Towards Gameplay Analysis via Gameplay Metrics

Anders Drachen
IT University of Copenhagen
Rued Langgaards Vej 7
2300 Copenhagen S, Denmark
+0045 72185038
drachen@itu.dk

Alessandro Canossa
IO Interactive/Danish Design School
Kalvebod Brygge 4
1354 Copenhagen K, Denmark
+0045 30233284
alessandroc@ioi.dk

ABSTRACT

User-oriented research in the game industry is undergoing a change from relying on informal user-testing methods adapted directly from productivity software development to integrating modern approaches to usability- and user experience testing. Gameplay metrics analysis form one of these techniques, being based on instrumentation methods in HCI. Gameplay metrics are instrumentation data about the user behavior and user-game interaction, and can be collected during testing, production and the live period of the lifetime of a digital game. The use of instrumentation data is relatively new to commercial game development, and remains a relatively unexplored method of user research. In this paper, the focus is on utilizing game metrics for informing the analysis of gameplay during commercial game production as well as in research contexts. A series of case studies are presented, focusing on the major commercial game titles *Kane & Lynch* and *Fragile Alliance*.

Keywords: gameplay metrics, game metrics, user experience, user testing, computer games.

1. INTRODUCTION

The game development industry has within the past decade established itself as a major component of the interactive entertainment industry, rivaling the movie industry in size [1]. Paralleling the dramatic development of the sector, computer games have gone from simple text-based adventures to almost photo-realistic renditions of virtual worlds within a variety of genres, offering a wealth of entertainment opportunities to their users. Contemporaneously with this development, a requirement for user-oriented testing methodologies that take into account the unique nature of digital games has become increasingly prevalent [19,27]. Furthermore, with the increasing complexity of contemporary computer games – in terms of the amount of possible user actions and –behaviors that they afford, as well as the breath of interaction options between the user and the software/hardware – the informal testing methods traditionally utilized in the game industry, which were adopted directly from productivity software testing, have come under pressure [24,27].

MindTrek 2009, September 30th-October 2nd 2009, Tampere, FINLAND.

Game testing can generally be divided into technical, functional and content testing. **Technical** testing is focused on issues relating to the game engine itself and associated hardware. Technical testing includes e.g. compliance, compatibility and soaking tests. **Functional** testing is concerned with bug hunting, stability, integrity of game assets, general problems with the game itself and gameplay or problems with the user interface, game controls etc. **Content** testing is oriented towards presentation and graphics, level design, the game story and how these elements are communicated to and perceived by the player (user). User-oriented testing (usually residing under Quality Assurance, QA) is essential to game production, because the quality of the product that a game is directly relates to the perceived user experience. Functional and content testing are therefore two areas that receives an increasing amount of attention from academic and industry environments alike [e.g. 13,15,17,18,19,23,24,26,31,32]**.** The purpose of user-oriented game testing is to see how specific components of, or the entirety of, a game is played by people; allowing designers to evaluate whether their ideas and work provides the experience they are designed for.

User-oriented game testing is normally carried out at different stages of the production cycle of digital games, which are commonly produced using agile methodologies. An organized, methodical approach to QA is vital, as this permits that issues are fixed as they arise rather than at the end of a production cycle. The major period of testing takes place during the alpha-stage, where a playable, full version of the game is (ideally) ready; and furthermore at the following beta-stage. These two periods of testing are usually intensive, with batteries of user-tests and a requirement for very fast turnaround on test results. Bugs and tweaks often need to be integrated within the span of hours or days, so that retesting of solutions can take place [31]. Testing also takes place earlier in the development phases, focusing for example on evaluating concept art and character designs in early phases, and gameplay balancing further into the production cycle. Following game launch, testing of software patches and updates is required, e.g. to insure that imbalances are not created – this is notably crucial to online games with elements of player vs. player competition.

The requirement for improving user-oriented testing methodologies within game development, the increasing complexity of digital games, the variety of aspects of the user interaction that needs to be tested, and the requirement for methods that do not require long turn-around times, has resulted in a variety of approaches from Human-Computer Interaction (HCI) research [13] being adapted user-oriented game testing and –research [6,17,19,24,26,28]. These include different forms of usability testing, ethnographic methods, experience testing etc. All of these have specific strengths and weaknesses, but are generally useful for capturing player feedback and subjective user experiences; and for acquiring in-depth feedback on e.g.

gameplay or design problems. However, these approaches are limited in that information is often hand-coded (surveys, analysis of audio-visual recording), meaning that getting highly detailed data about user behavior is either incredibly time consuming or downright impossible.

A potential source of supplementary data to accommodate this limitation is the automated collection and analysis of **instrumentation data**, an approach utilized within the general HCI field [e.g. 24,29], but however only recently adapted to computer game production [19,33]. Different types of instrumentation data can be recorded from player-game interaction. In game development, instrumentation data find uses within e.g. engine performance, sales across different countries or regions, project progress or user interaction with the game software, the latter category being of interest here. Within the context of user-oriented testing, instrumentation data related to player-game interaction are generally termed **gameplay metrics** [32,35], and serve to provide detailed quantitative information about the player (user) behavior. Gameplay metrics form objective data on the player-game interaction. Any action the player takes while playing can potentially be measured, from low-level data such as button presses to in-game interaction data on movement, behavior etc.

