Capturing Player Experience with Post-Game Commentaries

Jeremy Gow[1] Paul Cairns[2] Simon Colton[1] Paul Miller[3] Robin Baumgarten[1]

[1]Imperial College London
180 Queens Gate
London SW7 2RH, UK
+44 (0)20 75948315
jgow@doc.ic.ac.uk

[2]University of York
Heslington
York YO10 5DD, UK
+44 (0)1904 434751
pcairns@cs.york.ac.uk

[3]Rebellion Developments Ltd
Osney Mead
Oxford OX2 0ES, UK
+44 (01865) 797013
paul.miller@rebellion.co.uk

ABSTRACT
Player experience is at the heart of good game design, but designers typically have limited experience data to work with. Detailed and fine-grained accounts of gaming experience would be of great value to designers and researchers alike, but recording such data is a significant challenge. We describe an approach based on *post-game player commentaries*, retrospective verbal reports cued by video of the gaming session and a word list. A pilot study was carried out to capture player experience of a tutorial level for a third person shooter game. We show how the technique can be used to provide useful game design feedback.

Keywords
Player experience, game design, player commentaries

1. INTRODUCTION
Unlike many other computing applications, the point of playing a game is simply the experience it provides to the player. Whilst games designers are talented at providing games that do provoke a good experience in players, there is always commercial demand for greater insight into player experience. However, there is little systematic study of what constitutes good gaming experiences. Reliable knowledge of games tends to be restricted to what does not work [3] rather than what does work. Additionally, there is little in the way of fine-grained analysis of which parts of a game offer good experiences and which diminish the experience. Instead, most analyses of games are based on summative, broad-brush judgments on a single, sometimes protracted, instance of playing a game. Yet it is clear that detailed breakdowns of what contributes to a good gaming experience would be valuable to game designers and games experience researchers.

In this paper we argue that *post-game player commentaries* have the potential to provide more detailed, fine-grained and reliable data about player experience. We first review the range of methods researchers have employed to record player experience data (section 2). We then introduce post-game player commentaries (section 3), discussing their potential advantages and issues with reliability and methodology. In section 4, we describe the use of player commentaries in a small qualitative study of player experience in a single level of Rogue Trooper, a third-person shooter game, and show how commentaries can provide insights into level design and individual experiences.

CGAT Conference 2010, April 6–7, 2010, Singapore.

2. CAPTURING PLAYER EXPERIENCE
Understanding user experience, let alone designing for it, is a notoriously difficult task [3][7]. There are a variety of useful approaches to player experience in games, some of which can be used together for increased validity.

A straightforward approach is to ask players to compare two experiences, so that there is no need to ask players about details of their experience, but simply how they differ. Pedersen et al. [9] used a comparative approach where participants made retrospective relative decisions about their experience of two levels, e.g. the first level was more fun than the second. 181 participants played 4 levels of an online version of Super Mario Bros and completed an Internet survey about their pairwise preferences with respect to fun, challenge and frustration, using a 4-alternative forced choice protocol, i.e. for an experience X: A was more X than B, B was more X than A, A and B were the same X, or neither A nor B was X.

When using comparative methods the number of pairwise comparisons needs to be kept to a manageable level, so the approach is less suitable for investigating variations over many discrete periods of play, or distinguishing between many context-dependant experiences. The comparative approach is also coarse, as players must complete both segments before making a relative judgement.

In non-game contexts, it has been possible to take a more fine-grained approach using the technique of experience sampling, e.g. for flow [6] or presence [12]. People are interrupted in the course of their activities to answer five or six closed questions on their immediate experiences. However, in both cases of flow and presence the underlying concepts have been substantially studied and qualified prior to the use of experience sampling.

With regard to the gaming experience, it is not known exactly what are the important characteristics of the experience that need to be asked about. Thus, both the comparative and experience sampling methods must be restricted to examine the experiences that the researchers expect to see, and there is no opportunity to expand beyond these prior characteristics at the fine-level of detail that is required to really inform game design.

Some attempts to get a finer level of detail have been made using psychophysiological measurements e.g. [8]. The problem with such approaches is one of construct validity – how is the quantitative data obtained matched to the felt experience of playing? It is possible to make some claims that a player is happy or sad or emotionally aroused in some way based on such data. But the cause of these, actually very coarse, emotional states cannot be reliably ascribed to specific in-game elements without some more qualitative approach that reveals the felt experience of the player.

Thus, in order to arrive at some secure understanding of the gaming experience, it is first necessary to conduct some basic qualitative research into what it is that constitutes the gaming experience. This has been successfully done in previous contexts where grounded theory [13] has been used to scope out aspects of the gaming experience such as immersion [2], basic elements of the gaming experience [3] and people's experience of being "in the game" [5]. However, being based on interviews with gamers or reports of playing, this technique is still quite coarse, offering only a limited access to the experience of specific points of a game.

