Fusing Visual and Behavioral Cues for Modeling User Experience in Games

Noor Shaker, Stylianos Asteriadis, Georgios N. Yannakakis, *Member, IEEE,* and Kostas Karpouzis, *Senior Member, IEEE*

***Abstract*—Estimating affective and cognitive states in conditions of rich human-computer interaction, such as in games, is a field of growing academic and commercial interest. Entertainment and serious games can benefit from recent advances in the field as, having access to predictors of the current state of the player (or learner) can provide useful information for feeding adaptation mechanisms that aim to maximize engagement or learning effects. In this paper, we introduce a large data corpus derived from 58 participants that play the popular Super Mario Bros platform game and attempt to create accurate models of player experience for this game genre. Within the view of the current research, features extracted both from player gameplay behavior and game levels, and player visual characteristics have been used as potential indicators of reported affect expressed as pairwise preferences between different game sessions. Using neuroevolutionary preference learning and automatic feature selection, highly accurate models of reported engagement, frustration, and challenge are constructed (model accuracies reach 91%, 92%, and 88% for engagement, frustration, and challenge, respectively). As a step further, the derived player experience models can be used to personalize the game level to desired levels of engagement, frustration, and challenge as game content is mapped to player experience through the behavioral and expressivity patterns of each player.**

***Index Terms*—Content personalization, experience-driven procedural content generation, multimodal interaction, player experience modeling, visual cues.**

I. Introduction

VIDEO GAMES is a flourishing industry for more than three decades now, with revenues surpassing even those of the movie and music industries [1]. Due to their high popularity and huge computational demands, video games would always introduce leading technologies and pioneering methods in the field of human–computer interaction (HCI) at large. Today's technologies have reached a point where new addons can boost the gameplay experience, altering and guiding game content and evolution following affect-dependent strategies [2], [3]. To this aim, using context and behavior-related parameters to elicit information regarding the player's current state (and, consequently, obtain hints about her/his needs regarding interaction) is of primary importance for constructing personal behavioral and interaction-related models and guiding game adaptation in order to achieve maximum engagement [4] or possibly enable conditions of flow [5] and incorporation [6] and, ultimately, realize the affective loop [7] in games.

There is an abundance of studies presented in bibliography, dealing with the problem of user state estimation during HCI. Recent advances on computer vision techniques under uncontrolled conditions have allowed the proposal of techniques incorporating notions, such as body and head movements [4], eye gaze (with eye gaze usually necessitating specialized hardware, such as infrared eye trackers [8]), and facial expressions [9]. Typical works are those reported in [10] and [11], where the authors use Bayesian networking on gaze, postural, and contextual data for detecting user engagement with a robot companion [12] posing various expressions. In the domain of games, the increased diversification of human playing demographics, strategies, needs, skills, and preferences has increased the importance of experience personalization. Player experience modeling [3] studies that rely on single or multiple modalities of user input ([13]–[19] among many) have provided some initial benchmark solutions toward achieving such a goal.

Physiological signals are a popular modality in this framework; however, measuring affect using most physiological signals usually requires specialized hardware, which is often expensive, hard to calibrate and may result in cumbersome settings, which hamper interaction. As a result, related approaches may be efficient in terms of recognizing player affect but are extremely problematic to deploy in mass scales and for commercial uses. On the other hand, affect estimation approaches based on processing acceleration data, typically from mobile phones or accelerometer-equipped controllers (e.g., Nintendo's Wii-mote) or video sequences taken from low-end cameras (e.g. cameras mounted on top of the users' screen or Kinect sensors, typically sold for Microsoft's Xbox 360 platforms, but available for desktop computers, as well) use hardware that most gamers already possess and do not

Manuscript received April 9, 2012; revised November 23, 2012 and May 4, 2013; accepted June 13, 2013. Date of current version November 18, 2013. This work was supported in part by the FP7 ICT SIREN Project 258453, and by the European Union Social Fund (ESF). This work was also supported by the Greek National Funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF), Research Funding Program THALES. Recommended by Associate Editor S. Zafeiriou.

N. Shaker is with the Center for Computer Games Research, IT University of Copenhagen, Copenhagen 2300, Denmark (e-mail: noor.shaker@gmail.com).

S. Asteriadis is with the Informatics and Telematics Institute, Centre for Research and Technology Hellas, Thessaloniki, GR-57001, Greece (e-mail: stiast@image.ntua.gr).

G. N. Yannakakis is with the Department of Digital Games, University of Malta, Msida MSD 2080, Malta (e-mail: yannakakis@itu.dk).

K. Karpouzis is with the National Technical University of Athens, Zographou 15780, Greece (e-mail: kkarpou@image.ntua.gr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCYB.2013.2271738

Fig. 1. Snapshot from Infinite Mario Bros, showing Mario standing on horizontally placed boxes surrounded by different types of enemies.

impose any additional requirements, such as moving in confined spaces, since gamers carry controllers with them and do not usually move away from their screen or TV while playing. Buttussi [20] uses acceleration features to deduce motions and actions, besides physiological, in the framework of a fitness game, while Istance [21] and Nacke [22] use eye-gaze as a means of alternative game control. One of the issues of such approaches is what Almeida [23] refers to as the Midas touch problem, where eye gaze vectors are constantly used to issue commands, regardless of whether the user actually intends to do so or merely looks around at the game interface or is producing irrelevant fixations and saccades. To overcome this, several researchers focus on gamer attention and engagement, as a higher-level cognitive concept based on gaze. Seif El-Nasr [24] uses a commercial head-mounted eye tracker to identify points on a computer screen and then objects in the game world that attract the user's attention, while Sundstedt, Isokoski [25] and Smith [26] use eye gaze to control virtual and game characters. However, these approaches lie in-between those described before and a completely low-cost approach, since they do rely on visual features but require dedicated eye-tracking hardware to produce them. Kaiser and Wehrle [27] do rely on automatic visual estimation but concentrate on emotion labels, in order to produce an emotion-rich corpus, and do not delve into game-related concepts, such as flow and incorporation, nor do they attempt to adapt the game experience and close the affective loop based on the estimated user state.

Another direction that has received increasing attention is the procedural generation of content [2]. Artificial and computational intelligence methods have been used to generate different aspects of content with or without human interference [28]–[32]. The creation of personalized content for either the player or the designer [14], [28], [33]–[35] already shapes a leading research direction within procedural content generation (PCG). The first step toward creating personalized content is to effectively model the relationship between player experience and content. This can be achieved by constructing models on data collected throughout the interaction between the user and the digital content via the annotation of content with user experience tags [3].

Building on the experience-driven procedural content generation [3] framework, the presented work employs a fusion scheme of game-content parameters, game-performance indicators, and a series of visual features from the player's head in order to predict player preferences between different game variants. A large data corpus of behavioral and visual cues as well as game context and subjective experience annotations is collected from 58 users while playing variants of the popular Super Mario Bros platform game. Player subjective reports are identified via comparative questionnaires and different game variants are ranked with respect to frustration, engagement, and challenge. A coupling of automatic feature selection and neuroevolutionary preference learning is employed to select a subset of appropriate features that yield accurate predictors of the reported affect. Results show that highly accurate player experience models can be constructed as accuracies reach 91%, 92%, and 88% for engagement, frustration, and challenge, respectively. The models are used to generate a sample of maximally engaging, frustrating, and challenging levels for a number of players derived from our data corpus. The generated levels showcase the robustness of the algorithm and the personalization achieved in level design.

This paper builds on the authors' earlier study [36] and advances the current state-of-the-art in dissimilar ways. First, an extensive corpus of visual and behavioral data is used for the analysis of the cognitive state and behavior of the player; second, behavioral and visual cues are fused for the prediction of player experience in a single player game, producing concepts related to the gaming paradigm and moving forward from shallow emotional states by relating user states to particular in-game events; third, personalized levels are generated that potentially yield maximally engaging, frustrating, and challenging levels for a player; fourth, for the first time, procedural content generation is driven by computational models of fused modalities of player input.

The structure of this paper is the following. Section II describes the game platform used and the data collection strategy followed. Section III describes the gameplay and motion analysis features that have been considered for player experience model construction. Section IV introduces the methods that have been implemented to map player experience to reported affect. Section V gives the experimental results regarding player state prediction, while Section VII concludes the paper.

II. Dataset

This section presents the test-bed game used for data harvesting and the adopted protocol of the data collection experiment.

