User-Recordable Non-Player Characters for Croquet

Mark P. McCahill, Liz Wendland, Peter Moore
Office of Information Technology - University of Minnesota
mpm@umn.edu, wendland@umn.edu, pmoore@umn.edu

Abstract

Visually rich, immersive multi-user environments are commonly used for entertainment and online socializing. However, growing attention is being given to education and simulations for training using these immersive "game-oriented" technologies. In this paper, we describe extensions to Croquet facilitating the recording and playback of avatar interactions in 3D multi-user spaces. Applications of recordable user-controlled avatars include creation of Non-Player Characters (NPCs), and recording/playback of social interactions for education, such as foreign language instructional environments. By combining a set of user-controlled avatar gestures with live and pre-recorded avatars in a 3D Croquet space, we can create a virtual laboratory for users to practice both language and the social gestures which go with the language – and then replay the actions to both analyze the interactions and try different viewpoints and roles.

1. Introduction

As interest grows in using avatar-mediated immersive spaces for education a number of preliminary experiments are being conducted using game-oriented Massively Multiuser Online (MMO) systems. For instance, the Harvard Law School is using Linden Lab, Inc.'s MMO Second Life[1] in the fall of 2006 as a platform to teach a course which can be attended in avatar form[2]. This effort attempts to leverage the immersive and engaging nature of avatars in online social spaces for discussion and instruction.

Empirical confirmation of the folklore of game players - who feel that their avatars truly embody their virtual selves - is also becoming available. Nick Yee [3] presents evidence that suggests users of MMOs unconsciously adopt real world social conventions in virtual environments with the relative positioning of avatars and gaze - even though the avatars are controlled via mouse and keyboard and "talking" may consist of typed text chat.

All this suggests that an immersive 3D environment with avatar-mediated communications may be useful for simulation of social situations. To explore this area we have begun constructing a Croquet-based education environment to facilitate learning and study of gesture and social language (or pragmatics) – particularly for languages that are not the learner's native tongue.

Croquet [4] is well suited to construction of these sorts of learning environments because it provides both text and audio chat, and is easily modifiable. Previous work on Croquet laid the groundwork for delivering a persistent multi-user learning space [5] [6] [7] which can also extend to Jabber [8] chat users outside the learning space [9]. However, to be able to realistically model social interactions required significant improvements to Croquet's avatars to support life-like motion animations and a way to record and play back avatar motions and speech. This led us to extend Croquet to include BVH motion capture animation playback and user recordable avatars. The recordable avatars can then act as NPCs or script-animated avatars on playback.

2. BVH animations for Avatars

In the initial Hedgehog release of the Croquet software development kit, avatars are little more than 3D cursors – the only user-control of the avatar other than position/orientation in space is the avatar's head tracking of the 2D mouse pointer. Although crude, significant amounts of non-verbal information can be conveyed by intentionally moving the mouse and so moving the avatar's head/gaze. There are drawbacks to this approach, however. Since *any* 2D mouse motion is tracked, user interactions with menus are also reflected in the avatar's gaze in the 3D scene which can send mixed non-verbal messages to other players. We found that experienced users can deduce from other avatar's head motions the pull down menus are being used – this may break some of the immersive effect of the space. Ideally, the 2D user interface interactions in the private portion of the user interface should not muddle

Fifth International Conference on
Creating, Connecting and Collaborating through Computing(C5'07)
0-7695-2806-6/07 $20.00

the social cues we wish to project into the shared scene.

While shaking an avatar's head up and down or left and right can provide basic non-verbal cues, serious attempts at modeling social interactions require whole body animations. BVH motion capture data is a common approach in the game and entertainment realms for representing poses and animations of avatars, and a number of commercial and open source avatar modeling tools support capture and manipulation of BVH-coded avatar poses. By extending the avatar class for Croquet to support BVH motion capture animations, we were able to create scenes with significantly richer avatar gestures.

walking avatar moving his arms and legs

Our BVH-enabled avatars combine an avatar mesh and a collection of BVH motions to create a virtual character with a set of user-callable actions. Some of these animations can also be called automatically. For instance, a walking animation loop should run automatically when the avatar is being moved across the landscape. Similarly, a set of animations for an avatar standing in place should be run automatically, so that the user can concentrate on higher order interactions and avatar controls.

