Robot Gestures Make Difficult Tasks Easier: the Impact of Gestures on Perceived Workload and Task Performance

Manja Lohse Human Media Interaction PO Box 217, 7500 AE Enschede, The Netherlands m.lohse@utwente.nl, r.rothuis@student.utwente.nl Reinier Rothuis Human Media Interaction PO Box 217, 7500 AE Enschede, The Netherlands r.rothuis@student.utwente.nl Jorge Gallego Pérez Human Media Interaction PO Box 217, 7500 AE Enschede, The Netherlands j.gallegoperez@utwente.nl

Daphne E. Karreman Human Media Interaction PO Box 217, 7500 AE Enschede, The Netherlands d.e.karreman@utwente.nl

ABSTRACT

Gestures are important non-verbal signals in human communication. Research with virtual agents and robots has started to add to the scientific knowledge about gestures but many questions with respect to the use of gestures in humancomputer interaction are still open. This paper investigates the influence of robot gestures on the users' perceived workload and task performance (i.e. information recall) in a direction-giving task. We conducted a 2 x 2 (robot gestures vs. no robot gestures x easy vs. difficult task) experiment. The results indicate that robot gestures increased user performance and decreased perceived workload in the difficult task *but* not in the easy task. Thus, robot gestures are a promising means to improve human-robot interaction particularly in challenging tasks.

Author Keywords

human-robot interaction; gestures; perceived workload; task performance

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

INTRODUCTION

Gestures are important non-verbal signals in human communication (e.g., [19, 22]). Research with virtual agents and robots has started to add to the scientific knowledge about the

Copyright 2014 ACM 978-1-4503-2473-1/14/04\$15.00.

http://dx.doi.org/10.1145/2556288.2557274

Vanessa Evers Human Media Interaction PO Box 217, 7500 AE Enschede, The Netherlands v.evers@utwente.nl



Figure 1. Nao robot providing route directions to a user.

use of gestures in human-computer interaction (HCI). Complex gesture models for virtual agents have been developed and employed in HCI research (e.g., [1, 4, 20, 27]).

The evaluation of these models focused, among others, on the perceived naturalness of the gestures [20], on their expressivity [27], on users' trust in the agent [4], and on gesture production depending on the cognitive load of the speaker [2]. All these studies showed that virtual agents are perceived more positively when producing co-verbal gestures.

Gesture research with agents has attracted further attention in the recent years with the increasing number of robot research platforms that cannot only produce gestures in a virtual space but in the physical world that they share with humans (e.g., [3, 25, 28]). There is quite some work on gestures in human-robot interaction (HRI), particularly with respect to gesture-speech mismatches and synchronization (e.g. [25, 28]), and persuasiveness of gestures [6]. Common findings of these studies were that, if gestures are synchronized with

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org. *CHI 2014*, April 26 - May 01 2014, Toronto, ON, Canada

speech, they have positive effects on the evaluation (e.g., persuasiveness) of robots. Thus, HRI research has supported the findings from the virtual agents community.

Our research focuses on an effect of gestures that has largely been ignored in HCI in general as well as in HRI in particular: the influence of gestures produced by a system (here by a robot) on the perceived workload and task performance (here measured in terms of information recall) of the user in tasks with varying difficulty. While the impact of gesturing on objective task performance has found some attention in previous work (see Section "Related work"), perceived workload has not been taken into account. However, perceived workload is a highly important concept because it captures the users' subjective impression of the task (see Section "Related Work"). It has been shown that perceived workload affects the affective and physiological state of humans. High perceived workload can cause affective distress and increased blood pressure [8, 15] and has also been shown to influence job satisfaction [8].

To address this gap in the literature, we present an experiment on the influence of deictic (pointing) gestures that are redundant with speech in a direction-giving task. Users listen to directions provided by a robot in an easy or a difficult task. The descriptions are only accompanied by gestures for half of the participants. Based on our results, we show the importance of robot gestures particularly in difficult tasks.

