

User Interaction in Mixed Reality Interactive Storytelling

Marc Cavazza (1), Olivier Martin¹ (2), Fred Charles (1), Xavier Marichal (3) and Steven J. Mead (1)

(1) School of Computing and Mathematics, University of Teesside, Middlesbrough, TS1 3BA, UK.

(2) Laboratoire de Télécommunications et Télédétection, Université catholique de Louvain,
2 place du Levant, 1348 Louvain-la-Neuve, Belgium.

(3) Alterface, 10 Avenue Alexander Fleming, 1348 Louvain-la-Neuve, Belgium.

m.o.cavazza@tees.ac.uk, martin@tele.ucl.ac.be, f.charles@tees.ac.uk,

xavier.marichal@alterface.com, steven.j.mead@tees.ac.uk

1. Introduction

One promising application of Mixed Reality to entertainment has been the development of “interactive theatre” systems involving synthetic actors. However, these systems do not make use of the most recent advances in interactive storytelling technologies, which use Artificial Intelligence techniques to generate real time narratives featuring synthetic characters and supporting user intervention in the plot.

In this paper, we describe a Mixed Reality system based on a “magic mirror” model (Figure 1), in which the user’s image is captured in real time by a video camera, extracted from his/her background and mixed with a 3D graphic model of a virtual stage including the synthetic characters taking part in the story. The resulting image is projected on a large screen facing the user, who sees his/her own image embedded in the virtual stage with the synthetic actors. The graphic component of the Mixed Reality world is based on a game engine, Unreal Tournament 2003™. This engine not only performs graphic rendering and character animation but incorporates a new version of our previously described storytelling engine [1]. A single 2D camera facing the user analyses the image in real-time by segmenting the user’s contours [2]. The objective behind segmentation is twofold. Firstly, it extracts the image silhouette of the user in order to be able to inject it into the virtual setting on the projection screen. Secondly, it can recognise user behaviour, including symbolic gestures, in real-time, hence supporting a new interaction channel.

2. The Interactive Storytelling engine

Our approach is character-based, which means that the narrative is driven by the individual roles of each of the virtual actors, rather than by an explicit plot representation. The actors’ roles are formalised as plans,

which are executed in real time using a modified Hierarchical-Task Network planning algorithm. During execution, the planner selects the next action for a character, this action being played in the virtual environment, which is also updated to take into account its consequences. When an action fails (i.e. its intended outcome is not achieved), another course of action is generating through re-planning. The real-time selection of action supports interactivity, as the user can interfere with the environment, changing the executability conditions of potential actions.

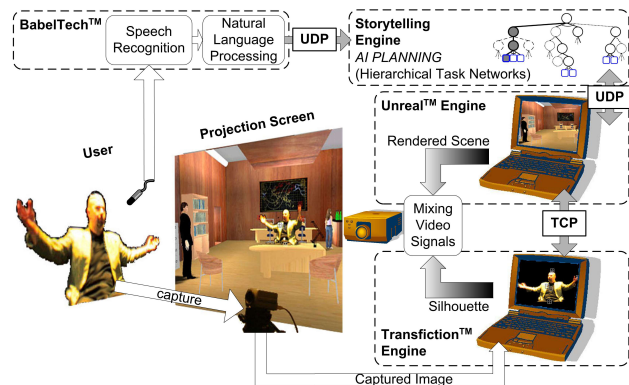


Figure 1. System overview.

3. User Interaction

The fact that the user visually takes part in the story presentation obviously affects the modes of user intervention with respect to other interactive storytelling systems: these will have to take the form of traditional interaction between characters. In other words, unlike in some of the systems previously developed, the user will have to *act*. This means that the actions by which he may interfere with the story should have a visual presentation that blends into it. In other words, the mechanisms of his normal acting (gestures, speech) should serve as a natural basis for his intervention in the storyline. These

¹ Olivier Martin is funded through a FIRST Europe fellowship from the Walloon Region (Belgium).

mechanisms comprise physical interaction (the direct contact between the user and virtual world entities), symbolic gestures, and speech (the latter two combining through various forms of multi-modal interaction).

