Squidball: An Experiment in Large-Scale Motion Capture and Game Design

Christoph Bregler, Clothilde Castiglia, Jessica DeVincezo, Roger Luke DuBois, Kevin Feeley, Tom Igoe, Jonathan Meyer, Michael Naimark, Alexandru Postelnicu, Michael Rabinovich, Sally Rosenthal, Katie Salen, Jeremi Sudol, Bo Wright

http://Squidball.net

Abstract. This paper describes Squidball, a new large-scale motion capture based game. It was tested on up to 4000 player audiences last summer at SIGGRAPH 2004. It required the construction of the world's largest motion capture space at the time, and many other challenges in technology, production, game play, and study of group behavior. Our aim was to entertain the SIGGRAPH Electronic Theater audience with a cooperative and energetic game that is played by the entire audience together, controlling real-time graphics and audio by bouncing and batting multiple large helium-filled balloons across the entire theater space. We detail in this paper the lessons learned.

Fig. 1. Electronic Theater audience playing Squidball

1 Introduction

Squidball is a large-scale, real-time interactive game that uses motion capture technology and computer graphics to create a unique and energetic experience for mass audiences. Using the motion capture volume with participating player audiences of up to 4,000 people, the game debuted on August 12th, 2004, at the Los Angeles Convention Center as pre-show entertainment for the SIGGRAPH Electronic Theater.

This paper describes the design criteria and technology behind this venture. It also explores the adventures and challenges that the production team had to overcome and the lessons learned.

SIGGRAPH audiences experienced a similar interactive Electronic Theater pre-show over a decade ago when the Cinematrix System was introduced in 1991 [2]. Cinematrix was an interactive entertainment system that allowed members in the audience to control an onscreen game using red and green reflective paddles. Other interactive entertainment systems have been tested on audiences in the hundreds to thousands, described in greater detail in section 2.

The success of Cinematrix was the original inspiration for our work, and we initiated our project to bring back this style of pre-show entertainment to SIGGRAPH. Although the 2004 Electronic Theater was our first public test, we envision *Squidball* being deployed in other large audience events for entertainment, social studies, team building exercises and other potential applications.

Developing and testing such a system was a very unique and high-risk venture with many challenges. Other games, graphics and interactive systems are usually designed for single user or a small group, and go through several test cycles. However, for the Squidball project, we were dealing with many factors orders of magnitude larger than standard environments, including a gathering of 4000 people, the construction of a system using a $240 \times 240 \times 40$ feet motion capture volume and a huge screen. Furthermore, the system had to work the first time, without the benefit of any full-scale testing.

During our initial brain-storming sessions, we decided to create a game that is played by bouncing and batting a number of balls that would be used as wireless joystick/mouse inputs to a game, across the entire audience. We also decided to track the balls using 3D motion capture technology and to use this data to drive real-time graphics and an audio engine.

We set out to design a game that:

- Requires no explanation of the rules – people must be able to pick it up and start playing;
- Is even more fun than just hitting a beach ball around an auditorium (we already know this is fun);
- Is fundamentally about motion-capture and takes full advantage of the capabilities of this technology;
- Can be played by 4,000 people simultaneously using a small number of input devices;
- Can be played by people standing, sitting or holding a beer in one hand (there was a cash bar in the Electronic Theater); and
- Involves people hitting balls AND looking at a projection screen.

