

# The Role of Embodiment and Perspective in Direction-Giving Systems

**Dai Hasegawa**  
Hokkaido University  
Sapporo, Hokkaido, Japan  
hasegawadai@media.eng.hokudai.ac.jp

**Justine Cassell**  
Carnegie Mellon University  
Pittsburgh PA, U.S.A.  
justine@cs.cmu.edu

**Kenji Araki**  
Hokkaido University  
Sapporo, Hokkaido, Japan  
araki@media.eng.hokudai.ac.jp

## Abstract

In this paper, we describe an evaluation of the impact of embodiment, the effect of different kinds of embodiment, and the benefits of different aspects of embodiment, on direction-giving systems. We compared a robot, embodied conversational agent (ECA), and GPS giving directions, when these systems used speaker-perspective gestures, listener-perspective gestures and no gestures. Results demonstrated that, while there was no difference in direction-giving performance between the robot and the ECA, and little difference in participants' perceptions, there was a considerable effect of the type of gesture employed, and several interesting interactions between type of embodiment and aspects of embodiment.

## Introduction

Increasingly, embodiment is a part of the design of intelligent systems, and many people can therefore expect to interact with embodied agents such as avatars, embodied conversational agents (ECAs) and robots. Robotocists often argue that the physical co-locatability of the robot will render it more effective and desirable as an intelligent system. Researchers in the field of Embodied Conversational Agents, on the other hand, often counter that the degrees of freedom and naturalness of movement and appearance of ECAs will allow them to be more effective and rated more highly. While both groups of researchers believe in the power of embodied systems, little research has actually looked at the actual benefit of embodiment (such as comparing a GPS to a direction-giving robot). Even less research has addressed the bone of contention between the two groups, to compare the different types of embodiment (such as comparing graphical agents to robots). And virtually no research has examined how those different types of embodiment may interact with the different aspects of embodiment (such as comparing particular uses of gesture or eye gaze in how efficient the robot and the ECA are at collaborating with a human to complete a task).

To help answer these questions, we conducted an experiment in which we evaluated a human's ability to complete

the task of giving directions to another human, and the human's experience of that task, when the directions are originally conveyed to the human by a GPS vs. an embodied conversational agent (ECA) vs. a robot, and when the versions of the system use speech only, speech + gesture oriented from the system's perspective (or speech + map in the GPS condition), or speech + gestures oriented to the listener's perspective. This design allows us to examine the impact of embodiment (robot / ECA vs. GPS), the impact of physical embodiment in the real world (robot vs. ECA), the impact of nonverbal behavior (no gesture vs. gesture), and the impact of situating the gestures in the listener's world (no gesture / speaker-perspective gesture vs. listener-perspective gesture), as well as the interaction among these factors.

## Comparing Robots to ECAs

Some previous work has examined the importance of a robot's physical existence when compared to ECAs (Powers et al. 2007; Kidd and Breazeal 2004; Shinozawa et al. 2005), with mixed results. Perhaps unsurprisingly, (Shinozawa et al. 2005) showed that an ECA is more influential in recommending action when the action involves the same screen real-estate that the ECA inhabits, and a robot is more influential when the action recommended involves the 3D world that the robot inhabits. (Powers et al. 2007) demonstrated that a people are more likely to disclose information about themselves to an ECA than to a robot, and more likely to forget what they spoke about to the robot. People did, nevertheless, prefer interacting with the robot. This research, however, displayed only the head and neck of the agent, and the entire body of the robot, and so it is unclear what effect embodiment plays. In the current research we move to a task where effectiveness can be judged (are directions correct), and where embodiment plays a clear role (both pointing and illustrative gestures play a key role in human-human direction giving).

## Direction Giving ECAs and Robots

Several ECAs have been developed to provide route directions using both speech and gesture (Theune, Hofs, and van Kessel 2007; Cassell et al. 2002; 1999). Perhaps the most elaborate is NUMACK (Kopp et al. 2007; Nakano et al. 2003; Stocky and Cassell 2002) which autonomously generates directions to landmarks on a college campus, directions

which include speech, eye gaze, and gesture. NUMACK provides directions from a route perspective, which is the perspective of the person following the route (Taylor and Tversky 1996; 1992; Stiegnitz, Lovett, and Cassell 2001), and describes landmarks and actions using iconic gesture which are automatically generated based on features of appearance of the landmarks or the actions. However, NUMACK has only been evaluated anecdotally to date.

