Towards Engagement Measures Using Dialogue Activities in Social Interaction

Guillaume Dubuisson Duplessis[1] Laurence Devillers[1,2]

I. Introduction

This paper presents an activity model for Human-Robot social dialogue exemplified by an entertaining interaction involving the Nao robot. We describe how this model could contribute to measuring engagement in H-R social dialogue.

II. Dialogue Activity Model

We envision H-R social dialogue as a set of joint activities that are activated and completed by dialogue participants. These activities can be viewed as *joint projects* [1]. A joint project is a bounded joint activity which can be broken down into an entry, a body, and an exit. Entry in a joint project is proposed by one participant and can be accepted or refused by the partner. Participants contribute to this activity through expected *participatory actions*.

In our approach, a dialogue activity is defined by a *type* (e.g., the "riddle" activity), a *conversational topic* (e.g., a specific riddle), an *initiator* and a *partner*. The initiator may either be the robot or the human. Several activities may be opened in parallel. A dialogue activity specifies expectations from dialogue participants in terms of *moves*. In our system, robot moves involve speech, affect bursts such as laughter, movements and eye colour changes. Human moves can be defined in terms of paralinguistic cues (e.g., expressed emotion in speech), linguistic cues (e.g., specific lexical entities) and extra-linguistic cues (e.g., a visual smile). Figure 1 presents the model of dialogue activity that we propose. Entry in the activity can either be *explicit* or *implicit*. An explicit entry consists in a proposition to enter the activity made by the initiator that can be accepted or rejected by the partner (cf. turns 4 and 5 in table I). Implicit entry consists in the initiator playing the first expected participatory action of the activity, thus making an implicit bid. Then, the partner can accommodate the activity by realising an *uptake* or a *rejection* [2]. Depending on the strategy adopted by the initiator, the establishment of the activity can lead to: (i) an explicit success or failure, or (ii) an implicit success or failure. Once the activity is established, our model considers two possible cases: a *dialogic success* or a *dialogic failure* (e.g., see turns 2 and 9 in table I). A dialogic success is reached when the activity is completed according to the expected participatory actions. In other words, a standard progress in the dialogue activity leads to a dialogic success. On the contrary, a *dialogic failure* happens when the activity takes an unexpected turn (e.g., non-fulfilment of an expected move, occurrence of an unexpected move, abandonment of the activity). Additionally, activities that reach a dialogic success can be assigned an *extra-dialogic status* referring to the conventional completion of the joint activity in terms of success or failure. For example, the activity of telling a riddle is a success if the partner discovers the right answer while it is a failure in the other case.

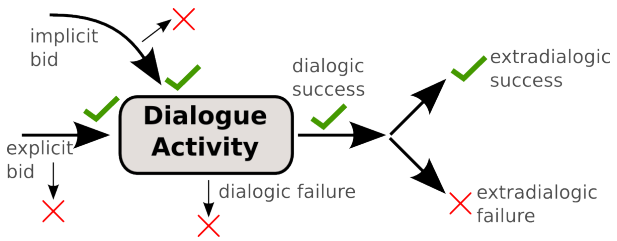

Fig. 1: Dialogue activity model for social dialogue in H-R interaction

III. Dialogue Activities and Engagement

Engagement can be defined as "the process by which two (or more) participants establish, maintain and end their perceived connection during interactions they jointly undertake" [3]. Engagement process involves nonverbal and verbal behaviours, as well as low-level processes (such as behaviour synchrony, mimetics) and high-level cognitive processes (such as answering a riddle). Recent work in H-R interaction aims at recognising and quantifying human engagement in dialogue in order to adapt the behaviour of the robot.

Our approach globally views the dialogue as the completion of dialogue activities. Our model discerns *fine levels of completion of an activity* which can turn out to be a success or a failure, namely: (i) the establishment of the activity (via an implicit or an explicit mechanism), (ii) the progress of the activity at the dialogic level, and (iii) the outcome of the activity at the extradialogic level. Accounting for these detailed levels of completion of activities in engagement measures can help to compute a high-level measure of engagement at the level of the social dialogue.

IV. Example in Entertaining H-R Dialogue

In our project, we aim to build a generic intelligent user interface which provides a multimodal dialogue system with social communication skills including humour and other social skills. We have developed a first prototype of the system that implements a system-directed dialogue. Implemented activity

* This work is part of the JOKER project (http://www.chistera.eu/projects/joker)

[1]LIMSI-CNRS, Paris, France, [2]Université Paris-Sorbonne 4, e-mail: {gdubuisson,devil}@limsi.fr

	Activity		Contribution				
	Name	**Expectation**	**Loc.**	**Transcription**	**Audio**	**Video**	**Activity Status**
1	Greetings		Nao	Hi, I'm Nao. I like to joke and I know riddles.			
2		Greetings	H		*silence*	*continued attention*	Dialogic failure
3	Dialogic failure recovery		Nao	Well, you do not want to tell me anything,			
4	Riddle			but let me tell you a riddle.			Explicit bid
5		Acceptance sign	H		*silence*	*smile*	
6			Nao	Who wrote the article "J'accuse?"			
7		Answer	H	Ah... I do not know.		*head movement*	
8			Nao	The answer was Émile Zola. (laugh)	laughter		
9		Positive reaction	H	*(laugh)*		*laughter*	Dialogic success Extra-dialogic failure
10	Riddle		Nao	I love riddles, let me tell you another one!			Explicit bid
				(...)			

