

Arrrgghh!!! - Blending Quantitative and Qualitative Methods to Detect Player Frustration

Alessandro Canossa
Center for Computer Games Research
IT University of Copenhagen
Rued Langgaards Vej 7
2300 Copenhagen S Denmark
alec@itu.dk

Anders Drachen
Aalborg University
Institute of Communication and
Psychology, Lautrupvang 15,
2750 Ballerup Denmark
drachen@hum.aau.dk

Janus Rau Møller Sørensen
Crystal Dynamics – IO Interactive
EIDOS – Square Enix
1300 Seaport Blvd. Suite 100
Redwood City, CA 94063
januss@crystald.com

ABSTRACT

Frustration, in small, calibrated doses, can be integral to an enjoyable game experience, but it is a very delicate balance: just a slightly excessive amount of frustration could compel players to terminate prematurely the experience. Another factor with high relevance when analyzing player frustration is the difference in personality between players: some are less willing to endure frustration and might give up on the game earlier than others. This article seeks to identify patterns of behavior that could point to potential frustration before players resolve to quit a game. The method should be applicable independently from the personalities of different players. Furthermore, in order for this method to be relevant during game production, it has been decided to avoid relying on large numbers of players, and instead depend on highly granular data and both qualitative approaches (direct observation of players) and quantitative research (data mining gameplay metrics). The result is a computational model of player frustration that, although applied to a single game (Kane & Lynch 2), is able to raise a red flag whenever a sequence of actions in the game could be interpreted as possible player frustration.

Categories and Subject Descriptors

I.5.2 [Pattern Recognition]: Design Methodology – *Pattern Analysis, Feature evaluation and selection.*

General Terms

Measurement, Design, Experimentation, Human Factors, Verification.

Keywords

Qualitative and Quantitative Research methods, User experience, Gameplay metrics, Player frustration, Small sample size, Patterns of behavior, Game design, Game development.

1. INTRODUCTION

Frustration is an emotional state that arises as a response to a perceived opposition towards the achievement of a goal, and it can either resolve in anger or disappointment [2] according to whether the level of perceived opposition is too high or too low

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

FDG '11, June 29–July 1, Bordeaux, France.

Copyright 2010 ACM 978-1-4503-0804-5/11/06...\$10.00.

and according to each individual's personality. Frustration is also a factor that has been recognized as pivotal in shaping or spoiling optimal experiences [5]: if the cause of frustration is mild and internal (laziness, lack of confidence, etc.) it can easily be a positive force to inspire and motivate, but if it is caused by external forces that are perceived to be outside an individual's control (i.e. a task too hard compared to the skills available) it can lead to feeling of powerlessness and eventually anger [8].

Frustration would normally be characterized as an unwanted component of user experience [3, 14, 23, 18]; however, frustration is also a recognized component of the experience of play [11, 12, 19, 28].

In the context of digital games, external frustration is caused mostly by bugs or balancing issues. Bugs are unintended faults in the code or assets that prevent players from advancing; they are addressed by the Quality Assurance department during development from very early stages of production. Balancing issues can cover anything between difficulty adjustment to navigation and world layout; this type of more phenomenological debugging is dealt with by the User Research department and it cannot begin until there is a version of the game complete enough to elicit a defined experience, usually at least a vertical slice, but at the same time it is recommended not to introduce major changes in the later stages of the development process.

These requirements strongly reduce the window of opportunity for the User Research department to catch instances of excessive frustration; therefore an automated system that can help detecting patterns of behaviors pointing towards probable player frustration is very welcome by user researchers and designers alike. Such a system could help achieve the difficult balance where frustration serves as a motivator and not as the reason leading players to stop the game.

This paper and the associated case study showcase the potential of behavioral data to enable modeling players' interaction and navigation through a game environment. This is of key interest to game designers because it allows them to observe how their games are being played. User-oriented methods such as playability testing [6, 19] can also locate gameplay problems, however, when integrating gameplay metrics in a data collection suite, it becomes possible to model the second-by-second behavior of players, from one to a few thousands subjects, simultaneously.