The term "metric" – as it is used here - stems from computer science, and denotes a standard unit of measure, with metrics generally being organized in systems of measurement, utilized in the evaluation and measurement of processes, events, interaction etc. [2,10,21]. Importantly, game metrics are not game heuristics – the latter are design principles that games can be build from, where game metrics are instrumentation data derived from game engines. In general, gameplay metrics can be recorded for any type of user-initiated behavior where interaction takes place in or with the virtual environment; as well as behaviors initiated by agents or systems operating in the virtual environment outside of the control of the player, e.g. autonomous agents [34]. The analysis of user behavior via gameplay metrics act as a supplement to the established methods for user-oriented research in the game industry and –research. For example, usability testing focuses on measuring the ease of operation of a game, while playability testing explores is users have a good playing experience [24]. Gameplay metrics analysis offers however insights into how the users are actually playing the games being tested. This makes this type of data uniquely suited to form the basis for gameplay evaluations.

In this paper, two case studies are presented that showcase novel forms of gameplay analyses that can be performed via the application of gameplay metrics. The case studies showcase new ways to analyze gameplay metrics and take advantage of the spatial dimension of certain metrics in order to target **gameplay analysis**. They are also indicative of the ability of gameplay metrics analysis to provide detailed data on player behavior, thereby providing a tool for not only game development and –design; but also general user-oriented research in interactive entertainment. While the case studies presented are based on two specific commercial titles, *Kane & Lynch* and *Fragile Alliance* (both: 2008, *IO Interactive*), the two games are examples of the highly popular shooter genre, and represent the two dominating modes of play within this genre: Single-player (*Kane & Lynch*) and multi-player (*Fragile Alliance*). The methods described in the case studies are therefore cross-game applicable, directly transferrable to other shooter games (e.g. *Doom 3*, *Unreal Tournament*, *Bioshock* and *Crysis*), as well as other game genres utilizing a central character as the vessel for game-player interaction, i.e. also many role-playing games and adventure games such as *Oblivion*, *Neverwinter Nights* and *Dreamfall*. Additionally, massively-multiplayer online games such as *Age of Conan* and *World of Warcraft* and online persistent worlds such as *Second Life*, as well as other Virtual Environments (VEs), have players taking control of a single 3D character, and therefore also form potential targets for the kinds of analysis presented.

The work presented here is being carried out in collaboration between Danish game developer *IO Interactive* (a subsidiary of *EIDOS Entertainment*), and the *IT University of Copenhagen.*

2. PREVIOUS WORK

The focus of this paper is not to review existing work on gameplay metrics but rather to showcase a series of developed analysis methods for evaluating gameplay. However, a brief overview of the related work is presented here.

The literature on game metrics is minimal, even including the information available from industry sources such as the Game Developers Conference. The handful of published examples of gameplay metrics analysis is restrictive in the amount of information that they provide [e.g. 5,19,17,25,35]. This is in part due to the novelty of the approach, in part because gameplay metrics data would normally be treated as confidential by a company. Furthermore, the few examples of systems developed for capturing game metrics-like data developed outside of the academia are generally targeted at Virtual Environments (VEs) rather than games specifically [e.g. 3,4,5,14]. Additionally, these applications are often targeted at analyzing specific features, such as movement, or developed for use with specific applications/games [e.g. 14], and therefore not portable across environments.

A substantial part of the current published material about gameplay metrics analysis stems from the work of *Microsoft Game Labs*, which perform game testing and user-oriented research for the various Microsoft-based game studios. *Microsoft Game Labs* developed e.g. the TRUE setup to capture gameplay metrics together with survey and video data. The approach has been applied to user testing of e.g. *Halo 2* and *Shadowrun* [19,33]. Among the most recent uses of game metrics in a research context is the work by Williams et al. [36], who examined user data from *EverQuest 2* in combination with survey-based information to evaluate player behavior, e.g. in terms of how many hours they spend gaming per week. Other examples include the data presented by Mellon [25] from the massively multiplayer online game (MMOG) *The Sims Online*; while Goetz [12] presented analysis of interface-based metrics from e.g. *Civilization IV*. Finally, DeRosa [8] reported from the use of time-spent reports developed at *Bioware* for their international hit game *Mass Effect.* Tychsen & Canossa [35] described a model for using gameplay metrics to describe personas of players in game design & -testing. What is apparent from the literature is that there is a need to open up the discussion about how to utilize gameplay metrics analysis in game production and –research, as the relatively limited size of the literature on the topic betrays the increasing amount of attention and use that the automated collection of gameplay metrics has gained within the past five years or so [32]. Notably within massively multi-player online games such as *World of Warcraft* and *Age of Conan* is the tracking of at least some basic gameplay metrics today a common practice during the development and live periods of the lifetime of these games [see e.g. 25]. In terms of using gameplay metrics as a user-oriented testing method, this is not without specific challenges.

