Aggregated Visualization of Playtesting Data

Günter Wallner
Department of Industrial Design
Eindhoven University of Technology
g.wallner@tue.nl

Nour Halabi
University of Ontario Institute of Technology
nour.halabi@uoit.ca

Pejman Mirza-Babaei
University of Ontario Institute of Technology
pejman@uoit.ca

ABSTRACT

Playtesting is a key component in the game development process aimed at improving the quality of games through the collection of gameplay data and identification of design issues. Visualization techniques are currently being employed to help integrate quantitative and qualitative data. Despite that, two existing challenges are to determine the level of detail to be presented to developers based on their needs and to effectively communicate the collected data so that informed design changes can be reached. In this paper, we first propose an aggregated visualization technique that makes use of clustering, territory tessellation, and trajectory aggregation to simultaneously display mixed playtesting data. Secondly, to assess the usefulness of our technique we evaluate it through interviews with professional game developers and compare it to a non-aggregated visualization. The results of this study also provide an important contribution towards identifying areas of improvement in the portrayal of gameplay data.

CCS CONCEPTS

• **Human-centered computing** → **Visualization techniques**; • **Applied computing** → **Computer games**.

KEYWORDS

Games User Research, Visual Game Analytics, Data Visualization, Mixed-methods, Physiological Measurements

ACM Reference Format:
Günter Wallner, Nour Halabi, and Pejman Mirza-Babaei. 2019. Aggregated Visualization of Playtesting Data. In *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland Uk.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3290605.3300593

CHI 2019, May 4–9, 2019, Glasgow, Scotland Uk

ACM ISBN 978-1-4503-5970-2/19/05…$15.00
https://doi.org/10.1145/3290605.3300593

1 INTRODUCTION

Developing a game that is fun to play is a complex endeavor due to reasons such as the interdisciplinary nature of game development (incorporating a diverse team of designers, programmers, and artists), the iterative character of the game development process, and the diversity of players who may interact with the game [34]. Playtesting aims to help developers to bring their game closer to their design intent and to deliver a satisfying player experience by providing insights into player behavior and their gameplay experiences in order to help identify and resolve potential problems before release [54].

There are two main data sources from which useful insights can be extracted: objective in-game data (e.g., avatar movement) and subjective player data (e.g., opinions, emotions) [27]. Previous academic work has demonstrated that these data sources are suitable for evaluating user experience in games, for example, by utilizing player movement data for improving level design [21, 31] or by applying physiological measures to assess user engagement in games with regard to the emotional component of their experience [25, 28]. Moreover, the game industry has shown interest in integrating these methods in game development and evaluation [6, 37]. However, there are several challenges for practitioners and researchers alike that need to be addressed before they are able to apply these measurements successfully. First, the sheer quantity of data that can and is collected nowadays needs to be efficiently analyzed and understood [53]. Second, playtesting has – as discussed above – come to rely on different data sources (e.g., game telemetry, physiological measures, interviews). These mixed datasets need to be integrated and tied together in a way such that the advantages of each can be exploited [29]. Third, to facilitate interpretation of the collected data by game developers both their tasks (e.g., improving a specific section in a level, adjusting the difficulty of the game) and their background (e.g., the needs of game programmer vs. the needs of an artist or a producer) [15] have to be taken into account when assimilating the data.

Visualization can greatly assist with these challenges and has thus gradually become an important tool to expedite exploration, analysis, and communication of playtesting data (see [51] for an overview). Despite increasing efforts in game data visualization there are still areas that would benefit from further investigation, including research on visual aggregation techniques. Visualizations simply overlaying the individual behavioral data of multiple players (e.g., [15, 20, 29]) make it difficult to observe common patterns and quickly suffer from overplotting and visual clutter – issues which are magnified when the size of the datasets increases. This makes it often difficult to read and interpret the data and, in turn, derive actionable insights for fixing gameplay issues or suggesting improvements. Aggregation thus plays an essential role for the above tasks as it is necessary for achieving a non-cluttered representation of the data in order to extract general features (cf. [2]) and to obtain an initial overview – one of the basic tasks in information visualization (cf. [17]).

In this paper, we contribute to this line of research by proposing an aggregated visualization which utilizes three different aggregation techniques (clustering, territory tessellation, and trajectory aggregation). These techniques can be used separately or in combination and can also be used to simultaneously display mixed datasets. Triangulating of mixed data sources has been acknowledged to be important in games user research (GUR) [32, 37] but visualization of it has received limited attention so far (see Section 2). We illustrate the proposed technique by using data gathered from *Infinite Mario* [38] and consisting of physiological, observational, and movement data. To study the effectiveness of our aggregated visualization in assisting developers in utilizing playtesting data we then interviewed nine professional game developers in various roles and compared it with a non-aggregated visualization. Thus, the second contribution of this paper (besides proposing a new visualization technique) is to provide an understanding and a supporting argument in when aggregated or non-aggregated visualizations are most appropriate to use. This is an important contribution to the fields of Games User Research and Human-Computer Interaction given the current move towards data-driven decision making and popularity of interactive data visualization.

2 RELATED WORK

Heatmaps are one of the most commonly employed visualization technique for gameplay analysis, the most popular example perhaps being death heatmaps (e.g., [1, 14]). Visually, one technique discussed in this paper bears resemblance with heatmaps. However, while heatmaps show the frequency distribution of a single variable our method visualizes metrics of the values of a variable by subdividing space into small cells. Beside showing death locations or hotspots of other events, heatmaps can also be used to visualize player movements to a certain extent. For example, Mueller et al. [33] used heatmaps to visualize player positions in *Minecraft* to detect locations where players frequently meet. Tremblay et al. [48], on the other hand, made use of heatmaps to depict player movements in a tool for analyzing combat and stealth behaviors. In both cases and in the work of Canossa et al. [1] heatmaps served as an indication of the amount of movement taken place in certain regions. Heatmaps, however, are not ideal for communicating the direction of movement and also smooth over individual differences [35]. For these reasons, we rely on an aggregated path visualization which also explicitly shows individual trajectories deviating from the general movement patterns.

Others have used line segments to represent the individual paths of players either in a 2D or 3D environment. Examples in this regard include the work of Dixit and Youngblood [12] and Wallner et al. [53] as well as Ubisoft's DNA suite [10]. In all three cases, color-coding was used to convey additional information, for example, to distinguish between the paths of different players [53] or to depict the flow of time [12]. Gagne et al. [20], on the other hand, used semitransparent lines to represent player movement to give a sense of the amount of movement in certain areas. Hoobler et al. [23], in turn, used two visual features of the path to encode more than one variable simultaneously, namely thickness to represent time and color to depict team membership. To indicate the direction of movement, Drenikow and Mirza-Babaei [15] – similar to Gagne et al. [20] – augmented the paths with arrow heads placed at regular intervals.

