

# Chapter 7

## Video Game Development and User Experience

Graham McAllister and Gareth R. White

**Abstract** The first step in understanding the user experience needs of the video games industry is to ascertain current practice. The following chapter gives an overview of the game development process and provides background on the time frame and roles involved. We present case studies from three world-class development studios and show how the user experience is currently addressed during a game's creation. The first case study with Disney's Black Rock Studio details the development of their most recent racing game, *Pure*, and describes the usability testing which the developer believes improved the game's Metacritic score by 10%. The second case study with Zoë Mode refers to several of their recent releases, *Rock Revolution*, *You're in the Movies*, and games in the *Eye Toy* series. Special consideration is given to understanding and addressing players in a language appropriate to their background as gamers. The third case study with Relentless Software concentrates on the studio's use of focus group testing and attention to the casual gamer demographic. In addition to showing real-world examples of current practice, this chapter identifies the contribution that HCI can make for user experience methodologies in the games industry. Recommendations are made for generally applying usability techniques earlier in development, and user experience testing later once a playable vertical slice is available. We conclude with some discussion of innovative methodologies and pose the need for a formalised framework for user experience in video game development.

### 7.1 Introduction

In order to design new methodologies for evaluating the usability and user experience of video games, it is imperative to initially understand two core issues. First, how are video games developed at present, including aspects such as processes and time scales and second, how do studios design and evaluate the user experience?

---

G. McAllister (✉)  
University of Sussex, Brighton, UK  
e-mail: g.mcallister@sussex.ac.uk

This chapter discusses the video game development processes and practices that studios currently use to achieve the best possible user experience. It will present three case studies from game developers Disney Interactive (Black Rock Studio), Zoë Mode, and Relentless Software, all based in Brighton, United Kingdom. Each case study will detail their game development process and also how this integrates with the user experience evaluation. In an attempt to represent a balanced view of state of the art in game development practices, the game studios chosen focus on different game genres and target user groups.

Reader's takeaway:

- Three concrete case studies of how video games are developed at world-leading studios
- Clear understanding of the game development life cycle
- Understanding of industry terminology, laying the foundations for a common language of user experience
- Understanding of industry needs, in terms of what they expect and require from usability and user experience evaluations

In summary, the key contribution that this chapter makes to the games usability community is an understanding of the game development process and how these studios currently involve the end user.

## 7.2 Previous Work

Although the topic of evaluating video game user experience is gaining more attention from both academia and industry, it is not a particularly new area. One of the earliest papers (Malone 1981) discusses which features of video games make them captivating and enjoyable to play. Today, this discussion still continues, and there is active research in determining which game features to evaluate and which approaches should be used.

Current approaches to evaluating the usability and user experience of video games have centred around themes such as mapping established HCI methods to video games, refining these methods, identifying guidelines and, perhaps most importantly, evaluating the overall player experience. A summary of relevant literature will be discussed below.

### 7.2.1 *Traditional HCI Approaches*

Due to the generic nature of the majority of usability methods, researchers have analysed how existing usability methods can be applied to video games (Jørgensen 2004). Others such as Cornett have employed usability methods such as observations, questionnaires, think aloud and task completion rate to determine

if they would be successful in identifying usability issues in MMORPGs (Cornett 2004). Without much, if any, modification, there is evidence to support the claim that conventional usability techniques can be successfully applied to video game evaluation.

### ***7.2.2 Refining Traditional Methods***

Although established usability methods can be directly applied to games, Medlock and others at Microsoft Game Studios have developed a usability process which is specific to games (Medlock et al. 2002). Their approach is called the Rapid Iterative Testing and Evaluation (RITE) method and although it is very similar to a traditional usability study, there are two key differences. First, the process is highly iterative, meaning that whenever a fault is found and a solution identified, it is immediately corrected and re-tested with the next participant. Second, the usability engineer and design team identify and classify each issue to determine if it can be resolved immediately, or if further data are needed before a solution can be found. This can be thought of as usability “triage”.

### ***7.2.3 Heuristics***

Nielsen’s usability heuristics (Nielsen 2005) have long served as general guidelines for creating usable applications or websites. However, various researchers (Federoff 2002, Desurvire et al. 2004, Schaffer 2008, Laitinen 2006) have constructed sets of heuristics which are specific to video games and compared their effectiveness to Nielsen’s. They found that their heuristics are most useful during the early phases of game development. Despite the existence of specific game heuristics, questions remain about their specificity and utility (Schaffer 2008), and feedback from developers suggests that they are too generic to be of much use.

### ***7.2.4 User Experience***

According to Clanton, the overall deciding factor of a good game is game play (Clanton 1998). Trying to specify what makes a good game is not a straightforward task, and Larsen has tried to unpack this problem by examining how professional game reviewers rate games (Larsen 2008).

Others have addressed the key criticism of heuristics by wrapping them up in a unified process (Sweetster 2005). This process, which they call GameFlow, can be used to design, evaluate and understand enjoyment in games. Meanwhile, Jennett has conducted a series of experiments to measure and define the immersion in video games (Jennett et al. 2008). The concepts of immersion and flow in games are often related to involvement or enjoyment, but they appear to have

subjective and imprecise definitions, thus making them difficult to design for and measure. Furthermore, flow seems to be applicable only to describing competitive gaming, despite other research pointing to the diversity of emotions a player can experience during play (Lazzaro 2004), and indeed the diversity of players and games (Schuurman et al. 2008).

### ***7.2.5 Game Development***

Most research in the area measures games towards the end of the development life cycle. Although this may be suitable for fixing small changes on time, it is not sufficient for altering key game mechanics. If new techniques are to be designed, which can evaluate a game during its development cycle as well as the final product, then a better understanding of the development life cycle needs to be obtained.

