Turn-Taking and Coordination in Commander-Robot Navigator Dialog Video Abstract

T. Cassidy and **C. Voss** and **D. Summers-Stay**

Computational and Information Sciences Directorate

Army Research Laboratory

Adelphi, MD

Abstract

The accompanying video captures the multi-modal data displays and speech dialogue of a human Commander (C) and a human Robot Navigator (RN) tele-operating a mobile robot (R) in a remote, previously unexplored area. We describe unique challenges for automation of turn-taking and coordination processes observed in the data.

Background Robots able to move into dangerous areas during emergencies and collaboratively explore and transmit timely information about these environments by teaming with remote humans have tremendous potential to impact search and rescue operations. Building such robots requires understanding of how human-robot dialogue interacts with processing of multi-modal perceptual data and motoric processes.

This video excerpt presents data collected in our previous study (Summers-Stay, Cassidy, and Voss 2014), highlighting challenges related to dialogue turn-taking patterns. A human commander (C) engages in speech dialog with a human robot navigator (RN) who tele-operates a robot (R) in an outdoor compound consisting of contiguous courtyards and buildings (C, R, and RN are remote from one another). RN is able to send to C a still image captured by R's video camera, or the most recent copy of R's laser-scan derived world map, with the push of a button. C and RN collaborate to locate and photograph as many building doorways as possible, and document their status (open or closed). The video shows, side by side, RN's (right) and C's (left) computer interfaces. RN's interface shows the world map as it is built live, R's pose (avatar in map with blue arrow), and the most recent still image sent to C. C's interface shows the most recent map and image sent by RN. This data collection is part of a larger effort toward iteratively automating motoric, linguistic, and perceptual capabilities performed by RN. We detail our current efforts toward a dialog management framework for this task in a position paper (Cassidy, Voss, and Summers-Stay 2015).

Video Description The ability to engage in human-like turn-taking behavior is a prerequisite for a useful, mobile, and communicative robot team member. Consider the task of finding and photographing building doorways in areas of interest, a plausible sub-task of a search and rescue mission: simply identifying *all* building doorways with no time constraints could be accomplished using a non-interactive drone tasked with generating a map of a new area (which could plausibly serve as input to a doorway detection algorithm). However, in a dynamic emergency scenario we need robots to relay information about their environment in a timely, natural way so that we can adjust their tasks accordingly (e.g., depending on the status of a given door). In this way the need for human-like turn-taking necessitates the following capabilities in a search and rescue robot:

1. Correctly timing the reporting of perceptual phenomena in accordance with dialogue context.

2. Quickly processing perceptual data to identify of a variety of perceptual phenomena with accurate confidence values.

3. Linguistic ability to generate important information about events in a natural way.

4. The ability to choose the correct modality for conveying information (speech, image, map)

5. A DMS that is aware of the relative difficulty of classifying various perceptual phenomena.

6. The ability to keep track of multiple objects of the same type while moving throughout a large scale environment.

For example, (1) is used to gauge when it is appropriate to report a possible new door sighting (i.e., not while currently pursuing a door sighting, but only after). The video shows two cycles in which C and RN look for a potential door sighting, find one, and home in on the sighting to photograph and describe it.

References

Cassidy, T.; Voss, C.; and Summers-Stay, D. 2015. Turn-taking and coordination in commander-robot navigator dialog. In *2015 AAAI Spring Symposium Series*.

Summers-Stay, D.; Cassidy, T.; and Voss, C. 2014. Joint Navigation in Commander/Robot Teams: Dialog & Task Performance When Vision is Bandwidth-Limited. In *COLING 2014 Proceedings of the Third Workshop on Vision and Language (VL'14)*.