The Effect of Head-Nod Recognition in Human-Robot Conversation

Candace L. Sidner* , Christopher Lee*, Louis-Philippe Morency**, Clifton Forlines*
Mitsubishi Electric Research Labs * and MIT CSAIL**
Cambridge MA 02139
+1 617-621-7594, +1 617-621-7585, +1 617-253-4278, +1 617-621-7553
{sidner, lee, forlines}@merl.com, lmorency@csail.mit.edu

ABSTRACT
This paper reports on a study of human participants with a robot designed to participate in a collaborative conversation with a human. The purpose of the study was to investigate a particular kind of gestural feedback from human to the robot in these conversations: head nods. During these conversations, the robot recognized head nods from the human participant. The conversations between human and robot concern demonstrations of inventions created in a lab. We briefly discuss the robot hardware and architecture and then focus the paper on a study of the effects of understanding head nods in three different conditions. We conclude that conversation itself triggers head nods by people in human-robot conversations and that telling participants that the robot recognizes their nods as well as having the robot provide gestural feedback of its nod recognition is effective in producing more nods.

Categories and Subject Descriptors
H.5.2 Information systems: **User Interfaces**, H5.1 Information systems: **Multimedia**, I.2.9 **Robotics**

General Terms: Measurement, Performance, Design, Experimentation, Human Factors.

Keywords: Human-robot interaction, collaborative conversation, nodding, conversational feedback, nod recognition.

1. INTRODUCTION
This paper reports on how people use visual feedback while conversing with robots. In face-to-face conversation, people use gestural information that provides feedback as part of the communication. One prototypical form of feedback is head nodding at the conversational partner. Nodding is used to support generally is accompanied by linguistic phrases such as "yes, uh-huh, mm-hm, right, okay" and the like, is used to supporting grounding, that is, the sharing of information considered in common ground between conversants [1], to answer yes/no questions, and to emphasize agreement with the conversational partner. Nodding among American speakers generally is accompanied by linguistic phrases such as "yes, uh-huh, mm-hm, right, okay," and the like, but nodding can also occur as the only communicative expression.

HRI'06, March 2–4, 2006, Salt Lake City, Utah, USA.

When people converse with robots, much is yet unknown about how they will produce gestural feedback to the robot. In particular, for head nods, it is unclear whether they would nod at the robot at conversationally appropriate times, and under what general circumstances they might do so. This paper begins to address these matters and reports on a set of experiments concerning nodding by human conversants with a conversational robot.

2. BACKGROUND
During a previous set of experiments to determine the effects of physical movement and gestures in conversation [2], we noticed that participants in our experiments nodded at the robot. In one condition, participants conversed with a robot whose mouth always moved (to indicate speaking) and whose body moved. In an alternate condition, they conversed with a robot whose body did not move except for mouth movement. Participants in both conditions nodded at the robot, even though the robot had no means whatsoever to understand nods, and though no participants were told the robot had any such understanding abilities. We found this behavior surprising, and wished to investigate it further.

Previous work in human-robot interaction has largely explored gaze and basic interaction behavior. Breazeal's work [3] on infantoid robots explored how the robot gazed at a person and responded to the person's gaze and prosodic contours in what might be called pre-conversational interactions. Other work on infantoid robot gaze and attention can be found in [4]. Minato et al [5] explored human eye gaze during question answering with an android robot; gaze behavior differed from that found in human-human interaction. More recent work [6] explores conversation with a robot learning tasks collaboratively, but the robot cannot interpret nods during conversation. Ishiguro et al [7] report on development of Robovie with reactive transitions between its behavior and a human's reaction to it; they created a series of episodic rules and modules to control the robot's reaction to the human. However, no behaviors were created to interpret human head nods. In other work, Sakamoto et al [8] have experimented with cooperative behaviors on the part of the robot (including robot nodding but not human nodding) in direction giving.

