Game and Player Feature Selection for Entertainment Capture

Georgios N. Yannakakis* and John Hallam[†]
Mærsk Mc-Kinney Møller Institute
University of Southern Denmark
Campusvej 55, DK-5230, Odense
{georgios*;john[†]}@mmmi.sdu.dk

Abstract— The notion of constructing a metric of the degree to which a player enjoys a given game has been presented previously. In this paper, we attempt to construct such metric models of children's 'fun' when playing the Bug Smasher game on the Playware platform. First, a set of numerical features derived from a child's interaction with the Playware hardware is presented. Then the Sequential Forward Selection and the n-Best feature selection algorithms are employed together with a function approximator based on an artificial neural network to construct feature sets and function that model the child's notion of 'fun' for this game. Performance of the model is evaluated by the degree to which the preferences predicted by the model match those expressed by the children in a survey experiment.

The results show that an effective model can be constructed using these techniques and that the Sequential Forward Selection method performs better in this task than n-Best. The model reveals differing preferences for game parameters between children who react fast to game events and those who react slowly. The limitations and the use of the methodology as an effective adaptive mechanism to entertainment augmentation are discussed.

Keywords: Entertainment modeling, intelligent interactive playgrounds, neuro-evolution.

I. Introduction

Cognitive modeling projects significant potential within digital interactive entertainment systems (such as computer games). Being able to model the level of user (gamer) engagement or satisfaction in real-time can provide insights to the appropriate AI methodology for enhancing the quality of playing experience [1] and furthermore be used to adjust digital entertainment environments according to individual user preferences.

The 'Playware' [2] intelligent interactive physical playground attempts to combine the advantages of both computer games and traditional playgrounds. On one hand, computer games keep children (among others) engaged more than other digital media because of their high degree of interactivity and the freedom for the child to develop and play a role within a fantasy world which is created during play [3]. On the other hand, traditional playgrounds offer the advantage of physical play, which furthermore improves the child's health condition, augment children's ability to engage in social and fantasy play [4], [5] and provide the freedom for children to generate their own rules for their own developed games. Experiments with children playing Playware games will be presented in this paper.

Following from the reported successful entertainment capture in physical interactive games [6], a further endeavor on capturing player satisfaction during gameplay (i.e. entertainment modeling) through more extensive methodology and experiments with children players is presented in this paper. As in [6], this is achieved by following the theoretical principles of Malone's intrinsic qualitative factors for engaging gameplay [3], namely *challenge*, *curiosity* and *fantasy*. Quantitative measures for challenge and curiosity are used from previous studies on quantitative reported entertainment capture [6] in the Playware playground.

A mapping between the aforementioned factors (game features) and the children's notion of 'fun' or entertainment (the two terms are used interchangeably herein) is derived using a game developed on the Playware playground as a test-bed. Each player's individual characteristics (player features), such as response time and foot pressure, are recorded during each during game. Feedforward artificial neural networks (ANNs) are trained using artificial evolution on this gameplay experimental data to construct a function mapping the examined game features and player features to the reported player satisfaction preferences. The n-Best (nBest) and the Sequential Forward Selection (SFS) [7] feature selection methods are used to extract the minimal subset of game and player features to be included in the ANN model.

Single feature experiments demonstrate that the average response time of children interacting with the playground is the feature that yields the best (highest-performing) mapping between game and player features and children's expressed preferences on entertainment. This result is consistent with the reported impact (i.e. significant linear correlation) of the average response time on reported entertainment in Playware games [6]. When more than one feature is examined, the SFS method generates feature subsets performing better overall than the subsets generated by the nBest method. More specifically, it finds a set of four features that yields the highest performance in matching opponent's and player's behavior to children's perceived entertainment. Player features include the player's average response time with the playground, the variance of the pressure force instances on the playground and the number of interactions with the playground. The game feature included in the most accurate model of player satisfaction obtained is the level of curiosity generated by the game opponents. Analysis of the obtained model shows that different children (classified by their response time) have different requirements on the levels of the curiosity factor for the game to be judged entertaining.