However, visualizations of individual trajectories are vulnerable to clutter if more than a few paths need to be drawn. To overcome this issue a variety of trajectory aggregation techniques have been proposed in the context of geographic visualization (see, e.g., [3, 43]). Notable examples in the games domain include the work of Moura et al. [32] who used lines of varying width to depict how many players moved between predefined areas of the level while we are concerned with the actual routes taken. Mitterhofer et al. [30], concerned with bot-detection in multiplayer online games, employed a clustering approach to detect frequently visited waypoints which are then used to derive an abstract path representation. Similar to Moura et al. [32] and this paper, the width of the path segments is used to convey the number of times a section was traveled. However, in contrast to us, Mitterhofer et al. rely on the general k-means algorithm to cluster points of the paths (which requires to specify the number of clusters beforehand) whereas we use a clustering algorithm [4] specifically developed for movement data. While not strictly games specific, Chittaro et al.'s work [8] on visualizing movement patterns in virtual environments should be mentioned as well as it discusses various way to visualize aggregated movement data, including a flow visualization based on vector fields.

As part of our use case we are also considering the visualization of physiological data, specifically galvanic skin response (GSR). Visualization of physiological measures has received some attention in personal visualization to raise emotional awareness. Typically, these efforts are rather focused on the visualization of individual data although solutions for aggregated visualization have been proposed as well. For example, Kucher et al. [24] use animated glyphs consisting of concentric circles and dot trails to represent GSR and accelerometer values of individuals which are then arranged using a dynamic layout to group people with similar excitement levels. While this gives an impression of the overall excitement of a group, aggregated values are not straightforward to infer and the abstract space representation causes a loss of spatial context which we deem important for our purposes. To make the emotional response to paintings in a museum visible, Du et al. [16] used histograms placed below the painting in a virtual 3D representation of the museum to convey individual GSR values. Furthermore, the floor in front of each painting was colored based on the average GSR values of all observers. This is similar to our approach as in both instances a color-overlay over the environment is used to reflect physiological data. However, in our case the areas to be colored are automatically derived from the players' movement data. Related to games, Robinson et al. [41] proposed an overlay visualization of physiological data atop the *Twitch* user interface to communicate the streamers' emotional state to their viewers. Perhaps most relevant to this paper is, however, the work of Mirza-Babaei et al. [28] who proposed biometric storyboards which relate the intended player experience with the actual physiological reaction of players. For that purpose, physiological data from only a single player is displayed as a line chart along a timeline and further data such as player comments. This time-centric approach makes it particularly suited for games where players experience the content in linear fashion. In contrast, this paper applies a spatial approach where data is displayed in relation to the environment. This makes our method also well-suited for games where the game world can be explored more freely.

While visualization of player data is receiving increasing attention (see Wallner et al. [51] for an overview), work explicitly exploring the visualization of mixed-data sets collected during playtests is still scarce. In our previous paper [29] we proposed a visualization which triangulates movement data, physiological data, and verbal comments on a per-player basis and which we will use for comparison purposes in our study. Recently, Drenikow and Mirza-Babaei [15] proposed an interactive visualization plug-in for the Unity3D engine which integrates, among others, movement data, in-game events, and facial expressions. As with the aforementioned approach, the system only focuses on displaying individual data and thus is prone to overplotting and clutter. The work presented in this paper can be seen as a continuation of these efforts by proposing ways to aggregate playtesting data in order to offer a comprehensive overview.

3 CASE STUDY - INFINITE MARIO

Throughout the paper we will use data gathered from *Infinite Mario* [38] to illustrate and evaluate the proposed visualization techniques. *Infinite Mario* is a 2D platformer inspired by the classic *Nintendo* game *Super Mario Bros.* We have chosen *Infinite Mario* because of the publicly available source-code and because levels can be completed within a relatively short time, in turn, reducing the time commitment required from participants. Moreover, most gamers are familiar with *Mario* games which decreases the time for getting acquainted with the game. While *Infinite Mario* produces the levels procedurally we have designed seven static levels to guarantee the same level geometry across all playtests.

Playtesting Sessions and Data Collection

We recruited six players who played video games on a regular basis and who had experience with platform games through a professional recruiter. Participation in the study was compensated with £30. The test sessions were conducted by a professional GUR experimenter with both industry and academic experience in running playtests. Each participant played the game on a PC and a 24" monitor. The order of the levels was predefined and was kept the same for each player. GSR signals were recorded with a NeXuS-10 MKII biofeedback system and video cameras captured the players while playing the game. In total, we collected data from three different sources, synchronized using an in-game timer:

In-game data: We have instrumented the source code of the game to track the in-game behavior of the player, most importantly movement data which was sampled in regular time intervals. In addition, events such as deaths, collected coins, bumped blocks, and casted fireballs together with positional information were recorded. All collected data was time-stamped using the in-game timer (which started at zero when commencing a level) before being written to a log-file stored on the server hosting the game.

Physiological data: We also captured physiological data in the form of galvanic skin response (GSR) to obtain a measure of the player's arousal state. As arousal levels differ from person to person, absolute values are not directly comparable and were thus converted to relative, normalized values in the range [0..1] using, as suggested by Mandryk et al. [26], the following equation:

$$\tilde{GSR}_t = \frac{GSR_t - GSR^{min}_{t\pm3.5}}{GSR^{max}_{t\pm3.5} - GSR^{min}_{t\pm3.5}}$$

where $GSR^{min}_{t\pm3.5}$ and $GSR^{max}_{t\pm3.5}$ are the minimum and maximum GSR values within a plus/minus 3.5 second window centered around the data point at time t. The 7-second window was chosen based on previous psychophysiological research [39] investigating phasic responses to events in video games. In our case the timestamps associated with the player positions were used to align the arousal values with the in-game location.

Observational data: Videos recorded during the playtest sessions were annotated using *VCode* [22] in an iterative process yielding a total of 12 different categories. Examples include: *tries to go down pipe*, *being careful*, *purposefully avoids enemies*, or *goes back for coins or blocks*. In addition, verbal comments made by the players were transcribed.

4 VISUALIZATION

In this paper we are concerned with visualizing data commonly collected during playtests, in particular discrete events (in our use case, the onset of video codings), player comments, continuous data in relation to player movement (GSR values in our particular example), and movement data itself. First, we will shortly describe – based on our previous work – a way to visualize the individual data before focusing on our proposed aggregation and visualization methods.

Non-aggregated visualization

Figure 1 shows the individual data we collected for one of our *Infinite Mario* levels. Visualization of player trajectories, GSR values, and player comments follows the method we proposed in our previous paper [29]. In particular, GSR values are color-coded using a yellow to red gradient and mapped to the player's trajectory. Comments made by players during the playtesting session are represented using speech bubbles. In addition to [29], discrete events – in our case onsets of video codings – are visualized using small icons.

Such a visualization is appropriate for examining details but it may be difficult to obtain an overview due to several reasons. First, drawing a large number of icons and trajectories can easily lead to visual clutter. Second, representing icons and trajectories individually increases the likelihood of overlappings. As a consequence, values mapped to the trajectories may be partly occluded and not visible anymore. Third, visualizing all trajectories individually makes it difficult to assess and compare the amount of movement in or between areas of a game level. This is further aggravated if trajectories overlap each other as it may give a skewed impression of the amount of movement as, for example, in area *A* in Figure 1.

In the following we address these issues by proposing three ways for aggregating discrete events, continues data in relation to player movement, and movement data itself.

Aggregated visualization

Figure 2 shows the aggregated visualization of the data depicted in Figure 1 using the three aggregation and visualization techniques described in the following.

