HAL
open science

Feedback-based gameplay metrics

Raphael Marczak, Gareth Schott, Pierre Hanna, Jean-Luc Rouas

To cite this version:

Raphael Marczak, Gareth Schott, Pierre Hanna, Jean-Luc Rouas. Feedback-based gameplay metrics. FDG Foundations of Digital Games, May 2013, Greece. pp.71-78. hal-00868293

HAL Id: hal-00868293

https://hal.archives-ouvertes.fr/hal-00868293

Submitted on 1 Oct 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Feedback-Based Gameplay Metrics

Raphaël Marczak
School of Arts
University of Waikato
New Zealand
raphaelm@waikato.ac.nz

Gareth Schott
School of Arts
University of Waikato
New Zealand
g.schott@waikato.ac.nz

Pierre Hanna
LaBRI
University of Bordeaux
France
pierre.hanna@labri.fr

Jean-Luc Rouas
LaBRI
University of Bordeaux
France
jean-luc.rouas@labri.fr

ABSTRACT
The application of gameplay metrics to empirically express a player's engagement with the game system has become more appealing to a broader range of researchers beyond the computer sciences. Within game studies, the appropriation and use of gameplay metrics not only further shifts these methods beyond formalized user testing (e.g. with the aim of product improvement) but creates a demand for a more universal approach to game metric extraction that can be applied to a wider range of games and account for different social science research agendas. Set in the context of a study into player experiences with participants aged 14-16 years of age, a method of articulating and identifying key moments of game-play pertinent to an analysis of experience with simulated violence was required [38]. This paper describes a pragmatic solution to the need to study a range of games that were determined not by the researchers but the study's participants. The solution exploits and examines the audio-video feedbacks presented to the player as part of their engagement with a game system. This paper broadly outlines this mode of metric extraction and its application to research that seeks to understand the nature of engagement and player motivations across a mixed methodology approach.

Categories and Subject Descriptors
K.8.0 [**Personal Computing**]: General – games
H.1.2 [**Models and Principles**]: User/Machine Systems
H.5.5 [**Sound and Music Computing**]: Methodologies and techniques
I.4.0 [**Image Processing and Computer Vision**]: General – *Image processing software*

General Terms
Algorithms, Measurement, Experimentation

Keywords
Audio-Visual Feedback, Gameplay Metrics, Game Studies, Player Experience.

1. INTRODUCTION
Whilst exhibiting clear associations with the aesthetic form and narrative structures of traditional games, fiction and film, we understand digital games differ by the way they require players to perform recursive actions [25] that lead to polysemic performances and readings [11]. Playing a game is predominantly a configurative practice, not an interpretative one like film or literature, requiring the player to "work with the materiality of a text, the need to participate in the construction of its material structure" [24]. Game research has consequently focused much attention on seeking to develop a better understanding of the different experiential components of gameplay, for example, through the development of concepts of enjoyment [43] flow, [23] or immersion [33]. Yet, players are often characterized within game studies literature either as 'implied' player [2] or 'active player' [17]. The former, simply acknowledging the predisposed need to have a player act on games in order for a game to function as intended and "exercise its effect" [1], while the latter refers to the "unexpected, aberrant, oppositional or directly subversive" [17] moments of play that researchers find interesting. Both are possibly one step removed from the actual the relationship between the "structure of a game and the way people engage with that system" [44].

Given that gameplay metric analysis originates from within the field of human-computer interaction (HCI), it has to-date mostly been applied with the intent of supporting game design [12, 13, 22, 30]. Yet, there is a growing interest in applying metrics to seek a better understanding of playing styles [42] and providing valuable empirical data to test conceptual and theoretical foundations of ludology (e.g., Bartle's early work on play styles in MUD's [4] or Juul's work on the meanings of failing in digital games [21]). Indeed, metrics possess the potential to enable game studies researchers to map and examine how players actually activate games of interest thereby contributing to a better understanding of the content encountered by players. Furthermore, in combining the gameplay metrics with other measures, such as biometrics derived from players as they play, there is potential to build an empirically derived picture of the impact that games exert during play [e.g. 29].

While the potential for integrating such quantitative measures to appraise and validate existing, and drive new theoretical work on player experiences is high, gameplay metrics do however possess certain limitations for humanities/social science oriented research. Metrics on their own do not account for reasons or motivations

behind the player's behavior [7] nor are there any companionable psycho-physiological measures, for instance, that differentiate the intentionality of the measured outputs. But these are limitations imposed by an exclusive use of these methods that can be overcome by their assimilation and triangulating with other qualitative and quantitative techniques [8]. More significant for some game researchers is the fact that the production of metrics is typically conditional on exploiting technology in ways that reduce the range of games that can then be studied, for example, in cases where generating metrics requires access to a game's source code. In this context it is possible to discern why gameplay metrics have so far mostly been used for playability research [12, 13, 22, 30] rather than research pursuing player experiences with games post-release [31].

To counteract the limitations associated with particular methods of metric production, our research project required that we study player experiences with participants engaged with a range of games that vary in genre, style of play, the presentation of data denoting player achievement/status (found in Head Up Displays) and accessibility (in terms of methods of metric collection). In order to combat the limitations (outlined above) and constraints on which games could be studied we sought to develop an alternative method for acquiring gameplay metrics through exploiting the audio-visual feedback experienced by players as they play (or screen and speakers/headphones). That is, information directly processed (or made available for processing) by the player as they play and therefore part of the conscious experience of the act of playing. The first aim, was to develop an automated data collection procedure that would permit a cogent representation, synthesizing hours of game-play using the flow of the game (determined by variables such as screen death, repetition, cut-scenes, health and stamina levels) that possess the capacity to be correlated with biometric data (e.g. heart rate, GSR), player commentaries and observation of both player (eye tracking, facial gestures) and text (textual and structural analysis). Secondly, provide a method that opens up the possibilities of collecting gameplay metrics in the study of any game that is commercially available without requiring access to the source code or a strong collaboration with the game company responsible for the game.

