluated more positively when non-verbal behaviors such as hand and arm gestures are displayed along with speech. These findings were found to be enhanced when the participants were explicitly requested to direct their attention towards the robot during the interaction.**

I. INTRODUCTION

One of the main objectives of social robotics research is to design and develop robots that can engage in social scenarios in a way that is appealing and familiar to users. For this, social robots should provide communicative functionality that is natural and intuitive to their interaction partners. The appropriate level of such communicative functionality, however, strongly depends on the appearance of the robot and attributions hence made to it. Given the design of humanoid robots, they are typically expected to expose humanlike communicative behaviors, using their body for non-verbal expression in a similar fashion as humans. Representing an important feature of human communication, co-verbal hand and arm gestures are frequently used by human speakers to illustrate what they express in speech. Crucially, gestures help to convey information which speech alone cannot provide, e.g. as in referential, spatial or iconic information. At the same time, human listeners have been shown to be well-attentive to information conveyed via such non-verbal behaviors [4]. Thus, humanoid robots that are intended to engage in natural human-robot interaction (HRI) should generate speech-accompanying gestures for comprehensible and believable behavior. Moreover, providing multiple modalities helps to dissolve ambiguity typical of unimodal communication and hence increase robustness of communication.

Gesture is a phenomenon of human communication that has been studied by researchers from various disciplines for many years. A multiplicity of hand, arm and body movements can be considered to be gestures, and although definitions and categorizations vary widely, much gesture research has sought to describe the different types of gestures ([11], [8]). In this paper, we use the term gestures to refer specifically to referential gestures, i.e. movements representing the content of speech by pointing to a referent in the physical environment (deictic gestures) or representational gestures, i.e. depicting a referent with the motion or shape of the hands (iconic gestures). Other types of gestures, which are not considered in the present work, include non-representational gestures such as emblems (movements that convey conventionalized meanings), beat gestures (movements that are performed along with the rhythmical pulsation of speech without conveying semantic information), and turn-taking gestures (movements that regulate interaction between multiple speakers) [6].

To endow a humanoid robot with communicative co-verbal gestures, a large degree of flexible control with regards to shape properties of the gesture is required. At the same time, adequate timing and natural appearance of these body movements are essential to add to the impression of the robot's liveliness. Based on the previous implementation of a speech and gesture production model for humanoid robot gesture [14], we exploit the achieved flexibility in communicative robot behavior for two controlled experimental studies by evaluating what humans perceive from a humanoid robot performing gestures in a situational context. This way, we try to shed light onto human perception and understanding of gestural machine behaviors and how these can be used to design more natural communication in social robots.

II. RELATED WORK

Although much of the robotics research has been dedicated to the area of gesture recognition and analysis, only few empirical findings focusing on the generation of humanoid robot gesture together with the investigation of human perception of such robot behavior have been presented. Many existing models of gesture synthesis typically denote object manipulation fulfilling little or no communicative function, e.g. [3], and are often based on the recognition of previously perceived gestures (imitation learning), e.g. [2]. In many cases in which robot gesture is actually generated with a communicative intent, these arm movements are not produced at run-time but are pre-recorded and simply replayed during human-robot interaction, e.g. [16]. Moreover, a major-

The work described is supported by the Honda Research Institute Europe.

M. Salem is at the Research Institute for Cognition and Robotics, Bielefeld, Germany msalem@cor-lab.uni-bielefeld.de

K. Rohlfing is at the Emergentist Semantics Group, Bielefeld University, Germany kjr@uni-bielefeld.de

S. Kopp is at the Sociable Agents Group, Bielefeld University, Germany skopp@techfak.uni-bielefeld.de

F. Joublin is at the Honda Research Institute Europe, Offenbach, Germany frank.joublin@honda-ri.de

978-1-4577-1573-0/11/$26.00 ©2011 IEEE

ity of approaches focusing on gesture synthesis for humanoid robots are limited to the implementation and evaluation of a single particular type of gestures, typically deictic (e.g. [17], [13]) or emblematic gestures (e.g. [7]) instead of providing a general framework that can handle all types of gestures.