One reason for the lack of tailored techniques which could be applicable to all stages of game development is that the game development process itself is not known in detail to the HCI community. Federoff's work shadowed a game development team for 5 days and has reported some details of the development process (Federoff 2002). However, the main focus of this research was to construct game heuristics, not report on the development process per se.

The next section discusses the general characteristics of game development including the development life cycle and relevant industry terminology.

## **7.3 Introduction to the Game Development Life Cycle**

In the video game industry, it is nominally accepted that the development life cycle is constructed from the following phases, though in practice those occurring prior to production are often contracted or skipped entirely.

### ***7.3.1 Concept***

Game concepts can be initiated either from the publisher, who provides finance, or from the development studio, which is responsible for the day-to-day production of the game. Once a general concept has been agreed between the two parties, a small development team of around five staff may spend 1–2 months producing an initial Game Design Document and visual representations such as still images or a movie to communicate the vision for the game. Additionally, a rough budget and plan is produced, including milestone agreements which define the production commitments of the developer and the corresponding financial commitments of the publisher. This would normally represent a phased or iterative delivery of the product, where only a first-pass of each feature is completed prior to evaluation and feedback from the publisher. Later in the schedule, a second delivery is made, which is a more concrete

implementation. Agreements made at this stage are still subject to adjustment at any future point.

### ***7.3.2 Prototyping***

During the early stages of development, many different aspects of the game may be prototyped simultaneously and independently. These provide examples of features such as menus, physics and vehicle handling, or could be technical demos such as grass rendering or other components of the game or graphics engine. In order to define a visual benchmark, the art team may construct virtual dioramas, which are models of events that players will experience during the game. Some of these prototypes could be interactive, others could be non-interactive movies demonstrating an example from which the interface for this part of the game could be developed.

This initial phase can take between 3 and 6 months, by the end of which these prototypes and concepts are evaluated, and if the project is given a green light, it moves into pre-production.

### ***7.3.3 Pre-production***

Following design approval, the game development team enter the important pre-production phase, during which time fundamental game mechanics are proven and problematic areas are identified. The purpose of this phase is to try out ideas quickly without getting bogged down in issues of final presentation quality, to identify risks and prove the important aspects of the game concept.

### ***7.3.4 Production***

During the main production phase, the team will be scaled up to full size and would tend to spend in the order of 12 months producing all of the characters, levels, front-end menus and other components of the game. Often during this stage, the team will produce a “vertical slice”, which is a high-quality, 10–15-min demonstration of a small sample of the game.

In addition to the core team of programmers, artists, designers and audio engineers, game developers also include a Quality Assurance (QA) group who are responsible for testing the game. This is essentially functional testing rather than usability or experiential testing. The QA team are keen gamers with a good understanding of the market and what to expect from a high-quality game. As such, in addition to functional bugs which are entered into a database and addressed in a formal process, testers may also identify “playability” issues which are informally discussed with the rest of the development team. Final issues of usability and playability are the responsibility of the producer and designers. QA teams are

often only scaled up to full size towards the end of the production phase, through Alpha and Beta.

### ***7.3.5 Alpha – Beta – Gold***

Towards the end of production, the game progresses through a series of statuses which indicate how close development is to completion.

In order to achieve Alpha status, all content in the game should be represented, but not necessarily be of final quality. Placeholder content is common, but the game should exist as a coherent whole.

From Alpha, a further 6 months would typically be spent advancing through Beta status until the game is finally available for release, with attention turned to bug fixing and finalising the quality throughout.

By Beta all content and features should effectively be finished, with all but final polishing still to take place. Nothing further will be added to the game, only tweaking and final adjustments. In particular, this phase is focussed on bug fixing. After Beta, the developer and publisher consider the game to be of a shippable quality and submit a Master candidate disc to the format holder (i.e., Microsoft, Nintendo or Sony) for approval.

Each game that is released on any of their consoles has first to be approved by the format holder's own QA team. Strict technical and presentation standards define how all games on the platform should deal with issues of brand recognition as well as certain HCI guidelines. For example, which controller buttons to use for navigating dialogue boxes, the format and content of messages to display to the player while saving games, or where to position important interface elements on the television screen.

The approval process usually takes 2 weeks, but in the event that the submission candidate fails, a further submission will have to be made once the development team have resolved all of the faults. How long this takes depends on the severity of the issues, but once the team have successfully dealt with them, another 5–10 days will be required for the re-submission approval process. Conceivably, further submissions could be required until all issues have been resolved.

Once approval has been given, the publisher uses the Gold Master disc to begin manufacturing and distribution, which takes between 1 and 4 weeks. Typically, a unified release date is agreed upon with all retail outlets, which requires an additional week to ensure stock is delivered from the distributors to all stores in time for simultaneous release. In the United Kingdom, this results in retail outlets releasing games on a Friday.

## **7.4 Case Studies**

This section presents three case studies from world's leading developers Black Rock Studio (part of Disney Interactive), Zoë Mode and Relentless Software. Each studio will discuss their development process, time scales and how they involve end users.

### 7.4.1 Case Study 1 – *Black Rock Studio*

Black Rock Studio specialise in developing racing games for Xbox, PlayStation (PS) and PC. Their latest game, *Pure*, was released in September 2008 to critical acclaim. We interviewed Jason Avent, the Game Director of *Pure*, who attributes the high ratings to not only the talent of his team, but also the usability evaluations that were conducted during development.

#### 7.4.1.1 Game Development at Black Rock Studio

Game development at Black Rock Studio typically takes between 18 and 24 months, with a phase breakdown as follows:

- Prototyping (3–6 months)
- Pre-production (6 months+)
- Production (6–12 months)
- Alpha, Beta, submission and release (4–9 months)
- Testing by the format owner (10 days)
- Disc manufacture (2–4 weeks)

The total development time for *Pure* was approximately 20 months, which was at the lower end of the range of each phase. Delivering the product while keeping to the lower end of the range was attributed to the team's experience and their agile development process. Each of these phases will now be explained in more detail.

