Roles of a Talking Head in a Cooperative Human-Robot Dialogue System⋆

Mary Ellen Foster

Informatik VI: Echtzeitsysteme und Robotik
Fakultät für Informatik, Technische Universität München
Boltzmannstraße 3, 85748 Garching bei München, Germany
foster@in.tum.de

The JAST human-robot dialogue system [1] is designed as a platform to integrate empirical findings on cognition and cooperative dialogue with research on autonomous robots by supporting multimodal human-robot collaboration on a construction task. The robot consists of a pair of mechanical arms with grippers, mounted in a position to resemble human arms, and a Philips iCat animatronic talking head [2]. The user and the robot work together to assemble construction toys on a common work area, coordinating their actions through speech, gestures, and facial displays. Fig. 1 shows the set-up of the robot.

The talking head has three distinct functions in the JAST system. It moves its lips to accompany the synthesised speech; it produces a range of facial expressions; and it modifes its gaze direction to look at the user or at specific locations on the table. Lip-synchronisation is straightforward to implement in the iCat software and plays a clear role in adding to the naturalness of the output. However, the other two functions of the animated head require more sophisticated implementations and have a more complex effect on the system output.

The facial expressions of the talking head may accompany specific parts of the speech—for example, the head might raise its eyebrows to emphasise a particular word or phrase in the output, or look pleased while thanking the user for helping with a task. Other displays may be produced independent of speech to provide non-verbal feedback or back-channeling, such as displaying a "thinking" expression while processing the user input, nodding to indicate agreement, or looking confused to indicate that there are problems with the speech-recognition system. Producing appropriate facial expressions requires a mechanism for choosing the correct commands and coordinating them with the output on other channels and with the user's behaviours.

The gaze direction of the talking head can be controlled to look at objects on the table when it manipulates or refers to them, and to look at the user's face when it addresses them directly. As well, gaze direction can be used as a form of fast, imprecise pointing, as the head is able to change position much more quickly than the robot arms. To include all of these behaviours in the system, the output planner must have access to the location of the user and of all objects in the world; it must also select appropriate gaze changes to include in the output
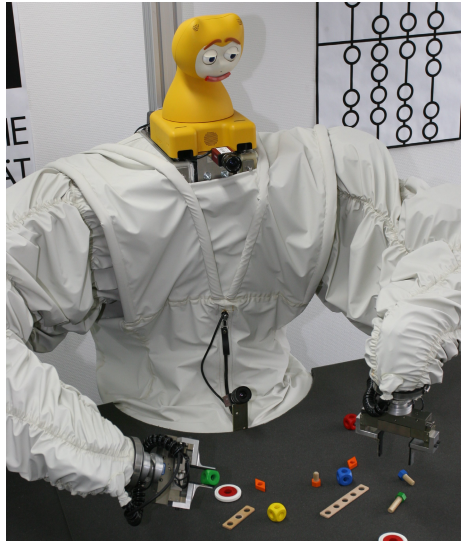

Fig. 1. Set-up of the robot in the JAST dialogue system

and, as for facial expressions, must coordinate the selected behaviours with the output on other channels and with the user's behaviours.

A number of studies have demonstrated that the expressiveness of an embodied agent—physical or virtual—has a significant effect on the experience of users of interactive systems. For example, a version of the Mel robot agent that used a full range of expressive and gaze behaviours increased user engagement the system [3]; the expressive facial displays of the animated virtual agent in the COMIC dialogue system mitigated the slowness of the system [5]; while the Leonardo robot [4] actually communicates entirely through body language and gestures. We expect the facial displays and gaze feedback of the embodied head in the JAST human-robot system to have a similarly positive effect on users' interactions with the system.

References

1. Foster, M.E., By, T., Rickert, M., Knoll, A.: Human-robot dialogue for joint construction tasks. In: Proceedings, 8th International Conference on Multimodal Interfaces (ICMI '06), pp. 68–71 (2006)
2. van Breemen, A.J.N.: iCat: Experimenting with animabotics. In: Proceedings, AISB 2005 Creative Robotics Symposium (2005)
3. Sidner, C.L., Lee, C., Kidd, C.D., Lesh, N., Rich, C.: Explorations in engagement for humans and robots. Artificial Intelligence 166(1-2), 140–164 (2005)
4. Breazeal, C., Brooks, A., Gray, J., Homan, G., Kidd, C., Lee, H., Lieberman, J., Lockerd, A., Chilongo, D.: Tutelage and collaboration for humanoid robots. International Journal of Humanoid Robotics 1(2), 315–348 (2004)
5. White, M., Foster, M.E., Oberlander, J., Brown, A.: Using facial feedback to enhance turn-taking in a multimodal dialogue system. In: Proceedings, HCI International 2005. (2005)