3. POST-GAME COMMENTARIES

What is needed is a way to break down the gaming experience of a whole session of play sufficiently to gain reliable qualitative data without destroying the gaming experience in the process. We propose *post-game player commentaries* as a method that allows players to talk about their whole gaming experience after the experience is over.

A typical data gathering procedure would start with the *play session*, where the participant plays a game as they would normally, without external interruption. During the play session the game's screen output is recorded on video. Play continues until a predetermined set of conditions are met, e.g. for a set period of time, until an objective is reached, or until the player decides to stop playing. Immediately following this is the *commentary session*. The player is shown the screen-capture video and asked to talk over it about what they can remember of their experiences. If required, the player can be prompted in various ways to ensure they talk as much as possible, e.g. by asking short neutral questions.

Post-game player commentaries are a form of cued retrospective think aloud (Cued RTA) protocol. The term *think aloud* (TA) is used for a protocol where a participant talks about the thought processes that occurred during the task in question. TA can take place either during or after the task, known respectively as concurrent (CTA) and retrospective think aloud (RTA) protocols [4]. RTA has the advantage that it does not affect task performance and allows participants to formulate more coherent *verbalisations* (reported thoughts). This is important for experience data, which can be difficult to interpret.

In Cued RTA the participant is shown a representation of their task session in order to cue memory, e.g. a video, screenshots or eye-tracking data. In player commentaries the cue is the screen-capture video. Cued RTA is known to generate more results, and more reliable results, than plain RTA [15], as the cue can aid participants' recall and helps them avoid unintentional fabrication. For problem-solving tasks cueing has been shown to increase the proportion of action-related verbalisations in RTA, as opposed to why (domain principles), how (strategic) or meta-cognitive (self-monitoring) verbalisations [14]. We speculate that action and experience are closely related in video games, so the recalling of actions may aid the recall of associated experience.

In general, RTA relies on the participant retrieving and verbalising memory traces of their thoughts from long-term memory (LTM) immediately after the task, unless the task is extremely short. In domains such as problem solving, the reliability of RTA is known to be highly sensitive to the procedure used and the kinds of information requested, with the work of Ericsson and Simon being particularly influential [4]. They distinguish between three types of verbal data: Type 1 is not transformed before verbalisation; Type 2 data must be transformed for verbalisation but is not otherwise processed, e.g. images; Type 3 data requires additional processing before verbalisation, which they regard as generally unreliable. Their work is not directly applicable to player commentaries because a) some experience thoughts may not be consciously attended to and remembered for later recall (they are not data) and b) video replay may induce processing of memories not related to the original task or verbalisation (they are unreliable Type 3 data) [1].

However, if we accept that some experiences will be lost or reinterpreted before verbalisation we can still use Ericsson and Simon's model and guidelines to inform the reliability of post-game player commentaries. Indeed, reinterpreted experience may be just as significant to the player as the original experience [3]. As with other forms of TA, there is tension between gaining useful design feedback and faithfully modelling cognition [1].

3.1 Methodology

We now look at how the way in which a player commentary study is conducted can affect reliability. The initial play session should be as natural as possible to encourage a normal gaming experience. To this end, the player should not be under an artificial pressure to perform and should be instructed that it does not matter how well they play the game. Players should also not be forced to continue playing when they would have normally stopped. They can be instructed that they play either as long as they like or until other termination conditions are met. As a failsafe, forced play is likely to be revealed by the commentary data.

The play session should also not be too long or repetitive. Retrieval from LTM is fallible, especially if it contains a number of similar memory traces. Between storage and verbalisation memories may become associated with other information stored in LTM, leading to participants incorrectly reporting things they "must have thought" [4]. Hence for long or repetitive session the player is more likely to confuse two similar episodes, or incorrectly report an episode. Retrieval may also be more difficult if the commentary session does not take place immediately after the play session, e.g. if other data gathering techniques, such as questionnaires, are being used in conjunction with commentaries.

Careful design of the commentary session can help players retrieve experiences more reliably. Ericsson and Simon state that simply instructing participants to only report experiences that were consciously heeded during the task can help [4]. Control of the video replay is also an issue: the pace of the action may be faster than participants can comfortably recall experiences, or they may wish to rewatch a segment to correct a verbalisation. Giving participants the ability to pause and rewind will allow them to comment at a comfortable pace. On the other hand, the experimenter may want greater control over the commentary even if it risks introducing bias, e.g. to keep the commentary short, or to focus on specific activities.

