Evaluating the Benefits of 3D Stereo in Modern Video Games

Tad Litwiller
University of Central Florida
Department of EECS
4000 Central Florida Blvd.
Orlando, FL 32816, USA
tad.litwiller@knights.ucf.edu

Joseph J. LaViola Jr.
University of Central Florida
Department of EECS
4000 Central Florida Blvd.
Orlando, FL 32816, USA
jjl@eecs.ucf.edu

ABSTRACT
We present a study that investigates user performance benefits of 3D stereo in modern video games. Based on an analysis of several video games that are best suited for use with commercial 3D stereo drivers and vision systems, we chose five modern titles focusing on racing, first and third person shooter, and sports game genres. For each game, quantitative and qualitative measures were taken to determine if users performed better and learned faster in the experimental group (3D stereo display) than in the control group (2D display). A game experience pre-questionnaire was used to classify participants into beginner, intermediate, and advanced gameplay categories to ensure prior game experience did not bias the experiment. Our results indicate that although participants preferred playing in 3D stereo for the games we tested, it does not provide any significant advantage in overall user performance. In addition, users' learning rates were comparable in the 3D stereo display and 2D display cases.

Author Keywords
3D stereo, video games, evaluation, user performance and experience

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous, K.8.0 Personal Computing: General – Games.

General Terms
Experimentation, Measurement, Performance

INTRODUCTION
3D stereo is making a huge push in the entertainment industry. With 13 3D movies released in 2009, 25 released in 2010, and 32 scheduled to come to theaters in 2011, as well as new television stations broadcasting entirely in 3D, movies and television are clearly pushing the technology. As the technology has started to become more available to consumers, game designers and hardware makers have started to take notice. For example, "The U-DECIDE Initiative" was an online survey run by Meant to be Seen, an advocate group for 3D stereo technology, that sought to determine consumers interest in 3D stereo gaming. The survey concluded that the overwhelming majority of users want game developers to natively support 3D stereo in their games [10]. Research has also shown that 3D stereo can be beneficial to user performance in certain, isolated tasks in the context of virtual reality and 3D user interfaces [2,6,14]. However, to the best of our knowledge, there has been little, if any, work that explores whether 3D stereo benefits a user playing video games in terms of performance and learning. Understanding these benefits could provide useful information to help understand and improve upon ways to increase motivation to play games. This motivation is important to game developers in the commercial sector, but, more importantly, it is important in the serious games domain where we can make use of these technologies to potentially improve learning in education-based games, training in military simulations, and obesity reduction and physical activity in exercise-based games.

To explore this question, we present a study investigating whether user performance and learning are enhanced when using 3D stereo over a traditional 2D monitor in modern video games. In our work, we define learning as becoming more proficient at the game and the tasks associated with the game. We analyzed several video games that are best suited for use with commercial 3D stereo drivers and vision systems (iZ3D and NVIDIA 3D Vision) and chose five modern titles including Left 4 Dead, Resident Evil 5, Flatout: Ultimate Carnage, Madden NFL 2008, and Major League Baseball 2K9 as a representative sample of modern video games. In choosing the games, we sought to have them spread across video game genres as well as have games that included tasks that, in our judgment, may benefit from 3D stereo. To evaluate player performance, we collected both quantitative data based on the tasks associated with each game and qualitative data based on post-questionnaires to gage user perception of their performance. We used between subjects design where the control group played the games using a 2D monitor and the experimental group played with a 3D stereo display. We also used a modified version of Terlekie and Newcombe's Video Game Experience survey [15] to classify participants into beginner, intermediate, and advanced categories to ensure prior game experience did not bias the experiment.

RELATED WORK

There has been a significant amount of work on the topic of the benefits of 3D stereo, especially in the virtual reality and 3D user interface communities, with mixed results [2,6,14]. Much of the research to date has focused on simple, isolated tasks in virtual environments, and there has been very little research involving more complex tasks and richer graphical environments, such as games. Menendez and Bernard have hypothesized that stereoscopic viewing would benefit a user in a flight simulation environment, but have yet to test the hypothesis [11]. Another study has concluded that binocular viewing in the real world as well as in virtual worlds may benefit the user over monocular viewing. While 3D stereo has been shown to be useful for depth ordering of objects in a virtual world, it may be impossible to measure how accurate a user's perception of 3D stereo is [13]. Research has also been conducted that looks into how shadows and stereo enhance user's perception of 3D. Positioning and resizing tasks were given to users with shadows on and off and with stereo and mono viewing. The researchers concluded that stereo viewing is more effective than shadows based on accuracy and speed with which users completed the tasks [6]. Stereo has also been found to help users playing a game in which they eliminate targets by moving objects into defined zones. The game was still a simple task of moving a cursor to a target in the virtual world that contains objects that needed to be maneuvered. To simplify the task only one object was present during the experiment [3]. This is still quite different from playing a modern video game, in which there is a lot more happening on screen and a lot more visual stimulus.

Other work has looked at benefits of 3D stereo and sought to separate the interaction technique from the stereo. In these studies, the interaction technique was found to be significant while stereo was not [9]. However, this finding has been somewhat contradicted by Teather and Stuerzlinger, who presented different positioning techniques that were dependent on the input devices used. They found that stereo was beneficial for accuracy in the tasks they presented to users, but not speed [14].

Specific work on comparing 3D stereo to 2D displays has also been done. However, very simple tasks were used to measure performance and other factors such as head tracking were included in the research. One of the more common display configurations used for this type of research is called Fish Tank Virtual Reality, where there is a desktop system with a stereoscopic display and head-tracking [2]. Arthur, Booth, and Ware used this setup to conduct experiments that compared viewing conditions of stereo display versus non-stereo display with head-tracking. In one such study, users thought that head-tracking created more compelling 3D perception than stereo viewing alone. Although head-tracking had better results in both user perception and performance in a tree tracing task, 3D stereo did show significant benefits over normal viewing [2]. While these types of studies have concluded that head-tracking produced better results than 3D stereo alone, we wanted to focus this study specifically on 3D stereo because it is more readily available to consumers.