Clustering of discrete events. To group discrete events (such as observations made when coding the videos – as in our case – or, e.g., automatically recorded events such as collected items) within the vicinity of each other we make use of clustering, specifically we employ the DBSCAN clustering algorithm [18]. We have chosen DBSCAN for several reasons: (i) the number of clusters does not need to be pre-specified, (ii) it can handle clusters of different shape and size, and (iii) it is mostly insensitive to noise and outliers. DBSCAN requires two parameters, namely the minimum number of points min_{pts} a point needs to have within a certain radius ϵ in order to be included in a cluster. Clustering is performed for each type of event separately based on the positions where the events took place. For each identified cluster the barycenter is calculated at which a glyph representing the event is placed. The size of the glyph corresponds to the number of events contained in the cluster.

Trajectory-based space tessellation. To provide the viewer with an overall impression of how a continuous player-related variable (e.g., health, GSR value) varies over the game environment we partition the environment into small non-overlapping regions. These regions are then color-coded based on the value to be represented. As movement is an essential part of the gameplay of many games we argue that it is beneficial to view such player-specific variables in dependence of a players' position. General purpose space partition techniques such as k-d trees [5] or binary space partitioning trees, however, aim for an approximately even distribution of data points to the resulting regions. This, in turn, leads to a fine tessellation of dense regions and a course subdivision of regions with few data points which makes such approaches not well-suited for our purposes. For example, if a region covers a large area of the game environment, the variable under investigation may vary quite considerable. These local, and perhaps important, variations will thus not be reflected in the visualization. For this reason, we are striving for regions of approximately equal size that are derived based on the movement data of the players. To achieve this we make use of the territory tessellation algorithm proposed by Andrienko and Andrienko [4]. In brief, the algorithm applies a specifically for this task developed clustering algorithm which groups the points of the trajectories in such a way that the resulting regions will be of appropriately the same size. The size of the cells is determined by a user-specifiable value r_t. The centroids of the groups are then used as generating points for a Voronoi tessellation. As suggested by Andrienko and Andrienko [4] we also introduce additional seed points in areas which are not covered by trajectories in order to achieve a more regular subdivision of these areas. Each spatial position of a trajectory is then assigned to the Voronoi cell within its boundaries it is located. Assuming

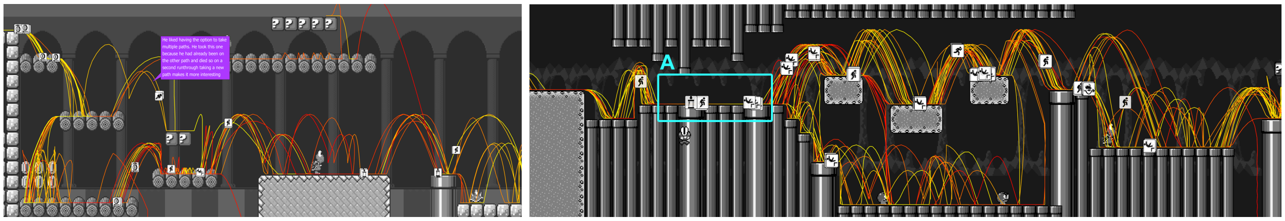

Figure 1: Individual player trajectories are visualized using color-coded connected line segments with color indicating arousal in this example (low high). Discrete events are represented through icons (in this case they correspond to the onset of video codings). Speech bubbles show comments made by the players during the playtest (colors indicate different players).

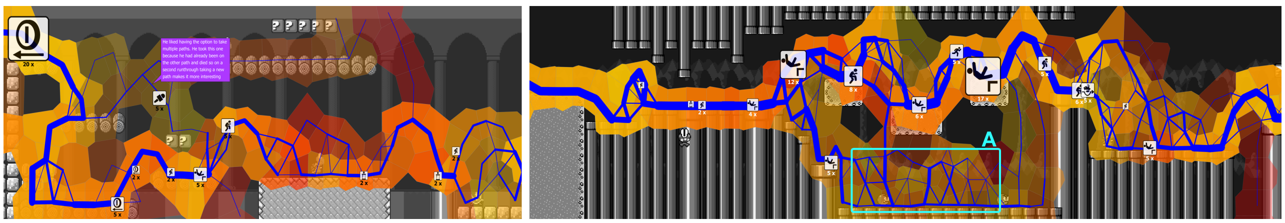

Figure 2: Aggregated visualization of the data depicted in Figure 1. Discrete events are clustered and represented using icons with size encoding the number of clustered events. The environment is spatially decomposed into small cells based on player movement with cell colors reflecting average arousal value (low high) within the cell. Movement between cells is aggregated and represented using lines of varying thickness to indicate the amount of movement. Less traversed cells are rendered more transparent than others to visually accentuate highly traveled areas. Player comments are not aggregated.

that for each position the respective value of the variable under investigation is available we calculate the average value[1] of the variable for each cell. Each cell is represented as a color-coded convex polygon reflecting the average number. In that sense, it can be viewed as a choropleth map where the regions are derived from the movement data. Cells are rendered semi-transparent such that the color-coded cells do not completely occlude the map of the game environment in order to provide spatial context for the analyst. In addition, the transparency is varied based on how often a cell has been passed through by all players (see Figure 2) in order to particularly highlight highly visited areas.

Trajectory aggregation. As pointed out above, drawing trajectories individually can be disadvantages as trajectories may occlude each other. As a consequence the information mapped to the occluded trajectories is not discernible. Furthermore, due to overlappings the number of trajectories in a certain area may not be conveyed properly (see Figure 1, area *A*). In order to address these issues we aggregate trajectories together to provide an overview of the distribution and quantity of movements over the game environment. For that purpose, we reuse the territory tessellation from the previous step and count how often a player crosses the border from one cell to another. This is done by iterating through the points of a trajectory and check in which cell it is located. If points p_i and p_{i+1} are located in different cells a transition takes place. Visually, the moves between the cells are represented by lines whose thickness represents the quantity of movement. It is important to note that this way outliers deviating from popular paths are not excluded but are instead also displayed. Such rare trajectories can be equally informative as popular paths.

5 EVALUATION

We designed a study using semistructured expert interviews and rating scales to evaluate our proposed aggregated visualization technique (Viz$_a$) and to compare it to the visualization showing individual data (Viz$_i$).

Procedure

Before conducting the actual interviews, a pretest with two visualization experts was conducted to assess the study design. After revising the study based on the provided feedback, the interviews were conducted online and by the same interviewer to ensure consistency. Participants were interviewed through either Skype [46] or Discord [11] as both platforms offered screen sharing capabilities. In addition, Realtime-Board [40], an online collaboration tool, was used to display the visualizations. We have opted for this tool because it offers easy access through a link (without requiring registration) and allows participants to zoom and move around to

[1]However, we should note that other measures such as maximum or minimum can be easily used instead as well.

closely examine each visualization. The participants received a consent form ahead of time to read and sign and the study was recorded using Open Broadcast Software [36] for future analysis. The interviews itself consisted of two stages:

In Stage 1 a semi-structured interview took place where participants were asked questions concerning their background in the game development industry and their experience with user testing and user test reports.