Physical interaction consists in all forms of contact between the user (or more precisely the user's embodiment through his video avatar) and virtual actors. This interaction is mediated by the low-level mechanisms managing interaction between actors in the Unreal Tournament 2003™ engine. Physical interaction is based on the use of a similar co-ordinate system for both the virtual world and the video information captured from the real setting. The exact form of physical interaction is triggered by the simultaneous recognition of a user's gesture and a collision between bounding boxes. This will in turn determine the action performed on a virtual actor and affect the storyline accordingly. For instance, depending on the gesture recognised (a handshake or a slap), the corresponding agent reaction is displayed (the appropriate animation being played) and the consequence of this reaction is used to update the agent's plan.

User gestures are recognised through a rule based system that identifies gestures from a gesture library, using data from image segmentation that provides in real time the position of user's extremities. One essential aspect of the interaction is that the system is actually tracking symbolic gestures corresponding to narrative functions, such as greetings, threatening (or responding to a threat, such as putting his hands up), offering, calling, dismissing, etc.

However, gestures are mostly used in conjunction with speech recognition. The speech recognition component is based on the Ear SDK from BabelTech™. Speech and gesture information is combined using temporal information as classically described for multi-modal utterances. The multi-modal parser recognises speech acts, such as e.g., greetings, questions, information

provision, or threats. These speech acts can then be mapped to the actor's plan to modify its progression.

4. An example

The storytelling scenario supporting our experiments is based on a James Bond adventure, in which the user is actually playing the role of the villain. Figure 2 illustrates the influence of user actions on the unfolding of the story, through their impact on the progression of the main character's (Bond) plan. The plan progression can be illustrated through the resolution of its various sub-tasks formalised as part of an HTN. The section of the plan represented in the figure corresponds to Bond's objective of acquiring a given piece of information, which can be obtained from various characters, depending on the circumstances. The user's action will precisely alter these circumstances (though not always conspicuously, due to the intricacy of the actors' plans). In the above example, when the user decides to call his assistant to his desk, this will open the possibility for Bond to attract her attention and obtain the coveted information from her. Naturally, the course of action can be further altered at a later stage by the user.

5. References

[1] M. Cavazza, F. Charles, and S.J. Mead, "Character-based Interactive Storytelling", *IEEE Intelligent Systems*, special issue on AI in Interactive Entertainment, 2002, pp. 17-24.
 [2] X. Marichal, and T. Umeda, "Real-Time Segmentation of Video Objects for Mixed-Reality Interactive Applications", Proceedings of the "SPIE's Visual Communications and Image Processing" (VCIP 2003) International Conference, Lugano, Switzerland, 2003.

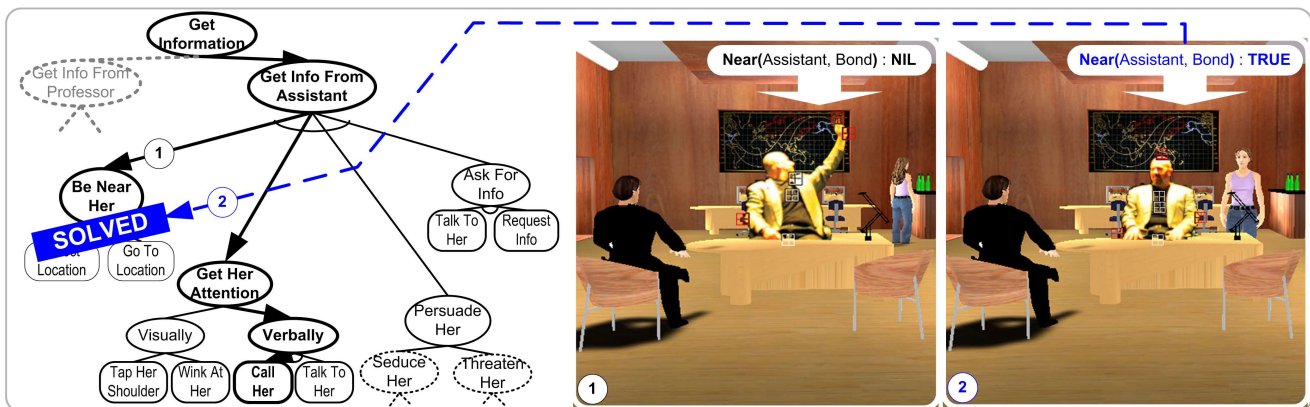


Figure 2. Mechanisms of Interactive Storytelling: User's actions interfere with the virtual actor's plans.