TOWARDS IMMERSIVE MULTIMODAL GAMEPLAY

Mitchel Benovoy,[1] Mark Zadel,[2] Rafa Absar,[3] Mike Wozniewski,[1] and Jeremy R. Cooperstock[1]
Centre for Interdisciplinary Research in Music Media and Technology
and [1]Centre for Intelligent Machines, [2]Schulich School of Music, [3]School of Information Studies
McGill University
Montreal, Quebec, Canada
{benovoym@cim, rafa.absar@mail, zadel@music, mikewoz@cim, jer@cim}.mcgill.ca

KEYWORDS
Immersiveness, multimodality, projection display, position tracking, gestural input, usability testing

ABSTRACT

We describe a computer game design that employs interface mechanisms fostering a greater sense of player immersion than is typically present in other games. The system uses a large-scale projection display, video-based body position tracking, and bimanual gestural input for interaction. We describe these mechanisms and their implementation in detail, highlighting our user-centered design process. Finally, we describe an experiment comparing our interaction mechanisms with conventional game controllers. Test subjects preferred our interface overall, finding it easier to learn and use.

INTRODUCTION

In game terminology, *immersiveness* is used to describe the degree to which a player feels a virtual environment mimics his or her experience with the real world. The video game industry has seen significant improvements in graphics detail and realism in recent years, but the use of standard displays and controllers continues to limit the degree of player immersion. Game controllers, keyboards and mice fail to exploit the rich capabilities of gestural expression and capture the subtleties of our interaction with the real world. This paper describes the design and implementation of a gaming system adapted to first or third person games, which offers a high sense of player immersion by using large-scale projection displays, multimodal feedback, body tracking, and bimanual gestural control. We demonstrate the capabilities of the system through a game we have designed: *Snowdown*.

Previous research has investigated how immersion is achieved and to what degree it succeeds. Brown and Cairns (2004) hypothesized that the more a game feels real, the greater the sense of immersion the player experiences. Cheng and Cairns (2005) observed the immersion of a player in a game and then introduced inconsistencies in its visual and physical laws. They found that once immersion was achieved, it was surprisingly difficult to break. These findings suggest that the quality of a game experience may be advanced significantly through improvements to the level of immersion.

Our approach to improving immersiveness is based in the belief that a system's user interface is its most important determinant of user experience. Starting from conventional game systems, we make improvements to both the input and output mechanisms. First, Snowdown is designed for a large-scale display, covering much of the player's field of view. Second, body tracking is used for controlling the avatar's position in space, while other actions are accomplished through gestures resembling their real-world equivalents. This engages the user in a highly physical interaction.

These interface designs are described in further detail below, with a focus on the importance of body tracking and gestural interaction. The user testing process is also discussed, and an experimental evaluation of the interface is presented, comparing our interaction model to the paradigm of traditional computer games.

RELATED WORK

The advent of low cost consumer hardware for human body tracking, multimodal input and output, and powerful 3D game development engines enables the deployment of computer games that are far more engaging than those of only a few years ago. This section provides an overview of related work in these fields.

Body Tracking

Estimating the pose of a moving body in six degrees-of-freedom (DOF) has been the subject of considerable previous literature, involving both outdoor and indoor tracking technologies. High quality hybrid GPS systems have been used in location-aware games, providing acceptable accuracy for the intended applications. For example, Piekarski and Thomas (2002) use differential-GPS and a digital magnetic compass to achieve suitable positional resolution in large outdoor environments (2-5 meters for position, $\pm$ 1 degree for orientation). Indoor tracking systems present technical challenges that are

often resolved with cumbersome or expensive infrared (IR) or radio (RF) systems (Want and Hopper, 1992; Philipose et al., 2000). Active Badges, introduced by Want and Hopper (1992), and similarly, the Local Positioning System (Shen et al., 2004) emit a unique optical signal at a regular frequency. Sensors or cameras mounted at fixed positions process the received signals to determine the identity and location of each tag, typically corresponding to a unique user. Such approaches are limited by the range of the optical signal, sensitivity of the receiver, in particular to ambient illumination, and occlusion effects, which often require installation of multiple receivers throughout the environment. For indoor environments, vision-based methods are generally less expensive and easier to deploy. Piekarski and Thomas (2002) use a fiducial marker system to register body position when the user enters into a building. An upward-pointing camera, mounted on the user's backpack observes fixed sized markers and from their geometry, determines position and orientation. Motivated by the ease of use, low cost, and flexibility of this approach, we implemented a similar tracking method for players in Snowdown.