2 Related Work

As mentioned above, the inspiration for Squidball came from the Cinematrix system [2] shown at the SIGGRAPH 1991 Electronic Theater and several other events. With this game, every audience member had red and green reflective paddles that control on-screen games, including a voting system, Pong, and a Flight Simulator. In the voting schema, the system counted how many red vs. green paddles were shown. In Pong, the left side of the audience played against the right side, and the position of the paddle controlled the ratio of red and green paddles on each side of the audience. It was surprising how quickly the audience learned to control the games and to jointly coordinate the mix between red and green paddles. Of course, the yelling and excitement of a large audience was also part of the show. Another set of similar interactive techniques was studied at student theater screenings at CMU [4]. Several computer-vision-based input techniques were tested on large audiences. The input technique most closely related to Squidball is the 2D beach ball shadow tracking, where the location of the shadow could be used as a cursor in several 2D games. D'CuCKOO (a music band that uses various kinds of new technological instruments) designed a gigantic beach ball that creates music as the audience bats it around. The MIDI-Ball [1], a wireless 5-foot sphere, converts radio signals into MIDI commands that trigger audio samples and real-time 3-D graphics with every blow. There have been other systems reported, that track small groups of people as they perform interactive music and dance activities [6], or are used for home video games [3] but none of them have been tested on thousands of players.

3 Large-Scale Motion Capture

Here we describe the challenges and experiments of building a large-scale motion capture space and how this ties into the Squidball game engine and game testing. In section 4 we describe additional details of the game design.

Our target venue was Hall K in the Los Angeles Convention Center, which was converted into a 4000-seat presentation environment to screen the Electronic Theater for the 2004 SIGGRAPH conference. The total space was 240×240 feet. We needed to build a motion capture volume that covered the entire seating area and allowed enough height above it to throw the balls up in the air: a capture volume of $190 \times 180 \times 40$ feet. To the best of our knowledge, no motion capture space of this size had been built before. One of the larger reported spaces was built for the Nike commercial by Motion Analysis Corp and Digital Domain [5]. It had dimensions of $50 \times 50 \times 10$ feet. It used 50 cameras to track six football players.

One of our design constraints was tracking multiple (up to 20) balls in 3D in real-time. We had a Vicon motion capture system [7] with 22 MCAM2 cameras, and each camera had a field of view of 60 degrees (12.5mm lens) and 1280×1024 pixel resolution. In its intended use, the system can track standard motion capture markers (0.5 inch) in a capture distance of up to 25 feet. The markers are made of retro-reflective material. Visible light illuminators placed around the camera lens shine light out, and almost all light energy is reflected back into the camera. This makes the retro-reflective markers appear significantly brighter than any other object in the camera view, and image processing (thresholding and circle fitting) is used to track those markers in each view. Triangulation of multiple camera views results in very accurate and robust 3D marker tracking. (We are currently also considering vision-based techniques on non-retro-reflective objects for future game experiments.)

We determined that the only way to utilize the Vicon motion capture system in a significantly larger space with the same number of cameras was to scale up each aspect of the system. The cameras' view scale up in an approximately a linear fashion; in other words, a marker 100 times larger in diameter and 100 times further away looks the same to the camera. Of course, because light intensity falls off with the square of the distance traveled, much greater illumination is necessary. With experimentation, we found that halogen stage lighting provided sufficient illumination for the Vicon tracker.

Three other challenges in scaling up the system were 1) producing the larger markers, 2) dealing with camera placement constraints, and 3) calibrating the space. All of them appear simple in theory but, in practice, these became critical production issues.

The Lemmon The Tomato The Orange The Orange Reflecting

Fig. 2. The evolution of our retro-reflective balls.

3.1 How to produce the balls (markers)

The motion capture system requires spherical markers for effective tracking. Normal marker sizes are 0.5 inch in diameter, which are covered with 3M retroreflective tape. In order to stage our event in a radically larger-than-normal space, with larger-than-normal camera view distances (max. 250 ft.) and using the balls as real-time inputs, we had to increase the size of those markers significantly. We determined that a 16-inch diameter marker was the smallest marker that could be robustly detected at 250 feet. In the final game, for dramatic effect and game-play, we opted for larger markers: 8-foot chloroprene bladders (weather balloons). In order to achieve the right “bounciness”, we under-inflated the balloons.

Each marker requires a retroreflective coating in order to be tracked by the Vicon motion capture system. Experiments using retroreflective spray-paint failed. The reflective intensity of the sprayed paint was 70% to 90% less than 3M retroreflective tape. After dozens of tests with tapes, and fabrics of varying color, reflectivity and weight, we settled on specific 3M retroreflective fabric (model # 8910). Figure 2 shows the results of these tests. The first test (the ”lemon”), the second iteration (the ”tomato”), and the final version (the ”orange”), which ultimately produced a perfectly round spherical shape.

