

# Sony EyeToy<sup>™</sup>: Developing Mental Models for 3-D Interaction in a 2-D Gaming Environment

Geanbry Demming

Amberlight Partners, Ltd, 3 Waterhouse Sq, 142 Holborn  
London, SW15 6EL, UK  
[gigi@amber-light.co.uk](mailto:gigi@amber-light.co.uk)

**Abstract.** The Sony EyeToy<sup>™</sup> is a webcam attachment for the Playstation2 (PS2) gaming console. The EyeToy<sup>™</sup> places a user's mirror image on a television screen while it detects their movements in order for them to manipulate objects and characters within a game. Although EyeToy<sup>™</sup> is immediately popular due to the novelty of interaction, users often have trouble forming a mental model of how to manipulate their own image in a 2-dimensional (2-D) 3<sup>rd</sup> person gaming scenario. Multiplayer testing sessions were carried out to explore the impact that a game's theme and a user's age and previous gaming experience would have on how users initially cope with this form of gaming interaction. A lack of tangible feedback and unrealistic game scenarios were the main hindrances to users reconciling the fact that they were in a 3-dimensional (3-D) space but had to behave as 2-D characters.

## 1 Introduction

Commercial gaming has seen a trend in the increasing demand for games that require physical interaction, mainly musical in nature. Users are looking more and more to full body movement to entertain themselves in the commercial world of video games. "Dance Dance[sic] Revolution" is a highly popular arcade game that's spawned an increase of dancemat purchases for home entertainment (Mixed Reality Lab). There is even an increasing amount of tournaments dedicated to dancing games (Diniz-Sanches, 2004). Users can also become virtual musicians by shaking maracas or playing with digital turntables (Walker, 2003). As removed as these games may seem from traditional handheld-controller gaming, they still rely on tactile interaction.

The EyeToy<sup>™</sup> is a product developed by Sony Computer Entertainment Europe (SCEE) for the Playstation 2 games consoles. The EyeToy<sup>™</sup> webcam attaches to the USB port of a PS2 gaming console. Users stand in front of the camera and see their inverted image (as if looking in a mirror) on screen. The camera then detects any movement, allowing users to manipulate various screen elements – effectively putting users in the game by having their movements dictate play. Considering the lack of commercial access users have had to this form of interaction<sup>1</sup> until recently, this research was conducted to investigate the usability and general appeal of EyeToy<sup>™</sup> and the games within the "EyeToy<sup>™</sup>: Play" package.

### 1.1 Nature of Interaction

Computer supported collaborative play (CSCP) is defined as "computer technology that enhance[s] physical exertion, social interaction, and entertainment in sport and play" (Ishii et al in Mandryk & Maranan, 2002). While EyeToy<sup>™</sup> descriptively fits in the CSCP paradigm, the games readily associated with CSCP (e.g. PingPongPlus or The Pirates) require tactile manipulation of artefacts to promote physical gameplay (Mandryk & Maranan, 2002). Therefore, EyeToy<sup>™</sup> presents an opportunity to define use cases, barriers to task completion, and determiners of a pleasurable experience that previous gaming research may not be able to account for.

With EyeToy<sup>™</sup>, whatever existing mental models about physical movement users adapt for gameplay can be complicated by the fact that, in several games, characters respond in a 3-D manner. For

---

<sup>1</sup> Logitech offered similar software with some of their webcams but gameplay and interaction is extremely limited compared to EyeToy<sup>™</sup>.

example, in the game “Kung Foo”, if a player hits a character, one of the programmed visual effects is having the character fly towards them and hit the screen. Users could simply flick a finger to their side to elicit this same type of behavioural feedback from a character. However, users could easily mistake a swift forward striking motion as the cause of a character behaving in this manner. Somewhat misleadingly, the game does not actually provide an analogous 3-D response to users 3-D actions.