Another issue is the use of prompting during the commentary session. If the player stops thinking aloud for more than a few seconds the experimenter may wish to prompt them, e.g. "keep talking" [4]. This may not be enough to keep the player talking specifically about experience, so short neutral questions could also be used, e.g. "what do you remember feeling here?" Alternatively, players can be prompted to talk by having access to a list of common experiences words, which can help them formulate descriptions quickly and clearly [10]. Prompt questions and word lists are both sources of bias, but may be considered worth the risk if it substantially increases experience data. Word lists have the advantage that their role is constant throughout, whereas question prompting can vary considerably during the commentary session.

4. ROGUE TROOPER STUDY

In order to explore the potential of post-game player commentaries for playtesting and research, we conducted a small (four participant) pilot study using Rogue Trooper, a third-person shooter game [11] (see Figure 1). The study is both a demonstration of the method and a means of assessing the kinds of experience data captured by the method. Gathering a large amount of data from a small number of participants is appropriate for a qualitative pilot study: the emphasis is on understanding the nature of the players' experiences rather than statistical significance.

Figure 1. A screenshot of Rogue Trooper [11]

We used open coding of think aloud data to classify player experiences of the first level of Rogue Trooper. This gives us an insight into both the level design and the utility of post-game player commentaries. Although our study uses open coding of video recordings, in general player commentary data is amenable to a range of data gathering and analytical methods, both qualitative and quantitative.

4.1 The Word List Prompt

To encourage players to talk about their experiences we developed a word list to act as an experience prompt. The aim is to help players quickly find a suitable description of their experience, preventing delays in the commentary and allowing clearer reporting of experiences [10]. However, the list must be short enough so that it can be quickly scanned without causing a delay itself.

We decided to compile of list of 20 words that would be 'most useful' to the player. An initial list of 33 words was drawn up by informally asking gamers what experiences they had whilst playing video games. We then designed a short online survey that asked "Which of the following words describe the experiences and feelings you regularly have while playing video games? Select ALL the words that you would find useful." The 33 words were presented to each respondent in a random order.

Table 1. The word list prompt: a list of words describing common gaming experiences, used to assist players during the commentary sessions.

Confused	Excited
Creative	Surprised
Satisfied	Powerful
Challenged	Confident
Bored	Determined
In Control	Curious
Immersed	Disappointed
Tense	Interested
Relaxed	Annoyed
Frustrated	Relieved

From 45 responses we compiled a ranking of the 33 words. We then selected 20 of the top 25 for the word list prompt. Table 1 shows the actual arrangement of words on the prompt sheet. In order to maintain diversity we rejected five of the top 25 (Amused, Intrigued, Happy, Pleased, Obsessed) because they had near-synonyms higher up on the list. The top rated words were Frustrated and Challenged, both chosen by 80% of respondents. The least popular words included in the list were Creative and Disappointed (both 33%), and the least popular overall were Frightened and Angry (both 16%).

4.2 Method

Four participants (A, B, C and D) played the first level of the PC version of Rogue Trooper in single-player mode, while the screen was recorded using Fraps video-capture software[1]. **Figure 1** shows a typical screenshot from the captured video: the player character 'Rogue' is in the centre-foreground, taking cover behind a rock whilst firing on an enemy NPC (centre). To the left an allied NPC can be seen. Also onscreen is a radar display showing nearby NPCs (bottom left), health and ammo display (bottom right) and contextual help for controls (top left).

Participants were asked to play as they would normally for up to 30 minutes, until they completed the first level, or until they

[1] http://www.fraps.com

wanted to stop. They were informed that the study was not about how well they played and that the game contained unrealistic violence typical of this genre. The level consisted of a linear tutorial mission designed to familiarise players with the basic game controls, and the experimenter gave no help during play unless it was explicitly requested. Immediately after the play session, participants completed a 44-item questionnaire as part of a separate study.

For the commentary session, the screen-capture video was replayed in Windows Media Player under control the participant. They were instructed to talk as much as they could during playback about what they were doing in the video and what they felt about it at the time, that it was acceptable not to remember something, and not to guess if they could not remember.

If the participant stopped talking during the commentary the experimenter prompted them with minor variations on two questions: *What were you doing here? What did you feel here?* Throughout the commentary, participants had in front of them a word list as a prompt (see section 4.1) and were told they could use these words, their negations and/or their opposites, to help describe how they had felt. It was stressed that list was an aid, and any description of how they had felt would be equally valid.

The commentary session was recorded on video. Both the participant and the screen capture video were filmed in order to capture gestures, e.g. pointing at objects on the screen. After the commentary session the participants completed a short questionnaire on themselves and their gaming background.