In Stage 2, the participants were presented with the different visualizations. Participants were asked to share their screens during this stage to facilitate the discussions and later analysis. For this part we prepared one aggregated and one non-aggregated visualization for two different levels of the game (yielding four visualizations in total[2]). Excerpts of these visualizations can be seen in Figure 1 and 2. Different levels were used for both visualizations to control for the possibility that participants' responses to one visualization would be influenced from insights extracted from the other visualization. Before a visualization was shown to the participants they were provided with a written explanation of the visualization and a legend describing the different parts.

After participants had sufficient time to view and get acquainted with both visualizations, we followed up with a discussion about their thoughts of the visualizations and provided them with tasks which required them to examine the visualizations in more detail. For example, we asked them if they got a sense of where players may have died or had trouble making a jump, and what they liked or did not like about each visualization. Finally, subjects were asked which of the two visualizations they prefer.

Afterwards, participants had to rate the visualization according to six quality measures on a five-point scale anchored by *poor* (1) and *excellent* (5). These measures were drawn from previous research on gameplay visualization from Wallner and Kriglstein [52] and include:[3] *clarity* (is the displayed data clearly interpretable or ambiguous), *readability* (are the visual elements easily legible and distinguishable), *informativeness* (does it provide interesting or new information), *aesthetic appeal* (is it visually appealing), *accurateness* (is the displayed data accurate enough), and *usefulness* (for which tasks is it useful). Responses to these scales will be treated as ordinal and analyzed using non-parametric statistics.

Participants

For the study we strove to recruit a diverse group of video game professionals that cover different roles in game development and have varying degree of visualization expertise. In total, nine video game professionals (P1: Designer/User Researcher, P2: Designer, P3: Gameplay Programmer, P4: Programmer, P5: Content System Designer, P6: Data Analyst, P7: Content Designer, P8: Designer, P9: Technical Producer) from Canadian and US companies took part in this study. One out of the nine participants was female. Four participants had less than 3 years, four had between 3 to 4 years, and one participant had 15 years of industry experience. Three participants considered themselves not or only slightly familiar with visualization, one as moderately familiar, and five as very or extremely familiar. Furthermore, seven participants worked in a mid-sized gaming company, one in a large company, and one in an indie company.

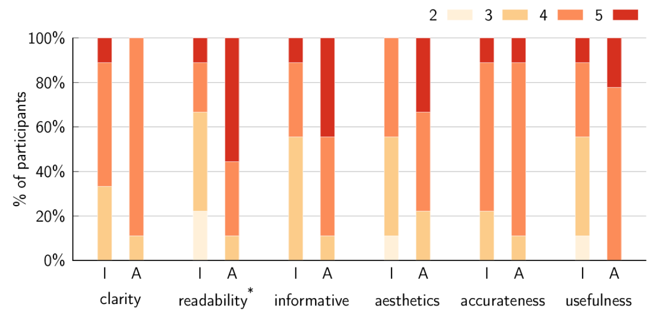

Figure 3: Participants' ratings of the two visualizations with respect to six criteria (I = individual, A = aggregated, *significant differences in ratings at $p < .05$).

Analysis

Transcripts of the recorded interview sessions were prepared and analyzed using MAXQDA [49]. For the analysis of the transcripts a deductive qualitative content analysis [44] with pre-defined categories derived from the six quality measures (readability, usefulness, accurateness, ...) was employed. We have chosen these categories in order to understand the participants reasoning behind their ratings of $\textsc{Viz}_i$ and $\textsc{Viz}_a$. Each statement was further labeled as either positive or negative to reflect the participant's sentiment. Furthermore, two additional codes were used to specify towards which visualization ($\textsc{Viz}_a$ and $\textsc{Viz}_i$) the statement was directed. The coding was performed independently by two coders with discrepancies being resolved through discussion.

6 RESULTS

In the following, the ratings and the results of the qualitative analysis of the interview transcripts will be presented with respect to the six quality measures.

Ratings

Figure 3 gives on overview of the distribution of the ratings of the individual and aggregated visualization according to the six quality measures. Ratings between the two versions

[2] Included in the supplementary material.

[3] Wallner and Kriglstein [52] list seven measures of which *readability* and *ease of extraction* have been replaced by a single *readability* category here as participants of the pretest struggled with the differentiation.

for each criteria were compared using Wilcoxon signed-rank tests using a significance level of $\alpha = .05$. Effect sizes were calculated following Rosenthal [42] using $Z/\sqrt{N}$ with N being the number of observations. Results indicated that the aggregated visualization was rated significantly higher in terms of *readability* ($Z = -2.209, p = .027$) with a large effect size (according to Cohen's criteria [9]) of $r = .52$. In terms of usefulness, analyses did not indicate a significant result ($Z = -1.933, p = .053$) but still yielded a medium effect of $r = .46$ in favor of the aggregated visualization. All other quality measure were statistically non-significant with effect sizes of $r < .4$.

Interviews

In the following we discuss participants' comments concerning the six quality measures. Table 1 provides a summary of the number of statements with respect to the six categories.

Readability. Viz$_a$ was better received (48 positive statements) than Viz$_i$ (51 negative statements) in regard to how easily different visual elements were extracted. This is also in line with the ratings, where Viz$_a$ received significantly higher scores than Viz$_i$.

Concerning Viz$_i$, comments were made relating to the lines creating too much clutter to extract information quickly and to cause clusters of illegible data which make it difficult to keep track of a single player's movement (24 statements from 8 participants). These occlusions also prevented participants from extracting data from the individual lines. Similarly, the readability of the icons and judging the frequency of occurrence of the depicted events was low either because of their small size or being occluded by other icons (16 statements; 8 participants). Additionally, Viz$_i$ was found easy to extract information about level pacing, deaths due to falling off the map, retracing player movements, and discerning enemy locations.

Viz$_a$, on the other hand, received predominantly positive statements with only 14 negative comments. In particular, subjects found that Viz$_a$ makes it easier to assess the overall arousal state of different sections of a level (4 statements; 3 participants) and to understand player behavior at a glance (23 statements; all participants) compared to Viz$_i$, for example, due to the frequency count being reflected by the size of the icons. In addition, player divergence from the main path and player struggles were easier to read due to varying thickness of the line and the change in opacity (10 statements; 5 participants).

What I liked the most of the aggregated view is the number of times individuals were confronted with a problem [icons] and the way that stacks and I also like that the size of the path is telling you what was the most commonly taken path. I think that's a little easier to read and it's a little more viable. [P5]

Table 1: Results of the deductive qualitative content analysis. Number of positive (+) and negative (-) statements concerning the six categories, grouped by visualization (Viz$_a$ = aggregated, Viz$_i$ = non-aggregated).

Category	+ Viz$_a$	-	+ Viz$_i$	-
Readability	48	14	24	51
Usefulness	35	5	24	6
Accurateness	1	4	4	4
Aesthetic Appeal	17	10	6	11
Informativeness	31	22	22	23
Clarity	26	40	33	27

Contrarily, some participants found it difficult to extract information in Viz$_a$ such as enemy and power-up locations due to the colored arousal data being superimposed on the level map or due to increased icon size (6 statements; 5 participants). This difficulty also appeared in connection with Viz$_i$ in areas with high movement.

