A Virtual Poster Presenter using Mixed Reality

Vanya Avramova ✉, Fangkai Yang, Chengjie Li,
Christopher Peters, and Gabriel Skantze

KTH, Stockholm, Sweden
avramova,fangkai,chengjie,chpeters,skantze@kth.se

Abstract. In this demo, we will showcase a platform we are currently developing for experimenting with situated interaction using mixed reality. The user will wear a Microsoft HoloLens and be able to interact with a virtual character presenting a poster. We argue that a poster presentation scenario is a good test bed for studying phenomena such as multi-party interaction, speaker role, engagement and disengagement, information delivery, and user attention monitoring.

1 Introduction

Interacting with a virtual character on a 2D display is very different from a physical robot or another human being. When viewed on a display, the animated agent does not share the same space as the observer. This makes it very hard to engage in *situated interaction* with the agent, where the shared space is of importance. For example, the agent cannot have exclusive mutual gaze with one of the observers: it will either appear to look at nobody, or at everyone in the room, the so-called Mona Lisa effect [2]. Also, it becomes hard to reach joint attention to objects in the shared environment. This is why situated interaction so far has been explored mostly in human-robot interaction scenarios, where these problems are avoided [2]. However, physical robot platforms are typically very expensive to build, alter and maintain. A consequence is that most researchers use the same platforms, which limits the exploration of possible designs, and by extension the possibilities for various forms of interactions. Some platforms provide flexibility in appearance, such as the Furhat robot head [1], where an animated face is back-projected on a mask, allowing for different animations to be displayed. Such flexibility makes it possible to study how the robot's appearance affect the interaction, including phenomena like the uncanny valley [4]. However, most robots still have very limited flexibility when it comes to the rest of the body, which can have important communicative functions (such as posture and pointing gestures).

We are currently developing a platform which will allow for situated interaction with virtual agents, and thereby provide a possibility to experiment with the effects of agent appearance on the interaction. The idea is to use Mixed Reality (MR), a technology that facilitates the overlay of computer graphics constructs onto the real world. Unlike Virtual Reality (VR), the graphics does not completely replace the environment. This is important for social interaction,

J. Beskow et al. (Eds.): IVA 2017, LNAI 10498, pp. 25-28, 2017.
DOI 10.1007/978-3-319-67401-8_3

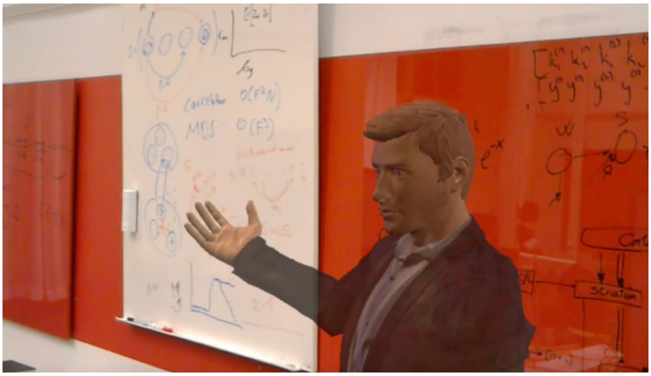

Fig. 1. The virtual presenter in mixed reality, as seen from the user's perspective.

since it allows the user to directly perceive her own body in a natural way. Also, physical objects (such as a table or objects under discussion) can be mixed with virtual objects, or augmented with virtual artifacts. Using this platform, we will investigate interactional phenomena that are specific to situated interaction, for example the impact of multimodal behaviour (speech, facial expression, full-body motions, conversational formations) and embodiment on turn-taking and joint attention.

A virtual agent in mixed reality can of course have important applications, such as virtual museum guides, tutors, pets, companions, etc. But the platform could also be used for the development of virtual replicas of robots, and test them out in interaction with humans, before they are manufactured, greatly reducing the development costs and increasing the room for experimentation.

In this "meta" demo, we will showcase the current status of the platform by allowing the user to interact with a virtual character which is presenting a poster that shows the current status of the platform, as illustrated in Figure 1. We think that a poster presentation scenario is a good test bed for studying phenomena such as multi-party interaction, speaker roles, engagement and disengagement, information delivery, and user attention monitoring (cf. [3]).

2 Platform

The platform involves four major components: The IrisTK multi-modal dialogue framework (www.iristk.net), a full-body animation controller [6], Unity3D, and the Microsoft HoloLens. An overview of the architecture is shown in Figure 2.

IrisTK has previously been used in several studies on human-robot interaction, and is especially targeted towards multi-modal, situated interaction [5]. IrisTK provides a set of modules for receiving sensory data from the users and the environment (e.g., speech recognition, body locations, item locations), and merges these into a 3D model of the physical situation (i.e., the status of agents