Multimodality

The HCI community has long maintained that major improvements in computing will be related not just to processing speed, but to interaction, responsiveness and transparency (Corradini et al., 2003; Dray, 1995; Carrol, 2002). One approach to interface improvement is to employ multimodal interaction techniques (Bernsen, 2002). For example, the PlayStation EyeToy (Sony, 2005) uses a webcam connected to the game console to track gestures using an illuminated wand held in one of the user's hands. The device allows the user to point and activate virtual objects, navigate through menus, and drag-and-drop content from one location to another using manual gestures. Furthermore, combining visual and auditory feedback in a location-based quiz game (Klante et al., 2005) was seen to result in improved performance and positive user experience. Our game takes a similar approach, incorporating both audible and visual cues as feedback wherever our usability tests indicated this to offer an improved experience. Bimanual input offers additional benefits, including time-motion efficiency through the increased degrees of freedom and a decrease in cognitive load (Leganchuk et al., 1998).

DESIGN AND IMPLEMENTATION

Snowdown uses a simple game concept, a snowball fight, that can be grasped easily by non-gamers, permitting a wide audience to begin playing without any instruction. As a research project, our goal was not the development of the game as a commercially viable end-product but as a study of the interaction experience and immersion paradigm offered by consumer technologies available today. However, one could imagine such a system eventually being deployed in public entertainment centres or home entertainment rooms. The physical space requirements of our prototype system are roughly fulfilled by the size of large living rooms.

Each player attempts to throw snowballs at the opponent, scoring points and lowering the opponent's health with every hit. Snowballs are gathered by a shoveling gesture. Players can also block incoming snowballs by raising a shield or ducking, both enacted by their corresponding physical gestures. The game ends when either player's health drops to zero. Players are represented by avatars, seen on the projected display from a third-person view of the environment, as illustrated in Figure 1.

The game uses two Wii-controllers (or Wiimotes) (Nintendo Wii Remote, 2007) as primary input devices. A fiducial tracking system, described in Section 3.3, tracks the three-dimensional position of both players, driving the on-screen characters. In the event that the tracking system is not deployed, alternatives were implemented to control player motion and other actions using only the Wiimotes.

Figure 1: In-Game Screenshot – The players can be seen standing in front of a snow fortress, which provides cover from the opponent's snowballs.

Gestural Interaction

An important objective of our game design was that interaction with the game should feel natural to the user. We thus make extensive use of real-world physical gestures, leveraging the kinesthetic feedback these provide, as well as audio and visual feedback from the game engine, to enhance the feeling of immersion. The actions used for throwing, shoveling, and activating the shield are relatively large-scale in motion. This both physically engages the user with the game and aids in gesture recognition, which is performed using only the onboard accelerometers of the Wiimotes. The gestures are illustrated in Figure 2 and described below.

Throwing The dominant hand swings overhead, ending with a quick snap, releasing the snowball.

Shoveling The non-dominant hand jerks toward the ground twice in succession, as if thrusting into the snow.

Blocking/Shield Both hands are crossed and held in front of the player's chest. To disengage, both hands are pointed toward the ground. The rationale for the choice of this somewhat unnatural gesture is provided below.

The two main features used for gesture recognition are the inclusion of a jerking motion and sensing the force of gravity on the controllers. These can be detected reliably using simple algorithms; for example, a large spike in the accelerometer data is indicative of either a throwing or shoveling gesture. The shield gesture detector looks for gravity acting on the controllers in a certain direction. For this gesture, added recognition robustness is achieved if both hands perform clearly distinguishable actions. In all cases, the accelerometer data is converted to spherical coordinates before processing. This permits an easy identification of the direction in which force is being applied to the controller. Forces can only be measured with respect to the Wiimote casing, so we assume that the controllers are held in their typical orientation. Obvious problems with recognition accuracy result when this assumption is violated.