Other studies have explored gaze and nodding in conversations with embodied conversational agents (ECAs), that is, on-screen 3D animated characters. Nakano et al [9] found that gaze at a map and lack of negative feedback were indicative that humans considered the previous utterance grounded and only looked at the interlocutor when they required more information to ground. They

reproduced this same behavior for an ECA. Researchers reporting in [10] have developed ECAs that produce gestures in conversation, including facial gestures, but to date, none have incorporated nod recognition in their interactions. Fujie et al [11] developed a nod and shake recognition algorithm and used it to improve the robot's interpretation of attitude in utterances that had been interpreted based on prosody alone. Morency et al [12,13] compared several different algorithms to interpret head nods and also incorporated speaking context to the interpretation. On the general topic of people and computer interactions, numerous studies (for example, [14,15]) have shown that people readily accept computers and embodied agents as social agents with whom they can interact.

3. A CONVERSATIONAL ROBOT

This research builds upon previous work with a humanoid robot depicted as a penguin, developed at MERL. Our research is focused on creating robots, with engagement capabilities [2, 16]. By engagement, we mean the process by which two (or more) participants establish, maintain and end their perceived connection. This process includes: initial contact, negotiating a collaboration, checking that the other is still taking part in the interaction, evaluating whether to stay involved, and deciding when to end the connection. The robot we have developed interacts with a single user in a collaboration that involves: spoken language (both understanding and generation), beat gestures with its arm, head gestures to track the user and to turn to look at objects of interest in the interaction, recognition of user head gestures in looking at objects, and recognition of user head nods. The robot also initiates interactions with users, and performs typical preclosings and goodbyes to end the conversation. All these capabilities increase the means by which the robot can engage the user in an interaction.

The robot, called Mel, has the following hardware:

- 7 DOF in the body (1 in the beak/mouth, 2 in the head, 2 in each of the 2 wings),
- body mounted on a Pioneer II mobile robot platform for floor navigation and body positioning,
- stereo camera,
- 2 far distance microphones (one for speech recognition, one for speech detection),
- 2 onboard laptop computers and an onboard PC-104 computer for all software.

The robot is able to converse with human participants by means of an architecture that includes (1) a conversational subsystem with speech recognition software, a conversational model based on Collagen ™ [17] and (2) a sensorimotor subsystem for sensor fusion of data from sound analysis and visual analysis and for robot control. The architecture is unusual because the two subsystems rely on information in the other to make decisions about the conversation state and robot control (see [2] for complete details). This architecture has been extended from previous work with a head gaze and gesture recognition software developed at the MIT Computer Sciences and Artificial Intelligence Laboratory [12,13]. The robot is depicted in Figure 1.

Our robot participates in two different interactions with human participants. The first of these is a demonstration of IGlassware, a technology developed at MERL with a specially designed cup that senses fluid level using capacitance and communicates the level to a special table by inductance. Mel was programmed to demonstrate this invention to a human visitor. A typical interaction is shown in Figure 2; the robot's gestures with its head towards the table are noted in the text. However, many different conversations are possible, including one that is not about the demo at all (if the human participant refuses to be shown a demo).

Figure 1 Mel the penguin robot

M: Hi, I'm Mel a robotic penguin.
U: Hi.
M: What's your name?
U: Sam.
M: Sam, I'd like to show you a demo. OK?
U: No.
M: But it's really interesting. Come on. Try it!
U: Okay.
M: It would be really nice to know what kind of progress people are making in their dining.
….
M: IGlassware stands for Instrumented glassware!
U: Okay.
M: Sam, right there <robot turns head to table> is the IGlassware cup <robot turns head to user>.
U: Ok.
….
M: All right, now we'll see how to use the cup.
U: All right.
M: Please pour enough water from the pitcher into the glass to fill the cup. Then put it on the table.
U: Okay.
……
M: I'm waiting for a reading from the table for the glass.
M: Good. Take a look at the bar on the readout.
U: Yes.
M: It shows that the cup is full, right?
U: Yes.

……
M: Would you like me to explain how this works?
U: Sure.
M: You see the copper in the glass? <robot turns head to glass>
…….
M: Well, Sam, that's about all. See Candy Sidner or Chris Lee for more about IGlassware. So long!
U: Good-bye.

