Hybrid Chat and Task Dialogue for More Engaging HRI Using Reinforcement Learning*

Ioannis Papaioannou[1] Christian Dondrup[1] Jekaterina Novikova[1] Oliver Lemon[1]

***Abstract*— Most of today's task-based spoken dialogue systems perform poorly if the user goal is not within the system's task domain. On the other hand, chatbots cannot perform tasks involving robot actions but are able to deal with unforeseen user input. To overcome the limitations of each of these separate approaches and be able to exploit their strengths, we present and evaluate a fully autonomous robotic system using a novel combination of task-based and chat-style dialogue in order to enhance the user experience with human-robot dialogue systems. We employ Reinforcement Learning (RL) to create a scalable and extensible approach to combining chat and task-based dialogue for multimodal systems. In an evaluation with real users, the combined system was rated as significantly more "pleasant" and better met the users' expectations in a hybrid task+chat condition, compared to the task-only condition, without suffering any significant loss in task completion.**

I. INTRODUCTION

Spoken dialogue systems, e.g. [1], [2], [3], are generally task-based and often fail to engage with users, concentrating instead on discovering a user's goal through multiple dialogue turns (such as booking a flight or finding a suitable restaurant). The same holds true for dialogue systems used in Human-Robot Interaction (HRI), e.g. [4], [5], [6]. On the other hand, chatbots (such as [7]) are focused on entertainment. They do not support execution of user tasks or transactions, because they have a limited memory and do not perform true language understanding, thus being unable to determine users' goals. Systems such as *Siri* and the *Amazon Echo* do combine some aspects of chat and task-based interaction, but generally only react to single user turns/commands, and do not support extended multi-turn dialogue to discover user goals. To create a more integrated approach to dialogue in HRI, the dialogue itself needs to support both entertaining chat and multi-step interaction for meeting user goals. To this end, the presented work focuses on enabling a robotic agent to combine a task-based dialogue system with chat-style interaction to fulfil the required tasks while at the same time being natural and engaging to interact with.

The problem domain addressed here is a shopping mall scenario where HRI should not only focus on fulfilling user tasks such as providing guidance to certain shops and giving vouchers, but also needs to be entertaining to create a natural and fun interaction for the visitors. Addressing this problem of carrying out a mixture of task-based and chat-style dialogues, we are combining the two technologies of task-based dialogue and chatbots resulting in a system that fulfils users' expectations and is (as we show) more pleasant to interact with. In contrast to previous approaches (e.g. [8]), we use Reinforcement Learning (RL) instead of hand-crafted rules to decide when to chat and when the execution of a task is required. This provides scalability and flexibility to the system making it applicable in more complex problems, as suggested in [9]. We then evaluated this first approach to applying such a hybrid dialogue system to HRI in a user-study with lay participants and a fully autonomous robotic system (see Fig. 1).

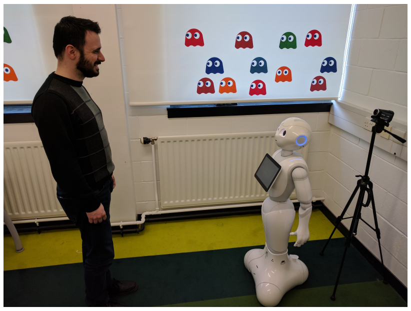

Fig. 1: User-study: a human interacting with Pepper at $\sim 1.5m$ distance. See Fig. 2 for experiment set-up.

II. RELATED WORK

There has been some limited work on combining chat- and task-based dialogue, though none of it has been in the context of HRI. In [8], a text-based hybrid system combines a Dialogue Manager with a Chatbot. The system implemented was rule-based and able to communicate only via text, and was evaluated by 3 different subjects of different educational background. The system showed promising results, being capable of holding long conversations, but only in *issue-based* dialogues (meaning that *"it views dialogue as the raising and resolving of questions"* [8]).