#### 7.4.1.2 Prototyping

During the prototyping phase, the publisher's marketing team employed a recruitment agency, which produced a detailed online questionnaire about the game concept and distributed it to approximately 100 people within the target demographic. Following the online survey, a focus group study was conducted with three groups of 4–5 participants to discuss the game concept. The study was run in a lounge environment, with two members of staff, one taking notes from behind a one-way mirror and the other sitting with the participants to facilitate the discussion.

The team also decided to build a "Pre-Vis" (pre-visualisation) prototype for one of *Pure*'s key features, player-customised vehicles. This took the form of a non-interactive movie showing each part of the bike attaching to a central frame. This was not a technical demo, but rather just a visual benchmark or reference from which the interface for this part of the game could later be developed.

#### 7.4.1.3 Pre-production

*Pure* had originally been intended to just be an incremental advance on the previous title in the series, but over the course of 7 months the game concept evolved through several different designs. The initial idea was to make the most authentic quad bike racing game on the market, but this was abandoned in favour of a concept

tentatively called *ATV Pure*, which avoided realistic sports simulation in favour of informal, social, off-road racing. The final concept, called simply *Pure*, was only settled upon by the time the team were halfway through pre-production. Each of the preceding versions contributed some design features to the final game, but many were prioritised so low that they were never included in the released title. The design strategy for *Pure* was to focus on implementing a few core features to a very high standard, rather than attempting many features to a lesser standard.

By 12 months into the development cycle, the team had fixed their vision for the game and were only addressing issues that supported four key aspects: massive airborne tricks, customisable vehicles, 16-player online games and so-called “FHM’s” (which for the purposes of this chapter we will describe as “Flying High” Moments). These FHMs represent the core experience of *Pure*: When the player drives round a corner at high speed, only to realise that what they thought was just a bump in the road turns out to be the edge of a cliff face, leaving them flying thousands of feet up in the air. This is not only a highly exciting part of the game, but also a key mechanic as it allows the player plenty of time to perform tricks in the air, which eventually results in a higher score and other rewards.

These concepts were graphically documented in a number of ways specific to Black Rock, but which have similar implementations in other studios. An important tool for summarising the concept of the game and keeping the team on track is the X-Movie, used by some of the world’s largest games publishers such as Electronic Arts. The X-Movie for *Pure* was a video showing a bike rider on an off-road race track jumping high into the air. Black Rock maintained the idea of “X marks the spot” when it came to their “Core Idea Sheets”. These were large display boards hung around the team’s development area showing the four key game aspects superimposed onto a bull’s-eye target. At the centre of the target was the FHM feature, as this was intended to represent the essence of the entire game. It is worth highlighting that this essentially puts the user experience (or excitement), as the single most important criteria for the game.

#### **7.4.1.4 Alpha to Release**

For *Pure*, Black Rock did not employ a QA team throughout the entirety of the project, instead they used only one or two testers in the month leading up to Alpha. From Alpha, this was increased to five staff to deal with both the Xbox 360 and PS3 versions of the game (which were simultaneously developed). Furthermore, the QA team were only concerned with addressing functional testing rather than questions of usability or user experience.

#### **7.4.1.5 Post-launch**

The publisher’s marketing team conducted studies after the game had been released and sold most of its units. The purpose of these studies was to identify what consumers liked or did not like about the game and what made them purchase it. Similar themes were discussed with consumers who did not purchase the game.



After release of the game, some informal analysis was conducted on official reviews and user comments on forums. Avent asserts that with the exception of some aspects of the garage, few usability issues were mentioned in game reviews. Most users' comments related to features that were intentionally excluded by the team for practical reasons (such as the excessive technical overhead of including replays and split-screen multiplayer).

#### 7.4.1.6 Understanding the User

*Pure* was the first title on which Black Rock employed usability tests. They began running tests with company staff, careful to choose people who were not in the development team. They then expanded to recruit other people who worked in the same building as them, but who were not part of their own company. Finally, the most substantial tests began with members of the public, recruited directly from the streets and local universities. In total, around 100 participants were involved over the course of 4 months, of which the final month was only concerned with quantitative analysis of the game's difficulty level and any issues that would prevent players from completing the game. The only requirements for recruitment were that participants were 14–24 years old, male, and owned at least one of the current generations of console, i.e. they were likely candidates to make a purchase. Tests were run in-house, in regular office space separated from the development teams. Up to eight players would be present simultaneously, with one supervisor for every two to three players. One of the supervisors was responsible for taking notes, and no video data were captured as this was considered too difficult to analyse due to the very large volumes produced.

Black Rock conducted “blind testing”, meaning that testers had never played the game before, and several different categories of testing were devised:

- Free flow: This is an unguided test where the player is encouraged to play the game however they wish. This is particularly useful for giving an impression of how much fun the game is, because as soon as the game becomes boring they would be inclined to stop playing.
- Narrow specific: In this mode, the player would only play a single level and they might play it multiple times in order to measure their improvement with familiarity. This appears to be similar to vertical prototype testing employed in usability evaluations.
- Broad specific: Similar to the narrow specific test, but playing over many levels. This seems similar to horizontal prototype testing in usability evaluations.

Most of the development team also had visibility of the playtests, but generally it was only designers who observed the sessions. Avent reflects that it may have been helpful for more programmers and artists to also have been involved with observation earlier on.

Despite the absence of video data, one of the programmers in the team had implemented “instrumentation” for the usability tests. This is the process of recording

quantitative data directly from the games console that describes the timings of events during the game session. This can be used to measure specific events such as lap times or how long it took to restart a game. Similar techniques have been employed by other studios such as Valve and Microsoft (Valve 2008, Thompson 2007).