In the area of embodied conversational agents, there has been active work in developing and evaluating complex gesture models for the animation of virtual characters. Several recent studies have investigated and compared the human perception of such traits as naturalness in virtual agents. In one such study [10], the conversational agent Max communicated by either utilizing a set of co-verbal gestures alongside speech, typically by self touching or movement of the eyebrows, or by using speech alone without any such accompanying gestures. Human participants were then asked to rate their perception of Max' behavioral-emotional state, for example, its level of aggressiveness, its degree of liveliness, etc. Crucially, the results of the study suggested that virtual agents are perceived in a more positive light when they are able to produce co-verbal gestures alongside speech (rather than acting in a speech-only modality). In [1] Bergmann et al. modeled the gestures of Max based on real humans' non-verbal behavior and subsequently set out to question the communicative quality of these models via human participation. The main finding was that Max was perceived as more likeable, competent and humanlike when gesture models based on individual speakers as opposed to a collection of speakers or no gestures at all were applied.

Despite the interesting implications of these studies, findings from the domain of virtual agents cannot be easily transferred to social robots. Firstly, the presence of real physical constraints can alter the perceived level of realism. Secondly, given the greater degree of embodiment and shared interaction space, interacting with a robot is potentially richer. This makes the HRI experience more complex and is thus expected to affect the outcome of the results.

One of the few models that resembles our approach in that it attempts to generate a multitude of gesture types for a humanoid robot was presented by Ng-Thow-Hing et al. [12]. Their proposed model reconstructs the communicative intent through text and parts-of-speech analysis to select appropriate gestures. The evaluation of the system, however, was merely undertaken using several video-based studies.

However, we argue that gesture scope and space can only be accurately observed and assessed in a true interaction. Thus, to obtain a representative assessment of robot gesture and the human perception thereof, it is necessary to evaluate such non-verbal behavior in actual interaction studies. Generally, only few systematic experiments have been presented for the evaluation of robot gesture during HRI; most approaches to generating multimodal robot behavior are only evaluated in observational user studies with small numbers of participants. Moreover, many studies that are investigating the effect of robot gesture employ non-humanoid robots as research platforms, such as the penguin robot used in [16]. However, to examine humanlike gesturing behavior, the robot's level of embodiment plays a crucial role.

To contribute to a basic understanding of gestural machine behaviors and their effects on human perception, we decided to conduct two controlled experimental studies using our speech-gesture synthesis model implemented on the Honda humanoid robot. Our major objective is to systematically investigate whether multimodal robot behavior, i.e. displaying gesture along with speech, is desired by human interaction partners and favored over unimodal communication.

III. Implementation of a Robot Control Architecture

The generation of communicative co-verbal gestures for artificial humanoid bodies demands a highly flexible control with regard to shape and time properties of the gesture, while ensuring a natural appearance of the movement. Ideally, if such non-verbal behaviors are to be realized, they have to be derived from conceptual, to-be-communicated information. Since the challenge of multimodal behavior realization for artificial humanoid bodies has already been tackled in various ways within the domain of virtual conversational agents, our approach exploits the experiences gained from the development of a speech and gesture production model used for embodied virtual agents. In particular, we build on the Articulated Communicator Engine (ACE), which is one of the most sophisticated multimodal schedulers and behavior realizers by replacing the use of lexicons of canned behaviors with an on-line production of flexibly planned behavior representations [9]. Having implemented an interface that couples ACE with the perceptuo-motor system of the Honda robot, this control architecture is now used as the underlying action generation framework for the humanoid robot. It combines conceptual representation and planning with motor control primitives for speech and arm movements of a physical robot body. Details of the implementation can be found in [14].

Using the framework described above, we conducted two experimental studies to investigate how communicative robot gesture might impact and shape human experience in human-robot interaction. In the following sections, we describe our experimental method and the results that we obtained.

