ResearchGate

See discussions, stats, and author profiles for this publication at: http://www.researchgate.net/publication/280931694

The Platformer Experience Dataset

CONFERENCE PAPER · SEPTEMBER 2015

READS
127

4 AUTHORS:

Kostas Karpouzis
Institute of Communications and Computer …
167 PUBLICATIONS **1,398** CITATIONS

SEE PROFILE

Georgios Yannakakis
University of Malta
145 PUBLICATIONS **1,820** CITATIONS

SEE PROFILE

Noor Shaker
Aalborg University
25 PUBLICATIONS **148** CITATIONS

SEE PROFILE

Stylianos Asteriadis
Maastricht University
40 PUBLICATIONS **265** CITATIONS

SEE PROFILE

All in-text references underlined in blue are linked to publications on ResearchGate, letting you access and read them immediately.

Available from: Stylianos Asteriadis
Retrieved on: 11 November 2015

The Platformer Experience Dataset

Kostas Karpouzis
Institute of Communication and Computer Systems
National Technical University of Athens

Georgios N. Yannakakis
Institute of Digital Games
University of Malta

Noor Shaker
Center for Computer Games Research
IT University of Copenhagen

Stylianos Asteriadis
Department of Knowledge Engineering
University of Maastricht

***Abstract*—Player modeling and estimation of player experience have become very active research fields within affective computing, human computer interaction, and game artificial intelligence in recent years. For advancing our knowledge and understanding on player experience this paper introduces the Platformer Experience Dataset (PED) — the first open-access game experience corpus — that contains multiple modalities of user data of Super Mario Bros players. The open-access database aims to be used for player experience capture through context-based (i.e. game content), behavioral and visual recordings of platform game players. In addition, the database contains demographical data of the players and self-reported annotations of experience in two forms: ratings and ranks. PED opens up the way to desktop and console games that use video from webcameras and visual sensors and offer possibilities for holistic player experience modeling approaches that can, in turn, yield richer game personalization.**

I. Introduction

Player experience has been investigated from a number of different perspectives as it defines the most critical factor and ultimate goal of the game design and development process. Player experience defines a core topic within game studies, game design and human computer interaction, and game artificial intelligence [1], [2]. The successful design of player experience can make the difference between a game that engages players for extended periods of time and a game that contains content — whether that is mechanics, levels, maps or narrative plot points — that fails to elicit appropriate experience patterns and rich affective experiences to its users [3]. If player experience estimates are employed in the game design process — and with particular quantitative or qualitative mechanisms in mind — designers may have the opportunity to change critical game parameters that hamper player experience and iteratively provide a better game to their users. We follow the definition provided in [1] and, thus, view player experience as the "synthesis of affective patterns elicited and cognitive processes generated during gameplay".

Classifying the available approaches for player experience modeling (PEM) one can distinguish between top-down approaches that are based on or derived from a theoretical model of player experience — such as Calleja's player involvement model [4], Koster's theory of fun [5], Lazzaro's four fun factors [6], the notion of *magic circle* of Salen and Zimmerman [7] and Csikszentmihalyi's theory to *flow* [8] — and bottom-up methods that construct player experience models via machine learning relevant and representative data of player experience usually mapping between experience manifestations to experience annotations [2]. The first (*model-based* [2], [1]) may offer a solid theoretical framework for the synthesis of computational models of experience but requires validation within games (which is not often performed in such approaches). Moreover, most of the aforementioned theoretical approaches are criticized as not being derived from or tested on interactive media such as games and concepts such as flow can only encapsulate limited aspects of the multifaceted nature of player experience [4]. *Model-free* approaches [1] on the other hand are making use of machine learning algorithms and they exploit the availability of large game datasets and improved sensor technology for the automatic construction of the computational models; the availability and quality of data are critical factors for the performance and usefulness of the derived computational models. It is important to note that any method that makes use of both a theoretical model and data to support it can be characterized as a *hybrid* between model-based and model-free [2].

