THE USER KNOWS BEST: REFINING A VR INTERACTIVE FICTION PROJECT

BY JOSEPHINE R. ANSTEY

Assistant Professor
Department of Media Studies: University at Buffalo

Introduction

Creating interactive work is an iterative process. The artist/programmer needs to solicit user feedback and observe users in the system in order to check whether the goals of the project are being realized and to refine the interaction. This paper is a case study of this process as it unfolded during the development of a virtual reality, interactive fiction piece, *The Thing Growing*, built at the Electronic Visualization Laboratory (EVL), University of Illinois at Chicago, between 1997-2000. The goal of *The Thing Growing* is to create an interactive story in which the user is the main protagonist and engages at an emotional level with a virtual character, the *Thing*.

The Thing Growing was originally designed for the CAVE(R) or ImmersaDesk(TM), projection-based, virtual reality systems. [1] These systems include a wand equipped with a joystick for navigation and three buttons for interaction, and tracking sensors that are attached to the user's body to feed information about her position and orientation to the computer. Typically in the CAVE the user's head and one hand is tracked. We added a tracker for the other hand. In projected VR systems, like the CAVE, the user sees her own body in the 3D environment.

Earlier Versions of The Thing Growing

The Thing Growing originated as a short story I wrote aimed at exploring a protagonist's emotions in a relationship with a clinging, demanding but loved other, an entity called the *Thing*. The protagonist is physically invaded by the *Thing*, yet grows emotionally attached to it. The story ends with the protagonist having the option of becoming detached from the *Thin*g - a process which would kill it - or living with it forever. On my first attempt at putting this story into VR, I made the *Thing* very amorphous, a collection of spheres that swirled around the user. I had imagined that I could convey the story just with the physical movement of this *Thing* and some sounds. It quickly became apparent that nothing of the story I wanted to tell was conveyed to the users unless I stood behind explaining what was happening. Some users never understood that the spheres were meant to be a discrete thinking and emotional entity. Feedback from this first experiment suggested that the experience had to be much more obvious if I wanted to convey my story, and that the *Thing* had to be more humanoid

The Thing Growing: initial plans

In Fall 1997, as I started work on the current version, I was therefore determined to create an experience that would be self-explanatory; an application smart enough to convey the story *and* explain to the user how they should interact with it. For this version I collaborated with Dave Pape, using XP (2) an authoring system for VR that he was developing at EVL. Among other things, the basic XP system took care of loading and transforming models; collision detection; navigation; triggering events; integrating sound and switching individual objects or whole scenes on and off. With Dave's help I extended the system to create the intelligence for the Thing, and the story environment as a whole.

I story-boarded a simple linear three act story to serve as the backbone of the experience. All the interactivity and any alternatives that were generated by the user's input were

organized around this core. In Act One the user finds a box in a shed. She opens the box. The box and shed explode outwards as rocks pour out of it. Then the *Thing* emerges from the bottom of the box, ecstatic to be free. It looks at the user and falls in love with her. In Act Two the *Thing* insists on teaching the use a dance that will express (in its eyes) the relationship between them. It becomes disenchanted with the user's attempts to dance and goes of in a huff. The rocks stalk and trap the user. The *Thing* frees her. The dancing begins anew. In Act Three, the *Thing's* four cousins interrupt the courtship, disgusted with the relationship they perceive between *Thing* and user. They imprison the two and threaten to kill them. The *Thing* passes a hidden gun to the user and implores her to shoot the cousins. When all the cousins are dead or have escaped, the *Thing* suddenly fears the user will shoot it. The user has the choice.

In this version the *Thing* has a body composed of multi-colored, semi-transparent, pyramids, one for the head, one for each arm, and several for the body and tail. Motion tracking is used to animate these pyramids and a strong illusion of a humanoid being with life-like gestures results. The *Thing* also has a voice. The voice became one of the most important tools in creating the experience. The *Thing's* speeches are used to tell the story and to interact with the user; the tone of the voice communicates the *Thing's* changes of mood; and we use these emotional changes to stimulate the user's emotions.