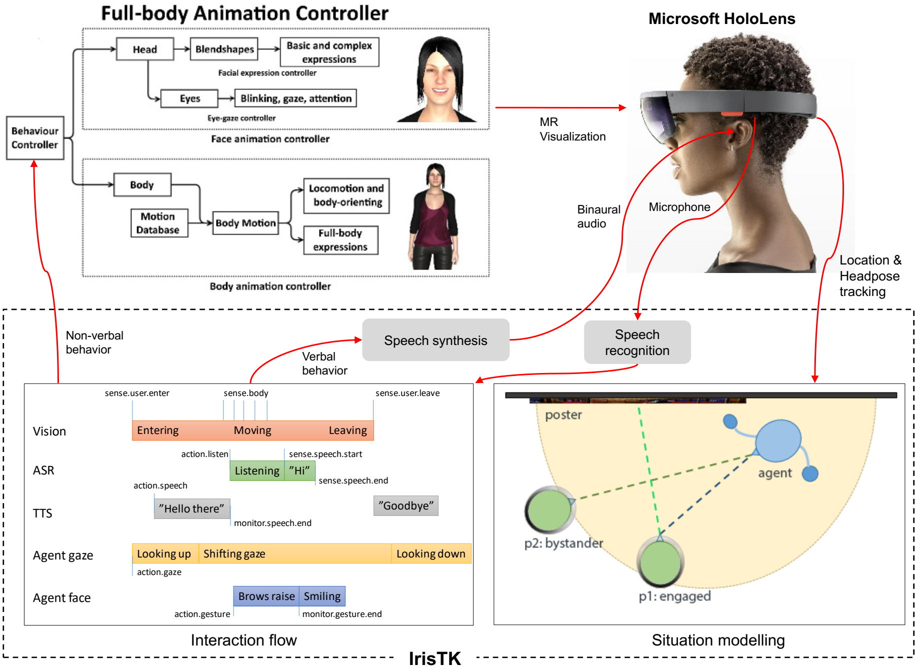

Fig. 2. Platform architecture.

and objects). This situation model can then be used to make higher-level inferences about when users enter and leave the interaction space of the robot, the conversational roles of the users, and where the agents' attention is targeted. When applied to human-robot interaction scenarios, IrisTK typically uses a sensor like Microsoft Kinect to track users and receive audio from the user. In this setting, we instead use the input and output modalities provided by the Microsoft HoloLens. The HoloLens scans the room in order to be able to track the users position and head pose, and contains gyro and accelerometer sensors. This allows it to position virtual agents and objects in the 3D environment and render them on the see-through display. The HoloLens can also give binaural audio to the user, effectively simulating the auditory experience of different locations of the virtual agent.

The full-body animation controller is part of a character behaviour toolkit supporting and enabling the development of expressive social interaction behaviours for small groups of virtual characters [6] . It supports parameterised control of their facial expressions, gaze, body gestures and posture using the face and body animation controller (see Figure 2), as well as locomotion, body orientation and free standing formations using a formation controller. Coordinated multimodal behaviours (facial expressions animated via blendshapes, gaze motions and full-body motion-captured animations) are controlled through a behaviour controller and defined using hierarchical FSMs. The controller and toolkit are embedded in Unity 3D, a real-time game engine providing sophisticated modern rendering, animation and audio capabilities. Unity supports real-time interaction through its support of a variety of virtual reality and augmented reality devices, including the Microsoft HoloLens.

3 Discussion and Future work

In addition to quick experimentation and control over the agent's embodiment, we plan to also use the mixed reality setting to explore unorthodox but situationally appropriate interaction cues, and study human acceptance and adaptability to these signals. For example, important objects in the setting can be highlighted instead of pointed to by the agent, or attention can be drawn by using sound localized on the target object. The agent also has access to unique ways of expressing its inner state, or its understanding of the affective state of their human counterpart. These methods will be explored to enable a platform which provides multimodal interactive settings unique to mixed reality environments.

While the HoloLens is capable of rendering holographic imagery at very high resolution, it does so at the expense of its field of view. The boundaries of the visible view screen are relatively small, and may often truncate holographic content as the wearer turns their head. This presents a drawback for the interaction, as the view cuts off objects that the wearer feels that they should be able to see, especially on the vertical view axis. We have to consider carefully the scale and placement of holograms in the interaction settings, so that the immersiveness of the experience is minimally affected.

References

1. Al Moubayed, S., Skantze, G., Beskow, J.: The furhat back-projected humanoid head-lip reading, gaze and multi-party interaction. International Journal of Humanoid Robotics 10(1) (2013)
2. Al Moubayed, S., Skantze, G.: Turn-taking Control Using Gaze in Multiparty Human-Computer Dialogue: Effects of 2D and 3D Displays. Proceedings of the International Conference on Audio-Visual Speech Processing 2011 pp. 99–102 (2011)
3. Eichner, T., Prendinger, H., André, E., Ishizuka, M.: Attentive Presentation Agents. In: Intelligent Virtual Agents, pp. 283–295. Springer Berlin Heidelberg, Berlin, Heidelberg (2007)
4. Paetzel, M., Peters, C., Nyström, I., Castellano, G.: Congruency matters - How ambiguous gender cues increase a robot's uncanniness. In: International Conference on Social Robotics. vol. 9979 LNAI, pp. 402–412 (2016)
5. Skantze, G., Moubayed, S.: IrisTK: A statechart-based toolkit for multi-party face-to-face interaction. In: ICMI'12 - Proceedings of the ACM International Conference on Multimodal Interaction (2012)
6. Yang, F., Li, C., Palmberg, R., der Heide, E.V., Peters, C.: Expressive Virtual Characters for Social Demonstration Games. In: 9th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games) (submitted)