Research has also been conducted on how well users perform with different types of 3D displays. Grossman and Balakrishnan looked at volumetric displays and concluded that for the simple tasks that were presented to users, 3D stereo always helped over simple perspective and although volumetric displays were more helpful for simple scenes, there was no benefit over normal 3D displays in more complex scenes [4]. Fully immersive virtual environments have also been shown to be more effective than stereoscopic desktop environments for certain tasks. In comparing a real world scenario of oil well path editing, researchers found that a fully immersive environment, such as a CAVE, was more effective than a stereoscopic desktop environment [5]. A similar study showed results that also suggested the immersive environment provided benefits to the user in analyzing data; however, it also concluded that users were more comfortable using the interaction techniques on the desktop environment [1]. Work has also been done that concluded users could identify targets better in an immersive 3D virtual world than with a desktop system [12]. Stereo has been shown to increase the size and amount of abstract data that can be viewed and understood, and the benefits were only increased with a higher resolution stereoscopic display [17]. We used a desktop setup for viewing the games in stereo on a 3D TV with a 1080p resolution. While some of the research may suggest that 3D stereo on different display types may be more effective, the desktop setup of a 3D TV with a computer and 3D glasses is more readily available for home use.

Many questions about the effectiveness of 3D stereo exist as the technology is far from perfect. It has been shown that depth perception tends to be underestimated by users in virtual environments [8], and also that for some selection tasks in 3D space, a one-eyed 2D cursor can be more beneficial than a 3D cursor [16]. While it may show some benefits depending on the task presented, it has also been shown to increase some negative symptoms as well. Stereoscopic viewing can have negative consequences and symptoms, such as additional eyestrain and simulator sickness. We wanted to see whether this trend of increased symptoms while viewing in 3D would also hold true while playing a game in stereo on a 3D TV, and we wished to see whether any of the benefits of 3D stereo that have been shown in virtual reality environment for simple tasks will translate to improved performance for users playing modern video games.

SELECTING THE GAMES

In order to look into possible benefits that 3D stereo provides to modern video games, we wanted to get a good sample of different game genres. We also wanted to make sure the games we used were highly rated for their use of 3D stereo. The technology we had available were iZ3D monitors and the NVIDIA 3D Vision kit, which are both products currently on the market. Both of these technologies provided a list of modern games that were compatible with the technology. We looked through both lists and picked 21 games that were rated highly for their use of 3D Stereo by both NVIDIA and iZ3D. We then played through sections of the 21 games in 3D stereo, taking notes about how they looked, any glitches that occurred, and any tasks we thought may benefit from 3D stereo. From our analysis, we decided to first remove the real-time strategy games as the 3D was not very convincing and it did not seem like any advantage could be gained in that genre. Based on our analysis, we felt a first person shooter, a third person shooter, a racing game, and sports games would be the most appropriate genres to explore. Of the 21 games that we tested, we found Left 4 Dead, Resident Evil 5, Flatout: Ultimate Carnage, MLB 2K9, and Madden NFL '08 performed the best in terms of visual quality and had tasks that we felt could benefit from 3D stereo.

Left 4 Dead is a first person shooter by Valve. Although we did not expect to see any difference in the shooting aspect of the game due to cross hairs being displayed on screen, which make aiming easier by just placing the cross hairs over the target, we wanted to explore the genre to see if 3D stereo may improve performance by helping with navigating the world or providing better sense of which enemies are closer and thus pose a more imminent threat.

Resident Evil 5 is a third person shooter by Capcom. It provided a mode in which the user is required to shoot with a bow and arrow that did not provide any cross hairs on the screen. This made targeting an enemy more difficult as the user is required to aim the arrow in 3D space. Third person perspective also adds to the aiming difficulty as the aiming is no longer only in the center of the screen as it is with first person shooters. We hypothesized that the added depth cue provided by 3D stereo may help the user with accuracy.

Flatout: Ultimate Carnage is a racing game by Empire Interactive. Although we did not expect to see any performance difference in racing games, we wanted to explore the possibility that the added depth perception of 3D stereo could aid maneuvering the course or help in judging corners.

MLB 2K9 is a major league baseball game by 2K Sports. In the baseball game we wanted to see if 3D stereo would help the user better judge where the ball was located in space as it came from the pitcher toward the batter in order to swing the bat at the appropriate time. We thought that the added depth perception could help users' timings when hitting the ball.

Madden NFL '08 is a football game by EA Sports. There was a punting mini-game in Madden that required the user to aim an arrow in 3D space to kick the ball in a desired direction toward a target. Because the aiming of the arrow was free in 3D space, we thought that 3D stereo may boost the user's performance in successfully hitting the targets.

USER STUDY

To explore whether there are any performance benefits to using 3D stereo with the video games we selected (discussed in the previous section), we conducted a usability evaluation where participants played each game using a 2D display (control group) or with a 3D stereo display (experimental group). We examined both quantitative metrics based on each game's goals and tasks and qualitative metrics based on whether participants preferred playing the games in 3D and whether they perceived any benefits. Based on our analysis of the games, we hypothesized that users would prefer playing in 3D stereo because it would provide a more engaging experience. However, there would not be any significant performance improvements in overall performance or in learning since the games were not specifically designed with 3D stereo in mind.

Participants and Equipment

Forty participants (30 males and 10 females ranging in age from 18 to 36 with a mean age of 23.15) were recruited from the University of Central Florida. We ranked the participants based on a modified version of Terlekie and Newcombe's Video Game Experience survey [15] that was used as a pre-questionnaire in which they answered questions about their previous gaming experience. Of the 40 participants, 9 were ranked as beginners, 23 as intermediate, and 8 as advanced. The experiment duration ranged from forty-five minutes to an hour and a half depending on how long the participants took to complete the tasks presented to them in the games and how much time was spent on the questionnaires. All participants were paid 10 dollars for their time.

