STC Usability SIG Newsletter
Usability Interface

What We Learned Evaluating the Usability of a Game

By Shannon Lucas and Denise Fulton

Papers often approach either the need for usability testing in games or the differences between what makes a game useful versus what makes a productivity application useful. This article focuses on the methods employed for the usability testing of Ion Storm's game Thief: Deadly Shadows.

Background

Deadly Shadows is the third installment of Ion Storm's stealth-based Thief game series. Rather than relying on the carnage associated with first person shooter games, the Thief series requires the user to practice stealth while eavesdropping and collecting loot in order to progress through the story; play involves evasion, espionage, and assassination. Players are immersed in a rich medieval fantasy world with large areas to explore, constant interaction with numerous characters and objects, and an engaging story line.

The primary goals for usability testing Deadly Shadows were:

1. To gauge overall player satisfaction with the game as it progressed through the alpha stage of development.
2. To identify what parts of the game were proving to be impediments to player enjoyment.

The approach we chose was the Rapid Iterative Testing and Evaluation (RITE) methodology used by Microsoft (Pagulayan, 2003). Using this approach, tests were to be conducted at short intervals with the results of each test being provided to the development team to apply modifications to the weekly builds. Thus we were pursuing a "find-and-fix" approach.

Play Tests

Gathering test subjects proved a relatively easy task, as the average college student met Ion Storm's target demographic for the game. Although we offered no compensation for participation, we collected a list of over 100 volunteers by sending a notice to a university mailing list. Gamers are eager both to be among the first to play an upcoming game and to be able to say they took part in a game's production.

Our tests consisted of three parts:

1. A game play session lasting four hour
2. A post-play group discussion lasting approximately 20 minutes
3. A web-based follow-up survey for any additional comments

The game play sessions were designed to accommodate three subjects—two on Xbox consoles and one on a PC. An observer remained in the room for the duration of the tests.

Usability & User Experience
An STC Community

Newsletter

This article was originally posted in the October 2004 issue (Vol 11, No. 2)

About the Authors

Shannon Lucas is a graduate student in the School of Information at the University of Texas at Austin.

Denise Fulton is Executive Producer at Ion Storm.

Xbox is a product of the Microsoft Corporation.

Related Topics

- Design and Documentation of Games

Players were encouraged to talk aloud both to each other and to the observer. The observer was not allowed to help the players except with technical difficulties, and was required to take notes on the players' game play and reactions to the game. We made a conscious decision not to include as an observer any member of the game's development team. We did this because we felt that people closely involved with the game's development may focus too closely on how they envision the game "should" be played rather than how the player is actually playing the game.

We conducted 14 play tests with 32 subjects. Each test was conducted using the most recent build of the game that had passed the quality assurance test suite. (Under most circumstances, this build would include changes based on the results of the prior week's play tests.) We captured the video output on the Xbox consoles on videotape. Our original intent was to review the game play videotape after each test, but it proved too time consuming. Instead, the tapes became a resource for the artificial intelligence developers; viewing the tapes allowed them to see how their artificial characters were interacting with the players in real game play.

Because the initial experience in a game is critical to keeping the player interested, our play tests focused on the training level and the first mission level of the game. (The training level introduces players to the various skills required in the game.) The importance of a good user experience here cannot be overstressed.

Halfway through the test cycle we began to play test some of the later levels in the game and called in some of our earlier subjects. This had the added benefit of getting feedback on how these repeat subjects felt the game had progressed since they last played.

Following the game play, the players were asked to participate in a discussion about the game. We found that this is where we got the most important feedback. Players were asked questions about where they had difficulties, what they liked and didn't like, and what they would add or take away.

We interviewed each session's subjects as a group. Our rationale was that it would reduce any intimidation the subjects might feel and draw out things that a single subject might have forgotten or not thought important. The planned 20 minutes often became an hour; our subjects seemed to enjoy talking about the game almost as much as playing it.

The online survey was the last component of our tests. Subjects were asked to complete the online survey within a few days of the play test to provide any further thoughts that may have fermented after the discussions or to voice any concerns they did not feel comfortable mentioning as a group. The survey provided more quantitative data and was weighed more heavily than the individual interview feedback in decisions to modify the game. This gave us information such as "8 out of 10 players had difficulty using the blackjack tool." We experienced a high response rate to the survey.

It may also be of interest that we did not have access to a traditional test room with a two-way mirror. This may have been to our advantage, as it resulted in an environment similar to that of a network gaming party in which players play with or against one another in a room of networked computers.

Conclusion

The biggest change we would make for next time would be to start planning and executing the testing earlier. Starting the planning earlier would have allowed us to work out some logistics problems sooner; instead, we had to work out the kinks during the first two test sessions. Executing the tests earlier would have provided more time to make modifications to the game based on the test results.

The gaming community warmly welcomed Thief: Deadly Shadows with glowing reviews. We believe that the usability tests contributed to its successful development. We relied on a mixture of qualitative (observation and interviews) and quantitative (survey results) data in this study.

The data we collected from our testing resulted in changes ranging from enhancing the tutorial level, changing camera controls, and revamping the menu system, to numerous smaller refinements such as changes to the user interface and how players interacted with small items such as boxes and candles. All of these contributed to an improved player experience.

References

Randy J. Pagulayan et al., "User-centered design in games" in Handbook for Human-Computer Interaction in Interactive Systems (Mahwah, NJ: Lawrence Erlbaum Associates, Inc., 2002), 883-906.

Thief: Deadly Shadows: www.thief3.com/

Ion Storm: www.ionstorm.com/

Comments or questions?
Please send your email to <stcusability at sufficiently.com>