The method and data coding described here was initially piloted with another participant. That data is not reported here due to variations in the protocol used.

4.3 Data Analysis

The main focus of our analysis is the use of open coding (see section 4.3.1) on the player commentaries to classify reported experiences into experience categories (section 4.3.2) and object categories (section 4.3.3). To get an overview of the reported experiences we conducted a *scene-level analysis* (section 4.3.4), looking at the categories of experience each distinct section of the level evoked.

4.3.1 Open Coding

To understanding the players' experiences, the video commentaries were transcribed and *coded* following an open coding methodology, inspired by Grounded Theory [13]. First, we identified all quotes in the transcribed commentary text that refer to some experience. Next, we tagged each quote with a descriptive code which closely matched its meaning, using either an existing code or creating a new one if a suitable one did not yet exist. Finally, we collected together similar codes into categories. These phases were somewhat interleaved, as codes and categories were adapted and reinterpreted as coding proceeded. A short written description of the in-game events of each play session was made for each player using their screen-capture video. These were particularly useful for clarifying and coding the player commentaries.

During coding we decided to develop two parallel sets of codes: one that described the kind of experiences players were reporting, and another that described what those experiences were about. Hence each quote was tagged with an experience code and an object code. For example, "I was really happy about this, as I love sniper games" (Player C) was tagged with the experience code Happy and the object code Sniper.

Categorisation involved bringing together similarly themed but distinct codes. For example, the codes Disappointed, Annoyed, Frustrated and Irritated were all categorised as Dissatisfied. This generalised the data at the cost of losing some of nuances of the original codes. There are no hard-and-fast rules for categorisation, but we aimed for a reasonable number of categories that generalised without misrepresenting, and avoided categories that only cover a very small amount of the original player commentary.

During categorisation we found that several experience codes required disambiguation. For example, the experience code Repetitive could be categorised as Bored or Easy depending on the context, and so was split into the codes Repetitive-Boredom and Repetitive-Easy.

The open coding process generated 14 experience categories (Aimless, Bored, Dissatisfied, Cautious, Confident, Confused, Controlled, Easy, Hard, In Control, Interested, Purposeful, Satisfied, Understand) and 11 object categories (Audio-Visual, Combat, Controls, General, Health, Goal, Interface, Mini-game, Misc-Gameplay, NPC and Story).

4.3.2 Experience Categories

The 14 experience categories were arranged into 7 dimensions as they emerged during open coding. Each dimension is pair of opposing categories: Challenge, Choice, Engagement, Knowledge, Pleasure, Power and Purpose. The dimensions, categories and original codes are shown in Table 2. The dimensions can be summarised as follows:

- **Challenge** experiences (Hard vs. Easy) are about the players' experience of their ability relative to the game.
- **Choice** experiences (Controlled vs. In Control) relate to the freedom the player has to decide what to do in the game. Note that experience of the actual game controls may be are categorised elsewhere, e.g. Knowledge.
- **Engagement** experiences (Interested vs. Bored) are about the game as the focus of the player's attention.
- **Knowledge** experiences (Understand vs. Confused) are about the player's understanding of the game and of their situation within the game.
- **Pleasure** experiences (Satisfied vs. Dissatisfied) are about the player's active pleasure or displeasure at aspects of the game.
- **Power** experiences (Confident vs. Cautious) are about the player's efforts to survive and achieve success within the game.
- **Purpose** experiences (Purposeful vs. Aimless) relate to the player's plans and goals within the game.

Table 2. The 14 experience categories generated by open coding. Pairs of opposing categories form 7 dimensions of experience. Each category has a valence (positive/negative or simple/difficult) and a set of original codes.

Dimension	Category	Valence	Codes
Challenge	Hard	Struggle	Challenging, Hard
	Easy	Cope	Easy, Repetitive-Simple
Choice	In-Control	Positive	In-Control, Interactive
	Controlled	Negative	Controlled, No-Choice
Engagement	Interested	Positive	Attached, Anticipation, Curious, Interested, Immersed, In-Zone, Focused
	Bored	Negative	Bored, Out-of-Game, Repetitive-Boredom, Break
Knowledge	Understand	Positive	Aware, Creative, Experimenting, Learning, Understanding
	Confused	Negative	Confused, Don't-Know, Overloaded, Unaware, Unsure-Know
Pleasure	Satisfied	Positive	Cool, Enjoyment, Fun, Happy, Satisfied
	Dissatisfied	Negative	Angry, Annoyed, Disappointed, Frustrated, Irritated
Power	Confident	Cope	Calm, Comfortable, Confident, Normal, OK, Powerful, Safe, Successful
	Cautious	Struggle	Afraid, Cautious, Reserved, Scared, Stressed, Tense, Worried, Useless
Purpose	Purposeful	Positive	Determined
	Aimless	Negative	Disoriented, Lost, No-Plan, No-Direction, Unsure-Plan