Thus, our study contributes first research on the influence of robot gestures on perceived human workload and task performance. We add to the understanding of user behavior and cognitive capabilities in tasks with different levels of difficulty. In this respect, our work lays the ground for the design of future robot systems.

RELATED WORK

In the following, we highlight related work on gestures in human interaction. Thereafter, we introduce the concept of perceived workload, before we discuss the relation between both (gestures and perceived workload) and establish the hypotheses for our experiment.

Gestures in human interaction

There is a vast body of research on gestures in human interaction. Based on Kendon's definition [19], they are understood here as deliberate movements with sharp onsets and offsets. They are an excursion in that a part of the body (usually the arm or the head) moves away from and back to a certain position. The movement is interpreted as an addressed non-verbal utterance that conveys information.

Gesture research has resulted in categorizations of types and functions of gestures. These are described in the following sections because they help to set the frame for our experiment.

Types of gestures

Several categorizations of types of gestures that accompany speech have been introduced in the literature (see [9] for an overview). One of the best known categorizations is by Mc-Neill [23]. He differentiates four types of gestures: *Iconic gestures* represent body movements, movements of objects or

people in space, and shapes of objects or people. *Metaphoric* gestures present an abstract idea rather than a concrete object. *Deictic gestures* are used to indicate objects, people, and locations in the real world that do not necessarily have to be present. Finally, *beat gestures* are beats with the rhythm of speech regardless of content.

As has been mentioned above, our experiment focuses on a direction-giving task. In this context, deictic gestures are the most relevant type of gestures because they connect utterances to the physical setting and the physical world in which the conversation takes place is the topic of the conversation [5]. Pointing is the most obvious way to create deictic references. For the use of gestures in route direction tasks, Kita [21] found that when people utter an expression such as "turn left", they also perform gestures to describe the direction of "left". Hato et al. [14] pointed out that also in HRI such gestures are important in deictic interactions that refer to a region in space.

Kendon [19] identified different pointing gestures in interaction data, extending the index finger being the most commonly used when a speaker singles out a particular object or place. Thus, in our experiment the robot will perform pointing gestures with its index finger extended while providing the directions to the user (see Section "Independent Variables").

Functions of gestures

Most of the research on human gestures follows one of two main traditions: the one claiming that the main function of gestures is to help the speakers think and structure their speech [9], or the one stressing the communicative importance of gestures [19]. We here follow the second tradition and summarize related work on the influence of gestures on the perception of a person by others.

Gesturing has been found to strongly affect human listeners because they pay close attention to information conveyed via such non-verbal behaviors [9]. The influence of the gestures is largely determined by their relation to the information encoded in verbal utterances. This relation can be reinforcing or supplementing [16]. Especially deictic gestures such as pointing, references to objects, locations, and actions, can be reinforcing (also called "redundant" in the literature). The gestures label what is pointed at, but the referent can be understood with speech only. In contrast, when the gesture is supplementing, the referent that is pointed at is not clear without the gesture (for example, if someone says "this", a pointing gesture is needed to clarify what is referred to with the utterance). For our experiment, we use reinforcing deictic gestures. Thus, the referent is also clear without the gesture which allows us to design and compare conditions with and without the robot gesturing (see Section "Independent Variables").

Perceived workload

Our work focuses on the effect of the reinforcing deictic gestures on the perceived workload of users. There are many definitions of the perceived workload concept in the psychology literature. One of the best known is by Hart and Staveland [12]. They define workload as a hypothetical construct that represents the cost incurred by a human operator to achieve a certain level of performance. This view is human-centered rather than task centered and captures a subjective experience rather than an inherent property of a task. In other words, the perceived workload depends on the circumstances, skills, behaviors and perceptions of the user.