First Steps and First Problem

We started by building Act One, and the dance sequence for Act Two. The main problem was to create the *Thing* as a seemingly intelligent, autonomous creature, able to react quickly to the user. Based on the story board, we recorded hundreds of phrases for its voice. At times these phrases are in a scripted sequence - for example when it is freed from the box it is trapped in. But mostly it speaks in response to the user - for example when it is teaching the user to dance. For the dance section we recorded different versions of words of praise, encouragement, criticism, and explanation. We also recorded the different types of utterance in different moods; happy, manic, whiny and angry. Each phrase lasts a few seconds and is linked to a motion-captured gesture. The phrase and gesture together make up an *action*. We built up a library of such actions, and then needed to build a brain for the *Thing*, which would determine when to use each action. The job of the brain was to select an appropriate action, according to the point in the narrative; the user's actions; and the *Thing*'s mood. We had to build a structure of rules for the brain to follow in determining its selection.

The narrative became a very useful tool for constraining the kind of action the brain can pick, thus simplifying the rule structure we had to build. For example when the *Thing* is attempting to teach the user to dance, it has a basic routine to follow:

- demonstrate a dance step
- observe user dancing or dance with user
- check whether user is dancing correctly
- praise or criticize user
- repeat dance step or teach new step

At the end of each action, the Brain looks at this routine to see what type of action comes next.

It was fairly simple to implement this basic routine, but one of the steps was more difficult to program - *check whether user is dancing correctly*. Since information comes in from the trackers about the user's position and orientation, it is possible to check whether the movements of the user's arms, body and head correspond to the movements of the *Thing's* arms, body and head. I imagined that to check the accuracy of the user's

movement I would have to divide the CAVE space into a 3D grid and continually check which box in the grid the user's arms and head were in. I would have to build up a series of check points for each of the dance actions that would correspond to the user moving their arms in and out of the boxes of the grid. This solution seemed complicated and a lot of work as I would need an individualized checking routine for every dance step.

Networked Thing

We were scheduled to show the first part of the project at SIGGRAPH 98. However, as the show approached, I was still unsure how to build an effective checking system. EVL was bringing a networked VR system to SIGGRAPH consisting of two ImmersaDesks. We decided to circumvent the checking problem by building a networked version of the project. Someone at the second ImmersaDesk would be an invisible voyeur on the scene between the Thing and an avatar of the participant, watching to see how well the participant was dancing. This networked user had a menu on his screen and used it to tell the *Thing* if the participant was dancing well or not, and also to control its moods. In this scenario, although the *Thing* had its inbuilt routines of behavior it was also getting help from a much more flexible intelligence system - with a wealth of fine-tuned interactive experience!

We ran with this "Wizard of Oz" brain for about nine months. During this period of time we observed both the users and ourselves and noted that both fell into very standard patterns of behavior. As the Wizard of Oz we fell into typical ways of altering the *Thing's* moods. The dancing interaction lasts for 2-3 minutes. The finale is the Thing running off in a huff. Essentially the mood always changes from good to bad over time. If the user is uncooperative - refuses to dance, runs away a lot - the moods becomes whiny or angry quicker. If the users are very good at dancing, they get more over-the-top praise from the thing in manic mood but eventually the *Thing* turns on them too. Users also had standard ways of reacting to the *Thing*. They either tried to obey it - or refused to dance and tried to get way from it. Those that tried to dance, varied widely between people who would copy exactly and those too shy to move very freely. As the Wizard of Oz we tended to treat these alike to encourage the timid.

We decided that we could take advantage of what we had observed and make the portion of the *Thing's* intelligence dedicated to checking much simpler than originally planned. The checker we implemented assumes any arm movement that travels more than an arbitrary minimum distance at times when the user is meant to dance is an attempt to dance and therefore counts as "dancing correctly". Over time the *Thing* is programmed to become randomly pickier, criticizing and rejecting movements that earlier it would have OK'd. The *Thing's* mood changes are also based on time and the level of user co-operation. In the end it will always get mad. But it will get whiny or angry much more quickly with users who won't dance with it. Users that move with more sweeping gestures - whose arms travel a greater distance over time - are considered more co-operative and its more likely to praise them lavishly.

For us the process of using a Wizard of Oz brain as a test bed before programming the more complicated parts of the intelligence system happened by chance, however I think that it points to an important lesson: fake the intelligence before you build it, so that you pinpoint where it is needed and what is needed.