The equipment used for the study consisted of a quad-core desktop PC with an NVIDIA GTX 260 graphics card and a Samsung 50 inch DLP 3D HDTV display, using the NVIDIA 3D Vision kit for the3D stereo gaming as can be seen in Figure 1. We decided to use the NVIDIA 3D setup over the iZ3D monitors because early pilot testing showed the NVIDIA solution had higher stereo quality with limited ghosting. In addition, the iZ3D monitors had variables that would need to be setup differently for each game and each user, where as the NVIDIA solution worked seamlessly with the games. The Xbox 360 Controller for Windows was used as the input device for the games.

Figure 1: *The experimental setup.*

Experimental Task

The participants were tasked with playing through sections of the five games that we selected. For each game, they were presented with a task specific to that game and a goal for completing each task.

Left 4 Dead. Participants had to finish the first section of "The Apartments" level. The goal was to finish the level as fast as possible. They were given three attempts with a goal of making it through the level faster each time. In the mode used for this experiment, there were three AI controlled teammates moving through the level with the participant.

Resident Evil 5. The task participants were given for Resident Evil 5 was to eliminate as many zombies as possible. The participants were instructed that the only weapon they could use was the bow-and-arrow. The moderator kept track of the time and the participants were given six minutes to play the game. They were told that deaths did not count against them.

Flatout: Ultimate Carnage. Participants were given a goal of finishing a lap in one minute and fifteen seconds. We determined that this was a challenging, but attainable goal through our time playing the game. They were given five attempts to reach the goal.

Madden NFL '08. Participants took part in a punting mini game. They were given five attempts to reach the goal of a gold medal, a score of 900 points. There were three zones in the targets they were aiming for: a 50 point zone, a 100 point zone, and a 200 point bull's eye. Each attempt consisted of six kicks to reach the desired point total.

MLB 2K9. Participants' tasks in this game were to hit 20 home runs in as few swings as possible during a home run derby. Participants controlled two hitters, both a right and left-handed batter that alternated every three outs. The pitcher would throw the ball and the participant would have to time the swing correctly to hit a home run.

Design and Procedure

Since the focus of our study was on user performance and the ability to learn to play the games dependent on whether participants were using the 2D display or 3D stereo, we chose a between subjects design. The independent variable was display mode (2D display or 3D stereo display) and the dependent variables were the various scoring metrics used in each game. A between subjects design was chosen over a within-subjects design in order to eliminate the bias and learning effects that would have followed from the subjects being exposed to the games in both 2D display and 3D stereo conditions. With a within subject design, the participants would have already been exposed to the games in a prior condition and we would not have been able to isolate their performance specifically for the 2D display and 3D stereo modes. We rejected this approach given it would have doubled the length of the study without providing much additional benefit in terms of quantitative data collection, making between subjects the better design choice. We did want to get some additional information about the use of 3D stereo in video games for those who played the games in the 2D display condition. Thus, we chose to have those participants who were in the 2D display condition, pick one game to try in 3D stereo to gather their reactions.

In order to group the participants into experience levels based on the pre-questionnaire data, we scored the questionnaire by assigning points to each question. Particular questions were given more points based on how the results fit within the context of our experimental setup. For example, participants who were familiar with the Xbox 360 controller or whose favorite games were first and third person shooters were considered to have a higher experience level. We then used the raw scores from adding up the points for each question to group the participants into the appropriate category. Both the quantitative and qualitative data was explored collectively as well as according to the three groupings.

Summary of Metrics	
Left 4 Dead	• Time for each run • Number of kills for each run • Kills per second
Resident Evil 5	• Accuracy • Number of Player deaths
Flatout	• Time for each attempt • Number of crashes • If a shortcut was attempted
MLB 2K9	• Number of swings taken to reach 20 HRs • Number of swings missing the ball
Madden '08	• Score for each attempt • Number of kicks hitting each target zone • Number of kicks missing target

Table 1: *Summary of Metrics.*

Quantitative Metrics

For each game, we tracked quantitative data that we felt was a good indication of how well the user performed. A summary of the quantitative metrics can be viewed in table 1.

Left 4 Dead. We collected times for each attempt at the level as well as the number of kills for each run. The number of kills could differ from run to run because there were three AI controlled friends helping the user in the game by shooting enemies, and the number of enemies spawned would change based on how long it took to get through certain areas of the level. We also looked at the ratio of kills per second.

Resident Evil 5. We tracked each shot the participant fired in Resident Evil 5, so we had the ability to look at the overall accuracy as well as the accuracy over the course of the participant's time playing the game. We also decided to break up the number of shots into thirds so we could track the participant's accuracy at the beginning, middle, and end of the gaming session. The number of deaths was also tracked.

Flatout: Ultimate Carnage. The time for each lap attempted was recorded along with the number of crashes, and whether or not a short-cut (i.e., a specific corner in the lap) was attempted.

MLB 2K9. The quantitative data we were tracking for MLB 2k9 was the total number of swings needed to reach 20 home runs. Each swing was tracked, so we also looked at the number of misses, both early and late.

Madden NFL '08. We recorded the score for each attempt, as well of the number of kicks that hit each section of the target and the number of kicks that missed the target completely.

Postgame Questions	
Q1	To what extent did the game hold your attention?
Q2	How much effort did you put into playing the game?
Q3	Did you feel you were trying your best?
Q4	To what extent did you lose track of time?
Q5	Did you feel the urge to see what was happening around you?
Q6	To what extent did you find the game challenging?
Q7	How well do you think you performed in the game?
Q8	To what extent did you feel emotionally attached to the game?
Q9	To what extent did you enjoy the graphics and the imagery?
Q10	How much would you say you enjoyed playing the game?
Q11	Were you disappointed when the game was over?
Q12	Would you like to play the game again?