Where methods such as usability-testing and playability-testing focus on establishing whether a player can interact with a game effectively, and if doing so is fun, gameplay metrics-analysis is targeted at clarifying what the player is actually doing, and sometimes why the behavior indicated arises. Importantly, the "why" is hypothesis-based: In order to assess the motivations and reasons behind player behaviors, it is often necessary to combine gameplay metrics-analysis with other user-oriented testing methods. Furthermore, gameplay metrics does not inform about the gender, age or other demographics of the player. As such, the approach supplements existing methods, addressing the key weakness of standard user-oriented game testing methods, namely the lack of detailed and objective data. Finally, it should be noted that order to enable metrics-based analysis, an infrastructure is needed to capture the data, which includes substantial storage needs in the case of large commercial titles such as *Kane & Lynch*.

In summary, the literature on gameplay metrics and game metrics analysis is largely based on data from in-house testing, with a few MMOG-related examples of data derived from installed clients or game servers such as Williams et al. [36]. Analyses have, with the exception of the work by *Microsoft Game Labs*, focused on high-level aggregate-counts of metrics. The spatial component of gameplay metrics have not been the focus of published work outside of the heatmaps (game level maps showing the locations of player character death aggregated for a number of players), which have been published for games such as *Half Life 2* and *Team Fortress 2;* and visualizations of player progression as a function of time for *Halo 3* [33]. Previous work with metrics has generally included only a single variable at a time, e.g. level completion times, and the potential for performing analysis on the spatial dimension of gameplay metrics (e.g. the position of the player within the virtual environment) limited to handful of industry- and academia-derived applications. There is therefore a substantial room for development of novel approaches towards utilizing gameplay metrics, and this forms the overarching goal of the research presented here.

3. CASE STUDIES

The data used as a basis for the case studies presented here is test data, i.e. data from in-house testing with professional testers or representatives of the target audience for the games in question. A varied amount of game sessions/number of people are used in the datasets for the different case studies, however, it should be noted that essentially the data suffer from a bias because they all stem from in-house testing at *IO Interactive*, not testing carried out in an independent laboratory. However, the purpose of the case studies is not to present conclusions from experiments performed, but to showcase techniques and methodologies, which can be employed in both an industry and research-oriented context. It should also be noted that the data are derived from tests run on unfinished versions of the games involved, and the data therefore do not represent material from the finished games. This is in the current context not a problem in terms of the quality of the analyses presented, rather it adds strength to the argument of e.g. [19,35] that gameplay metrics form a valuable supplement to other user-oriented research methods during the production phases of computer game development (where testing is carried out iteratively with more and more complete versions of the game/game levels). Due to the confidential nature of the gameplay metrics data, no absolute numbers are presented.

Data for the case study is drawn from the *EIDOS* metrics suite, developed by the *EIDOS Online Development Team* at *IO Interactive*. The metrics suite is an instrumentation solution engineered to provide *EIDOS* studios with metrics data for e.g. user behavior analysis. Logged information is streamed to a central server for extraction and analysis. The metrics suite can collect data from game engine software during production as well as from installed clients during the live period of the lifetime of a game (i.e. following launch and distribution to the customers) [19,35]. In addition to utilizing gameplay metrics in user-oriented testing, *IO Interactive* and other *EIDOS* developers potentially involves a battery of methods including audiovisual recording and analysis, survey-based approaches, expert testing, different forms of usability testing, biofeedback, eye tracking etc. [27]. The combination of gameplay metrics analysis with existing methods for user-oriented game testing permits strengthens across-method triangulation of results.

Following transmission from the game engine, the data is transformed through and ETL process (Extract, Transform and Load) and imported to an SQL server. From the server, analysts extract data for analysis and visualization**,** creating reports for the interested parties (generally QA, designers, producers and marketing). The transformation of raw gameplay metrics data drawn from the EIDOS metrics suite, into reports disseminated to the interested stakeholders requires a series of potentially iterative steps.

When performing analyses of gameplay metrics with a spatial component, data are imported into a geodatabase system from where files are drawn, analyzed and visualized using a Geographical Information System [22], ArcGIS. Some of the gameplay metrics can be quickly visualized using a custom in house-produced tool which adds data on top of level maps. This combination of tools permits both rapid visualization of basic information, as well as in-depth analysis of user-behavior data performed by analysts.

Case study 1: Level analysis by sub-sector in *Fragile Alliance*

Fragile Alliance is similar to many other team-based multi-player shooters, e.g. *Unreal Tournament* and *Quake Arena*; however, it has a twist: It pitches players as mercenaries trying to accomplish a heist. A game session will typically consist of multiple rounds being placed on the same map (scenario) and/or different maps. The winner of a round is the player who leaves with the most money, irrespective of how these are obtained. At the end of the round, all money stolen is divided among surviving mercenaries. If a mercenary dies, they respawn (are reinstated in the game universe) as police officers, working along with a group of AI-controlled autonomous agents ("bots") playing the same role. This creates a situation where the balance of power initially will typically be on the side of the mercenaries, but shift towards the police (AI and players). After the second death, the player will typically not respawn, but will have to wait for the game round to end (usually after a few hundred seconds depending on the map). Mercenaries run the risk of being killed by AI-police, as well as police players; however, there is a nother substantial threat is other mercenaries: *Fragile Alliance* allows mercenaries to betray each other. If for example a mercenary player had managed to secure a sum of money from a bank vault, another mercenary could kill the first, and steal his/her money. If a mercenary kills another mercenary he becomes a "traitor" but is allowed to keep all the money collected without sharing (Figure 1).