IV. Methodology

To learn about the effects of communicative robot gesture on human interaction partners, we conducted two between-subjects experimental studies using the Honda humanoid robot. For this purpose, a suitable scenario for gesture-based human-robot interaction was designed and benchmarks for the evaluation were identified.

Study 1: Unimodal vs. multimodal robot behavior in human-robot interaction

The study scenario comprised a joint task that was to be performed by a human participant in collaboration with the Honda humanoid robot. In the given task, the robot referred to various objects by utilizing either unimodal (speech only) or multimodal (speech and gesture) utterances, based on which the participant was expected to perceive, interpret and perform an according action.

1) Participation: In the first study, a total of 40 subjects (20 female, 20 male) participated in the experiment, ranging in age from 20 to 63 years (*M* = 31.3, *SD* = 10.55). All subjects were native German speakers who were recruited at Bielefeld University and had never participated in a study involving robots before. Based on five-point Likert scale ratings, participants were identified as having negligible experience with robots (*M* = 1.23, *SD* = 0.42), while their computer and technology know-how was moderate (*M* = 3.68, *SD* = 0.94). Participants were randomly assigned to the different experimental conditions while maintaining gender- and age-balanced distributions.

2) Experimental Design: The experiment was set in a kitchen environment in which the humanoid played the role of a household robot. Participants were told they were helping a friend move house and were tasked with helping to empty a cardboard box of kitchen items. The box contained nine kitchen items whose storage placement is not typically known a priori (unlike plates, e.g. which are usually piled on top of each other). Specifically, they comprised a thermos flask, a sieve, a ladle, a vase, an eggcup, two differently shaped chopping boards and two differently sized bowls. The objects were to be removed from the box and arranged in a pair of kitchen cupboards (upper and lower cupboard with two drawers). Given the participant's non-familiarity with the friend's kitchen environment, the robot was made to assist the human with the task by providing information on where each item belongs.

Conditions: We manipulated the robot's non-verbal behavior in two experimental conditions:

- In **condition 1**, the *unimodal (speech-only)* condition, the robot presented the participant solely with a set of verbal instructions to explain where each object should be placed. The robot did not move its body during the whole interaction; no gesture or gaze behaviors were displayed.
- In **condition 2**, the *multimodal (speech-gesture)* condition, the robot presented the participant with the identical set of verbal instructions used in condition 1, however, accompanied by corresponding gestures, to explain where each object should be placed. Simple gaze behavior supporting hand and arm gestures (e.g. look right when pointing right) was displayed during interaction.

Verbal Utterances: In order to keep the task solvable under both conditions, we decided to design the spoken utterances in a self-sufficient way. This means that the gestures used in the multimodal condition contained redundant information that was also conveyed via speech. Each instruction presented by the robot typically consisted of two or three so-called *utterance chunks*. Based on the definition provided by [9], each *chunk* refers to a single idea unit which is represented by an intonation phrase and, optionally in a multimodal utterance, by an additional co-expressive gesture phrase. The verbal utterance chunks used in our study are based on the following syntax:

- **Two-chunk utterance:**
  `<Please take the [object]> <and place it [position+location].>`
  Example: *Please take the thermos flask and place it on the right side of the upper cupboard.*
- **Three-chunk utterance:**
  `<Please take the [object],> <then open the [location]> <and place it [position].>`
  Example: *Please take the eggcup, then open the right drawer and place it inside.*

An example of a multimodal three-chunk utterance delivered by the robot is illustrated in (Fig. 1).

Gestures: In the multimodal condition, the robot used three different types of gesture along with speech to indicate the designated placement of each item:

- **Deictic gestures**, e.g. to indicate positions and locations
- **Iconic gestures**, e.g. to illustrate shape/size of objects
- **Miming gestures**, e.g. hand movement using a ladle or opening cupboard doors

Examples of the three gesture types are displayed in (Fig. 1).