Table 2: *Postgame Questionnaire. Participants answered these questions on a 7 point Likert scale after playing each game.*

Qualitative Metrics

Our initial approach for the qualitative aspects of the study was to provide each participant with a post-questionnaire after they played all 5 games. From early pilot studies, we determined that we should reduce the number of qualitative questions and give the survey after each game rather than one larger questionnaire at the end of playing all the games. After playing each game, participants filled out a short, 12 question survey that was based on an immersion questionnaire from Jennet et al. [10] (see Table 2) aimed at gathering their ideas on how involved or immersed in the game they became. The questions were the same for each game and responses were measured on a 7 point Likert scale (1 = most negative response, 7 = most positive response). An additional 11 question survey was also given to participants in order to gather their opinions on how playing the games in 3D stereo affected their experience (see Table 3). These questions included whether they preferred to play the games in 3D stereo and whether 3D stereo helped or hurt their performance when playing the games.

Procedure

The experiment began with the participant seated in front of the computer and the moderator seated to their side. Participants were given a standard consent form that explained the study and what they would be asked to do. They were then given a pre-questionnaire that focused on their gaming experience. Participants were then presented with the games, in random order. Half the participants played the games in 2D display mode (control group) and half played in 3D stereo (experimental group). The moderator would present the game and give instructions to the participant as to what they needed to accomplish in the game and what their goals were. After each game, the participant filled out a post-questionnaire with questions about their experiences with the game. If the participants played the five games in the 2D display group, they then selected one game to play in 3D stereo. Thus, all participants were given a final post-questionnaire about their experiences with the 3D stereo display.

3D Stereo Questions	
Q1	3D stereo improved the overall experience of the game.
Q2	I would choose to play in 3D stereo over normal viewing.
Q3	I felt that stereo enhanced the sense of engagement I felt.
Q4	3D stereo is a necessity for my future game experiences.
Q5	What would you be willing to spend on a 3D stereo device?
Q6	Did 3D stereo help you perform better in the games?
Q7	Which games did it help you in?
Q8	How did it help you in those games?
Q9	Did 3D stereo hurt your performance in the games?
Q10	Which games did it hurt your performance in?
Q11	How did it hurt you in those games?

Table 3: *3D Stereo Questionnaire. Participants responded to statements 1-4 on a 7 point Likert scale. Questions 5-11 were multiple choice and open ended questions to gauge the users' perceptions of the effects of 3D stereo.*

RESULTS AND ANALYSIS

To analyze the performance data, we used independent sample t-tests to look for significance between groups. We also wanted to see whether there was learning taking place in the form of gameplay improvement. We looked at the improvement in the performance measures for each game from the first user run to their last run using a repeated measures ANOVA. Finally we wanted to look at the participant's perception of their performance through the post questionnaires. To analyze this Likert scale data, we used the Wilcoxon Signed Rank test. For all of our statistical measures, we used α=0.05.

Left 4 Dead

There was a significant difference in average time (t_{37} = -2.626, $p < 0.05$) and average kills per second (t_{37} = 2.334, $p < 0.05$). The 2D display group was actually faster in this game with a mean completion time of 214.36 seconds (σ=116.73), compared to 329.14 seconds (σ=152.74) for the 3D stereo group. In addition, participants had significantly

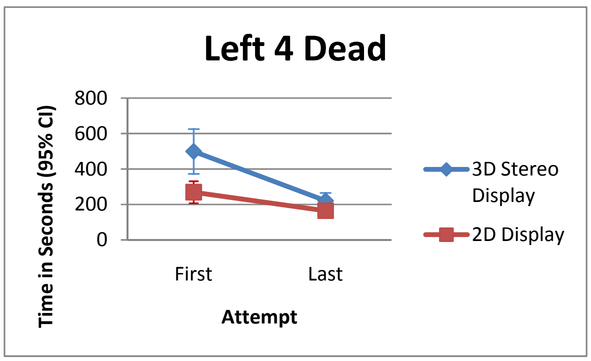

Figure 2: *Left 4 Dead Improvement. This graph shows the improvement in times from the first to the last run.*

more kills per second for the 2D display group as well with a mean of 0.204 (σ=0.063) kills/sec to the 3D stereo display group's mean of 0.155 (σ=0.066). There was no statistical difference in the average number of kills (t_{37} = -0.981, p = 0.333). After looking at the overall average of the three runs, we decided to look at the third and final run in an attempt to remove some of the experience factor, as by this time everyone would have knowledge of the level and the controls. We thought that this might be a good judgment of raw performance for the two groups. As we had originally expected for Left 4 Dead performance, there was no statistically significant difference in the completion times (t_{37} = -1.89, p = 0.067) or number of kills (t_{37} = - 0.268, p = 0.79) for the last attempt.

When isolating the three gamer ranks, the beginner group and the advanced group showed no differences for any of the statistics. However, the 2D display group performed significantly better for participants in the intermediate rank for worst time ($t_{10.48}$ = -2.875, $p < 0.05$), best time (t_{21} = -2.432, $p < 0.05$), average time ($t_{11.37}$ = -3.021, $p < 0.05$), average kills per second (t_{21} = 2.351, $p < 0.05$) and the third attempt's kills per second (t_{21} = 2.29, $p < 0.05$).[1]

Because 22 out of the 40 participants had played Left 4 Dead previously, we separated the participants based on whether or not they had played the game. This resulted in no significant differences for any of the metrics. Since this result is what we originally expected, previous game experience may have affected the overall performance statistics.

Participants significantly improved their times between the 3 runs for both 2D display ($F_{2,17}$ = 16.64, $p < 0.05$) and 3D stereo display groups ($F_{2,18}$ = 14.00, $p < 0.05$). This shows that there was some learning taking place between runs for both groups. The 3D stereo display group improved their time from 499.54 (σ=281.42) seconds to 220.82 (σ=97.90) seconds while the 2D display group improved their run. time from 269.09 (σ=141.38) to 164.64 (σ=87.08) seconds. This translates to a 55.8% improvement for the 3D stereo display group compared to a 38.8% improvement for the 2D display group (see Figure 2).