First Draft Problems

By the summer of 1999 we had a first version of the complete story working. However, feedback from colleagues indicated several major problems that needed to be addressed for the project to successfully communicate its story. Watching users interact with the

project also revealed what was not working, and helped us to identify key areas where we needed to get information across with more punch and clarity - whether it was information about the story, or information about the interaction. Some of these problems had existed from the very beginning, now was the time to address them.

Dancing Scene Problems

First, some users had difficulty discerning the *Thing's* body. It is a collection of detached pyramids "tied" together by the motion captured movement. We expected people to move their arms in the same way that the *Thing* moved it's arms. Some users, however, simply wriggled about like hypnotized snakes. After observing a number of these users, we realized that they were trying to copy the *Thing's* tail movements. Adding a verbal injunction to "move your arms and body like mine - don't worry about the tail," did not fully fix the problem. Dan Sandin suggested outlining the *Thing's* body and de-emphasizing the tail. We also added eyes to the head to make it more obviously a head. This new look helped people to correctly identify the *Things's* body parts and move their own bodies' accordingly.

A second and related problem was to clearly indicate to people that they had to move their *bodies* during the dancing sequence, and that the *Thing* "knew" what they were doing. Some users, especially expert users, thought that they should use the joystick and buttons to dance - this resulted in the *Thing* rushing after them and chiding them for moving away. They would then try to use the joystick to drive close to the *Thing*, which it would continue to interpret as "driving away" behavior. Confusion would ensue with the user muttering, 'I'm just trying to get close." Other users would only dance with one arm - although in the final version both arms are being tracked.

To solve these problems we inserted some new behavior at the very beginning of the dance sequence. The *Thing* announces that it will show the user how to dance, then says that they must "loosen up a little first." It instructs the user to wave her arms above her head. The program performs checks to see :

1. If the joystick or buttons are pressed - if so the user is instructed not to touch the buttons but to move only the body

2. If the user is moving both arms - if not the user is instructed to move both arms

3. If the user is trying hard enough, i.e.. are they waving there arms high enough - if so the *Thing* proceeds to teach the first dance step

This new sequence establishes at the outset the rules for the dancing and that the *Thing* knows what the user is doing.

Content Problems

Other problems were more to do with content than technical issues. In the first draft the *Thing* was simply too unpleasant and people nearly always shot it as soon as they were given the chance. Ideally we wanted the user to be more ambivalent. There was also a timing problem. In the second act the *Thing* danced with the user, then stormed off in a huff. The rocks on the plain then chased and trapped the user. The user had barely been released from that trap when the third act was triggered and the user was again trapped while the four cousins berated the *Thing*. This period of inactivity (in-interactivity) was too long and many users' attention wandered. The transition between the second and third acts was also too abrupt - people were confused about what had happened and did not understand who the cousins were. Some users believed that they had come to save them

from the *Thing*, rather than feeling that they were now on the same side as the *Thing* and in deep trouble.

The solutions for these problems were interdependent, but I will describe them as they happen chronologically in the story. First we added a new behavior in which the *Thing* now copies the user's movements and inserted this after the user is released from the rock. The *Thing* says, "Let's do something different - you dance and I'll copy." The user discovers that as she moves her arms and head the *Thing* moves with her (Basically we took the tracking data from the user and applied it to the *Thing's* body with a slight delay.) It is strangely flattering - because the *Thing* becomes a mirror of one's own idiosyncratic movement. The *Thing* becomes very seductive while this is going on and the user feels that she is finally in control. So this small new scene, lasting about 40 seconds serves two purposes. It makes the user "like" the *Thing* more - we like the thing that does what we want - and it breaks up the two periods where the user is trapped and unable to interact.

Second we added a more elaborate transition scene between the second and third acts. The user is now happily dancing with the *Thing* copying her. Suddenly lightning crashes across the sky, the sky darkens and a god-like voice out of nowhere booms, "What's going on here?" Another higher voice cries, "We must stop this evil!" A lightning bolt flashes towards the user, cracks appear in the ground and the user and the *Thing* fall through a red trench into a world below. This is the territory of the four cousins.

Third we revised the *Thing's* speeches and the four cousins behavior throughout the third act to clarify the back story- the taboo relationship which user and *Thing* have been caught in. Now as the four cousins approach, the *Thing* whispers asides that reveal that the dancing behavior that they were indulging in is heresy if it takes place between a meat-object, the user, and a *Thing*. As before the four cousins berate the *Thing*, but now they are more vicious and pull off its arm. This is to encourage the user to feel more protective of the *Thing*.