The opposing experience categories for each dimension can be distinguished further into those that, in general, affect the player's attitude to the game (positive vs. negative) and those that relate to how difficult the player finds the game (cope vs. struggle). Hence every category has a *valence*:

- **Positive** experiences: In-Control, Interested, Purposeful, Understand and Satisfied.
- **Negative** experiences: Controlled, Bored, Aimless, Confused and Dissatisfied.
- **Cope** experiences: Easy and Confident
- **Struggle** experiences: Hard and Cautious

4.3.3 Object Categories

Every coded experience has an object code that described what that experience was about or what caused it. These object codes were organised into 11 categories:

- **Audio-Visual** experiences were about the graphics and sound within the game.
- **Combat** experiences are about aspects of combat outside the mini-games (see below).
- **Controls** experiences refer to the controls for movement, weapons etc.
- **General** experiences have no obvious referent and are taken to refer to the general experience at that moment, e.g. "I'm happy."
- **Health** experiences refer to the player character's changing health levels.
- **Goal** experiences refer to the objectives set by the game and the player's awareness of them.
- **Interface** experiences relate to the details of the games information displays (health, ammo, map, help messages, tutorial dialogs etc.)
- **Mini-game** experiences are about the two special combat sections in the first level, The Lazooka (M1) and The Flak Cannon (M2). See section 4.3.4 for details.
- **Misc.** experiences are about any aspect of gameplay not covered in other categories, e.g. bombs dropped from aircraft, level layout.
- **NPC** experiences relate to enemy NPCs (Non-Player Characters) or allied NPCs, especially Rogue's comrades Gunnar and Bagman.
- **Story** experiences are about the level's narrative, which is established in a number of cutscenes (see section 4.3.4).

4.3.4 Scene-Level Analysis

In summary, after open coding we had a collection of experience quotes for each player commentary, where each quote was tagged with an experience and object code, and with codes arranged in categories as described above. This annotated data is more readily interpreted than the original player commentaries, but the complexity of data still makes it difficult to get an overview of players' changing experiences and how they relate to the events within each play session, and to the level design.

In order to get an overview of the data we chose to look at experiences on a 'scene-by-scene' basis. First we divided the level into distinct sections of combat (C), navigation (N), cutscene (T) and mini-game (M). There are 37 scenes in total: 17

Table 3. Overview of the four play sessions: Total playing time (mins:secs); Shots fired, number hitting enemy NPCs, and percentage success; Grenades thrown; Enemy NPCs killed in total, killed with sniper fire, and killed by a shot to the head; Damage taken, inflicted and efficiency (damage inflicted per damage taken); Number of times player died.

Player	Time	Shots	Hits	% Hit	Gren.	Kills	Sniper Kills	Head Kills	Damage Taken	Damage Inflicted	Damage Efficiency	Died
A	37:06	629	292	46.4	2	34	0	3	386	1469	3.8	9
B	24:06	545	265	48.6	2	35	6	1	390	1509	3.9	2
C	16:14	514	257	50.0	9	36	7	5	126	1601	12.7	0
D	15:49	363	231	63.6	1	33	1	4	208	1503	7.2	0

navigations, 12 navigations, 6 cutscenes and 2 mini-games (the Lazooka and the Flak Cannon).

The level is designed proceed in a set sequence of scenes:

T1, N1, C1, **T2**, N2, C2, N3, C3, T3, N4, C4, N5, **T4**, N6, M1, N7, **T5**, N8, C5, N9, C6, N10, C7, N11, C8, N12, C9, N13, C10, N14, M2, N15, C11, N16, C12, N17, **T6**

If the player character dies then they may be moved back the start of the scene, or to the start of an earlier scene. Players A and B had their characters die and so attempted some scenes multiple times, and/or out-of-order. C and D did not die and so encountered each scene only once in the standard order. For this level, the design predetermines a linear sequence of activities, e.g. via location-based combat triggers, so the division is quite straightforward. However, a scene analysis would still be possible with non-linear level designs e.g. based on each player's phases of activity and location.

Once a sequence of scenes had been established for a player we determined the list of associated quotes for each scene S. We then defined the list of experience-object categorisations C(S) for that scene. Note that a scene may be associated with the same experience several times but with different objects, or the same object several times but with different experiences. Furthermore, a single object may also have opposing experiences associated with it in the same scene. This can represent a changing experience, or alternatively a contradictory experience, e.g. a combat described as "stressful but normal" (player B) is categorised as both Cautious and Confident.