Intuitive interaction is also hindered by the fact that there is no tangible feedback for EyeToy™. For more traditional games in which users’ input is mediated through handheld controls, or a dancemat, users can more readily associate their tactile interaction with the feedback on the screen (Diniz-Sanches, 2004). With EyeToy™, however, the visual and aural feedback users receive is not an absolute indicator by which they can determine the exact timing and 3-D location of their input controls (i.e. body parts). The exact path from cause to effect is further camouflaged when taking into consideration interference from other moving objects and persons visible to the camera, or the lack of contrast between the user’s skin/clothing colour and their background. Both instances make it difficult for the camera to detect movement and subsequently cause confusion amongst users as to why the game is not behaving in the way they anticipated.

## 1.2 Games Themes

The set of 12 games known as “EyeToy™: Play” was used for testing. The games range both in theme and the nature of the movement that they require. All the games in the Play package can be divided across 2 high-level categories in terms of their objectives: Realistic and Abstract. Realistic games easily map onto real world scenarios or actions popularised by television or films (e.g. Kung Foo or Boxing Chump). Abstract games are those that bare no relationship to a real world scenario or situation (e.g. UFO Juggler). Realistic games allow for users to access inherent mental models of movements in order to emulate realistic scenarios. For abstract games, users were initially less sure of how to behave or move without the aid of a tutorial or help.

There also appears to be an unintentional correlation between game themes and the nature of the movement required for *intended* gameplay [for games that do not require precision hand movements, it is easy to cheat simply by standing within 1 meter of the camera while waving arms frantically]. Realistic games most often require fast long movements with the arms. Abstract games most often require precision hand placement. Given the limited number of games available for testing, it is hard to say if this trend is significant. It is addressed for the purpose of investigating the impact the type of movement required might have on users’ ability to intuitively engage in gameplay. Table 1 describes how each game fits into a matrix of game themes and the movement required for intended gameplay.

**Table 1.** The 12 Play games classified in terms of theme and type of movement required for successful gameplay.

Games theme	Long quick motions	Short quick motions	Precision hand placement
Realistic	Kung Foo, Boxing, Wishi Washi, Keep Ups	Plate Spinners	
Abstract	Beat Freak, Slap Stream	Ghost catcher, UFO Juggler	Rocket Rumble, Mirror Time, Boogie down, Slap Stream

## 2 Method

Testing was aimed at 3 objectives: 1) scoping initial reactions towards the technology and general appeal of EyeToy™ 2) determining users mental models for navigation and gameplay 3) determining to what extent the factors that contribute to a pleasurable experience with EyeToy™ differ, if at all, from those for traditional console games.

Several assumptions were made in order to formulate a viable framework for testing. While players generally play computer games on their own (Mandryk & Maranan, 2002), EyeToy™ best affords

social gameplay – even though the multiplayer mode is designed for asynchronous inter-player gameplay. Therefore, groups of 4 users were recruited for multiplayer testing sessions. Four was determined as an ideal number: a) to foster a highly sociable environment and b) to accommodate what would customarily be the absolute maximum number of players normally allowed for co-located multiplayer console games. Finally, each session was held at the home of one of the users taking part in the session. All participants within a session were also acquainted with one another to promote a comfortable social environment for testing.

## 2.1 User Groups

Children and teenagers were SCEE's assumed target audience, due to the look and feel of the games to be tested. However, groups for testing were extended to incorporate adults to see what appeal EyeToy™ might have for them. The final groups were: Children aged 5 – 10; Children aged 11 – 14; Family of 4; University Students (aged 18 – 25); Young professional men (aged 24 – 28). Aside from the young professional men each group had an equal male-female ratio and all groups contained 4 participants. Young professionals was delineated as a same sex group based on fact that even though 46% of console purchases are by women, women only account for 28% of regular games players (Hafer, GamesFirst.com).

