Managing dialog and joint actions for virtual basketball teammates

Divesh Lala[1,2] and Tatsuya Kawahara[1]

[1] Kyoto University, Graduate School of Informatics
[2] International Research Fellow of the Japan Society for the Promotion of Science (JSPS)

Abstract. Research on embodied teammate agents which use dialog and gesture to coordinate their activities with the user is relatively sparse compared to conversational agents. We propose a dialog management model to handle interactions between user and agent in a virtual basketball environment. The model describes how a joint action should be initialized and executed through dialog, and how it should handle new dialog interruptions. The model also allows the agent to be parameterized to exhibit different combinations of speech and gestural behavior over repeated joint actions. We propose that this model allows us to conduct several types of unique experiments in this environment.

1 Introduction

Many sophisticated systems in agent research have been built for the purpose of providing face to face interactions between humans and virtual agents [1–3]. However these types of interactions are not the only form of communication. In team sports, interactions occur over a wide area and are relatively infrequent, but the same interactions often reoccur.

In this paper we describe the development of agents who will act not as conversational partners, but as teammates in a basketball environment. There are unique challenges related to this type of environment. Players use shorter utterances and expect that the meaning can be inferred from the game context. For example the utterance "Pass" has a different meaning according to whether or not the speaker has the ball. The management of task dialog is also important in basketball. For example, an agent should know that moving to the left could be a sub-task of a passing joint action.

Our long-term goal is to implement communicative behavior for teammate agents which allows them to be perceived as intelligent as opposed to merely reactive to inputs. In real basketball teammates do not often use explicit signals (such as saying "Pass" when calling for a pass) because of their shared experience. Accordingly, our ideal agent should modify their behavior to gradually reduce their use of explicit signals. We propose that this will be indicative of teamwork between human and agent.

2 Dialog and joint action management

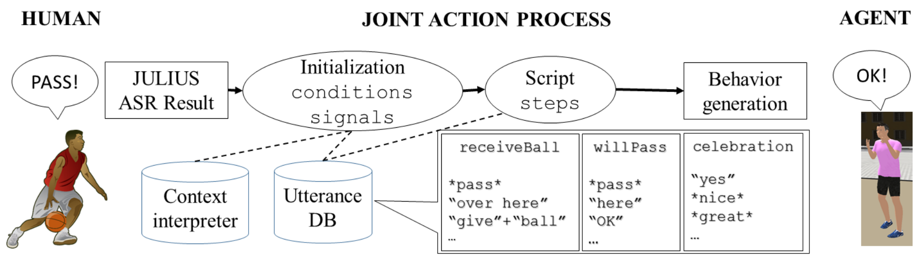

Fig. 1. The general architecture of the dialog manager.

Our general architecture for dialog management is shown in Fig. 1. We consider an interaction between teammates as being a joint action (JA), triggered by either verbal or non-verbal signals. For now we consider only verbal signals and describe an example JA "Call for pass". The JA is initialized as follows:

```
conditions=possession, liveGame, teamAttack
signals=V[receive ball], NV[wave arms], NV[turn to partner]
```

Assume that a human utterance has been received. If the conditional contexts (in possession, a live game, and team on attack) have been met, the agent will check that the utterance (`V`) is related to the `receive ball` action in the utterance database. The goal of this action is, as the name suggests, to receive the ball. This database defines abstract actions and their related utterances. An utterance can be part of one or more actions. For unmatched utterances, the agent hypothesizes the intended JA using the context or analyze prosodic features to filter out irrelevant utterances such as self talk. Once the JA is initialized, we define a script for how the agent behaves:

```
Step1: {V wait}find space
Step2: {V willpass}turn to partner
Step3: {V throwpass}pass
```

Each `Step` describes a specific action for the agent, in this case finding space, turning to a partner, then throwing a pass. During these actions, the agent *may* select a corresponding utterance (indicated by `V`). The JA also contains success and failure states to end the JA, although this is not shown here for brevity.

This structure can be used to modify the type of expressive signals of the agent. The script contains information on the verbal utterances an agent may wish to use, but they do not need to. The decision on whether or not to use an utterance is dependent on the beliefs of the agent and can be parameterized.

