Faster Pace in Human-Robot Dialogue Leads to Fewer Dialogue Overlaps

Cassidy Henry[1], Carla Gordon[2], David Traum[2], Stephanie M. Lukin[1], Kimberly A. Pollard[1], Ron Artstein[2], Claire Bonial[1], Clare R. Voss[1], Ashley Foots[1], and Matthew Marge[1]

[1]U.S. Army Research Laboratory, Adelphi, MD 20783
[2]USC Institute for Creative Technologies, Playa Vista, CA 90094
cassidy.r.henry.ctr@mail.mil

Abstract

In this paper, dialogue overlap at the transaction unit structure level is examined. In particular we investigate a corpus of multi-floor dialogue in a human-robot navigation domain. Two conditions are contrasted: a human wizard typing with a keyboard vs using a constricted GUI. The GUI condition leads to more utterances and transaction units, but leads to less overlap at the transaction unit level.

1 Introduction

Although dialogue is often thought of as a sequential process, with orderly transition of speaker turns and discussion topics, there are multiple overlap types that sometimes occur. These are instances in which a new unit is started before the previous one is complete. Perhaps most obvious is overlap between utterances, when different speakers speak at the same time. However, there is also dialogue structure overlap where one new structure starts before the previous one is completed. Here, we examine overlap between different *transaction units* (Sinclair and Coulthard, 1975; Carletta et al., 1996; Traum et al., 2018), abbreviated as "TUs."

A single TU contains all utterances relevant to executing a single task or intention. This often includes an initial instruction, statement, or question, followed by other utterances to extend, clarify, or give feedback. An example of TU overlap is a new instruction being issued before all feedback is received about the previous instruction. Some overlaps are minor and should not cause deterioration in dialogue, but others may cause confusion, e.g., in the above example, feedback occurring after a second instruction might refer to either the first or second instruction.

This paper explores dialogue overlap in a corpus of Wizard-of-Oz human-robot interactions collected under two conditions: typing responses using a keyboard vs. selecting utterances and templates with a GUI. The GUI condition leads to significantly less overlap at the TU level structure. Decrease is noted primarily in long overlaps.

2 Related Work

Previous exploration of overlap in dialogue has mostly been at the turn-taking level e.g., (Schegloff, 2000; De Ruiter et al., 2006; Heldner and Edlund, 2010). Some relevant work has explored higher level patterns in dialogues, such as those involving conversational thread disentanglement in Internet relay chats (Elsner and Charniak, 2010) and extracting collaborative patterns in tutorial dialogue (D'Mello et al., 2010). In these cases, there may be overlap in the topic sequences even if no one is sending text at the same time (though, there might also be both turn and topic overlap). We focus specifically on multi-floor dialogue (Traum et al., 2018), in which one participant is multi-communicating (Reinsch et al., 2008), and not all participants have access to all communications by other participants. In particular, we focus on physically situated human-robot collaborative dialogue involving multi-communicating parties that transmit information across multiple conversational floors.

3 Domain and Data

To investigate overlap, we use an annotated corpus of human-robot dialogues from a Wizard-of-Oz (WoZ) study of a collaborative search-and-navigation task between a human commander and a mobile robot that can navigate within a remote physical space (Marge et al., 2016) (see Figure 2, page 2). Our human participant, the Commander (or CMD), directs a remotely-located robot using speech. In our initial phases, two Wizards

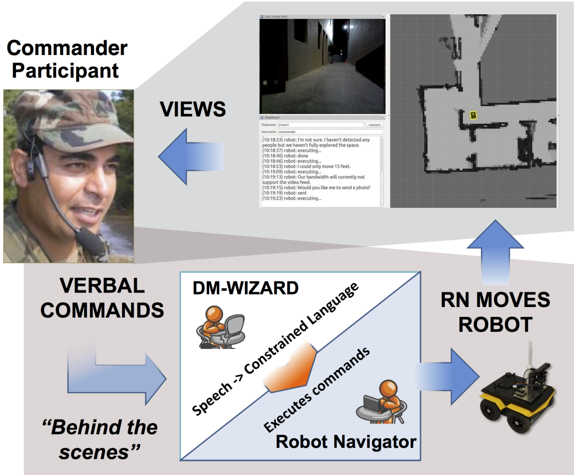

Figure 1: The Wizard of Oz experimental setup, where Wizards are responsible for emulation of the "robot's" capabilities of navigation and dialogue.

control robot intelligence: one for processing dialogue (called Dialogue Manager, or DM) and one for teleoperation (called Robot Navigator, or RN). The RN also provides photos upon request from the robot's camera and the Commander has access to a continuously-updating 2D map from the robot's LIDAR.