One problem with C(S) is that scenes often contain multiple quotes with the same experience-object categorisation. These are mostly, but not exclusively, multiple descriptions of the same experience or closely related experiences. Hence C(S) will contain repeated elements for the same experience. We make the simplifying assumption that all repetitions are repeated descriptions, and so defined C'(S) as the set of experience-object categorisations, i.e. C(S) with no repeated elements. We call C'(S) the *scene experiences* for S. We use this measure for the remainder of the paper.

As redundant multiple experiences have been removed we regard the scene experiences as a good indication of the 'amount of reported experience' for a scene S, even though it fails to distinguish genuinely repeated experiences. There are of course other features of the original experience that scene experiences fail to capture, such as the scope and intensity. In the results section below we use the scene experiences for all comparisons, regarding each scene experience as of 'equal value' and ignoring the fact that scene length can vary.

4.4 Results

All four participants were male, between 26 and 45 and had played video games for at least 10 years. Players A and B currently played very infrequently (a few times a year or less) and favoured games in genres distinct from Rogue Trooper, i.e. not first- or third-person shooters. C and D played very frequently (every week and every day respectively) and included shooters (e.g. Halo) in their list of favourite games.

4.4.1 Gameplay Statistics

In total, we recorded around 1h30m of player commentary. Table 3 gives the end-of-level statistics for the four play sessions, along with playing time, defined as the time from the end of the introductory cutscene (T1) to the beginning of the end-of-level cutscene (T6). All four participants completed the level, with one (A) exceeding the requested 30 minutes, instead choosing to complete the level.

From Table 3 we get a rough idea of the relative success of the participants at playing this level of Rogue Trooper: C and D completed the level in roughly 15 minutes without dying and have higher damage efficiency (they inflict more damage points for each point taken). C used more diverse weaponry than D (sniper rifle, grenades), took the least and inflicted the most damage, but D was a more accurate shot. A and B both have low damage efficiencies, took longer and died multiple times. Of these two, A took longer due to a large number (9) of player character deaths. Also, B killed with the sniper rifle while A did not.

4.4.2 Overall Experience

Table 4 shows the number of scene experiences (see section 5) for each player and valence. Overall for the level there were more negative scene experiences (141) reported than positive (98), and a similar number of cope (36) and struggle (27) experiences.

Table 4. Scene experiences by player and valence.

	A	B	C	D	Total
Positive	52	7	21	18	98
Negative	44	46	17	34	141
Cope	9	9	16	2	36
Struggle	5	14	3	5	27
Total	110	76	57	59	302

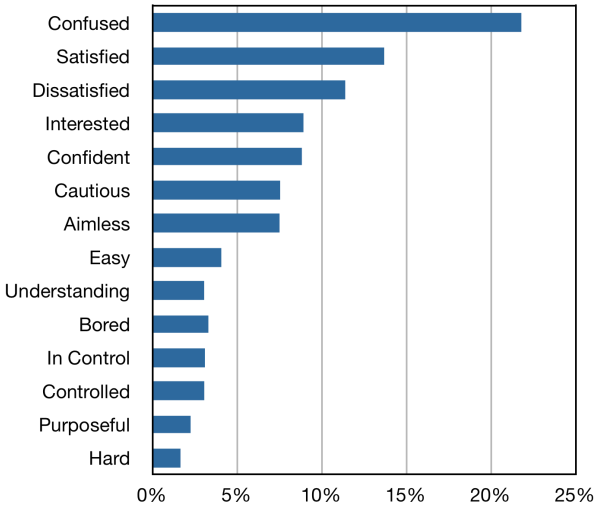

Figure 2. Mean share for scene experiences.

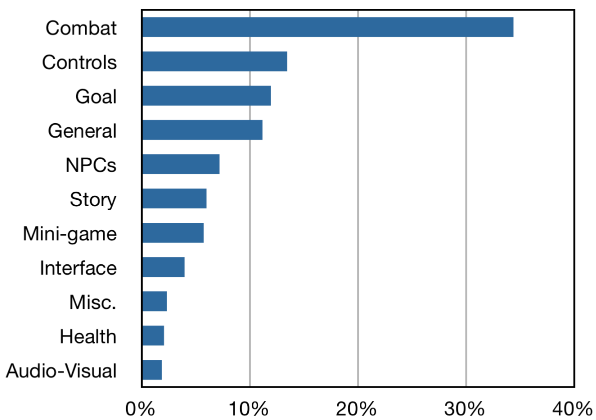

Figure 3. Mean share for objects of scene experiences.