A. *Testbed Platform Game*

The testbed platform game used for our study is a modified version of Markus Persson's Infinite Mario Bros (see Fig. 1), which is a public domain clone of Nintendo's classic platform game Super Mario Bros. The original Infinite Mario Bros and its source code is available on the web.[1]

The gameplay in Super Mario Bros consists of moving the player-controlled character, Mario, through 2-D levels. Mario can walk, run, duck, jump, and shoot fireballs. The main goal of each level is to get to the end of the level. Auxiliary goals include collecting as many coins as possible, and clearing the level as fast as possible.

While implementing most features of Super Mario Bros, the standout feature of Infinite Mario Bros is the automatic generation of levels. Every time a new game is started, levels are randomly generated. In our modified version, we

[1] http://www.mojang.com/notch/mario/

concentrated on a few selected parameters that affect gameplay experience.

B. Dataset Design

To assess the players' affective state during play, the following experiment protocol was designed. We seated 58 volunteers (28 male; player age varied from 22 to 48 years) in front of a computer screen for video recording. Experiments were carried out in Greece and Denmark. Lighting conditions were typical of an office environment, and for capturing players' visual behavior, a high definition camera (Canon Legria S11) was used.

We designed a postexperience game survey to collect subjective affective reports expressed as pairwise preferences of subjects playing different variants (levels) of the test-bed game by following the experimental protocol proposed in [37]. The detailed description of the procedure followed is described here:

1) An introduction scene presents the game to the player and contains information about the procedure that will be followed. The player is being told that during the session she will play two short games and will be asked to answer a few questions about her game experience.
2) Then, a demographics questionnaire is presented, used to collect the following data: age, whether the player is a frequent gamer, how much time she spends playing games on a weekly basis (0, 1, 2, or more than 3 hours per week), and whether she had played Super Mario Bros before.
3) The player is introduced to the keys that can be used to control Mario.
4) The player is then informed that her game sessions will be video recorded and analyzed.
5) After these introductory steps the player is set to play the first game (game A). The player is given three chances to complete the short game level of Super Mario Bros. If she fails in the first trial the game is reset to the starting point and the player is set to try again. The game ends either by winning one of the three trials or by failing the third one.
6) After finishing game A, a Likert questionnaire scheme is presented to the player [38]. The player is asked to express her emotional preferences of the played game across the three different emotional states (engagement, frustration, and challenge). The questionnaire is inspired by the game experience questionnaire (GEQ), according to which a likert scale from 0 to 4 represents the strength of the emotion (four means extremely; 0 means not at all).
7) A second short game (game B) is then presented to the player and she is set to play. The player is given three chances (i.e., Mario lives) to complete the level and the same rules apply as in game A.
8) After finishing game B, the GEQ questionnaire is presented to the player (as in game A).
9) After completing a pair of two games A and B, the player is asked to report the preferred game for the three emotional dimensions through a four-alternative forced choice (4-AFC) questionnaire protocol (i.e., A is preferred to B, B is preferred to A, both are preferred equally, neither is preferred (both are equally not-preferred)) [39]. Please note that the questionnaire presented to the players is the following: Which game was more x, where x is one of the three emotional states under investigation.
10) The player then has the choice to either end the session or to continue. In the latter case, a new pair of two games is presented and the procedure is repeated.

Each participant played from two to five pairs of games on average, resulting to a total of 380 games (more than 6 hours of recordings). In most cases, players were left alone in the rooms they were playing and, whenever this was not possible, everyone was asked not to distract them. The game sessions presented to players have been constructed using a level width of 100 Super Mario Bros units (blocks), about one-third of the size usually employed when generating levels for Super Mario Bros game in previous experiments [40], [41]. The selection of this length was due to a compromise between a window size that is big enough to allow sufficient interaction between the player and the game to trigger the examined affective states and a window, which is small enough to set an acceptable frequency of an adaptation mechanism applied in real-time aiming at closing the affective loop of the game [7].

After removing interaction session instances for which visual data was corrupted the full dataset considered in this paper consists of 167 pairs of games. In addition, a preprocessing step was applied to remove the game pairs for which players reported unclear preferences (those that were equally preferred or equally not-preferred). After this step 127, 121, and 144 game pairs remain for engagement, frustration, and challenge, respectively. Those game pairs are used to train models of player experience based on clear reported preferences as described in Section V.

III. Feature Extraction

The following subsections describe the features that have been extracted and used in this paper as predictors of reported experience. This includes game level (content) features, gameplay behavioral features, and head movement features. The section ends with the description of the player experience annotations.

A. Content Features

The level generator of the game has been modified to create levels according to the following six controllable (game content) features:

1) the number of gaps in the level, G;
2) the average width of gaps, $\bar{G}_w$;
3) the number of enemies, E; this parameter controls the number of goompas and turtles scattered around the level, changing the level difficulty;
4) enemies placement, E_p; the way enemies ares placed around the level is determined by three probabilities, which sum to one.
   a) Around horizontal boxes, P_b: Enemies are placed on or under a set of horizontal blocks (a number of blocks placed horizontally without connection to the ground).
   b) Around gaps, P_g: Enemies are placed within a close distance to the edge of a gap.

Fig. 2. Enemies placement using different probabilities. High probability is given to placement around horizontal boxes. (a) P_b. (b) Around gaps, P_g. (c) Random placement, P_r.

c) Random placement, P_r: Enemies are placed on a flat space on the ground.

Fig. 2 illustrates positioned enemies by giving different values for P_b, P_g, and P_r. Fig. 2(a) of the figure shows enemies placed by setting P_b to 80%. Fig. 2(b) illustrates the result of setting P_g to 80%. Fig. 2(c) is the result of $P_r = 80\%$.

5) The number of powerups, N_w. Mario can collect powerup elements hidden in boxes to upgrade his state from little to big or from big to fire.
6) The number of boxes, B. We define one variable to specify the number of the two different types of boxes that exist in Super Mario. We call these two types blocks and rocks. Blocks contain hidden elements, such as coins or powerups. Rocks may hide a coin, a powerup or they can be empty. Mario can smash rocks only when he is in big mode.

According to the methodology presented in [41], two states (low and high) are set for each of the controllable parameters above except for enemies placement, which has been assigned three different states allowing more control over the difficulty and diversity of the generated levels.

The selection of these particular controllable features was made after consulting game design experts, and with the intent to cover the features that have the most impact on the investigated affective states [40], [41].

B. Gameplay Features

While playing the game, different player actions and interactions with game items and their corresponding time-stamps have been recorded. These events are categorized in different groups according to the type of the event and the type of interaction with the game objects. The events recorded are the following: level completion event; Mario death event and cause of death; interaction events with games items, such as free coins, empty rock, coin block/rock and power-up rock/block; Mario enemy kill event associated with the type of actions performed to kill the enemy and the type of enemy; changing Mario mode (small, big or fire) event; changing Mario state (moving right, left, jump, run, duck) event; and the full trajectory of Mario as a combination of events.

Several features have been directly extracted from the data recorded. Most of these features appear in our previous studies [40]–[42] and their selection is made in order to be able to represent the difference between a large variety of Super Mario Bros playing styles. The full list of gameplay features is presented in Table I.

C. Head Movement Features

In our experiments, as subjects were seated in front of a computer monitor, the upper part of their body was monitored by a camera, while head motion was of particular importance for creating behavioral correlations to game events and levels of difficulty. Being in line with Csikszentmihalyi's flow theory [43], visual features related to arousal were searched for [44] and combined with expressed player states and experience reports. It was noticed that head movement was of primary importance, and different patterns of motion were correlated to different player states and preferences (Figs. 3 and 4). For example, frustration was observed to be linked to sudden and very quick head movements while low levels of challenge would normally be associated with smoother movements, probably due to lack of high interest [36]. For the above reasons, in this paper we examine the relation of a series of head movement features [4], [45] with experience models, along with gameplay and content features. In particular, we track player's head motion through head horizontal and vertical (yaw and pitch) rotational movements. These are extracted using the method proposed in [4], due to its efficiency in terms of computational complexity, accuracy, and robustness to various lighting conditions and spontaneous movements. The values of the extracted features are considered, both throughout whole game sessions (mean head movement features) and during small periods of critical events (visual reaction features).

1) *Mean Head Movement Features:* As head movement, here, we considered the first derivative of the norm of the head pose vector [4] and use the average (Avg) of its absolute values throughout whole game sessions. A series of further head movement features [45] have also been considered in order to elicit emotional information of the player during each game session (Mean Head Movement Features). More specifically, we considered:

a) Overall Activation (OA), which comes as the sum of quantities of motion [45] for each rotational movement, separately. In other words, OA stands for the quantity of movement during certain periods of time. Let H be a sequence of head pose cues for the corresponding session, consisting of T frames, as in 1

$$H = [(y_1^H, p_1^H), (y_2^H, p_2^H), \ldots, (y_T^H, p_T^H)] \tag{1}$$

where y_i^H, p_i^H are the absolute yaw and pitch angles, respectively. Head Pose Overall Activation for sequence H is

$$OA = \sum_{i=1}^{T}(dYaw + dPitch) \tag{2}$$

with

$$dYaw = \frac{dy}{dt} \tag{3}$$

and

$$dPitch = \frac{dp}{dt} \tag{4}$$

b) Temporal Expressivity (TE) parameter, which denotes the speed of movement and dissociates fast from slow head gestures, is the average of OA during periods T.
c) Spatial Extent (SE) parameter is considered as the maximum value of the instantaneous expansion of head from a frontally posed position (y=p=0).