For more situation-specific animation, user control is essential to convey appropriate social meaning. While using the mouse to select from a menu of available expressions and animations may seem appealing, this suffers from the problem of diverting the avatar's gaze as the avatar's head tracks the mouse motion.

Another approach is to use the keyboard to enter commands in a chat window to trigger animations. This approach is used in some MMO game systems and is appealing in Croquet since if we use audio chat for voice, the keyboard may be used to issue commands. Unfortunately, it can be difficult to remember all the available commands so this approach has drawbacks as well.

A third approach that we have considered but not yet implemented is using a sort of flick of the mouse to call a menu of available motions into being. By trapping the mouse flick gesture, it should be possible to provide menu selection of animations without leaking 2D user interface mouse motions into the avatar gaze tracking in the shared 3D world.

an idle avatar: checking his wristwatch

With the basic tools in place for creation of more realistic avatars, we turned our attention to providing tools for recording avatar actions.

3. Recordable Avatars

Croquet's message-based architecture allows for relatively straightforward capture of avatar motions and actions by capturing the replicated messages used to control the avatar in the shared space and storing these messages to a file for later playback. For ease of synchronization it is simplest to record both the motion and the audio chat for an avatar at the same time. However to enable re-recording or overdubbing of avatar speech and motion we treat the motion and audio chat as two separate channels or "tracks" which is similar to multi-track audio recording. By storing the motions and audio events separately it is possible to place a different voice on an avatar recording by playing back the motion track while recording the voice over.

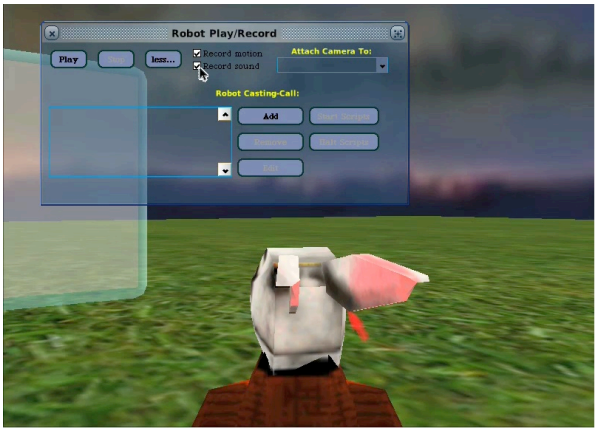
Users can control recording and playback by calling up a "Robot Play/Record" dialog. In minimized state, the dialog allows starting/stopping of record and playback and switching the user's camera between their avatar and any NPCs (robots) in the space with the attach camera popup.

Being able to switch your virtual viewpoint from your avatar to any of the NPCs playing in the space makes it possible to experience a recorded scene from

several different perspectives. This concept is useful for creating guided tours and allowing users to select from a number of virtual tour guides. More importantly, it allows wearing another avatar's skin to observer a social interaction in an immersive way from a new perspective.

Robot play/record dialog minimized

The maximized state of the dialog includes options for adding prerecorded motions, audio tracks, and looped sounds to NPCs – the "Robot Casting Call". Users select an avatar mesh for the NPC, and then are prompted to add actions and triggers.



Robot play/record dialog maximized

Several types of triggers can fire recorded sounds and motions. The *play* trigger corresponds to the play button in the record/play dialog – this a particularly useful when creating a narrative where a number of avatars and interactions will be added, once track at a time to the scene. In this case, it is best to record the longest running avatar first. Another technique for simple synchronization is to record a throw-away avatar to use as a visual cue for recording other avatars. Once the other avatars have been recorded, the throw away track can be discarded – something like a *click track* in mulitrack audio recording.

For NPCs that should run on a timed repeating cycle, the *timer* trigger is starts the animations on a recurring basis. This sort of trigger is useful for NPCs which follow a repeated path such as birds or avatars representing people milling about in a crowd scene.

The *click* trigger starts the animation when the NPC is clicked on by another user in the space. This is useful for creating NPCs that are guides or helpers that much be interacted with to start their spiel.