2. BACKGROUND

The first attempt at a model for detecting player frustration presented here has its roots in a case study that originated during the testing phase of the game "Kane & Lynch 2: Dog Days"

(KL2) developed by IO Interactive (IOI), a Square Enix studio. IOI has long-standing traditions of user research established by running high quality usability and user experience evaluation sessions during development, so much so that several other Square Enix studios have made use of IOI's User Research services. Integral part of the user research practice is the "Game Metric Suite" developed by the Online Team in an effort to provide quantitative input to the primarily qualitative methods already in place in User Research [27]. A task force was established: the User Research manager Janus Rau Sørensen, the captain of the Online Team Thomas Hagen and the researchers Alessandro Canossa (IT University of Copenhagen) and Anders Drachen (Aalborg University). The task force had access to cutting edge technology, years of experience in user research and notably resources, in terms of time and focus, afforded by the inclusion of academic scholars, which is rarely available in a purely industrial context, highlighting the usefulness of industry-academia collaboration in the games sector. The first step was the development of a metrics analysis and visualization system that allowed tracking the location and time of events such as player deaths, weapon-use and NPC kills gathered from the in game behavior of the play-session participants before the release of a game.

Unfortunately it was not possible to rely on the same quantity of testers involved in play sessions as described by Kim et al. [15] where the fact that numerous projects are carried out at the same time could warrant a large user research team taking in between 30 and 50 users per session (in the current case study 22 users are involved). In fact, numbers in this scale are required for an adequate sample size and create meaningful heat-maps for death numbers or location. With the visualization system in place at IOI it was possible to see where play testers had died and used a specific weapon, but due to the very limited number of data points rarely there was a clear pattern or even a connection between these metrics and the level of enjoyment. Even the few patterns individuated by a purely quantitative approach rarely added new information to the results of qualitative research and a low statistical significance loomed upon each result.

On the opposite end of the scale, it is possible to employ advanced data-mining and statistical methods on metrics data generated by thousands of players once the game has been shipped [9, 10] and the results have a potentially high impact on the design both in terms of recognition and prediction of patterns of behavior [30, 31], but it is already too late in the production cycle to be directly applicable for development. Furthermore there was a methodological gap between the qualitative approach normally in use at the User Research department and the new metrics tools.

Quantitative methods such as heat-maps and reports either did not add any new information due to too few data points available during early phases of development or were made available too late to make any real difference. At the same time these methods are highly attractive for a production environment since they produce generalizable, objective and actionable information.

Given this background the main challenges faced by the task force were:

1) Establishing processes for successfully interfacing metrics analyses with qualitative methods

2) Individuating procedures to utilize metrics reliably with small sample sizes

3) Implementing these procedures and producing results in ways that respect the fast rhythms to which the game industry is subjected to.

The case study presented attempts to deal with all challenges listed above.

3. METHODOLOGIES FOR GAME USER RESEARCH

In user research all tools and techniques come with their own theoretical assumptions that function as a filter to the observed world: according to the tool employed some features will slip through, others will not. As an example it is revealing to look at questionnaire as a tool: answers are given to questions asked, not to the questions that have never been uttered - it is not just a matter of designing the questionnaire properly: it is a defining characteristic of any method, theory or tool – it is only possible to see something because other things are kept out. That is why most user research studies make use of more than one tool so the phenomenon under scrutiny is triangulated.

It becomes important then to define which constellation of tools and techniques gives the best coverage. Historically a big part of this discussion has revolved around a clash between qualitative and quantitative approaches. Even though in practice most user researchers use both, there are distinct mismatches between some quantitative and qualitative methodologies and awareness of the difference between methods is essential. Most quantitative methodologies are founded on a logical-positivistic paradigm, while most qualitative research tools are grounded in interpretive social sciences. The differences are highlighted in the table below [adapted from 24]:

Table 1. A comparison of key features of qualitative and quantitative approaches to game user research

Foundation	<i>Interpretive social science</i>	<i>Positivism</i>
Concepts	Social constructions, meanings, inter-subjectivity	Social structure, social fact
Methods	Qualitative hypothesis generation	Quantitative hypothesis testing
Driver	Emergent patterns, theory	What can be measured and sampled
Domain preference	People in their own environment	Lab
Status of data	Situated experiences	Facts about behavior and attitudes

Classical usability testing with users behind a one-way mirror performing predefined tasks and answering questionnaires belongs primarily with positivism since usability is highly connected with behaviorism or behavioral psychology which is directly born from positivism.