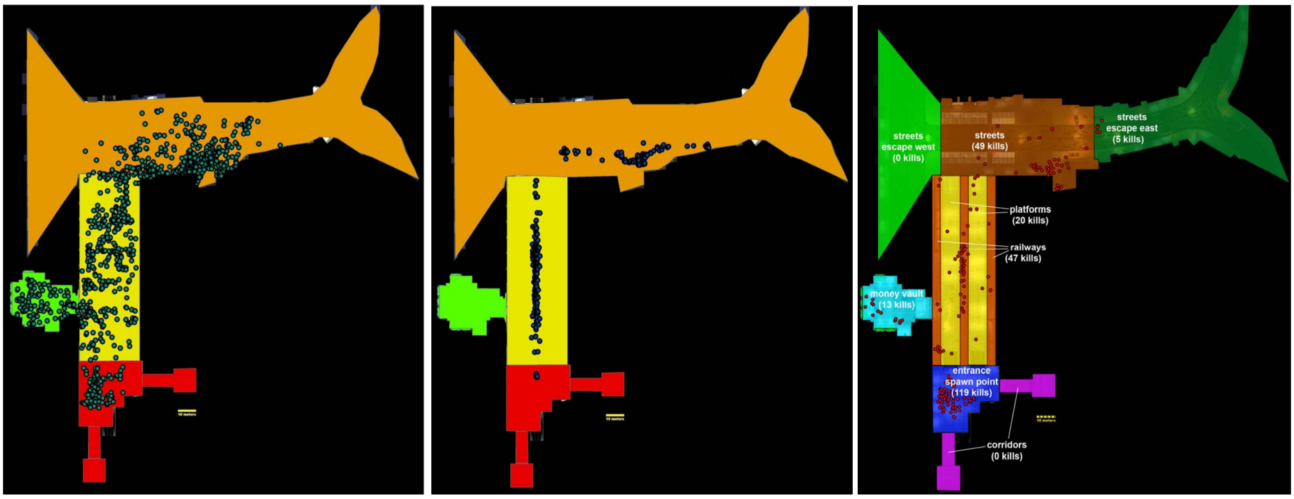


Figure 2: The *Fragile Alliance* level map divided into four sub-sections: Red = spawning area; Yellow = subway; Green = vault area; Orange = road/exit area. Left: The locations where police officers were the cause of death have been mapped on top of the sub-sector map using overlay analysis in ArcGIS (data from approximately 50 game sessions, 6000 total kills recorded). A broad distribution is apparent indicating that police officers can reach the entire map. Middle: Locations of suicide-caused death events. The events in the yellow sector of the map are caused by players being run over by a subway train while crossing a set of tracks, while the suicides in the orange zone are caused by exploding scenery (e.g. cars that explode after becoming too damaged by weapons fire). Right: An example of feedback to the game designers following metrics analysis of the game level. The map shows the distribution of about 250 player death occurrences overlain the level map, and has added explanations to guide the interpretation of the map. The map is developed in ArcGIS. Two layers of data have been added to a map of a level from the *IO Interactive* produced game *Fragile Alliance*. One layer subdivides the game level map into sub-sectors; a second consists of point-based plots of the locations of player character deaths. Using ArcGIS, the number of points in one layer intersecting with the polygons in the other has been calculated.

If a player kills a traitor they receive an instant reward; however, if a police player kills the traitor that killed him as a mercenary, he will reap a bigger revenge-reward. If a police player secures an amount of loot money from a mercenary, he keeps a percentage as a finder's fee. The purpose of the mercenaries is to escape with as much money as possible – either by working independently or together.

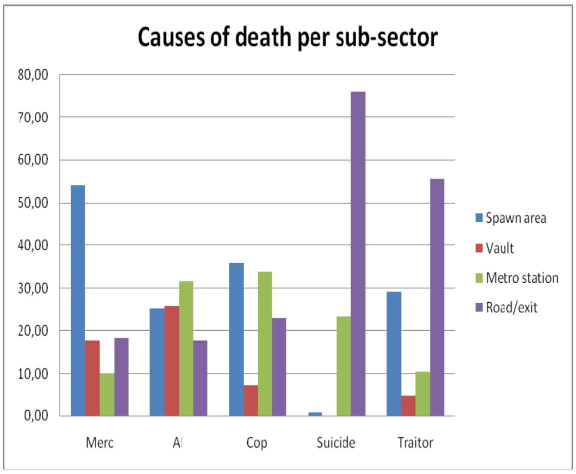
Figure 1: Screenshot from Fragile Alliance, showing a Traitor clearly marked to the remaining mercenary players.

One of the key design concerns when creation scenario-based FPS-levels is ensuring that the right events take place in the right locations. For example, in the current map the mercenaries should at least in some cases reach the exit and complete the mission, rather than being gunned down by the police in the vault area (Figure 2). This question was addressed by analyzing the **spatial patterns** of player death. This is a powerful tool for analyzing player behavior, as the gameplay metrics data permit plotting of locations of player death – and the causes – with pinpoint accuracy. This allows fine tuning of the design of the game level.