Evolution of the Gestural Vocabulary

The gestures evolved over the course of project development due to technological constraints. This evolution serves to illustrate the limitations of gesture recognition using only accelerometer data, as well as what designs may be more challenging. We had originally intended to provide each player with a virtual shovel, held in the non-dominant hand. This would be used both for picking up snowballs with a real-life shoveling action and acting as a shield when raised. However, because accelerometer data can only sense relative forces and orientations, these gestures proved difficult to detect. In particular, without the absolute position of the controller available, it was not possible to place the shield correctly in space.[1] For ease of prototyping, gesture parsing was carried out in Pure Data (Puckette, 1996), using the Wiimote external (Wozniewski, 2007) to obtain sensor information. The gesture messages are forwarded to the main C++ game application using a UDP loopback socket architecture, supported natively by Pure Data. Further improvements to gesture recognition would be possible using machine learning techniques such as neural networks or discriminant analysis.

[1]This could be resolved by the addition of an absolute orientation sensing device, such as Nintendo's recent unveiled Wii MotionPlus add-on.

Video-Based User Tracking

To track the players in 6 DOF, we implemented a video-based system utilizing fiducial markers, as pictured in Figure 3. The markers can be any asymmetric patterns surrounded by a black square. To reduce cost and implementation complexity, the markers are detected by a single head-mounted camera oriented toward the ceiling, as shown in Figure 4. In our system, we distributed 26 markers, each measuring 8 x 8 cm, over a ceiling area of approximately 4.5 x 2 m. Using the freely available ARToolkit API (2007), a program was developed that analyzes each captured frame to find markers and output the position and pose of the camera with respect to the detected marker.

Figure 3: Prototypical Fiducial Marker

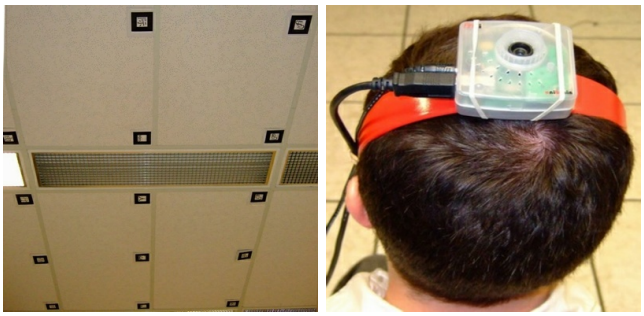
Once the physical pose of the player is computed, the avatar is updated accordingly. At present, we only map the positional x, y, and z coordinates, as the roll (x-axis rotation) and pitch (y-axis rotation) parameters were found to be unstable. The z (height) value is used as an indication of the player's crouching state. The mapping scales physical parameters to fit the range of motion in the virtual space, using the same scaling factor to each axis for coherency. This implies that the allowable virtual play area should be proportional to that of the physical space. No kinesthetic animations, such as throwing or shovelling motions, were portrayed by the avatar.

Our prototype implementation operates at 15Hz on an Athlon64 1.6Ghz system combined with a Unibrain Fire-i Firewire camera. Worst-case tracking inaccuracy constrained to translational movements was measured as approximately 10 cm in all directions at a distance of 1.6m from the ceiling. Unfortunately, any out-of-plane rotation of the camera results in significant positional error, due to numerical instability in the transformations employed by the fiducial marker tracking (the developers of this API have indicated that a revision to address this problem is forthcoming). Although we instructed our participants to avoid tilting their heads during testing, this constraint is clearly unacceptable for an immersive

(a) Throwing (b) Shoveling (c) Shielding

Figure 2: Example Interaction Gestures

game experience.
Although beyond the scope of our work, increased accuracy and robustness could be obtained by combining Kalman filtering and robust statistical methods (Park et al., 1999).



(a) Tagged Ceilling (b) Head-Mounted Camera

Figure 4: Fiducial Marker Tracking Setup

USER TESTING AND EVALUATION

Before committing to any design decisions, we iteratively evaluated and refined the system to address usability issues. This was accomplished by a series of tests as the game evolved from a low-fidelity prototype to a fully functional one. Since the overall concept diverged significantly from conventional game interaction, our usability testing required different paradigms.
Whereas most games or software applications require physical contact with the system through standard input devices, such as a keyboard, mouse, gamepad, joystick or even touchscreen, each of which have familiar affordances, our game design involves minimal use of familiar objects such as buttons or keys. Using gesture recognition and body tracking as the main means of input, we cannot assume a high likelihood for transfer of knowledge from other interactive computing applications. However, we consider this an advantage, since our goal is to exploit the gestural vocabulary from real-life interaction with everyday physical objects.
Following established principles of user centered design, we first developed a low-fidelity 3D prototype (Snyder, 2003) adequate for testing the system design on users. This took the form of a game environment built from styrofoam, cardboard and toy figures, and incorporated visual and audio output using the Wizard-of-Oz technique (Salber and Coutaz, 1993). A snapshot of the physical mock-up of the game and one of the user test sessions is shown in Figure 5. From this initial prototype testing phase, we evolved to a minimal computer prototype and then to increasingly functional ones, each time iteratively testing and incorporating changes.