In order to achieve a perfect spherical shape and to spread the force evenly throughout the surface, the fabric was cut on the bias, in panels like those of a beach ball. At this large scale, any of these shapes were adequate for the Vicon system to track them. The advantages of a perfect sphere were both aesthetic and functional. With the force evenly distributed, one spot is no more likely to rip the fabric than any other spot. Similarly, hitting the ball anywhere has the same predictable result. The balls were inflated with helium to reduce their weight. Because the fabric was heavy, they did not float away when filled with helium, fortunately. In future generations of this game, we plan further reduce the weight of the balls in using a different material or smaller sized balls.

3.2 Camera Placement

Standard camera placement for a motion capture system is an “iterative refinement process” dependent on several site-specific aspects. For standard motion capture, cameras are usually placed on a rectangle around the ceiling, all facing into the capture space. Sample motion capture markers are distributed over the capture space, and cameras are adjusted so that each marker is seen by as many as possible cameras from as many as possible directions. Additionally, the tracking software is checked for each camera during placement.

In our scaled-up system, camera placement was a significant challenge. We could not afford as many trial-and-error cycles in camera placement that would be possible in standard-sized motion capture labs since our time was limited in the final space and each adjustment took a significant amount of time. Other logistical constraints affecting camera adjustment included: A) cooperating with the Union LACC workers schedule to get access to the ceiling and catwalks, B) coordinating between people on the 40-foot high ceiling and people on the floor up through radio-communication for each re-mount and re-alignment of a camera, C) getting live feedback from the Vicon PC station in the control booth to people on the ceiling so they could see the effects of their adjustments, D) camera view limitations - the 60 degree wide-angle lenses did not actually see a full 60 degree angle of view; even with the extra heavy studio lights mounted next to the cameras the visibility of the weather balloons dropped off after 250 feet in the center (and at even shorter distances at the perimeter of the camera view), and E) scale issues - moving balls on the ground takes much longer because of their large size and the distance to be covered. In a standard mo-cap studio, you pick up a marker and lay it down a few seconds later; in this space, we had to move a shopping cart with a ball or drive an electric car across the hall.

3D Simulation in Maya In anticipation of all those problems, we designed a 3D model in Maya for all the target spaces, including one for the campus theater (our first test), one for the campus sports center (our second, third and fourth test), and one for Hall K at the LACC (Figure 3), our final show. This final model was derived from blueprints we obtained from the building maintenance team.

We also built a 3D model of the “visibility area” to determine the sight lines of the cameras. We determined that the cameras could not “see” of center at distances of 250 feet. The further out we moved the balls, as shorter the visibility became. Based on experimentation, we built a 3D Maya model for the camera visibility volume (Figure 3 green). This volume was then used in our Maya building model to simulate several camera placement alternatives. Our goal was that each point in the capture volume should be seen at least by 3 cameras, given the constraints on the lengths of video cables. The final configuration we used was pretty close to our simulation. We settled on mounting evenly all 22 cameras around the left and right catwalk, and along the back-end catwalk and the center catwalk, but not the frontal catwalk. We didn’t want to mount any cameras and high-intenstiy lights above the screen, so the audience would not be distracted. (Consult our website to see the final camera-placement in detail).

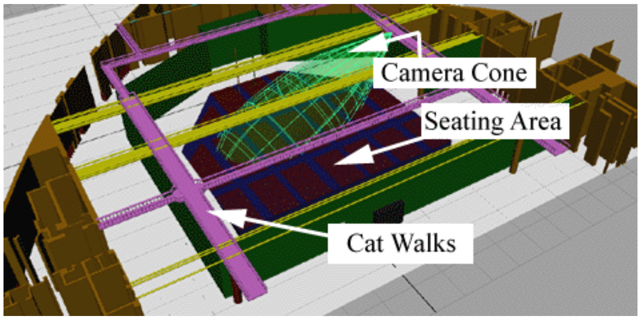

Fig. 3. This shows our Maya model of Hall K and the visibility cone (green) for one of the cameras.