TABLE I
GAMEPLAY AND CONTENT FEATURES EXTRACTED FROM DATA RECORDED

Category	Feature	Description
		Content (Level) Features
Content (Level) Features	G	Number of gaps
	$\bar{G}_w$	Average width of gaps
	E	Number of enemies
	E_p	Placement of enemies
	N_w	Number of powerups
	B	Number of boxes
		GamePlay Features
Time	t_{comp}	Completion time
	$t_{lastLift}$	Playing duration of last life over total time spent on the level
	t_{duck}	Time spent ducking (%)
	t_{jump}	Time spent jumping (%)
	t_{left}	Time spent moving left (%)
	t_{right}	Time spent moving right (%)
	t_{run}	Time spent running (%)
	t_{small}	Time spent in Small Mario mode (%)
	t_{big}	Time spent in Big Mario mode (%)
Interaction with items	n_{coins}	Free coins collected (%)
	$n_{coinBlocks}$	Coin blocks pressed or coin rocks destroyed (%)
	$n_{powerups}$	Powerups pressed (%)
	n_{boxes}	Sum of all blocks and rocks pressed or destroyed (%)
Interaction with enemies	$k_{cannonFlower}$	Times the player kills a cannonball or a flower (%)
	$k_{goombaKoopa}$	Times the player kills a goomba or a koopa (%)
	k_{stomp}	Opponents died from stomping (%)
	$k_{unleash}$	Opponents died from unleashing a turtle shell (%)
Death	d_{total}	Total number of deaths
	d_{cause}	Cause of the last death
Miscellaneous	n_{mode}	Number of times the player shifted the mode between: Small, Big, and Fire
	n_{jump}	Number of times the jump button was pressed
	n_{gJump}	Difference between the # of gaps and the # of jumps
	n_{duck}	Number of times the duck button was pressed
	n_{state}	Number of times the player changed the state between: standing still, run, jump, moving left, and moving right

d) Energy Expressivity parameter (Power) of head movement(PO) during the stroke phase of the head gesture. Head gestures (similar to hand gestures) are considered to constitute of three phases, namely, the preparation, stroke, and withdrawal. The message is primarily conveyed during the stroke phase, while the phases of preparation and withdrawal occur while the head moves from and to its neutral position, respectively. The formalization of this parameter, according to this definition, however, is far from trivial, since the automatic detection of these stages is quite a challenging task. Alternatively, we opted to associate this parameter qualitatively, with the first derivative of speed (acceleration), during certain periods of time (5)

$$PO = \frac{\sum_{i=1}^{T}(\frac{d^2 y_i}{dt^2} + \frac{d^2 p_i}{dt^2})}{T}. \quad (5)$$

e) Fluidity of head movement (FL) distinguishes between smooth and abrupt movements. Under this prism, the variation of speed was considered for the two components of head pose used in this paper. This concept attempts to denote continuity of movements, regardless of the magnitude of speed. Equation 6 shows the calculation of the fluidity parameter

$$FL = \frac{var(dYaw) + var(dPitch)}{2}. \quad (6)$$

The reader is prompted to note that the above quantity takes high values for periods of time containing abrupt/sudden/unforseen movements, while small values are considered for gestures of higher continuity.

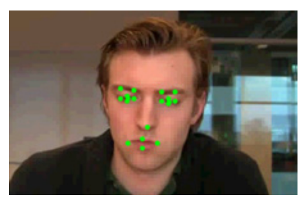

Fig. 3. Facial feature tracking for head movement features extraction.

The detailed list of extracted head movement features is presented in Table II. For more details regarding the extraction of the above criteria, please refer to [45]. In this paper, in addition to the above features, the median values of horizontal, $M_{horizontal}$, and vertical, $M_{vertical}$ rotations, as well as medians of head rotation norms M are also considered.

2) *Visual Reaction Features:* As players' expressivity appears to increase during certain events, we also considered the above features for certain gameplay events as described below.

a) when the player loses a life;
b) when the player kills an enemy by stomping on it;
c) when the player starts or ends a critical move: jump, duck, run, and move left or right;
d) when the player interacts with an object.

These features are calculated for periods of ten frames before and after the corresponding events. Subsequently, their mean values were compared to the corresponding average values (by calculating fractions) during normal gameplay, for each game session separately. A detailed list of the features used can be seen in Table II.

TABLE II

HEAD MOVEMENT FEATURES. MEAN HEAD MOVEMENT FEATURES EXTRACTED THROUGHOUT WHOLE SESSIONS AND VISUAL REACTION FEATURES DURING GAMEPLAY EVENT HAVE BEEN PRESENTED. THE GAMEPLAY EVENTS CONSIDERED INCLUDE: LOSING, STOMPING, (START/END) JUMPING, DUCKING, RUNNING LEFT, RUNNING RIGHT, AND INTERACTING WITH ITEMS

Category	Feature	Description
		Head Movement Features throughout whole sessions
Mean Head Movement	Avg	Absolute first order derivative of Head Pose Vector
	OA	Overall Activation
	SE	Spatial Extent
	TE	Temporal Expressivity parameter
	PO	Energy Expressivity parameter
	FL	Fluidity
	$M_{horizontal}$	Median value for horizontal head rotation
	$M_{vertical}$	Median value for vertical head rotation
		Head Movement Features during gameplay events
Visual Reaction Features	Avg_a	Absolute first order derivative of Head Pose Vector when the gameplay event, a occur
	OA_a	Overall Activation when the gameplay event, a occur
	SE_a	Spatial Extent when the gameplay event, a occur
	TE_a	Temporal Expressivity parameter when the gameplay event, a occur
	PO_a	Energy Expressivity parameter when the gameplay event, a occur
	FL_a	Fluidity when the gameplay event, a occur
	M_a	Median value for head rotation norm when the gameplay event, a occur

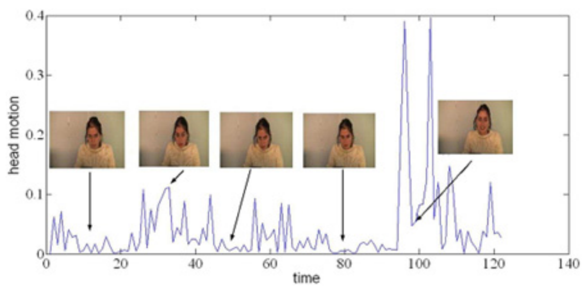

Fig. 4. Typical head expressivity of player reacting to certain game events.

D. Player Experience

As mentioned earlier, player experience is measured through four-alternative forced choice questionnaires, presented to the player after playing a pair of games generated by a different set of controllable feature values. The questionnaire asks the player to report the preferred game for three user states: engagement, challenge and frustration. The selection of these states is based on earlier game survey studies [40] and our intention to capture both affective and cognitive/behavioral components of gameplay experience [3]. Moreover, we want to keep the self-reporting as minimal as possible so that experience disruption is minimized. Pairwise preferences have been adopted for this paper because of their numerous advantages over rating-based questionnaires: a recent comparative study among the two schemes [46] shows that rating yields significant order and inconsistency effects as it is biased by a number of factors including personality and culture.

IV. PREFERENCE LEARNING FOR MODELING PLAYING EXPERIENCE

Neuroevolutionary preference learning [47], [37] has been used to construct models that approximate the function between gameplay, head movement features, content features, and reported affective preferences. In neuroevolutionary preference learning, a genetic algorithm (GA) evolves an artificial neural network (ANN) so that its output matches the pairwise preferences in the data set. The input of the ANN is a set of features that have been extracted from the data set. The GA implemented uses a fitness function that measures the difference between the reported emotional preferences and the relative magnitude of the model output. A sigmoid-based fitness function has been adopted as its shape has been optimized for maximum model performance in earlier studies [48], [37].