Figure 2. Excerpts from the IGlassware Demo with Mel

The robot can also demonstrate its own capabilities and does so in an interactive conversation. A portion of one such conversation can be found in Figure 3. During this conversation, the robot gestures at its body parts when relevant, and in the case of its mobile base, moves around to demonstrate its locomotive abilities. We have observed that people interacting with the robot during this conversation maintain their gaze on the robot through most of the conversation with relatively few looks away (in contrast to the IGlassware conversations in which users direct much of their attention to the IGlassware cup and table). We surmise that gaze maintenance is due to the fact that there is nothing else that is relevant to the conversation, and due to the fact that the robot is pointing out parts of itself in the conversation.

M: Hello, I'm Mel, a robotic penguin.
…….
M: Mike, I'd like to tell you about myself. okay?
U: Ok.
M: First I'll explain my body.
U: Ok.
M: Well. I have 7 degrees of freedom in my body parts plus my mobile base can turn wherever I want it to. <robot rotates base during conversation>
U: Ok.
..……
M: How about we discuss my vision and speech systems?
U: Ok.
M: Shall I explain my vision system,, or my speech abilities first?
U: Your vision system.
M: Well,, above my head is a black camera with two lenses. Do you see it? <robot looks up above head during this exchange>
U: yes
M: I see with my two lens eyes.
..…….
M: Well, that's all my parts. Are you curious about what I can do with all that?
U: yes
M: Mostly I can talk about things. And I give demos. And I'd like to learn how to tell jokes well.
U: Ok.
M: So you wanna hear a joke?
U: No.
M: Well, Mike that's about all.
U: Ok.
M: See Candy Sidner or Chris Lee, for more about me! So long!
U: Good-bye.
M: Good-bye.

Figure 3: Excerpts from a Mel self demo

When initiating both interactions, Mel uses his vision system to find a conversational partner. Thereafter Mel tracks the conversational partner's face and adjusts his "gaze" towards the partner, even when the partner moves about. Mel has eyes in his head, but they do not see, as his cameras are above his head, so his gaze merely communicates his focus of attention to the partner. Mel does not look at the human partner at all times. During the demos, he turns to look at the table and its contents or to parts of himself. Mel also prompts a partner who fails to look at the table to notice the objects there. After the demo and explanation conclude, Mel wishes the partner goodbye, waves and drops his head to his chest to indicate that he is no longer available.

4. EXPERIMENTAL APPROACH

Our experiments were carried out in the context of research on engagement between humans and robots. Head nodding is one of the many means by which participants indicate that they are engaging in the conversation. When used to ground the previous turn of the other participant, head nodding and accompanying linguistic gestures tell the last participant that their information is being admitted to the common ground of the participants (cf [1]). When used to answer affirmatively in a yes/no question, nodding provides supporting information to the affirmative answer.[1] Answers also convey that the conversation should continue. When nodding is used to emphasize agreement, it serves as an indirect communication that the conversation is still relevant to the nodder. Hence, nodding is an engaging behavior. Furthermore, because it is a natural one that occurs in normal human conversation, it behooved us to consider it as a behavior that a robot could interpret.

However, we are unaware of any research on the effects of robot recognition of head nods in human-robot conversation. To explore this problem, our robot held conversations with human participants during which it recognized human head nods. The goal of the research was to determine if a robot's recognition of head nods caused any effect on its human conversational partners. Since our human participants nodded at a robot during a hardware demonstration conversation when the robot did not understand nods [2], we sought to determine if they would nod more when the robot actually understood their nods. If not, then one could conclude that nodding would occur as long as conversation was present between the two participants. In addition, we wanted to know if there were effects resulting from people knowing that the robot could nod and getting feedback during the conversation about the fact that the robot recognized a nod. Thus we wanted to compare the situations where the participant had a robot with no nod recognition with one who did and with one in which the human knew the robot's ability and got feedback about it during the conversation.

5. EXPERIMENTAL STUDY

The robot was equipped with a stereoptic camera and used the Watson system and associated algorithms for nod recognition reported by Morency et al [12,13]. The robot used the conversational system and architecture described earlier in the

[1] We are not aware of any literature that provides statistics on the frequency of nods only in human conversation.

paper. Head nods were reported from the sensorimotor subsystem to the vision system in the same way that other gestures (for example, looking at an object) were.