During the sequence where the user shoots at the cousins, all the *Thing's* phrases were revised. Many of them give more information about the back story. The *Thing's* mood and behavior also alter more radically depending on the user. If the user can't or won't shoot the cousins, the *Thing* begs her to protect it, and finally becomes furious and abusive. On the other hand if the user does shoot the cousins, and the faster she shoots, the *Thing* begins to have second thoughts and begs the user to spare her family members. All this is designed to increase the user's ambivalence.

Adding the Intro Scene

Finally, we added a scene at the beginning that served to introduce the user to the system and teach her how to use it. A human helper is needed to help the user to put on the tracked glasses and get hold of the wand, but then the helper steps away. The user finds herself standing in front of an archway with a billboard in it which has a picture of the wand on it. The wands three buttons and joystick are identified in the picture. The billboard's text reads, "Press the left button to begin." When the user pushes the left button, the billboard slides away to reveal the Title of the Piece and names of the collaborators. Those too slide away and arrows appear on the ground. The user is told to use the joystick and follow the arrows. The arrows snake through the arch and across the ground and finally point through another arch into blackness. As the user drives through this arch, she drives into Act One. The introduction is designed to give the user the expectation that the system will tell her how to operate it, and to ensure that her focus is on the Thing and her relationship with it.

These fixes comprised a second draft of the project. Feedback from users, and observing

users in the application, suggested that this draft communicated the story much more successfully and that the users were much more comfortable with the interface.

Conclusion

User testing consistently pointed out weaknesses in both the form and content of this interactive fiction. Some problems had to be addressed two or three times, with successive revisions of the code, before they were solved. Every time we showed the piece we tried to leave users alone with the experience and not jump in to help them if they became confused. This strategy meant that it became painfully apparent when the instructions for the interaction - which were embedded in the narrative of the story - were not working, and made us realize it was crucial to demark any changes in the interaction. In this piece the user navigates through the VR world with the joystick, but is also expected to move her body when dancing with the thing. Communicating this change in interactive mode clearly and concisely was probably the most persistent development problem.

Remarks that the users spontaneously made to the Thing during the experience showed if they were immersed in and following the story, however, it was also important to question them to see if the points I wanted to make were getting across. (Because the story was so obvious to me, I was often surprised by the different interpretations that people came up with.) When these interpretations obscured the story we wanted to tell we made changes designed to clarify the story-line and to elicit from the user the kind of emotions that we wanted.

It was particularly interesting to discover how important the transitional sections were for the experience. Users became disoriented by abrupt scene shifts and it was important to signal or foreshadow an impending change in some way. An example of this comes between the Introductory Act and Act One - the user sees the arrows pointing through an arch into a black void - they don't know what is over the threshold but the threshold itself signifies a change. Another example is the transition from Act Two to Act Three - a foreshadowing of the four cousins power and disapproval was added at the end of Act Two to ready the user for the events of Act Three. Every single transition point in the story - in terms of the narrative and the interaction - had to be worked and reworked before it successfully did its job.

The general lesson of this experience was "Don't be afraid to be obvious." As VR becomes more ubiquitous, people will become more comfortable with it, but at present most people do not know what to expect, and what is expected of them. In order for the experience to win out over the discomfort with the technology, the technology has to explain itself and reassure the user; and the interaction has to be easy to grasp. Then the story can draw the user into a space that is playful but powerful, and which depends for its meaning on the amount the user is prepared to give of herself.

References

1. Cruz-Neira, C., Sandin, D.J., DeFanti, T.A., Kenyon, R.V., and Hart, J.C., The CAVE: Audio Visual Experience Automatic Virtual Environment, Communications of the ACM, Vol. 35, No. 6, June 1992, pp. 65-72.

2. Pape, D., Imai, T., Anstey, J., Roussou, M., DeFanti, T., XP: An Authoring System for Immersive Art Exhibitions, in the Proceedings of Fourth International Conference on Virtual Systems and Multimedia, Gifu, Japan, Nov 18-20, 1998

Acknowledgments

The Thing Growing Project was created at the Electronic Visualization Laboratory, University of Illinois at Chicago.