TABLE I: Excerpt of a dialogue from our corpus of entertaining interactions between a human and our first automatic prototype of the system (translated from French to English). H=Human, Nao is the robot.

types include: greetings, goodbye, telling a riddle, telling a funny short story, telling a positive or negative comment about the human participant, telling punctual idiomatic expressions. This automatic system has been involved in a data collection with 37 human volunteers.

An excerpt of the collected corpus is presented in table I. It shows the occurrence of various activity states during dialogue, and illustrates the chaining of activity. Dialogue starts with a greetings activity which is implicitly established by the robot. The human participant is expected to return the greetings (or to explicitly reject the activity). However, he does nothing except showing continued attention. The activity is a dialogic failure, notified by the robot in turn 3. Then, the system explicitly introduces a riddle activity, which is accepted by a positive sign from the human (a smile). This activity progresses as expected, and terminates on a dialogic success and an extra-dialogic failure (the answer to the riddle has not be found).

The determination of the completion levels can take advantage of activity-specific knowledge. Indeed, our model provides a local interpretation context of human contribution that could be fruitfully used to fuse verbal and non-verbal channels for social behaviour perception and interaction capabilities as well as for assessing engagement. Indeed, it makes it possible to specify expectations from the human participant *relatively to the status of the activated dialogue activity*. Specifications of human moves and engagement measures can take advantage of various interactional, emotional and spoken paralinguistic cues in human audio activity. In our project, we are currently studying cues such as: audio duration activity (speech and non-speech), speech duration, duration of speech involving a positive/negative/neutral emotion, speech reaction time (duration between the end of the speech produced by the robot and the start of speech produced by the human), and presence of affect bursts (e.g., laughter, onomatopoeia). Our current goal is to assess engagement by taking into account both the levels of completion of dialogue activities and the mood expressed by the human participant.

V. Related Work

A similar dialogue activity-based approach has been presented in the context of social interaction with an embodied conversational agent [2], [4]. Our model extends what was proposed in that: (i) it provides a refined model of the activity status that goes beyond the implicit entry by discerning implicit/explicit entry, dialogic success/failure and extradialogic success/failure, and (ii) it aims at taking into account paralinguistic, linguistic and extra-linguistic cues.

In the context of collaborative task-oriented interaction between a human and a robot, Rich & Sidner [3] have identified four types of connection event (directed gaze, mutual facial gaze, delay in adjacency pair, and backchannel) involved in the computation of statistics on the overall engagement process. Our focus is on social dialogue rather task-oriented one. We consider paralinguistic, linguistic and extra-linguistic cues that include the four types of connection event.

VI. Conclusion and Future Work

We have presented a model of dialogue activity in the context of H-R social dialogue that discerns various success and failure states relatively to the entry, the body and the exit of the activity. This model has been illustrated in the context of an entertaining interaction between a robot and a human. We have emphasised how this model could be advantageously used to contribute to the computation of engagement measures in dialogue. This model seems promising to manage multimodal social dialogue between a human and a machine. A dialogue planner can take into account the possible outcomes of a dialogue activity in order to fruitfully combine them. Next, this model provides a rich interaction history footprint as a sequence of past dialogue activities along with their outcome. This could be usefully exploited to enrich a representation of the H-R relationship [2], [5].

References

[1] H. Clark, *Using language*. Cambridge University Press, 1996, vol. 4.

[2] T. Bickmore and D. Schulman, "Empirical validation of an accommodation theory-based model of user-agent relationship," in *Intelligent Virtual Agents*. Springer, 2012, pp. 390–403.

[3] C. Rich and C. L. Sidner, "Collaborative Discourse, Engagement and Always-On Relational Agents." in *AAAI Fall Symposium: Dialog with Robots*, 2010.

[4] J. Cassell and T. Bickmore, "Negotiated collusion: Modeling social language and its relationship effects in intelligent agents," *User Modeling and User-Adapted Interaction*, vol. 13, no. 1-2, pp. 89–132, 2003.

[5] A. Delaborde and L. Devillers, "Use of nonverbal speech cues in social interaction between human and robot: Emotional and interactional markers," in *Proceedings of the 3rd International Workshop on Affective Interaction in Natural Environments*. ACM, 2010, pp. 75–80.