Assessing Differences in Multimodal Grounding with Embodied and Disembodied Agents

Eugenia Hee, Ron Artstein, Su Lei, Cristian Cepeda, David Traum

University of Southern California, Institute for Creative Technologies, Los Angeles, CA
Email: {ehee, artstein, slei, ccepeda, traum}@ict.usc.edu

1 Introduction

Establishing common ground is an essential part of any collaboration process and can be critical in the success of the desired task at hand. With the increased introduction of artificial agents into society, understanding the way that we interact with both embodied and disembodied versions of these agents becomes even more critical. While people are getting more comfortable with using machines for accessing information and providing services, it is less clear to what degree people strive for common ground with these machines and provide feedback related to their reactions to the provided information. We look at the question of how people provide grounding-related feedback when in conversation with a robot and a virtual human in a variety of tasks and modalities. We examine several different types of activities, including first-contact social dialogue, and several item-ranking tasks, in which participants can reveal their own rankings and rationales and potentially influence others. We also examine several kinds of feedback, including positive and negative signals of understanding and agreement. Finally, we examine verbal utterances and non-verbal signals for these functions. We look at whether different tasks or agent types influence the amount and modalities of different kinds of feedback behaviors. We also look at whether feedback patterns are correlated with different amounts of influence that the agents exert on humans.

2 Methods

[1] reports on an experiment looking at rapport and influence with different kinds of artificial agents and different tasks. The recordings from this experiment provided an opportunity to examine various feedback behaviors. In this data, 40 participants interact with both Niki, a Nao Robot, and Julie a virtual human (Figure 1) in one icebreaker task and three ranking tasks.
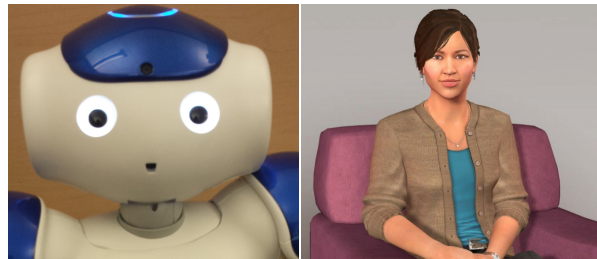

Figure 1: Niki and Julie

The ranking tasks consisted of lunar and desert survival situations and a "Save the Art" task, and participants would interact with an agent to discuss each other's rankings of 10 items. Influence was calculated by how much the participants changed their rankings after the conversation with the agent. In the icebreaker task, participants would share information about themselves and learn more about the agent. Influence was not calculated in this task. Julie was visually present only in the icebreaker task, and available only by voice in the other tasks. Niki had a physical presence in all tasks.

We annotated videos of the participants using the ELAN software to make note of each type of feedback and trace two of the four basic communicative functions discussed in [2]: understanding and attitudinal reactions. We looked at positive and negative feedback at both understanding and agreement levels. The annotation scheme that we developed was

based on a simplified version of the MUMIN multimodal coding scheme [3]. Although there are a wide variety of signals used to display grounding, we chose to look at facial displays that included head shake, head nod, and eyebrow movement, as well as verbal feedback like utterance and laugh.

3 Results

In total across all experiments, the 40 human participants indicated understanding 3420 times and non-understanding only 106 times. The agents could not only establish common ground and share information with the human participants, but participants also behaved towards both systems with natural dialogue mechanisms. There was a significant difference between signs of agreement and disagreement $F(1,76)=3.06$, $p=0.08$, with agreement occurring much more frequently. In the icebreaker task, this could be attributed to the fact that icebreakers in human-agent interaction tend to mirror human-human icebreakers, where topics are easy to agree with and disagreement is small. Spoken utterances were the most common type of feedback and occurred 3078 times, while head shakes were the least common, occurring only 85 times.

When examining responses to Niki and Julie separately, there were significantly more head nods with Niki than with Julie, $F(1,114) = 4.79$, $p = 0.03$, perhaps because Julie was disembodied for all ranking tasks. There was not a significant difference in the amount of utterances between Niki and Julie, although utterances were a larger percentage of total feedback actions with Julie (Figure 2). While individual results varied from participant to participant, all participants had understanding as their most common action category. Within the different tasks, the icebreaker task had the highest level of understanding overall, with both Niki and Julie.

Analyzing the relationship between influence and feedback showed that Julie tended to

have a positive correlation between level of influence and number of feedback actions, but Niki did not. This could be because sometimes eye gaze and other feedback signs are used as for non-understanding or disagreement [4].
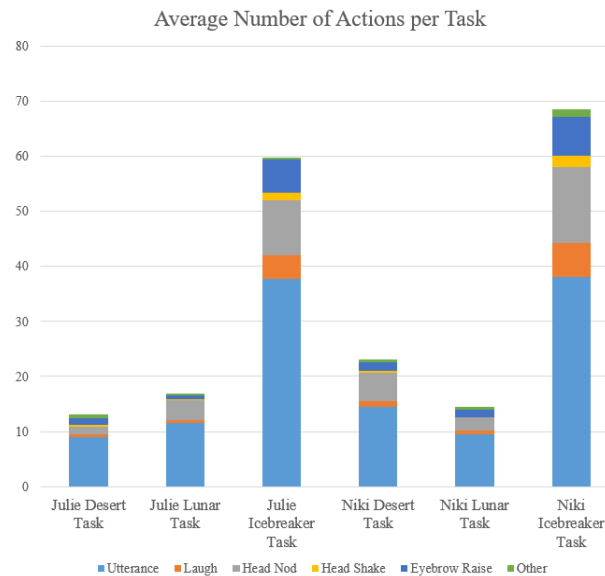

Figure 2: Difference in Feedback Actions

4 Conclusion

In summary, we discovered that there were more feedback actions overall with Niki, the embodied robot, than with Julie, the disembodied virtual agent. Participants tended to build feedback with non-verbal motions more with Niki, perhaps because he occupied a physical space and had a physical presence. Despite not having as many non-verbal actions as Niki, the Icebreaker task with an embodied version of Julie proved to have the highest level of understanding, perhaps because of the human-like conversation and the rapport-building nature of the task. It is important to note that Julie was also more human-like than Niki and this could have affected the results along with differences in embodiment. Moving forward, it would be interesting to look more into the relationship between feedback actions and level of rapport or trust, as well as underlying notions of conflict avoidance and how this varies between embodied and disembodied agents.

References

1. R. Artstein, D. Traum, J. Boberg, A. Gainer, J. Gratch, E. Johnson, A. Leuski, and M. Nakano. Listen to my body: Does making friends help influence people? In *Proceedings of the Thirtieth International Florida Artificial Intelligence Research Society Conference*, pages 430–435, May 2017

2. J. Allwood, J. Nivre, and E. Ahlsén. On the Semantics and Pragmatics of Linguistic Feedback. In *Journal of Semantics, 1992, Vol. 9 No. 1*, p. 1-26, 1992.

3. J. Allwood, L. Cerrato, K. Jokinen, C. Navarretta, and P. Paggio. The MUMIN Multimodal Coding Scheme. *NorFA yearbook,* p. 129-157, 2005.

4. Y. Nakano, G. Reinstein, T. Stocky, and J. Cassell. Towards a Model of Face-to-Face Grounding. In *ACL '03 Proceedings of the 41st Annual Meeting on Association for Computational Linguistics – Volume 1*, p. 553-561, July 2003.