To compare the occurrence of scene experiences across players, we look at the percentage share of each player's total scene experiences. The mean share across all four players is shown in **Figure 2**. By far the most popular was Confusion with 21% mean share of reported scene experiences, followed by Satisfied (14%) and Dissatisfied (11%). Interested, Confident, Cautious and Aimless also had more than a 5% mean share each.

The mean share for objects of scene experiences is shown in Figure 3. Combat was by far the most frequent object of scene experiences with 34% mean share. Also common were references to Controls (13%) and Goal (12%), and General experience reports with no specific object (11%).

4.4.3 Experience by Player

The dominant scene experiences for **Player A** were Interested (19%) and Confused (19%), followed by Satisfied (14%). He had the most diverse experience, with three dimensions having both opposing experiences at 5% or more: Interested-Bored, Satisfied-Dissatisfied and Aimless-Purposeful. Out of all the players he also had the highest percentage of Interested, Purposeful and In Control. After Combat, his experiences are mostly about Goal (15%, both successes and failures). Examining his scene-by-scene breakdown, he has mixed reactions to his first death and the immediately following mini-game (scenes N6 and M1) and strong Confusion and Aimless feelings when meeting the NPC Bagman (N8). He is regularly Interested for the first half of the session, but after a positive combat C7 becomes Confused and negative at C8-C9, then Bored when he begins to die regularly (7 times in all) around combats C9-C10. After completing that section he is Purposeful and In Control, with a positive C11-C12.

Player B is mainly Confused (24%), Dissatisfied (21%) and Cautious (18%). He has opposing experiences of Confident-Cautious, the highest share of Dissatisfied, Cautious and Controlled and the lowest Satisfied. After Combat, his experiences are mostly about Controls (22%), the highest of any player. His scene-by-scene breakdown shows he has very negative experiences of the initial orientation scene (N1) and his first death just before the first mini-game (N6). He is regularly Confused up to combat C7, has some positive experiences around C7-C8, but is Confused again after C10. He is most negative about combat C12 (his second death), but Satisfied with its resolution.

In contrast, the majority of **Player C**'s experiences are Confused (19%), Satisfied (18%) and Confident (16%). He has some opposing experiences of Satisfied-Dissatisfied and the highest share of Confident and Easy. After Combat, his experiences are mostly General. Focusing on the scene-by-scene picture, positive and negative experiences are fairly evenly distributed throughout the play session, although with no negative experiences until the 10th scene (N4).

Finally, **Player D** experienced mostly Confused (23%), Satisfied (17%) and Aimless (15%). He has highest Aimless and the lowest Confident of the players. After Combat, his experiences are about Goal (17%), the highest of any player. He also has the highest share of Story experiences. Looking scene-by-scene, he has a positive start with scenes N1-C1 as he discovers sniper combat. The cutscenes experience is Dissatisfied and Bored, and navigation is often Aimless. In combat C8 he is very negative, Confused about the situation and the controls. Here he switches to pistol by accident and is unaware he can switch back.

4.4.4 Experience by Scene

Combining all the players' scene-by-scene experiences, 7 scenes stand out as invoking a large amount of a particular experience:

- The initial orientation (N1) produces a range of negative experiences;
- The first mini-game (M1) and the preceding scene (N6) invoke Confused, Dissatisfied and Aimless;
- Meeting the NPC Bagman (N8) is Confused;
- Combat C7 produces Satisfied experiences;
- Combat C8 produces Confused experiences;
- Combat C9 induces a lot of positive and negative experience, and the most struggle experience.

4.5 Discussion

The most striking feature of the results is the dominance of Confused experiences and Combat as an object of experience. Combat is designed to be the central feature of gameplay, so it reassuring that it attracts the most scene experience reports, both positive and negative. None of the players had played Rogue Trooper before, and it was a tutorial level that introduced a variety of game mechanics and controls, which may account for the Confused experiences across all players. This fits in with Goal and Control being the two most frequent subjects of experience after Combat. These problems with Knowledge may be part of an engaging learning experience or poor level design, and further work could examine this issue more closely. The two more able players (C and D) both had some (5%) Understanding experience, perhaps reflecting that learning was more significant for them. All the players experienced a number of Dissatisfied and Satisfied experiences, with everyone but Player C reporting more Satisfied.

The study also gave us a good picture of the individual players' changing experiences. The less able players have very different play sessions: although Player A finds it harder to play he is Interested from the start, becoming quite positive after overcoming adversity and the associated negative experience. In contrast, B is put off from the start and has a negative experience throughout dominated by issues with Controls. The more able players are both Satisfied, but whereas C is Confident, finds the level Easy and engages in all aspects of combat, Player D is more Aimless, negative about the narrative cutscenes and has a unusually difficult experience in one combat.