At the start of a game, *Pure* does not present the player with the traditional selections of Easy, Medium and Hard difficulty levels, but rather dynamically adjusts the AI vehicles to suit the players' performance while they play. During the final 2 weeks of testing, the Black Rock team focused only on balancing this dynamic difficulty system. The team were able to approximate a learning or performance curve by comparing the players' finishing position after their initial and subsequent playthroughs of a given track. By tweaking the dynamic balance system, they were able to adjust this difficulty curve to keep players challenged and engaged enough to replay races and improve their performance.

Avent strongly believes that these tests were crucially important in order to ensure a Metacritic score of 85% and that without them he felt that the game would have been reviewed at around 75% instead. This is a strong recommendation for even simplistic and lo-fi usability testing.

However, reflecting on the quality of results, Avent does recognise some of the limitations with a principally quantitative approach. In particular, he comments that even with a large dataset, results could be misleading or misinterpreted. For example, if the data show that players consistently come last in a race, one response may be to reduce the AI vehicles' performance to be more in line with that observed from the players. However, this may be inappropriate if the real cause lies not with the AI over performing per se, but perhaps with bugs in the vehicle physics system which bias in favour of computational rather than human control.

In order to identify such cases, Avent recognises that qualitative approaches would be beneficial and this is where more effort will be focused in the future. He hopes that future projects will take the agile development model even further, incorporating user testing throughout the development life cycle, including during prototyping and pre-production.

Furthermore, usability testing could begin as soon as the team produce lo-fidelity prototypes. For example, an HTML mock-up of a front-end menu, a 2D interactive demo of vehicle handling and a white box of a level where the visuals are of the minimum possible standard in order to test game play (i.e., typically without colour and with purely blocky, geometric shapes). Indeed, while only a small amount of usability testing was carried out from the vertical slice, Avent believes that the team should be conducting at least one test per week from that point onwards, with two to three playtests per week by the time Alpha has been reached.

During *Pure's* development, the key team member to be formally concerned about usability was Jason Avent, the Game Director. While individuals in the design team specialised in difficulty and tracks, Avent talks about the possibility of hiring an HCI specialist in the future, who would be the team's user experience expert by the time a vertical slice is produced. Eventually, he imagines multiple usability designers for different aspects of the game, one to specialise in track design, others for difficulty curve, vehicles or avatar design (for more on this topic see

Isbister 2006.) While these may seem highly specific, the team already speak about a nomenclature for the design of tracks, a “track language” involved in the dialogue between game and player. An example of good communication in track language would be a vivid distinction between areas the player can drive on and areas they cannot. This “language” must be clear and “readable” to the player, otherwise their mental model of the game will be inaccurate. Of course, there are occasions when communication in this language is intentionally obfuscated, for example where there are secret shortcuts that should be indicated in a much more subtle way. Avent hopes that for future projects semiotic codes can be defined for each of these areas of game design.

#### **7.4.1.7 Pure Development Summary**

Off-road racing game for Xbox 360, PS3 and PC  
Target demographic: young, male console owners  
60–70 staff  
Agile development model  
20-month development lifecycle  
5 functional testers from Alpha  
100 playtesters  
Custom quantitative usability analysis  
Usability studies increased Metacritic score by 10%

### **7.4.2 Case Study 2 – Zoë Mode**

In 2007, Zoë Mode rebranded itself as a casual games developer, with a mission to become the world’s leader in music, party and social games. In general, the company tends not to design for a specific age group, but rather hopes to create games that anyone can pick up and play easily. Such a broad target can present obvious challenges, such as a development team who are largely in their 20s and 30s trying to design for consumers who could be 6–60 years old.

We spoke to Martin Newing, Executive Producer, Karl Fitzhugh, Design Director and Dan Chequer, Lead Designer about the studio’s recent games including *You’re in the Movies*, *Rock Revolution* and a number of titles in the *EyeToy* and *SingStar* series of games.

#### **7.4.2.1 Understanding the User**

The studio head, Ed Daly, was the core driver for involving usability and quality control in the studio’s development process, though a number of usability techniques came to Zoë Mode from outside the company, and Microsoft were a particularly strong influence. Despite Codemasters being the publisher, Microsoft were involved in the testing of *You’re in the Movies* as the game ran on their Xbox 360 platform. Fitzhugh comments that the quantification of results was particularly useful for the

team, with Microsoft presenting data in the form X out of Y participants found this aspect of the game problematic. Twelve groups of four participants were involved during the course of a week. Recognising that this is a small sample size, particularly given the intention to sell the game to millions of consumers, Fitzhugh would like to involve more people but does reflect on the difficulty of analysing a much greater volume of data. In particular, he comments that participant selection is critical, so while the kind of ad hoc testing they have performed with friends and families can be productive, formal testing should be of a much higher standard.

Some focus testing was also conducted with *EyeToy Play Sports*. Sessions were led and run by the publisher, Sony, at a location in London rather than in the developer's studio. Newing is keen to point out, as have others in our case studies, that publisher interest in this kind of testing is very welcome but currently uncommon. Furthermore, Newing is also aware that focus group testing can be misleading if the sample size is too small to draw general conclusions or so wide that the results suggest the game be reduced to the lowest common denominator. Chequer is also anxious about individual participants controlling focus group discussions by exerting their influence on other participants who may have valuable feedback which is never revealed. This is a concern for all focus group studies, however, we would argue that good moderation should be able to overcome this challenge and draw out all issues that any of the participants may have.

As with other studios in this series of case studies, Zoë Mode points out the problems of presenting pre-release games to focus groups. In particular, Newing relates an anecdote about showing an early version of the game to personal friends and family who were simply too distracted by the poor quality of early artwork to want to play the game.