Perceived workload is a well-studied concept in HCI which is underlined by the fact that validated standard scales like the Nasa TLX exist. In [13], Hart reviews a selection of 550 papers that have used the Nasa TLX. It needs to be mentioned that the review was conducted some years ago, in 2006, and even at that time papers were selected and the overall number was higher. According to Hart's analysis, the research areas where the scale has been employed include manual control tasks such as flying (14%), driving (9%), data entry (10%) and others. However, there are much fewer studies addressing the perceived workload caused by virtual agents and robots. The Nasa TLX has mainly been used to measure human performance and workload in teleoperation scenarios (with one human teleoperating one or multiple robots) [29]. The perceived workload concept has also been mentioned in some other contexts. Gu and Badler [10] propose to research the effects of workload in studies with virtual agents by manipulating the number of agents one interacts with. However, they present no evaluation. In HRI, Yoshiike et al. [31] suggest to reduce workload by minimizing the social cues of a robot so that the user does not have to follow eye-gaze, facial expressions, life movement, etc. Again, there is no study to test the effect of this minimal design on the perceived workload of the user.

The influence of robot gestures on users' task performance and perceived workload

In the following, we establish a connection between gestures, performance and perceived workload, and introduce our hypotheses.

For human interaction, Kendon [18] proposed that gestures along with speech make utterance units more effective. In an HRI direction-giving task (as in our experiment), Okuno et al. [26] discovered that gestures had a positive effect on recall of the information. Also McNeil, Alibali, and Evans [22] suggested that redundant gestures facilitate understanding in adult-child interaction. They assumed that indexical (deictic) gestures may facilitate comprehension most when speakers make reference to particular objects and locations. However, their findings only partially underlined these assumptions. They found that speakers' reinforcing gestures facilitated listeners comprehension of the accompanying speech when the verbal message was highly complex (in their study this was the case for younger children compared to older ones) but not when it was simple. Also for HRI, Häring et al. [11] found that additional modalities (gaze and gesture) did not help to increase performance in the simple condition of a puzzle but in the complex condition. This leads us to our first hypothesis:

easy task /	easy task /
no gestures	gestures
difficult task /	difficult task /
no gestures	gestures

Table 1. Experimental conditions

H1: The participants' task performance (in terms of correct recall of directions) will improve significantly when the robot produces redundant gestures in the difficult task but not in the easy task.

However, task performance is not our main interest here. In contrast to previous studies, we are also interested in the effects of gestures on the perceived workload of the users. As has been mentioned above, we are not aware of any related work addressing this relation directly. However, it has been shown that visual feedback can decrease perceived mental workload [30]. Also gestures are a visual cue which leads us to our second hypothesis:

H2: The participants' perceived workload will be significantly lower when the robot produces redundant gestures, both in the difficult and in the easy task.

METHOD

We designed a 2 (gesture vs. no gesture) x2 (easy vs. difficult task) between subjects experiments (see Table 1), in which the participants played a memory game with the robot Nao. The robot told the participants a seemingly random set of directions that they were later asked to recall. The directions consisted of phrases used to navigate a building. Each direction included two pieces of information: one directional information (up, down, left or right) and a number specifying the amount of levels the participant needed to go up or down, or the number of turns the participant needed to take to left or right. The phrases we used were:

- "Take the 'nth' left"
- "Take the 'nth' right"
- "Go up 'n' levels"
- "Go down 'n' levels"

In all utterances, n was a number between one and five. The amount of directions per round was varied between difficulty levels (see Section "Independent Variables").

Robot

The robot used in this experiment is the Nao robot by Aldebaran Robotics (see Figures 1, 2, and 3). It is a bipedal humanoid robot that measures 57.3 cm in height when in an upright, standing position. The robot has 25 degrees of freedom to move its head, arms, hands, legs, and hips. It has two speakers integrated in its head.

In our experiment, the robot was connected to a laptop from where the researcher could remote control its behavior. Scripts for the utterances and gestures were prepared beforehand and played during the interaction. All participants received the same set of directions in a randomized order.



Figure 2. Nao in the condition without gestures (static pose).