2. PLAYER EXPERIENCE AS OUTPUT STREAMS

In the course of playing digital games we are presented with two main output streams: the audio stream (sound from speakers or headphones) and the video stream (moving image on screen). These streams contain all the key communication components necessary for the player to interact efficiently and with purpose in the game-world. Both video and audio streams are capable of producing two levels of analysis: at the diegetic level, which portrays information about the game's story world and the interface layer containing non-diegetic information about the game system [14, 16]. At the interface level, the player is informed both about abstract game elements (e.g., number of lives, score etc.) and meaningful game design parameters players do not directly experience (as they are not physically present in the game world) but require feedback on in order to adjudge anything from immediacy or urgency of their actions, strategic success to sense of presence [35]. Further rationale for considering audio and video streams as a data source is that it presents a new layer of analysis. In addition to the game level containing diegetic information and the interface level containing non-diegetic information, audio and video streams also communicate more general information such as the overall color of a game sequence, which has the potential to be linked with the degree of emotional impact on the player [19]. Additionally, video and sound quality [6, 40] and different sound parameters [32] can be related to degree of player involvement.

Audio and video signal processing of multimedia content constitutes an established and active research field within computer sciences. Numerous algorithms have been created to detect the presence of specific objects in a video [5], automatically segment a film into sub-scenes [18, 37] and extract symbolic information such as annotations [9] or jersey numbers in sport videos [45]. Signal processing of audio streams also constitutes an established and active research field within computer science. Algorithms have been created to detect speech and music in radiophonic streams [36], to retrieve artist and song title information from music recognition [3, 34] or to indicate moments of 'story intensity' in movies through audio tempo analysis [46]. Yet, algorithms of this nature have yet to be applied to digital interactive games.

In processing digital games, we have sought to exploit similar algorithms as they present large amounts of information in an abstract and schematic way (e.g., via life-bar representation, icon of carried weapon, on-screen blur, screen desaturation, heartbeat sound). Different visual or representational design elements can be extracted from output streams and analyzed to create a picture of the nature and developmental changes in players' interaction and experience with a particular game. Although game audio still remains a hugely under-researched subject, it is understood to play a vital role in gameplay experience. Sound serves to orient players actions in a virtual space thus increasing the 'immersiveness of the gameplay experience' [16] by allowing players to identify certain game functions [20], triggering emotions [28] and offering an essential 'preparatory function' [10].

So far, however, audio and visual streams have yet to be applied in an extensive analysis of player interaction with digital games. The reason for this can be attributed directly to its underdeveloped nature combined with the complexity of extracting data from audio or video streams when compared to established methods for extracting metrics directly logged by the game engine itself. In taking this path, we are not arguing that audio-visual output streams constitute a superior method but serves to prevent those seeking to understand the cultural and social impact of games from being constrained in their game choices by the need for developer cooperation (e.g., licensing) or technically determined accessibility constraints. By extracting game metrics from audio and video output streams the aim is to open the doors to researchers to be able to profit from the advantages of quantitative data collection with any game of their choice.

Key to any automated collection method however is the manner in which it can considerably speed up the coding and analysis of gameplay footage, allowing large amounts of data to be coded across numerous parameters (see Section 3). In the context of our study [38], the audio-visual output streams achieved from +40 hours of game play per participant (with each game studied) are currently employed in a two-stage process. Firstly, graphic representation of the events experienced and the particular course taken by the player are examined alongside other data sets (e.g. GSR and heart rate) in order to discern possible relevance to the player. The regression of variables ranging from stamina/health levels, menu activation or the expressiveness of music together with heart rate/GSR suggest which moments of game play

activation offer challenge, excitement, frustration and/or reward for the player. The second stage entails corroboration of the metrics, by textual and structural analysis and player commentary (in which players recall and narrate extracts of their play) of the sections of game play selected.

In order to test the feasibility of this approach, a software system was developed for processing both audio and video feedback streams that were predicted to be most useful and representative of the gameplay experience (a process that is still ongoing). In the following sections, three video filters based on simple image processing features: color detection, image similarity and global hue detection are presented. We outline how even simple image characteristics can be exploited to produce data that is accurate, meaningful and easy to interpret. Attention is then turned to the demands and techniques required for audio processing.

3. AUTOMATIC VIDEO FEEDBACK STREAM ANALYSIS

Beginning with a 'color ratio' filter, we found it possible to detect the variability of any information in the HUD presented in the form of a bar, for example, commonly applied to express health (reduction most indicative of combat), stamina (activity levels), weapon deterioration (player strategy) but also applied in loading icons during further initialization of the game (indicative of player progress). In executing this filter, it was also necessary for it to be able to detect specific color schemes, as the presence of which indicates levels within a bar (high or low). Reflecting on health bar variability, we are able to present changes in the player's health state through the game. Analytically this supports the identification of hazardous moments that can be distinguished from moments of relative safety for the player. In these phases of play the player's activity can veer between combat and exploration or performativity and knowledge gathering (e.g. narrative back-story) to cite just two examples.

The game *Dead Island* (Deep Silver, 2011) was used to first trial the method (with a sample of participants over 18 years of age) before commencing the study. In this game the health bar would disappear from screen, indicative of a conversation taking place between characters, a cut-scene had been triggered, or the in-game menu activated. These gaps were not misinterpreted in our analysis as 'deaths' since death is preceded by a progressive loss of health that does not occur when the life bar totally disappears from the screen. Switching to 'stamina bar' progression, the manner in which the variability is expressed similarly indicates different activities, for example, running was found to be represented by a regular decrease in stamina levels, while combat is expressed by a more sudden and chaotic drop.