When broken down based on the gamer ranks, the beginning 3D stereo display ($F_{2,3}$ = 8.448, p = 0.059), beginning 2D display ($F_{2,1}$ = 17.59, p = 0.166), the advanced 3D stereo display ($F_{2,3}$ = 4.452, p = 0.127), and the advanced 2D display ($F_{2,1}$ = 0.586, p = 0.679) participants showed no significance in improving their times, while the intermediate 3D stereo display ($F_{2,8}$ = 10.88, $p < 0.05$) and 2D display ($F_{2,11}$ = 13.99, $p < 0.05$) groups appeared to show the same results as the overall learning. As with the overall learning rates, the intermediate 3D stereo display group appeared to outpace the intermediate 2D display group by about the same amount with a 57.8% improvement in time compared to the 2D display group's 38.7% improvement.

Participants in the 3D stereo display group who had not played the game previously saw a significant improvement in their times ($F_{2,11}$ = 13.72, $p < 0.05$) from 610.16 (σ=255.44) seconds to 257.16 (σ=91.83) seconds (σ=91.83), a 57.8% improvement. The 2D display group who had not played the game previously showed no statistically significant improvement between runs ($F_{2,16}$ = 7.92, p = 0.112). For participants who had previously played the game, both the 3D stereo display group ($F_{1.04,6.26}$ = 6.10, $p < 0.05$) and the 2D display group significantly improved their times ($F_{2,13}$ = 12.18, $p < 0.05$).[2] The 3D stereo display group saw a 47.86% improvement with their average time improving from 294.07 (σ=211.02) to 153.33

[1] For worst time and average time, Levene's test for equality of means was significant so a correction was used.

[2] The 3D stereo display group's test violated the sphericity assumption, therefore, we applied a Greenhouse-Geisser correction.

seconds while the 2D display group improved from 258.25 (σ=136.12) to 148.16 seconds (σ=60.94), a jump of 42.6%.

There were not many statistically significant differences in the qualitative data. Overall, the game was found to be significantly more challenging ($Z = -2.394$, $p < 0.05$) for the 3D stereo display group ($\bar{x}$ =4.90, σ=1.25) than the 2D display group ($\bar{x}$ =4.05, σ=0.89). This would be in line with the overall average time being worse for the 3D stereo display group. For those participants that had played the game before, the game significantly held the attention more ($Z = -1.981$, $p < 0.05$) for the 3D stereo display group ($\bar{x}$ =6.86, σ=0.38) than the 2D display group ($\bar{x}$ =6.2, σ=0.86). The same trend was seen for participants in the advanced rank where the game significantly held the attention more ($Z = -2.049$, $p < 0.05$) for the 3D stereo display group ($\bar{x}$ =6.8, σ=0.45) than the 2D display group ($\bar{x}$ =6.0, σ=0.00).

Resident Evil 5

Contrary to what we expected to see in Resident Evil 5, there was very little difference between the groups. There was no overall difference in the number of player deaths ($t_{38} = -0.62$, $p = 0.539$) or accuracy ($t_{38} = 0.024$, $p = 0.981$). This also held for the beginner player deaths ($t_7 = -0.743$, $p = 0.482$) and accuracy ($t_7 = -0.779$, $p = 0.461$) as well as the intermediate group's player deaths ($t_{21} = 0.206$, $p = 0.839$) and accuracy ($t_{21} = -0.617$, $p = 0.544$). While there was also no difference ($t_6 = -1.067$, $p = 0.327$) for the advanced group in player deaths, the group did show a difference ($t_6 = 2.794$, $p < 0.05$) in accuracy, but it was the opposite of the difference we thought we may see. The 2D display group displayed a higher accuracy of 71.8% compared to that of the stereo 3D group at 59.2%.

In order to test for improvement in accuracy throughout the participant's time playing the game, we divided each user's attempted shots into the first, second, and third group of shots, with each grouping of shots being equal in number for the participant. We then looked at the accuracy change from the first third to the second third to the last third. There were no significant differences found for either the 2D display ($F_{2,18} = 0.898$, $p = 0.425$) or the 3D display group ($F_{2,18} = 0.651$, $p = 0.533$) in the changes of the user's accuracy over the course of their time playing the game. This trend held for each gamer rank, both 2D display and 3D stereo display groups.

The only difference found for the qualitative data for Resident Evil 5 was in the group that had not played the game before. In this group, there was a difference ($Z = -2.104$, $p < 0.05$) for the question about how much they enjoyed playing the game. The 2D display group actually rated that they enjoyed playing the game more ($\bar{x}$ =5.33, σ=1.54) than the 3D stereo display group ($\bar{x}$ =4.41, σ=1.32).

Flatout: Ultimate Carnage

As we expected from this genre, in all of the quantitative data that we tracked for Flatout, which included average time, average time in runs without a crash, number of crashes, and best time, there were no differences found between the 3D stereo display and 2D display groups overall or at any experience level.

From looking at the difference in times from the first attempt through the fifth attempt, there did appear to be significant learning taking place in both the 2D display ($F_{2.85,51.26} = 12.35$, $p < 0.05$) and 3D stereo display groups ($F_{1.72,29.29} = 5.85$, $p < 0.05$).[3] As with Left 4 Dead, the rate of learning did look to be slightly higher in the stereo group who improved their time from 90.47 (σ=12.54) seconds to 81.49 (σ=5.51) seconds, compared to an improvement from 88.71 (σ=5.44) to 82.36 (σ=4.44) seconds for the 2D display group. The 3D stereo display group showed more improvement with a 9.93% gain compared to a 7.16% gain for the 2D display group (see Figure 3). When broken down by game ranks, the only significance shown in learning was for the intermediate 2D display group ($F_{4,9} = 20.55$, $p < 0.05$) who improved their time by 6.20%.