In the map level selected for this case study, the mercenaries spawn in the bottom of the map, the police AI agents to the top right (Figure 2). The objective of the mercenaries is a vault, located to the left in the map, and thereafter to reach the level exist, in the top right corner, behind the spawning area of the police. The game level consists of four sub-sectors: The **spawning area**, where the mercenary players enter the game and begin addressing the heist objective (red in Figure 2). The **vault area**, where the money they need to steal are located (green in Figure 2), a **subway station** area approximately in the middle between the spawning area of the police (AI and players) and mercenaries (yellow in Figure 2) and finally an area at street level (orange in Figure 2), through the rightmost side of which the mercenary players go if they want to escape the map (i.e. complete the heist or run away).

Dividing the level map into sub-sectors permits an even more detailed analysis of the distribution of the death events (Figure 3). By combining visualization of the spatial behavior of players with statistics of their temporal (and spatial) behavior, a more thorough understanding of the player behavior is gained as it is now possible to compare spatial and temporal behavior. The analysis shows that mercenaries primarily turn traitor in the beginning of the game in the spawn area, but most commonly (55.72%) in the road/exit area – i.e. when the mercenaries are close to getting out with their stolen booty. This pattern of behavior confirms to the designers at *IO Interactive* that at least in this respect, the behavior of the traitors is as expected from the design of the game level. For the mercenaries, the majority of the kills occur in the spawning area, where mercenaries enter the game and later Figure 2). The AI agents are spread across the entire map, indicating that their search & destroy behavior is working excellently. Suicides occur in the vast majority of cases (76.04%) in the road/exit area, where a series of cars are placed which can explode if taking too much damage. A smaller part takes place in the metro station area, where players can be hit by metro trains while crossing the tracks coming from the vault to the exit/road area to the north in the map (Figure 2).

In terms of the roles played by players when they are killed, the pattern is generally as would be expected from the gameplay design. Police (players and AI) are typically killed in the road/exit area where they spawn (69.32%), and very few are killed in the spawn and vault areas, where instead the mercenaries are under pressure from the police (44.17%). Interestingly, traitors are typically killed in the spawn area (61.25%), but rarely in the road/exit area (8.81%), which indicates that it is a much more risk-filled endeavor to turn traitor early in the game rather than later (note that spatial analysis shows that mercenaries turning traitor outside of the spawning area rarely move into the spawning area again – by this point the action has moved to the other segments of the map).



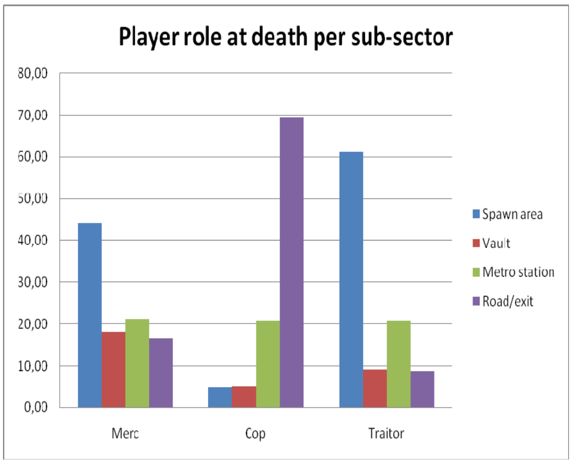


Figure 2: (Top) diagram showing the causes of player death as a function of the *Fragile Alliance* level map sub-sector. (Bottom): The role of the player at the time of death as a function of the map sub-sector. Same data source as for Figure 2.

In summary, the behavioral patterns of the players in terms of death events in the specific level of *Fragile Alliance* appear to follow closely the intents of the designers. It should be noted that a somewhat larger amount of death events occur in the spawn area (lowermost part of the map) than is ideal, which could indicate that mercenaries are perhaps a bit too eager to turn traitor early in the game. This may not be a problem to the user experience but rather the opposite. However, gameplay metrics cannot inform about the quality of the user experience on its own, meaning that this question is a subject that should be analyzed in combination with other user-oriented game testing. This example highlights the key limitation of gameplay metrics analysis: It can inform what players are doing, but not always why.

While the causes of death in *Fragile Alliance* are specific to the game, the basic multi-player shooter game platform is among the more popular among online games. Furthermore, while the players have specific roles unlike in other multi-player shooters such as *Unreal Tournament,* it is increasingly common to see multi-player shooters employing character classes, i.e. the player chooses a character class with a specific team (assault, sniper, special agent, etc.). This is known from games such as *Call of Duty*, *Team Fortress 2* and the *Battlefield*-series. The approach is therefore directly transferrable to these games. In single-player games (even non-shooters), causes of death will not be other players, but for example different types of computer-controlled enemies or environmental effects. The principle of the analysis remains the same, as to the goal of locating "trouble spots" in game levels/areas where the patterns of death are not as intended by the game design.