Mounting and Networking We knew we had to set up the system in LACC very quickly, so we ran multiple practice sessions for camera mounting in New York, first for 10 cameras and then 22 cameras.

The setup required careful cabling. Vicon sends camera data first through analog wires to a Datastation, which thresholds video frames and compresses the resulting binary images. The Datastation then sends all 22 Video Streams at 120Hz over gigabit Ethernet to the Vicon PC, which does the real-time 3D tracking. In order to have the shortest possible video cable length, the Datastation had to be close to the cameras. In Hall K at LACC, the Datastation was placed 40 feet above the audience on one of the catwalks. The compressed video data was then sent via gigabit Ethernet down to the "control booth" in the back of the audience on the floor. The control booth contained all the workstations, including the real-time 3D tracker and the game system.

During camera placement we operated the Vicon PC in the control booth through a wireless laptop and Remote Desktop. This allowed us to walk from camera to camera on the catwalk, and do all adjustments, while remotely monitoring what the camera "sees" and how the tracking software performs.

3.3 Calibrating the space

The final challenge for the motion capture setup was camera calibration. Using the Vicon software, the calibration process in a standard space is done by waving a calibration object throughout the entire capture volume. Usually, this is a T-shaped wand that has 2 or 3 retroreflective markers placed on a straight line. (Figure 4) The 2D-tracking data for the calibration object from each camera is then used to compute the exact 3D locations, directions and lens properties of the cameras. This is called the calibration data, which is crucial for accurate 3D tracking.

Of course, the standard T-wand would not be seen by any camera in such a large target space (below pixel resolution). We determined that a 16-inch marker was the smallest marker that could reliably be seen and tracked from 250 feet. To overcome this, we built several "calibration T-wand" versions. Figure 4 shows one version that allowed us to "wave" the calibration object as high as 30 feet. We conducted initial tests on how much time an "exhaustive volume coverage" would take and how physically exhausting it would be using the roof of our lab. In the campus theater space and the campus sports center, we either walked the wand around holding it at several heights or skate-boarded through the space. In the final test at Hall K, we first used a crane and ropes. Ultimately, we ended up using a T-wand constructed out of light weight bamboo sticks lashed together using a traditional Japanese method and then drove that around at several heights on an electric car. A calibration run took around 30 minutes. Tensions ran high during calibration in Hall K, but the process was a success (see video on squidball.net) and we had a spot-on reading of Hall K right up to the periphery of the seating areas. In actuality, we were able to track the balls beyond the boundaries of the game "board".

4 Software Integration of Motion Capture and Game Engine

Before we could start the various game tests, we needed a rapid prototyping environment connected with the real-time motion capture input, one that generated real-time computer graphics and sound effects. The real-time visualization system and game engine were written using the

Fig. 4. Left: The standard sized calibration objects of length 15 inch. Right: Large 15 feet high wand.

Fig. 5. Calibration in Hall K using a crane.

Max/MSP/Jitter development environment distributed by Cycling74. The system consisted of five main components:

- A TCP socket communication system, which distributed motion capture data from the Vicon system to the tracking, game-engine and audio subsystem in real-time.
- A real-time tracking module that would take the raw motion capture data, filter it using Kalman filters and then extract useful metrics such as object velocity and collision detection. (This was necessary to produce consistent sound effects with the balls flying high up in the air).
- A game engine written partly in Java and partly as Max patches, which drove the game simulation. Submodules of this system included components that handled the basic game narrative, the media files, and the interface to the graphics engine. The graphics engine used Jitter to render the game using OpenGL commands.
- An audio subsystem resident as a Max patch on a second computer (and receiving forwarded motion capture information as well as scene control from the main Max computer). It produced sound for collision, bounces, flying noises, etc.

The prototyping environment proved robust and fast enough to use in the final show, and we continued to tweak the game until the night before the opening.