A hybrid system was also proposed in [10] merging a chatbot with a dialogue manager in a rule-based manner. Their proposed system would have access to both a local and external knowledge base, that along with the user dialogue input would be able to create a model of the user's intention and simultaneously keep track of the user's interactions with the world. This prototype system was evaluated using actual users, conversing with the system *"following a predefined*

*The research leading to the these results has received funding from the European Commission's H2020 programme under grant agreement No 688147, MuMMER. http://mummer-project.eu/

[1]Ioannis Papaioannou, Christian Dondrup, Jekaterina Novikova, and Oliver Lemon are with the School of Mathematical and Computer Sciences, Heriot-Watt University, Edinburgh EH14 4AS, Scotland, UK {i.papaioannou, c.dondrup, j.novikova, o.lemon}@hw.ac.uk

TABLE I: Example of state transitions using the hybrid (task+chat) system. The spoken utterance and which action was selected using the MDP policy (actionTaken) is shown on the right. Keywords in the utterance and the resulting state changes in the next turn are highlighted in bold.

Turn	Distance	Mode	usrEngChat	timeout	tskFilled	tskCompl	usrEng	sBye	usrTerm	turnTaking	prevAct	ctxTask	actionTaken	Utterance
(1)	1	0	0	0	0	0	1	0	0	0	0		Greet	SYS: “Hi there”
(2)	1	0	0	0	0	0	1	0	0	1	2		Wait	USR: “Hello. What is your name”
(3)	1	**1**	**1**	0	0	0	1	0	0	0	6		Chat	SYS: “My name is Pepper”
(4)	1	1	1	0	0	0	1	0	0	1	4		Wait	USR: “My name is John”
(5)	1	1	1	0	0	0	1	0	0	0	6		Chat	SYS: “Nice to meet you John”
(6)	1	1	1	0	0	0	1	0	0	1	4		Wait	USR: “Do you know any place to get a **cappuccino**?”
(7)	1	1	**0**	0	**1**	0	1	0	0	0	6	**coffee**	taskConsume	SYS: “There are 2 coffee shops nearby. These are Starbucks, and Costa”
(8)	1	**0**	0	0	0	**1**	1	0	0	1	1		Wait	USR: “How can I get to **Starbucks**?”
(9)	1	0	0	0	**1**	0	1	0	0	0	6	**directions**	giveDirections	SYS: “To get there you need to go left, then straight ahead”
(10)	1	0	0	0	0	**1**	1	0	0	1	5		Wait	**[User starts moving away from the agent]**
(11)	**2**	0	0	**1**	0	0	1	0	0	0	6		reqTask	SYS: “Is there anything I can help you with?”
(12)	2	0	0	0	0	0	1	0	0	1	8		Wait	USR: “No thank you. **Goodbye**!”
(13)	2	0	0	0	0	0	1	**1**	0	0	6		goodbye	SYS: “Have a nice day!”

context and set of example tasks” [10]. The system was successfully achieving the objectives for creating a conversational agent with an external knowledge base.

Both these systems, although somewhat successful in providing semantic representations combined with chatbot conversation, are lacking in terms of extensibility and maintainability, due to their rule-based development process. New rules must be written by expert developers for each new domain and task, and there is no guarantee of optimality for the dialogue management decisions defined in this way. By contrast, systems using Reinforcement Learning (RL) can be trained on data, or via interaction with users, and can learn optimal dialogue policies, e.g. [2], [3], [6]. Therefore it is of interest to determine whether combined chat/task multimodal dialogue systems for HRI can also be trained using RL methods.

Concurrently with our work, [11] also explore an RL method for combing chat and task-based dialogue, although in the context of a movie promotion system rather than for multimodal HRI. They found that the addition of chat behaviour enhanced user engagement.

Note that none of the above systems use **multimodal** information (e.g. from a robot's sensors) to enrich the dialogue, as is required for a holistic approach to HRI. To show the feasibility of including such information, our system uses the distance between the robot and the human during dialogue management as one of the features for adaptation, but RL also provides the capability to use a wider variety of sensory inputs, since RL dialogue management permits highly complex decision making over large state-spaces.