Zoë Mode's latest release, *Rock Revolution*, incorporated observational playtesting as a key part of the development process. Chequer argues that running these studies was essential for Jam Mode, a more free-form music making part of the game. Although the team already had some ideas about aspects that needed to be changed, observing real players struggle gave them the impetus to actually make the changes before release. Indeed, the bulk of revisions came from focus group and playtesting, with Chequer mentioning two issues in particular: help text to explain what on-screen buttons do and further encouragement for first-time players to just get in and start making music. Despite *Rock Revolution* being a challenging project with some poor reviews, and only one iteration for testing, the team were pleased that a GameSpot preview praised the game's Studio, in which Jam Mode takes place, saying "The most fun we found is in the jam session".

#### 7.4.2.2 Game Language

Discussing the value of observing playtests, Newing and Chequer reflect on how easy it is for developers to overlook the simplest of things which can impede novice players from even starting to play. Like the other studios in this study, the team are now aware of the problems associated with using traditional terminology for games that are intended to appeal to a broader demographic.

Newing referred to the distinction between “Challenge Mode” and “Story Mode”, which are clear labels for the team who’ve discussed and created these different play modes, but for the player who’s just interested in playing for the first time they can be a source of uncertainty. Newing goes on to discuss the difference between the traditional “Easy”, “Medium” and “Hard” difficulty levels, commenting that typically these offer effectively the same game with more or less and harder or weaker enemies. Critiquing the terminology of difficulty further, Newing also refers to *Brothers In Arms – Hell’s Highway*, which offers “Casual”, “Veteran” and an unlockable setting, “Authentic”. However, the descriptive text for even the casual setting gives the impression that it is a challenging setting even for experienced players. The alternative he proposes is, rather than varying the number and strength of the opposition, to vary the game mechanics. This way, the easiest levels would accommodate only a simple subset of possible mechanics which could be opened up to the player at harder levels. A case in point is the Nintendo 64’s *Golden Eye* where higher difficulty levels provided extra challenges by adding more demanding success criteria. Levels could be completed by beginners, but on the more advanced settings players would be required to complete additional goals.

The *EyeToy* series of games gave Zoë Mode a particular challenge due to their unique input mechanism. Players interact with these games primarily through their real-world movements which are seen by a camera attached to the console. For gamers and non-gamers alike, this is a novel mode of interaction which introduces the potential for many problems. Chequer remembers one particular scenario where the player was in-game, then the game cut to a non-interactive animation before returning to interactive play. During the observational sessions, it became clear that the players did not understand the difference between the interactive and non-interactive sections and were unclear when their physical movements in front of the camera would have an effect or not.

Fitzhugh likewise discusses some of the challenges of *You’re in the Movies* and communicating technical instructions to the player. For this game, there is a calibration process which requires the player to evaluate whether the game has successfully identified and separated the camera’s image of the player from the background of the room they’re playing in. The developers refer to this process as “segmentation”, a technical term that comes from the field of image processing – a concept and term that’s almost certainly alien to players of this casual game. Despite using everyday language both in voiceover and on-screen texts, the team had to iterate through several different phrases during focus testing, for example “have you been successfully cut out from the background?” Unfortunately, even this apparently straightforward question was inappropriate for the audience who were concentrating more on the fact that they appeared on-screen than on attending to the game’s needs of identifying whether its algorithm had been successful or not. Finally, the team settled on the presentation of images showing examples of what successful and unsuccessful results would look like, which is effectively a tutorial or training session for the player. This is particularly significant for casual games which are intended to be played by anyone, and especially people unwilling to invest much time and effort in learning how to effectively use them.

### 7.4.2.3 Game Complexity and Accessibility

Note that in the games industry, the term “accessibility” usually does not refer to disability as is often the case in the HCI community, but rather any player’s initial contact with a game, and especially so for casual games (see also Desurvire and Wiberg 2008). Throughout our text here, we keep to this meaning.

Newing is keen to point out that while games should be open to play without having to read through complex instructions or manuals, having a degree of hidden depth behind the scenes is still important for the longevity of the title. Chequer sums up the issues by stating that games should try to avoid any issues that would block the player from play and should also provide interesting secondary systems and mechanics for advanced players. Fitzhugh goes on to point out the success of games like *Guitar Hero* which offer notoriously challenging difficulty levels for the most experienced players, but which also appeal to beginners on easy levels.

Chequer reflects that in retrospect some of the minigames in *EyeToy Play 3* required too much learning through trial and error before players could really experience them. In contrast, *EyeToy Play Sports*, which featured 101 minigames, lacked some of the depth but perhaps was more accessible to beginners. Finding the middle ground is where the art of balancing comes in, and we would suggest that usability and user experience testing provides a number of approaches to facilitate this.

Chequer points out that the most accessible games in the Wii are relatively instinctive and easy for beginners to play. This can be observed in *Wii Sports* where most of the games are based directly on real-life actions that non-gamers are familiar with, such as swinging your arm for tennis, and which serve to give the impression of a transparent interface. The boxing minigame, however, is significantly less accessible due to its more abstract input mechanism which responds less well to natural movements.

Another example of accessibility is navigation flow through menus. Proficient gamers are used to a certain set of conventions for menu screens, such as where to find controller options, etc. but observing non-expert players can reveal that this is a learnt association that may be at odds with the assumptions novice players bring with them.