All features extracted are uniformly normalized to [0, 1] using standard max-min normalization. After normalization, these values are used as inputs for feature selection and ANN model optimization. Our modeling approach contains three following steps (Fig. 5):

1) Feature selection: We use sequential forward selection (SFS) [49] to select the relevant subset of features for predicting each emotional state [37]; this is achieved by training single-layer perceptrons (SLPs) as a mapping between selected features and reported preferences. SFS is a bottom-up approach where a feature is chosen to be added to the current set of selected features, such as the new subset of features yields a maximum possible performance. The quality of a feature subset is determined by three-fold cross-validation on unseen data.
2) Feature space expansion: The feature subset derived from the first phase is used as the input to small multilayer perceptron (MLP) models of one two-neuron hidden layer and SFS selects additional features from the remaining set of features during the training of these small MLPs.
3) Optimizing topology: In the last phase of the modeling process, the topology of the MLP models is optimized using neuroevolutionary preference learning. The network topology optimization process starts with a small two hidden-neuron MLP and the network topology gradually increases up to two hidden layers consisting of ten hidden neurons each.

The quality of a feature subset and the performance of each MLP is obtained through the average classification accuracy in three independent runs using three-fold cross validation across ten evolutionary trials. Parameter tuning tests have been conducted to set up the parameters' values for neuroevolutionary user preference learning that yield the highest accuracy and minimize computational effort. As a result of this parameter tuning process, we use a population of 100 individuals and

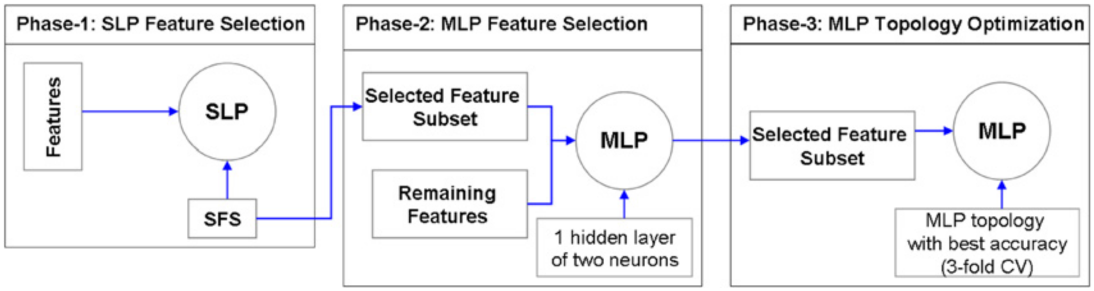

Fig. 5. Three-phase player experience modeling approach followed.

we run evolution for 20 generations. A probabilistic rank-based selection scheme is used, with higher ranked individuals having higher probability of being chosen as parents. Finally, reproduction is performed via uniform crossover, followed by Gaussian mutation of 1% probability.

V. Experiments and Analysis

The following sections describe the experiments that have been conducted to construct and compare different models of player experience derived from the features extracted (as described in the previous sections). We construct models based on gameplay and content features only, models from mean head movement features only and models from visual reaction features. We then investigate models constructed from fusing different modalities of player input.

We start by analyzing the features selected and the models' accuracies obtained from each feature set, then we further investigate the differences on significance between the models constructed on the different categories of features.

A. Player Experience Modeling Through Gameplay and Content Features

Modeling player experience from gameplay and content highlight important aspects of player behavior and game design that have strong impact on the gameplay experience. For this purpose, all features presented in Table I are set as inputs for feature selection and model optimization. The subsets of features selected, the models' accuracies and the best MLP topologies obtained vary across the three emotional states under investigation as can be seen in Table III. By constructing models based only on gameplay and content features, we are able to predict the three affective states with average accuracies (across 20 trails) higher than 72% while the best performances obtained exceed 89% for engagement and frustration. The best accuracy obtained for predicting challenge is 80.6%, which is significantly lower than the ones obtained for predicting engagement and frustration (significance is set to 1% in this paper).

It is worth observing that out of 30 different gameplay and content features, a maximum of five features only have been considered to be important for predicting each affective state. However, different feature subsets have been picked for each emotional state with only one common feature between engagement and challenge, namely, the time spent jumping t_{jump}. Three out of the six controllable features appear in the subsets of selected features for predicting engagement and challenge, namely, the number of enemies, E, the placement of enemies, E_p and the number of powerups, N_w. Note that frustration can be predicted with the smallest subset of features (only three features have been selected), nevertheless, the prediction accuracy for this emotional state is significantly higher than the ones obtained for predicting engagement and challenge.

Although high accuracies have been obtained for predicting the three emotional states, challenge appears the hardest to model from gameplay features, while frustration is the easiest.

B. Player Experience Modeling Through Mean Head Movement Features

In order to map visual behavior to players' reported affect, the mean head movement features presented in Table II are used as inputs to select the relevant features for predicting players' affect and optimizing the players' experience models. The results presented in Table III show that the models constructed from the head movement features, extracted throughout whole game sessions yield accuracies that are as good as the ones obtained from gameplay features, or slightly lower.

An analysis on the selected features shows that the median horizontal head rotation ($M_{horizontal}$) is an important feature for all three states, while OA and ($M_{vertical}$) are only to be found as predictors of engagement and frustration. Moreover, PO is a common predictor of both engagement and challenge.

The significance test shows that the model constructed for predicting frustration significantly outperforms the two other models for predicting engagement and challenge. Note that this also applies for the models constructed from gameplay features, which implies that single input modalities (behavioral or visual) are better for predicting engagement and frustration than for predicting challenge.

C. Player Experience Modeling Through Visual Reaction Features

It was our assumption that visual reaction features during certain events (losing, making critical moves, etc.) used as the only input channel for estimating affective states would yield more accurate results when compared to mean head movement features (which refer to the overall visual behavior during whole game sessions) or gameplay features. Affective states seem to be mostly correlated with events occurring at certain instances during the game, rather than whole game durations-related visual features. Visual reaction features are fused on the feature level before feeding the predictive models and feature fusion is expected to boost the model's predictive power.

Accuracy obtained for frustration yields higher values when using visual reaction features: visual behavior during jumping, losing, running and interacting with various items appear to be good predictors of frustration. More specifically, it is typical that the Energy Expressivity parameter during interaction with

TABLE III

FEATURES SELECTED FROM THE SET OF GAMEPLAY, MEAN HEAD MOVEMENT (DURING WHOLE GAMES) AND VISUAL REACTION FEATURES (DURING CERTAIN EVENTS) FOR PREDICTING ENGAGEMENT, FRUSTRATION, AND CHALLENGE. TABLE ALSO PRESENTS THE CORRESPONDING AVERAGE ($\bar{P}$) AND BEST (P_{max}) PERFORMANCE VALUES OBTAINED FROM THE ANN MODELS' AND THE BEST MODELS' ANN TOPOLOGIES. THE ANN TOPOLOGIES ARE PRESENTED IN THE FORM: NUMBER OF NEURONS IN THE FIRST HIDDEN LAYER−NUMBER OF NEURONS IN THE SECOND HIDDEN LAYER. BEST PERFORMANCE VALUES OBTAINED (THAT DON'T SHOW SIGNIFICANT DIFFERENCE) FOR EACH EMOTIONAL STATE APPEAR IN BOLD. CONTENT FEATURES ALSO APPEAR IN BOLD

	One Modality		Bimodality		
	Gameplay/Content	Mean Head Movement	Visual Reaction	Gameplay/Mean Head Movement	Gameplay/Visual Reaction
			Engagement		
Selected features	t_{jump} k_{stomp} $\boldsymbol{N_w}$ t_{run} $\boldsymbol{E_p}$	OA Avg $M_{vertical}$ $M_{horizontal}$ SE PO	OA_{endRun} FL_{endRun} FL_{stomp} $PO_{startLeft}$ PO_{move} $TE_{endRight}$ $Avg_{endJump}$	t_{jump} k_{stomp} $\boldsymbol{N_w}$ $M_{horizontal}$ t_{comp} n_{boxes} $M_{vertical}$ FL	$TE_{endRight}$ t_{jump} $\boldsymbol{B}$ PO_{stomp} $TE_{endJump}$
ANN topology	6	4	$4-6$	4-6	2
P	78.69%	74.23%	78.06%	77.78%	**83.97%**
P_{max}	89.68%	78.57%	86.51%	89.68%	**91.27%**
			Frustration		
Selected features	$t_{lastLife}$ n_{boxes} t_{left}	$M_{horizontal}$ OA TE FL $M_{vertical}$	OA_{lose} Avg_{stomp} SE_{endRun} $PO_{startRun}$ $M_{endJump}$ PO_{item}	$t_{lastLife}$ TE n_{gJump} n_{boxes} PO d_{total} OA	FL_{item} FL_{stomp} t_{small} FL_{lose} n_{boxes}
ANN topology	$8-2$	4	$8-10$	$4-4$	8
P	83.5%	83.04%	**86.21%**	83.71%	85.92%
P_{max}	89.17%	89.17%	**92.5%**	92.5%	89.17%
			Challenge		
Selected features	t_{jump} $\boldsymbol{E}$ $k_{unleashed}$ d_{total}	$M_{horizontal}$ FL PO	FL_{lose} $PO_{startRun}$ $FL_{startRun}$ $Avg_{endLeft}$	$M_{hotizontal}$ t_{jump} t_{run} t_{big} $k_{unleashed}$ t_{small}	FL_{item} $FL_{startRun}$ $M_{startJump}$ FL_{lose} $SE_{endLeft}$ t_{left} $Avg_{startJump}$
ANN topology	$4-2$	$4-8$	$10-8$	$10-10$	$10-10$
P	72.36%	75%	**84.13%**	77.36%	78.40%
P_{max}	80.56%	79.17%	**88.88%**	85.41%	86.81%

items (PO_{item}) and starting to run ($PO_{startRun}$), as well as the Overall Activation when losing (OA_{lose}) are related to the notion of frustration. In addition to frustration, very good accuracies have been obtained when using the visual reaction features for predicting challenge with both frustration and challenge significantly outperforming the accuracies obtained for predicting engagement.