Further building on the vast potential of model-free player modeling, this paper introduces a database of 58 subjects created for the study of user experience and the advancement of user experience and user modeling research in games and beyond. The Platformer Experience Dataset (PED) is publicly available[1] and contains multiple modalities of user input from a game survey in which participants played the most popular (and arguably the most representative) linear platformer game: a variant of *Super Mario Bros* (Nintendo, 1985). In particular, the database contains visual information from the players (obtained though a high-definition camera), in-game behavioral data (logged during gameplay), game context (i.e. game level content) representation, self-reported experience via rating and ranking questionnaire schemes and player demographics. The experimental results already obtained on this dataset [9], [10], and briefly presented here, indicate that the visually extracted information and the gameplay features recorded are appropriate for multimodal player modeling and for testing player experience models using different sets of features and data representations.

A. Novelty of this Paper

This paper describes the *first* multimodal, large and *open-access* player experience dataset which contains synchronized affective and player behavior, that can be used for research

[1]PED is available at: http://game.edu.mt/PED/

on player experience modeling, player modeling, procedural content generation, multimodal interaction, affective computing, image analysis and machine learning — the introduction of such a dataset defines the core contribution of this paper. While open-access multimodal corpora exist in the literature (e.g. [11], [12]) none is extracted from in-game behavior or focuses on player experience. In addition a number of large game datasets are nowadays publicly available — e.g. the *StarCraft* (Blizzard Entertainment, 1998) dataset [13] — but these are solely constrained to player actions and in-game decisions which limit the capacity of any derived player models. Commercial tools are available for game data logging and processing within the game industry (e.g. see [14]) but none offers an open-source dataset along with it. Arguably, the key contribution of PED is the opportunity to investigate any relationships between multiple parameters of player models and input types including visual information (movement, expressivity and eye gaze), game behavior, preferences, prior experience and demographics. While the majority of player experience modeling studies focus on the relationship between in-game behavioral data and experience annotations (e.g. see [15], [16], [17], [18], [9] among others), the literature is sparse investigating the interplay between visual and in-game behavior of users and its association to self-reported experience and game content.

II. Background

This section provides the necessary background on player experience modeling based on objective, non-verbal, measures of player experience and focuses on approaches for behavioral and visual analysis of player experience.

A. Behavioral Analysis of Reported Player Experience

Admittedly, player experience modeling methods that rely on gameplay behavior are comparatively easy to deploy, since there is no need for additional hardware, other than what is needed to play the game. As a result, researchers can easily and quickly attract large numbers of players, compile a large number of game sessions, crowdsource player experience annotations and train generalisable models via machine learning. A popular and highly publicized example of this approach is the *Restaurant Game* [19], which generated more than 1,000 games in the first two weeks of its release and 10,000 crowdsourced games within three years. While crowdsourcing defines a powerful player experience modeling tool [16] it relies on self-reporting for capturing particular dimensions of player experience. Due to the numerous limitations of self-reporting, Yannakakis and Hallam [20] opted not to choose ratings but pairwise preferences, where players are asked to compare their player experience in terms of different dimensions between two or more levels of the game. A similar *preference learning* approach is also employed in [21] to learn deep physiological models of affect.

Besides the inherent subjectivity of linguistic information, where the language barrier also exists for players whose native language is different to that used in the questionnaire or during the interviews, self-reports can be either intrusive if they are provided during game sessions or suffer from post-experience effects [22] where episodic memory, learning or social, cognitive and personality factors may bias the players' responses.

B. Non-verbal Cues for Player Experience Modeling

Complementary to self-reports, a number of non-verbal cues from the players during gameplay can be exploited to enrich the feedback received from them regarding their player experience. In addition to approaches derived from product marketing [22], techniques from user experience with new media, virtual and augmented reality also lend themselves well to this end [23], [24], while human-human interaction may also offer alternative approaches [25]. Physiological signals are the modality of choice in most cases, since their interpretation is relatively straighforward and less subjective compared to other modalities [26]: besides [21], where the authors' test bed is a maze-like game, Tognetti et al. [15] investigated player experience in the context of a racing game, using a sensor worn on the player's hand, Drachen et al. [27] focused on first-person shooter games and Tijs' work [28] was based on a variant of Pac-Man.