Robot control and behavior: During the study, the Honda robot was partly controlled using a Wizard-of-Oz technique. This way, minimal variability in the experimental procedure was ensured to allow for a consistent and comparable interaction experience across all participants. The robot's speech was identical across conditions and was generated using the text-to-speech system MARY [15] set to a neutral voice. Speech recognition was not used during the experiment, again to minimize variability in the participants' experienced interaction. Instead, the experimenter initiated the robots interaction behavior from a fixed sequence of pre-determined utterances. Once triggered, a given utterance was generated autonomously at run-time. The ordering of this sequence remained identical across conditions and experimental runs.

In this study, the robot delivered each two-chunk or three-chunk instructional utterance as a singular one-shot expression without any significant breaks in the delivery process. Successive chunks indicating object, position and location were delivered contiguously in the manner of natural speech. Moreover, in the co-verbal gesture condition, gestures became confluent with the utterance process. Participants were instructed to indicate when they had finished placing an item and were ready for the following item by saying "next".

3) Hypothesis: Based on findings from gesture research in human-human as well as in human-agent interaction we developed the following hypothesis for gesture-based human-robot interaction:

Subjects who are presented with multimodal instructions by the robot (using speech and gesture) will evaluate the robot more positively than those who are presented with unimodal information by the robot (using only speech).

4) Experimental Procedure: Participants were first given a brief written scenario and task description to read outside the experiment room. They were then brought into the experiment room where the experimenter verbally reiterated

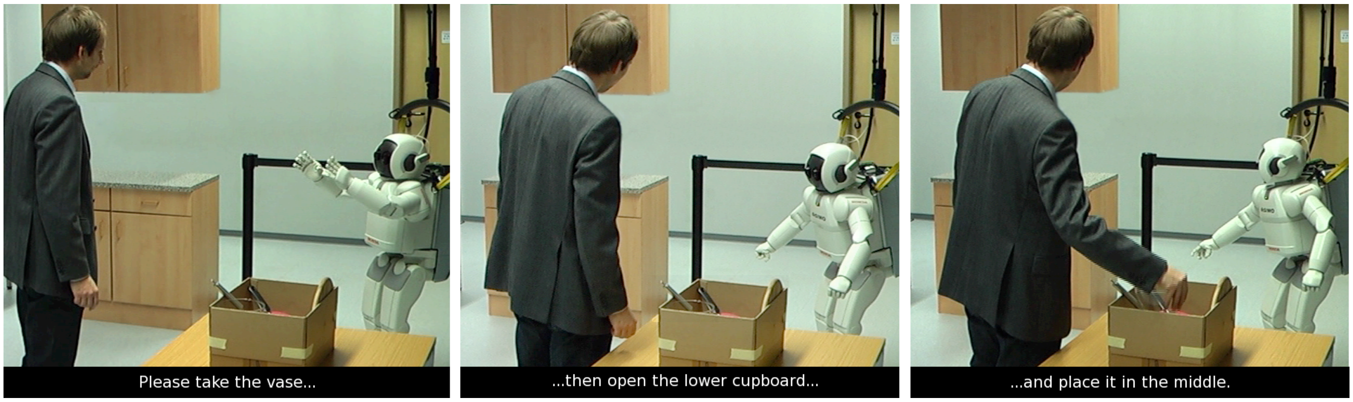
Please take the vase...
...then open the lower cupboard...
...and place it in the middle.

Fig. 1. Example of a multimodal three-chunk utterance delivered by the robot during interaction. Three different types of gesture are used (left to right): *iconic gesture* illustrating the shape of the vase; *miming gesture* conveying the act of opening the cupboard; *deictic gesture* pointing at designated position.

the task description to ensure the participants' familiarity and to give them the opportunity to ask any clarifying questions. The experimenter then left the participant to begin the interaction with the robot. At the beginning of the experiment, the robot greeted the participant and gave a verbal introduction to the task. It then presented the participant with individual utterances as described in the experimental design, each of which was triggered by the experimenter sitting at a control terminal. The participant attempted to follow the uttered instructions and place each item into its correct location. At the end of the interaction, the robot thanked the participant for helping and bid them farewell.