When combined with other measures, the health bar is also indicative of the flow and pace of the game, enabling the identification of mastery (e.g. stamina reduction indicating combat matched with stable health levels) screen death and repetition, object discovery (e.g. health packs), or key dramatic moments in linear gaming (e.g. boss fights & climatic sections of game play that increase the demands on the player).

Secondly, an 'image similarity' filter was employed in order to enable the detection of icons, for example, an activated flashlight, weapon hold, use of first aid kit etc., or any static HUD information in an area of the video. Again it was possible to apply this information source in conjunction with information gathered from another video filter, for example 'health bar' levels to determine the nature of play. Drawing on *Dead Island* again, we were able to determine that the first aid icon appeared each time the player used a curing item (e.g. energy drink). This information enabled us to highlight and distinguish moments when the player was exploring the environment as opposed to being engaged in intense action (fighting) sequences, gathering resources and replenishing health as a result of moments of discovery.

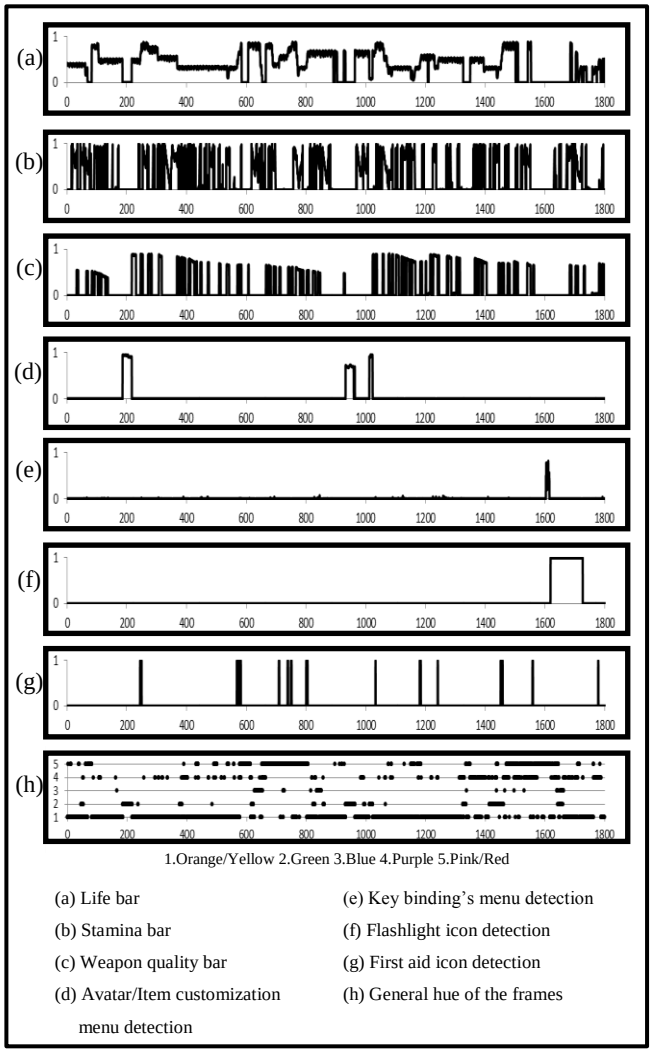

Figure 1. Example of Video Stream Analysis Produced for a Single Participant

Finally, a 'global hue filter' was employed to analyze the graphical dimensions and design of a game and its relationship with player response and also identify specific sequences of the game that possess their own color code, for instance switching to the in-game menu. Unlike the other two filters that analyze the interface layer, this filter is linked to the video parameters layer. It computes the hue histogram of each frame, and then searches for the most present hue regarding 5 classes: orange/yellow, green, blue, purple, pink/red. The general hue of the frames helped us to quickly distill a wide range of information from the gameplay footage. With *Dead Island*, for instance, the main hue of orange/yellow and blue helped us to establish the types of environments and content the player was encountering in its tropical island setting (e.g. beach, wooden houses, sea, swimming pools); with the pink/red hue supporting the identification of moments where violence or otherwise 'disturbing' imagery is presented.

This section has sought to demonstrate the kind of data that has

been possible to generate automatically requiring no manual coding (an example of the video stream analysis is presented in Figure 1). The only necessary user input consists of specifying in advance what screen area to analyze, norm threshold to consider and colors, color error tolerance and icons to detect. After these considerations these inputs can be used without modification with any footage captured using the same game. This becomes particularly useful when faced with many of hours of footage that requires analyzing and coding, particularly in larger studies or pseudo-longitudinal studied that seek to examine the lifecycle of a particular game from beginning to end/completion.

4. AUTOMATIC AUDIO FEEDBACK STREAM ANALYSIS

In the following sections we outline the application of two algorithms that were employed to acquire gameplay metrics from the game's audio feedback stream. Using examples derived from the implementation of the algorithms during the analysis process for *Bioshock 2* (2K Games, 2010), we are able to demonstrate the role of the audio feedback stream analysis as both a search mechanism for pulling out particular sections of game play across a sample of participants, and also a mechanism for accurately charting individual progression and the pacing of their encounters with key structural elements of any game.

4.1 Audio Detection Methods

The exploitation of musical expression and the atmospheric soundscape that expresses space and meaning in games was tested as a means of segmenting gameplay sessions into the differing environments encountered by the player or key moments necessary for progression or comprehension. Music employs dynamic indications, phrasing, differing qualities of touch and articulation, color, intensity, energy and excitement to differing effects in order to communicate to audiences. In order to utilize this communication device and for a detection process to function automatically, our processing system was required to possess prior knowledge of the different musical phrases of the games under investigation. Thus, the system employed had to first be given 20 second sound extracts taken from the game's sound output in order for it to be able to classify and identify particular sections or musical phrases employed throughout the game.