While some of those sensors, also including accelerometers and tilt sensors, are easy to deploy, they provide large quantities of data (in some cases, more than 50 measurement per second of use), which require extensive preprocessing for data-noise removal. In addition, interpreting the resulting features is two-fold: one may choose to interpret them qualitatively or, alternatively, seek for particular movements and gestures with clear semantics. Expressivity features are low-level statistical indicators of the perceived qualities of movement and offer a possible solution to the issue of subjective interpretation. Hartmann first proposed expressivity features in the context of synthesizing avatar movement [29], while Caridakis [30] used them for visual analysis. In a nutshell, expressivity features are based on 2D or 3D information extracted either visually or via accelerometer sensors and map ranges of time-varying statistical features to qualitative movement characteristics, such as power, fluidity and overall activation. Although their interpretation is still subjective when compared to universal facial expressions, it has been shown that automatically estimated overall activation is correlated with human perception and, thus, can be used as a robust indicator for annotating human movement [31]. Other expressivity features also show strong correlation with measurements from human annotators [32], further illustrating the validity of this approach to quickly and robustly interpret large quantities of low-level data related to affect.

C. Visual analysis for Player Experience Modeling

The visual estimation of player experience has been mostly following studies in market or product research, with gaze fixations and saccades indicating the level of interest and engagement to the game's particular objects or features [33], [34], [35], [36]. Regarding affective analysis, not much work

has been done, mostly due to the lack of universally recognizable facial expressions during game play and the subjective meaning of universals (a smile may indicate irony, while raised eyebrows may correspond to anticipation of an "epic win") and lower-level facial cues (winks, head nods), which cannot be resolved without looking into the game context that triggered them.

Studies have shown that when one narrows down the context of interaction, useful information may be extracted regarding the users' cognitive state. For example, Asteriadis et al. [37] use a web camera mounted on top of the screen to estimate the user's head pose, eye gaze and engagement towards a reading interface in real time, while Peters et al. [38] extend the same concept towards shared attention to virtual objects with an embodied conversational agent. Eye gaze research has progressed a lot in the past years, thanks mostly to specialized hardware for eye tracking, but also to high quality, yet inexpensive web cameras which cater for real-time, software-based processing of visual streams. Henderson et al. [39] use a hardware gaze tracker to estimate the user's cognitive state (e.g., confusion, inattention) in the context of reading software, while Steichen et al. [40] study gaze features and cognitive abilities in information visualization and Palinko et al. [41] estimate cognitive load in a driving simulator.

III. Recording and Processing the Database

The Platformer Experience Dataset was recorded in the framework of the Siren project [46], during which we investigated the role of affect in serious games. Our initial objectives were to identify which facial expressions, facial cues and changes of body posture (see Fig. 1 and Fig. 2) can be observed during gameplay and whether those can be used to model player experience, along with the characteristics of the game level, player demographics and preferences and in synchrony with player behavior and performance. Since we also wanted to log the parameters used to generate each game level, as well as player behavior, we chose a public domain clone of Nintendo's classic platform game *Super Mario Bros* (Nintendo, 1985), named Infinite Mario Bros (IMB). Game level generation was based on the work of Shaker et al. [16] where hundreds of pairs of IMB game levels were analyzed in terms of self-reported affective states, game level size and features of the level which impact player experience the most.

In terms of technical equipment, we used a camera (Canon Legria S11) capable of recording high definition footage and good quality video in different lighting conditions, so as to minimize gain-related noise in the video signal when filming in environments with moderate lighting; the camera was placed behind the players' screen, so as to capture body and head movement, as well as facial expressivity. HD video is a welcome feature for visual analysis of facial expressions and, especially, eye gaze, since the eye region is far too small and noisy to robustly estimate eye gaze when using even high quality web cameras. Synchronization of the video stream and the data logged by the game was based on the characteristic

Fig. 1. Typical player responses to losing in IMB

Fig. 2. Typical player responses to hard situations in IMB

sound played by IMB when the level begins, which was captured by the camera.

Fifty eight volunteers participated in the recording sessions (28 male, with player age ranging from 22 to 48 years), which took place in Greece (with mostly Greek participants) and in Denmark (with participants of different ethnic backgrounds). The experimental protocol was based on [42] and started with a consent form, signed by the participants, informing them that they will be recorded during playing IMB and that the videos would be used for research purposes, but with no further information about the aims of the experiment, i.e. analysis of player experience, so as to not trigger unnatural behavior. Then, players were presented with a playable introductory scene, so as to acquaint them with the game, its objective and controls, the successive levels they would have to play and the player experience questionnaire they would have to fill in. They were also presented with a demographic data collection form, which collected information about their gender, age, frequency of playing games, hours spent on gameplay on a weekly basis and any previous experience with the particular game (see Fig. 3).