Several robots have also been developed to give route directions. Okuno et al. (Okuno et al. 2009) showed that route directions from a robot using gesture was better understood by users than route directions not using gesture. In their experiment, all of the robots took an "align perspective" in which the the robot and the user stood side-by-side, and both were looking at the same poster which showed a part of town. The robot gave directions using deictic gestures, such as "go straight this way" while pointing at the poster. In the "align perspective", the robot's right is the same as user's. As we can see in (Ono, Imai, and Ishiguro 2001), in the situation where two people are collocated inside of an environment and share a perspective (for example, standing on a street corner), people usually take the "align perspective" to explain directions, and it is natural for a robot to take the same perspective as the human.

However, Okuno did not compare robots to graphical agents. And direction-giving in the real world more commonly occurs in the absence of an image or map of the space in which directions are being given. In this situation there is no natural side-by-side alignment of perspective, and it is most natural for the two interlocutors to stand facing one another. In this situation, it is difficult to use deictic gesture to show directions (i.e. it is inappropriate to say "go this way" accompanied by pointing). We believe there are two possible ways to provide directions in this situation. The first one is "speaker perspective" where a speaker who is going to give directions faces a listener and takes his/her own perspective to show directions (if the speaker says "turn right," the speaker should point to his/her own right, which is the listener's left). The second one is "listener perspective" where a speaker turns slightly to his/her right/left and takes the same perspective as the listener (if he/she says "turn right," he/she should point at a direction which looks like the listener's right).

In the current work, we compare "listener-perspective gesture" to "speaker perspective gesture." We compare ECAs to robots. And we evaluate how effective these embodied systems are compared to conventional direction giving methods - a GPS map accompanied by speech guidance.

## Experiment

In sum, we report here on an experiment where we investigate the effects of an "agent factor" (robot vs. ECA vs. GPS) and a "gesture factor" (no gesture vs. speaker-perspective gesture vs. listener perspective gesture) on direction-giving.

### Settings

In order to ensure parity across conditions, we recorded one single set of direction-giving utterances using the Festival

Table 1: A Set of Utterances

- U1: Hello there.
- U2: Let me give you walking directions to your destination.
- U3: Leave this building by the side exit and **turn right**.
- U4: Then **go straight**.
- U5: Then you will see a **tall building in front of you**.
- U6: **Turn left** before it.
- U7: Next there will be a house with a **round-shaped porch** on your left.
- U8: **Turn slightly right** after it.
- U9: Next your destination building is a **small building on your left**.

speech synthesis tool (Black and Taylor 1999), and coordinated each utterance with gestures in the ECA and robot conditions, and map segments in the GPS condition. The utterances described directions from one building to another building across an imaginary campus, and are shown in Table 1 where the bold parts are the directions synchronized with a redundant gesture (that conveyed the same information as speech). Following (Okuno et al. 2009), the duration between utterances was six seconds, except between U1 and U2, and between U2 and U3. In our pilot studies, this duration led to optimal memory for the directions given. Likewise, each system produced only 8 utterances because our pilot studies showed that this number neither led to a ceiling nor floor effect for participants. Figure 1c shows the imaginary campus, where the building marked by an inverted triangle is the building where the participants begin. The building second from the bottom on the far right is the destination.

We used a KHR2-HV<sup>1</sup> as a humanoid robot, with a height of 30 cm and 17 degrees of freedom (see Figure 1a). For the ECA, we used NUMACK, shown in Figure 1b, projected on a 60-inch plasma screen. As a conventional system that gives route directions, we used a TomTom GPS visible in front of participants with the map projected on the plasma screen. In the GPS condition the map is used instead of gestures. All participants stood at the same distance from the plasma screen/robot, and the robot was elevated on a black box so that its eye position was at the same height as NUMACK's. Figure 2a shows the robot condition with a participant listening to directions. Figure 2b shows the NUMACK condition and Figure 2c shows the GPS condition. For each condition, we used the same speakers.