D. Fusing Features for Modeling Player Experience

This subsection presents experiments with bimodal features as inputs to the predictive models. We first fuse the gameplay/content with the mean head movement features and we then examine the impact of the fusion between gameplay/content and the visual reaction features on the prediction accuracy of the models.

1) *Modeling Through Gameplay/Content and Mean Head Movement Features:* Using head movement features throughout whole game sessions along with gameplay/content features yield accurate results for predicting engagement, frustration and challenge.

Different gameplay and head movement features have been selected for predicting each emotional state. Median horizontal and vertical head directionality, together with fluidity in motion, along with gameplay/content features (number of killed enemies by stomping, time spent jumping and completing the whole game and powerups) resulted in a model for predicting engagement with up to 89.68% accuracy. Some of these features (such as the number of powerups, N_w, the time spent jumping, t_{jump}, the median horizontal and vertical head direction, $M_{horizontal}$ and $M_{vertical}$) also appear in the subset of features selected when constructing models from each one of these two modalities on its own. This indicates the importance of the features as predictors of player engagement.

The subset of features selected for predicting frustration includes: temporal (TA), energy, and OA expressivity parameters being used along with $t_{lastLife}$, n_{gJump}, n_{boxes} and d_{total}. The TA and OA features also appear in the subset of features selected for predicting frustration from only mean head movement features. Unsurprisingly, the time spent playing during the last life ($t_{lastlife}$) and the number of boxes pressed or destroyed (n_{boxes}) are important predictors of frustration. These gameplay features also appear in the model constructed on gameplay features only.

The features selected for predicting challenge are mainly time-related gameplay features, which are fused with the mean head horizontal rotation ($M_{horizontal}$). The gameplay features

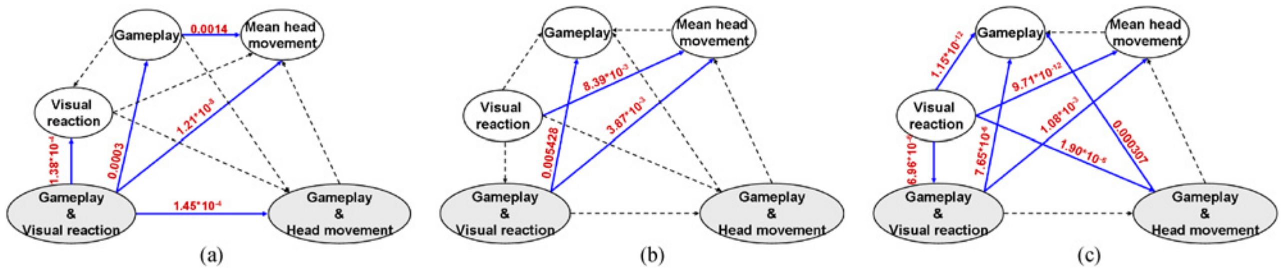

Fig. 6. Testing for statistical significance between the obtained performance of the different sets of features examined for modeling player experience. Solid arrows between two feature sets depict a significant difference on the average performance between them. dash arrows depict average performance differences of no statistical significance. P-values are added next to significant differences. (a) Engagement. (b) Frustration. (c) Challenge.

selected that also appear in the subset of features selected for predicting challenge with only gameplay as input include the time spent jumping (t_{jump}) and the number of opponents that were killed by unleashing a turtle shell ($k_{unleased}$). The new time-related gameplay features selected (t_{run}, t_{big} and t_{small}) result in an average performance increase of 5% (compared to the average performance of the models built on gameplay features only) indicating the importance of time spent running and time Mario being in large or small mode as predictors of player challenge.

The t-test shows that the accuracies obtained from the model constructed for predicting frustration are significantly higher than the ones for predicting engagement and challenge (Note that this finding is similar to the ones observed when testing for differences of significance in mean performance values between the models constructed from gameplay features only and from mean head movement features only).

2) *Modeling Through Gameplay/Content and Visual Reaction Features:* Combining gameplay/content features and visual reaction, results in the appearance of features not used when using each one of the two modalities by themselves. This may be attributed to the fact that there are correlations between features used by gameplay/content and visual reaction features alone. As feature selection seeks beyond linear correlated features, new selected feature subsets are expected to be derived for maximizing performance accuracy.

For engagement, a smaller subset of combined features resulted into a higher accuracy than using larger sets of features from each of the two input modalities alone. Most of the features selected do not appear in the subset of features selected for predicting engagement from each of these two modalities at a time. The majority of the features selected are directly or indirectly linked to head movement and gameplay events while jumping; t_{jump} is an indication of the time spent jumping, PO_{stomp} is the head movement energy while stomping on an enemy, which is an action that requires jumping, $TE_{endJump}$ is the temporal expressivity parameter when landing, and B refers to the number of boxes, which require a jump to interact with. It therefore expectedly appears that the jump event is a contributor for the prediction of engagement in platform games as the average accuracy achieved for engagement (83.97%) via the bimodal fusion of gameplay and visual reaction features is the best obtained across any other feature type as model input.

The selected subset of features for predicting frustration also contains less features than the ones selected individually for each modality. It is interesting to note that there is no overlap between the features selected from the fused features and the ones selected from the visual reaction features while there is only one common feature (n_{boxes}) between the selected fused features and the features selected from gameplay.

The feature subset selected for predicting challenge contains a larger number of features when compared to the ones selected from each modality alone. By looking at the features selected for the three modes—the models constructed from gameplay features, the model constructed from visual reaction features, and the model constructed from fusing these two modalities—it appears that there are two overlaps between the visual reaction features selected ($FL_{startRun}$ and FL_{lose}) and there is no gameplay feature in common. The resulting average performance for challenge (78.4%) suggests that the new features selected do not improve the predictive power of the model when compared to the corresponding performance of the visual reaction features. The statistical analysis shows no significant performance difference between the models constructed for predicting engagement and frustration while these two models' performances are significantly higher than the performance of the model constructed for predicting challenge.

E. Statistical Analysis

We perform a statistical analysis to test for significant differences in the accuracies obtained from the models constructed on all different categories of features. Fig. 6 presents the results obtained from testing for significant performance differences between the models constructed on all categories of features across the three emotional states. A significant difference on average performance is illustrated with a solid arrow, while a dash arrow depicts average performance differences of no statistical significance. The p-values obtained from the statistically significant differences are also presented.

As can be seen from Fig. 6, mean head movement features do not yield high performances compared to the other features when used on their own; all models constructed from other feature sets yield higher or significantly higher performances than the model constructed based on the mean head movement features for engagement. These features, however, outperform (with no significant difference) the models constructed from gameplay features for predicting frustration and challenge. Fusing the mean head movement features with gameplay features, nevertheless, resulted in better accuracies than the ones obtained when only mean head movement features are used to construct the player experience models for all emotional states. The accuracies obtained are even better than the ones

TABLE IV
P-VALUES OBTAINED FROM THE TWO-WAY ANOVA TEST. TWO FACTORS CONSIDERED ARE THE INCLUSION/EXCLUSION OF THE GAMEPLAY FEATURES (A) AND TWO TYPES OF VISUAL CUE FEATURES (B). SIGNIFICANT EFFECTS APPEAR IN BOLD

Factors	Engagement	Frustration	Challenge
(A)	**0.00001**	0.78	**0.03**
(B)	$\mathbf{4.21 * 10^{-6}}$	**0.0004**	$\mathbf{3.54 * 10^{-9}}$
(A*B) Interaction	0.13	0.512	$\mathbf{1.05 * 10^{-6}}$

obtained from gameplay features for predicting frustration and challenge.