After this, players were presented with level A and three chances to finish it (see, for instance, Fig. 4). When players lost all three lives or completed the level, they were presented with a rating questionnaire which asked them to report their level of engagement, frustration and challenge in a scale between 0 to 4 (0 denoting "not at all'', 4 meaning "extremely''). The process was then repeated with a different game level (level B) and another rating questionnaire. After completing these two games, players were asked to report which of the two games they preferred for each of the three player experience dimensions, via a 4-alternative forced choice (4-AFC) questionnaire protocol: A over B; B over A; both equally engaging/frustrating/challenging; both equally not preferred. Finally, players were given the option to play more pairs of games (most of them did) or quit the game. These levels

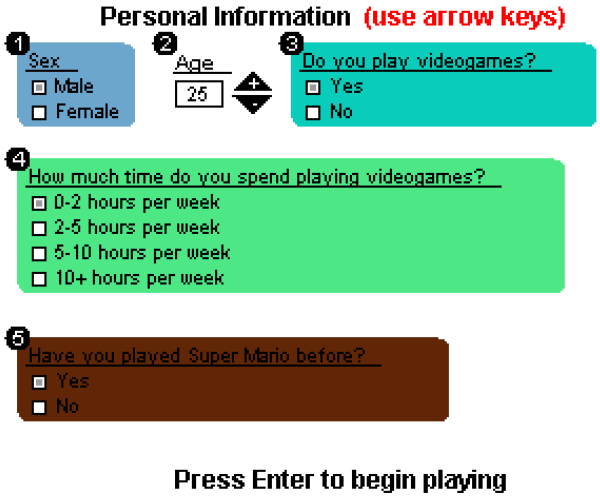

Fig. 3. The game screen collecting demographic and game-related information

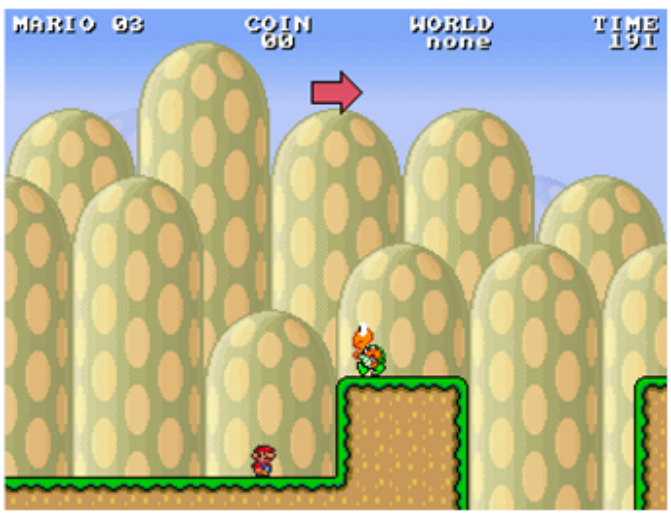

Fig. 4. A game level in Infinite Mario Bros

were sufficiently long to evoke affective experiences, but small enough to cater for possible adaptation and seamless procedural generation of the subsequent levels (size and positioning of gaps, number and placement of enemies, positioning of rewards), after they are completed.

Participants played a total of 380 games (more than 6 hours of recording in total). In most cases, they were isolated in the playing room, but even when that was not possible, distractions were observed only between (and not during) successive games, when they were still recorded but this visual data was subsequently removed. In some cases, game logging or visual feature extraction was not possible due to camera failures and freezing of the game, leaving us with 321 video-recoded game sessions in total.

A. Recorded Game Context and Player Behavior features

Since the game level generator is open source, we had the opportunity to choose which level generator parameters to log. After consulting game design experts and following earlier work [43], [16], we opted to control the number of gaps in the level and their average width, the number of enemies and their placement (around horizontal boxes, around gaps or random), the number of powerups and the number of boxes which make up platforms where bonuses (coins or powerups) may be hidden.