III. SYSTEM COMPONENTS

To create a more engaging intelligent agent for HRI, we developed a multimodal dialogue system which combines chat and task-based dialogue, based on work presented in [9]. In contrast to previous approaches, we are using Reinforcement Learning (RL) to switch between the conditions (chat or task) in an adaptive and scalable manner. In the following, we describe the dialogue management process, the robot hardware, and the global planner used for the experiment in Sec. IV. All components have been developed for the Robot Operating System (ROS).

A. Robot

To test the system, we used the Pepper robot[1]. Pepper is a 1.21m tall humanoid robot (see Fig. 1) with a holonomic mobile base, two arms with 4 Degrees of Freedom (DoF), two hands with 2 DoF, and a 2 DoF head. The mobile base hosts 3 laser scanners and 2 sonars. The head is comprised of an ASUS Xtion depth camera and two RGB cameras with VGA resolution all of which are used for face and people detection. The head also hosts a 4 microphone array used for audio recording and two loudspeakers (one on either side) for speech synthesis. For speech recognition, Nuance Cloud[2] is used.

B. Speech Processing

To enable chat-style dialogue, the system uses a collection of AIML files forming the chatbot Rosie[3] using the *Program-Y*[4] AIML 2.0 interpreter. The utterance string (*pattern*) is encoded and send to the chatbot via REST calls, where an appropriate response (*template*) is formulated and fed back to the call. To switch from chat to task-based dialogue, the task-related state variables as shown in the

[1] `http://doc.aldebaran.com/2-5/home_pepper.html`
[2] `http://www.nuance.com`
[3] `http://github.com/pandorabots/rosie`
[4] `http://keiffster.github.io/program-y/`

example dialogue in Tab. I are switched based on specific words and phrases used during the interaction.

C. Reinforcement Learning

On each consequent turn, the dialogue manager decides which action to take based on a trained Markov Decision Process (MDP) policy π^*. The policy was designed and trained using *BURLAP* [12], a Java-based framework for RL. The standard *Q-Learning* algorithm [13] was used to train the agent, using hand-crafted simulated users emulating how they could react to each action the agent takes. For example, if the agent responds to a user task utterance (such as "Where can I get a coffee?") with chat, the simulated user will leave the conversation with probability 0.9.

For training, the discount factor γ was set to 0.99, since the agent should care about long-term rewards, while the learning rate α was kept fixed at 0.1. In order for the agent to explore as much as possible during the early stages of the training, an *ϵ-greedy* policy is followed with an *initial* ϵ of 0.9, decaying after each turn i, $\epsilon = \epsilon \cdot \frac{1}{i+1}$.

The system's *states*, denoting the agent's knowledge about its environment at any given time, are represented with 12 features e.g.: *Distance, TaskCompleted, UserEngaged, etc.* (see Tab. I for the full list), resulting in a state-space of approx. 82944 states for policy learning.

The action space consists of 8 actions $a_t \in A$ where A = [*PerformTask, Greet, Goodbye, Chat, GiveDirections, Wait, RequestTask, RequestShop*]. Most of these actions are converted to text using a mixture of template-based generation and database lookup, and are then synthesised as combinations of speech and robot gestures. *PerformTask* can be unpacked in several other tasks, depending on the context of the information given (*CtxTask*). For example if the user requested a discount or a voucher for a specific shop, the robot will present the image of a voucher in QR code on its screen.

The *reward function* is optimising for successful task completion as well as higher engagement, and thus awards each completed task with +10, and +5 for each consequent turn. It also penalises when the user abruptly leaves (i.e. without a 'goodbye' phase) with -100. Starting the training process, the initial Q-values were set to 0 ($Q(s_0, a_0) = 0$).