Participants held one of two conversations with the robot, one to demonstrate either its own abilities or to demonstrate collaboratively the IGlassware equipment. During these conversations people nodded at the robot, either because it was their means for taking a turn after the robot spoke (along with phrases such as "ok" or "yes" or "uh-huh"), or because they were answering in the affirmative a yes/no question and accompanied their linguistic "yes" or "ok" with a nod. Participants also shook their heads to answer negatively to yes/no questions, but we did not study this behavior because too few instances of no answers and headshakes occurred in our data. Sometimes a nod was the only response on the part of the participant.

A total of 49 participants interacted with the robot. None had interacted with our robot before. Most had never interacted with any robot. One participant had an abbreviated conversation because the robot misrecognized the user's intention and finished the conversation without a complete demo. However, the participant's conversation was long enough to include in the study. Thirty-five participants held the IGlassware demo with the robot, and fourteen participants held the self demo.

The participants were divided into two groups, called the MelNodsBack group and the MelOnlyRecognizesNods group. The MelNodsBack group with fifteen participants, who were told that the robot understood some nods during conversation, participated in a conversation in which the robot nodded back to the person every time it recognized a head nod. It should be noted that nodding back in this way is not something that people generally do in conversation. People nod to give feedback on what another has said, but having done so, their conversational partners only rarely nod in response. When they do, they are generally indicating some kind of mutual agreement. Nonetheless, by nodding back the robot gives feedback to the user on their behavior. Due to nod mis-recognition of nods, this protocol meant that the robot sometimes nodded when the person did not nod.

The MelOnlyRecognizesNods group with fourteen subjects held conversations without knowledge that the robot could understand head nods although the nod recognition algorithms were operational during the conversation. We hypothesized that participants might be affected by the robot's nodding ability because 1) when participants nodded and spoke, the robot took another turn whereas without a response (verbal and nod), the robot waited a full second before choosing to go on and similarly, 2) when participants responded only with a nod, the robot took another turn without waiting further. Again nod mis-recognition occurred although the participants got no gestural feedback about it. The breakdown of groups and demo types is illustrated in Figure 4.

These participants are in contrast to a base condition called the NoMelNods group, with 20 subjects who interacted with the robot in a conversation in which the robot did not understand nods, and the participants were given no indication that it could do so. This group, collected in 2003, held only the IGlassware equipment conversation with the robot.

	IGlassware	Self	Total
MelNodsBack	9	6	15
MelOnlyRecognizesNods	6	8	14
NoMelNods	20	0	20

Figure 4: Breakdown of participants in groups and demos

Protocol for the study: The study was a between subjects design. Each participant was randomly pre-assigned into one of the two nodding conditions (that is, no subjects had conversations in both nodding conditions). Video cameras were turned on after the participant arrived. The participant was introduced to the robot (as Mel) and told the stated purpose of the interaction (i.e. to have a conversation with Mel). Participants were told that they would be asked a series of questions at the completion of the interaction. Participants were also told what responses the robot could easily understand (that is, "yes, no, okay, hello, good bye," their first names, and "please repeat"), and in the case of the MelNodsBack condition, they were told that the robot could understand some of their nods, though probably not all. They were not told that the robot would nod back at them when they nodded. Participants in the 2003 study had been told the same material as the MelOnlyRecognizesNods participants.

When the robot was turned on, the participant was instructed to approach Mel. The interaction began, and the experimenter left the room. Interactions lasted between 3 and 5 minutes. After the demo, participants called in the experimenter and were given a short questionnaire, which is not relevant to the nodding study.

6. ISSUES IN DATA ANALYSIS

Every conversation was videotaped and annotated for participant utterances in the conversation, change of gaze of the participants, participant head nods and in the MelNodsBack condition, robot head nods. The number of nods on the part of a participant in a conversation varies greatly from participant to participant for reasons we have yet to understand. Some participants nod freely in conversation, as many as 18 times, while some do not--as few as 0, 1, or 2 times. However, as we discovered, determining when head nods occurred is extremely difficult.