In the unimodal (speech-only) condition all utterances including the greeting and farewell were presented verbally; in the multimodal (speech-gesture) condition, all utterances including the greeting and farewell were accompanied by co-verbal gestures.

After completing the task, subjects filled out a post-experiment questionnaire that recorded their demographic background and, based on a five-point Likert scale, measured their affective state, evaluation of the task and interaction, and perception of the robot. Among other items, they were asked to rate the robot's appearance, naturalness, liveliness and friendliness. Upon completion of the questionnaire the participants were de-briefed and received a chocolate bar as a thank-you. The questionnaire data was collated and analyzed, the results are presented and discussed in the following.

5) Results: We assessed how the humanoid robot was perceived by participants using several items, e.g. 'active', 'communicative', 'competent', 'engaged', 'friendly', 'lively', and 'sympathetic' on five-point Likert scales with endpoints 1 = *not appropriate* and 5 = *very appropriate*. We conducted independent-samples t-tests with 95% confidence intervals (CI). On average, all qualities were rated higher, i.e. more positively, in the multimodal condition. At a significant level, participants assessed the robot as more active in the multimodal condition (M = 3.10, SD = 1.11) than in the unimodal condition (M = 2.35, SD = 0.88), $t(38)$ = -2.70, p = 0.005. Similarly, the robot was perceived as more lively when its utterances were accompanied by gestures (M = 3.12, SD = 0.97) than when it was only speaking (M = 2.52, SD = 0.84), $t(38)$ = -2.09, p = 0.02. Moreover, participants rated the robot using multimodal behaviors as more sympathetic (M = 4.20, SD = 0.95) than when it relied on unimodal communication only (M = 3.60, SD = 1.05), $t(38)$ = -1.90, p = 0.03. Significantly at the 10% level, the gesturing robot was rated as more competent (M = 4.26, SD = 0.87) than the robot that relied on speech only (M = 3.85, SD = 0.93), $t(37)$ = -1.43, p = 0.08. Mean values and statistical p-values for all measured characteristics are visualized in Fig. 2.

These results support our hypothesis and suggest that the inclusion of gestural behavior casts the robot in a more positive light than in the speech-only condition. The significantly higher rating of the characteristics 'lively' and 'active' in the gesture condition can be attributed to the robot's gestural movements. The rating of the characteristic 'sympathetic' suggests that humanlike non-verbal behaviors including gestures actually trigger a higher positive emotional response within the human participant. In practice, this could intuitively result in the participant responding to the robot more positively.

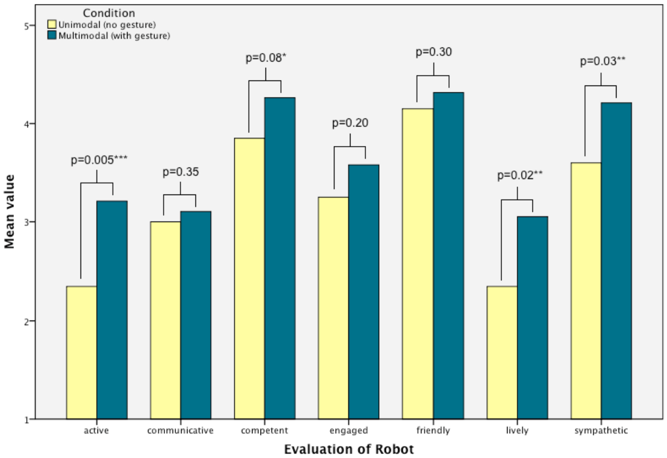

Fig. 2. Evaluation of the humanoid robot in Study 1, based on a 5-point Likert scale; *=p<0.10, **=p<0.05, ***=p<0.01.