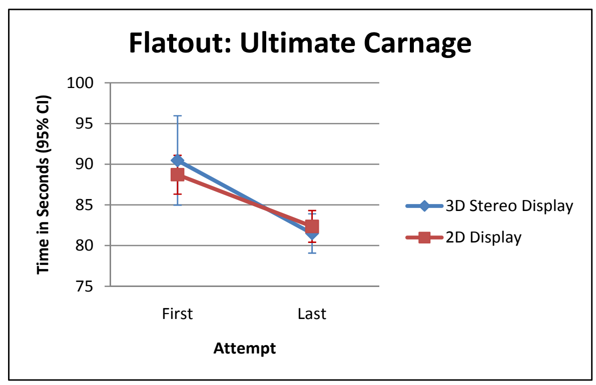

Figure 3: *Flatout: Ultimate Carnage Improvement.*

Like the previous games, there was not much difference in the answers received for the qualitative questions. The only differences came when divided into the 3 game ranks. For the intermediate rank, the 3D display group put significantly more effort ($Z = -2.35$, $p < 0.05$) into the game ($\bar{x}$ =6.5, σ=0.73) than the 2D display group ($\bar{x}$ =5.77, σ=0.73). In addition, the 3D display group gave significantly higher ratings ($Z = -2.344$, $p < 0.05$) for trying their best ($\bar{x}$ =6.6, σ=0.70) than the 2D display group ($\bar{x}$ =5.85, σ=0.80).

MLB 2k9

As we expected due to the task being more of a timing task than a spatial 3D task, there was no significant difference in the performance data for missing early ($t_{38} = 0.214$, $p =$

[3] For these tests, the sphericity assumption was violated, therefore, we applied a Greenhouse-Geisser correction.

0.832), missing late (t_{38} = -0.908, p = 0.370), outs (t_{38} = -0.141, p = 0.889), and total number of swings (t_{38} = -0.593, p = 0.556). This trend held across the gamer ranks as well.

Similar to what we did for Resident Evil 5, we broke the swings into thirds to evaluate the presence of any learning that may have been happening. Overall there did not appear to be any improvement as far as the number of home runs from the first third of the swings to the last third for either the 2D display ($F_{2,18}$ = 1.878, p = 0.182) or 3D stereo display groups ($F_{2,18}$ = 1.277, p = 0.303). There was significant improvement for both groups, 2D display ($F_{2,18}$ = 8.078, $p < 0.05$) and 3D stereo display ($F_{2,18}$ = 7.811, $p < 0.05$), when we looked at the number of swing-and-misses in each third. This time, the non-stereo group had a slight advantage in the improvement as they went from 3.6 to 2.0 misses while the 3D stereo group dropped to 3.0 misses from 4.2. This translated to a 44.4% improvement for the 2D display group compared with a 28.6% improvement for the 3D stereo display group (see Figure 4). When broken down by gamer ranks, however, the intermediate stereo 3D participants were the only group that exhibited this learning ($F_{2,8}$ = 5.954, $p < 0.05$) on the number of misses decreasing throughout their swings with a 26.3% improvement.

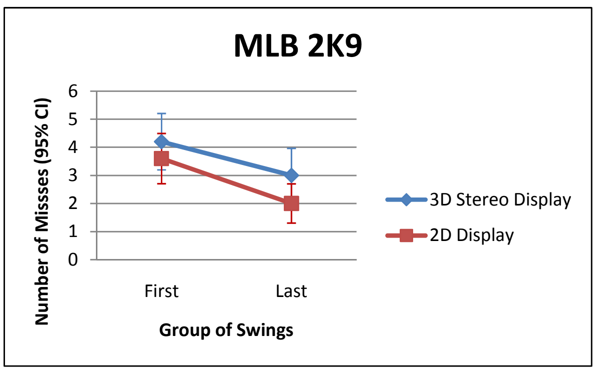

Figure 4: *MLB 2K9 Improvement.*

In line with what we have seen in other games, there was not much difference in the user's responses to the qualitative questions. The only significant difference (Z = -0.488, $p < 0.05$) was seen for the intermediate rank, where the 3D stereo display group was less likely to be distracted by what was happening around them ($\bar{x}$ =1.70, σ=0.67) than the 2D display group ($\bar{x}$ =2.85, σ=1.34).

Madden NFL '08

Going against what we expected to see from the task in Madden, there was no difference between the groups in number of kicks that missed the target (t_{38} = 0.64, p = 0.526), the number of 200 point target hits (t_{38} = -0.534, p = 0.597), 100 point target hits (t_{38} = -0.337, p = 0.738), 50 point target hits (t_{38} = 0.525, p = 0.603), best score (t_{38} = -0.858, p = 0.396), worst score (t_{38} = -0.135, p = 0.893), or average score (t_{38} = -0.62, p = 0.539). This held across the gamer ranks when broken down, there was no difference.

Overall, there did appear to be some learning happening when we looked at the difference in the user's score from the first try to the fifth and final try for both the 2D display ($F_{4,13}$ = 4.604, $p < 0.05$) and 3D stereo display groups ($F_{4,12}$ = 3.495, $p < 0.05$). As with Left 4 Dead and Flatout, the 3D stereo group demonstrated a more drastic change in this game as their scores grew to 406.25 (σ=297.7) from 171.87 (σ=146.02), a 136.4% improvement, while the 2D display group demonstrated a 103.2% improvement, increasing their score from 185.29 (σ=189.37) to 376.47 (σ=222.28) (see Figure 5). The only group that demonstrated significant improvement when isolated was the intermediate non-stereo group ($F_{4,8}$ = 4.53, $p < 0.05$). They made a 170% improvement as the increased their score from 170.83 (σ=151.44) points to 470.83 (σ=187.64).