Results obtained from models constructed on visual reaction features, on the other hand, are better than the ones obtained from the models constructed on mean head movement features or on gameplay features for predicting frustration and challenge. These models also improve upon the models constructed on the fused features of gameplay and mean head movement for all emotional states.

By fusing visual reaction features with gameplay features, we were able to construct models with higher performance in predicting engagement than any other models constructed from any other feature sets. This argument also holds for frustration and challenge except for the model constructed from visual reaction features, which outperforms the model constructed from fusing these features with gameplay features.

Fusing features from different modalities, in general, appears to result in more accurate models for predicting players' affect than the ones obtained when constructing models from features extracted from one modality. Fusing the features (i.e., visual reaction features) empowers the models with implicit knowledge about more than one channel of information, which appears to have a positive impact on the models' performance.

We have anticipated that fusing gameplay and visual reaction features would yield higher accuracies than when using any other feature set. But our assumption does not hold for the state of challenge. Analyzing the features selected and their correlations with players' preferences would help us shed some light on this effect. However, the models constructed for predicting challenge from visual reaction features and from fusing these features with game play features are multilayered perceptions of two hidden layers, which further implies that the relationship between the features selected and the reported players' preferences is more complex than simple linear correlations.

We anticipate that the performance decrease obtained when fusing the features is the result of the feature selection approach followed, which fails to select the optimal subset of features for prediction when the pool of features to select from become large. For instance the total number of 114 features is reached when fusing gameplay features with visual reaction features.

To further analyze the effect of the interaction between the features on the models' accuracies, we run a two-way ANOVA test. For this test, two factors have been considered: 1) the existence (versus non existence) of the gameplay features for the prediction of affect, and 2) the existence of visual reaction features (versus head movement features). Such an analysis would help us investigate whether the use of visual, or alternatively head movement, features or the fusion of gameplay with visual cues would yield significant changes in the models' performance. The results of a 2×2 [(gameplay and no-gameplay) x (visual reaction and head movement)] between-groups two-way ANOVA are presented in Table IV.

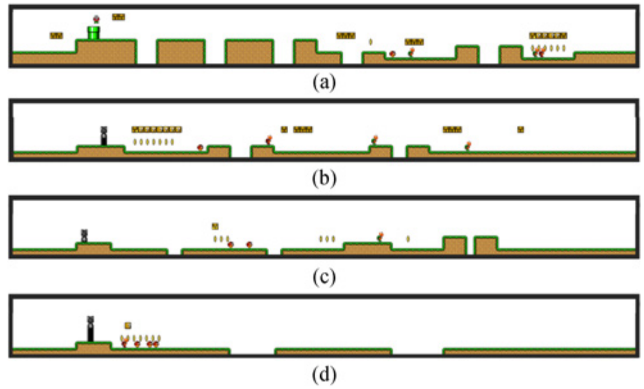

Fig. 7. Example levels generated to maximize predicted frustration and challenge for two human players with different visual reaction features. (a) Generated level for maximum frustration (Subject no. 1). (b) Generated level for maximum frustration (Subject no. 2). (c) Generated level for maximum challenge (Subject no. 1). (d) Generated level for maximum challenge (Subject no. 2.

Both independent variables seems to have an impact on engagement prediction with p-values of 0.0001 and $4.21{*}10^{-6}$, respectively. However, no significant effect was identified when analyzing the interaction between the variables ($p-value = 0.13$). As for frustration, the results showed significant difference only for the second factor ($p-value = 0.0004$) while no significant effects were observed for the first factor ($p-value = 0.78$) or for the interaction between the factors ($p-value = 0.512$). Finally, for challenge, significant effects were observed for both factors ($p-value = 0.03$ and $p-value = 3.54 * 10^9$) and for the interaction between the factors ($p-value = 1.05 * 10^{-6}$). These results suggest that the type of the visual cues has a significant impact on the prediction accuracies for the three emotional states, while the inclusion of the gameplay features was found to have a significant effect on predicting engagement and challenge. The interaction between gameplay and visual cues features, on the other hand, was found to have a significant effect only on the prediction of challenge.

VI. USE OF PLAYER EXPERIENCE MODELS FOR PERSONALIZED LEVEL GENERATION

The ultimate aim for constructing data-driven player experience models is to use these models to close the affective loop [7], [50], [51] in the game by tailoring the game content generation according to each individual players' needs and playing characteristics and realizing the experience-driven PCG [3] core principle. In the proof-of-the-concept experiments presented in this section, we describe the method followed for tailoring content generation driven by the player experience models constructed in the previous section. We focus on the models built on selected features from gameplay and visual reaction as these models give the best accuracy for predicting engagement and high accuracies with rich information about player behavior and visual cues when predicting frustration and challenge.

The player experience models constructed are used to tailor the content of the game to individual players. As a first step toward this process we adopt the methodology proposed in [41] to build models that permit control of content by forcing controllable features in the input of the ANNs. Then, in order to generate levels that are tailored to an individual player, we exhaustively search the content space seeking for a combination of values for the content features that yields (together with the selected gameplay and visual reaction features) the highest ANN output value for the examined affective or cognitive state (i.e., engagement, challenge, and frustration). The details of this approach can be found in earlier work of the authors [41]. Indicatively for the player experience models built in this paper, the search space consists of a maximum of five content features: number of gaps, average width of gaps, number of enemies, enemies placement and number of boxes with value ranges of [2, 6], [5, 15], [3, 7], [0, 2], and [0, 15], respectively. The search space is explored by starting from the minimal possible values and at each step the values are increased by one. With such a small search space (13 200 configurations) we can find the optimal configuration almost instantly, allowing real time level generation.

As a proof-of-the-concept experiment, we generate levels that maximize the predicted frustration and challenge for two human players having different visual reaction features that are not used for model construction. Using the experience-driven PCG mechanism proposed in [41], we were able to generate a new level for each player that optimizes those two states of predicted player experience (Fig. 7). It is apparent that the experience-driven PCG (i.e., adaptation) mechanism generates a variety of personalized levels depending on the behavioral and visual cues of the player. For example it seems that a level can be more frustrating for the first player when it contains more gaps with small width, a large number of boxes, and enemies scattered randomly around. A level with less gaps having small width and enemies around them is found to be more frustrating for the second player. Likewise, a challenging level for the first player is the one containing small width gaps, a small number of enemies scattered randomly around the level, and no boxes. A level with slightly more challenging aspects has been generated for the second player where a smaller number of gaps has been chosen but with larger width, and enemies placed around collectible items.

Note that neither player behavioral data or self-reported experience is available for the generated levels; hence, there is no guarantee that the adaptation mechanism generates higher levels of challenge and frustration. However, the highly accurate ANN models built (above 80% accuracy)—that drive the generation of levels—suggest that higher values are most likely achieved for all emotional states. Moreover, an earlier user study on Super Mario Bros [41]—where the same exhaustive search approach was followed to generate personalized levels based on simpler player models—demonstrated that personalized levels are preferred from the majority of players.

VII. Conclusion

We have presented an extensive set of experiments for modeling player experience in games by relying on two modalities of player input: behavioral data from gameplay and the player's visual behavior. A large corpus of behavioral, visual and player experience report data of 58 Super Mario Bros players has been collected and predictors of player experience have been constructed using a coupling mechanism of automatic feature selection and neuroevolutionary preference learning. It was shown that players' visual reactions fused against certain game events can provide a rich source of information regarding preferences with respect to challenge and frustration (reaching model accuracies of 88.88% and 92.5%, respectively). However, engagement (best model accuracy obtained was 91.27%) seem to be a notion related both to the way a game has been designed, played, as well as to the visual information coming from the player himself.

Future work also includes testing for the generality of the proposed methodology and the results obtained. While Super Mario Bros defines more or less the platform game genre, it would be interesting to investigate to which extent the methodology proposed can be generalized to other game genres, such as first person shooter (FPS) or serious games. We argue that the approach presented has a great potential to be applied successfully to such games since most of the gameplay features defined can be easily generalized to capture playing styles in a variety of other games. The applicability of the visual reaction features (which proven to be efficient predictors of player's affect) appears to be a trivial process since the extraction of these features depends on key performance events of the context (such as indicators of losing and winning).

There are a number of limitations inherent in the player experience modeling approach followed. The feature selection method provides an efficient mechanism for selecting relevant features when the size of the search space is rather small. This method, however, results in a suboptimal subset of features when searching a large space. Automatic feature selection is an essential step when constructing the experience models since selecting the correct subset of features may have a great impact on the prediction accuracy obtained. Improving on the global search abilities of the feature selection process is one way to improve the prediction accuracy of the models. Algorithms relying on meta-heuristic search, such as genetic-based feature selection [52] can improve the detection of more appropriate feature subsets.