In the production mode, we ran two duplicate sets of the game and audio computers for redundancy, with a switch, but fortunately this was never needed. The system required five human operators during the shows one on sound, one on the game system, one monitoring the Vicon PC, one watching the audience, and a master show controller who coordinated the team.

5 Large-Scale Game Design

We wanted to design a game that worked well in the poorly-understood dynamics of cooperation and competition of a large-scale group, but we knew we had a limited number of markers that we could track. We also understood that however we used the markers, the game had to run well with no full-scale testing. We wanted everyone to "win the game" as a single body, not as many small groups. This meant we had to discourage degenerate strategies.

We decided that the rules of the game should be discovered on the fly as people played, rather than through an instruction sheet. Finally, we wanted to create a game that was more rewarding than simply bouncing balls around in a space.

Playtesting was a major challenge we faced. Gathering a 4000-player audiences is difficult and expensive, so we had limited opportunities to test the game at full-scale. This led to a number of creative solutions in testing. Though we were able to use spaces of approximately the same size as the final space for testing, we were not able to test with a full-scale audience. We came up with a number of innovations, including clumping groups in various locations in the testing space and organizing the clumps strategically to make them appear to be a larger audience. However, the first test of the game with a full audience was not until its premiere at SIGGRAPH.

After many discussions and iterations with the prototyping system, we settled on the game rules described below.

Fig. 6. Example Screen shot of Squidball game. Please see video (on squidball.net)for game in action.

The Game: The rules for the game were simple, and had to be discovered by participants through gameplay. The twelve weather balloons in physical space were represented within the digital game space as green spheres on the screen. Players moved the weather balloons around the auditorium (whose space corresponded to a 3D space onscreen), in order to destroy changing grids populated by 3D target spheres. The game had 3 levels of increasing complexity, and each level could be replayed 3 times before a loss condition was reached. The second level introduced an element of time pressure, so players had to complete the game challenge within the allotted period of time. This was the level that really taught players how the game worked, as most audiences failed to clear the level on the first try.

Through repetition and the existence of a loss condition, the players eventually discovered the victory condition, as well as the correspondence between the weather balloons and their representation within the virtual game space. In the third level, players worked to clear special colored spheres that, when activated, revealed a composite image. Players quickly discovered a range of social strategies that emerged from their physical proximity with other players; in each instance of the game (the game was played 6 times over the course of 4 days) the 4,000 or so players came together organically to collaborate in the play of the game as they discovered what the gameplay required of them. It was an interesting first step in designing a kind of game that was extremely simple in its rules and interaction but extremely complex in the forms of social dynamics it spawned.

6 Gameplay in Practice

In the control both, we had a control which we could adjust to alter the sensitivity of the game during play. Turning up the sensitivity made the virtual target sizes larger and therefore gameplay easier. Turning down the sensitivity made the virtual target sizes smaller and gameplay harder.

Fig. 7. Squidball Gamers

For good gameplay, we felt it was essential that the players be able to make mistakes and learn from them. So, by default, we set the sensitivity fairly low. However, in some circumstances we increased the sensitivity to temporarily make the gameplay easier, giving the audience a little ”boost”. A person in the control booth was responsible for watching the progress of the game and making these ”group mind” decisions regarding when to adjust gameplay. We believe that similar controls were included in Cinematrix

Even with a sensitivity control, there were some issues. One problem was that the audience at the start of show was less than a full house, which we had not anticipated in our game design. Because of this, some of the game levels proved hard to clear, because virtual targets were located in places where few audience members could reach them. This problem could be addressed by creating multiple configurations for different audience sizes.

A second issue was uneven audience distribution. Some people were in sparse sections of the audience, and did not get to participate as much as others. To address this concern, we enlisted student helpers to help move balls around.

A third issue was that, during the game, people had to divide their attention between the screen and the balls. Some people decided to only watch the screen. Others ignored the screen and simply pushed the balls towards the center of the room. Initially, a relatively small percentage of ”aware” players actually watched both and drove the gameplay forward. The number of ”aware” players increased dramatically towards the end of the game, demonstrating that the game design principles were working. However, the split attention issue remains a challenge for any game design involving thousands of people, balls and a single screen.