The principle of the detection relies on an estimation of similarity between an extract of music taken from the game, which is then located and identified within the context of other game play footage achieved across a sample of participants. Here the automated process is serving to find the same instance across the wider data set. A significant moment of game-play signified by its dramatic or tension building music was arrived at through the automated video processing method (comprised of varying combinations of variables) described above, which were then detectable for comparison across the sample if also accompanied by a particular musical signature. Thus, for each comparison, an identification process had to be computed that indicates which part of the musical piece has been recognized.

Estimating the similarity between two musical pieces is a complex task [15]. Yet, this task becomes even more challenging when the musical pieces being sought, differ in their presence due to the varying audio levels overlaying the music (representative of the player's presence and action in the game world). The intensity and volume of diegetic sounds (gunshots, screams, speech, etc.) is typically higher than that of the musical backdrop to any scene therefore requiring the audio recognition system developed and applied to be extremely robust. In this study we opted to identify music extracts based on their tonal properties as it is possessed the potential to function in a robust manner irrespective of the presence of additional sound [3, 39]. In this approach musical pieces are encoded as sequences of symbols corresponding to the distribution of the energy in the amplitude spectra. The comparison of these symbols is computed by aligning the sequences with the use of local alignment algorithms [41]. Local alignment algorithms compute a score similarity. The higher the score computed then the greater the similarity between the segments.

Game segmentation using speech detection is based on the Hidden Markov Models (HMM). Four HMM states were used for this task, each state being modeled by a Gaussian mixture with 256 components. The features used are classical speech processing features, Perceptual Linear Prediction coefficients (PLP). Twelve coefficients together with their first and second-order derivatives were utilized.

5. IMPLICATIONS AND INFERENCES DRAWN FROM AUDIO DETECTION

During the first 45 minutes of *Bioshock 2*, players play through several memorable moments that are important for player comprehension of the game world and narrative context connected to the aims of the game. These moments were identified as:

- The player acquires his first super human power during a non-interactive cut-scene where he painfully injects a plasmid (a specific power) in his arm.
- The player encounters the first boss fight (Big Sister), introduced by a dark non-interactive cut-scene. The player is unable to finish the battle, as the Big Sister eventually flees from the scene.
- The player is sent underwater after a second confrontation with the Big Sister ends with a window breaking and consequently a flooding of the room.
- The player is presented an overview of the abandoned underwater city of Rapture.

All these sequences are recognizable through their accompanying sound scheme: In the first condition, the player's character screams in agony during the injection; Second, the fight with the Big Sister is preceded by suspense building music; Third, the breaking window makes a loud jarring cracking noise; and fourthly, the overview of the city is accompanied by majestic music that rises to coincide with the big reveal of the cityscape and to denote the magnitude of Rapture and the task ahead of the player.

In successfully detecting these four key moments and their temporal location in terms of when it occurs in the player's gameplay timeline (i.e. progression) and for how long (duration), the metrics also convey a sense of pace and level of engagement as players move through the game. In this way the detection process also enabled evaluation and insights into player behavior with *Bioshock2*. For example, it was possible to examine the average duration of key moments for our sample. In doing so, it revealed which players took more time to work their way through sequences. Further analysis of their footage revealed that these players performed a greater desire to explore the game world for resources and information, constantly scanning and searching their environment, attuning to minor details and the general mis-en-scene of the game. A good example of this occurred in the last of the key moments identified above, in which there is a point after the player is submerged underwater, just prior to the scene in which the player is treated to the vista of Rapture. In this moment

there is detail that is neither vital nor necessary, but is nevertheless made available to the player to observe if they opt to. Should the player gaze back inside the city through the windows, the player can observe another Big Daddy (for they themselves play as one) engaged in combat in the protection of a Little Sister. Because the underwater sequence is extremely linear and no other action other than movement from point a to b is required from the player in this game sequence, we can identify the players that either missed or ignored this sequence and instead proceeded directly to the overview of the city as those who completed this sequence between 75s and 90s. With other players who experienced an extended sequence of the game between 119s and 147s more likely to have stopped to watch the fight unfold.

Figure 2 (below) demonstrates how nine of our participant cohort traversed the four key scenes identified above. In the case of the Big Sister fight (key moment #2), the antagonist in this first boss fight flees the scene. Post-fight, most of our participants were observed leisurely exploring the environment for resources, operating under the assumption that the fight was over. However, Participant 1 (Figure 2) went off in pursuit of the Big Sister, as demonstrated by their arrival at the second confrontation (key moment #3) only 70s later (while the others participants typical took up to 2 minutes).

In contrast, Figure 2 also revealed that Participant 3 expanded time, of the very linear game-play of *Bioshock 2* as it took 10 minutes to reach the second confrontation. Indicating that Participant 3 either possess a strong desire to explore the game world in detail or has a tendency to get lost or stuck in several places as it took this participant almost 1400s to encounter and experience the four key events. This particular kind of information is useful in automatically segmenting the data from game sessions, when seeking to conduct correlation analysis between these moments or sequences and other feedback-based gameplay metrics (e.g. speech detection) or psycho-physiological responses

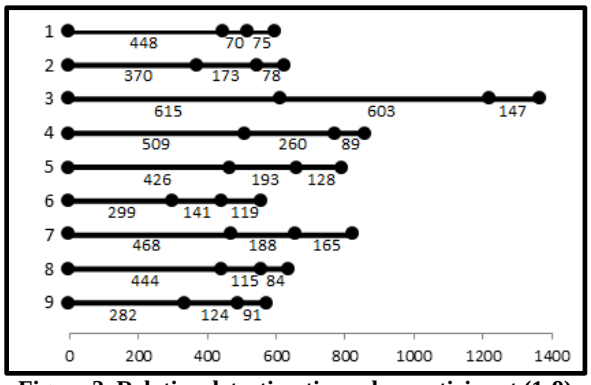

Figure 2. Relative detection times, by participant (1-9)