Usefulness. Both visualizations received a large number of positive comments regarding their usefulness for gameplay analysis with both being deemed helpful for getting an idea of which issues need to be fixed and for identifying possible solutions. Concerning Viz$_i$ subjects declared they found it useful for analyzing how to provide incentives to players to perform a certain action (2 statements; 2 participants) and for assessing difficult areas of a level (5 statements; 3 participants), as exemplified by the following quote:

I can see all the individual players and what exactly they are doing. I find it more useful for gauging difficulty. [P8]

Viz$_a$ was considered effective for identifying and reporting main issues of a level (17 statements; 7 participants) and understanding pacing (climax/cooldown) (6 statements; 3 participants). Moreover, three participants found the frequency of behavior portrayed through the icons useful for comparing the actual to the intended experience, and two participants were able to distinguish between level design issues and player skill in Viz$_a$.

So it's more of like did we want that to happen because that's going to happen to a lot of our players. [P3]

One participant considered Viz$_a$ less useful for determining level difficulty. Additionally, one subject mentioned that the visualization would be even more useful if it would be interactive to be able to adjust the displayed data based on the analysis task to facilitate better decision making.

Accurateness. In general, subjects did not talk much about the accuracy of both visualizations, possibly indicating that accurateness has not been a major point of concern for both visualizations.

Concerning accurateness of Viz$_i$ participants had diverging opinions as reflected by an even number of positive and negative statements (by three participants). One participant felt it is not necessary to have pixel accurate information on player traversal of the level and that aggregated data is sufficient. Another participant questioned the accurateness of the data in Viz$_i$ when there was a mismatch between the encoded event and the possible player behavior, or when there was not enough supportive data to back up expectations.

I don't really understand how people can miss a jump over there. I don't know if that's even possible. [P3]

In other instances, participants found Viz$_i$ more accurate than Viz$_a$ due to the higher level of detail and clarity of the individual player paths.

Viz$_a$ received five statements pertaining to accurateness with one being positive. Critical comments in this regard, mainly questioned Viz$_a$ due to the participants not understanding how the data was aggregated or when they had problems when icons did not reflect player events as expected (requiring them to refer to Viz$_i$ for further clarification).

I'm not sure there but it was a tricky jump [...] it doesn't look like he got caught in this corner because we don't have positional data in that corner [...] So that's maybe something I would have to look at in an individual play session to see okay how did we code this or why did we code it. [P6]

Aesthetic Appeal. Each visualization received positive and negative statements referring to the aesthetics of the visualization. However, Viz$_a$ received almost three times more positive statements then Viz$_i$. Both visualizations received positive comments relating to the color choices being complementary allowing for information to stand out. At the same time, two participants pointed out that the red and yellow colors may cause problems for individuals with color vision deficiency.

Concerning Viz$_i$, subjects found that the smoothness of the lines was aesthetically pleasing as it better portrayed the jumps. Some of the negative comments regarding aesthetics were related to icons not being easy to spot to due their small size or low contrast or how they added noise to the visualization (5 statements; 2 participants). It was also mentioned that areas with increased movement negatively affected the aesthetics due to clutter (3 statements; 2 participants).

Viz$_a$ was mainly praised for presenting information in a way such that it clearly stands out, for example, through the contrast in color and the aggregated icons.

They [icons] actually pop out really well because they are still black and white, but they are layered on top of the colors so they stand out really well. [P6]

However, some participants felt that the cells from the territory tessellation and the thin lines showing minor cell transitions (such as in area *A* in Figure 2) added noise to the visualization (4 statements; 3 participants).

Informativeness. Both visualizations received over 20 positive and negative statements concerning the visualizations' ability to provide new or interesting information with Viz$_a$ receiving the most positive statements (31).

Pertaining to both visualizations, two individuals stated that they do not find the arousal data informative. Additionally, two participants commented that more interesting information could be derived from the icons such as "avoided power-ups" or "avoided enemy" if they were more specific (i.e., icons which differentiate between different enemies and power-ups or between players going back for either a coin or a power-up). Some positive points that were highlighted in both visualizations are their ability to inform where players missed a jump and had to go back and how the icons and arousal data point to areas of player struggle (Viz$_i$: 12 statements; 5 participants, Viz$_a$: 14 statements; 8 participants).

Specifically in Viz$_i$, player deaths through lines trailing off the map or sudden line ends provided information on areas of struggle. However, the interviews highlighted how certain information is lost due to clutter or player trajectories occupying the same space (cf. *readability*, Figure 1 – area *A*).

So any individual user data becomes lost because as more people are traveling through there you can get to the point where some people turn around or some people die and you don't know that because it's just one thin line. [P2]

In addition, participants found that information cannot be deduced in high density areas in Viz$_i$, and that individual lines may not be as informative as Viz$_a$ (10 statements; 6 participants). Participants, for example, commented:

I can't tell what happened here [cluster of lines in Viz$_i$]. It just looks like a cluster. Yeah, it's informative but again not as much as the last one [Viz$_a$]. [P9]

Regarding Viz$_a$, seven subjects said that the icons and their corresponding frequency value provide interesting information about players' behavior and where the main issues are.

The icon and the thought bubble you know made it really obvious that hey people were trying and not only that but why they were trying. [P7]

Another point that was expressed during the interview by one participant was that information regarding the exact number of players diverging from the main path becomes lost as it is not depicted textually.

Clarity. Viz$_i$ received slightly more positive (33) than negative statements (27) concerning how clearly the information is interpretable. In contrast, participants were more critical concerning Viz$_a$ with 40 statements classified as negative and 26 as positive. Participants found it is clear what both visualizations are trying to communicate such as players' experiences within a level, points of difficulty, and player behavior (e.g., where players fell off the map). However, it was not clear which items were hidden in the power-up boxes.

Adding where power-ups are would be helpful as well as [...] to say where there is a danger so that if there is a failed jump – its because of danger there and not random failed jumps. [P2]

The level geometry in Viz_i is clear as it is not obstructed by larger icons, but in areas of high density information it becomes difficult to interpret player behavior. It was mentioned that the number of players who pass through an area is not clear as the individual lines occlude one another (resembles one thin line) resulting in loss of information (10 statements; 5 participants). Two subjects, for instance, reverted to counting the number of lines in Viz_i to get an idea about the number of players the data was from.

In relation to Viz_a, it was noted that the visualization provides a clear overview of the level and that the arousal areas are easily interpretable. The majority of the negative comments (15 statements; 5 subjects) were in relation to how level information was obstructed by iconography and arousal data making it less clear where enemies or power-ups are.

Participants' preferences. The results indicate that the aggregated visualization (Viz_a) was highly preferred by the majority of the participants. Two participants (P1 and P4) preferred to use both visualizations for game evaluation and only one (P8) preferred Viz_i. P4's reasoning was that Viz_a could be used to get a general idea of issues but when specific data is required he would switch to Viz_i. P1 stated that his choice would depend on the specific case he is dealing with:

If I was reporting to a designer who said: Hey what are the main issues? Then, Viz_a would be the one I chose because it is easier to see the icons [...] it's easier seeing that single picture and having that single picture style (referring to Viz_a). [P1]

P8 preferred to use Viz_i as she found it was easier to gauge difficulty and was able to better differentiate between players.