Study 2: Enhancing the effect of robot gesture by increasing participants' attention towards the robot

In Study 1, it was often observed that participants immediately turned to the object being referred to by the robot within the first chunk of the utterance. In such cases the participants' attention typically shifted from the robot to the named object while the robot was still delivering the following chunk(s) of the instructional utterance. Such behavior is in line with findings from human gesture research, showing that addressees rarely gaze directly at the speaker's gesture, while typically spending as much as 90-95% of the total viewing time fixating the speaker's face [5]. Generally, this behavior displayed by participants during human-robot interaction can be viewed as a positive finding indicating that they interacted in a fairly natural way. However, participants of Study 1 frequently reported that they had difficulty in assessing the robot's behavior after completing the task, since they had not consciously paid attention to it. As a consequence, we decided to modify the design of the first study so that the participants' attention would be directed towards the robot for a longer period of time during the interaction.

1) Participation: We tested a total of 41 participants (21 female, 20 male), ranging in age from 20 to 61 years (*M* = 31.54, *SD* = 10.96), with similar preconditions to Study 1. Participants were again identified as having negligible experience with robots (*M* = 1.39, *SD* = 0.67) and moderate computer and technology know-how (*M* = 3.73, *SD* = 0.92).

2) Experimental Design: The general set-up, scenario and conditions in Study 2 were similar to the design of Study 1. However, in order to increase the participants' attention towards the robot, we decided to separate the utterances delivered by the robot into two parts. The first part comprised the object, i.e. the first chunk of a two-chunk or three-chunk utterance. The second part comprised the item's designated location and position, i.e. the second chunk of a two-chunk utterance or the second and third chunk of a three-chunk utterance. In the co-verbal gesture condition, the gestures maintained their synchronization with the verbal chunks, thus gestural behavior was effectively paused whenever there was a break in the delivery of the utterance.

Our primary motivation in splitting the utterances was to increase the participant's attention directed towards the robot. The second part of the utterance was only triggered once the participant had picked up the object from the box and had returned to stand in front of the robot, while directing their gaze at the robot in anticipation of the next instruction.

3) Hypothesis: Based on the findings from Study 1 we developed the following hypothesis for the utilization of split utterances in Study 2:

Increasing the participant's attention on the robot will result in an enhancement of the effects found in Study 1.

4) Experimental Procedure: The experimental procedure in Study 2 was almost identical to Study 1, with the only difference being the modified delivery of utterance chunks. Furthermore, the participants were not required to verbally indicate when they were ready for the robot to proceed with the next piece of information, but instead were asked to stand in front of the robot to trigger subsequent instructions. Finally, in contrast to Study 1 in which a more subtle and natural perception of the robot was desired, participants in Study 2 were explicitly asked to dedicate their attention towards the robot in the process of solving the given task.

5) Results: In Study 2, we investigated the same general effects of gesture on the interaction process, although we now focused on how the split-utterance procedure would enhance the effects observed in Study 1. As with Study 1, we conducted independent-samples t-tests with 95% confidence intervals (CI). Our hypothesis holds true in that results of Study 2 show a significant effect of condition on more dependent measures than in Study 1: the 'sympathetic' characteristic shows slightly greater significant difference between conditions, with higher ratings in the multimodal condition (*M* = 4.05, *SD* = 0.92) than in the unimodal condition (*M* = 3.30, *SD* = 1.34), *t*(39) = -2.09, *p* = 0.02. Similar to the results of Study 1, the robot was perceived as more lively when its utterances were accompanied by gestures (*M* = 3.15, *SD* = 0.99) than when it was using only speech (*M* = 2.55, *SD* = 0.82), *t*(39) = -2.10, *p* = 0.02. In Study 2, the characteristics 'friendly', 'communicative', and 'engaged' were also identified as being significantly different between conditions at the 5% and 10% levels respectively, where there had been no significant differences in Study 1. Specifically, participants assessed the robot as more friendly in the multimodal condition (*M* = 4.25, *SD* = 0.97) than in the unimodal condition (*M* = 3.65, *SD* = 1.18), *t*(38) = -2.76, *p* = 0.04. Furthermore, the gesturing robot was rated as more communicative (*M* = 3.50, *SD* = 1.28) than the robot that relied on speech only (*M* = 2.90, *SD* = 1.33), *t*(38) = -1.45, *p* = 0.08. Finally, participants perceived the robot as more engaged in the speech-gesture condition (*M* = 3.75, *SD* = 1.33) than in the speech-only condition (*M* = 3.16, *SD* = 1.30), *t*(37) = -1.40, *p* = 0.08. Mean values and statistical p-values of all measured dimensions are displayed in Fig. 3. Interestingly, there was no enhancement in the difference between levels of the 'competent' characteristic, and in fact, a decrease in difference between conditions with regard to the characteristic 'active'. We can possibly attribute this to people using the lively characteristic as a measure of 'activity', thus reducing the impact of the actual 'active' evaluation. Overall, the results demonstrate that co-verbal gesture results in a more positive HRI experience, i.e. the robot is observed more positively. In line with our latter Study 2 hypothesis, splitting the utterance chunks such that participants' focus on the robot is increased leads to a significant effect of condition on more dependent measures.