Regarding gameplay and player behaviour, the logged features are: level completion; Mario death and the cause of death; picking up blocks and bonuses as an absolute number and percentage of total; killing enemies; changing mode and time spent in small, big or fire mode; changing direction and time spent moving left, right, jumping, ducking and running; and the full trajectory of Mario as a combination of events. Specific visual features, other than the video stream, were not recorded, leaving the field open for researchers to investigate any visual manifestations of player experience or any affective characteristics which may be used for player modeling.

As mentioned before, PED offers researchers the opportunity to investigate visual behavior and player experience not only aggregated across a game level, but also at specific occasions within each level, associating expressed behavior with the particular game state. This, for instance, may be used to correlate a surprised face with the player finishing a particular stage for the first time or facing a seemingly impassable gap surrounded by enemies. In PED, video files are sorted in different folders, based on a player identification tag (e.g. “ama_16_03”). These folders contain video files, each of them corresponding to a game level played by the specific player (e.g. “1_A”, “1_B” for the first set of games, “2_A”, “2_B” if they played a second set, etc.), and associated with a CSV file containing the ID code of time-stamped game events. Table I shows the logged events and their ID codes.

Each AVI video has been processed using the software described in [37], providing visual features and eliminating the need for further processing to detect facial features. Detected features from each video are contained in a text file which contains the following tab-separated data: a timestamp (time from the beginning of the level); two 2-D vectors describing the player’s eye gaze and head pose horizontally and vertically; a float value corresponding to whether the player stands still, approaches the screen or moves back; head roll in degrees; and a list of prominent facial features around the player’s eyes (top, bottom, left, right corners), mouth (top, bottom, left, right corners), nose (nose tip), and eye brows (left corner, mid point, right corner).

Finally, each player folder also contains a CSV file containing player demographics and the player responses to the questionnaire about prior experience and another CSV file containing self-reported rating values for each level and self-reported preferences between the two levels of the same set. Possible rating values correspond to “Extremely”, “Fairly”, “Moderately”, “Slightly” and “Not at all”, while the preference value denotes which of the two levels was more engaging, challenging and frustrating. As noted earlier, players can choose between levels A and B or, alternatively, report no clear

TABLE I
THE LOGGED GAME EVENTS AND THEIR IDS CONTAINED IN THE CSV FILE OF THE PLATFORMER EXPERIENCE DATASET

Game event name	Game event ID
ARMORED TURTLE KILLSTOMP	311
JUMP FLOWER KILLSTOMP	411
CANNON BALL KILLSTOMP	511
CHOMP FLOWER KILLSTOMP	611
RED TURTLE UNLEASHED	021
GREEN TURTLE UNLEASHED	121
GOOMPA UNLEASHED	221
ARMORED TURTLE UNLEASHED	321
JUMP FLOWER UNLEASHED	421
CANNON BALL UNLEASHED	521
CHOMP FLOWER UNLEASHED	621
LITTLE START	020
LARGE START	120
FIRE START	220
LITTLE END	021
LARGE END	121
FIRE END	221
JUMP START	030
DUCK START	130
RUN START	230
LEFT MOVE START	330
RIGHT MOVE START	430
JUMP END	031
DUCK END	131
RUN END	231
LEFT MOVE END	331
RIGHT MOVE END	431

preference (levels are "Equally" ranked or "Neither" level is ranked).

B. Initial Analysis of PED

The PE dataset has been used in a number of studies which illustrate its usefulness in multimodal user modeling and user experience research, and prediction of player experience in the framework of experience-driven procedural content generation [1].