7 DISCUSSION

The most pronounced differences between the two visualizations appeared in terms of **readability**, with participants rating the aggregated visualization significantly higher than the visualization of the individual data. This better performance in terms of readability was also evident in the participants comments with Viz_a being received more positively, mainly because of the lines in Viz_i causing clutter, lines being superimposed and thus occluding information, and the large amounts of – sometimes overlapping – small icons making it difficult to judge the frequency of occurrence of events. At the same time, Viz_a was considered easier to read due to icons being aggregated and their size reflecting occurrence and the varying thickness of the lines assisting in perceiving main paths and divergences from it. Concerns with respect to Viz_a were mostly voiced in regard to occlusions of the level information caused by larger icons and the colored cells of the territory tessellation. These issues could be solved by allowing to tune icon sizes and the opacity of icons and cells.

Surprisingly, the **accuracy** of the aggregated visualization seems not to have been a major source of concern. Both visualizations were rated very similarly and favorably (cf. Figure 3). Participants also did not express much concern during the interviews – at least in light of the tasks for which the aggregated visualization was deemed useful. Any form of aggregation goes in hand with a loss of detail to be able to highlight patterns and general features. Consequently, Viz_a was mostly deemed **useful** for identifying main issues (e.g., areas of struggle, major paths, comparing expected to actual behavior). Besides, Viz_a was also considered useful to more easily communicate identified issues to stakeholders which do not require excessive detail (e.g., designers). Viz_i, on the other hand, was preferred for getting details and inspecting players individually, for example, to gauge difficulty. However, with increasing data the issue of clutter (already present in our dataset) gets aggravated and will make it increasingly hard to follow individual player traces. In this sense, aggregated visualization also scale better to large datasets. In our case the level of abstraction can be adjusted by changing the parameters of the clustering and territory tessellation.

Participants found the data displayed to be **informative**, with Viz_a receiving slightly higher ratings and comparatively more positive feedback. One pertinent issue was that defiances in readability (e.g., clutter, occlusions) directly affected how instructive the visualization is. Some participants also found individual player traces not as informative as the aggregated ones but this is likely a matter of the analysis goal as is the request of some subjects for more specific icons.

One important aspect in the design of Viz_a was that the analyst can see all the gathered data in relation to the game environment. However, when asked about **clarity** the interviews revealed that a major issue in terms of Viz_a was level information (e.g., location of enemies) not being visible due to being obstructed by large icons and color-coded cells. This resulted in a loss of context which is essential to interpret behavior and to draw meaningful conclusions. This issue could be remedied by specifically highlighting objects which are considered vital for interpreting the data (e.g., power-ups).

Concerning the **aesthetics** of the visualization, Viz_a received considerably more positive feedback than Viz_i. Approximately two-thirds of the statements concerning Viz_a were positive while only about one-third was positive for Viz_i. This is also partly reflected in the ratings where Viz_a received more favorable scores, although the differences were not statistically significant. From the statements it is evident that subjects especially appreciated that information in Viz_a was clearly differentiable and perceivable due to the contrast in color and aggregation. However, some subjects found it noisy in areas with many but minor cell transitions.

Ultimately, it is clear that each visualization has advantages and drawbacks. In this regard, one point that repeatedly emerged during the interviews was to make the visualization interactive to leverage the benefits of both. Some of the suggestions included being able to select a line or a few lines to view just a selection, user-definable colors, and being able to toggle visibility of selected elements such as turning off the arousal data in the aggregated visualization. In the present study, our aim was to assess the appropriateness of the proposed techniques and to identify areas of improvement in the portrayal of gameplay data and thus we purposefully kept the visualization static. However, our future goal is to use the insights gathered in this study to turn our approach into an interactive system to allow for an *overview first, details on demand* [45] exploration process.

While we have used a specific case study with specific data for each of the three aggregation techniques described in Section 4 we would like to remark that these are not confined to these particular data types. For example, instead of clustering video codings, in-game events such as casting a certain spell or teleporting to another location could be represented with glyphs too. Automatically detected facial expressions could be another type of data which could be visualized with this method. Similarly, the space tessellation approach could be used to visualize continuous variables other than GSR as well. For instance, health could be displayed to get an overview of the average healthiness of players in relation to their in-game position. While in terms of visual appearance this approach is similar to heatmaps we would like to reemphasize that a heatmap shows frequency of occurrence while the proposed technique can be used to show descriptive measures of a variable (average GSR value in our use case). With respect to trajectory aggregation and visualization we have opted to omit directional information for our particular case study but it should be noted that such kind of information can be easily added to the visualization, for instance, by using bi-directional arrows (cf. [4, 47]).

Since we are focusing on three frequently occurring kinds of data (movement, discrete events, and player-related measures) our approach readily translates to other games where gameplay is movement-driven and where the player controls a single character. We would also like to point out that although the visualization is currently realized in 2D, it also applies to 3D games as in many cases the data can be projected to 2D (as it us usually done in case of heatmaps).

Due to our desire to visualize mixed playtesting data we used a rather small dataset (qualitative data is usually collected on a smaller scale than game telemetry) to illustrate the proposed visualization. It should be stressed, however, that the algorithms employed also scale well to larger data sets. The algorithm for grouping the points of the trajectories as part of the territory tessellation (see Section 4) has a runtime linear to the number of spatial positions to be processed. We did not perform any pre-processing on the input trajectories but in cases where large numbers of trajectories need to be treated or have been recorded at a resolution higher than necessary, performance can be improved by simplifying the trajectories beforehand. This can be achieved either with general line simplification algorithms such as Douglas-Peucker [13] or Visvalingam-Whyatt [50] or with specifically designed trajectory simplification algorithm such as those proposed by Andrienko and Andrienko [4] or Chen et al. [7]. For generating the Voronoi tessellation we currently use Fortune's algorithm [19] which performs the task in $O(n \log n)$ time with n being the number of generating points. DBSCAN also has an average runtime of $O(m \log m)$ (cf. [18]) where m is the number of geospatial events to be clustered. Lastly, aggregating the trajectories based on the territory tessellation is linear with respect to the number of points in the trajectories. Lastly, we should note that DBSCAN is not well-suited if there are large variations in density. However, as we are mainly interested in small, locally confined, clusters we do not expect this to be a major deficiency.

8 CONCLUSIONS

The paper proposed a new aggregated visualization and presented a study that provides support for employing aggregated visualization techniques to provide game developers with actionable insights as they will yield higher readability and a more efficient overview of the collected playtesting data. A successful visualization approach would utilize both aggregated and non-aggregated techniques as game developers, depending on their role, require different levels of abstraction. For example, a level designer may benefit from non-aggregated data to fix a specific issue in a level, whereas a creative director may take advantage from a high-level visualization to quickly identify levels that require design review. Visualization techniques are crucial to make comparisons between design intent and players actual experience data as concise as possible. Furthermore, we have evidence that our technique successfully overcomes some of the drawbacks we identified regarding the individual representation of data. Moreover, it emphasizes the need for user- and task-adaptive visualizations to accommodate the wide range of users and analysis tasks arising in game development.

ACKNOWLEDGMENTS

We would like to thank the interviewees for their time and feedback. This research was supported in part by NSERC Discovery (GPIN-2014-05763) and NSERC CREATE CLUE.

REFERENCES

[1] Canossa Alessandro, Nguyen Truong-Huy D., and Seif El-Nasr Magy. 2016. G-Player: Exploratory Visual Analytics for Accessible Knowledge Discovery. In *DiGRA/FDG'16 - Proceedings of the First International Joint Conference of DiGRA and FDG*. Digital Games Research Association and Society for the Advancement of the Science of Digital Games, Dundee, Scotland.

