Towards the Consideration of Dialogue Activities in Engagement Measures for Human-Robot Social Interaction

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

***Abstract*— This paper addresses the problem of evaluating Human-Robot spoken interactions in a social context by considering the engagement of the human participant. We present an activity model for Human-Robot social dialogue, and show that it offers a convenient local interpretation context to assess the human participation in the activity. We describe an engagement score based on this model, and discuss results obtained by its application on a corpus of entertaining interactions involving the Nao robot in a cafeteria. We show how this approach makes it possible to discern three groups of participants in terms of engagement.**

***Index Terms*— Human-Robot Interaction, Social Dialogue, Engagement**

I. Introduction

Our research effort aims at evaluating Human-Robot spoken interactions and Human-Robot relationships in a social context. In this direction, recent work in Human-Robot interaction (HRI) intends to recognise and quantify human engagement in dialogue in order to adapt the behaviour of the robot. 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" [1], [2]. 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).

From our perspective, evaluation of HRI should be multifaceted. To that purpose, we propose to evaluate human engagement by combining verbal and nonverbal behaviour along with contextual information. In this paper, we describe a dialogue activity model for social HRI that we are currently investigating, and we report our first results in trying to quantify human engagement based on this model. We show promising results regarding the ability of our model to provide a local interpretation context of human behaviour in order to infer his engagement. Our research effort is part of the JOKER project which aims to build a generic intelligent user interface providing a multimodal dialogue system with social communication skills including humour and other social behaviours [3]. Results presented in this paper are exemplified on a corpus collected in a cafeteria setting as part of the JOKER project.

Section II provides links and discussion of related work. Section III describes our proposition of dialogue activity model for H-R social dialogue. Section IV presents a corpus collection of entertaining H-R interactions performed with an automatic data collection system. We describe our system which features an audio-based paralinguistic detection module, a dialogue manager based on the activity model and a communicative behaviour synthesis via the Nao robot. Section V is interested in the contribution of the dialogue activity model to measure engagement in H-R social dialogue. It presents a proposition of engagement score to assess the participation of the human in terms of audio activity in the dialogue activities lead by the robot. This approach is implemented on the collected corpus, and first results are presented and discussed. Section VI concludes this paper and presents perspectives.

* 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

II. Related Work

A similar dialogue activity-based approach to the one that we describe in section III has been presented in the context of social interaction with an embodied conversational agent [8], [5]. 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 bid, 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 [2] 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 than task-oriented one. We consider the interpretation of cues in the context of a dialogue activity. Interestingly, cues such as the four types of connection event could be integrated to our approach.

III. 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* [4]. 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. We view dialogue as emerging via dialogue

activity combinations (e.g., sequence, embedding, parallel execution). 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 may 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 II). 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* [5]. Depending on the strategy adopted by the initiator, the establishment of the activity can lead to: (i) an explicit success (acceptance of an explicit bid) or failure (rejection of an explicit bid), or (ii) an implicit success (uptake of an implicit bid) or failure (rejection of an implicit bid). 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 II). 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 *extradialogic 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.

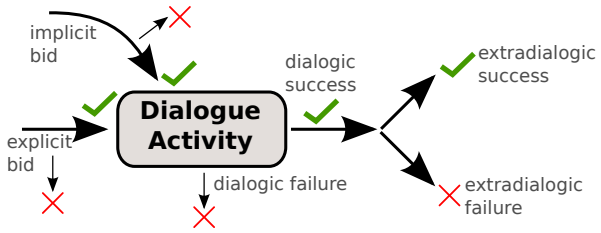


Fig. 1: Dialogue activity model for social dialogue in HRI

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. The first level is the *establishment of the activity* via an implicit or an explicit mechanism. It captures a part of the effort of the participants to co-construct and co-control the dialogue, seen as an opportunistic joint activity [4]. Failure and success at this level have been interpreted as an evidence of the intimacy level in Human-Agent social relationship [5] but could also be considered at the level of engagement. The second level is the *progress of the activity* at the dialogic level. It is chiefly related to the participation of interlocutors in the dialogue activity. In other words, the dialogic status reflects whether the participants have performed their moves in order to get to the end of the activity or not. For instance, a human participant not taking a turn after the robot signals its end of turn in the conversation has been interpreted as a clear sign of disengagement [1]. While previous levels are dedicated to the dialogue activity in itself, the last level somewhat takes a "task-oriented perspective" by dealing with the outcome of the activity. Thus, the extradialogic status is the result of a dialogue activity that has been successfully carried out by dialogue participants at the dialogic level.