Case study 2: Perfect paths in *Kane & Lynch*

A key strength of gameplay metrics analysis is the ability to recreate the playing experience of the player in detail. Being able to model the navigation of players through a game environment is of interest to game development for a number of reasons, not the least because it allows designers to observe how their games are being played. Other methods of user-oriented testing of computer games can also locate problems with gameplay, e.g. reporting that a specific encounter is too difficult. However, when integrating gameplay metrics, the second-by-second behavior of the players can be modeled – this enables much more detailed explanations for the observed behaviors. Furthermore, patterns of behavior are much easier to establish (and less time consuming to produce) when based on gameplay metrics as compared to data from play- or usability testing. Gameplay metrics analyses that focus on recreating the player experience generally belongs to the more in-depth types of analysis, where the focus is on capturing and analyzing metrics data in great detail and across a large number of variables, typically from a limited number of users. An example is the analysis of whether and how players move through a game environment varies from the path/-s intended by the designers. In contemporary major game titles where there is a trend towards flexible, open-ended gameplay (e.g. *Grand Theft Auto IV*), this type of gameplay metrics analysis can be complex.

Kane & Lynch is a shooter-type game where the player takes control of one of a pair of criminals, playing through a series of missions in order to secure the freedom of his family (Figure 4). The game is brutal and the language harsh, depicting the more hard-core element of the criminal/terrorist world with abandon. In terms of gameplay, the game follows previous shooters in that the player controls a single character and mainly has to worry about staying alive, eliminate enemies and solve specific tasks. In certain missions, the player commands a squad of characters, with however the squad operating semi-independently under AI-control.

The focus of this case study is to show how gameplay metrics can be applied to examine player navigation through level maps, and whether the players deviate from the intended path of the level design. Furthermore, if there are any "trouble spots" where the gameplay might be too challenging (i.e. players die/get injured) – i.e. moving beyond the navigation mapping of Chittaro et al. [3,4] to include additional variables on the player state. These are questions relevant to all games and virtual worlds where navigation forms a part of the gameplay, and the player controls a single character/avatar (or small group).

The type of analysis required to address the questions of the designers requires and event-set approach to data collection, where gameplay metrics are recorded in sets associated with specific events. For example, looking not only at where players die, but also where they crouch and cover, the speed of their movement at different intervals through game levels, where they get injured and how much; and so forth. In order to analyze whether players experience the content of the map is a potentially complex analysis, depending on the specific features that are sought investigated. Initially however, content exposure is a matter of **placement** within the virtual universe – the player has to be there in order to experience the content.

Figure 4: Screenshot from *Kane & Lynch*, showing the two main characters. The player controls Kane, the character in the front (image © IO Interactive, 2008, used with permission).

For this case study, a section of one of the many level maps from *Kane & Lynch* was selected. The level features a series of interconnected fights and different locales through which the player (or two players in co-op mode) must navigate and survive. The locales include both near-ranged combat in enclosed spaces, as well as more open environments. The total playtime for an experienced player is about 12-14 minutes.

The first problem addressed was whether the players deviated from the intended path of the designers. This was approached by defining a "perfect path" – the one that players were expected to follow – and calculate the difference between the player paths and the perfect path. The perfect path was defined with the lead designer of the level and a 2 meter wide path defined on top of it using ArcGIS (Figure 5).

Following this, players were asked to play through the level segment, and the following gameplay metrics recorded:

1) **Navigation**: The path of the players recorded as (X,Y,Z)-coordinates every second. Unlike Chittaro et al. [3,4], it was chosen to use timed coordinates rather than vectors because point data are easier to label and take up less space in the data stream.
2) **Health**: The (X,Y,Z)-coordinates for the locations where the player is injured, and the degree of injuring. Health is in *Kane & Lynch* divided into five degrees, depending on the level of injury the player has sustained. These data are tracked as a feature of player location, and can be mapped as a secondary variable together with the location point data. These types of plots are useful to investigate where players are injured and the causes.
3) **Crouch and cover:** The (X,Y,Z)-coordinates for the locations where the player crouches or uses the snap-to-cover system that is a hallmark of *Kane & Lynch*.
4) **Speed:** The movement modifier of the players, i.e. whether the player was standing still, walking, running etc.

By plotting the navigation coordinates on top of the perfect path, it is in ArcGIS possible to calculate the extent to which players follow this path, i.e. behave as intended by the design (Figure 5). In the example provided here, roughly 85% of the player´s path falls within the 2-meter perfect path, i.e. the player is more or less behaving as intended with respect to navigation. The same procedure can be run with data from thousands of players at the same time plotted on top of the perfect path. If such datasets are available, it becomes especially valuable to examine if there are sections where numerous points or clusters of navigation points fall outside the perfect path, as it indicates a deviation by a substantial number of players from the intended path/-s.

In order to address the second question, if there are areas where the player is heavily challenged, the health of the players was added as a color code on top of the navigation point data. The analysis showed that for this segment of the game level, the player went mostly uninjured or lightly injured. However, in two positions the player´s health drops rapidly, and often players would die in these locations. These are locations where considerable opposition is supplied via AI-controlled opponents (Figure 6). This analysis assists in clarifying whether the level is too hard to play (players often dying/getting injured heavily), or too easy. Conversely, if there is a good balance in the level between easy and hard segments (a typical design goal for shooter-type games). The use of crouching and the cover system was also examined, as was the speed with which the player moved through the level. During most of the investigated segment, movement was rapid, which corresponds with the player not being substantially challenged. However, in the two "trouble spots" the player would slow down to a standstill and take cover, navigating between different cover options in order to eliminate the opposition. The analysis showed that for the first "trouble spot", the players generally take little cover; however, cover is also more scarcely available.