Another limitation of the proposed modeling method concerns the expressiveness of the player experience models. By using neuroevolutionary preference learning, we gain the advantage of universal approximation capacity for constructing accurate nonlinear models but we loose the ability of easily analyzing the cause-effect relationships between the features selected and the models' prediction of each emotional state. Thus, exploiting the use of more expressive model representations, such as decision trees or fuzzy neural networks for modeling player experience constitutes a future direction.

As demonstrated with a proof-of-the-concept experiment in this paper, a level designer can use the derived player experience models and automatically generate personalized levels for each player. Given a set of behavioral and visual reaction features of a player, the ANN player experience models can inform the designer about the set of game level features (such as the number of enemies and gaps) that can maximize (or indeed minimize) the modeled experience state (ANN output) for that particular player. The personalized generated Super Mario Bros levels show that the experience-driven procedural content generation framework [3] can be realized, the affective loop can be closed in games and provides a novel approach for control and adaptation in computer games. The adaptation methodology proposed, however, needs to be validated with human players in actual gameplay sessions where players get to play and compare randomly generated levels against levels, which are optimized for a player's modeled experience. Results in earlier studies on a small group of human players showcase that the exhaustive search adaptation framework is effective in generating levels, which are preferred by the majority of players [41].

The exhaustive search adaptation method presented in this paper is appropriate due to the relatively small size of the search space explored. As this paper did not focus on experience model-driven adaptation—but on the fusion of modalities for the creation of reliable player experience models—future work includes the construction and validation of more general methods for game adaptation, which are effective in larger search spaces. Evolutionary methods, for instance, can be utilized for this purpose; previous studies have demonstrated the potential of meta-heuristics in exploring large content spaces by integrating the adaptation mechanism within the content generation process [53], [54].

REFERENCES

[1] Reuters. (2007). *Video-game sales overtaking music* [Online]. Available: http://articles.moneycentral.msn.com/Investing/Extra/VideoGameSalesOver%takingMusic.aspx

[2] G. N. Yannakakis, "Game AI revisited," in *Proc. ACM Comput. Frontiers Conf.*, May 2012, pp. 285–292.

[3] G. N. Yannakakis and J. Togelius, "Experience-driven procedural content generation," *IEEE Trans. Affective Comput.*, vol. 2, no. 3, pp. 147–161, Jul.–Sep. 2011.

[4] 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 Applicat.*, vol. 41, no. 3, pp. 469–493, 2009.

[5] M. Csikszentmihalyi, *Flow: The Psychology of Optimal Experience*. New York, NY, USA: Harper Perennial, 1991.

[6] G. Calleja, *In-Game: From Immersion to Incorporation*. Cambridge, MA, USA: MIT Press, 2011.

[7] K. Höök, "Affective loop experiences—What are they?," in *Proc. PERSUASIVE*, LNCS 5033. 2008, pp. 1–12.

[8] C. Jennett, A. L. Cox, P. Cairns, S. Dhoparee, A. Epps, T. Tijs, and A. Walton, "Measuring and defining the experience of immersion in games," *Int. J. Human Comput. Stud.*, vol. 66, no. 9, pp. 641–661, Sep. 2008.

[9] S. Ioannou, G. Caridakis, K. Karpouzis, and S. Kollias, "Robust feature detection for facial expression recognition," *J. Image Video Process.*, vol. 2007, no. 2, p. 5, Jul. 2007.

[10] G. Castellano, A. Pereira, I. Leite, A. Paiva, and P. W. McOwan, "Detecting user engagement with a robot companion using task and social interaction-based features," in *Proc. Int. Conf. Multimodal Interfaces*, 2009, pp. 119–126.

[11] J. Sanghvi, G. Castellano, I. Leite, A. Pereira, P. W. McOwan, and A. Paiva, "Automatic analysis of affective postures and body motion to detect engagement with a game companion," in *Proc. 6th Int. Conf. Human Robot Interaction*, 2011, pp. 305–312.

[12] A. J. N. van Breemen, X. Yan, and B. Meerbeek, "iCat: An animated user-interface robot with personality," in *Proc. AAMAS*, 2005, pp. 143–144.

[13] A. Drachen, A. Canossa, and G. N. Yannakakis, "Player modeling using self-organization in tomb raider: Underworld," in *Proc. IEEE Symp. Comput. Intell. Games*, Sep. 2009, pp. 1–8.

[14] N. Shaker, J. Togelius, G. N. Yannakakis, B. Weber, T. Shimizu, T. Hashiyama, N. Sorenson, P. Pasquier, P. Mawhorter, G. Takahashi, G. Smith, and R. Baumgarten, "The 2010 Mario AI championship: Level generation track," *IEEE Trans. Comput. Intell. Games*, vol. 3, no. 4, pp. 332–347, Dec. 2011.

[15] G. Yannakakis, H. Martínez, and A. Jhala, "Towards affective camera control in games," *User Model. User-Adapted Interaction*, vol. 20, no. 4, pp. 313–340, 2010.

[16] S. Tognetti, M. Garbarino, A. Bonarini, and M. Matteucci, "Modeling enjoyment preference from physiological responses in a car racing game," in *Proc. IEEE Symp. Comput. Intell. Games*, Aug. 2010, pp. 321–328.

[17] R. Mandryk, K. Inkpen, and T. Calvert, "Using psychophysiological techniques to measure user experience with entertainment technologies," *Behav. Inform. Technol.*, vol. 25, no. 2, pp. 141–158, 2006.

[18] A. Kapoor, W. Burleson, and R. W. Picard, "Automatic prediction of frustration," *Int. J. Human Comput. Stud.*, vol. 65, no. 8, pp. 724–736, Aug. 2007.

[19] J. Sykes, "Affective gaming: Measuring emotion through the gamepad," in *Proc. CHI: New Horizons*, 2003, pp. 732–733.

[20] F. Buttussi, L. Chittaro, R. Ranon, and A. Verona, "Adaptation of graphics and gameplay in fitness games by exploiting motion and physiological sensors," in *Proc. Smart Graphics, 7th Int. Symp.* (ser. Lecture Notes in Computer Science), vol. 4569. Jun. 2007, pp. 85–96.

[21] H. O. Istance, A. Hyrskykari, S. Vickers, and T. Chaves, "For your eyes only: Controlling 3-D online games by eye-gaze," in *Proc. INTERACT*, vol. 1. 2009, pp. 314–327.

[22] L. Nacke, S. Stellmach, D. Sasse, J. Niesenhaus, and R. Dachselt, "Laif: A logging and interaction framework for gaze-based interfaces in virtual entertainment environments," in *Proc. Interactive Cultures Conf.*, Sep. 2010, pp. 19–28.

[23] S. Almeida, A. Veloso, L. Roque, and O. Mealha, "The eyes and games: A survey of visual attention and eye tracking input in video games," in *Proc. SBGames: X Brazilian Symp. Comput. Games Dig. Entertainment Arts Design Track*, Nov. 2011.

[24] M. S. El-Nasr and S. Yan, "Visual attention in 3-D video games," in *Proc. Adv. Comput. Entertainment Technol.*, 2006, p. 22.

[25] P. Isokoski, M. Joos, O. Spakov, and B. Martin, "Gaze controlled games," *Universal Access Inform. Soc.*, vol. 8, no. 4, pp. 323–337, 2009.

[26] J. D. Smith and T. C. N. Graham, "Use of eye movements for video game control," in *Proc. ACM SIGCHI Int. Conf. Adv. Comput. Entertainment Technol.*, Jan. 2006, pp. 20–28.

[27] S. Kaiser, T. Wehrle, and S. Schmidt, "Emotional episodes, facial expressions, and reported feelings in human-computer interactions," in *Proc. Xth Conf. Int. Soc. Res. Emotions*, 1998, pp. 82–86.

[28] A. Liapis, G. N. Yannakakis, and J. Togelius, "Adapting models of visual aesthetics for personalized content creation," *IEEE Trans. Comput. Intell. AI Games, Special Issue Comput. Aesthetics Games*, vol. 4, no. 3, pp. 213–228, Sep. 2012.

[29] J. Doran and I. Parberry, "Controlled procedural terrain generation using software agents," *IEEE Trans. Comput. Intell. AI Games*, vol. 2, no. 2, pp. 111–119, Jun. 2010.

[30] M. Frade, F. F. de Vega, and C. Cotta, "Evolution of artificial terrains for video games based on accessibility," in *Proc. Evol. Applicat.*, vol. 6024. 2010, pp. 90–99.