Conceptually, these three levels form a scale in which a success at a level allows to pass to the next one. Conversely, a failure at a given level prevents from moving on to the next.

Evaluation of the social interaction between a human and a robot via engagement measures can take advantage of this dialogue activity model notably by taking into account these detailed levels of completion. In this paper, we present a first step in this direction in which we mainly focus on the dialogic level, i.e. in the participation of the human in the dialogue activity proposed by our system.

IV. Implementation in an Entertaining H-R Dialogue

A. Automatic Data Collection System

The automatic data collection system used in this work is a first prototype exploring the paralinguistic aspect of the JOKER project. This system currently takes as input an audio signal (this will be extended with visual input in the next step of the project). It features a recognition module involving emotion detection, laughter detection and speaker identification from audio signal [6]. This module provides paralinguistic cues that are advantageously used by the dialogue management process. Notably, our system does not include an automatic speech recognition (ASR) module yet. This is due to the lack of open and free systems for French language. However, an adapted ASR for French is currently being developed in the JOKER project. This limitation is overcome by a careful design of dialogue activities (described in section IV-B) and a simplified specification of expected moves from human participant (see section V-A). The dialogue manager deals with system-directed social interaction dialogues, and is an initial prototype based on the dialogue activity model that takes into account paralinguistic cues. Eventually, the communicative behaviour of the system is synthesised through the Nao robot via speech, laughter, movement and eye colour variations.

B. Dialogue Activities, Interaction Scenario and Limitations

The scenario implements a system-directed entertaining interaction dialogue that includes the telling of riddles and other humorous contributions. Dialogue is in French. Examples presented in this paper have been translated from French to English.

Type	Description	w
Greetings	Greeting and presentation phase	2
Riddle	An activity where the robot asks a riddle, waits for an answer from the human, reveals the right answer and lets the human react	4
Teasing	Positive or negative comment about the human participant	2
Strength	A question about the strength of the human participant followed by a funny comment of the robot about itself	4
Memory	A question about the effectiveness of the memory of the human participant followed by a funny comment of the robot about itself	4
Goodbye	Closing phase of the dialogue	2

TABLE I: Dialogue activities involved in the entertaining HRI (see section IV-B). w is the total number of turns in the body phase of the dialogue activity for both dialogue participants.

This scenario implements the following dialogue activities (summarised in table I): greetings, telling a riddle, telling a positive or negative comment about the human participant (teasing), challenging the human about his strength or his memory, and goodbye. The "greetings" and "closing" dialogue activities are meant, respectively, to open the dialogue and to close the dialogue. They each consist of two turns: the first from Nao, the second from the human participant. The "riddle" activity allows Nao to tell different types of riddle (absurd, social, general knowledge) to the human, e.g., "How do you know there are two elephants in your fridge? – You can't close the door.". Riddles follow a common structure. First, the system asks a question forming the riddle (which are made so that the answer is not expected to be found). Then, the participant reacts to the riddle. Finally, the robot provides the right answer and lets the human react. In the "teasing" activity, Nao makes a positive comment (e.g., "You're doing really well! Congratulations!") or a negative one (e.g., "A child could answer that!") about the previous human contribution. This activity involves two turns (one from Nao and the other from the human). The "strength" and "memory" activities follow a common structure in which: (i) Nao asks a question to the human (e.g., "Have you a good memory?", "Are you very strong?"), (ii) lets the human answer, then (iii) makes a funny comment about itself (e.g., "I have a small head, but it does not matter.", "I am not very strong, look at my muscles!"), and (iv) lets the human react.