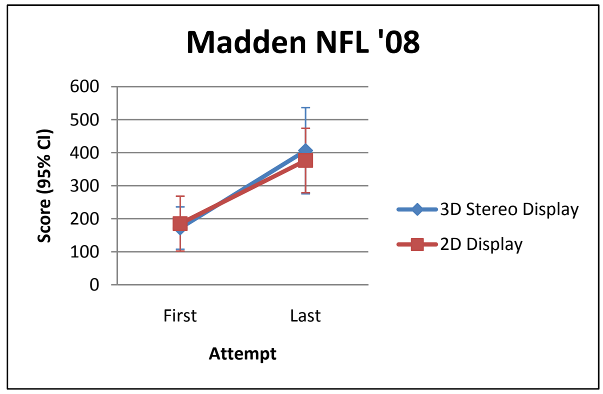

Figure 5: *Madden NFL Improvement.*

Madden NFL '08 actually did show some differences in the qualitative section. Overall, the 3D stereo display group gave significantly higher ratings (Z = -2.279, $p < 0.05$) for trying their best ($\bar{x}$ =6.05, σ=1.15) than the 2D display group ($\bar{x}$ =5.2, σ=1.20). The 3D stereo display group was also significantly more likely (Z = -2.337, $p < 0.05$) to want to play the game again ($\bar{x}$ =4.1, σ=1.41) than the 2D display group ($\bar{x}$ =2.95, σ=1.73). There were no differences for beginning users, but there were differences on 7 questions for intermediate users. The 3D stereo users had a more favorable view of the game (Z = -2.37, $p < 0.05$) in responding that the game held their attention more ($\bar{x}$ =5.3, σ=1.7) than the 2D display group ($\bar{x}$ =3.9, σ=1.38), they were more likely (Z = -2.337, $p < 0.05$) to lose track of time ($\bar{x}$ =4.6, σ=0.84) than the 2D display group ($\bar{x}$ =3.2, σ=1.63), enjoyed the graphics (Z = -2.5, $p < 0.05$) more ($\bar{x}$ =4.6, σ=1.07) than the 2D display group ($\bar{x}$ =3.2, σ=1.59), and were more likely (Z = -2.04, $p < 0.05$) to want to play the game again ($\bar{x}$ =4.5, σ=1.51) than the 2D display group ($\bar{x}$ =2.4, σ=1.56). For advanced users, there was a difference (Z = -2.037, $p < 0.05$) for 1 question in which the 3D stereo display group responded that they enjoyed the graphics and imagery more ($\bar{x}$ =4.4, σ=0.89) than the 2D display group ($\bar{x}$ =2.66, σ=0.58).

Stereoscopic 3D Questions

Three people from the 2D display group who played the five games in non-stereo then played one game in 3D stereo thought that the stereo provided them an advantage. Of the 20 participants in the 2D display group, four chose to play Resident Evil 5. Of those four people, two thought that the stereo 3D helped them. The other person who thought that stereo helped was one of the five people who chose to play Flatout: Ultimate Carnage. Of those same people who played in non-stereo, seven thought the technology hurt their performance when they got the chance to replay a game in stereo. Three of the seven participants who thought the technology hurt them were playing Left 4 Dead and another three of the participants were playing Flatout. The remaining participant who thought 3D stereo hindered their performance was playing Resident Evil 5.

Of the participants from the 3D stereo display group that played all the games in stereo, 10 of them thought that it gave them an advantage in at least one of the games, while seven of them thought that it hurt them in at least one of the games. In this group, nine participants thought the technology helped them in Left 4 Dead, eight thought so in Resident Evil 5, five felt it benefited their performance in Flatout, and another 5fivefelt the same in MLB 2K9. In the same group, three felt it hurt them in Left 4 Dead, three more participants felt it hurt their performance in Resident Evil 5, and another participant thought it hindered them in MLB 2K9. No participants from the 3D stereo display group thought that 3D hurt their performance in Flatout, and no participants thought that it helped or hurt them in Madden NFL '08.

Despite the fact that 3D stereo did not seem to impact performance and had very little impact on how the participants rated their experience with the games, the participants still preferred to play in 3D stereo. As part of the questionnaire relating to the stereoscopic aspect of the study, participants responded to four statements on a 7 point Likert scale (1 = Strongly Disagree, 7 = Strongly Agree). Participants agreed that 3D stereo improved their experience ($\bar{x}$ =5.13, σ=1.40), they would choose to play video games in 3D stereo over the 2D display ($\bar{x}$ =5.13, σ=1.47), and that it enhanced the sense of engagement they felt ($\bar{x}$ =5.58, σ=1.11). Although preferred, participants responded that it was not a necessity ($\bar{x}$ =3.85, σ=1.51).

DISCUSSION AND FUTURE WORK

Although there seemed to be a few tasks in these games in which participants could gain an advantage from added depth perception (e.g., tasks like hitting a baseball and aiming an arrow in 3D space), the quantitative data showed that 3D stereo did not provide any significant performance benefits to the user. One reason is that current video games are adapted to work with 3D stereo, but not built to take advantage of 3D stereo in terms of game mechanics and user interface. Additionally, with almost every game on the list of the best games in 3D stereo, there were settings that needed to be adjusted to maximize the 3D aspect or reduce artifacts that it would create. Most of the games needed shadows to be turned off as they would not be rendered correctly with 3D stereo enabled. Work has shown have that while not as beneficial as stereo, other depth cues such as shadows can increase the benefits for some 3D tasks [6]. Even with the settings tuned the way they were requested, some games still had noticeable glitches at times. Left 4 Dead would periodically create a flash effect in which the screen would go really bright if a light caught the camera the wrong way, which was hard on the eyes.

Another factor may have been the controls. Studies have shown that the interaction devices can have a significant impact on user performance [9]. This factor was most evident in the beginner group of participants, as it appeared that a lot of times they were struggling much more with manipulating the controls than anything that was being viewed on the screen. In addition, the controller used in our study was a standard Xbox 360 controller. Other controllers that provide 3D spatial input such as the Playstation Move or Microsoft's Kinect device could provide users with control mechanics that are more conducive to performing game tasks in 3D stereo.

One possible reason we did not see the benefits that other research has shown with 3D stereo may be that in video games, the tasks are not as cleanly isolated and evident. There is much more going on in the environment and the scenes are much more complex with a significant amount of animations than most previous research on the topic. Left 4 Dead and Resident Evil 5 are the two notable examples from our study.

We did see some differences start to show up when we looked at the possible learning effects taking place in the games. It was clear that learning occurred in most of the games for users regardless of whether they were viewing in 3D stereo or not. What is interesting in these results is that in three of the games, the learning that occurred was greater for users viewing in 3D. Thus, it is possible that 3D stereo may help users in learning the game environments or tasks in the games. Similar results have been shown by other research in which it was determined that 3D stereo allowed users to grasp larger, more complex scenes with more understanding [17]. In those studies, the benefits were only increased as the resolution of the display increased. We were running the study on a 3D TV in 1080p with 120Hz refresh rate. 3D TVs are starting to come out now with 240 and 480Hz refresh rates, so more benefits may occur as display technology advances allowing for more detail to be shown as well as motion on the screen to be cleaner.