### Conditions

Below we describe the conditions we implemented in this experiment:

#### The robot + listener perspective gesture

The robot slightly turns to its right and uses gestures from the same perspective as the listener. When the robot says "turn right," it points to a direction aligned with the users' right (Figure 3a).

<sup>1</sup>Kondo Kagaku Co. Ltd, <http://www.kondo-robot.com/>

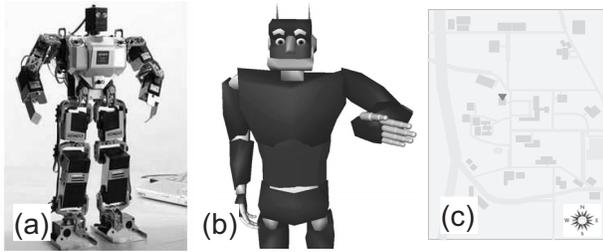


Figure 1: (a) A Robot, (b) NUMACK and (c) A Map

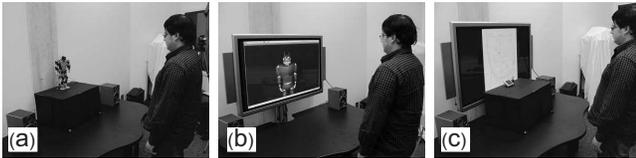


Figure 2: Snapshots of (a) Robot condition, (b) NUMACK condition, and (c) GPS condition

### The robot + speaker perspective gesture

The robot faces the participant and uses gestures from the robot's perspective. When the robot says "turn right," it points to its own right (the listener's left).

### The robot + no gesture

The robot is facing the user and does not use any gestures.

### NUMACK + listener perspective gesture

NUMACK slightly turns to its right and uses gestures from the same perspective as the listener. When NUMACK says "turn right," it points to a direction aligned with the user's right (Figure 3b).

### NUMACK + speaker perspective gesture

NUMACK faces the participant, and uses gestures from the robot's perspective. When NUMACK says "turn right," it points to its own right (the listener's left).

### NUMACK + no gesture

NUMACK is facing the user and does not use any gesture.

### The GPS + a map

The GPS map is projected on the large screen while the system speaks the directions (Figure 2c).

### The GPS + no map

In this condition, audio is played without any map.

## Evaluation Methods

Because we are interested in the actual effectiveness of these systems as well as user perception, we rely on three types of data to answer our three questions (the impact of embodiment, the comparison of different types of embodiment, and the relationship between the types of embodiment and the different aspects of embodiment). Those three types of data are: performance on a retelling task where participants repeat the directions heard from the system to a naive partner, performance on a map task, where participants use cut-outs

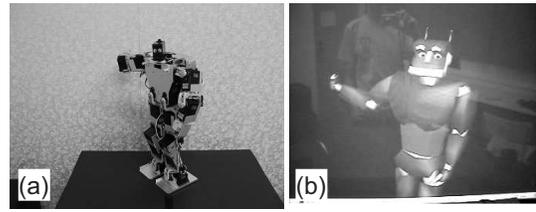


Figure 3: Gesture of "Turn Right" from Listener Perspective: (a) Robot, (b) NUMACK

Table 2: Evaluation Points of Retelling

- U1: Hello there.
- U2: Let me give you a walking direction to your destination.
- U3: Leave this building by the side exit and turn right.
- U4: Then go straight.
- U5: Then you will see a tall building in front of you.
- U6: Turn left before it.
- U7: Next there will be a house with a round shaped porch on your left.
- U8: Turn slightly right after it.
- U9: Next your destination building is a small building on your left.

of landmarks and paths to display the route they heard, and answers to a 26 item questionnaire, grouped into the categories: naturalness, presence, engagement, understandability, familiarity, reliability and enjoyment. On the two objective performance measures, the underlined twenty direction points in Table 2 are evaluated as "correctly mentioned", "not mentioned", or "wrongly mentioned" by the participant in the retelling task.