This system was able to discover optimal actions which human designers would have difficulty anticipating, for example: to trigger the chat behaviour in particular multimodal state configurations where the user is moving away from the robot and a task is incomplete.

D. Execution Framework

The actions of the robot that are not executed as a dialogue action, e.g. finding a person to interact with and starting the dialogue, are controlled by a combination of two existing approaches, i.e. ROSPlan [14] and Petri-Net Plans (PNP) [15]. The benefit of using PNPs is the fast and robust execution of sequences of actions comparable to a Finite State Machine (FSM). In order to prevent hand-crafting these PNPs, we employ ROSPlan to create an appropriate sequence of actions required to achieve the given goal, e.g. engage a human in dialogue, which is subsequently transformed into a PNP and immediately executed. All this happens online and in real-time, i.e. $t \leq 1s$ to generate an action sequence, transform it into a PNP, and start execution. While this allows for fast and flexible sequencing and execution of actions in real-world scenarios, the highly controlled nature of the experimental environment does not necessarily require such a system. Hence, the execution framework is only described for the sake of completeness to allow for reproduction of the described experiment.

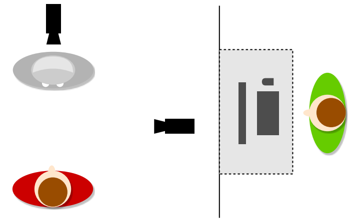

Fig. 2: The participant (red) and the robot (grey) were positioned face to face. The scene was recorded from the robot's perspective focusing on the face of the participant and from the side, showing the whole scene. The experimenter (green) was seated behind a divider.

IV. EXPERIMENT DESIGN AND EVALUATION

The human-robot dialogue system was evaluated via a user study in which human subjects interacted with the Pepper acting autonomously using the system described above; all interactions were in English. The physical setup of the experiment can be seen in Fig. 2.

A. Experimental Scenario

The task and the setup chosen in the study were considered as first steps towards understanding how a humanoid social robot should behave in the context of a shopping mall while providing useful information to the mall's visitors. To this end, participants were asked to imagine that they were entering a shopping mall they had never been to before where the robot was installed in the entry area interacting with visitors one at a time. Participants were asked to complete as many as possible of the following five tasks:

- Get information from the robot on where to get a coffee.
- Get information from the robot on where to buy clothes.
- Get the directions to the clothing shop of their choice.
- Find out if there are any current sales or discounts in the shopping mall and try to get a voucher from the robot.
- Make a selfie with the robot.

Instructions were given to use natural language spontaneously while interacting with the robot.

B. Participants and Experimental Design

In order to explore the benefits of using a combination of task-based and chat-style interaction, two conditions were compared, i.e. purely task-based dialogue (T) and task-based dialogue combined with a chatbot (C) as described in Sec. III. In the first condition, the robot would simply reply

"I am afraid I cannot help you with that" wherever the *Chat* action would have been triggered in the second condition. A within-subject design was used to compare these two conditions. The order of conditions was assigned pseudo-randomly to each participant, avoiding a learning bias in the collected data, and ensuring that half of the participants start with condition T and the other half with condition C.

41 people (13 females, 28 males) agreed to participate in our study, ranging in age from 18 to 38 (M=24.46, SD=4.72). The majority of them were students that had no or little previous experience with robots.

Participants were initially given a briefing script describing the goal of the task and providing hints on how to better communicate with the robot, e.g. "wait for your turn to speak" and "please keep in mind that the robot only listens to you while its eyes are blinking blue"[5]. We reassured our participants that we were testing the robot, not them, and controlled environment-introduced biases by avoiding non-task-related distractions during the experiment. During experimental sessions, participants stood in front of the robot and the experimenter was hidden in another corner of the room but available in case the participant would need any help (see Figure 2).At the end of the experiment participants were debriefed and received a £10 gift voucher. The duration of each session did not exceed thirty minutes.