An 'interpretive social science' approach to analyzing the same player-game interaction could be e.g. observing the game being played in the participant's home, interacting with the participant while playing, using loosely-structured interview guides, letting the participant take part to the design phase etc. As such, looking at the same phenomena, different methods often reveal different things - there will be overlaps, but also discrepancies.

Nielsen [33], maps the multitude of methods and approaches available to user research specialists along two dimensions: qualitative/quantitative and attitudinal/behavioral.



Figure 1. Some of the methods utilized in user research mapped over “approaches” and “sources” dimensions. [adapted from 21]

The qualitative/quantitative axis describes the type of approach: qualitative data is gathered directly (observation on the field) while quantitative data is gathered indirectly (through an instrument). Also the questions answered by the two approaches are quite different: quantitative data answer questions regarding what players do, where they do it, when and how much, while qualitative enquiries can answer why they do it and how.

The attitudinal/behavioral axis attempts to map different sources of data: what players say and what players do and quite often these two sources do not match at all.

Gameplay metrics are quantitative data gathered through automatic instrumentation about the interaction between the player and the game [4]. Since starting up in 1997, Microsoft Game User Research (MSGUR) has taken an early lead in the field [19, 6, 32]. One of the most well-known MSGUR cases is their work with Halo 3 [26]. In parallel with the work of MSGUR, other game studios have described work with gameplay metrics

and data mining player behavior [17]. Once in a while reports from developers emerged to the wider community [7, 25] as well as from the research community [29, 13].

IO Interactive had predominantly used qualitative approaches born from the Scandinavian Interaction Design tradition (itself derived from interpretive social sciences), partly because qualitative approaches do not necessarily require large sample sizes and partly because of its participatory nature. At the same time it was evident that quantitative methodologies could provide a level of objectivity that would prove invaluable when convincing designers to modify their work. Another benefit of quantitative research is the parametric nature of the models generated, this allows for generalization, recognition and prediction of patterns.

One of the biggest problems was coordinating the different tools - positivistic approaches were applied to metrics and a social-interpretive approach were applied to the rest of the research; this discrepancy manifested itself when reflecting on the number of participants used for testing: it was adequate for the qualitative research, but inadequate for the metrics analysis, resulting in poor statistical significance.

Metrics analysis as a tool is mostly used in a positivistic way; but there is really no reason why it could not be used in connection with an interpretive social science approach, and this is partly what this case study is about.

4. APPROACH AND METHOD

4.1 Background of the case study: During the regular user-oriented testing of KL2 at IO Interactive in the later development stages of the game (where vertical slices were available and the game for these slices mostly functional), several participants were observed manifesting clear signs of frustration at particular times throughout the sessions with a seemingly consequent lack of commitment.

In most of such test situations, the reasons for these points of frustration appeared to be related to bugs in the game or unfinished design – for instance a point in the game where there was a weapon that did not work or a checkpoint that malfunctioned, causing the player to respawn too far back after dying. This was not something that would have any value to the development team at IOI, but there were traces of a pattern of in-game behavior connected to the experience of frustration that theoretically might as well occur in connection to a ‘real’ design issue – for instance the participants rushing forward after dying over and over, not really looking around for danger, increasing the risk of dying again, causing even more frustration, after which the participants slowly regained their composure and concentration and took a more careful approach.

However, with a large number of participants playing at the same time, it becomes cumbersome for resident game user research personnel in a development company to detect the various hints in the behavior of players pointing towards frustration. A method for detecting and defining behavior that occurs when players are frustrated is therefore desirable. Relying on game metrics permit pinpointing the exact location in the game world where such behaviors occur.

This requirement from the user research team at IOI led to the case study described here. The key questions attempted answered are as follows:

- What are the patterns of interaction and navigation in the game that point towards a state of frustration in that player?
- Could these symptoms be observed also in different players?
- Can they be generalized, identifying that type of frustration throughout the whole game?

A mixed-methods approach based on general usability and playability principles, focusing on a traditional small lab-based sample size (Pagulayan et al., 2003) was applied in order to examine player frustration while playing the game “Kane & Lynch: Dog Days” (KL2), a third-person shooter developed by IOI. The player controls a single character and mainly has to worry about staying alive, eliminating enemies and solving specific tasks. 22 people participated in the study. To put this in context, usability testing experts [33] report that cost-benefit analysis of user testing provides the optimal ratio around three or five users for each highly distinctive group of users.