An interaction scenario consists of 8 dialogue activities that are chained one after the other. It starts with the "greetings" activity and stops with the "goodbye" one. A scenario necessarily includes a "riddle" activity, a "teasing" activity and either a "memory" or a "strength" activity.

An excerpt of the collected corpus is presented in table II. 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 extradialogic failure (the answer to the riddle has not be found).

The inability of our system to extract linguistic content from the speech contributions of the human participant introduces some limitations in our study. As a consequence, our system does not exploit the full richness of the dialogue activity model. Indeed, all activities except the "riddle" one are introduced by an implicit bid at the establishment level. On the contrary, the "riddle" activity is introduced via an explicit bid (e.g., see turns 4 and 10 in table II). However, a bid (either implicit or explicit) cannot be refused since it often requires linguistic understanding (e.g., "I don't want to hear one of your silly riddle!", see [5] for a study on this subject). Additionally, our system cannot truly evaluate the answer of the human participant in the "riddle" activity (and therefore, the extradialogic status of the activity). To overcome this limitation, riddles have been designed so that the answer is not expected to be found by the participant. As a result, Nao can reveal the answer in the third turn of the activity without sounding strange.

C. Collected Data

The experimentation took place in the cafeteria of the LIMSI laboratory with 37 French-speaking participants. Volunteers were 62% male, 38% female, and their ages ranged from 21 to 62 (median: 31.5; mean: 35.1). Participants were seated facing the Nao robot at around one meter from it. Audio data have been acquired thanks to a high-quality AKG Lavalier microphone. Audio tracks of 16kHz have been recorded internally by the system for each interaction between a volunteer and the robot. A total of 1h 30min 5sec of audio data has been collected (average session duration: 2min 26sec ; standard deviation: 14sec).

Despite the lack of linguistic understanding, participants have reported in questionnaire right after the interaction that they were satisfied and amused by the interaction with the robot, and that they felt the desire to talk to the robot [3].

V. Preliminary Results: Engagement Score based on Dialogue Activities

A. Expectations in Dialogue Activities and Dialogic Status

Specifications of human moves can take advantage of various interactional, emotional and spoken paralinguistic cues in human audio activity. In this first attempt, we restricted our expectations specification of human moves to straightforward cues available in the audio channel. Our expectation specifications distinguish three levels of audio activity:

1) Silence: the human participant stays silent during his turn.

	Activity Type	Contribution				
		Loc.	Transcription	Audio	Video	Activity Status
1	Greetings	Nao	Hi, I'm Nao. I like to joke and I know riddles.			
2		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		H		*silence*	*smile*	
6		Nao	Who wrote the article "J'accuse?"			
7		H	Ah... I do not know.		*head movement*	
8		Nao	The answer was Émile Zola. (laugh)	laughter		
9		H	*(laugh)*	*laughter*		Dialogic success Extra-dialogic failure
10	Riddle	Nao	I love riddles, let me tell you another one!			Explicit bid
			(...)			

TABLE II: 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.

Activity	t_2	t_4
Greetings	Speech	*NA*
Goodbye	Speech	*NA*
Riddle	Speech	Audio activity
Teasing	Audio activity	*NA*
Strength	Speech	Audio activity
Memory	Speech	Audio activity

TABLE III: Specification of expectations in dialogue activities. NA=Not Applicable, audio activity = speech or non-speech audio contribution. t_1 and t_3 are Nao turns (respectively, first and third turns in the activity).

Activity Type	Dialogic failure	Dialogic success
Greetings	18.4%	81.6%
Riddle	19.7%	80.3%
Teasing	5.3%	94.7%
Strength	74.1%	25.9%
Memory	27.3%	72.7%
Goodbye	50.0%	50.0%

TABLE IV: Distribution of dialogic success/failure in the collected corpus (see section V-A)

2) Non-speech audio activity: the human participant contribution does not contain speech but other human sounds (e.g., laughter, breath, sigh).
3) Speech activity: the human participant produces a contribution containing speech.

This coarse-grained level of expectations follows the limited capabilities of our paralinguistic dialogue system, and can also be explained by the relatively constrained system-directed interaction scenario that we use in this work.