C. Dependent Variables

We collected a range of objective measures from the log files and audio recordings of the interactions. Following [16], we considered two categories of the objective measures based on those used in the PARADISE framework [17]: conversational efficiency and dialogue quality. In addition, we also considered a range of subjective measures for a qualitative evaluation. For that, after each interaction session participants were asked to fill in a questionnaire to assess their perception of the robot.

Conversational Efficiency was assessed based on how successful participants were in performing the given tasks with the help of the robot. Conversational efficiency was evaluated using the number of tasks completed during a session, the number of performed robot actions, the number of system turns, and the number of tasks completed per turn.

Dialogue Quality concentrated on the success of interaction and was evaluated using the duration of a dialogue, the number of human turns, the human turns per system turn, and the confidence scores from speech recognition.

Perception of Robot was assessed using responses on the questionnaire filled by participants at the end of each interaction session. The questionnaire was based on a combination of the User Experience Questionnaire UEQ [18] and the Godspeed Questionnaire [19]. It consisted of 21 pairs of contrasting characteristics that may apply to the robot, specifically: *fake – natural, machinelike – humanlike, unconscious – conscious, artificial – lifelike, unfriendly – friendly, unkind – kind, unpleasant – pleasant, awful – nice, annoying – enjoyable, disliked – liked, incompetent – competent, ignorant – knowledgeable, irresponsible – responsible, unintelligent – intelligent, foolish – sensible, does not meet expectations – meets expectations, obstructive – supportive, unpredictable – predictable, confusing – clear, complicated – easy, not understandable – understandable.* Validity of the used questionnaire was tested by measuring its internal consistency with Cronbach's α, which was equal to 0.93 (high consistency). Based on the high value of the Cronbach's α, we assume that that our participants in the given context interpreted the robot characteristics, provided in the questionnaire, in an expected way.

[5]Pepper's default way of communicating that it is listening.

V. RESULTS

Results of the data analysis revealed that participants were successful in performing the given tasks in both experimental conditions, with the average number of completed tasks being 3.98 (SD = 0.95) in the task-only condition and 3.93 (SD = 1.10) in the hybrid task+chat condition (see Table II). The number of completed tasks was not significantly different between the two conditions (one-sided T-test, p = 0.83), the same holds for the number of tasks per turn (p = 0.68), system turns (p = 0.28), and the number of actions performed by the robot (p = 0.25). To summarise, the findings showed no significant difference between the two conditions in terms of conversational efficiency, and on average the participants were not adversely affected by the style of conversation when performing their tasks.

TABLE II: Results of conversational efficiency and dialogue quality in two conditions. M denotes mean value, SD - standard deviation.

measure	**task only**		**chat+task**	
number of tasks	M=3.98,	SD=0.95	M=3.93,	SD=1.10
number of actions	M=5.85,	SD=0.80	M=5.65,	SD=0.74
system turns	M=18.75,	SD=7.74	M=21.05,	SD=10.73
tasks per turn	M=0.24,	SD=0.11	M=0.23,	SD=0.12
duration, sec	M=203.99,	SD=80.43	M=228.99,	SD=117.02
human turns	M=20.03,	SD=10.69	M=17.10,	SD=7.10
human turns per system turn	M=0.91,	SD=0.07	M=0.94,	SD=0.05
confidence of speech recognition	M=0.51,	SD=0.01	M=0.51,	SD=0.02

In terms of dialogue quality, the results of the data analysis show that the number of human turns, human turns per system turn, and confidence of speech recognition all have very similar values in both conditions. However, the duration of interaction, although being not significant (one-sided T-test, p = 0.21), differed more and was on average 25 sec (12%) longer in the hybrid task+chat dialogue condition, compared to the task only condition, with a maximum duration of 578 sec in the hybrid condition vs 409 sec in the task only condition (41% longer).