The nature of nodding is far more complex than expected. Different participants have different types of head nods (for example, head up then down, head down then up, head toss from the side up and then down, etc), the duration of their nods varies (1 up/down versus several), and angle of the nod varies greatly. In fact, our informal sense is that nods vary from very small angles (3-4 degrees) for acknowledgement of what the robot says, to larger angles for yes answers, to large swings of the head when expresses emphatic affirmation or agreement. Furthermore, because participants are not always looking directly at the robot, they may nod as they turn to look at another object, while they are looking at another object, or as they turn back to the robot from viewing another object.

We found that a single annotator was not sufficient for interpreting head nods because that one annotator found additional nods in a second viewing after viewing several different subjects. Therefore, two annotators annotated all 49 videos. However, the annotators did not always agree. For some participants, annotation agreement was quite high (complete agreement or as many as 15/17 nods). For other participants, a

total of 9 of the 29 participants, agreement was much lower at about 50%. To control for this variation, our results make use of only those instances where annotators agreed.

The variation in head nod detection by human annotators helps explain why the vision alone nod recognition algorithm, based on SVM learning techniques, also misrecognized head nods. While we collected 30 additional conversations (not reported in this work) of participants talking to our robot, in order to improve head gesture recognition using dialog context [13, 18], the robot's ability to understand nods for our study was less than perfect. The MelNodsBack group had a mean nod recognition rate of about 48%, and the MelOnlyRecognizesNods group had a mean recognition rate of 41%. A comparison of means between these two groups shows no significant difference in nod recognition rate ($t(25) = 0.82$, $p = 0.42$).

Every conversation with the robot in our total of 49 participants varied in the number of exchanges held. Hence every participant had a varying number of opportunities to give feedback with a nod depending on when a turn was taken or what question was asked. This variation was due to: different paths through the conversation (when participants had a choice about what they wanted to learn), the differences in the demonstrations of IGlassware and of the robot itself, speech recognition (in which case the robot would ask for re-statements), robot variations in pausing as a result of hearing the user say "ok," and instances where the robot perceived that the participant was disengaging from the conversation and would ask the participant if they wished to continue.

In order to normalize for these differences in conversational feedback, we coded each of the individual 49 conversations for feedback opportunities in the conversation. Opportunities were defined as the end of an exchange where the robot paused long enough to await a response from the participant before continuing, or exchange ends where it waited only briefly but the participant chose to interject a verbal response in that brief time.

So for each participant, the analysis below uses a "nod rate" as a ratio of total nods to feedback opportunities, rather than the raw number of nods in an individual conversation. Furthermore, the analysis makes three distinctions: nod rates overall, nod rates where the participant also uttered a verbal response (nod rates with speech) and nod rates where no verbal response was uttered (nod only rates).

7. EXPERIMENTAL RESULTS

Our study used a between-subjects design with *Feedback Group* as our independent variable, and *Overall Nod Rate*, *Nod with Speech Rate*, and *Nod Only Rate* as our three dependent variables. In total, 49 people participated in our study, fifteen in the MelNodsBack group, fourteen in the MelOnlyRecognizesNods group. An additional twenty participants served in the NoMelNods group.

A one-way ANOVA indicates that there is a significant difference among the three feedback groups in terms of Overall Nod Rate ($F_{2,46} = 5.52$, $p < 0.01$). The mean Overall Nod Rates were 42.3%, 29.4%, and 20.8% for MelNodsBack, MelOnlyRecognizesNods, and NoMelNods groups respectively. A post-hoc LSD pairwise comparison between all possible pairs shows a significant difference between the MelNodsBack and the NoMelNods groups ($p=0.002$). No other pairings were significantly different. The mean Overall Nod Rates for the three feedback groups are shown in Figure 5.