Figure 5: The path of the tester has been overlain the 2-meter wide perfect path. As would be expected from the relatively linear design of *Kane & Lynch,* the paths almost overlap (85%). However, in a few places the paths diverge (red circles).

Figure 6: Visualizing the health of a single player. The points are spaced one second of playtime apart. Color signifies the health of the players (five categories). Locations of lowest possible health (right before death) have been marked with orange crosses.

This kind of gameplay metrics analysis is useful in recreating the player experience in detail, and evaluating if there are areas of the level design that are problematic – e.g. lack of cover options, encounters that are too challenging or not, and so forth. The kind of feedback provided via gameplay metrics is very precise and detailed - this down to a level where the position of the player can be mapped to the specific corner he or she was covering behind, for how many seconds the player spent at the location and what the player was looking at, etc. Using gameplay metrics in this fashion – ideally in concert with attitudinal data - it is possible to analyze in detail the behavior of players and normally also the causes, for example by considering the spawning points of enemies in concert with player health data and movement patterns. Using ArcGIS, it is possible to plot different variables of player behavior in successions of layers, perform calculations across them, and to dynamically add/remove the individual layers.

4. CONCLUSIONS AND PERSPECTIVES

Gameplay metrics form a novel approach within game development and addresses one of the major challenges to games-oriented user research, namely that of tracking and analyzing user behavior when interacting with the very complex systems that contemporary computer games are. As a user-oriented approach, it complements existing methods utilized in the industry very well, providing detailed and quantitative data to supplement attitudinal and semi-quantitative data from e.g. usability testing and playtesting [19,24,27,30]. In summary, the main benefits of gameplay metrics are as follows:

- **Quantitative** and **highly detailed data** on player behavior
- **Objective** way of visualizing and analyzing play-session data
- **Detailed feedback** on game design and mechanics
- **Supplements existing methods** for user experience testing and bug-tracking (data for both purposes can be collected simultaneously)
- **Assists the location of game problems:** E.g. bugs and faulty patterns of play, and helps with evaluating fixes
- **Progressive detail:** Gameplay metrics permit analysis from a top-down approach working towards progressively more detailed layers of analysis.

Instrumentation data from users form an important contribution to not only user research and –testing during the development phases of game production, but also in monitoring and evaluating user (player) behavior during the extended usage, i.e. during the live periods of games, where given the right tools, data can be obtained directly from the users operating within their natural environments – at home, in internet cafés, LAN-parties etc. This is particularly useful for academic purposes, e.g. where the aim is to examine how people play games in their own environments [e.g. 36]. As shown by e.g. 14, it is possible within an academic context to develop metrics tracking systems. However, in order to take advantage of gameplay metrics, methods need to be developed that can be used to decide which data to track and how to analyze them. Furthermore, because the use of game metrics for gameplay analysis is a relatively recent innovation, it remains uncertain what the limits are for their application and which methods that are cross-game applicable. The use of gameplay metrics is in itself novel, being hitherto applied in about a dozen different publications, despite the widespread use of instrumentation data within related fields such as application software production and website design. Existing approaches towards gameplay metrics analysis generally focus on single variables and rarely focus on the spatial dimension of the game worlds that players navigate. In this paper, two case studies have been presented which showcases the usefulness of working with multiple variables and within spatial environments, providing an approach for analyzing gameplay and recreating the playing experience. The case studies are based on common features of shooter-type games, navigation and death, and are therefore cross-applicable across games of this genre. Furthermore, the approaches presented are relevant for other kinds of games and VEs where the player controls a single character/avatar. The case studies indicate the usefulness of gameplay metrics in being able to recreate the play experience, as a method for evaluating game design, e.g. figuring out where the challenge level is too high or low, where the player has problems navigating, and if the player does what is intended from in the design, or goes outside it. In games where multiple means of completing the game are possible, using different strategies, such as *Deus Ex*, the *Tomb Raider*-series and role-playing games such as *Neverwinter Nights*, this latter potential is especially promising.

ACKNOWLEDGEMENTS

The current work would not have been possible without the stellar work of the *EIDOS Online Development Team*, who created the metrics server and –system, and who maintain it to the joy of the authors and other data analysts. Also sincere thanks to the many other colleagues at *IO Interactive, EIDOS*, the *IT University of Copenhagen* and the *Danish Design School*, for assistance, support and interest.