The focus of the case study was to identify which behaviors led to player frustration, not the type and/or magnitude of frustration experienced. This is a topic for future work.

This case study contains several steps of observation and experimentation, exemplifying a game user research process in an industry context, with the difference that the industry-academic collaboration permitted a focus not just on solving the specific user-experience issue encountered in KL2, but also to spend resources on investigating the potential for model-building on player frustration. The steps in the case study can be summarized as follows:

- 1) **Game user testing:** A problem with players getting frustrated was identified during the regular user-research testing of Kane & Lynch: Dog Days. Observing this problem led to the initiation of empirical analysis.
- 2) **Analysis of recorded session video:** Initially, a video recording and screen capture from one (1) play session where a player got frustrated was examined in order to identify segments of play sessions of interest (where players were frustrated) (qualitative hypothesis forming).
- 3) **Game metrics analysis:** The behavior of the player (1) during the segments of time where they were frustrated, as extracted in the form of detailed game metrics. Game metrics data were analyzed conjointly with video recordings to identify markers of behavior leading to player frustration (quantitative hypothesis forming).
- 4) **Experimentation:** Game metrics data from 22 randomly selected players among the IOI tester base were analyzed, identifying behaviors similar to the ones identified previously in six (quantitative verification).
- 5) **Follow-up interviews:** The 22 sampled players were identified about their feelings of frustration during their specific game sessions. Unstructured interviews combined with video recordings from the game sessions in question, were employed to maximize depth of understanding generated. The same six players whose patterns of behavior were found to carry the markers of frustration identified

earlier confirmed that they felt frustrated at those times during their play sessions. None of the other 22 players felt frustrated (qualitative verification).

4.2 Analysis of recorded session video and game metrics: There are different theories of frustration [1, 22], for example failure to understand goals, failure to communicate means available to achieve goals and repeated failure to overcome challenges. For the case study, frustration was defined using the following definition: repeated failure to overcome challenges. This definition formed a compromise between the nature of frustration and the limitations of user-instrumentation data (which cannot show what users feel, only provide indication based on defined guidelines).

The initial analysis was as mentioned above, carried out on the video recordings from one game session featuring one participant, where segments were identified where the participant in question exhibited aggravation. The corresponding game metrics were pulled from the IOI Game Metrics Suite (Figure 2), enabling joint analysis.

In the current case, the programmers delivered a version of the game in which a check-point malfunctioned forcing players to repeat a fairly long and challenging segment of play. In that session, the user researchers managing the tests observed a test participant, who considered himself fairly proficient, become more and more aggravated as he failed to complete a level of the game, dying several times in the same area. The participant manifested frustration through body movements, facial expressions and verbalizations, arriving to the point of throwing the controller away.

The malfunctioning checkpoint exasperated the situation because every failure was further punished with a lengthy navigation of the same environment and facing the same challenges without any sort of achievement or feeling of progression.

Analysis of the video recording and in-game behavior via the game metrics data resulted in the following indicators for frustration (Figure 3):

- 1) The player died in the same location four consecutive times, actually regressing in the second, third and fourth attempt.
- 2) The number of enemies killed decreased considerably in each play through.
- 3) The pace of movement of the player becomes considerably faster in each play through, and the same route is repeated with no variation (third and fourth death).
- 4) Also lacking is the presence of special events such as triggering environment explosions or picking up weapons dropped by enemies.

Another pattern located in the data is the progressively higher coincidence of the camera vector (where is the player looking at) with the character vector (in which direction is the player moving), showing little interest, in the later deaths, towards checking around corners. The fourth attempt proved to be the most unsuccessful, lasting only few seconds (Figure 3), showing the player rushing into the enemies and failing almost instantly. After this failure the player appears to regain control, slowing the pace of movement, attempting a new route (leftward turn), killing a considerable amount of enemies and taking the time to pick up dropped weapons.

A second and a third set of data where the participant under investigation reported frustration in a follow-up interview after the session, were analyzed to see if the same behavioral patterns were visible in the metrics in different game locations, with positive results (same patterns of behavior in-game and ex-game were located) (Figure 3).