By aggregating players' experiences we are able to identify scenes with particular experience characteristics. Level designers could use similar techniques to see if the experiences match their expectations for the level. For instance, the negative experiences, especially Confused, in scenes N1, N6, M1, N8 and C8 did not seem to be part of the design, which suggests these would benefit most from redesign effort. Negative experience is not necessarily a design problem: combat C9 induced a lot of positive, negative and struggle experience in what seemed to be a test of ability. Indeed, working to overcome it resulted in significant positive experience for Player A.

5. CONCLUSIONS

Post-game player commentaries aim to increase the retrospective recall of player experience by cueing with a video replay. It has the potential to be a useful technique for gathering experience data in playtesting or research, having the advantage of giving relatively fine-grained access to a wide variety of detailed experience data without destroying the experience itself. We have outlined a general methodology and argued for its reliability, although there is a lack of research into experiential think aloud.

Further research into player commentaries could be taken in a number of directions: rigorous exploration of their relative reliability and utility; investigation of methodological issues such as the use of prompts, including developing better word list prompts; better techniques and tool support for commentary data analysis, including improved modelling of experience frequency, granularity, intensity and scope, and analysis of the relationships between activity and experience.

We have also demonstrated the use of player commentaries in a small study, which successfully obtained a range of experience data. In general, the method could be used with a wide variety of data analysis techniques. Our qualitative analysis used open coding to obtain a scene-level analysis that gave an overview of varying individual and aggregated experience, and allowed key experiences to be focused in on. We also showed how aggregated scene experience data could be useful tool for focusing game design efforts.

6. ACKNOWLEDGMENTS

This research was supported by EPSRC grants TS/G002835/1 and TS/G002843/1 and the Technology Strategy Board.

7. REFERENCES

[1] Boren, M. T., Ramey, J. 2000. Thinking Aloud: Reconciling Theory and Practice. IEEE Transactions on Professional Communication, 43, 261-278.

[2] Brown, E., Cairns, P. 2004. A Grounded Investigation of Immersion in Games. In Proceedings of the ACM Conference on Human Factors in Computing Systems, CHI 2004, ACM Press, 1297-1300.

[3] Calvillo-Gámez, E. H. 2009. The Core Elements of the Experience of Playing Video Games: Studying the Gaming Experience, Lambert Academic Publishing.

[4] Ericsson, K. A. and Simon, H. A. 1993. Protocol Analysis: Verbal Reports as Data. Revised Edition. The MIT Press.

[5] Jennett, C., Cox, A., Cairns, P. 2009. Being 'in the game.' In Gunzel, S., Liebe, M., Mersch, D. (eds) Proceedings of Philosophy of Computer Games 2008, Potsdam University Press, 210-227.

[6] Larson, R., Csikszentmihalyi, M. 1983. The Experience Sampling Method. New Directions for Methodology of Social and Behavioral Science, 15:41-56.

[7] McCarthy, J., Wright, P. 2004. Technology as Experience. MIT Press.

[8] Nacke, L., Lindley, C. A. 2008. Flow and Immersion in First-Person Shooters: Measuring the player's gameplay experience. In Proceedings of FuturePlay 2008, November 3-5, Toronto, Ontario, Canada, 81-88.

[9] Pedersen, C., Togelius, J., Yannakakis, G. 2009. Modeling Player Experience in Super Mario Bros. In Proceedings of the 2009 IEEE Symposium on Computational Intelligence and Games, September 7-10, 2009, Politecnico di Milano, Milano, Italy.

[10] Petrie, H., Harrison, C. 2009. Measuring Users' Emotional Reactions to Websites. CHI Ext. Abstracts 2009, 3846-3852.

[11] Rebellion Developments Ltd 2006. Rogue Trooper (Microsoft Windows version). Eidos Interactive.

[12] Slater, M., Usoh, M. and Steed, A. 1994. Depth of Presence in Virtual Environments, Presence: Teleoperators and Virtual Environments, 3.2 MIT Press, 1994, 130-144.

[13] Strauss, A., Corbin, J. 1990. Basics of Qualitative Research: Grounded Theory Procedures and Techniques. Sage.

[14] Van Gog, T., Paas, F., van Merrienboer, J. J. G., Witte, P. 2005. Uncovering the Problem-Solving Process: Cued Retrospective Reporting versus Concurrent and Retrospective Reporting. Journal of Experimental Psychology: Applied, 11(4): 237-244.

[15] Van Someren, M. W., Barnard, Y. F., Sandberg, J. A. C. 1994. The Think Aloud Method: A Practical Guide to Modeling Cognitive Processes. London: Academic Press.