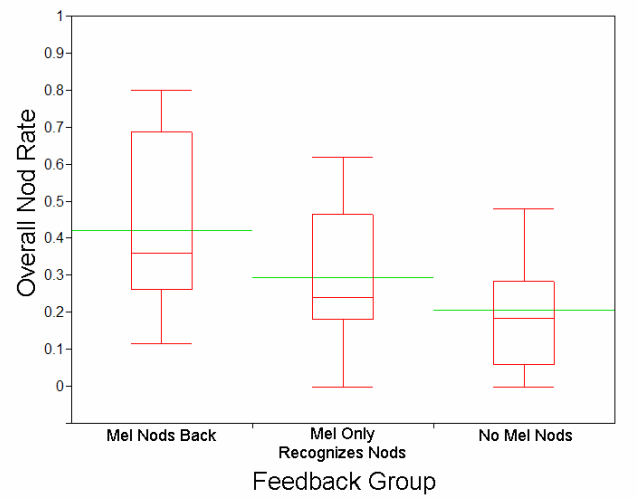

Figure 5: Overall Nod Rates by Feedback Group. Subjects nodded significantly more in the MelNodsBack feedback group than in the NoMelNods group. The mean Overall Nod Rates are depicted in this figure with the wide lines.

A one-way ANOVA indicates that there is also a significant difference among the three feedback groups in terms of Nod with Speech Rate ($F_{2,46} = 4.60$, $p = 0.02$). The mean Nod with Speech Rates were 32.6%, 23.5%, and 15.8% for the MelNodsBack, MelOnlyRecognizesNods, and NoMelNods groups respectively. Again, a LSD post-hoc pairwise comparison between all possible pairs of feedback groups shows a significant difference between the MelNodsBack and NoMelNods groups ($p=0.004$). Again, no other pairs were found to be significantly different. The mean Nod with Speech Rates for the three feedback groups are shown in Figure 6.

Finally, a one-way ANOVA found no significant differences among the three feedback conditions in terms of Nod Only Rate ($F_{2,46} = 1.08$, $p = 0.35$). The mean Nod Only Rates were much more similar to one another than the other nod measurements, with means of 8.6%, 5.6 and 5.0% for the MelNodsBack, MelOnlyRecognizesNods, and NoMelNods groups respectively. The mean Nod Only Rates for the three feedback groups are shown in Figure 7.

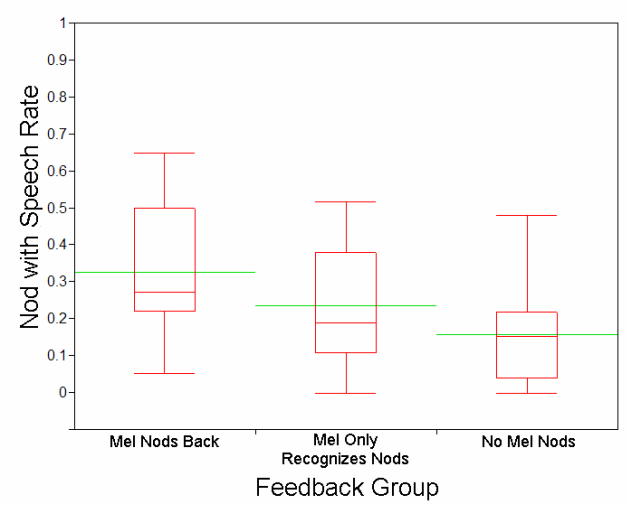

Figure 6: Nod with Speech Rates by Feedback Group. Again, subjects nodded with speech significantly more frequently in the MelNodsBack feedback group than in the NoMelNods group.

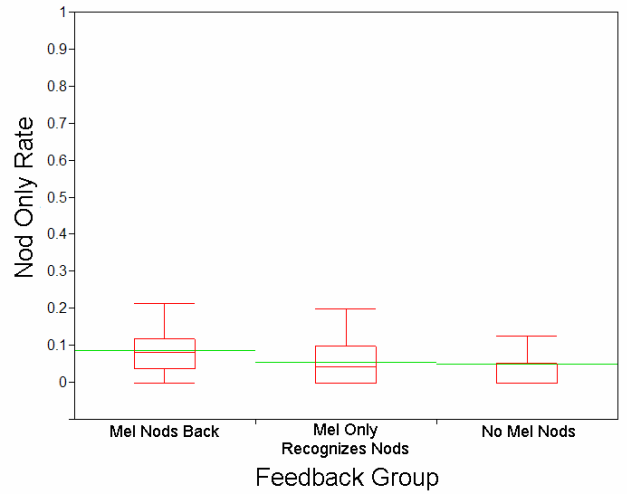

Figure 7: Nod Only Rates by Feedback Group. There were no significant differences among the three feedback groups in terms of Nod Only Rates. [2]