Figure 2. A simple visualization using ArcGIS, showing the path of a player navigating a level in Kane & Lynch: Dog Days, until the first occurrence of death (red circle in the lower center of the map). The location of the player is plotted at each second of playtime, and a color scale applied to show the dimension of time along the path. Various events are plotted as symbols: enemy kills (blue dots), weapon pickups (red triangle) and taking cover (green squares). Spatial metrics visualizations such as this one are highly useful for the detailed evaluation of gameplay and balancing in shooter games.

4.3 Eksperimentation: Following identification of a possible pattern in player behavior (in-game, as expressed via game metrics data) indicating frustration, metrics data from KL2 from a sample of 22 randomly selected players among the IOI tester base were analyzed.

The metrics data gathered from each of the 22 players was comprehensive. The temporal sampling interval was one second, which include the location of the players, the vector of the avatar, the vector of the virtual camera, the health of the player, movement modifiers (walking, running, sprinting), whether the players were crouching or not and whether the players were “in cover” or not. In addition, triggered (event-based) were analyzed, which include checkpoint activation, picking up weapons and ammo, making use of exploding objects in the level, being “down but not dead”, killing of Non-Player Characters (NPCs) and player deaths.

The gameplay metrics data were investigated using procedural algorithms, with the aim of exploring whether or not the in-game

behavioral patterns correlating with player frustration in the initial test participant, could be located in the data from the 22 additional testers. It is understood that other players might also have experienced frustration but reacted in the game in very different ways.

The first step was to generate a query that followed the parameters identified initially, and applying it to the data from the 22 participant sample.

The second step was to dynamically visualize the tracked variables, to allow visual inspection and spatial analysis. There are many ways to visualize navigational data and various ways of handling data visualizations, which are flexible enough to handle a variety of contexts. In the current case, a Geographic Information System (GIS), built using the package ArcGIS (Figure 3), was created to provide spatial visualization [9, 10, 16]. This form of visualization allows experimenters to see through the eyes of the player in a manner similar to a video recording of a game session, but with the added benefit of having recorded metrics mapped within the game environment, and the ability to draw quantitative results from these.

The spatial analysis of the metrics data from the 22 participants, during the times when their behavior expressed similarities with the periods where the initial participant experienced frustration, indicated that several factors of in-game behavior appeared to be radically different during these intervals as opposed to other randomly selected segments of play, notably number of death events, the movement speed of the participant in the virtual environment of KL2, etc. The factors are generally correlated – for example, if a player death event (Pd) occurs within 2 minutes, the average pace (speed) of movement of the player (Pm) will increase compared to the average movement pace for the entire game, and the number of NPCs killed decrease progressively between each death happening. Similarly, the number of weapons and ammunition supplies picked up (WApu) decreases progressively as players die more and more often within shorter time intervals.

Based on these specific behaviors, a model was developed specifying the timing and frequency of the behaviors identified, as follows:

$$\begin{aligned} t_n < t_{n+1} \\ Pd > 2 \\ 0 < Pdl < 20 \\ Pmf > Pm \\ NPCd(t_n) > NPCd(t_{n+1}) \\ WApu(t_n) > WApu(t_{n+1}) \end{aligned}$$

Where:

- Timestamp (t). The timestamp is set to zero the moment a new play session begins. <tf> describes a time interval that has been identified as “frustrated”
- Number of player’s deaths (Pd), <Pdl> expresses location of player deaths in world units.
- Player’s pace of movement (Pm) measured as distance in space travelled in one second, averaged for the whole play session. <Pmf> defines the average pace of player movement during an interval of time identified as “frustrated”
- Number of NPCs killed (NPCd).
- Number of weapons or ammo picked up (WApu)

Importantly, all conditions need to occur simultaneously for the model to contain all the indications of player frustration reported in the initial part of the case study.