Independent variables

The following sections describe how we manipulated the robot behavior and the task between the conditions.

Gesture vs. no gesture

We designed two different robot behaviors with respect to gesturing. In the "no gesture" condition (also referred to as "static condition" in the following), the robot stayed in one posture throughout the experiment and did not gesture (see Figure 2). For the condition with gestures (also called "animated condition" in the remainder of the paper), we developed movements that were in accordance with the directions that the robot provided to the users. Thus, all gestures were pointing (deictic) gestures redundant with speech (see "Related Work" Section). Their design was based on work by Kanda et al. [17]. However, as the Nao robot cannot rotate its torso, we replaced this motion by the rotation of the head. This should allow the participants and the robot to establish a joint perspective. We implemented the following gestures:

- 1. The robot points its arm and rotates its head to its right (left for the participant) or left (right for the participant).
- 2. The robot points its arm and tilts its head upwards.
- 3. The robot points its arm and tilts its head downwards.

Figure 3 depicts examples of the gestures of the robot.

Easy vs. difficult task

We chose a direction-giving task because the embodied robot could indeed be credible in such a task pointing out directions in space. Furthermore, this task allowed for manipulation of the difficulty by influencing the amount of information provided to the participants. As has been mentioned above, each direction consisted of two pieces of information. We conducted a pretest with five students of our university to determine how many directions appeared easy or difficult to them.



Figure 3. Gestures of the Nao robot (upper left: pointing to the left; upper right: pointing to the right; lower left: pointing upwards; lower right: pointing downwards).

All participants of the pretest should be able to solve the easy task. The difficult task should cause some problems to some of them but still be solvable for the majority of participants. Based on the results of the pretest, we decided for two directions (with two pieces of information each) in the easy task and four directions (with two pieces of information each) in the difficult task. This decision is in line with Miller's Law that claims that humans are able to hold 7 ± 2 objects in their working memory [24].

Next to the pieces of information within one direction, we also analyzed how long the participants paid attention to the task and the robot before getting bored or distracted. Based on this information, we decided for five rounds that the participants had to complete with the robot in both conditions.

Dependent variables

Our dependent variables included measures of perceived shared reality, task performance, and perceived workload. We also collected demographic information on age, gender, and experience with robots.

Perceived shared reality

We employed a scale on perceived shared reality because it has been found that robot gestures and shared reality are closely intertwined in HRI [28]. Thus, this scale basically serves as a manipulation check to ensure that the manipulations between the static and animated conditions were actually perceived by the users and caused the potential differences in the task difficulty and perceived workload ratings.

We used the experienced shared reality scale previously employed by Salem et al. [28]. The scale contains three items:

• I feel close to the robot

- I found the interaction with the robot pleasant
- I had fun while interacting with the robot

The scale was a 5-point Likert scale (1: strongly disagree - 5: strongly agree). Its internal reliability (Cronbach's α) was .822.

Task performance

The task performance was an objective measure. We counted the number of correct recalls of directions by the users based on video data that we recorded during the trials. The maximum score was 5 (minimum 0) because the users played five rounds and only if the recall of the information presented in one round was 100% correct, the recall was counted as being correct (thus, increasing the score by one point per round).

Perceived workload

To capture the users' subjective impression of the workload (or in other words the perceived workload), we used the Nasa TLX scale which is a standard in recording subjective task workload [12]. We included all subscales (mental demand, physical demand, temporal demand, performance, effort, and frustration) into the analysis, each of them being rated on a 10-point Likert scale ranging from "low" to "high".

We employed the weighted version of the Nasa TLX which allows to take the importance of each subscale in a specific scenario into account. Thus, the questionnaire included the scale to weigh the subscales which asked the participants to compare all individual subscales with all others and to indicate which one they found more important or relevant.