Longer duration of conversation, on the one hand, might mean more frustration and inefficiency in the conversation. On the other hand, longer conversation might mean a more engaging and entertaining interaction experience for a human. The results of the data analysis reveal that a longer

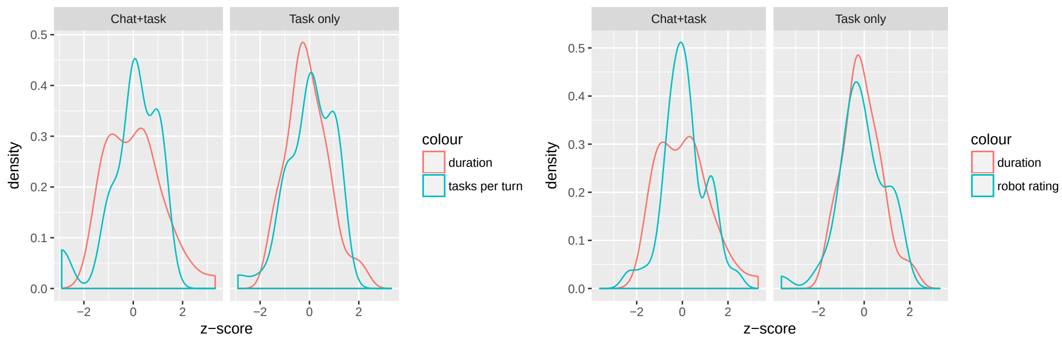

Fig. 3: Density plots representing: Left - distribution of dialogue duration and number of tasks per turn in the two conditions; Right - distribution of dialogue duration and robot ratings in the two conditions. The values of dialogue duration, tasks per turn, and robot ratings are z-normalised to make them comparable.

duration of a conversation does not have a significant influence on the number of completed tasks during interaction (Granger Causality test, p = 0.57), as shown in Fig. 3, left. In general, the results show that longer conversations do not affect efficiency of interaction in a negative way.

In order to understand *whether longer conversations make a better interaction experience* for the human, we analysed the results of the robot perception questionnaire. In general, participants more often gave higher scores and less often give lower or average scores to the robot in the hybrid task+chat condition, compared to the task-only condition (see Fig. 4).

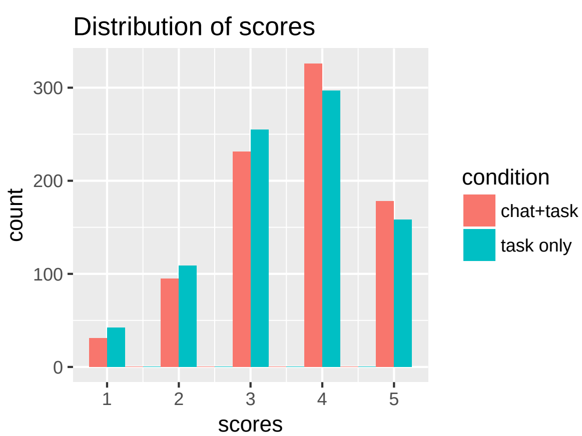

Fig. 4: Distribution of robot ratings (all 21 ratings from the questionnaire are taken into account) in the task-only and hybrid conditions.

The results of the Wilcoxon Signed Rank Tests revealed that humans rated the robot as significantly more "pleasant" ($p < 0.005$, average of 0.42 point higher) in a hybrid task+chat condition, compared to the task-only condition. In addition, the robot was assessed as "it met expectations" significantly better in the hybrid condition ($p < 0.05$, average of 0.39 point higher), compared to the task-only condition. Moreover, the right part of Fig. 3 shows that a longer duration of the conversation with the robot does not result in lower ratings neither in the task-only nor in the hybrid condition. These results suggest that longer conversation with the robot might mean a more engaging and entertaining interaction experience for the human. All other questions did not show any significant differences between the conditions.