The purpose for the first analysis of the PE dataset was to investigate whether visually manifested patterns are correlated with player demographics (age, gender and familiarity with games) and preferences or performance [10]. Player modeling [2] has been mostly researched in terms of player behavior (not visual or affective behavior) in games where the player model, personality and player aims influence in-game decisions, given specific or free choices [44]; in general, player modeling and clustering may also be used to identify players of similar skill and have them play together in a multi-player game, or robustly estimate player experience and adapt the game accordingly. The study in [10], instead, mostly focuses on spontaneous movements, either over the whole game level — regardless of how this ends (i.e. the player finishes the level successfully or loses a life) — or towards the last seconds before a loss. In that study facial expressions (for example, indicating joy or disappointment [32]) were not investigated since their application in gaming contexts may be ambiguous: players may smile not only after completing a level, but also when losing, as an ironic response. In addition to this, after visually inspecting the videos, distinguishable expressions forming any pattern with respect to game events have not been noticed. The visual feature used in the analysis was a heuristic of head movement, defined as the 2D transformation of the position of the player's eyes within a bounding box around the player's head at any given frame. We chose to measure player performance based on time spent playing a level, since the number of collected rewards, which is an alternative measure of player performance, varies between player styles (some opt to complete the game level without collecting any). Player data was clustered [45] using visual and game play features: in the three player clusters that were identified, older players (average age: 33.9 y.o. ± 6.3) mostly belonged to one cluster (named Cluster 3), while players of Cluster 1 (demonstrating longer play times) had reported some experience with the particular game genre. Play time was a statistically significant factor (p=0.002) with 65% of players in Cluster 3 reporting that they do not play games at all, while 83.3% of players in Cluster 2 played games at least occasionally.

Regarding player experience, Shaker et al. [9] conducted a set of experiments on the PE dataset to test the hypothesis that the use of multimodal channels of information about player behavior would result in more accurate models of player experience than the ones constructed from a single input modality. A number of player experience models were constructed from four different subsets of the information extracted from the rich information provided in the dataset. The subsets were comprised of gameplay and content features; visual reaction features; mean head movement features; and a set of fused features of content, gameplay and visual reactions. The analysis showed differences among the emotional states investigated; while players' visual reactions fused against certain game events provided the richest source of information regarding reported preferences with respect to challenge and frustration (with accuracies up to 88.88% and 92.5%, respectively). On the other hand, engagement (predicted with models of up to 91.27% accuracy) seems to be a notion related to the way a game has designed, played, as well as to the visual information coming from the player. In general, the results suggest that combining the two modalities of gameplay and the visual reactions to specific game events would yield more accurate estimators of player experience than relying on a single input modality.

IV. CONCLUSIONS

This paper introduced a multimodal database, named the Platformer Experience Dataset, available for the study of player experience, computational and artificial intelligence in games, affective computing and affect modeling, and multimodal interaction. The dataset consists of visual data, player demographics, preferences, in-game behavior and parameters used to create game content. The corpus can be used to model

affective and game behavior of players, based on their profiles, predict their behavior and use constructed player experience models for the complete, or mixed-initiative generation of game levels that optimize aspects of player experience [1] for a single player or for a group of players with similar characteristics for cooperative or competitive play.

ACKNOWLEDGMENTS

This research was partially supported by the FP7 ICT project SIREN (project no: 258453). K. Karpouzis was funded by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program "Thales. Investing in knowledge society through the European Social Fund".

REFERENCES

[1] G. N. Yannakakis and J. Togelius, "Experience-driven procedural content generation," *Affective Computing, IEEE Transactions on*, vol. 2, no. 3, pp. 147–161, 2011.

[2] G. N. Yannakakis, P. Spronck, D. Loiacono, and E. André, "Player modeling," *Artificial and Computational Intelligence in Games*, vol. 6, pp. 45–59, 2013.

[3] G. N. Yannakakis and A. Paiva, *Handbook of Affective Computing*. Oxford Press, 2013, ch. Emotion in Games.

[4] G. Calleja, *In-Game: From immersion to incorporation*, 2011.

[5] R. Koster, *Theory of fun for game design*. O'Reilly Media, Inc., 2010.

[6] N. Lazzaro, "Why we play games: Four keys to more emotion without story," 2004.

[7] K. Salen and E. Zimmerman, *Rules of play: Game design fundamentals*. The MIT Press, 2004.

[8] M. Csikszentmihalyi, *Flow: The Psychology of Optimal Experience*. Harper Perennial, Mar. 1991.

[9] N. Shaker, S. Asteriadis, G. N. Yannakakis, and K. Karpouzis, "Fusing visual and behavioral cues for modeling user experience in games," *Cybernetics, IEEE Transactions on*, vol. 43, no. 6, pp. 1519–1531, 2013.