Data analysis

Before analyzing the data, we checked for normality of distribution. If the data was normally distributed, we used two-factorial ANOVAs and we report main- and interaction effects. In case of non-normally distributed data, nonparametric tests were applied, namely Kruskall-Wallis tests followed up by post-hoc Mann-Whitney tests as proposed by Field [7].

Procedure

At the start of the experiment, each participant was greeted by the experimenter who provided some explanation and asked the participant to fill out a consent form. Thereafter, the participant sat down in front of the table on which the Nao was standing (see Figure 1). Thus, both were facing each other (rather than having the participants look down on the small robot). The experimenter explained that when the robot would either point to the right or left, it would comply to the spatial perspective of the participants, i.e. if it used its left arm to point, it would point to the right. After providing this information, the experimenter withdrew to a laptop that was used to control the robot in a Wizard of Oz fashion. The participants could not see the experimenter from where they were sitting.

In all conditions, the interaction started with the robot uttering "Let's begin the game". As mentioned above, the robot's actions were pre-programmed and triggered by the experimenter at the correct moments. Every directional phrase was randomly separated by one of the following words: "next", "then", "following that", "subsequently", or "after that". When the participant completed the round successfully, the robot would say "That is correct". If the participant made an error, the robot said "That is not correct." Both events were triggered by the experimenter. Each new round started with the robot announcing "Let's begin a new round". After having completed five rounds, the experimenter asked the participant to fill out the questionnaire and thanked him/her.

Sample

The experiment was conducted with 32 participants distributed equally between the conditions. All participants were students of the university (3 female, 29 male). Their age ranged between 18 and 28 years with a mean age of 21 years (standard deviation (sd) = 2 years). The participants had hardly any experience with robots: 27 had never seen a robot in real life, 3 had only seen one, and just 2 had interacted or worked with robots before the experiment.

RESULTS

Perceived shared reality

A Shapiro-Wilk test showed that the data for perceived shared reality were normally distributed. Thus, we conducted a twoway ANOVA to discover differences between the conditions. This ANOVA revealed that there was no statistical difference between the conditions regarding experienced shared reality, F(3, 28) = 1.865.p = 0.159. In fact, both the static / easy and the static / difficult conditions had a mean of 2.96 (easy: sd = 0.81, difficult: sd = 0.63). The means for the animated / easy and animated / difficult conditions were the same as well with a value of 3.58 (easy: sd = 0.42, difficult: sd =(0.89). Thus, the task difficulty did not have an impact on the perceived shared reality which is positive because only the gestures should have. There was indeed a main effect for the gestures F(1, 29) = 5.594, p = .025. Based on this finding, we can conclude that the manipulation was successful and the gestures led to higher perceived shared reality.

Task performance

A Shapiro-Wilk test showed that the data on task performance were not normally distributed. Hence, we conducted a Kruskal-Wallis test to compare the conditions. The test revealed a significant difference between them (p=.001) (see Figure 4). Post-hoc Mann-Whitney tests showed that participants who completed one of the easy conditions had significantly higher scores (m = 4.35, sd = 0.70) than participants who completed one of the two hard conditions (m =3.06, sd = 1.29, U = 47.50, p = .002. This finding indicates that the manipulation of the difficulty of the task was successful. Tests between the individual conditions showed that the difference was statistically significant between the static conditions (U = 5.00, p = .003) as well as between the animated conditions (U = 14.00, p = .045). Participants who completed an animated condition had significant higher scores (m = 4.23, sd = 0.76) than participants who completed a static condition (m = 3.19, sd = 1.38), U = 68.00, p = .023. Thus, not only the difficulty of the task, but also the gestures played an important role for task

Subscale	Mean weight (sd)
Mental demand	4.34 (0.90)
Physical demand	0.41 (0.67)
Temporal demand	2.19 (1.06)
Performance	3.97 (0.93)
Effort	2.81 (1.00)
Frustration	1.31 (1.06)

Table 2. Mean weights (standard deviations) of the Nasa TLX subscales

performance. To look at this in more detail, we compared all individual conditions. We found that the gestures had no significant influence on task performance in the easy condition but in the difficult condition (U = 7.50, p = .007). In other words, in the easy condition the recall of the participants did not improve if the robot gestured, but in the difficult condition it improved significantly. This finding is in line with related work and **H1** (The participants' task performance (in terms of correct recall of directions) will improve significantly when the robot produces redundant gestures in the difficult task but not in the easy task).