With a *proximity* trigger, any avatar approaching the NPC within a specified trigger distance starts the action scripts. This is particularly useful for creating NPC guides or sages who provide advice and instructions to avatars that wander near them.

The *chat* trigger is fired when an escaped command is typed in the croquet chat – for instance "/meow", "/run" or "/fly". This sort of trigger is useful for scripting NPC pets and companions.

connecting triggers to actions

While a full-fledged e-toys style tile/text scripting environment would be an obvious enhancement, surprisingly rich environments can be build with these simple tools for animating and recording NPCs. We expect to add an option to include more traditional scripting options as an adjunct to the tool presented here. However, for many of the social interactions we need to model, it is more important to retain a human feel to the avatar gaze and positioning of the NPCs, so recording human's controlling the virtual actors is probably the most efficient approach.

4. Machinima

We were pleasantly surprised to realize that production of movies of the virtual space or *machinima* is greatly simplified by use of recorded avatars. Typically, machinima are recorded in real time with live user running the avatars, but this limits the options

0-7695-2806-6/07 $20.00

for movie capture and camera angles. By recording the avatar actors, it is possible to replay the avatar's interactions as many times as necessary to get the proper camera angles. Moreover, capturing camera shots from the viewpoint of each of the players is simplified. Finally, by creating an invisible avatar, it is possible to pre-define a number of camera tracks through the scene independently of any of the visiblew avatar – and use this as a camera path.

5. Summary and future directions

We now have the tools in place to construct social spaces with realistic avatars, collections of gestures, poses, and animation. Combined with the ability to record and playback users' avatar interactions this allows for role-playing scenarios that can be replayed, studied, discussed, and analyzed. Virtual tours of spaces and social scenarios can also be created and explored by following the viewpoint of a recorded player.

We are now in the process of collecting libraries of gestures and building a language learning space combining social gesture and speech for non-native Spanish speakers. This proof of concept work is designed to validate our theory that combining a persistent virtual environment with a series of tasks structured as game-like quests can be an effective space to learn, reflect on, and perfect social communication in a new language and culture.

5. References

[1] Second Life, Linden Lab, Inc. *http://www.secondlife.com/*

[2] C. Nesson, R. Nesson, and G. Koo, Harvard Law School course "CyberOne: Law in the Court of Public Opinion", *http://blogs.law.harvard.edu/cyberone/*

[3] N. Yee, J. Bailenson, M. Urbanek, F. Chang, D. Merget, "The Unbearable Likeness of Being Digital: The Persistence of Nonverbal Social Norms in Online Virtual Environments", *CyberPsychology and Behavior,* in press.

[4] D. A. Smith, A. Raab, and D. P. Reed, A. Kay. "Croquet: A Menagerie of New User Interfaces." In: *Proceedings of the Second Annual Conference on Creating, Connecting, and Collaborating through Computing*, , IEEE Computer Society Press, 2004.

[5] M. P. McCahill and J. Lombardi, "Design for an Extensible Croquet-Based Framework to Deliver a Persistent, Unified, Massively-Multiuser, and Self-Organizing Virtual Environment." *Proceedings of the Second Conference on Creating, Connecting and Collaborating through Computing,* IEEE Computer Society Press, 2004.

[6] J. Lombardi and M.P. McCahill, "Enabling Social Dimensions of Learning Through a Persistent, Unified, Massively Multi-User, and Self-Organizing Virtual Environment", *Proceedings of the Second Conference on Creating, Connecting and Collaborating through Computing (C5'04*), IEEE Computer Society Press, 2004.

[7] J. Lombardi and M.P. McCahill, "User Interfaces for Self and Others in Croquet Learning Spaces", *Proceedings of the Third Conference on Creating, Connecting and Collaborating through Computing (C5'05*), IEEE Computer Society Press, 2005.

[8] Jabber Software Foundation: http://www.jabber.org/

[9] M.P. McCahill, P. Moore, L. Wendland and A. Zampagna, "Extending Croquet Spaces with Vector Fields, Vehicles, and Virtual Presence" *Proceedings of the Fourth Conference on Creating, Connecting and Collaborating through Computing,* IEEE Computer Society Press, 2006.

0-7695-2806-6/07 $20.00 © 2007 IEEE