## Participants

Seventy five participants were recruited. Sixty-six native English-speaking participants were recruited by flyers on the campus of an American University and were paid US\$10 in cash after the session. For pragmatic reasons, nine subjects participated in Japan; all nine were also native English-speakers. They were recruited by flyers on the campus of a Japanese University and were paid 1,000 Yen (approximately equal to US\$10) in cash after the session. We excluded the data collected from one person since the procedure was not conducted appropriately. Participants were 35% male, 65% female, 87% aged 18-22 and 10% aged 23-28. The experiment was conducted using a between-subject design, with 10 participants per condition (except for robot+listener-gestures which had 9, GPS+map which had 7, and GPS+no map which had 9).

## Procedure

Participants listened to a Festival-generated speech-only introduction to the experiment to habituate them to the synthesized voice. Participants were told they would be given directions to a building across campus, and would need to

repeat those directions to somebody who didn't know the campus, and would also need to make a map displaying the route. After listening to these instructions, participants began the study. The questionnaire, which was not introduced beforehand, was given to them at the end of the experiment.

## Hypothesis

Among the three authors on the study, two hypothesized that the higher the physicality presence of the "agent factor," the stronger the impact of the information exchange (resulting in more correct retelling, more correct map, and higher subjective judgments). That is, even though the robot has restrictions in terms of degrees of freedom, its physicality is hypothesized to outweigh the limitations posed by those restrictions. The third author hypothesized that the more natural the movements and appearance of the "agent factor," the stronger the impact of this agent on the exchange of information. All three authors hypothesized that in terms of the "gesture factor", the presence of gesture would aid in performance, and that the listener perspective gesture would be the most effective. We also hypothesized that there would be an interaction such that whichever agent was more effective in information exchange would benefit most from the listener perspective gesture.

## Results

As the design of the experiment was an incomplete block design, we conducted two-way factorial ANOVAs in three ways: (1) 3 agents (the robot vs. NUMACK vs. GPS) and 2 gestures (listener-gesture/map vs. no-gesture), and (2) 3 agents (the robot vs. NUMACK vs. GPS) and 2 gestures (speaker-gesture/map vs. no-gesture), to check impacts of embodiment, and interaction with aspects of embodiment; (3) 2 agents (the robot vs. NUMACK) and 3 gestures (listener-perspective gesture vs. speaker-perspective gesture vs. no-gesture) to compare kinds of embodiment (robot vs. NUMACK) and aspects of embodiment (listener-perspective gesture vs. speaker-perspective gesture vs. no gesture), and to test for interaction effects among them. In our post-hoc analyses, we used t-test with Bonferroni method for multiple comparisons.

### Effects of Embodiment

From the results of ANOVA (1) comparing the robot to NUMACK to a GPS and listener gestures/map to no gestures/audio only, participants judged both embodied agents as more natural ( $F(2, 49)=5.066, p<.01$ , Bonferroni  $p<.05$ ) and more present ( $F(2, 49)=5.066, p<0.1$ , Bonferroni  $p<.05$ ) than the GPS, and felt that they would be more likely to ask the embodied agents for directions than the GPS ( $F(2, 49)=3.187, p<.05$ , Bonferroni  $p<.05$ ). They also judged both embodied agents as more familiar and more reliable than the GPS, but only when the embodied systems were using listener-oriented gesture ( $F(2, 49)=5.066, p<.01$ , Bonferroni  $p<.05$ ). No significant difference was found on performance however.

From the results of ANOVA (2) comparing the robot to NUMACK to a GPS and speaker gestures/map to no gestures/audio only, participants judged both embodied agents

as more present ( $F(2, 52)=5.038, p<.01$ , Bonferroni  $p<.05$ ) than the GPS. However, they judged NUMACK as more natural ( $F(2, 52)=3.175, p<.05$ , Bonferroni  $p<.05$ ) and more enjoyable ( $F(2, 52)=3.175, p<.05$ , Bonferroni  $p<.05$ ) than the GPS. Once again, there was no significant difference in performance due to the agent factor.