Perceived workload

As has been mentioned in the "Dependent Variables" Section, we used the weighted Nasa TLX scale to assess perceived workload. The results indeed revealed that some subscales were more relevant than others to the participants, e.g., physical demand had almost no influence with a mean weight of 0.41 (sd = 0.67, on a scale of 0 to 5). Table 2 depicts the weights of all subscales.

Overall, the perceived workload was average (m = 5.32)(sd = 1.17) on a scale of 1:low - 10: high). Again, we conducted a Shapiro-Wilk test for normality which showed that the data for perceived workload were normally distributed. A two-way ANOVA revealed a significant difference of the perceived workload between the four conditions (F(3, 28) =6.905, p = .001). We discovered main effects for animation (F(1, 31) = 6.438, p = .017) and for task difficulty (F(1, 31) = 11.018, p = .003). In other words, the perceived workload in the easy task (m = 4.77, sd = 0.90) was significantly lower than in the hard task (m = 5.87, sd = 1.18); and it was significantly lower in the animated conditions (m = 4.90, sd = 0.83) compared to the static conditions (m = 5.74, sd = 1.33). There was no interaction effect between animation and difficulty.

However, Figure 5 shows that the perceived workload decreased much more due to the gestures in the difficult task $(m_{animated} = 5.15, sd = 0.81, m_{static} = 6.58, sd = 1.07)$ than it did in the easy task $(m_{animated} = 4.65, sd = 0.82; m_{static} = 4.89, sd = 1.01)$. In fact, unpaired T-tests showed that the difference in the easy task was not significant, while the difference in the difficult task was (T(14) = 3.031, p = .009). Therefore, **H2** (The participants' perceived workload will be significantly lower when the robot produces redundant gestures, both in the difficult and in the easy task) was only supported for the difficult but not for the easy tasks.

This finding is in line with the results regarding task performance. It appears that in the case of our experiment, the task



Figure 4. Task performance (mean number of directions that the participants recalled correctly) in the conditions.



Figure 5. Perceived workload in the conditions.

performance was actually correlated with the perceived workload. Indeed, we discovered a significant negative correlation between the weighted Nasa TLX scores and the task performance (number of correct answers), r(30) = -0.695, p = 0.000. In other words, the higher the perceived workload, the less correct answers were provided by the participants.

DISCUSSION

This paper set out to answer the question whether robot gestures can decrease the users' perceived workload and increase their performance on a task. To find an answer to this question, we conducted a direction-giving experiment in which we manipulated the gestures that a robot performed (no gestures vs. deictic gestures) and the difficulty of the task (easy vs. difficult task). Our results showed that the robot gestures did indeed support human task performance and lowered the perceived workload of the users but only in the difficult task. This result is in line with research on gestures in human interaction that came to the same conclusion (see "Related work", [22]). We believe that this is due to the easy task being simple enough for the participants to be completed with very little information. This little information could easily be kept in mind even if it was only presented with one modality. However, the difficult task required the participants to recall more information and multimodality was useful to convey it.

Session: Human-Robot Interaction

But what does this imply for the design of robots or virtual agents? As has been mentioned in the introduction, perceived high workload can have a serious negative impact on humans' psychological and physical well-being. Thus, if we are building systems, we have to strive to keep the perceived workload low enough to avoid such effects. This goal has to be achieved despite the fact that the difficulty of the task itself in many cases cannot be reduced. However, we can influence how a system and a human solve the task. Our results show that equipping systems with the capability to gesture is a promising way to reduce perceived workload while keeping the task difficulty stable.