V. Conclusion and Future Work

We conducted two experimental studies using a humanoid robot, in order to investigate how humans perceive various gesture types performed by the robot during a task-related interaction. The findings from our first study suggest that the perception and evaluation of the robot is more positive when it displays non-verbal behaviors in the form of co-verbal gestures along with speech. The results from our

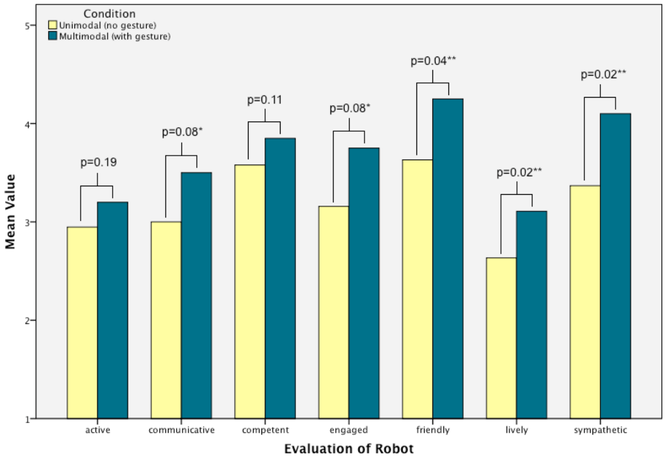

Fig. 3. Evaluation of the humanoid robot in Study 2, based on a 5-point Likert scale; *=p<0.10, **=p<0.05.

second study support these findings; in addition, they reveal that a more concious perception of the non-verbal behaviors displayed by the robot leads to an enhancement of this effect.

Several limitations apply to our studies: first, all our participants were native German speakers; hence it is possible that similar experiments conducted in a different cultural environment might yield different results. Second, despite the deliberate choice to set the studies in a kitchen environment to create a sense of familiarity for the participants, these were nevertheless experiments conducted in a laboratory with visibly installed video cameras. Finally, our results apply to the Honda humanoid robot whose appearance might have had an impact on the participants' perception and evaluation of the robot. However, in addressing this issue, the modifications that we carried out for the Study 2 procedure effectively allowed us to shift the participants' attention to the robot's behavior. We can therefore mark the observed differences between the two conditions as being elicited by the robot's behavior and not its appearance.

In the studies presented, the robot's gaze behavior was modelled in a very simplistic way in the multimodal condition; robot gaze in the unimodal condition was static throughout the interaction. These design choices were deliberately made to direct the participants' attention to the hand and arm movements performed by the robot in the multimodal condition. As a consequence, however, the robot's gazing behavior did not appear very natural during the interaction, since the robot did not follow the human interaction partner with its gaze. In future studies it will be desirable to systematically investigate the effect of the robot's gaze behavior alone in an isolated experimental set-up without hand and arm gesture. This way, it can be examined to what extent these findings are determined by the robot's hand arm gestures as opposed to the gaze behavior.