5.1 Speech Detection & Narrative Engagement

In *Bioshock 2*, large numbers of monologs play an important role in communicating the backstory of the game, providing players with the option of delving into the detail and context beneath the current game state or at the very least diegetic ambient sounds of the game world. Speech in *Bioshock 2* is experienced in a number of different forms, as propaganda messages broadcasted over the public address system by *Sophia Lamb* (leader of Rapture and ideologist), radio messages sent by (seemingly) friendly characters, and diary tapes containing short spoken messages by previous citizens of Rapture whose narrate their personal stories during the demise of the city.

Similar to the application of the game sequence identification system, the results of speech detection can be used to segment a player's gameplay session. In many cases the audio narratives, like cut-scenes, are triggered by the game (denoting progression) but are then subsequently optional as to whether the player attends to the messages, ignores or skips them. In the case of the public address system, or any broadcast communication within *Bioshock 2*, players may remain close to the output source (taking note of the information) or continue to move forward and risk noise interference or volume reduction (disregarding the information). Such moments of knowledge transfer from game to player thus denote where players are in terms of the structure of the game but also level of player engagement/interest in the game's backstory. Indeed, in the context of our project this automatic segmentation method has been employed in the selection of key-moments for post-game commentary session in which players' recall and understanding of the game is assessed.

Figure 3 presents activity again taken from the first 45 minutes of the *Bioshock 2*. The game begins with a cut-scene that orients the player and conveys a dramatic encounter between their character and Sofia Lamb, probable antagonist, but assured manipulator and key figure in the destiny of the player and the future of Rapture. For most participants in our sample, the cut scene is evident around 200s, yet, there are those for which the speech detection does not note the cut-scene, suggesting and later confirmed as participants skipping this cut-scene.

As highlighted in Figure 3, Participant 3 took 10 minutes to progress from the first Big Sister fight to a second encounter and key moment #3 in which the water penetrates the city. For this participant there is also no speech detection between 1650s and 1950s. Furthermore, the waveform illustrates silence. Examination of the game footage for this participant revealed that they had decided to enter the game's story menu to read up on information provided on entities and objects existent within Rapture. The menu does not contain any sound or music.
When counting the speech moments between key moment #1 (the injection scene) and key moment #4 (the overview of the city), we can see that Participant 1 only encountered four instances of speech while Participant 3 encountered thirteen. Participant 1 also arrived at key moment #4 within the same time frame as Participant 3 took to encounter the first key moment. It thus seems that counting the moments of speech can validate inferences about certain play styles; Participant 1 possessing a more action driven play style, trying to move quickly through the game not taking much note of diaries scattered around the environment, whereas the behavior of Participant 3 shows a much greater interest in the story elements of *Bioshock 2*.

In order to assess the validity of the above presented methods, we compared the automatically processed results with their corresponding hand coded versions. A full presentation of this process is not within the scope for this paper, but more details can be found in Marczak et al. [26].

6. CONTEXTUALIZING METRICS

Although the video and sound analysis presents information that holds value on its own, greater context is required to fully convey players' gameplay experience. The metrics outlined in this paper play a pivotal role in identifying key sequences of game-play for further analysis and investigation that exploit other data collection methods (player observation, game-play capture, bio-metric feedback or player commentaries).

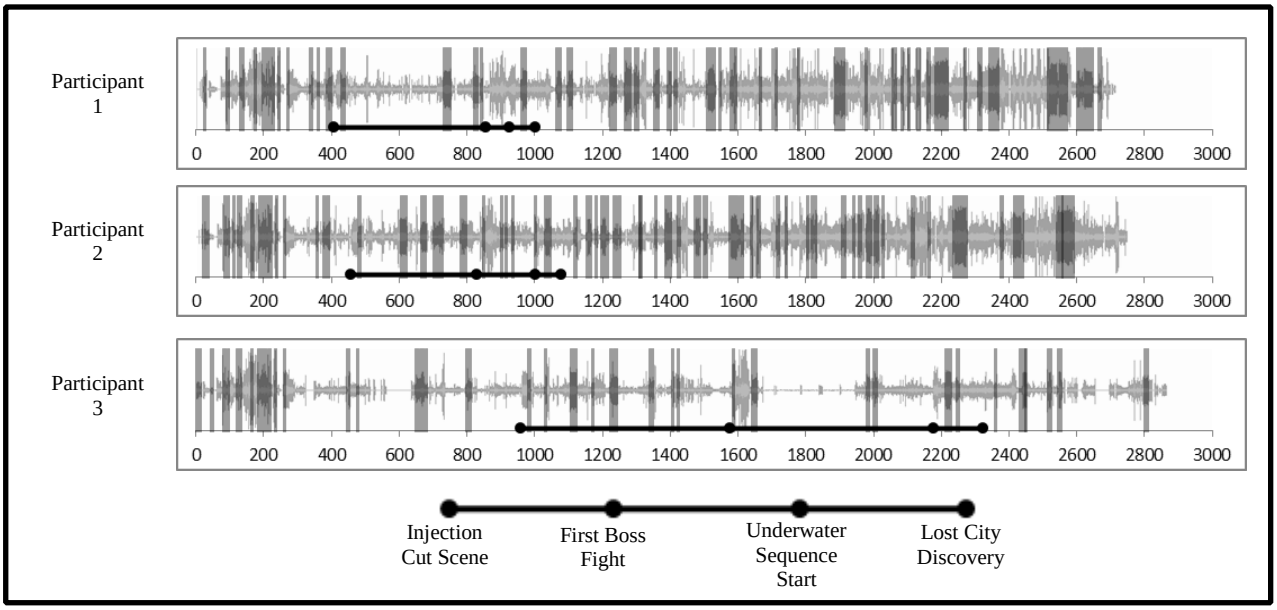

Figure 3. Sound Processing Results (Speech detection = Grey Highlighted Areas on Waveform)