In multiple phases (or Experiments), automation progressively increases (Henry et al., 2017). In Experiment 1, the DM freely typed responses, following a dialogue management protocol (Marge et al., 2016). In Experiment 2, the DM used a GUI with buttons to facilitate faster response times. The buttons generated responses constructed from Experiment 1 DM text responses, with variable input for numeric values (e.g. ___ feet) (Bonial et al., 2017). In addition, floor-holding strategies were implemented, including statements like "processing" and "moving" to acknowledge that an action was in progress.

For both experiments, the Commander and Robot Navigator speech was transcribed and then aligned with DM text messages. Four separate communication channels were tracked, consisting of two floors (with the DM participating in both as a mediator of information exchange). The DM exchanges information between the CMD and RN by conveying commands in simplified language to the RN and by communicating status updates back to the CMD. This creates more opportunity for overlap in communication, given that RN and CMD do not have access to all of the dialogue informa-

Commander (CMD)	DM-CMD	DM-RN	RN	TU
could i get a picture				1
		send image		1
			sent	1
	sent			1
move forward to the second doorway				2
	processing...			2
		move to alley Door 2		2
	moving...			2
			alright done	2
	done			2

Table 1: Example with two TUs that do not overlap.

Commander	DM-CMD	DM-RN	RN	TU
turn ninety degrees left				1
	ok			1
		turn left 90 degrees		1
	I will turn left 90 degrees			1
	turning...			1
			done	1
take a picture				2
	done			1
		send image		2
			sent	2
	sent			2

Table 2: Simple case with one utterance interrupted.

Commander	DM-CMD	DM-RN	RN	TU	
go into the room in front of you				1	
and face south				1	
		move into room in front of you, face south		1	
	executing...			1	
take a picture				2	•
			done	1	
	done			1	
		image		2	•
face the doorway in front of you and to the right				3	
			image sent	2	•
and then take a picture				3	
	sent			2	•

Table 3: An example of a complex overlap scenario. Areas where overlap is occurring, and can be confusing are marked with "•".

tion that the DM does. The transcripts were annotated for meso-level dialogue structure (Traum et al., 2018). The annotations include transaction units (TUs), relation types based on how a new utterance is connected to its antecedent within the same TU. Table 1 shows a simple example fragment with the four dialogue streams. CMD and DM-CMD constitute one floor between Commander and "robot" (DM), while the DM-RN and RN streams constitute the second floor for "robot internal" dialogue. The TU column shows which unit each utterance is part of. In this case there

	Experiment 1		Experiment 2		p
Total TUs	919		1202		
All Overlaps	259		234		*.010*
Single Utterance Overlap	164	*(63.3%)*	166	*(70.9%)*	*.004*
Multi-Utterance Overlap	95	*(36.7%)*	68	*(29.1%)*	*.004*

Table 4: Number and size of overlaps between TUs.

is no overlap, as TU 2 starts after TU 1 has been completed. Table 2 shows a minimal amount of dialogue structure overlap, where the first utterance of TU 2 occurs before the final utterance of TU 1. Table 3 shows a more profound amount of overlap, where two TUs are overlapping TU 2, causing multi-utterance overlap. The ambiguity arises when the second image request is made in TU 3, and shortly thereafter confirmation comes from TU 2 that the image was sent, which could potentially be unclear as to which command was carried out.

4 Dialogue Overlap Analysis

The corpus contains data from 20 users (∼20 hours of audio; 3,573 utterances; 18,336 words), plus 13,550 words from DM to user in text messages, 9,643 words from DM to RN, and 3,485 words from RN to DM.

In total, the data contained 2,121 TUs, with 493 total observed TU overlaps. The majority of overlaps (330 or ∼67% of total overlaps) had only one utterance overlapped, like Table 2. Here, the phenomenon is relatively unproblematic and does not suggest communication or action completion difficulties. The remaining ∼33% of cases are more complex, such as that in Table 3.