### 7.4.2.4 Usability Tests

During 2005, Sony employed a usability company to run tests for *EyeToy Play 3* with children and families, which were observed by Chequer and others from the team. The results of the sessions were encouraging for the team both as a morale boost to show real players having fun with their work and as a keen insight into some big design flaws they hadn’t considered before. Unfortunately, these sessions occurred late in development as the game approached Beta and so the team didn’t have sufficient time to address some of the more significant issues. Newing mentions that the quality of reports from these sessions was very high, providing recommendations for the team in a non-prescriptive way. For the focus group sessions, reports also provided background on individual participants and interpretations for events

during play as well as a description of the overall mood. Finally, the large amount of video data gathered was invaluable for demonstrating and resolving problems with the game.

#### **7.4.2.5 Changing Demographic**

Chequer comments on the development of the industry as a whole and points out that in earlier times the market was predominantly made of a small core of people who were experienced players, of which the game developers themselves were part. In those times, it was relatively easy for developers to make games they liked and be more confident that they would appeal to the market, as the team represented that market (for more discussion of the background and effect of these “cultures of production”, see Dovey and Kennedy 2006). Now with a broader market and games that are particularly intended to be played by less experienced players, the distance between the developers and market means that fewer assumptions can be made and more attention has to be paid to testing. Chequer describes observing his own mother playing a game (see also the “mum-factor” in the Relentless case-study), who wanted to stop playing but wasn’t familiar with the convention that the “Start” button not only starts games but also stops them. These kinds of conventions can clearly be confusing to the casual audience who may need additional assistance and explanation.

#### **7.4.2.6 Studio-Wide Quality Review**

Currently, Zoë Mode do not have an official mandate to conduct usability studies, though a new initiative in the company does incorporate usability techniques as part of their new studio-wide quality review, which also includes focus group testing and their standard postmortem of the development process itself. This is a relatively new initiative which was begun only a year ago, and which itself is currently under review. Under this process, qualitative comments about the development process are collated and summarised, then anonymously reported back to the team. Previously, only senior management were involved in the process of deciding whether a game was of sufficient quality to ship, but a new model for this process additionally involves members of the team. These include senior staff such as the discipline leads, but some people from other teams in the company are also brought in for a fresh perspective. Newing points out that internal reviews can be problematic, whereas bringing in external reviewers helps to provide fresh, impartial and unbiased assessment. We would agree and further recommend considering feedback from players *external to the game development industry*.

#### **7.4.2.7 Postmortem**

Games finished in the previous 2–3 years have run postmortems, but not in a standardised manner that would allow the team to quantify and compare their successes and failures with previous projects. The definition of a standardised postmortem

template is one of the goals of the quality review process. Typically, Zoë Mode's previous postmortems have been conducted 1–3 months after each game has been finished, and only circulated internally after 3–6 months, by which time some of the team may have moved on and others may have simply forgotten important issues that arose during the 1–2 years development period.

The ability to quantify data is also considered important by the Zoë Mode team. For instance, the games industry is prone to underestimating the amount of time required for tools and technology production, so in that sense the same sort of issues commonly cause the same sort of scheduling problems. However, due to the R&D and creative endeavour involved, the specific instances of these problems are hard to estimate in advance. By taking an approach similar to that used in agile software development, comparing the amount of time initially estimated for a given task and the amount of time actually required, overall trends become apparent that could be used to plan future projects. Fitzhugh states that quantifiable measurements should allow the postmortem to identify three to five specific goals that should be addressed in the next project and provide conditions by which to measure success.

#### **7.4.2.8 Summary**

Casual game developer

Casual (non-technical/non-traditional) terminology

12 months from pitch to release

QA for functional bugtesting

12 groups of four participants for playtesting

Video data invaluable

Usability and focus group testing around Beta

Can anticipate future need for in-house usability expert

Postmortem circulated 3–6 months after release

### **7.4.3 Case Study 3 – Relentless Software**

Relentless are an independently owned developer, working exclusively for Sony Computer Entertainment, manufacturers of the PlayStation series of games consoles.

Following the release of Relentless' first title, *DJ Decks and FX* for the PS2 in 2004, the creation of the *Buzz!* franchise began when Sony approached Relentless with the proposal to develop a music-based game. As a result, *Buzz! The Music Quiz* was released in October 2005. Their most recent title, *Buzz! Quiz TV*, was released in July 2008 and is a continuation of the *Buzz!* series. Casper Field, Senior Producer, discussed with us the process of designing a new game and where user experience currently fits into their development strategy.



### 7.4.3.1 Internal Testing

In addition to the core team of programmers, artists, designers and audio engineers, Relentless have a QA group who are responsible for testing the game throughout production. This is essentially functional testing rather than usability or experiential testing. As an example, Field comments that network functionality is a perennial problem and that their testers try to identify scenarios under which the current implementation will fail – such as how to handle a matchmaking case where one party loses connectivity.

Relentless have no formal procedures for dealing with these concerns, particularly towards the beginning of a project, rather relying on the skill of the producer to recognise what the audience want. Later on, when the game is of a sufficient standard that people external to the team are brought in for focus group testing, the producer's earlier decisions are put to the test.

### 7.4.3.2 Understanding Users

In addition to internal and external QA, the producers at Relentless decided to employ external focus group testing for *Buzz! Quiz TV*. This study was conducted during February 2008, 8 months prior to the game's eventual retail release date, which Field describes as being approximate 75% of the way through production. Based on data from previous games in the series, Sony's marketing team had identified three demographic groups, from which the focus group test company sourced four individuals each, totalling 12 participants:

1. Social players (mid-20 s, the “post-pub crew”)
2. Family players (mother, two children and family friend)
3. Gamers (students, late teens)

Field devised 64 questions for the focus test, which were grouped into the following 8 categories,

- Instructions in the intro sequence (4 questions)
- Using the game's menus (12 questions)
- News page (4 questions)
- First impressions (16 questions)
- The overall experience (14 questions)
- Enjoyment (5 questions)
- User-created quizzes (5 questions)
- Future purchases (4 questions)

Each participant rated their response on a 4-point Likert scale, with an additional non-numerical code for no data. The responses were analysed as ordinal data and metrics were produced per question and per participant. In addition to this numeric

analysis, mean responses were also presented back to the development team in bar graph form, whose value axis ranged from 1.00 to 4.00.