[10] S. Asteriadis, K. Karpouzis, N. Shaker, and G. N. Yannakakis, "Towards detecting clusters of players using visual and gameplay behavioral cues," *Procedia Computer Science*, vol. 15, pp. 140–147, 2012.

[11] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "Deap: A database for emotion analysis; using physiological signals," *Affective Computing, IEEE Transactions on*, vol. 3, no. 1, pp. 18–31, 2012.

[12] J. Fierrez, J. Ortega-Garcia, D. Torre Toledano, and J. Gonzalez-Rodriguez, "Biosec baseline corpus: A multimodal biometric database," *Pattern Recognition*, vol. 40, no. 4, pp. 1389–1392, 2007.

[13] B. G. Weber and M. Mateas, "A data mining approach to strategy prediction," in *Computational Intelligence and Games, 2009. CIG 2009. IEEE Symposium on*. IEEE, 2009, pp. 140–147.

[14] S. McCallum and J. Mackie, "Webtics: A web based telemetry and metrics system for small and medium games," in *Game Analytics*. Springer, 2013, pp. 169–193.

[15] S. Tognetti, M. Garbarino, A. Bonanno, and M. Matteucci, "Modeling enjoyment preference from physiological responses in a car racing game," *IEEE Transactions on Computational Intelligence and AI in Games*, 2010.

[16] N. Shaker, G. Yannakakis, and J. Togelius, "Crowdsourcing the aesthetics of platform games," *Computational Intelligence and AI in Games, IEEE Transactions on*, vol. 5, no. 3, pp. 276–290, 2013.

[17] H. Martinez and G. Yannakakis, "Mining multimodal sequential patterns: A case study on affect detection," in *Proceedings of the 13th International Conference in Multimodal Interaction, ICMI 2011, Alicante*. ACM Press, November 2011.

[18] P. Spronck, M. Ponsen, and E. Postma, "Adaptive game ai with dynamic scripting," in *Machine Learning*. Kluwer, 2006, pp. 217–248.

[19] J. Orkin and D. Roy, "Automatic learning and generation of social behavior from collective human gameplay," in *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems-Volume 1*. International Foundation for Autonomous Agents and Multiagent Systems, 2009, pp. 385–392.

[20] G. N. Yannakakis and J. Hallam, "Ranking vs. preference: a comparative study of self-reporting," in *Affective Computing and Intelligent Interaction*. Springer, 2011, pp. 437–446.

[21] H. P. Martinez, Y. Bengio, and G. N. Yannakakis, "Learning deep physiological models of affect," *Computational Intelligence Magazine, IEEE*, vol. 8, no. 2, pp. 20–33, 2013.

[22] A. L. McGill and D. Iacobucci, "The role of post-experience comparison standards in the evaluation of unfamiliar services," *Advances in Consumer Research*, vol. 19, no. 1, pp. 570–8, 1992.

[23] W. A. IJsselsteijn, H. de Ridder, J. Freeman, and S. E. Avons, "Presence: concept, determinants, and measurement," in *Electronic Imaging*. International Society for Optics and Photonics, 2000, pp. 520–529.

[24] C. Gale and A. F. Monk, "Where am i looking? the accuracy of video-mediated gaze awareness," *Perception & psychophysics*, vol. 62, no. 3, pp. 586–595, 2000.

[25] M. Argyle, L. Lefebvre, and M. Cook, "The meaning of five patterns of gaze," *European journal of social psychology*, vol. 4, no. 2, pp. 125–136, 1974.

[26] R. L. Mandryk and M. S. Atkins, "A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies," *International Journal of Human-Computer Studies*, vol. 65, no. 4, pp. 329–347, 2007.

[27] A. Drachen, L. E. Nacke, G. Yannakakis, and A. L. Pedersen, "Correlation between heart rate, electrodermal activity and player experience in first-person shooter games," in *Proceedings of the 5th ACM SIGGRAPH Symposium on Video Games*, ser. Sandbox '10. New York, NY, USA: ACM, 2010, pp. 49–54.

[28] T. J. Tijs, D. Brokken, and W. A. IJsselsteijn, "Dynamic game balancing by recognizing affect," in *Fun and Games*. Springer, 2008, pp. 88–93.