[2] Gennady Andrienko and Natalia Andrienko. 2008. Spatio-temporal aggregation for visual analysis of movements. In *IEEE Symposium on Visual Analytics Science and Technology*. IEEE, Washington, DC, USA, 51–58. https://doi.org/10.1109/VAST.2008.4677356

[3] Gennady Andrienko and Natalia Andrienko. 2010. A General Framework for Using Aggregation in Visual Exploration of Movement Data. *The Cartographic Journal* 47, 1 (2010), 22–40. https://doi.org/10.1179/000870409X12525737905042

[4] Natalia Andrienko and Gennady Andrienko. 2011. Spatial Generalization and Aggregation of Massive Movement Data. *IEEE Transactions on Visualization and Computer Graphics* 17, 2 (Feb 2011), 205–219.

[5] Jon Louis Bentley. 1975. Multidimensional Binary Search Trees Used for Associative Searching. *Commun. ACM* 18, 9 (Sept. 1975), 509–517. https://doi.org/10.1145/361002.361007

[6] James Berg. 2018. Dissecting the Dragon: GUR for Dragon Age: Inquisition. In *Games User Research*, Anders Drachen, Pejman Mirza-Babaei, and Lennart Nacke (Eds.). Oxford University Press, Oxford, 403–416. https://doi.org/10.1093/oso/9780198794844.003.0023

[7] Yukun Chen, Kai Jiang, Yu Zheng, Chunping Li, and Nenghai Yu. 2009. Trajectory Simplification Method for Location-based Social Networking Services. In *Proceedings of the 2009 International Workshop on Location Based Social Networks (LBSN '09)*. ACM, New York, NY, USA, 33–40. https://doi.org/10.1145/1629890.1629898

[8] Luca Chittaro, Roberto Ranon, and Lucio Ieronutti. 2006. VU-Flow: A Visualization Tool for Analyzing Navigation in Virtual Environments. *IEEE Transactions on Visualization and Computer Graphics* 12, 6 (Nov. 2006), 1475–1485. https://doi.org/10.1109/TVCG.2006.109

[9] Jacob Cohen. 1988. *Statistical Power Analysis for the Behavioral Sciences* (2 ed.). L. Erlbaum Associates.

[10] Jonathan Dankoff. 2014. Game Telemetry with DNA Tracking on Assassin's Creed. https://www.gamasutra.com/blogs/JonathanDankoff/20140320/213624/Game_Telemetry_with_DNA_Tracking_on_Assassins_Creed.php Accessed: September, 2018.

[11] Discord Inc. 2015–2018. Discord. https://discordapp.com/ Accessed: September, 2018.

[12] Priyesh N Dixit and G Michael Youngblood. 2008. Understanding playtest data through visual data mining in interactive 3d environments. In *12th International Conference on Computer Games: AI, Animation, Mobile, Interactive Multimedia and Serious Games (CGAMES)*. 34–42.

[13] David H. Douglas and Thomas K. Peucker. 1973. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica: The International Journal for Geographic Information and Geovisualization* 10, 2 (1973), 112–122. https://doi.org/10.3138/FM57-6770-U75U-7727

[14] Anders Drachen and Alessandro Canossa. 2009. Analyzing Spatial User Behavior in Computer Games Using Geographic Information Systems. In *Proceedings of the 13th International MindTrek Conference: Everyday Life in the Ubiquitous Era (MindTrek '09)*. ACM, New York, NY, USA, 182–189. https://doi.org/10.1145/1621841.1621875

[15] Brandon Drenikow and Pejman Mirza-Babaei. 2017. Vixen: Interactive Visualization of Gameplay Experiences. In *Proceedings of the 12th International Conference on the Foundations of Digital Games (FDG '17)*. ACM, New York, NY, USA, 3:1–3:10. https://doi.org/10.1145/3102071.3102089

[16] Shen Du, Edouard Shu, Feifei Tong, Yinghao Ge, Lu Li, Jingbo Qiu, Philippe Guillotel, Julien Fleureau, Fabien Danieau, and Daniel Muller. 2016. Visualizing the Emotional Journey of a Museum. In *Proceedings of the 2016 EmoVis Conference on Emotion and Visualization (EmoVis'16)*. Linkoping University, Linkoping, Sweden, Sweden, 7–14. https://doi.org/10.3384/ecp10302

[17] N. Elmqvist and J.D. Fekete. 2010. Hierarchical Aggregation for Information Visualization: Overview, Techniques, and Design Guidelines. *IEEE Transactions on Visualization and Computer Graphics* 16, 3 (2010), 439–454. https://doi.org/10.1109/TVCG.2009.84

[18] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. 1996. A Density-based Algorithm for Discovering Clusters a Density-based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD'96)*. AAAI Press, 226–231.

[19] S Fortune. 1986. A Sweepline Algorithm for Voronoi Diagrams. In *Proceedings of the Second Annual Symposium on Computational Geometry (SCG '86)*. ACM, New York, NY, USA, 313–322. https://doi.org/10.1145/10515.10549

[20] André R. Gagné, Magy Seif El-Nasr, and Chris D. Shaw. 2012. Analysis of Telemetry Data from a Real-time Strategy Game: A Case Study. *Computers in Entertainment* 10, 1 (2012), 2:1–2:25. https://doi.org/10.1145/2381876.2381878

[21] Pietro Guardini and Paolo Maninetti. 2013. Better Game Experience Through Game Metrics: A Rally Videogame Case Study. In *Game Analytics: Maximizing the Value of Player Data*, Magy Seif El-Nasr, Anders Drachen, and Alessandro Canossa (Eds.). Springer, London, 325–361. https://doi.org/10.1007/978-1-4471-4769-5_16

[22] Joshua Hailpern and Joey Hagedorn. 2008. VCode. http://social.cs.uiuc.edu/projects/vcode.html (Accessed September, 2018).

[23] Nate Hoobler, Greg Humphreys, and Maneesh Agrawala. 2004. Visualizing competitive behaviors in multi-user virtual environments. In *IEEE Visualization*. IEEE, Washington, DC, USA, 163–170.

[24] Kostiantyn Kucher, Daniel Cernea, and Andreas Kerren. 2016. Visualizing Excitement of Individuals and Groups. In *Proceedings of the 2016 EmoVis Conference on Emotion and Visualization (EmoVis'16)*. Linkoping University, Linkoping, Sweden, Sweden, 15–22. https://doi.org/10.3384/ecp10303

[25] Regan L. Mandryk and M. Stella Atkins. 2007. A Fuzzy Physiological Approach for Continuously Modeling Emotion During Interaction with Play Technologies. *Int. J. Hum.-Comput. Stud.* 65, 4 (April 2007), 329–347. https://doi.org/10.1016/j.ijhcs.2006.11.011

[26] Regan L. Mandryk, M. Stella Atkins, and Kori M. Inkpen. 2006. A Continuous and Objective Evaluation of Emotional Experience with Interactive Play Environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*. ACM, New York, NY, USA, 1027–1036. https://doi.org/10.1145/1124772.1124926