Focus group testing in the games industry is generally approached with a degree of trepidation. Most developers are sceptical about the quality of the processes, participants, their feedback and interpretation of data. Subsequently, it can be hard to get buy-in from the development team, and most importantly from the senior members who have the authority to make decisions relating to them. Field's answer is to prove the quality of these issues to the major stakeholders in the team. For example, during their recent testing sessions the lead programmer and artist were actively involved, visited the testing site and gave their feedback about the research questions the study was intended to address.

Field praised the work of the focus group company and, despite commenting that it was an expensive process, would consider doubling the number of participants for their next game. In particular, he pointed out that all of the participants were already aware of *Buzz!*, so an additional control group who had never played before would be beneficial.

The identification of demographic groups does guide the development process, and Field points out the importance of understanding the context and manner in which the game will likely be played. Throughout the development life cycle, the team try to bear in mind what Field calls "the mum-factor"; an informal persona-based approach where they try to imagine their own mothers holding the controllers and enjoying playing the game. Similarly, they have a "drunk factor" scenario, for groups of gamers who come home to play after a drinking session at the pub. Similar to Zoë Mode, it is acknowledged that the Relentless development team do not represent the typical consumer and that features which individual developers might enjoy are not necessarily appropriate to include in the final game.

This attention to players permeates the whole design process to the extent that the designers try to use a more conversational language when addressing players, such as avoiding conventional game terms like "Loading". This terminology could potentially alienate players for whom *Buzz!* might be their first video game experience, so the team prefer to speak to the players informally with phrases like "How would you like to play this game?" instead of the terse but typical "Select Mode".

### 7.4.3.3 Post-Launch

Not unlike the technique of instrumentation discussed earlier in the Black Rock case study, *Buzz! Quiz TV* captures data which allow the team to identify what, how and when the game is played. However, rather than being captured and used only internally with pre-release versions of the game, Relentless capture telemetry data remotely from players of the final, released game as they play in their own homes. Sony's legal department understandably limits what kind of data can be collected, but clearly this still continually produces a vast quantity of data, and Field comments that this does make it difficult to filter and analyse.

Relentless also analysed the 50 or so reviews available after the game shipped in an attempt to identify problematic issues and incorporate this feedback into future developments. This process involved a frequency analysis of comments about

specific individual areas such as menus systems and the user interface, but Field is more interested in whether reviewers understood the game generally. Additionally, not all reviews are treated with the same significance. For the *Buzz!* series, reviews from casual or mainstream media like The Sun newspaper in the United Kingdom are considered more important than niche or hardcore gaming publications. However, once again the issue of historical context is pertinent – as *Buzz! The Music Quiz* was released during the early part of the PS3's life cycle, the market is more likely to be early adopters who have paid more to purchase the console and who have different interests and concerns than the more casual or mainstream market that typically adopts a platform later in its life cycle when the price point has reduced. As such, they are more likely to read website reviews and comment in online forums, so these sites are of more importance than they might be for future games released later in the console's life.

#### 7.4.3.4 Relentless Software Typical Development Summary

12 months production

3 months Alpha – Release

Functional QA

Target demographic: social, family and gamers

Everyday language in games

Focus group test conducted 75% through development

Three groups of four focus group test participants

Content analysis of reviews

## 7.5 Discussion

The single most important issue that has emerged from the case studies is that the studios are testing too late in the life cycle (sometimes as late as Beta). This means that any feedback they obtain from usability studies is unlikely to make it into the final game.

Fitzhugh discussed some of the problems with testing, in particular highlighting the apparent paradox of when to test. Testing later on in the life cycle ensures that the game is more representative of the final product, and hence improves the validity of test results, but from a production point of view this is the worst time to find out about problems. Newing also comments on the scope of testing and mentions that for both *Rock Revolution* and *You're in the Movies*, only parts of the game were tested due to constraints on time and budget. We would suggest that a productive solution would be to embrace testing as part of an agile development process, whereby discrete aspects of the game are tested individually during prototyping, vertical slice and throughout the remainder of development. To that end, by the time of release all aspects of the game should have been tested individually and in coordination as a whole – with the usual proviso that the finished game may have a tighter scope or size than originally intended in order to ensure that quality is maintained.

All studios agreed that they should be testing sooner, and approaches such as EA's method of using focus groups early on to decide on key game concepts could easily be integrated into the development plan.

The vertical slice could be used as an approximate measure for dividing usability testing from user experience testing. All studios acknowledge that testing a game's overall user experience can only be measured once all the components are in place (final artwork, audio, game mechanics, etc.), and the earliest that this can be achieved is at the vertical slice. During interview, Zoë Mode mentioned that they were considering writing mock reviews before a game is released, and we feel that the vertical slice is a useful point at which reasonably representative and valuable data could be generated early in the lifecycle.

If everything after the vertical slice is user experience oriented, then before that milestone the focus should be on usability issues. This would typically mean issues such as user interface layout, game controls or menu navigation.

However, the usability/user experience divide around the vertical slice is not so clear cut. Usability issues will still need to be evaluated after this point (such as game flow and pace), and it is possible to evaluate user experience before the vertical slice (such as game concept focus group test at the start of a project).

## 7.6 Future Challenges

Player enjoyment is currently understood by observing or asking participants for their reactions to a game. One of the key future challenges is to capture, measure and understand a player's body data. Signals such as heart rate (ECG), skin conductance (GSR), facial muscle tension (EMG) or eye tracking may become integrated into commercial game usability evaluations in the future. Indeed, studios such as Valve have already expressed that bio-data could help them to better understand the game play experience (Newell 2008).