[31] L. Cardamone, G. Yannakakis, J. Togelius, and P. Lanzi, "Evolving interesting maps for a first person shooter," in *Proc. Int. Conf. Appl. Evol. Comput.*, vol. 6624. 2011, pp. 63–72.

[32] J. Togelius, M. Preuss, and G. Yannakakis, "Towards multiobjective procedural map generation," in *Proc. Workshop Procedural Content Generation Games*, 2010, p. 3.

[33] S. Kazmi and I. Palmer, "Action recognition for support of adaptive gameplay: A case study of a first person shooter," *Int. J. Comput. Games Technol.*, p. 1, 2010.

[34] J. Togelius, R. De Nardi, and S. Lucas, "Towards automatic personalised content creation for racing games," in *Proc. IEEE Symp. Comput. Intell. Games*, Apr. 2007, pp. 252–259.

[35] G. Smith, J. Whitehead, and M. Mateas, "Tanagra: A mixed-initiative level design tool," in *Proc. Int. Conf. Foundations Dig. Games*, Jun. 2010, pp. 209–216.

[36] N. Shaker, S. Asteriadis, G. Yannakakis, and K. Karpouzis, "A game-based corpus for analysing the interplay between game context and player experience," in *Affective Computing and Intelligent Interaction*, vol. 6975. Berlin, Germany: Springer, 2011, pp. 547–556.

[37] G. N. Yannakakis, M. Maragoudakis, and J. Hallam, "Preference learning for cognitive modeling: A case study on entertainment preferences," *IEEE Trans. Sys., Man, Cyber. A, Syst., Humans*, vol. 39, no. 6, pp. 1165–1175, Nov. 2009.

[38] K. Poels and W. I. Jsselsteijn, "Development and validation of the game experience questionnaire," in *Proc. FUGA Workshop Mini Symp.*, Apr. 2008.

[39] G. N. Yannakakis, "Preference learning for affective modeling," in *Proc. Int. Conf. Affective Comput. Intell. Interaction*, Sep. 2009, pp. 126–131.

[40] C. Pedersen, J. Togelius, and G. N. Yannakakis, "Modeling player experience for content creation," *IEEE Trans. Comput. Intell. AI Games*, vol. 2, no. 1, pp. 54–67, Mar. 2010.

[41] N. Shaker, G. N. Yannakakis, and J. Togelius, "Towards automatic personalized content generation for platform games," in *Proc. AAAI Conf. Artif. Intell. Interactive Dig. Entertainment*, Oct. 2010, pp. 63–68.
[42] N. Shaker, G. Yannakakis, and J. Togelius, "Digging deeper into platform game level design: Session size and sequential features," in *Proc. Eur. Conf. Applicat. Evol. Comput.(EvoApplications)*, 2012.
[43] M. Csikszentmihalyi, *Finding Flow: The Psychology of Engagement with Everyday Life*. New York, NY, USA: Basic Books, 1997.
[44] G. Caridakis, G. Castellano, L. Kessous, A. Raouzaiou, L. Malatesta, S. Asteriadis, and K. Karpouzis, "Multimodal emotion recognition from expressive faces, body gestures and speech," in *Proc. AIAI*, 2007, pp. 375–388.
[45] G. Caridakis, S. Asteriadis, and K. Karpouzis, "User modeling via gesture and head pose expressivity features," in *Proc. 5th Int. Workshop Semantic Media Adaptation Personalization*, 2010, pp. 9–10.
[46] G. Yannakakis and J. Hallam, "Erratum: Ranking vs. preference: A comparative study of self-reporting," *Affective Comput. Intell. Interaction*, p. 1, 2011.
[47] J. Fürnkranz and E. Hüllermeier, "Pairwise preference learning and ranking,"in *Proc. Mach. Learn. ECML*, 2003, pp. 145–156.
[48] G. N. Yannakakis and J. Hallam, "Game and player feature selection for entertainment capture," in *Proc. 5th Int. Conf. Comput. Intell. Games*, 2007, pp. 244–251.
[49] G. N. Yannakakis and J. Hallam, "Entertainment modeling through physiology in physical play," *Int. J. Human Comput. Stud.*, vol. 66, no. 10, pp. 741–755, Oct. 2008.
[50] P. Sundstrm, "Exploring the affective loop," Stockholm Univ., Stockholm, Sweden, Tech. Rep., 2005.
[51] E. Hudlicka, "Affective computing for game design," in *Proc. 4th Int. North Amer. Conf. Intell. Games Simulation*, 2008, pp. 5–12.
[52] H. Martinez and G. Yannakakis, "Genetic search feature selection for affective modeling: A case study on reported preferences," in *Proc. 3rd Int. Workshop Affective Interaction Natural Environments*, 2010, pp. 15–20.
[53] N. Shaker, G. Yannakakis, J. Togelius, M. Nicolau, and M. O'Neill, "Evolving personalized content for super mario bros using grammatical evolution," in *Proc. AAAI Conf. Artif. Intell. Interactive Dig. Entertainment*, Oct. 2012, pp. 48–54.
[54] E. J. Hastings, R. K. Guha, and K. O. Stanley, "Evolving content in the galactic arms race video game," in *Proc. 5th Int. Conf. Comput. Intell. Games*, 2009, pp. 241–248.

Noor Shaker received the five-year B.A. degree in IT engineering from Damascus University, Damascus, Syria, in 2007, the M.Sc. degree in artificial intelligence from Katholieke Universiteit Leuven, Leuven, Belgium, in 2009, and the Ph.D. degree from the IT University of Copenhagen, Copenhagen, Denmark, in 2013.

She is a Post-Doctoral Researcher at the IT University of Copenhagen. Her current research interests include player modeling, procedural content generation, computational creativity, affective computing, and player behavior imitation.

Stylianos Asteriadis graduated from the Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Thessaloniki, Greece, in 2004, received the master's degree in digital media from the Aristotle University of Thessaloniki, in 2006, and the Ph.D. degree from the National Technical University of Athens, Athens, Greece, in 2011.

He is currently a Post-Doctoral Researcher at the Informatics and Telematics Institute, Centre for Research and Technology Hellas, Thessaloniki, Greece. His current research interests include image processing, machine learning, 3-D reconstruction, affective computing, and activity recognition in human-computer interaction scenarios. He has also published over 30 journal and international conference papers in the aforementioned areas.

Georgios N. Yannakakis (S'04–M'05) received the Ph.D. degree in informatics from the University of Edinburgh, Edinburgh, Scotland, in 2005.

He is an Associate Professor with the University of Malta (UoM), Msida, Malta. Prior to joining the Department of Digital Games, UoM, in 2012, he was an Associate Professor at (and still affiliated with) the Center for Computer Games Research at the IT University of Copenhagen, Copenhagen, Sweden. He has published over 100 journal and international conference papers in the aforementioned fields. His current research interests include AI (computational intelligence, preference learning), affective computing (emotion detection, emotion annotation), advanced game technology (player experience modeling, procedural content generation, personalization), and human–computer interaction (multimodal interaction, psychophysiology, user modeling).

Dr. Yannakakis is an Associate Editor of the IEEE TRANSACTIONS ON AFFECTIVE COMPUTING and the IEEE TRANSACTIONS ON COMPUTATIONAL INTELLIGENCE AND AI IN GAMES, and the chair of the IEEE CIS Task Force on Player Satisfaction Modeling.

Kostas Karpouzis (SM'13) received the Ph.D. degree from the National Technical University of Athens, Athens, Greece, in 2001.

He is currently an Associate Researcher with the Institute of Communication and Computer Systems, Athens. He has published over 110 papers in international journals and proceedings of international conferences. He has contributed to the *Humaine Handbook on Emotion Research* and K. Scherer's *Blueprint for Affective Computing: a Sourcebook*. Since 1995, he has participated in over 12 research projects at Greek and European levels—most notably, the FP6 Network of Excellence Humaine and the FP7 ICT Siren projects on children with issues in conflict resolution (http://www.sirenproject.eu), where he is the Technical Manager. His current research interests include human–computer interaction, affective and assistive computing, multimodal sensing, understanding and reasoning about emotion, and affect-aware games.

Dr. Karpouzis is an Elected Member of the Executive Committee of the Humaine Association and the Student Activities Chair for the IEEE Greece Section. In 2010, he was invited to present affective computing as a potential technology at the Social Inclusion and Related Technologies Workshop organized by the ICT for Inclusion Unit. He was the Keynote Speaker of the EvoGames 2010 Conference, Istanbul, Turkey, and the SMAP 2010 Conference, Cyprus.