Comparison Between Paradigms

In order to measure the overall effectiveness of bimanual interaction and physical body tracking of our game, we devised an experiment to compare our system to an equivalent keyboard-mouse interface without automated body tracking. The goal was to measure and assess different aspects of gameplay, immersion and overall user experience, both quantitatively and qualitatively.

Overview
In the keyboard-mouse setup, pointing and selecting the various menu items is accomplished with the mouse while snowball fighting gameplay is limited to keyboard entry. Predefined keys were assigned to navigate, throw

a snowball, crouch, activate shield, collect snowballs or pause the game. When possible, these key mappings were chosen by following the conventions of first-person-shooter type games. In the multimodal version of the game, bimanual input and body tracking are used, as previously described. In both scenarios, test subjects were given five minutes to familiarize themselves with the controls by playing the game against a randomly moving opponent. This was followed by a ten-minute trial in which the subject was asked to play (and win) as many games as possible within the allocated time. A post-test questionnaire was administered to each participant to rate different aspects of gameplay and immersion. The subjects were asked to give a subjective rating, on a scale of 0 to 5, of the *level of immersion*, *level of enjoyment*, and *ease of learning*. This was followed by an informal open-ended interview where the participants were encouraged to share their impressions of the two systems and qualitatively characterize them.

Participants

Seven participants (four male and three female) aged between 20 and 49, each with varying levels of gaming experience were studied. All had at least five years of computing experience and indicated themselves to be moderate to frequent computer game players. The style of games they preferred varied from sports to role-playing to action games, but all indicated mainly using the keyboard and mouse to play, with one user also having limited experience with a game controller similar to a Sony PlayStation type controller.

Results

The multimodal version of the game was preferred by all test subjects, some of whom commented on what they called a "refreshing new gameplay experience". Ratings for both the level of immersion (4.6 out of 5) and level of enjoyment (4.1 out of 5) were rated higher than the keyboard-mouse equivalent (3.3 and 3.1, respectively). Similarly, the multimodal system was judged to be easier to learn, receiving a score of 4.8 vs. 3.1 for the keyboard-mouse version. These results are all found to be statistically significant ($p < 0.01$) using Student's *t-test*. These results are summarized below in Table 1. Interestingly, there was little difference between the averages of number of games won in the ten-minute trials for the multimodal and keyboard-mouse versions of the game, which were 13 and 15, respectively, and not statistically significant ($p > 0.05$). This suggests that *winning* a game and *enjoying* the actual gameplay should be viewed independently. Clearly, the kinesthetic advantages of the multimodal version led to a more effective user experience, but not necessarily improved competence. . The informal interview provided added insight to the user's experience. Most participants vocally expressed the "natural feel" and "intuitiveness" of the multimodal system. Although none of the users commented on the inaccuracies of the motion tracking, two participants mentioned the lack of robustness for the shield activation gesture. No quantitative evaluations of the gesture recognition rate was performed. As a cautionary note, one user mentioned that the physical effort required to accomplish some of the gestures might lead to fatigue, and thus, require him to stop playing earlier than with the keyboard-mouse setup. This, of course, is a natural and expected consequence of our intended interaction paradigm. Another user questioned the practicality of the tracking system for home use, given the space requirements. With regard to areas for possible improvement, greater robustness of gesture recognition, improved graphics effects, and incorporation of the haptic feedback capabilities of the Wiimotes were suggested, the last by veteran gamers.