From the results of ANOVA (1), participants judged the systems that employed listener-oriented gestures as more natural ( $F(1, 49)=7.182, p<.01$ ), more understandable ( $F(1, 49)=7.182, p<.01$ ), more familiar ( $F(1, 49)=7.182, p<.01$ ), more reliable ( $F(1, 49)=7.182, p<.01$ ), and more enjoyable ( $F(1, 49)=4.038, p<.05$ ) than those same systems without gesture. In addition, the presence of listener-perspective gestures tended to reduce errors in the retelling task with respect to no gesture ( $F(1, 49)=2.811, p<.10$ ).

From the results of ANOVA (2), participants judged the presence of speaker-perspective gestures as more natural ( $F(1, 52)=4.027, p<.05$ ) than no gesture. However, there was no significant difference in performance due to the gesture factor when comparing the robot to NUMACK to the GPS.

### Kind of Embodiment

From the results of ANOVA (3), when NUMACK was compared to the robot, no significant performance results were found. However, interestingly, while the agent factor did not result in a difference in direction *effectiveness* on either retellings or the map task, it did result in a difference in *behavior*: when comparing NUMACK to the robot, all NUMACK conditions resulted in a statistical significantly higher use of complementary gestures ( $F(1, 49)=7.182, p<.01$ ). That is, when retelling directions told by NUMACK, participants were more likely to increase their information exchange by conveying one aspect of the directions in speech and a different, but complementary, aspect of the direction in gestures.

### Aspects of Embodiment

Compared to the paucity of significant performance differences due to kind of embodiment (NUMACK vs. robot), a number of results demonstrated significance due to the kind of gesture (listener-perspective gesture vs. speaker-perspective gesture vs. no gesture) the system used. Thus, comparing NUMACK to the Robot, we find that for both embodied systems, the presence of listener-oriented gestures resulted in a reduction of errors in the direction retelling task with respect to both speaker-perspective gestures and no gestures at all ( $F(2, 49)=3.187, p>.05$ , Bonferroni  $p<.05$ ). In the map task, too, only the presence of listener-oriented gestures resulted in a reduction of map errors ( $F(2, 49)=5.066, p<.01$ , Bonferroni  $p<.05$ ) – neither the kind of agent, nor any other gesture type resulted in a significant effect.

In responses to the questionnaire, listener gestures led participants to judge both agents as more reliable ( $F(2, 49)=5.066, p<.01$ ) and more natural ( $F(2, 49)=3.187, p<.05$ ) than the same system using either speaker-perspective gestures or no gestures. And the use of

listener-perspective gestures led participants to evaluate directions as more informative ( $F(2, 49)=3.187, p<.05$ , Bonferroni  $p<.05$ ) and easier to understand ( $F(2, 49)=5.066, p<.01$ , Bonferroni  $p<.05$ ) than speaker-perspective gestures (which, in turn, were judged more informative and easier to understand than no gesture).

Additionally, NUMACK and the robot were judged as more human-like when they used listener-perspective gestures than when they used speaker-perspective gestures ( $F(2, 49)=5.066, p<.01$ , Bonferroni  $p<.05$ ). To follow up on this result, we looked at whether participants themselves (the *real* human agents) were more likely to use listener-oriented gestures than speaker-oriented gestures . . . and we found out that they were not. Thus, although a number of subjective judgments led to higher evaluation of the systems when they used listener-oriented gestures, only 5 of the 74 participants used any listener-oriented gestures in their own direction-giving, while 61 participants used speaker-oriented gestures (the remainder used no gesture at all). No correlation existed between the experimental condition and the kind of gesture used in retellings.

### Interaction Effects

A number of questionnaire response results supported our hypothesis that there would be an interaction between the aspects of embodiment and the type of embodiment. Thus, listener gestures resulted in a perception of familiarity, but only for the robot ( $F(2, 49)=7.182, p<.01$ , Bonferroni  $p<.05$ ). Likewise, listener gestures led the user to judge the system as more enjoyable ( $F(2, 49)=7.182, p<.01$ , Bonferroni  $p<.05$ ), more understandable ( $F(2, 49)=3.187, p<.05$ , Bonferroni,  $p<.05$ ), and more co-present ( $F(2, 49)=3.187, p<.05$ , Bonferroni  $p<.05$ ) but only for the robot.