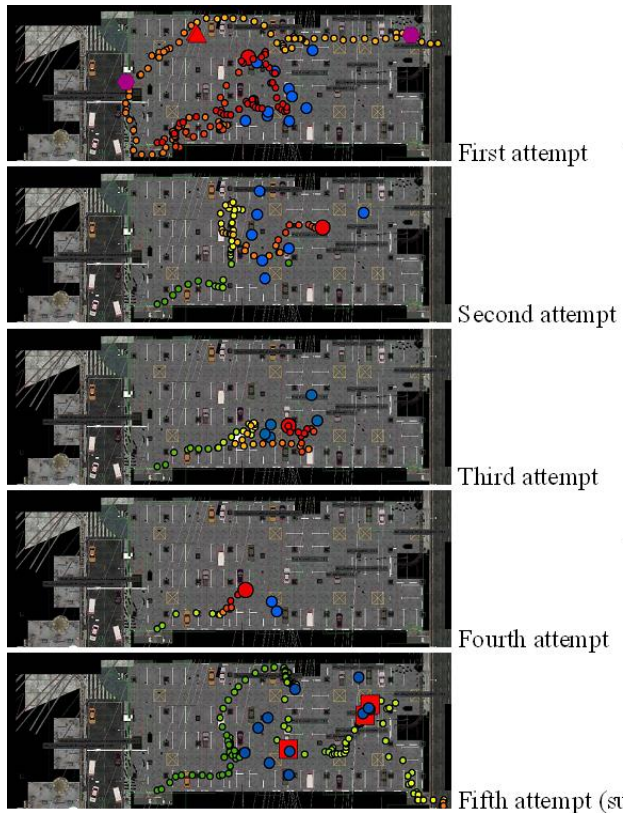


Figure 3. The images show the path of a single player (participant) and specific events that occurred during the gameplay session used in the case study. Each image represents the time segment from one instance of player death to the next, showing decreasingly less progress in the game from death 1-4; indicative of a behavioral pattern pointing towards player frustration (red dots: player deaths, blue dots: NPC kills, red square: weapon or ammo pickup, red triangle: environment explosion, purple dots: checkpoints, small dots: player position in time).

4.4 Follow-up interviews: Following analysis of the metrics data and definition of the model indicating frustration, the model was applied to the data from the 22 participants. In six of these participants, patterns similar to the ones identified previously were identified using relational queries (quantitative verification). I.e., these six participants displayed the simultaneous reoccurring patterns of play behavior that the model defined specifies indicate player frustration.

The six participants were called in for open interview sessions, which aided by video recordings of their play sessions and game metrics visualizations attempted to uncover if they had experienced undesirable frustration during the period identified by the quantitative model. A browser-based tool (Figure 4), “G-player” was developed to show animated replays of test sessions, and proved invaluable when reconstructing the experience with

the players. The participants confirmed that in all of the segments of play identified by the model, they had experienced undesirable frustration (i.e. frustration contrary to the user experience).

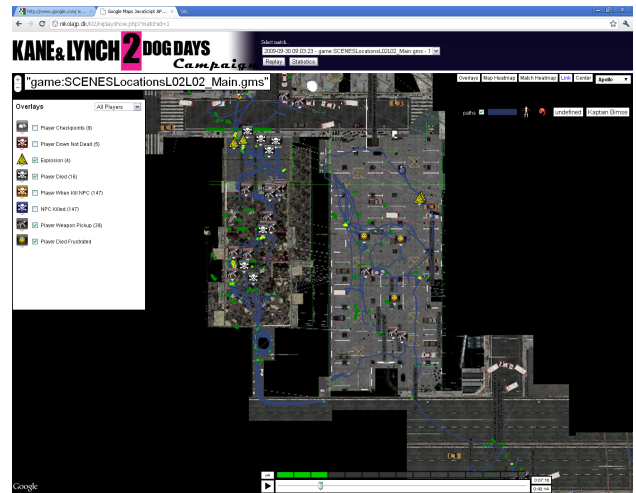


Figure 4. The G-Player dynamic visualization tool [20]

5. DISCUSSION AND CONCLUSIONS

The model of frustration developed for this KL2 case study is specific to the game, and no claims are made here as to its validity outside the confines of KL2.

The results of the case study are encouraging in that they indicate that it may be possible – for specific games at least – to develop behavioral models that can be implemented in automatic warning systems that raise red flags in user-testing situations (or to designers directly). Such automated systems might not capture all instances of player frustration, but only those integrated in the model used as the basis. However, considering how fast and inexpensive it would be for frustration detection systems to operate, it might be worthwhile the initial analysis. This notably because in a real-life user testing situation, it is not realistic to expect user researchers to keep track of all of these variables while running user tests, highlighting the usefulness of automated game metrics tracking and recording as a tool for game user research/user-testing.

Future research will investigate whether the model translates to games of the same genre (3rd person action games), or perhaps even other genres. Ultimately, it may be possible identify universal markers enabling the automatic detection of frustration problems during instrumentation-based evaluation of play experiences in all phases of game development.