Extensive academic research has been conducted on psychophysiological metrics (Mandryk 2005, FUGA 2006) and excellent tools and techniques for the capturing and analysis of such data are available (Nacke et al. 2008). However, the focus of these research projects has been on using biofeedback for automatic adaptation of game AI, rather than as a tool with which to iterate on the design of games prior to release. Furthermore, such studies tend to analyse very short periods of game play with small numbers of participants. It remains to be demonstrated whether such approaches can scale to be applicable for games that may be played by millions of diverse players and whether such techniques could be used for representative longitudinal studies of potentially many tens of hours.

In addition to gathering data from the player's body, the studios in our case studies have already begun to automatically capture player performance data directly from the game (such as *Pure's* dynamic difficulty system). This makes it straightforward to capture an enormous amount of quantifiable metrics, making comparison across a large number of players easier.

This chapter has presented three case studies on how world-class games are currently developed. Although studios are keen to integrate usability evaluations into their life cycle, they are not certain how this can best be achieved. As such, one of the main barriers to conducting usability evaluations is the lack of a formal process that studios can follow. However, traditional usability has a similar issue where there is no strict process that can be followed, rather there are a toolbox of methods that exist which practitioners can use when needed. Future work may involve moving towards a general framework of game usability, which would detail not only the usability techniques which can be used, but also where in the life cycle they should be ideally applied.

**Acknowledgements** The authors would like to thank interviewees Casper Field, Jason Avent, Martin Newing, Dan Chequer and Karl Fitzhugh, and additional feedback from Thaddaeus Frogley.

## References

- Clanton C (1998) An interpreted demonstration of computer game design, CHI '98 Conference Summary on Human Factors in Computing Systems.
- Cornett S (2004). The usability of massively multiplayer online roleplaying games: Designing for new users. In: Proceedings of SIGCHI Conference Human Factors in Computing Systems, pp. 703–710. ACM, New York. doi:10.1145/985692.985781.
- Desurvire H, Caplan M, Toth JA (2004). Using heuristics to evaluate the playability of games. CHI '04 Extended Abstracts on Human Factors in Computing Systems. doi:10.1145/985921.986102.
- Desurvire H, Wiberg C (2008). Master of the game: Assessing approachability in future game design. CHI '08 Extended Abstracts on Human Factors in Computing Systems.
- Dovey J, Kennedy HW (2006) Game Cultures: Computer Games as New Media. Open University Press, Maidenhead.
- Federoff M (2002) Heuristics and guidelines for the creation and evaluation of fun in video games. Indiana University.
- FUGA (2006) Fun of gaming. <http://project.hkkk.fi/fuga/>. Accessed 13th December 2008.
- Isbister K (2006) Better Game Characters by Design. Morgan Kauffman, San Francisco, CA.
- Jennett C, Cox AL, Cairns P (2008) Being in the game. Proceedings of the Philosophy of Computer Games 2008, Potsdam University Press, pp. 210–227.
- Jørgensen AH (2004) Marrying HCI/Usability and computer games: A preliminary look. Proceedings of NordiCHI 2004. doi:10.1145/1028014.1028078.
- Laitinen S (2006) Do usability expert evaluation and test provide novel and useful data for game development? Journal of Usability Studies 2(1): 64–75.
- Larsen JM (2008) Evaluating user experience – How game reviewers do it. CHI Workshop.
- Lazzaro N (2004) Why we play games: Four keys to more emotion without story. Game Developers Conference.
- Malone T (1981) Heuristics for designing enjoyable user interfaces: Lessons from computer games. Proceedings of the Conference on Human Factors in Computer Systems, pp. 63–68.
- Mandryk RL (2005) Modeling user emotion in interactive play environments: A fuzzy physiological approach. Ph.D. Dissertation, Simon Fraser University, Burnaby, Canada.
- Medlock MC, Wixon D, Terrano M, Romero R, Fulton B (2002). Using the RITE method to improve products: A definition and a case study. Usability Professionals Association, Orlando FL.
- Nacke L, Lindley C, Stellmach S (2008) Log who's playing: Psychophysiological game analysis made easy through event logging. In: Proceedings of 2nd International Conference Fun and Games 2008, Eindhoven, The Netherlands.

- Newell G (2008) Gabe Newell writes for edge. Edge online. <http://www.edge-online.com/blogs/gabe-newell-writes-edge>. Accessed 27th March 2009.
- Nielsen J (2005) Heuristics for user interface design. [http://www.useit.com/papers/heuristic/heuristic\\_list.html](http://www.useit.com/papers/heuristic/heuristic_list.html). Accessed 13th December 2008.
- Schaffer N (2008) Heuristic evaluation of games. In: Isbister K, Schaffer N (eds) *Game Usability*. Morgan Kaufmann, Burlington, MA.
- Schuurman D, De Moor K, De Marez L, Van Looy J (2008) Fanboys, competers, escapists and time-killers: A typology based on gamers' motivations for playing video games. *Proceedings of the 3rd International Conference on Digital Interactive Media in Entertainment and Arts*, ACM, pp. 46–50.
- Sweetster P, Wyeth P (2005) GameFlow: A model for evaluating player enjoyment in games. *ACM Computers in Entertainment* 3(3): 1–24.
- Thompson C (2007) Halo 3: How microsoft labs invented a new science of play. *Wired Magazine* 15(9). [http://www.wired.com/gaming/virtualworlds/magazine/15-09/ff\\_halo](http://www.wired.com/gaming/virtualworlds/magazine/15-09/ff_halo). Accessed 13th December 2008.
- Valve (2008) Game and player statistics. <http://www.steampowered.com/v/index.php?area=stats>. Accessed 13th December 2008.