Users were also recruited on the basis of their previous console gaming experience ranging from novice (play console games at *most* once a month) to 'hard-core' gamers (play at least 3 times a week and across various gaming genres). The definition for novice and hard-core gamers was decided in conjunction with SCEE. Finally, to capture novice feedback and requirements, all users had no previous experience playing with EyeToy™.

## 2.1 Session Agenda

Each testing session lasted for approximately 90 minutes. The majority of each session (approx. 60mins) was devoted to users interacting with EyeToy™ with little to no interference from the facilitator. Users were given an 'out of the box' scenario. Their foreknowledge about the nature of interaction was limited to understanding that they will see themselves on the television and they can use their various body parts to play the game. Users were then left to their own devices to navigate through the menu structure and to play games both of their choosing and those chosen by the facilitator. The facilitator noted comments and behavioural signs of frustration and positive responses.

## 3 Reactions to Technology

The novelty of interaction was well-received by all participants – regardless of their statistical background. There is not enough conclusive evidence to suggest gender has an effect on users understanding of the nature of interaction, despite males' inherent advantage over females when it comes to spatial manipulation (Kimura, 2001). However, age and previous gaming experience appear to be the two main contributors to how readily users understand the nature of interaction needed to play EyeToy™ games.

Users who have difficulty with EyeToy™ appear to experience a conflict between seeing themselves on screen, yet having to ultimately play in a 3<sup>rd</sup> person scenario. In other words, they become the avatar they have to manipulate, but not by means of traditional analogue control mechanisms.

### 3.1 Impact of Age

Most of the children (aged 5-14) saw EyeToy™ almost as a touch screen interface. The more negative feedback children would receive from the game, the more they approached the screen, until they were actually touching it. This suggests they may have been trying to resort to the more familiar tactile interaction they are accustomed to with traditional video games or most other interfaces they encounter – i.e. they feel if they apply physical pressure to a mechanism of control they will see a response on screen.

Furthermore, for children, seeing their image on screen was often incidental. All their attention was devoted to using the characters and elements in the games as spatial cues about where to place their hands in a 3-D space, not knowing they needed to manipulate their image in a 2-D plane. Therefore, they concentrated solely on where the characters were being generated from, rather than the relationship between their image on the screen and the characters. At times only the top of their heads would be visible on screen, and yet they would still exhibit extreme concentration while enthusiastically trying to bat away characters.

### 3.2 Impact of Previous Gaming Experience

Those users with more previous console gaming experience performed overwhelmingly better at games than their novice counterparts. It may not be surprising that those with more extensive gaming experience would be better at gameplay due to their relatively sharpened skills in hand-eye coordination. However, the concept of their relationship to the characters on screen should still be fairly abstract as they are accustomed to an avatar performing on their behalf not manipulating their own image in real-time. Nevertheless, they tended to pick up on the proper use of EyeToy™ more readily than their novice counterparts. One suggestion is that hardcore gamers are used to seeking out gaming cues to frame interaction – they may more readily notice the positioning outlines players are supposed to stand in or the fact that their head accidentally selected a button.

### 3.3 Impact of Games

Table 2 reflects the popularity of the 12 games, across all users, within the context of their demarcation in Table 1. There was very little variation between user groups when it came to gaming preference. Kung Foo and Boxing Chump were unanimously listed as the 2 most popular games. Some games with mixed popularity, such as Beat Freak, did not elicit a strong positive or negative response from users in terms of preference. For other games, which received a mixed response, some users really enjoyed a particular game while others disliked it (e.g. Slap Stream).

**Table 2.** Game popularity based on game theme. 1 = highly popular; 2 = mixed popularity; 3 = unpopular.

Games theme	Long quick motions	Short quick motions	Precision hand placement
Realistic	1	3	
Abstract	2	3	2

Users' preference in terms of movement had more to do with physical fatigue than how intuitive it was to engage in game play. In contrast, their preference in game themes actually had an impact on how readily they came to understand the nature of interaction needed for successful gameplay.