Besides the obvious advantage for game development of being able to recognize patterns of frustration in the metrics to fix troublesome areas in the level, these patterns could also theoretically be used as an element for adaptive gameplay.

Similarly, it may be possible to identify and isolate markers for other types of experiences: fear, joy, triumph, etc. It is however doubtful whether the full set of emotions that make up the play experience can be detected via in-game behavioral data, but if correlates for some even types of experiences and for some players can be found, that would be enough to establish automated diagnostic systems that could raise red flags pointing to potential problems. Due to the fact that not all detected cases

necessarily lead to frustration and also to the fact that not all episodes of frustration might be detected, the intervention of a human expert, whether it is a user researcher, a level designer or a game designer, will always be needed to evaluate each case and make the final call. Nevertheless such a model can ease the burden on the shoulders of user researchers during play test observations since in modern games the number of factors that must be monitored in order to evaluate the player experience is very high. Automated detection systems allow researchers to focus on the physical side of players, their bodies, gestures and mannerisms that occur during play.

On a final note, the patterns identified in the case study described here are based on the definition of frustration as being failure to overcome challenges. Other forms of frustration were not considered.

In conclusion, the most valuable outputs from this case study are:

- The 'new' tools (metrics/quantitative) were anchored in the concerns and observations of the existing (qualitative) research framework and therefore can be more easily integrated, and could more easily feed back into the framework and actually be used as an interview-prop.
- Focusing on a ubiquitous phenomenon in game experience design (frustration/loss of commitment) rather than a game or level specific point of analysis allows industry and academia, normally processing at very different speeds, to function in unison bridging that gap.
- Data mining extremely dense play sessions, even if from a small number of participants, and combining qualitative methods, enables use of metrics with smaller participant-sample sizes than usual.

6. ACKNOWLEDGMENTS

The current work with gameplay metrics conducted with Square Enix Europe showcased in this article is carried out by a variety of people, from the Square Enix Europe Online Development team to the user-research experts at IO Interactive and Crystal Dynamics; the authors would like to thank Thomas Hagen, Jim Blackhurst, Tim Ward, Ian Thomkins, Kim Krogh and Nicolai Pedersen.

7. REFERENCES

- [1] Amsel, A. 1990. Arousal, suppression, and persistence: Frustration theory, attention, and its disorders. In *Cognition & Emotion*, Psychology Press.
- [2] Berkowitz, L. 1969. The frustration-aggression hypothesis revisited. In: *Roots of aggression*, Atherton Press, New York.
- [3] Brave, S. B. and Nass, C. 2003. *Agents that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent*, Doctoral Dissertation, Stanford University Stanford, CA, USA.
- [4] Canossa, A. and Drachen, A. 2009. Patterns of Play: Play-Personas in User-Centered Game Development. In *Proceedings of DIGRA '09* DiGRA Publishers, London, UK.
- [5] Csikszentmihalyi, M. 1990. *Flow: The Psychology of Optimal Experience*. New York: Harper and Row.
- [6] Davis, J. P., Steury, K. and Pagulayan, R. 2005. A survey method for assessing perceptions of a game: The consumer playtest in game design. In *Game Studies*, 5, 1.
- [7] DeRosa, P. 2007. *Tracking Player Feedback to Improve Game Design*. Gamasutra. Available from: http://www.gamasutra.com/view/feature/1546/tracking_player_feedback_to_.php.
- [8] Dollard, J. and Miller, N. E. (Eds.). 1939. *Frustration and aggression*, Yale University Press, New Haven, USA.
- [9] Drachen, A. and Canossa, A. 2011. Towards Gameplay Analysis via Gameplay Metrics. In *Media in the Ubiquitous Era: Ambient, Social and Gaming Media*. IGI Global Publishers.
- [10] Drachen, A. and Canossa, A. 2011 Evaluating Motion: Spatial User Behavior in Virtual Environments. In press for *International Journal of Arts and Technology* - Special issue on ACM Academic MindTrek '09: Everyday Life in the Ubiquitous Era. Inderscience Publishers.
- [11] Hazlett, L. R. 2006. Measuring emotional valence during interactive experiences: boys at video game play. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*. ACM New York, NY, USA.
- [12] Ijsselstein, W et al. 2007. Characterizing and Measuring User Experiences in Digital Games. In *Proceedings of ACE Conference'07*, Salzburg, Austria .
- [13] Isbister, K. and Schaffer, N. 2008 (Eds.). *Game Usability: Advancing the Player Experience*. Morgan Kaufman.
- [14] Kiel, M., Gilleade, K. M. and Dix, A. 2004. Using frustration in the design of adaptive videogames. In *Proceeding of the 2004 ACM SIGCHI International Conference on Advances in computer entertainment technology*, ACM New York, NY, USA .
- [15] Kim, J. H., Gunn, D. V., Schuh, E., Phillips, B., Pagulayan, R. J. and Wixon, D. 2008. Tracking real-time user experience (TRUE): a comprehensive instrumentation solution for complex systems. In *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems '08*, Florence, Italy.
- [16] Longley, P. A., Goodchild, M. F., Maguire, D. J. 2005. *Geographical information systems: principles, techniques, management, and applications*. John Wiley & Sons.
- [17] Mellon, L. 2004. *Metrics in MMP Development and Operations*. Game Developer Conference presentation. Available from: <http://www.maggotranch.com/gdc2004.ppt>.
- [18] Norman, D. A. 1988. *The psychology of everyday things*. Basic Books, New York, USA.
- [19] Pagulayan, R.J., Keeker, K., Wixon, D., Romero, R.L. and Fuller, T. 2003. User-centred design in games. In *The Human-Computer Interaction Handbook*, Erlbaum, Mahwah, NJ, USA.
- [20] Pedersen, N. 2010. *G-Player*. Available from: <http://nikolajp.dk/k12/replayshow.php?matchid=1>.
- [21] Rohrer, C. 2008. *When to Use Which User Experience Research Methods*. Available from: <http://www.useit.com/alertbox/user-research-methods.html>.
- [22] Rosenzweig, S. 1944. *Personality and the behavior disorders*. Ronald Press, New York, USA.