Expectations specification per dialogue activities are presented for each human turns in table III. In this study, expectations in dialogue activities are based on the following assumptions: (i) An explicit solicitation from Nao via a question or a riddle should trigger a speech contribution from the human, namely the second pair part of the adjacency pair [7] (this is the case after Nao first turns in the "riddle", "strength" and "memory" activities). (ii) "Greetings" and "goodbye" uttered by Nao should be conventionally returned by the human participant. (iii) Punctual funny/teasing contributions should trigger a reaction from the human participant, e.g., laughter (this is the case after the first Nao turn in the "teasing" activity, and after the third turns in the "strength" and "memory" activities). Lastly, human participants are expected to react either with speech or with other human sounds to the revealing of the right answer of the riddle (e.g., "I knew it!", laughter).

These expectations have been used to compute the dialogic status of dialogue activities occurring in the corpus, presented in table IV. Expectations have been met in more than 70% of the cases for the "greetings", "riddle", "teasing" and "memory" activities. On the contrary, the "strength" activity has mainly been a dialogic failure. One explanation can be that human participants preferred to react extra-linguistically (e.g., via a smile or a pout) rather than to respond to a challenging question about their strength. As such, this activity did not succeed to engage many dialogue participants. Participation to the "goodbye" activity is mixed as well. Half of the participants did not participate linguistically in this activity. We see two main reasons. First, participants may have preferred a goodbye gesture rather than saying something. Then, participants may have chosen not to react knowing that this was the end of the interaction and that they could not get a reaction from Nao.

B. Engagement Score

In this paper, we focus on the dialogic status of dialogue activities and how it could contribute to measure engagement from the human participant in the social interaction. We propose to use as an engagement measure the proportion of dialogue activities that have been a dialogic success in a given interaction, weighted by the length of the body of the activity (specified in table I). For each participant p, we computed an *engagement score* corresponding to the proportion of dialogue activities (da_i^p, $1 \leq i \leq n$) that have been a dialogic success over all the dialogue activities that occurred in this specific interaction, weighted by the length of the activity (see equation 1).

$$\text{score}_p = \frac{\sum_{i=1}^{n} \text{success}(\text{da}_i^p) \times w_i}{\sum_{i=1}^{n} w_i} \quad (1)$$

where:

$$\text{success}(\text{da}_i^p) = \begin{cases} 1 & \text{if } \text{da}_i^p \text{ is a dialogic success} \\ 0 & \text{if } \text{da}_i^p \text{ is a dialogic failure} \end{cases}$$

The score quantifies the participation of the human in the dialogue activities initiated by Nao during the interaction. The lower the score, the less the human participated in the dialogue activities compared to what was expected, the less engaged he was. On the contrary, the higher the score, the more the human participated in the dialogue activities, the more engaged he was. Interestingly, this score is not a task completion score but a dialogue activity completion score. As such, it is concerned with the activity participation rather than the activity outcome. Nevertheless, a task completion score based on the dialogue activity model could take advantage of the extradialogic status.

C. Results

The distribution of engagement scores that we obtained is presented on figures 2. Scores range from 0.11 to 1.0 (mean=median=0.67, standard deviation=0.22). The distribution of scores shows a wide variety of cases, from human participants that have mainly reached dialogic failures to ones that have reached a dialogic success for all the dialogue activities. We pursued our analysis by trying to determine if the computed scores make it possible to detect groups of participants. To that purpose, we performed a cluster analysis based on the k-means clustering method. This method aims to partition a set of points into k groups such that each point belongs to the group with the nearest mean. The number of clusters $k = 3$ was determined by looking for a bend in the sum of squared error as a function of the number of cluster. The clustering method was automatically performed using the algorithm of Hartigan and Wong via the R software (version 3.0.2, see `https://www.r-project.org/`). Computed clusters are presented on figure 2a. Three groups have been automatically detected that could be informally qualified as the "little engaged" group (red square), the "engaged" group (green triangle) and the "very engaged" group (blue circle). The first group contains 8 participants (21.5% of the total) with the lowest engagement scores ranging from 0.11 to 0.44. The second group includes 21 participants (57% of the total) with engagement scores ranging from 0.55 to 0.78. The third group includes 8 participants (21.5% of the total) with the highest scores ranging from 0.88 to the maximum 1.0.