Table 1: Subjective Rating Results of Game Playing Modalities

	Multimodal	Keyboard-Mouse
Level of immersion	91%	66%
Level of enjoyment	82%	62%
Ease of learning	96%	62%

CONCLUSION AND FUTURE WORK

We have presented the design and implementation of a computer game intended to give players a more immersive experience than is typically possible with standard input and output devices. This is done by exploiting the capabilities of gestural control through the Wiimote, tracking of body position, and incorporating a large-scale projection display. User testing found a significant preference for our combination of input and output modalities, although future possibilities exist to enhance the level of multimodality and immersiveness of the game, including the use of spatialized audio, stereo video, and haptic feedback. As a benefit to future game developers, more extensive studies with a broad range of gamer types will be helpful to determine the value of each of the multimodal features of the system in isolation, as well as the cognitive load introduced through their combination, in particular in a highly interactive gaming environment. This latter concern has potential implications to a wide variety of applications beyond that of games.

In further development, it will be desirable to tune the current set of gestures, ideally with the goal of matching them more closely with their real world analogs. The interaction can also be expanded with additional actions, mapped to other motions appropriate to the game context. Major enhancements will include the incorporation of continuous parametric information from the gesture recognizer to drive the game elements more responsively and the use of pattern recognition techniques to

(a) Physical Mock-Up

(b) Wizard-of-Oz Testing Technique – The tester (crouching) controls the physical avatar to mimic the test subject's actions.

Figure 5: Usability Evaluation

improve recognition performance.

REFERENCES

ARToolKit, 2007. `http://www.hitl.washington.edu/artoolkit`.

Bernsen N., 2002. *Multimodality in language and speech systems: From theory to design support tool*, Kluwer Academic Publications. 93–148.

Brown E. and Cairns P., 2004. "A grounded investigation of game immersion". In *Proceedings of Conf. on Human Factors in Computing Systems (CHI)*.

Carrol J.M., 2002. *Human-Computer Interaction in the New Millenium*, ACM. xxvii–xxxvii.

Cheng K. and Cairns P., 2005. "Behaviour, realism and immersion in games". In *Proceedings of CHI 2005*.

Corradini A.; Mehta M.; Bernsen N.O.; Martin J.C.; and Abrilian S., 2003. "Multimodal input fusion in human-computer interaction". In *Proceedings of NATO-ASI Conf. on Data Fusion for Situation Monitoring, Incident Detection, Alert and Response Management*.

Dray S., 1995. "The importance of designing usable systems". *interactions*, 2, no. 1, 17–20.

Klante P.; Kroesche J.; and Boll S.C., 2005. "Evaluating a mobile location-based multimodal game for first-year students". In *Proceedings of SPIE*.

Leganchuk A.; Zhai S.; and Buxton W., 1998. "Manual and Cognitive Benefits of Two-Handed Input: An Experimental Study". *ACM Transactions on Human-Computer Interaction*, 5, no. 4, 326–359.

Nintendo Wii Remote, 2007. `http://wii.nintendo.com/controller.jsp`.

Park J.; Jiang B.; and Neumann U., 1999. "Vision-based Pose Computation: Robust and Accurate Augmented Reality Tracking". In *Proceedings of the 2nd IWAR'99, IEEE Computer Society*.

Philipose M.; Fishkin K.P.; Fox D.; Hahnel D.; and Burgard W., 2000. "Mapping and Localization with RFID Technology". In *Proceedings of IEEE Intl. Conf. on Robotics and Automation*.

Piekarski W. and Thomas B., 2002. "ARQuake: the Outdoor Augmented Reality Gaming System". *Communications of the ACM*, 45, no. 1, 36–38.

Puckette M., 1996. "Pure Data". In *Proceedings of the Intl. Computer Music Conf.* 269–272.

Salber D. and Coutaz J., 1993. "Applying the Wizard of Oz Technique to the Study of Multimodal Systems". In *Proceedings of EWHCI*. 219–230.

Shen C.; Wang B.; Vogt F.; Oldridge S.; and Fels S., 2004. "RemoteEyes: A Remote Low-Cost Position Sensing Infrastructure for Ubiquitous Computing". In *1st International Workshop on Networked Sensing Systems*. 31–35.

Snyder C., 2003. *Paper Prototyping: The Fast and Easy Way to Design and Refine User Interfaces*, Morgan Kaufmann.

Sony, 2005. EyeToy by Sony Computer Entertainment Inc. `http://www.eyetoy.com`.

Want R. and Hopper A., 1992. "Active badges and personal interactive computing objects". *IEEE Transactions on Consumer Electronics*, 38, no. 1, 10–20.

Wozniewski M., 2007. “Wiimote external for Pure Data”. http://mikewoz.com/index.php?page=pd-stuff.