One solution might be to place multiple screens on all sides of the audience. However, this introduces another difficulty: coordination. Even with a single screen, players had difficulty coordinating the balls and target locations. The problem is that the player must face one direction, look at a screen over their shoulder, and then punch a ball in a third direction towards a target. Since few people have much practice at this activity, balls were popped left when they should have been popped right, or forward rather than back. Adding more screens would confound this issue. The problem of having to watch the balls and the screen is fundamental. Originally we also considered audio-only games, or games in which balls hitting each other is the point. We plan to reconsider those ideas in future experiments, and more detailed evaluations of the audience interaction.

7 Conclusions

After all the hard work to create and setup Squidball for SIGGRAPH 2004, the roar of the crowd at the end of each a level was gratifying validation of our efforts.

Since Squidball, we have been discussing possible iterations for future games. One option we have discussed is to use spotlights shone onto the crowd as targets, rather than using targets on a virtual screen. This would address some of the gameplay issues we encountered. The audience would have a physical cue showing where they are trying to get the balls to, rather than a virtual cue shown on a screen over their shoulder. Using spotlights, it would be possible to create roving patterns, enabling the spotlights to be moved in a pattern which ensures that everyone gets a chance to participate, taking into account the audience density and distribution. Of course, spotlights introduce a whole new set of technical challenges, though none that are insurmountable. We are considering this and other game design changes for Squidball 2.

Acknowlegments This work has been partially supported by NYU, NSF, ACM SIGGRAPH, Vicon Motion Systems Ltd, Apple Computer Inc, Advanced Micro Devices Inc, Alienware, Cycling 74, David Rokeby: very nervous systems, NVIDIA Corp., Segway Los Angeles. The audio was supplied for “Insert Coin” and ”Re-Atair” by Skott, “Superkid” by Max Nix, and “Watson Songs” by The Jesse Styles 3000. Furthermore we would like to thank especially our friends from AVW - TELAV, specificially Jim Irwin, Gary Clark, John Kennedy, Tom Popielski, Gerry Lusk, Mark Podany and Mike Gilstrap. Without their production efforts this would have been impossible. Also special thanks to Debbi Baum, Robb Bifano, Alyssa Lees, Jared Silver, Lorenzo Torresani, Gene Alexander, Boo Wong, Damon Ciarrelli, Jason Hunter, Gloaria Sed, Toe Morris, Scott Fitzgerald, Ted Warburton, Chris Ross, Eric Zimmerman, Cindy Stark, Brian Mecca, Carl Villanueva, and Pete Wexler for all their great help. And the SIGGRAPH chair Dena Slothower, who let us use their venue for the first big test.

References

1. BLAINE, T. 2000. The outer limits: A survey of unconventional musical input devices. In *Electronic Musician*.
2. CARPENTER, L., 1993. Video imaging method and apparatus for audience participation. US Patent #5210604, #5365266.
3. FREEMAN, W. T., TANAKA, K., OHTA, J., AND K.KYUMA. 1996. Computer vision for computer games. In *2nd International Conference on Automatic Face and Gesture Recognition, Killington, VT, USA, IEEE*.
4. MAYENES-AMINZADE, D., PAUSCH, R., AND SEITZ, S. 2002. Techniques for interactive audience participation. In *IEEE Int. Conf. on Multimodal Interfaces, Pittsburgh, Pennsylvania*.
5. MOTIONANALYSISSTUDIOS, 2004. Largest mocap volume. http://www.motionanalysis.com/about_mac/50x50volume.html.
6. ULYATE, R., AND BIANCIARDI, D. 2004. The interactive dance club: Avoiding chaos in a multi participant environment. In *Int. Conf. on New Interfaces for Musical Expression*.
7. VICON, 2004. Sponsor of siggraph electronic theater. http://www.vicon.com.