LIMITATIONS AND FUTURE WORK

Our work suffers from some limitations that will be addressed in future research. First of all, the impact of gesture in interaction is highly dependent on the task. Our research here is restricted to one direction-giving task and to deictic gestures. Thus, future work needs to focus on the influence of other types of gestures (iconic, metaphoric, and beat) on performance and perceived workload in a variety of tasks. Furthermore, future research would profit from larger sample sizes and samples that better capture representative user groups of robots, i.e. samples with better gender and age distribution.

Nevertheless, we believe that this work addresses a novel aspect in HRI that comes with valuable implications for the larger HCI community: it introduces the relation between robot gestures and perceived workload to the field and shows that the interaction can actually profit from robot gestures with respect to perceived workload, particularly in difficult tasks.

ACKNOWLEDGMENTS

The research leading to these results was partially supported by the European Community's 7th Framework Programme under Grant agreements 600877 (SPENCER - Social situation-aware perception and action for cognitive robots) and 288235 (FROG - Fun Robotic Outdoor Guide).

REFERENCES

- Bergmann, K., Eyssel, F. A., and Kopp, S. A second chance to make a first impression? how appearance and nonverbal behavior affect perceived warmth and competence of virtual agents over time. M. Walker, M. Neff, A. Paiva, and Y. Nakano, Eds., Proceedings of the 12th International Conference on Intelligent Virtual Agents, Springer (2012), 126–138.
- Bergmann, K., Kahl, S., and Kopp, S. Modeling the semantic coordination of speech and gesture under cognitive and linguistic constraints. In *Intelligent Virtual Agents*, R. Aylett, B. Krenn, C. Pelachaud, and H. Shimodaira, Eds., vol. 8108 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2013, 203–216.
- 3. Breazeal, C., Kidd, C., Thomaz, A., Hoffman, G., and Berlin, M. Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. In *Intelligent Robots and Systems, 2005. (IROS 2005).*

2005 IEEE/RSJ International Conference on (2005), 708–713.

- Buisine, S., Abrilian, S., and Martin, J.-C. From brows to trust. Kluwer Academic Publishers, Norwell, MA, USA, 2004, ch. Evaluation of multimodal behaviour of embodied agents, 217–238.
- 5. Cassell, J. Embodied conversational agents. MIT Press, Cambridge, MA, USA, 2000, ch. Nudge nudge wink wink: elements of face-to-face conversation for embodied conversational agents, 1–27.
- 6. Chidambaram, V., Chiang, Y.-H., and Mutlu, B. Designing persuasive robots: how robots might persuade people using vocal and nonverbal cues. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, HRI '12, ACM (New York, NY, USA, 2012), 293–300.
- 7. Field, A. *Discovering statistics using SPSS*. Sage publications, 2009.
- Fox, M. L., Dwyer, D. J., and Ganster, D. C. Effects of stressful job demands and control on physiological and attitudinal outcomes in a hospital setting. *Academy of Management Journal 36*, 2 (1993), 289–318.
- 9. Goldin-Meadow, S. *Hearing Gesture: How Our Hands Help Us Think*. Belknap Press of Harvard University Press, 2005.
- Gu, E., and Badler, N. I. Visual attention and eye gaze during multiparty conversations with distractions. In *IVA*, J. Gratch, M. Young, R. Aylett, D. Ballin, and P. Olivier, Eds., vol. 4133 of *Lecture Notes in Computer Science*, Springer (2006), 193–204.
- Häring, M., Eichberg, J., and André, E. Studies on grounding with gaze and pointing gestures in human-robot-interaction. In *Social Robotics*, S. Ge, O. Khatib, J.-J. Cabibihan, R. Simmons, and M.-A. Williams, Eds., vol. 7621 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2012, 378–387.
- Hart, S., and Staveland, L. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Human mental workload*, P.A. Hancock and N. Meshkati (Eds.), Amsterdam: Elsevier (1988), 139–183.
- Hart, S. G. Nasa-task load index (nasa-tlx); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting 50*, 9 (2006), 904–908.
- Hato, Y., Satake, S., Kanda, T., Imai, M., and Hagita, N. Pointing to space: modeling of deictic interaction referring to regions. In *Proceedings of the Conference on Human-Robot Interaction*, P. J. Hinds, H. Ishiguro, T. Kanda, and P. H. K. Jr., Eds., ACM (2010), 301–308.
- Ilies, R., Dimotakis, N., and De Pater, I. E. Psychological and physiological reactions to high workloads: Implications for well-being. *Personnel Psychology* 63, 2 (2010), 407–436.