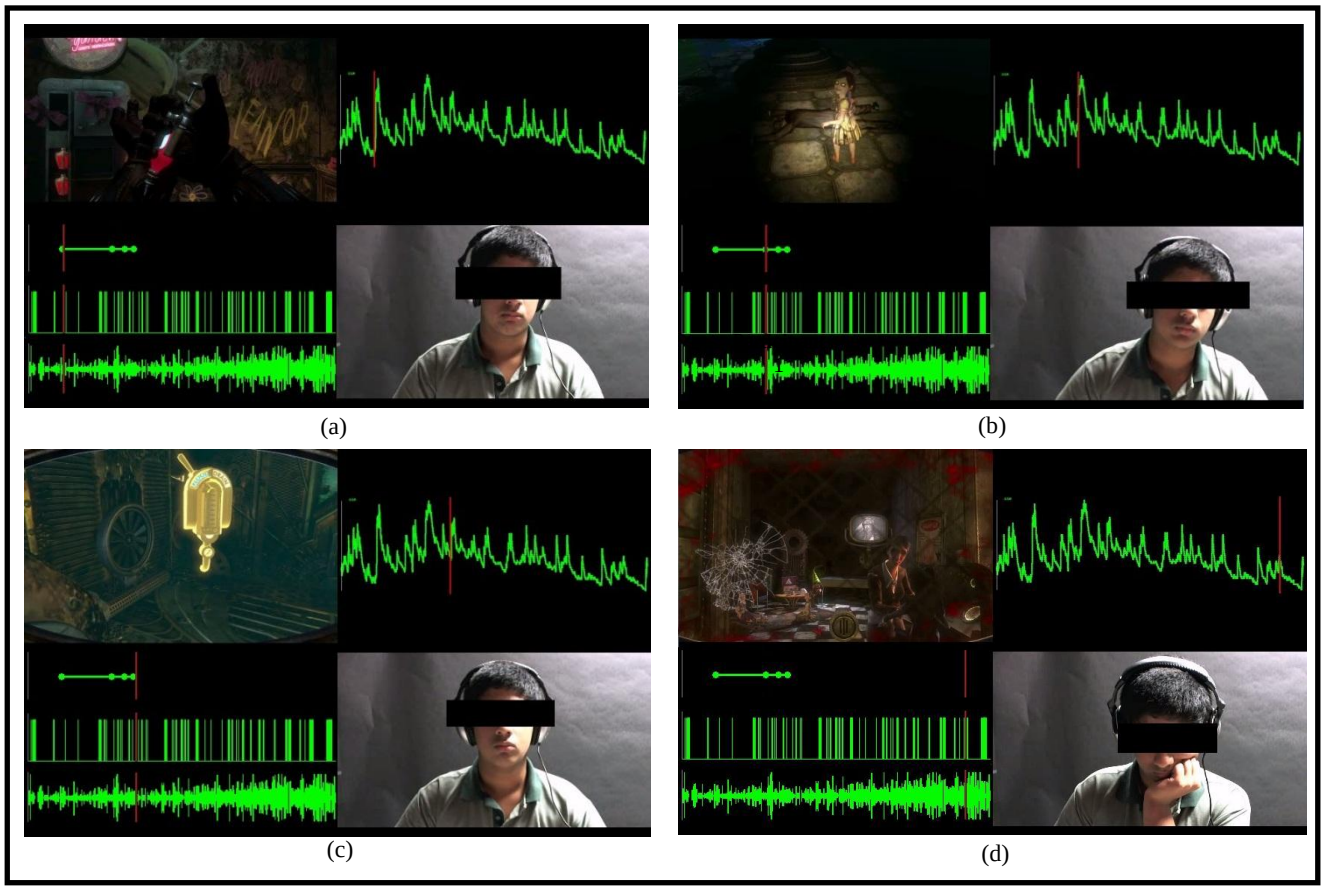

Figure 4. Synchronized video presentation (*Bioshock 2* screenshots ©2K Games)

In order to aid this process a software system for synchronized presentation has been instigated to generate an audio-video output that includes the gameplay footage, player's facial expression (captured with an HD-webcam), the player's galvanic skin response (GSR) (which indicates physical arousal) together with either video or sound analysis results (see Figure 4). During playback a moving cursor is presented on each curve to locate and connect the different data streams.

This approach can be compared to the "Biometric Storyboard" processing and presentation method developed by Mirza-Babaei [27] in which psycho-physiological data is combined with players' self-drawn diagrams relating to their play experience, and a manual coding of the game session. However, the synchronized presentation introduced here deconstructs the game play session alongside footage of the game session in a real-time audio-video presentation. Both the access and visualization of data aids analysis of the game sessions allowing the researcher to pinpoint potential relationships between different strands of the data that can be examined further both qualitatively and quantitatively.

Figure 4 illustrates several frames extracted from the synchronized video generated for Participant 8 (anonymized). For illustrative purposes and simplicity sake, the lower-left hand corners of the screens only contain a strand of the sound analysis (omitting video filter results). GSR measures are presented on the upper-right hand corner. Quadrants (a) and (b) display sequences that have been selected with the aid of the audio sequence detection system which demonstrate the injection cut scene (#1) and the first boss fight (#2). At the same moments, a peak in the GSR curve is detected which represents a high level of physical arousal. Further (qualitative) analysis may enable us to attribute this arousal to either the suspenseful nature of these sequences, or to the presentation of a non-interactive cut scene that removes player agency leaving him helpless to intervene in the situation.

