Disfluency Handling for Robot Teammates

Felix Gervits
Tufts University
Medford, MA 02155
Felix.Gervits@tufts.edu

1. INTRODUCTION AND BACKGROUND

Robots are being increasingly used as partners in mixed-initiative teams with humans (e.g., urban search and rescue, space robotics, etc.). Given the emphasis that human teams place on verbal communication, much prior work in human-robot teaming has focused on the role of task-oriented dialogue as a means of *grounding*, or establishing mutual knowledge. However, grounding in typical human-robot task domains is complicated by various constraints, including time pressure, workload, and lack of visual access. Many of these factors have been shown to affect language in human teams, resulting in increased disfluency rates, miscommunication, and overlapping speech. Though some of these features can be seen as "noise" in the speech channel, others (such as disfluencies) may provide a benefit to the interaction. Evidence from the Psycholinguistic literature shows that disfluencies may actually serve a coordination function, both for speech production and comprehension [5].

The above studies suggest that human-robot teaming may benefit from the ability of the robot to interpret disfluencies in spontaneous speech. Since this is unexplored territory, my work addresses this challenge through both empirical and computational means. On the empirical end, I am conducting studies that test for the benefit of disfluencies on team coordination and performance in human-robot domains. On the computational end, my works seeks to move past traditional "detect-and-remove" approaches to disfluency handling [1], instead focusing on identifying the function of disfluent utterances to exploit their utility in the way that humans do.

2. INITIAL STUDY

My initial study addresses the empirical challenge by examining how disfluencies and grounding strategies interact in human teams. To this end, I used a unique corpus of spontaneous, task-oriented dialogue (CReST corpus [2]), which was annotated for disfluencies, and conversational moves. In CReST, human dyads performed a Cooperative Remote Search Task, in which they had to complete a variety of interdependent objectives while communicating through remote headset. The team members had asymmetrical roles, with one person serving as the director, and the other as the searcher. This task was designed to simulate the structure of teams in which a robot may play the searcher role, and so it has many of the factors that are of interest to us (e.g., remote communication, workload, hierarchical structure, etc.).

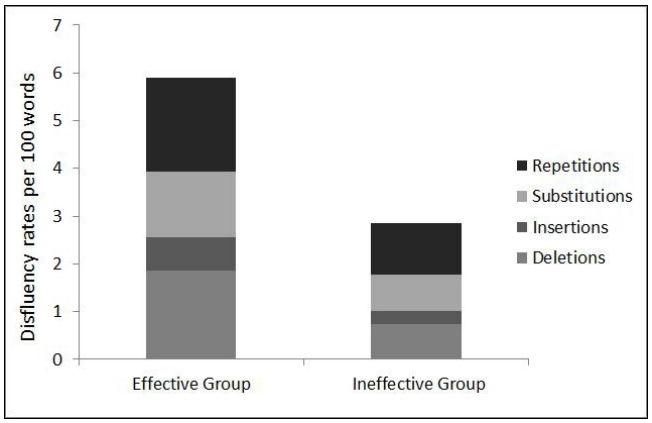

Figure 1: Group effect for disfluency (self-repair) rate

I performed a detailed quantitative analysis of this corpus which yielded novel results about factors that influence effective team communication [4]. The results showed that the best-performing teams used specific grounding strategies, including: establishing shared referents to describe locations, predicting and completing one another's turns, and taking their partners' perspective for referential descriptions. Effective teams also displayed specific dialogue patterns which helped to maintain common ground during periods of workload, including: showing greater responsiveness to their teammate (*Ready* moves), consistently seeking confirmation of understanding (*Check moves*), and repairing their own speech for clarity (*self-repairs*; see Fig. 1).

3. SELF-REPAIR FUNCTIONS

My study was the first to find that self-repairs (i.e., disfluencies) are used as grounding tools to facilitate coordination in naturalistic collaborative tasks. Though disfluencies for effective teams in the study facilitated grounding in various way (see [3]), here I will focus on two in particular - clarification and prediction - which can be exploited by robotic

HRI '17 Companion, March 06-09, 2017, Vienna, Austria

ACM ISBN 978-1-4503-4885-0/17/03.
DOI: http://dx.doi.org/10.1145/3029798.3034806

systems to improve coordination. Consider the following example from the CReST corpus which highlights a clarification disfluency:

(1) S: Well [pause] see the two pink boxes?
D: Yes
S: **On the right corner - the inside corner**
D: Yes

In this example, the bolded utterance contains an Insertion self-repair which was produced by the searcher in describing the location of a box. Though the searcher meant to describe the location as "on the right, inside corner", such an interpretation is only possible if the disfluency is identified as an Insertion in which "right" and "inside" are both treated as adjectival modifiers for "corner". Consider another example, highlighting the predictive benefit of disfluency:

(2) D: There's also one in the second - [pause] uh, we only have three minutes to do this, okay.
S: Okay, second cubicle I got that.