[27] Graham McAllister, Pejman Mirza-Babaei, and Jason Avent. 2013. Improving Gameplay with Game Metrics and Player Metrics. In *Game Analytics: Maximizing the Value of Player Data*, Magy Seif El-Nasr, Anders Drachen, and Alessandro Canossa (Eds.). Springer, London, 621–638. https://doi.org/10.1007/978-1-4471-4769-5_27

[28] Pejman Mirza-Babaei, Lennart E. Nacke, John Gregory, Nick Collins, and Geraldine Fitzpatrick. 2013. How Does It Play Better?: Exploring User Testing and Biometric Storyboards in Games User Research. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 1499–1508. https://doi.org/10.1145/2470654.2466200

[29] Pejman Mirza-Babaei, Günter Wallner, Graham McAllister, and Lennart E. Nacke. 2014. Unified Visualization of Quantitative and Qualitative Playtesting Data. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1363–1368.

https://doi.org/10.1145/2559206.2581224
[30] Stefan Mitterhofer, Christian Platzer, Christopher Kruegel, and Engin Kirda. 2009. Server-Side Bot Detection in Massively Multiplayer Online Games. *IEEE Security & Privacy* 7, 3 (2009), 29–36. https://doi.org/10.1109/MSP.2009.78
[31] Dinara Moura and Magy Seif El-Nasr. 2015. Design Techniques for Planning Navigational Systems in 3-D Video Games. *Comput. Entertain.* 12, 2 (Feb. 2015), 2:1–2:25. https://doi.org/10.1145/2701657.2633421
[32] Dinara Moura, Magy Seif el Nasr, and Christopher D. Shaw. 2011. Visualizing and Understanding Players' Behavior in Video Games: Discovering Patterns and Supporting Aggregation and Comparison. In *Proceedings of the 2011 ACM SIGGRAPH Symposium on Video Games (Sandbox '11)*. ACM, New York, NY, USA, 11–15. https://doi.org/10.1145/2018556.2018559
[33] Stephan Mueller, Barbara Solenthaler, Mubbasir Kapadia, Seth Frey, Severin Klingler, Richard P. Mann, Robert W. Sumner, and Markus Gross. 2015. HeapCraft: Interactive Data Exploration and Visualization Tools for Understanding and Influencing Player Behavior in Minecraft. In *Proceedings of the 8th ACM SIGGRAPH Conference on Motion in Games (MIG '15)*. ACM, New York, NY, USA, 237–241. https://doi.org/10.1145/2822013.2822033
[34] Lennart E. Nacke, Pejman Mirza-Babaei, Magy Seif El-Nasr, Heather W. Desurvire, and Regina Bernhaupt. 2014. Games and Entertainment Community SIG: Reaching Beyond CHI. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1123–1126. https://doi.org/10.1145/2559206.2559216
[35] Jakob Nielsen and Kara Pernice. 2010. *Eyetracking web usability*. New Riders.
[36] OBS Studio Contributors. 2012–2018. Open Broadcaster Software. https://obsproject.com/ Accessed: September, 2018.
[37] Randy J. Pagulayan, Kevin Keeker, Dennis Wixon, Ramon L. Romero, and Thomas Fuller. 2003. User-centered Design in Games. In *The Human-computer Interaction Handbook*, Julie A. Jacko and Andrew Sears (Eds.). L. Erlbaum Associates Inc., Hillsdale, NJ, USA, 883–906.
[38] Markus Persson. 2008. *Infinite Mario Bros.* Game [Web].
[39] Niklas Ravaja, Marko Turpeinen, Timo Saari, Sampsa Puttonen, and Liisa Keltikangas-Järvinen. 2008. The psychophysiology of James Bond: Phasic emotional responses to violent video game events. *Emotion* 8, 1 (2008), 114–120.
[40] RealtimeBoard Inc. 2011–2018. RealtimeBoard. https://realtimeboard.com/ Accessed: September, 2018.
[41] Raquel Robinson, Zachary Rubin, Elena Márquez Segura, and Katherine Isbister. 2017. All the Feels: Designing a Tool That Reveals Streamers' Biometrics to Spectators. In *Proceedings of the 12th International Conference on the Foundations of Digital Games (FDG '17)*. ACM, New York, NY, USA, 36:1–36:6. https://doi.org/10.1145/3102071.3102103
[42] Robert Rosenthal. 1994. Parametric measures of effect size. In *The Handbook of Research Synthesis*, Harris Cooper and Larry V. Hedges (Eds.). Russell Sage Foundation, New York, 231–244.
[43] Roeland Scheepens, Niels Willems, Huub van de Wetering, Gennady Andrienko, Natalia Andrienko, and Jarke J. van Wijk. 2011. Composite Density Maps for Multivariate Trajectories. *IEEE Transactions on Visualization and Computer Graphics* 17, 12 (Dec. 2011), 2518–2527. https://doi.org/10.1109/TVCG.2011.181
[44] Margrit Schreier. 2012. *Qualitative Content Analysis in Practice*. Sage Publications, London, UK.
[45] Ben Shneiderman. 1996. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In *Proceedings of the 1996 IEEE Symposium on Visual Languages*. IEEE, Washington, DC, USA, 336–343. https://doi.org/10.1109/VL.1996.545307
[46] Skype Technologies. 2003–2018. Skype. https://www.skype.com/ Accessed: September, 2018.
[47] Waldo R. Tobler. 1987. Experiments In Migration Mapping By Computer. *The American Cartographer* 14, 2 (1987), 155–163. https://doi.org/10.1559/152304087783875273
[48] Jonathan Tremblay, Pedro Andrade Torres, and Clark Verbrugge. 2014. An algorithmic approach to analyzing combat and stealth games. In *IEEE Conference on Computational Intelligence and Games (CIG)*. IEEE, Washington, DC, USA, 1–8. https://doi.org/10.1109/CIG.2014.6932898
[49] VERBI GmbH. 1989–2018. MAXQDA. https://www.maxqda.com/ Accessed: September, 2018.
[50] Maheswari Visvalingam and James Duncan Whyatt. 1993. Line generalisation by repeated elimination of points. *The Cartographic Journal* 30, 1 (1993), 46–51. https://doi.org/10.1179/000870493786962263
[51] Günter Wallner and Simone Kriglstein. 2013. Visualization-based analysis of gameplay data - A review of literature. *Entertainment Computing* 4, 3 (2013), 143–155. https://doi.org/10.1016/j.entcom.2013.02.002
[52] Guenter Wallner and Simone Kriglstein. 2016. Visualizations for Retrospective Analysis of Battles in Team-based Combat Games: A User Study. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '16)*. ACM, New York, NY, USA, 22–32. https://doi.org/10.1145/2967934.2968093
[53] Günter Wallner, Simone Kriglstein, Florian Gnadlinger, Michael Heiml, and Jochen Kranzer. 2014. Game User Telemetry in Practice: A Case Study. In *Proceedings of the 11th Conference on Advances in Computer Entertainment Technology (ACE '14)*. ACM, New York, NY, USA, 45:1–45:4. https://doi.org/10.1145/2663806.2663859
[54] Veronica Zammitto, Pejman Mirza-Babaei, Ian Livingston, Marina Kobayashi, and Lennart E. Nacke. 2014. Player Experience: Mixed Methods and Reporting Results. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems*. ACM, New York, NY, USA, 147–150. https://doi.org/10.1145/2559206.2559239