- [23] Scheirer J., Fernandez R., Klein J. and Picard R.W. 2002. Frustrating the user on purpose: A step toward building an affective computer. *Interacting with Computers - the interdisciplinary journal of Human-Computer Interaction*, 14, 2, 93-118.
- [24] Silverman, D. 2001. *Interpreting Qualitative Data: Methods for Analysing Talk, Text and Interaction*. Sage Publications Ltd; 2nd edition.
- [25] Swain, C. 2009. Improving academic-industry collaboration for game research and education. In *Proceedings of the 4th International Conference on Foundations of Digital Games*. ACM New York, NY, USA.
- [26] Thompson, C. 2007. How Microsoft Labs invented a New Science of Play. *Wired Magazine* 15, 9. Available from: http://www.wired.com/gaming/virtualworlds/magazine/15-09/ff_halo?currentPage=all.
- [27] Tychsen, A. and Canossa, A. 2008. Defining Personas in Games Using Metrics. In *Proceedings of FUTURE PLAY '08* ACM publishers, 73-80, Toronto, Canada.
- [28] Vorderer, P. and Jennings, B. 2006. *Playing Video Games: Motives, Responses, and Consequences*. Lawrence Erlbaum Associates, Oxford, UK.
- [29] Williams, D., Yee, N. and Caplan, S. E. 2008. Who plays, how much, and why? Debunking the stereotypical gamer profile. *Journal of Computer-Mediated Communication* 13, 993-1018.
- [30] Drachen, A., Canossa, A. and Yannakakis, G. 2009. Player Modeling using Self-Organization in Tomb Raider: Underworld. In *Proceedings of IEEE Computational Intelligence in Games (CIG) '09*, IEEE Publishers, pp. 1-8. Milan, Italy.
- [31] Mahlmann, T., Drachen, A., Togelius, J., Canossa and A. Yannakakis, G. N. 2010. Predicting Player Behavior in Tomb Raider: Underworld. In *Proceedings of IEEE Computational Intelligence in Games (CIG) '10*, IEEE Publishers. Copenhagen, Denmark.
- [32] Pagulayan R. and Keeker, K. 2007. Measuring Pleasure and Fun: Playtesting. In *Handbook of Formal and Informal Interaction Design Methods*. Morgan Kaufmann Publishers.
- [33] Nielsen, J. and Landauer, T. K. 1993. A mathematical model of the finding of usability problems. In *Proceedings of ACM INTERCHI'93 Conference*. Amsterdam, The Netherlands.