[29] B. Hartmann, M. Mancini, and C. Pelachaud, "Implementing expressive gesture synthesis for embodied conversational agents," in *gesture in human-Computer Interaction and Simulation*. Springer, 2006, pp. 188–199.

[30] G. Caridakis, J. Wagner, A. Raouzaiou, F. Lingenfelser, K. Karpouzis, and E. Andre, "A cross-cultural, multimodal, affective corpus for gesture expressivity analysis," *Journal on Multimodal User Interfaces*, pp. 1–14, 2013.

[31] J.-C. Martin, G. Caridakis, L. Devillers, K. Karpouzis, and S. Abrilian, "Manual annotation and automatic image processing of multimodal emotional behaviors: validating the annotation of tv interviews," *Personal and Ubiquitous Computing*, vol. 13, no. 1, pp. 69–76, 2009.

[32] K. Karpouzis, G. Caridakis, R. Cowie, and E. Douglas-Cowie, "Induction, recording and recognition of natural emotions from facial expressions and speech prosody," *Journal on Multimodal User Interfaces*, pp. 1–12, 2013.

[33] L. Nacke, A. Drachen, and S. Göbel, "Methods for evaluating gameplay experience in a serious gaming context," *International Journal of Computer Science in Sport*, vol. 9, no. 2, pp. 1–12, 2010.

[34] S. Stellmach, L. Nacke, and R. Dachselt, "Advanced gaze visualizations for three-dimensional virtual environments," in *Proceedings of the 2010 symposium on eye-tracking research & Applications*. ACM, 2010, pp. 109–112.

[35] J. Broekens, W. Kosters, and T. de Vries, "Eye movements disclose decisions in set." BNAIC, 2009.

[36] S. Asteriadis, K. Karpouzis, and S. Kollias, "Visual focus of attention in non-calibrated environments using gaze estimation," *International Journal of Computer Vision*, vol. 107, no. 3, pp. 293–316, 2014.

[37] S. Asteriadis, P. Tzouveli, K. Karpouzis, and S. Kollias, "Estimation of behavioral user state based on eye gaze and head pose application in an e-learning environment," *Multimedia Tools and Applications*, vol. 41, no. 3, pp. 469–493, 2009.

[38] C. Peters, S. Asteriadis, and K. Karpouzis, "Investigating shared attention with a virtual agent using a gaze-based interface," *Journal on Multimodal User Interfaces*, vol. 3, no. 1-2, pp. 119–130, 2010.

[39] J. M. Henderson, S. V. Shinkareva, J. Wang, S. G. Luke, and J. Olejarczyk, "Predicting cognitive state from eye movements," *PloS one*, vol. 8, no. 5, p. e64937, 2013.

[40] B. Steichen, G. Carenini, and C. Conati, "User-adaptive information visualization: using eye gaze data to infer visualization tasks and user

cognitive abilities," in *Proceedings of the 2013 international conference on Intelligent user interfaces*. ACM, 2013, pp. 317–328.

[41] O. Palinko, A. L. Kun, A. Shyrokov, and P. Heeman, "Estimating cognitive load using remote eye tracking in a driving simulator," in *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*. ACM, 2010, pp. 141–144.

[42] G. N. Yannakakis, M. Maragoudakis, and J. Hallam, "Preference learning for cognitive modeling: a case study on entertainment preferences," *Trans. Sys. Man Cyber. Part A, IEEE Transactions on*, vol. 39, pp. 1165–1175, November 2009.

[43] C. Pedersen, J. Togelius, and G. N. Yannakakis, "Modeling player experience in super mario bros," in *CIG'09: Proceedings of the 5th international conference on Computational Intelligence and Games*. Piscataway, NJ, USA: IEEE Press, 2009, pp. 132–139.

[44] R. A. Bartle, *Designing virtual worlds*. New Riders, 2004.

[45] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, no. 2, pp. 224–227, 1979.

[46] G.N Yannakakis, J. Togelius, R. Khaled, A. Jhala, K. Karpouzis, A. Paiva, A. Vasalou, "Siren: Towards adaptive serious games for teaching conflict resolution," in *Proceedings of ECGBL*, 2010.