On the other hand, NUMACK scored higher than either the robot or the GPS in familiarity ( $F(2, 49)= 5.066, p<.01$ , Bonferroni  $p<.05$ ) and enjoyability ( $F(2, 49)= 5.066, p<.01$ , Bonferroni  $p<.05$ ) in the no gesture/no map conditions. This interaction is exemplified by participant judgments of empathy which are higher for NUMACK when it did not use gesture, and higher for the robot when it used listener-perspective gestures ( $F(2, 49)=3.187, p<.05$ , Bonferroni  $p<.05$ ); and participant judgments of desire to interact with the system again, which was also higher for NUMACK than the robot in the absence of gesture, and higher for the robot than NUMACK when listener-perspective gesture was used ( $F(2, 49)=5.066, p<.05$ , Bonferroni  $p<.05$ ).

## Discussion

### Summary

In this study, we addressed three questions about direction-giving systems: the impact of embodiment, the comparison between different kinds of embodiment, and the relationship between kinds of embodiment and different aspects of embodiment. The findings from the experiment can be summarized as follows:

(1) Embodiment (robot and NUMACK vs. GPS) did have a positive effect on participants' perceptions of the systems, but not on their performance.

(2) Kind of embodiment (robot vs. NUMACK) had no effect on retelling or map errors and correctness. However, NUMACK was more likely to result in efficient and rich information exchange, through the increased use of complementary gestures.

(3) Type of gesture had a strong effect on both performance and perception, with listener-perspective gestures resulting in reduced errors in both retelling and map-building, and in increased judgements of reliability, naturalness, informativeness, understandability, and human-likeness. Interestingly, while listener-oriented gestures evoked a strong positive result in participants, virtually none of the participants used listener-oriented gesture in their own direction-giving.

(4) Listener-perspective gestures and speaker-perspective gestures were differentially judged in the different kinds of embodied systems. While the robot was judged more positively than NUMACK when listener-perspective gestures were used, NUMACK was judged more positively than the robot in the no gesture condition. In fact, participants attributed empathy to the robot when it used listener-perspective gestures and to NUMACK when it used no gestures.

### Limitations

Our results demonstrate a gain for listener-perspective gesture in both performance and perception. This result is in accord with findings from (Okuno et al. 2009), although our research was also able to demonstrate a comparison not just with no gestures, but also with the more common conversational speaker-perspective gestures. However, while we looked at the presence of redundancy vs. complementarity in the gestures of our participants, we did not vary this factor in the gestures of our agents. Our future work will need to address this important aspect of embodied information exchange.

We compared a robot, which has physical co-locatability, and NUMACK, which has naturalness of movements and appearance, and found no differences in performance. One reason for this may be that we carefully controlled the speech to be identical across conditions, and the gesture to be as similar as possible, and we compared the two systems on a task where physicality and virtuality do not come into play - that is, directions to a place that is not visible to either participant. However, we did not control every potentially relevant variable. For example, the size of robot was slightly smaller than the projection of NUMACK, which may have disadvantaged the robot and, on the other hand, the robot's entire body was visible, while NUMACK was cut off below the torso.

### Conclusions

While it has been claimed that a visible map will trump any use of embodiment in direction giving, the research reported here demonstrated that embodied systems are judged more positively than a GPS map + speech system and, more importantly, that embodied systems carefully designed (with gestures that align with the user's perspective) are more

effective in reducing error and increasing correctness in direction-giving. Results on a map construction task demonstrate that not only are users better able to remember and repeat directions given by embodied systems with listener-perspective gesture, they are also better able to build (cognitive) maps of the spaces that were described to them.

We chose to evaluate the impact of embodiment, comparisons between different kinds of embodiment, and the effect of different aspects of embodiment, using a rich direction-giving task. Our methodology allowed us to examine perceptions of the agents, but also effects on participants' memory for the directions and their representation of the directions on a map. The methodology also allowed us to examine aspects of participants' behavior that contribute to their effectiveness as direction-givers themselves. In this regard one of the most evocative results is that only in the NUMACK condition did participants use complementary gesture (an example being one participant who said "you'll see a small building" and illustrated with her hands the position of the building with respect to the other landmarks). It is possible that participants were able to read more information from NUMACK's gestures with respect to the robot's gestures because NUMACK's gesture was more natural.