Users performed best on the games Kung Foo, Boxing Chump and Wishi Washi (objective is to clean as many windows as possible) simply by emulating the motions they would expect to execute in a comparable real world situation. There is not enough evidence, however, to discount the fact that the direction and timing of characters' behaviour in these games did not affect users' level of understanding about intended gameplay. For example, in Boxing Chump, users are encouraged to fight in a 2-D plane due to the positioning of the robot they are sparring with and the visual feedback for successful blows. The robot never falters to the left or the right, but always doubles over in a way that suggests it is being hit in the stomach, directly from the front. Hence, users are given cues that punching directly at the robot, and not to its side, is the best way to defeat him.

### 3.4 Bridging the 2D/3D Divide

A user's statistical background (e.g. age and previous gaming experience) cannot be ignored when it comes to how they will cope with trying to form a mental model for 2-D interaction in a 3-D world.

However, in terms of future development of games that will sit on this type of interaction, the findings of this study point to 2 key requirements to promote the successful acquisition of mental models for gameplay by novice users:

1. Develop games around real world and practical scenarios
2. Do not provide visual cues that appear to respond as a 3-D analogue to users' movements

The first requirement allows novice users to bring to the game pre-existing mental models of movement and use them to their advantage for gameplay. The second requirement acts as a guidepost for interaction, albeit retroactively at first, by illustrating how best to get a reaction from characters and other gaming elements. The second requirement also opens the door for games with more abstract themes by easing users into developing a mental model of 2-D movement.

## 4 Future Research

For users that have trouble with the relationship of their image to the characters on the screen, it would be of interest to see if creating an avatar that responded directly to users' movements would work to reconcile the conceptual rift between their 3-D environment and the game's 2-D world. The integration of EyeToy™ with other accessories is also an area of interest. Children had the hardest time grasping the concept of interaction needed, potentially due to the lack of tangible feedback. The fact that the most highly quoted suggestions from children were the integration of a dancemat for dancing games and a number of fighting accessories for combative games lends credence to this argument.

Also of interest is exploring various themes for EyeToy™ games and their subsequent effect on users' understanding of intended gameplay. Many users expressed an interest in having more games relating to sports themes or an exercise video based around EyeToy™ technology. Games with these themes have an inherent set of rules and movements that are implicit to most users, which could allow for a more intuitive interaction. Finally, continued investigation is needed for the visual and aural feedback users receive to drive gameplay. Continuing with the aforementioned example of the sparring robot in Boxing Champ, its responsive behaviour was analogous to the type of movement the user needs to employ to consistently score points. This could explain why users preferred and performed better on this game. However, in order to isolate this type of feedback as having a significant impact on how users come to understand intended 2-D interaction, further testing needs to be carried out.

## 5 Conclusion

The nature of interaction mediated by the EyeToy™ has opened up numerous opportunities for growth in the gaming industry in terms of broadening the appeal of games and subsequently the gaming market. Its appeal is also strengthened by providing domestic access to games that aid in alleviating the sedentary lifestyle that is often associated with long-term game use and its subsequent effect on the health and social interaction of gamers – especially children. Optimising game design to make for a more intuitive and compelling interaction should then be a priority to promote more sociable and healthy gameplay.

## References

1. Diniz-Sanches, J. EyeToy:Groove. *Edge* 132 (2004), 103.
2. Hafer, M. GamesFirst.com  
<http://www.gamesfirst.com/articles/monica/womenwant/womenwant.htm>
3. Kimura, D. Biological constraints on parity between men and women. *Psynopsis* 23,1 (2001), 3.
4. Mandryk, R. and Maranan, D. False prophets: Exploring Hybrid Board/Video games. *Proc. CHI 2002*, Interactive Poster: Fun, 640-641.
5. Mixed Reality Research Lab – Research – Human Pacman.  
<http://mixedreality.nus.edu.sg/research-HP-inform.htm>
6. Walker, I. (2003). Gamerswave.com. <http://www.gamerswave.com/musicspecial.htm>