Since our expectations specification of a dialogic success are dependent on the presence or absence of speech from the human, we investigated the impact of speech duration on the engagement score. Not surprisingly, a study of the linear correlation between the engagement score and the accumulated speech duration from the human participant via the Pearson's correlation coefficient reveals the existence of a moderate but significant correlation ($\rho = 0.68$, p-value $= 3.27 \times 10^{-6}$). Next, we adopted a similar clustering procedure but we replaced the engagement score by the accumulated speech duration of human participant during the interaction with Nao. The number of clusters k turned out to be 3 as well. Computed clusters are presented on figure 2b. Clustering based on speech duration reveals 3 groups. The first one contains 12 participants (32% of the total) with speech durations ranging from 2110ms to 12430ms. The second one includes 21 participants (57% of the total) with speech durations ranging from 14270ms to 26610ms. The third group contains 4 participants (11% of the total) with speech durations ranging from 29460ms to 39960ms.

A comparison between the two clustering results shows that the engagement score computed by using dialogue activities cannot be reduced to a score solely based on speech duration. Indeed, we can observe that some participants belonging to the less talkative groups can get high scores (see the two red squares around 0.8, and the green triangle at 1.0 on figure 2b). Conversely, we can observe a participant belonging to the more talkative group who does not obtain an engagement score above 0.8.