Here, the pause and hesitation marker at the end of the deleted segment served to indicate that the director was describing an object - which the searcher accurately guessed was a cubicle. This kind of fast-paced prediction from self-repairs was common in the corpus, and increasingly so for effective teams. However, existing dialogue systems would be unable to interpret it since they would simply delete the initial incomplete segment to produce a "clean transcript".

4. DISFLUENCY HANDLING

I have begun to implement mechanisms in a robotic architecture to identify the type and function of a disfluency. My approach involves the use of an incremental dependency parser that builds up a syntactic and semantic interpretation of partial utterances using the combinatorial categorial grammar (CCG) formalism. I am also developing mechanisms that allow the system to interpret basic types of self-repairs based on their type (see Table 1), along with specific parts of the utterance that correspond with the characteristic features of self-repairs: *reparandum*, *interregnum*, and *repair segment*. This process works slightly differently for each type of self-repair. For Substitutions, the reparandum is identified when the parser detects an ungrammatical segment or an interruption point (e.g., "uh/um"). Everything before the ungrammatical point is treated as the reparandum, which is replaced by the subsequent repaired segment. The parts of speech of each of these segments are tagged by the parser so that matching phrases can be identified in the reparandum and in the repair segment. Repetition self-repairs are identified by repetitive strings, whereas Insertions are identified by filled pauses ("uh/um") which serve as repair markers to indicate that the next word or segment is meant to replace the previous one. Finally, Deletions are detected through a combination of ungrammatical parses ("We don't have -") and repair marker placement ("uh"), which serves to indicate that the original segment was abandoned and replaced by a new one ("let's hurry up").

5. FUTURE WORK

The next step moving forward is to develop mechanisms that will enable a physical robot to utilize the knowledge gained from disfluency to improve interaction. For clarification self-repairs (Ex. 1), the additional information will need to be incrementally incorporated into the referential description, enabling the robot to add to it's representation. For prediction self-repairs (Ex. 2), the robot will need to update a search space that tracks the probability of the next word in the utterance given the partial parse and context.

Table 1: Four types of self-repair disfluencies

Disfluency Type	Example
Repetition	"Look- look in the box"
Substitution	"Pink- I mean blue box"
Insertion	"In the room- the nearby room"
Deletion	"We don't have- uh let's hurry up"

On the empirical side, follow-up experiments will be necessary in order to discover new ways in which disfluencies can improve coordination, and also to evaluate the mechanisms that I have developed. For evaluation, I will need to test the mechanisms in the context of an unscripted human-robot collaborative task. This type of evaluation will allow us to see if our preliminary human results apply to human-robot teams, and to test various hypotheses about how self-repair strategies affect team performance.

6. CONCLUSION

In order to make the goal of true human-robot teamwork a reality, robots need to exploit the wealth of information contained in natural language. My initial empirical work has identified some of this information, and shown the importance of it for supporting team coordination. This work serves as the basis to inform the types of mechanisms needed in collaborative robot teammates. Overall, I believe that these unique disfluency handling mechanisms will take us a step closer towards natural and effective robot teammates.

7. ACKNOWLEDGMENTS

This work was supported in part by ONR grants N00014-07-1-1049, N00014-11-1-0493, and N00014-14-1-0149.

8. REFERENCES

[1] R. Cantrell, M. Scheutz, P. Schermerhorn, and X. Wu. Robust spoken instruction understanding for HRI. In *Proceedings of the 2010 Human-Robot Interaction Conference*, pages 275–282, March 2010.

[2] K. M. Eberhard, H. Nicholson, S. Kübler, S. Gundersen, and M. Scheutz. The Indiana "Cooperative Remote Search Task" (CReST) corpus. In *Proceedings of the International Conference on Language Resources and Evaluation, LREC 2010, 17-23*, 2010.

[3] F. Gervits, K. Eberhard, and M. Scheutz. Disfluent but effective? a quantitative study of disfluencies and conversational moves in team discourse. In *Proceedings of the 26th International Conference on Computational Linguistics*, 2016a.

[4] F. Gervits, K. Eberhard, and M. Scheutz. Team communication as a collaborative process. *Frontiers in Robotics and AI*, 3:62, 2016b.

[5] V. L. Smith and H. H. Clark. On the course of answering questions. *Journal of Memory and Language*, 32(1):25–38, 1993.