REFERENCES

1. Bateman, C., Boon, R.: 21st Century Game Design. Charles River Media (2005)
2. Blythe, M. A., Overbeeke, K., Monk, A. F. Wright, P.C.: Funology – from usability to enjoyment. Springer (2004)
3. Börner, K., Penumarthy, S.: Social Diffusion Patterns in Three-Dimensional Virtual Worlds. Information Visualization, 2 (2003), 182-198
4. Chittaro, L. & Ieronutti, L.: A visual tool for tracing users´ behavior in virtual environments. In Proc. Working Conference on Advanced Visual Interfaces (2004), 40-47
5. Chittaro L., Ranon R., Ieronutti L.: VU-Flow: A Visualization Tool for Analyzing Navigation in Virtual Environments. IEEE Transactions on Visualization and Computer Graphics, 12,6 (2006), 1475-1485
6. Davis, J.; Steury, K., Pagulayan, R.: A survey method for assessing perceptions of a game: The consumer playtest in game design. Game Studies: The International Journal of Computer Game Research, 5 (2005)
7. De Kort, Y., Ijsselssteijn, W: People, places and play: Player experience in a socio-spatial context. ACM Computers in Entertainment 6, 2, article no. 18 (2007)
8. DeRosa, P.: Tracking Player Feedback to Improve Game Design. Gamasutra, Aug. 7 (2007)
9. Desurvire, H., Caplan, M., Toth, J. A.: Using heuristics to evaluate the playability of games. In CHI '04 extended abstracts on Human factors in computing systems (2004)
10. Dumas, J. S. 2003. User-Based Evaluations. In: The HCI handbook. Lawrence Erlbaum Associates (2003), 1093-1117
11. Gilleade, K. M., Dix, A.: Using frustration in the design of adaptive videogames. In: Proceedings of the ACM SIGCHI International Conference on Advances in Computer Entertainment Technology, Singapore (2004), 228-232
12. Goetz, P.: Too Many Clicks! Unit-Based Interfaces Considered Harmful. Gamasutra, August 23 (2006)
13. Hilbert, D. M. and Redish, J. C.: A practical guide to usability testing. Intellect books (1999)
14. Hoobler, N., Humphreys, G., Agrawala, M.: Visualizing Competitive Behaviors in Multi-User Virtual Environments. In Proc. of the IEEE Visualization Conference (2004), 163-170
15. Hurst, A., Hudson, S. E., Mankoff, J.: Dynamic detection of novice vs. skilled use without a task model. In Proceedings of CHI 2007, ACM press (2007), 271-280
16. Ijsselsteijn, W., de Kort, Y., Poels, K., Jugelionis, A., Bellotti, F.: Characterizing and Measuring User Experiences in Digital Games. In: Proceedings of ACE, Toronto, Ontario (2007)
17. Isbister, K., Schaffer, N.: Game Usability: Advancing the Player Experience. Morgan Kaufman (2008)
18. Jørgensen, A. H.: Marrying HCI/Usability and Computer Games: A Preliminary Look, Proceedings of NordiCHI, Tampere, Finland (2004)
19. Kim, J. H., Gunn, D. V., Schuh, E., Phillips, B. C., Pagulayan, R. J., Wixon, D.: Tracking Real-Time User Experience (TRUE): A comprehensive instrumentation solution for complex systems. In Proceedings of CHI 2008, Florence, Italy (2008), 443-451
20. Kort, J., Steen, M. G. D., de Poot, H., ter Hofte, H., Mulder, I.: Studying usage of complex applications. In: Proceedings of Measuring Behavior. Noldus IT (2005), 266-269
21. Kuniavsky, M.: Observing the User Experience: A Practitioner´s Guide to User Research. Morgan Kaufman (2003).
22. Longley, P.; Goodchild, M. F., Macquire, D. & Rhind, D.: Geographic Information Systems and Science. Wiley & sons (2005).
23. Mandryk, R., Atkins, M.S., Inkpen, K.: A continuous and objective evaluation of emotional experience with interactive play environments. In: Proceedings of CHI (2006.
24. Medlock, M.C., Wixon, D., Terrano, M., Romero, R.L., Fulton, B.: Using the RITE method to improve products: A definition and a case study. In: Proceedings of UPA (2002)
25. Mellon, L.: Metrics in MMP Development and Operations. Presentation at the GDC, Los Angeles, USA (2004)
26. Nørgaard, M., Rau, J.: User Testing in the Combat Zone. In: Proceedings of ACE, Toronto, Ontario (2007)
27. Pagulayan, R. J., Keeker, K., Wixon, D., Romero, R. L., Fuller, T.: User-centered design in games. In: The HCI handbook. Lawrence Erlbaum Associates (2003), 883-906
28. Pagulayan, R. J., Steury, K. R., Fulton, B., Romero, R. L.: Designing for fun: user-testing case studies. In: Blythe, M. A., Overbeeke, K., Monk, A. F. Wright, P. C. (Eds.): Funology: From usability to enjoyment. Human-Computer Interaction series. Kluwer Academic Publishers (2005).
29. Renaud, K. and Gray, P.: Making sense of low-level usage data to understand user activities. In: Proceedings of SAICSIT 2004, ACM Press (2004), 115-124.
30. Rouse, R. Game Design: Theory and Practice. Wordware publishing Inc. (2001).
31. Schultz, C. P.: Game Testing All in One. PTR (2005)
32. Swain, C. 2008. Master Metrics: The Science Behind the Art of Game Design. Presentation at NLGD Conference, Utrecht, Holland (2008)
33. Thompson, C.: Halo 3: How Microsoft Labs Invented a New Science of Play. Wired Magazine 15, 9 (2007)
34. Trappl, R., Petta, P.: Creating Personalities for Synthetic Actors: Towards Autonomous Personality Agents. Springer Publishers, Berlin (1997)
35. Tychsen, A., Canossa, A.: Defining Personas in Games Using Metrics. In: proceedings of Future Play 2008, Toronto, Canada. ACM publishers (2008), 73-80
36. Williams, D.; Yee, N., Caplan, S. E.: Who plays, how much, and why? Debunking the stereotypical gamer profile. Journal of Computer-Mediated Communication 13 (2005), 993-1018