7.1 Discussion:

These results above indicate that under a variety of conditions people will nod at a robot as a conversationally appropriate behavior. Furthermore, these results show that even subjects who get no feedback about nodding do not hesitate to nod in a conversation with the robot. We conclude that conversation alone is an important feedback effect for producing human nods, regardless of the robot's ability to interpret it.

It is worthwhile noting that the conversations our participants had with robots were more than a few exchanges. While they did not involve the human participants having extended turns in terms of verbal contributions, they did involve their active participation in the purposes of the conversation. Future researchers exploring this area should bear in mind that the conversations in this study were extensive, and that ones with just a few exchanges might not see the effects reported here.

The two statistically significant effects for nods overall and nods with speech that were found between the NoMelNods group and the MelNodsBack group indicate that providing information to participants about the robot's ability to recognize nods and giving them feedback about it makes a difference in the rate at which they produce nods. This result demonstrates that adding perceptual abilities to a humanoid robot that the human is aware of and gets feedback from provides a way to affect the outcome of the human and robot's interaction.

It is important to consider the fact that the participants in this study were novices at human-robot conversational interaction. Their expectations about their interactions with the robot only had a few minutes to develop and change. It may well be that in multiple interactions over a long period of time, people will infer more about the robot's abilities and be able to respond without the need for the somewhat artificial gestural feedback we chose.

The lack of statistical significance across the groups for nod rates without any verbal response (nod only rates) did not surprise us. The behavior of only nodding in human conversation is a typical behavior, although there are no statistics we are aware of about the rates. It is certainly not as common as nodding and making a verbal response as well. Again, it is notable that this behavior occurs in human-robot interaction and under varying conditions. By count of participants, in the NoMelNods group, 9 of 20 participants, nodded at least once without speech, in the MelOnlyRecognizesNods group, 10 of 14 did so and in the MelNodsBack group, 12 of 15 did so. However, when using normalized nod rates for this behavior for each group, there is no statistical difference. This lack is certainly partially due to the small number of times subjects did this: the vast majority participants (44/49) did a nod without speech only 1, 2, or 3 times during the entire conversation. A larger data set might show more variation, but we believe it would take a great deal of conversation to produce this effect.

8. CONCLUSIONS

Conversation is a powerful device for eliciting nods from human participants in human robot interaction. It is powerful enough to elicit nods even when the robot cannot interpret the nods for their conversational feedback purposes. In fact, we found no statistical difference in the amount of head nodding when the robot recognized a head nod from when it could not, a finding that countered our initial hypothesis. However, when human participants know that their robot partner recognizes nods and when they also receive gestural feedback for their nods, they nod more than when the robot cannot understand their nods, where the difference is statistically significant.

[2] There is a slight difference in the totals of the MelNodsBack group for rates of total nods, nods with speech and nods only. The 1.1% error is due to 3 subjects who nodded without speech and without a taking a turn—they were nods back to the robot's nods (which were in response to human nods). This slightly "recursive" response was very amusing to see, and told us that the participants paid attention to the robot's nods. We coded these nods in total nods, but did not code them with the other without speech nods.

Building robots that can converse and interpret human gestures may not be enough; people must have a means of learning what the robot can do, at least until robots and people interact often in everyday life. Until then, feedback that indicates when the robot understands gestures are useful just as feedback is for when the robot understands spoken utterances.

9. ACKNOWLEDGMENTS

Our thanks to Tira Cohene and Charles Rich for assistance in annotating video sequences.

10. REFERENCES

[1] Clark, H.H., *Using Language*. Cambridge University Press, Cambridge, 1996.

[2] Sidner, C., Lee, C., Kidd, C., Lesh, N. and Rich, C. Explorations in Engagement for Humans and Robots, *Artificial Intelligence*, 166(1-2): 140-164, August, 2005.