VI. DISCUSSION

Our evaluation shows that the participants rated the hybrid task+chat condition significantly higher on a 5 point Likert scale regarding pleasantness of the interaction and meeting their expectations, while all other questions did not show any significant differences between the conditions. Especially the rating for meeting the users' expectations suggests that a system that is able to chat in addition to just fulfil given tasks is more natural to interact with. Moreover, the higher rating for pleasantness of the interaction indicates that this kind of dialogue management system is more engaging to interact with. This is supported by the fact that the duration of the interaction was longer in the hybrid task+chat condition even if not significantly. However, this longer duration might also be an indication that the system was more cumbersome to use and not as efficient in fulfilling the required tasks – but our data does not support this argument, because there is no statistically significant difference between the number of system or human turns and tasks completed between the two conditions, as can be seen from Tab. II. Moreover, none of the questionnaire results concerning this issue show significantly different results either.

In general, the hybrid task+chat system received higher ratings in the questionnaires more often than the task only conditions, as can be seen from Fig. 4. This is another indication that the robot is more engaging to interact with when being able to chat. One of the participants put it nicely when commenting on the hybrid task+chat system: *"It was interesting to see that the more I interacted with the robot the more I could discover new possible questions and answers. This made me feel that I could actually try to make a conversation with the robot."*

Limitations The system presented here is currently rather simple in terms of the chat ability. The chatbot is standard software that has not been fully adapted to the domain of a shopping mall. This has led to confusion once in a while, where participants wanted directions to a non-existent shop

X and asked "How do I get to X" and the chatbot replied "A lot of hard work". While this might be perceived as funny, it is not particularly helpful as a means of clarifying if there is a shop X in the shopping mall or not. Future work will have to address this issue by being able to classify even non-existent items or shops as such, and trigger a task that informs the user of the fact that this shop is not present.

The control condition was a task-based dialogue system which is the current standard in HRI. However, the option of just replacing every chat action of the hybrid task+chat system with saying "I am sorry. I cannot help you with that." is a rather simple way of dealing with the problem of task unrelated statements or questions. A more sophisticated system that is able to inform the user of possible tasks the robot can perform instead might have scored higher in the ratings. However, we also had feedback that welcomed this simple form of reply about the incapability of dealing with the recognised input: *"If it didn't know the answer, it wouldn't give me a related answer, it would just tell me it didn't know."* which was categorised as positive by one of the participant.

A technical limitation of the presented system, as for most dialogue systems, is the reliability of the Automatic Speech Recognition (ASR). The Nuance ASR as used for the presented experiment often misunderstood given commands or would interpret even short pauses as the end of a statement, and participants often mentioned this issue in the questionnaire. In these cases the task-only condition would produce the statement that the robot cannot help with this whereas the chatbot would answer the partially heard sentence in some way. Since this problem is common to all dialogue systems using voice input and the target domain of a shopping mall generally is a very noisy environment, the chatbot offers a possible mean of dealing with this issue and still being able to produce some form of reply other than stating not being able to help.

VII. CONCLUSION AND FUTURE WORK

We presented and evaluated a first approach of a fully autonomous robotic system using a novel combination of task-based and chat-style dialogue. We employed Reinforcement Learning (RL) to create a scalable approach to combining these two modalities, while being able to easily enrich the used feature vector by including information from the robot's sensors in addition to verbal information. Our experiment showed that our participants found the proposed system more pleasant to interact with and had the feeling that it met their expectations better than a purely task-based version of the same approach. On average, participants interacted longer with the robot without impeding the overall task efficiency, which indicates that this kind of robotic agent is more engaging than a purely task-based one. This presents a first step towards a holistic approach to HRI, being able to not only respond to utterances related to an a-priori defined set of tasks but also being able to chat with the human interaction partner.

To further improve the user experience and cope with uncertainty that would arise in real life applications in a shopping mall, in future work the MDP model for RL will be substituted with a POMDP or a related approach such as deep reinforcement learning [20]. This might also provide better results when additional multimodal inputs are employed (e.g. sentiment/emotion analysis from video). The chatbot will also be replaced with one trained on data more suitable to the domain. A more accurate user model will be designed based on collected data, as opposed to the hand-crafted user simulation utilised in this research.