Looking at behavior, we also found that virtually no participants used listener-oriented gestures even though those gestures were most effective for performance on the direction-giving task, and were most highly rated along a number of different dimensions. We also found an interaction between gesture type and embodiment type, and it remains a puzzle as to why the listener-oriented gestures were rated more highly for the robot, while the no gesture condition was rated more highly for NUMACK. In addition to examining gesture redundancy vs. complementarity, this interaction effect will also be further examined in our future work. In the meantime, we hope to have demonstrated that the comparison between visible maps (GPS systems) and human-like behavior in embodied agents is not as straightforward as it has been portrayed, nor is the comparison between robots and ECAs a clear win for either side. What is clear is that the details of embodiment and their relationship to task and talk are central to the design of embodied systems.

## References

- Black, A. W., and Taylor, P. 1999. *The Festival Speech Synthesis System*. University of Edinburgh.
- Cassell, J.; Bickmore, T.; Cambell, M.; Chang, K. L.; and Vilhjalmsson, H. 1999. Embodiment in conversational interfaces: Rea. In *Proceedings of the CHI'99 Conference*, 520–527.
- Cassell, J.; Bickmoreand, T.; Nakano, Y.; Ryokai, Y.; Tversky, D. K.; Vaucelle, C. D.; and Vilhjalmsson, H. 2002. Mack: Media lab autonomous conversational kiosk. In *Proceedings of Imagina02*, 224–225.
- Kidd, C., and Breazeal, C. 2004. Effect of a robot on user perceptions. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'04)*, 3559–3546, 4.
- Kopp, S.; Tepper, P. A.; Ferriman, K.; and Cassell, J. 2007. Trading spaces: How humans and humanoids use speech and gesture to give directions. In *T. Nishida (ed.), Conversational Informatics (John Wiley, 2007), chap. 8*, 133–160.
- Nakano, Y. I.; Reinstein, G.; Stocky, T.; and Cassell, J. 2003. Towards a model of face-to-face grounding. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics*, 553–561.
- Okuno, Y.; Kanda, T.; Imai, M.; Ishiguro, H.; and Hagita, N. 2009. Providing route directions: design of robot's utterance, gesture, and timing. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, 53–60. New York, NY, USA: ACM.
- Ono, T.; Imai, M.; and Ishiguro, H. 2001. A model of embodied communications with gestures between humans and robots. In *Proceedings of the Annual Meeting of the Cognitive Science Society (CogSci2001)*, 732–737.
- Powers, A.; Kiesler, S.; Fussell, S.; and Torrey, C. 2007. Comparing a computer agent with a humanoid robot. In *Proceedings of the ACM/IEEE international conference on Human-robot interaction (HRI'07)*, 145–152. New York, NY, USA: ACM.
- Shinozawa, K.; Naya, F.; Yamato, J.; and Kogure, K. 2005. Differences in effect of robot and screen agent recommendations on human decision-making. *International Journal of Human-Computer Studies* 62(2):267–279.
- Stiegnitz, K.; Lovett, A.; and Cassell, J. 2001. Knowledge representation for generating locating gestures in route directions. In *Proceedings of the Workshop on Spatial Language and Dialogue (5th Workshop on Language and Space)*, 732–737.
- Stocky, T., and Cassell, J. 2002. Shared reality: Spatial intelligence in intuitive user interfaces. In *Proceedings of Intelligent User Interfaces*, 224–225.
- Taylor, H. A., and Tversky, B. 1992. Spatial mental models derived from survey and route descriptions. *Journal of Memory and Language* 31:261–292.
- Taylor, H. A., and Tversky, B. 1996. Perspective in spatial descriptions. *Journal of Memory and Language* 35:371–391.
- Theune, M.; Hofs, D.; and van Kessel, M. 2007. The virtual guide: A direction giving embodied conversational agent. In *INTERSPEECH-2007*, 2197–2200.