Generally, our results suggest that a robot presenting social cues in the form of co-verbal hand and arm gestures is perceived in a more positive way than a robot whose means of communication is limited to a single modality, namely speech. Specifically, they suggest that humanlike behavior is expected from a humanoid robot and has a positive impact on the way humans perceive the robot in an interaction. Although these findings might not be too surprising, they contribute to an advancement in HRI research: since our results are based on two experimental interaction studies involving a total of 81 participants, they provide a fundamental and meaningful basis for future studies on robot gesture. Ultimately, these results will allow us to design and build better artificial communicators in the future.

References

[1] K. Bergmann, S. Kopp, and F. Eyssel. Individualized gesturing outperforms average gesturing – evaluating gesture production in virtual humans. In *Proceedings of the 10th Conference on Intelligent Virtual Agents*, pages 104–117. Springer, 2010.

[2] A. Billard, S. Calinon, R. Dillmann, and S. Schaal. Robot Programming by Demonstration. In B. Siciliano and O. Khatib, editors, *Handbook of Robotics*, pages 1371–1394. Springer, Secaucus, NJ, USA, 2008.

[3] S. Calinon and A. Billard. Learning of Gestures by Imitation in a Humanoid Robot. In K. Dautenhahn and C.L. Nehaniv, editors, *Imitation and Social Learning in Robots, Humans and Animals: Behavioural, Social and Communicative Dimensions*, pages 153–177. Cambridge University Press, 2007.

[4] S. Goldin-Meadow. The role of gesture in communication and thinking. *Trends in Cognitive Science*, 3:419–429, 1999.

[5] M. Gullberg and K. Holmqvist. Keeping an eye on gestures: Visual perception of gestures in face-to-face communication. *Pragmatics and Cognition*, 7:35–63, 1999.

[6] A. B. Hostetter and M. W. Alibali. Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin and Review*, 15(3):495–514, 2008.

[7] K. Itoh, H. Matsumoto, M. Zecca, H. Takanobu, S. Roccella, M.C. Carrozza, P. Dario, and A. Takanishi. Various emotional expressions with emotion expression humanoid robot we-4rii. In *Proceedings of the 1st IEEE Technical Exhibition Based Conference on Robotics and Automation Proceedings TExCRA 2004*, pages 35–36, 2004.

[8] A. Kendon. Gesture: Visible action as utterance. *Gesture*, 6(1):119–144, 2004.

[9] S. Kopp and I. Wachsmuth. Synthesizing Multimodal Utterances for Conversational Agents. *Computer Animation and Virtual Worlds*, 15(1):39–52, 2004.

[10] N. Krämer, N. Simons, and S. Kopp. The effects of an embodied conversational agents nonverbal behavior on users evaluation and behavioral mimicry. In *Proceedings of Intelligent Virtual Agents (IVA 2007)*, pages 238–251. Springer, 2007.

[11] D. McNeill. *Hand and Mind: What Gestures Reveal about Thought*. University of Chicago Press, Chicago, 1992.

[12] V. Ng-Thow-Hing, P. Luo, and S. Okita. Synchronized gesture and speech production for humanoid robots. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 4617–4624, 2010.

[13] Y. Okuno, T. Kanda, M. Imai, H. Ishiguro, and N. Hagita. Providing route directions: Design of robot's utterance, gesture, and timing. In *Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction*, HRI '09, pages 53–60. ACM, 2009.

[14] M. Salem, S. Kopp, I. Wachsmuth, and F. Joublin. Towards an integrated model of speech and gesture production for multi-modal robot behavior. In *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication*, pages 649–654, 2010.

[15] M. Schröder and Jürgen Trouvain. The German Text-to-Speech Synthesis System MARY: A Tool for Research, Development and Teaching. 6:365–377, 2003.

[16] C.L. Sidner, C.D. Kidd, C. Lee, and N. Lesh. Where to look: a study of human-robot engagement. In *Proceedings of Intelligent User Interfaces*, pages 78–84. ACM Press, 2004.

[17] O. Sugiyama, T. Kanda, M. Imai, H. Ishiguro, and N. Hagita. Natural deictic communication with humanoid robots. In *Proceedings of the IEEE International Conference on Intelligent Robots and Systems*, pages 1441–1448, 2007.