- Iverson, J. M., Capirci, P., Longobardi, E., and Caselli, M. C. Gesturing in mother-child interaction. *Cognitive Development 14* (1999), 57–75.
- 17. Kanda, T., Shiomi, M., Miyashita, Z., Ishiguro, H., and Hagita, N. A communication robot in a shopping mall. *Robotics, IEEE Transactions on 26*, 5 (2010), 897–913.
- Kendon, A. Gesture and speech: How they interact. In Nonverbal Interaction (Sage Annual Reviews of Communication, Volume 11), J. Wiemann and R. Harrison, Eds. Sage Publications, Beverly Hills, California, 1983, 13–46.
- 19. Kendon, A. *Gesture Visible Action as Utterance*. Cambridge University Press, Cambridge, 2004.
- 20. Kipp, M., Neff, M., Kipp, K. H., and Albrecht, I. Towards natural gesture synthesis: Evaluating gesture units in a data-driven approach to gesture synthesis. In *Proceedings of the 7th international conference on Intelligent Virtual Agents*, IVA '07, Springer-Verlag (Berlin, Heidelberg, 2007), 15–28.
- 21. Kita, S. Pointing: Where Language, Culture, and Cognition Meet. PSYCHOLOGY Press, 2003.
- McNeil, N., Alibali, M., and Evans, J. The role of gesture in children's comprehension of spoken language:now they need it, now they don't. *Journal of Nonverbal Behavior 24*, 2 (2000), 131–150.
- 23. McNeill, D. *Hand and Mind: What Gestures Reveal about Thought*. University of Chicago Press, Chicago, 1992.
- Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *The Psychological Review* 63, 2 (March 1956), 81–97.

- 25. Ng-Thow-Hing, V., Luo, P., and Okita, S. Y. Synchronized gesture and speech production for humanoid robots. In *IROS*, IEEE (2010), 4617–4624.
- 26. Okuno, Y., Kanda, T., Imai, M., Ishiguro, H., and Hagita, N. 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, ACM (New York, NY, USA, 2009), 53–60.
- Pelachaud, C. Studies on gesture expressivity for a virtual agent. *Speech Commun.* 51, 7 (July 2009), 630–639.
- Salem, M., Eyssel, F., Rohlfing, K., Kopp, S., and Joublin, F. To err is human(-like): Effects of robot gesture on perceived anthropomorphism and likability. *International Journal of Social Robotics* 5, 3 (2013), 313–323.
- Steinfeld, A., Fong, T., Kaber, D., Lewis, M., Scholtz, J., Schultz, A., and Goodrich, M. Common metrics for human-robot interaction. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, HRI '06, ACM (New York, NY, USA, 2006), 33–40.
- Vitense, H. S., Jacko, J. A., and Emery, V. K. Multimodal feedback: an assessment of performance and mental workload. *Ergonomics* 46, 1-3 (2003), 68–87. PMID: 12554399.
- Yoshiike, Y., Silva, P. R. S. D., and Okada, M. Mawari: A social interface to reduce the workload of the conversation. In *ICSR*, B. Mutlu, C. Bartneck, J. Ham, V. Evers, and T. Kanda, Eds., vol. 7072 of *Lecture Notes in Computer Science*, Springer (2011), 11–20.