Quadrant (c) illustrates the selection of footage based on a spike in the GSR curve. This time the GSR spike does not correspond to any notable results from the sound analysis. However, analysis of the gameplay footage appears to clarify the player's arousal levels. The footage shows the player's arrival in the decompression room, which indicates the successful end of a mission (this requires information from other automated data such as the 'color hue' video filter). Finally quadrant (d) displays a sequence selected for its speech detection results. By jumping to the speech segments, it is possible to enhance the interpretation of the acquired data by identifying what type of speech the player is encountering. In picture (d), speech comes from an encounter with a friendly character that gives advice to the player and further information about the game world.

6. CONCLUSION

This paper has aimed to show the potential value of acquiring gameplay metrics through an automated procedure utilising both video and sound outputs achieved from activating and engaging with games. We have sought to demonstrate an alternative approach to gameplay analysis which permits researchers to assimilate the data extracted from audio and video outputs that illustrate movement, flow, progress and focus of engagement that can be matched specifically against biometric and psycho-physiological measures or investigated further using interviews, textual and structural analysis. Visual and graphical representation presented in a synchronized form, creates a platform for annotating and coding the data with the advantage of having a first layer of analysis already added, capable of guiding researcher interests in specific sequences. Finally, a further consequence of this approach is the capacity for researchers to be able to profit from this brand of metrics in the analysis of player engagement with any game that has received a commercial release.

7. ACKNOWLEDGMENTS

The Royal Society of New Zealand, Marsden Grant for funding 'Videogame Classification: Assessing the experience of play' (10-UOW-024).

The University of Waikato, International Scholarship Grant, for funding 'Quantitative Data and Gameplay Segmentation Applied to the assessment of Players' Experience with Digital Games'.

Thank you to Jasper van Vught for his valuable help and support.

8. REFERENCES

[1] Aarseth, E. (1997) Cybertext. Perspectives on Ergodic Literature, London: The John Hopkins University Press, p.127.

[2] Aarseth, E. (2007) I Fought the Law: Transgressive Play and the Implied Player, Proceedings of Digital Game Research Association Conference: Situated Play.

[3] Bandera, C.D.L., Barbancho, A.M., Tardón, L.J., Sammartino, S., and Barbancho, I. (2011) Humming Method for Content-Based Music Information Retrieval, In Proc. ISMIR, pp. 49-54.

[4] Bartle, R. Hearts, (1996) Clubs, Diamonds, Spades: Players Who Suit MUDs. The Journal of Virtual Environments, Vol. 1, No. 1.

[5] Bay, H., Ess, A., Tuytelaars, T., Gool, L.V. (2008) SURF: Speeded Up Robust Features, Computer Vision and Image Understanding (CVIU), Vol. 110, No. 3, pp. 346--359,

[6] Bracken, C.C., Skalski, P. (2009) Telepresence and Video Games: The Impact of Image Quality. PsychNology Journal, 7(1), pp. 101–112.

[7] Canossa, A. (2009) Play-Persona: Modeling Player Behaviour in Computer Games. Unpublished PhD thesis.

[8] Canossa, A., Drachen, A., Sørensen., J.R.M. (2011) Arrrgghh!!! – Blending Quantitative and Qualitative Methods to Detect Player Frustration, Proc. of FDG 2011, ACM.

[9] Chen, D., Shearer, K., Bourlard, H. (2001) Video OCR for Sport Video Annotation and Retrieval, Proceedings of the 8th IEEE International Conference on Mechatronics and Machine Vision in Practice, pp. 57-62

[10] Collins, K. (2008) Game Sound. An Introduction to History, Theory, and Practice of Video Game Music and Sound Design, Cambridge, Massachusetts: The MIT Press.

[11] Consalvo, M. (2003) Hot Dates and Fairy-Tale Romances: Studying sexuality in videogames, in J. P. Wolf and B. Perron (Eds.) The Video Game Theory Reader, New York: Routledge.

[12] Drachen, A., and Canossa, A. (2009) Towards Gameplay Analysis via Gameplay Metrics, Proc. of the 13th MindTrek, ACM-SIGCHI Publishers, pp. 202-209.

[13] Drachen, A., Canossa, A. (2009) Analyzing Spatial User Behavior in Computer Games using Geographic Information Systems, Proc. of the 13th MindTrek, ACM, pp. 182-189.

[14] Fagerholt, E., Lorentzon, M. (2009) Beyond the HUD. Master of Science Thesis, Department of Computer Science

and Engineering, Chalmers University of Technology, (Göteborg, Sweden).

[15] Ferraro P., Robine M., Allali J., Hanna P., and Rocher T. (2011) Detection of Near-Duplicate Musical Documents from a Multi-Level Comparison of Tonal Information, Information Extraction from the Internet, pp. 129-143.

[16] Grimshaw, M., Schott, G. (2007) Situating Gaming as a Sonic Experience: The Acoustic Ecology of First-Person Shooters, Proc. Situating Play DiGRA 2007.

[17] Heide Smith, J. (2006) Plans and Purposes: How videogame goals shape player behaviour, Unpublished PhD Manuscript, IT University, Copenhagen, pp. 30-31

[18] Huang, C.L, Liao, B.Y. (2001) A robust scene-change detection method for video segmentation. Circuits and Systems for Video Technology, IEEE Transactions, vol.11, no.12, pp.1281-1288, Dec 2001

[19] Joosten, E., Lankveld, G.V, Spronck, P. (2010) Colors and Emotions in Video Games, 11th International Conference on Intelligent Games and Simulation GAME-ON 2010 (ed. Aladdin Ayesh), pp. 61-65.