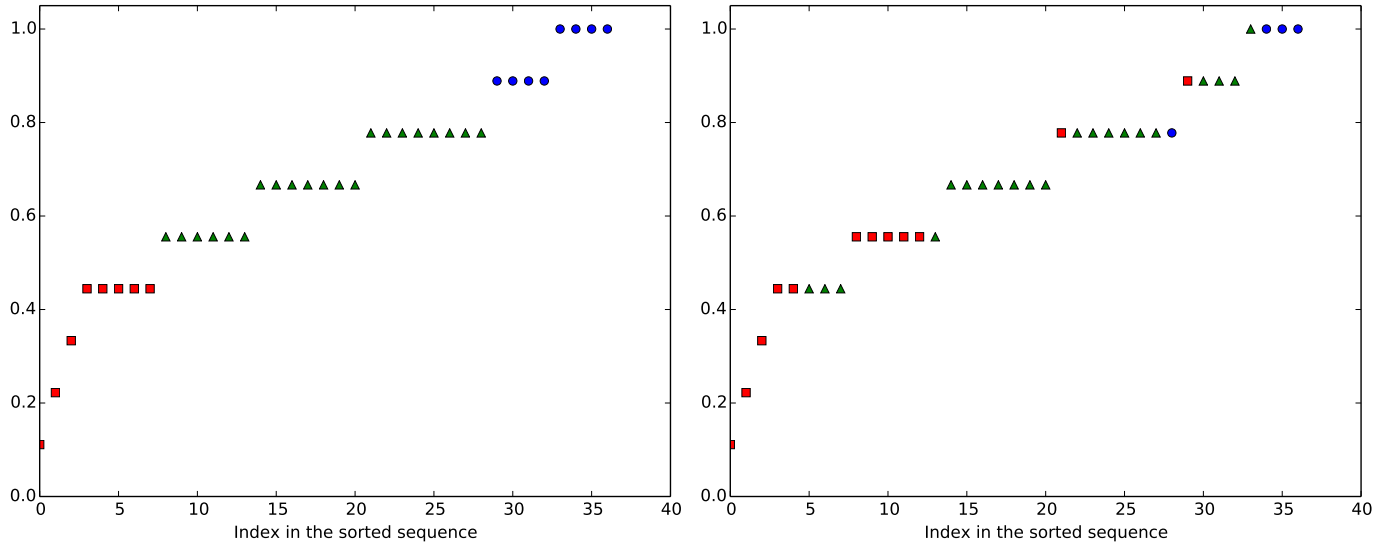
D. Discussion

We have presented an engagement score that takes advantage of the dialogic status of dialogue activities initiated by Nao and involving a human partner. We have computed this engagement score for the 37 participants that have been interacting with Nao on a social entertaining dialogue. A cluster analysis using the common k-means method has made it possible to automatically discern 3 groups of participants in terms of engagement: the "little engaged" (21.5% of all participants), the "engaged" (57%) and the "very engaged" (21.5%) groups. Despite the fact that the dialogic status of an activity is mainly dependent on human speech production, a similar cluster analysis based only on human speech duration did not yield the same results. In other words, human participants speaking the most were not necessarily the ones with the highest engagement scores (and conversely). From our perspective, this result highlights an important feature of the dialogue activity model which is that it offers a *local interpretation context* to infer the engagement of human participant. As a consequence, the engagement score based on this model proposed in this paper favours participants that coordinate their contributions with those of Nao in order to interact *at the right time*.

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 statuses relatively to the entry, the body and the exit of the activity. We have developed a first automatic prototype dialogue system based on this model, that implements a system-directed entertaining dialogue between the Nao robot and a human. This system involves a dialogue manager based on the extraction of paralinguistic information in the audio signal. A corpus of 37 interactions with human participants have been collected in which volunteers reported to be satisfied and amused by the interaction.

In this first step, we have shown how the proposed dialogue activity model provides a local interpretation context



(a) Clustering by engagement score

(b) Clustering by speech duration

Fig. 2: Engagement score distribution clustered into 3 groups (red square, green triangle, blue circle) with two different features.

that could be used to infer the human engagement in the activity, and on the overall interaction. In this paper, we have focused on the human participation in the dialogue activities initiated by Nao via their dialogic statuses. We have shown how they could be exploited to compute an engagement score for the human participants. These scores have made it possible to automatically discern three groups of participants in terms of engagement: the "little engaged" group, the "engaged" group and the "very engaged" group.

Future work includes many challenging perspectives. First, it consists in extending the perceptual capabilities of our system in order to more accurately assess engagement from human. To that purpose, we are currently exploring the specifications of human moves 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). We believe that our model could be fruitfully used to fuse verbal and non-verbal channels for social behaviour perception and interaction capabilities.

Next, the dialogic status refers to whether the interlocutors produced their moves as expected or not. This status may seem too radical, and we believe that it could fruitfully be extended with an additional indicator that would assess how well the produced moves have been performed. It would help, e.g., to differentiate an answer to a question that is delivered with a sigh or with a smile.

Then, the study presented in this paper focuses on the dialogic status of an activity, and therefore partially exploits the richness of the model. In particular, the establishment level has been left aside whereas accepting or refusing to enter an activity is an important clue to take into account for assessing engagement. Furthermore, one simplification of our study is the system-directed management of the dialogue. We are planning to waive this limitation to some extent in order to the let the possibility to the human to initiate his own activities.

Eventually, the dialogue activity 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. Moreover, 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 [9], [5].

REFERENCES

[1] C. L. Sidner, C. Lee, C. D. Kidd, N. Lesh, and C. Rich, "Explorations in engagement for humans and robots," *Artificial Intelligence*, vol. 166, no. 1, pp. 140–164, 2005.

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

[3] L. Devillers, S. Rosset, G. Dubuisson Duplessis, M. Sehili, L. Béchade, A. Delaborde, C. Gossart, V. Letard, F. Yang, Y. Yemez, B. Türker, M. Sezgin, K. El Haddad, S. Dupont, D. Luzzati, Y. Estève, E. Gilmartin, and N. Campbell, "Multimodal data collection of human-robot humorous interactions in the joker project," in *6th International Conference on Affective Computing and Intelligent Interaction (ACII)*, 2015.

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

[5] 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.

[6] L. Devillers, M. Tahon, M. A. Sehili, and A. Delaborde, "Inference of human beings' emotional states from speech in human–robot interactions," *International Journal of Social Robotics*, pp. 1–13, 2015.

[7] E. A. Schegloff and H. Sacks, "Opening up closings," *Semiotica*, vol. 8, no. 4, p. 289–327, 1973.

[8] 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.

[9] 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.