Table 4 summarizes the breakdown between Experiment 1 and Experiment 2. There is an increased pace of dialogue from Experiment 1 to Experiment 2, with 1202 TUs in Experiment 2, compared to just 919 in Experiment 1. However, even with this increased pace, there are notably fewer overlaps in Experiment 2. Experiment 2 has 234 instances, reducing by 25 from Experiment 1. In particular, Experiment 2 has fewer overlaps involving multi-utterance interruptions, 27 instances fewer. Most cases in Experiment 2 involved just one utterance remaining after interruption (∼71%).

Statistics were conducted using independent samples t-tests to compare overlaps in the free-response (Experiment 1) vs GUI conditions (Experiment 2). Significant differences were observed in all overlaps between Experiment 1 trials (M=0.27, SD=0.16) and Experiment 2 trials (M=0.17 SD=0.12); t(58)=2.667, p=.010). There also was significant difference between Experiment 1 and Experiment 2 in terms of proportion of single utterance overlaps (Experiment 1 trials (M=0.48, SD=0.28), Experiment 2 trials (M=0.70, SD=0.29); t(58)=-3.006, p=0.004).

5 Conclusion

We presented an initial investigation of dialogue overlap in multi-floor multi-wizard human-robot communication. The results show that the faster-paced dialogues of our second experiment, where the responses were facilitated by GUI, reduced transaction unit overlap. Understanding these overlaps can also help shape future automated system behaviors to include protocols for avoiding overlap or resolving potential ambiguities caused by overlap.

References

Claire Bonial, Matthew Marge, Ashley Foots, Felix Gervits, Cory J Hayes, Cassidy Henry, Susan G Hill, Anton Leuski, Stephanie M Lukin, Pooja Moolchandani, Kimberly A. Pollard, David Traum, and Clare R. Voss. 2017. Laying down the yellow brick road: Development of a wizard-of-oz interface for collecting human-robot dialogue. *In Proc. of AAAI Fall Symposium Series*.

J. Carletta, A. Isard, S. Isard, J. Kowtko, G. Doherty-Sneddon, and Anne Anderson. 1996. HCRC dialogue structure coding manual. Technical Report 82, HCRC.

Jan-Peter De Ruiter, Holger Mitterer, and Nick J Enfield. 2006. Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language*, 82(3):515–535.

Sidney D'Mello, Andrew Olney, and Natalie Person. 2010. Mining collaborative patterns in tutorial dialogues. *Journal of Educational Data Mining*, 2(1):1–37.

Micha Elsner and Eugene Charniak. 2010. Disentangling chat. *Computational Linguistics*, 36(3):389–409.

Mattias Heldner and Jens Edlund. 2010. Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, 38(4):555–568.

Cassidy Henry, Pooja Moolchandani, Kimberly A. Pollard, Claire Bonial, Ashley Foots, Ron Artstein, Cory Hayes, Claire R. Voss, David Traum, and Matthew Marge. 2017. Towards Efficient Human-Robot Dialogue Collection: Moving Fido into the VirtualWorld. In *Proceedings of the WiNLP workshop*, Vancouver, Canada.

Matthew Marge, Claire Bonial, Kimberly A Pollard, Ron Artstein, Brendan Byrne, Susan G Hill, Clare Voss, and David Traum. 2016. Assessing agreement in human-robot dialogue strategies: A tale of two wizards. In *International Conference on Intelligent Virtual Agents*, pages 484–488. Springer.

N Lamar Reinsch, Jeanine Warisse Turner, and Catherine H Tinsley. 2008. Multicommunicating: A practice whose time has come? *Academy of Management Review*, 33(2):391–403.

Emanuel A Schegloff. 2000. Overlapping talk and the organization of turn-taking for conversation. *Language in society*, 29(1):1–63.

J. M. Sinclair and R. M. Coulthard. 1975. *Towards an analysis of Discourse: The English used by teachers and pupils.* Oxford University Press.

David Traum, Cassidy Henry, Stephanie Lukin, Ron Artstein, Felix Gervits, Kimberly Pollard, Claire Bonial, Su Lei, Clare Voss, Matthew Marge, Cory Hayes, and Susan Hill. 2018. Dialogue Structure Annotation for Multi-Floor Interaction. In *LREC*.