As for the qualitative data, we had thought we would see more of a difference in the responses for the 3D stereo group. We assumed this based on previous work in which user preference was clearly for 3D stereo such as the U-Decide initiative released my Meant to be Seen 3D [10]. Overall, it did not appear that the user's perception of their performance was affected by 3D stereo as there were not very many instances where any of the qualitative data was found to be statistically different between the 2D display and 3D stereo display groups. One reason for the difference might be because of the relatively short time that our users played each game. With such short play times, it may have been difficult for them to become immersed in the game whether or not they were viewing it in stereo.

Research that has studied the benefits of 3D stereo has compared it to normal viewing as well as viewing with head tracking [2]. An important area for future work is to explore how head-tracking with and without 3D stereo affects user performance and learning with modern video games. From our results, we also postulate that if games are designed with 3D spatial tasks in mind (e.g., users must actually move in 3D space), 3D stereo could provide performance benefits, especially if 3D spatial input devices are used. We plan to explore this idea in future studies.

CONCLUSION

We have presented a study exploring whether 3D stereo provides performance benefits to users playing modern video games. Overall, the results showed that 3D stereo did not provide any significant advantage in performance over a 2D display and learning rates were comparable between display modes for the games we tested. Our qualitative data suggests that there was not much difference in perception of the gameplay experience between the 3D stereo display group and the 2D display group. Despite these results, participants indicated that they preferred playing the games in 3D stereo over playing on the 2D display. Our study indicates that the status quo for video games that make use of 3D stereo does not provide any additional user performance benefits. This means that game developers will need to develop new strategies for game interaction and potentially game narrative if they want to have 3D stereo provide any performance benefits to gamers. It remains to be seen if user performance gains can be made with games specifically designed for 3D stereo coupled with 3D spatial interaction.

ACKNOWLEDGEMENTS

This work is supported in part by NSF CAREER award IIS-0845921 and NSF Awards IIS-0856045 and CCF-1012056. We wish to thank the anonymous reviewers for their valuable suggestions.

REFERENCES

1. Arns, L., Cruz-Neira, C., and Cook, D. The Benefits of Statistical Visualization in an Immersive Environment. In *Proceedings of the IEEE Virtual Reality* (1999). VR. IEEE Computer Society, Washington, DC, 88.
2. Arthur, K. W., Booth, K. S., and Ware, C. Evaluating 3D task performance for fish tank virtual worlds. *ACM Trans. Inf. Syst.* 11, 3 (1993), 239-265.
3. Fujimoto, M. and Ishibashi, Y. The effect of stereoscopic viewing of a virtual space on a networked game using haptic media. In *Proceedings of the ACM SIGCHI international Conference on Advances in Computer Entertainment Technology* (2005). ACM, New York, NY, 317-320.
4. Grossman, T. and Balakrishnan, R. An evaluation of depth perception on volumetric displays. In *Proceedings of the Working Conference on Advanced Visual interfaces* (2006). ACM, New York, NY, 193-200.
5. Gruchalla, K. Immersive Well-Path Editing: Investigating the Added Value of Immersion. In *Proceedings of the IEEE Virtual Reality* (2004). VR. IEEE Computer Society, Washington, DC, 157.
6. Hubona, G. S., Wheeler, P. N., Shirah, G. W., and Brandt, M. The relative contributions of stereo, lighting, and background scenes in promoting 3D depth visualization. *ACM Trans. Comput.-Hum. Interact.* 6, 3 (1999), 214-242.
7. Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., and Walton, A. Measuring and defining the experience of immersion in games. *Int. J. Hum.-Comput. Stud.* 66, 9 (2008), 641-661.
8. Jones, J. A., Swan, J. E., Singh, G., Kolstad, E., and Ellis, S. R. The effects of virtual reality, augmented reality, and motion parallax on egocentric depth perception. In *Proceedings of Applied Perception in Graphics and Visualization* (2008). ACM, New York, NY, 9-14.
9. McMahan, R. P., Gorton, D., Gresock, J., McConnell, W., and Bowman, D. A. Separating the effects of level of immersion and 3D interaction techniques. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology* (2006). ACM, New York, NY, 108-111.
10. Meant To Be Seen. The U-Decide Initiative. http://www.mtbs3d.com/findings.shtml, 2010.
11. Menendez, R.G.; Bernard, J.E., "Flight simulation in synthetic environments," *Aerospace and Electronic Systems Magazine, IEEE* , vol.16, no.9, pp.19-23, 2001.
12. Pausch, R., Proffitt, D., and Williams, G. Quantifying immersion in virtual reality. In *Proceedings of the 24th Annual Conference on Computer Graphics and interactive Techniques* (1997). ACM Press/Addison-Wesley Publishing Co., New York, NY, 13-18.
13. Treadgold, M., Novins, K., Wyvill, G., and Niven, B. What Do You Think You're Doing? Measuring Perception in Fish Tank Virtual Reality. In *Proceedings of the international Conference on Computer Graphics* (2001). IEEE Computer Society, Washington, DC, 325.
14. Teather, R. J. and Stuerzlinger, W. Guidelines for 3D positioning techniques. In *Proceedings of the 2007 Conference on Future Play* (2007). ACM, New York, NY, 61-68.
15. Terlecki, M.S. & Newcombe, N.S. How important is the digital divide? The relation of computer and videogame usage to gender differences in mental rotation ability. *Sex Roles (2005)*, 53, 433-441.
16. Ware, C. and Lowther, K. Selection using a one-eyed cursor in a fish tank VR environment. *ACM Trans. Comput.-Hum. Interact.* 4, 4 (1997), 309-322.
17. Ware, C. and Mitchell, P. Reevaluating stereo and motion cues for visualizing graphs in three dimensions. In *Proceedings of Applied Perception in Graphics and Visualization* (2005). ACM, New York, NY, 51-58.