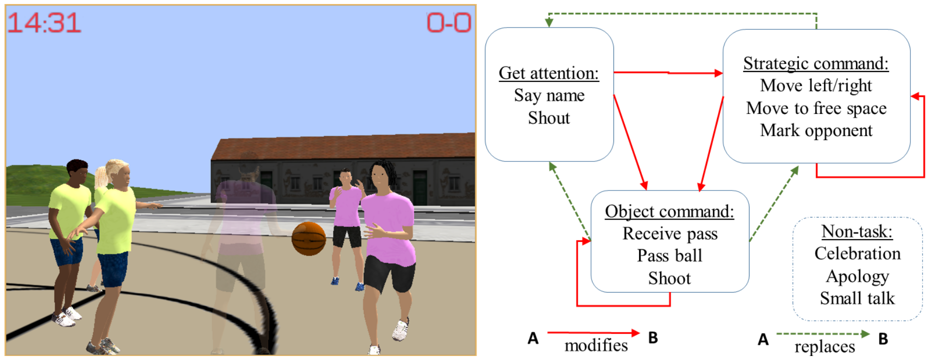

Fig. 2. A screen shot of the game (left) and the model for joint action categorizations and interruption handling (right).

A feature of basketball is that new dialog may interrupt a JA even before it has finished. The agent must decide whether this interruption *replaces* an existing JA, or *modifies* it. For example, saying "Go left" replaces an existing JA of "Go right", but modifies a "Pass ball" JA, because it assumes that going left is a sub-task to be completed in order to pass the ball. We handle interruptions by categorizing JAs into four types as in Fig. 2. Three of the types are related to task behavior - getting attention, object commands and strategic commands. Object commands are joint actions those which involve moving the ball, while strategic commands involve only the player. Through this structure the agent decides how JAs should be handled given a new utterance.

Our system implementation uses the VISIE system described in previous work [4]. This system allows the user to play basketball without handheld peripherals by recognizing gestures of passing, shooting and dribbling. The user navigates throughout the environment by walking on a pressure sensor. We integrated a Japanese speech recognition system, Julius [5], which allows the user to communicate through spoken commands via a headset.

We created two agents, Akira and Tamako, to act as teammates for a human player. These agents can also recognize non-verbal communication signals [6]. To implement the above dialog management system, we conducted Wizard-of-Oz experiments to collect data on the speech and gesture used by humans during the basketball game. We then created a speech corpus and categorized the utterances which will be recognized in the system.

3 Experimental scenarios

Our system allows humans to play with both agents and compare the two directly. With our architecture we are able to parameterize the communicative behavior of the agents in the game. More specifically, we can address the following interrelated research issues on the perception of agents in this environment:

- The ideal ratio of verbal to non-verbal signals. Do users have a preference for agents responding with speech, gesture, or both?
- The use of signals as interaction progresses. Is it more natural for the agent to slowly transition - from explicit utterances and gestures to more implicit signals such as body rotation - as the user becomes familiar with them during the interaction?
- The matching of expressive signals with that of the user. Does the type of modality used for signaling affect the user's perception of the agent?

Our model allows us to store a history of communicative acts, and based on the target phenomena decide the method of communication during a joint action. For the first and second questions, the agent decides the form of their signal based on the target ratio of modality types and historical signal explicitness, respectively. The third question involves recognizing the modality used by the human and responding using the same modality.

4 Conclusion

We described how our virtual basketball agents manage user dialog and how they can flexibly use expressive signals to coordinate joint actions. The framework can also be generalized to other domains outside of basketball, particularly for those which require multimodal coordination of tasks. Our future work is to conduct experiments testing various parameterizations and determine what kind of signals are most suitable for human-agent interaction in this environment.

References

1. Niewiadomski, R., Bevacqua, E., Mancini, M., Pelachaud, C.: Greta: An interactive expressive ECA system. In: Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems-Volume 2, International Foundation for Autonomous Agents and Multiagent Systems (2009) 1399–1400
2. Schroder, M., Bevacqua, E., Cowie, R., Eyben, F., Gunes, H., Heylen, D., ter Maat, M., McKeown, G., Pammi, S., Pantic, M., Pelachaud, C., Schuller, B., de Sevin, E., Valstar, M., Wollmer, M.: Building autonomous sensitive artificial listeners. IEEE Transactions on Affective Computing **3**(2) (2012) 165–183
3. DeVault, D., Artstein, R., Benn, G., Dey, T., Fast, E., Gainer, A., Georgila, K., Gratch, J., Hartholt, A., Lhommet, M., et al.: Simsensei kiosk: A virtual human interviewer for healthcare decision support. In: Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems, International Foundation for Autonomous Agents and Multiagent Systems (2014) 1061–1068
4. Lala, D., Nishida, T.: VISIE: A spatially immersive interaction environment using real-time human measurement. In: 2011 IEEE International Conference on Granular Computing (GrC). (2011) 363–368
5. Lee, A., Kawahara, T., Shikano, K.: Julius – an open source realtime large vocabulary recognition engine. In: EUROSPEECH. (2001) 1691–1694
6. Lala, D., Nishida, T.: A data-driven passing interaction model for embodied basketball agents. Journal of Intelligent Information Systems (2015) 1–34