[3] Breazeal, C. and Aryananda, L. Recognizing affective intent in robot directed speech, *Autonomous Robots*, 12:1, pp. 83-104, 2002.

[4] Miyauchi, D., Sakurai, A., Makamura, A., Kuno, Y. *Active eye contact for human-robot communication. In: Proceedings of CHI 2004--Late Breaking Results*. Vol. CD Disc 2. ACM Press, pp. 1099—1104, 2004.

[5] Minato, T., MacDorman, K., Simada, M., Itakura, S., Lee, K. and Ishiguro, H. Evaluating humanlikeness by comparing responses elicited by an android and a person, *Proceedings of the Second Interantional Workshop on Man-Machine Symbiotic Systems*, pp. 373-383, 2002.

[6] Breazeal, C., Brooks, A., Gray, J., Hoffman, G., Kidd, C., Lee,H., Lieberman, J., Lockerd, A., and Chilongo, D. Tutelage and collaboration for humanoid robots. International Journal of Humanoid Robotics, Vol. 1, No. 2 (2004) 315-348.

[7] Ishiguro, H., Ono, T., Imai, M., and T.Kanda, 2003. Development of an interactive humanoid robot “Robovie”---an interdisciplinary approach. In: Jarvis, R.A., Zelinsky, A. (Eds.), *Robotics Research.* Springer, pp. 179--191.

[8] Sakamoto, D., Kanda, T., Ono, T., Kamashima, M., Imai, M., and Ishiguro, H. Cooperative embodied communication emerged by interactive humanoid robots. *Int. J. Human.-Computer Studies.* 62(2): 247-265, 2005.

[9] Nakano, Y., Reinstein, G., Stocky, T., and Cassell, J. Towards a model of face-to-face grounding. In: *Proceedings of the 41st meeting of the Association for Computational Linguistics*. Sapporo, Japan, pp. 553—561, 2003.

[10] Gratch, J., Rickel, J., Andre, E., Badler, N., Cassell, J., and Petajan, E., 2002. Creating interactive virtual humans: Some assembly required. IEEE Intelligent Systems, 54--63.

[11] Fujie, S. Ejiri, Y., Nakajima,K., Matsusaka, Y. and Kobayashi, T. A Conversation Robot Using Head Gesture Recognition as Para-Linguistic Information, *Proc. 13th IEEE Intl. Workshop on Robot and Human Communication, RO-MAN 2004*, pp.159-164, Kurashiki, Japan, Sept. 2004.

[12] Morency, L.-P., and Darrell, T. From conversational tooltips to grounded discourse: head pose tracking in interactive dialog systems, *Proceedings of the International Conference on Multi-modal Interfaces*, pp. 32--37, State College, PA, October, 2004.

[13] Morency, L.-P., Lee, C., Sidner, C., and Darrell, T. Contextual recognition of head gestures, *Proceedings of the Seventh International Conference on Multimodal Interfaces (ICMI'05)*, pp.18-24, 2005.

[14] Reeves, B., and Nass, C., *The media equation: how people treat computers, television, and new media like real people and places*, Cambridge University Press New York, NY, USA, 1996.

[15] Lee, K. M. and Nass, C. (2003). Designing social presence of social actors in human computer interaction. *Proceedings of the Computer-Human Interaction (CHI) Conference, 2003*, ACM Press.

[16] Sidner, C.L. and Dzikovska, M. Human-Robot Interaction: Engagement Between Humans and Robots for Hosting Activities. In the *Proceedings of the IEEE International Conference on Multimodal Interfaces*, pp. 123-128, 2002.

[17] Rich, C.; Sidner, C.L., and Lesh, N.B., "COLLAGEN: Applying Collaborative Discourse Theory to Human-Computer Interaction", *Artificial Intelligence Magazine*, Winter 2001 (Vol 22, Issue 4, pps 15-25).

[18] Morency, L.-P., Sidner, C., and Darrell, T. Towards context-based visual feedback Recognition for Embodied Agents, *Proceedings of the Symposium on conversational Informatics for Supporting Social Intelligence and Interaction*, AISB-05, pp 69-72, 2005.