REFERENCES

[1] M. F. McTear, "Spoken dialogue technology: enabling the conversational user interface," *CSUR*, vol. 34, no. 1, pp. 90–169, 2002.

[2] V. Rieser and O. Lemon, *Reinforcement learning for adaptive dialogue systems*. Springer, 2011.

[3] S. Young, M. Gasic, B. Thomson, and J. D. Williams, "Pomdp-based statistical spoken dialog systems: A review," *Proceedings of the IEEE*, vol. 101, no. 5, pp. 1160–1179, 2013.

[4] T. Kollar, S. Tellex, D. Roy, and N. Roy, "Toward understanding natural language directions," in *Human-Robot Interaction (HRI), 2010 5th ACM/IEEE International Conference on*. IEEE, 2010, pp. 259–266.

[5] D. L. Chen and R. J. Mooney, "Learning to interpret natural language navigation instructions from observations." in *AAAI*, vol. 2, 2011, pp. 1–2.

[6] E. Ferreira and F. Lefèvre, "Reinforcement-learning based dialogue system for human-robot interactions with socially-inspired rewards," *Computer Speech & Language*, vol. 34, no. 1, pp. 256–274, 2015. [Online]. Available: http://dx.doi.org/10.1016/j.csl.2015.03.007

[7] ALICE AI Foundation, "S.u.p.e.r. aiml bot," n.d. [Online]. Available: https://code.google.com/archive/p/aiml-en-us-foundation-super/

[8] A. van Woudenberg, "A chatbot dialogue manager," Ph.D. dissertation, Open University of the Netherlands, 2014.

[9] I. Papaioannou and O. Lemon, "Combining chat and task-based multimodal dialogue for more engaging hri: A scalable method using reinforcement learning," in *Proceedings of HRI 17 Companion, Vienna, Austria*, 2017.

[10] A. Dingli and D. Scerri, "Building a hybrid: Chatterbot - dialog system," *Text, Speech, and Dialogue*, pp. 145–152, 2013.

[11] Z. Yu, A. Black, and A. Rudnicky, "Learning conversational systems that interleave task and non-task content," in *arXiv*, no. arXiv:1703.00099, 2017.

[12] J. MacGlashan, "Burlap," 2016. [Online]. Available: http://burlap.cs.brown.edu/

[13] A. G. Sutton, Richard SBarto, *Reinforcement learning*. MIT Press, 1998.

[14] M. Cashmore, M. Fox, D. Long, D. Magazzeni, B. Ridder, A. Carrera, N. Palomeras, N. Hurtós, and M. Carreras, "Rosplan: Planning in the robot operating system." in *ICAPS*, 2015, pp. 333–341.

[15] V. A. Ziparo, L. Iocchi, P. U. Lima, D. Nardi, and P. F. Palamara, "Petri net plans," *Autonomous Agents and Multi-Agent Systems*, vol. 23, no. 3, pp. 344–383, 2011.

[16] D. J. Litman and S. Pan, "Designing and evaluating an adaptive spoken dialogue system," *User Modeling and User-Adapted Interaction*, vol. 12, no. 2-3, pp. 111–137, 2002.

[17] M. A. Walker, D. J. Litman, C. A. Kamm, and A. Abella, "Paradise: A framework for evaluating spoken dialogue agents," in *Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, 1997, pp. 271–280.

[18] B. Laugwitz, T. Held, and M. Schrepp, "Construction and evaluation of a user experience questionnaire," in *Symposium of the Austrian HCI and Usability Engineering Group*. Springer, 2008, pp. 63–76.

[19] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi, "Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots," *International journal of social robotics*, vol. 1, no. 1, pp. 71–81, 2009.

[20] H. Cuayahuitl, S. Keizer, and O. Lemon, "Strategic dialogue management via deep reinforcement learning," in *Proc. NIPS workshop on Deep Reinforcement Learning*, 2015.