[20] Jørgensen, K. (2006) On the Functional Aspects of Computer Game Audio. Proc. of the Audio Mostly Conference. Interactive Institute, Piteå, Sweden.

[21] Juul, J. (2009) Fear of Failing? The Many Meanings of Difficulty in Video Games, in Mark J. P. Wolf & Bernard Perron (Eds.), The Video Game Theory Reader 2, New York: Routledge, pp. 237-252.

[22] Kim, J.H., Gunn, D.V., Schuh, E.,Phillips, B.C., Pagulayan, R.J, Wixon, D. (2008) Tracking Real-Time User Experience (TRUE): A comprehensive instrumentation solution for complex systems, Proc. CHI 2008, ACM.

[23] Kivikangas, J. M. (2006) Psychophysiology of Flow Experience: An Explorative Study, Faculty of Behavioural Sciences, Department of Psychology. Helsinki, Finland, University of Helsinki.

[24] Klevjer, R. (2002) In Defence of Cut-Scenes, in F. Mäyrä (Ed.) Proc. Computer Games and Digital Cultures Conference 2002. Tampere University Press.

[25] Lauteren, G. (2002) Pleasure of the Playable Text: Towards an aesthetic theory of computer games, in F. Mäyrä (Ed.) Proc. Computer Games and Digital Cultures Conference 2002. Tampere University Press.

[26] Marczak, R., Vught, J.V., Schott, G. and Nacke, L.E. (2012) Feedback-based gameplay metrics: measuring player experience via automatic visual analysis. Proc. of the IE 2012 Conference (Auckland, New Zealand).

[27] Mirza-Babaei, P., and McAllister, G. Biometric Storyboards: visualizing meaningful gameplay events. BBI Workshop CHI 2011. ACM. (2001)

[28] Moffat, D., and Kiegler, K. (2006) Investigating the effects of music on emotions in games, Proc. of the Audio Mostly Conference.

[29] Nacke, L. E. (2009) Affective Ludology: Scientific Measurement of User Experience in Interactive Entertainment, Ph.D. Thesis. Blekinge Institute of Technology, Karlskrona, Sweden.

[30] Nacke, L. E., Drachen, A. and Goebel, S. (2010) Methods for Evaluating Gameplay Experience in a Serious Gaming Context. International Journal of Computer Science in Sport, vol. 9 (2).

[31] Nacke, L., Drachen, A., Korhonen, H., Kuikkaniemi, K., Niesenhaus, J., van den Hoogen, W. Poels, K., IJsselsteijn, W., de Kort, Y. (2009) DiGRA Panel: Playability and Player Experience Research. Proc. DiGRA 2009.

[32] Nacke, L., Grimshaw, M. (2011) Player-game Interaction Through Affective Sound, Game Sound Technology and Player Interaction: Concepts and Developments, Hershey, PA, USA: IGI Global Publishing.

[33] Nacke, L., and Lindley, C. A. (2008) Flow and Immersion in First-Person Shooters: Measuring the player's gameplay experience, Proceedings of the Conference on Future Play, pp. 81-88.

[34] Orio. N. (2006) Music Retrieval: A Tutorial and Review, Volume 1 of Foundations and Trends in Information Retrieval, Now Pub.

[35] Ruch, A.W. (2010) Videogame Interface: Artefacts and Tropes. Videogame Cultures and the Future of Interactive Entertainment Global Conference, Oxford, UK.

[36] Richard, G., Ramona, M., and Essid, S. (2007) Combined Supervised and Unsupervised Approaches for Automatic Segmentation of Radiophonic Audio Streams. Proc. ICASSP 2007. IEEE International Conference.

[37] Saraceno, C., Leonardi, R. (1997) Audio as a Support to Scene Change Detection and Characterization of Video Sequences: Acoustics, Speech, and Signal Processing, ICASSP-97, IEEE International Conference, vol.4, pp. 2597-2600.

[38] Schott, G., van Vught, J. & Marczak, R. (2012) The 'Dominant Effect' of Games: Content vs. Medium, CoLab: Journal of Creative Technologies, Issue 3, Paper 25.

[39] Serrà, J., Gómez, E., Herrera, P., and Serra, X. (2008) Chroma Binary Similarity and Local Alignment Applied to Cover Song Identification, IEEE Trans. on Audio, Speech and Language Processing, 16(6): 1138-1152.

[40] Skalski, P., Whitbred, R. (2010) Image versus Sound: A Comparison of Formal Feature Effects on Presence and Video Game Enjoyment. PsychNology Journal, 8(1), 67 – 84.

[41] Smith, T.F., and Waterman, M.S. (1981) Identification of Common Molecular Subsequences. Journal of Molecular Biology, 147, pp. 195-197.

[42] Tychsen, A. & Canossa, A. (2008) Defining Personas in Games Using Metrics. FuturePlay Proceedings, ACM. (Toronto, Canada).

[43] Vorderer, P., Hartmann, T., and Klimmt, C. (2003) Explaining the Enjoyment of Playing Video Games: The Role of Competition. In *Proc.* ICEC '03, pp. 1-9.

[44] Waern, A. (2012) Framing Games, Proceedings of Nordic DiGRA 2012, p.1.

[45] Ye, Q., Huang, Q., Jiang, S., Liu, Y., Gao, W. (2005) Jersey Number Detection in Sports Video for Athlete Identification, Visual Communications and Image Processing, Proceedings of the SPIE, Vol. 5960, pp. 1599-1606

[46] Yeh, C.H., Kuo, C.H., and Liou, R.W. (2009) Movie Story Intensity Representation Through Audiovisual Tempo Analysis, Multimedia Tools Appl, 44(2)