The work reported here is novel in that it isolates game and player features attributed to reported entertainment in physically demanding games and demonstrates a way of constructing a subjective model (a predictor of user preferences) of reported entertainment grounded in statistical features obtained from child-game interaction. The limitations of the proposed methodology and its extensibility to other genres of digital entertainment are discussed. Its generic use as an efficient baseline for capturing reported entertainment in physical interactive games in real-time is also outlined.

II. ENTERTAINMENT CAPTURE

There have been several psychological studies to identify what is 'fun' in a game and what engages people playing computer games. Theoretical approaches include Malone's principles of intrinsic qualitative factors for engaging game play [3], namely challenge, curiosity and fantasy as well as the well-known concepts of the theory of flow [8] incorporated in computer games as a model for evaluating player enjoyment, namely *GameFlow* [9]. A comprehensive review of the literature on qualitative approaches for modeling player enjoyment demonstrates a tendency of overlapping with Malone's and Csikszentmihalyi's foundational concepts. Many of these approaches are based on Lazzaro's 'fun' clustering which uses four entertainment factors based on facial expressions and data obtained from game surveys on players [10]: hard fun, easy fun, altered states and socialization. Koster's [11] theory of fun, which is primarily inspired by Lazzaro's four factors, defines 'fun' as the act of mastering the game mentally. An alternative approach to fun capture is presented in [12] where fun is composed of three dimensions: endurability, engagement and expectations.

Vorderer et al. [13] present a quantitative analysis of the impact of competition (i.e. challenge) on entertainment and identify challenge as the most important determinant of the enjoyment perceived by video game (Tomb Raider) players. They claim that a successful completion of a task generates sympathetic arousal, especially when the challenge of the task matches the player's abilities. According to Choi et al. [14], challenge and satisfaction appear as independent processes, in contrast to the views of Malone [3] and Yannakakis et al. [6] where satisfaction derives from the appropriate level of challenge and other game components.

Iida's work on metrics of entertainment in board games was the first attempt in the area of quantitative 'fun' modeling. He introduced a general metric of entertainment for variants of chess games depending on average game length and possible moves [15]. Other work in the field of quantitative entertainment capture is based on the hypothesis that the player-opponent interaction — rather than the audiovisual features, the context or the genre of the game — is the property that contributes the majority of the quality features of entertainment in a computer game [16]. Based on this fundamental assumption, a metric for measuring the real time entertainment value of predator/prey games was designed, and established as efficient and reliable by validation against human judgement [17], [18]. Further studies by Yannakakis

and Hallam [19] have shown that Artificial Neural Networks (ANN) and fuzzy neural networks can extract a better estimator of player satisfaction than a human-designed one, given appropriate estimators of the challenge and curiosity of the game and data on human players' preferences.

A step further to entertainment capture is towards games of richer human-computer interaction and affect recognizers which are able to identify correlations between physiological signals and the human notion of entertainment. Experiments by Yannakakis et al. [20] have already shown a significant effect of children's average heart rate on children's reported entertainment in action games played in interactive physical playgrounds. Moreover, Rani et al. [21] propose a methodology for detecting anxiety level of the player and appropriately adjusting the level of challenge (e.g. speed) in the game of 'Pong'. Physiological state (hear-rate, galvanic skin response) prediction models have also been proposed for potential entertainment augmentation in computer games [22]. Similar work in adjusting a game's difficulty include endeavors through reinforcement learning [23], genetic algorithms [24], probabilistic models [25] and dynamic scripting [26]. However, the aforementioned attempts are based on the assumption that challenge is the only factor that contributes to enjoyable gaming experiences while results reported have not been cross-verified by human players.

Following the theoretical principles reported from Malone [3], Koster [11] and Yannakakis [18], this paper is primarily focused on the contributions of game opponents' behavior to the real-time entertainment value of the game. We argue that among the three dimensions of 'fun' (endurability, engagement, expectations) defined in [12] it is only *engagement* that is affected by the opponent since both *endurability* and *expectations* are based primarily on the game design *per se*. Given a successful interactive game design that yields high expectations and endurability, we only focus on the level of engagement that generates 'fun' (entertainment). However, instead of being based on empirical observations of children's entertainment, the work presented here uses quantitative measures for Malone's entertainment factors of challenge and curiosity (as introduced in [6]). On that basis, a mapping between the two aforementioned factors, children's play recorded features and their expressed preferences is constructed using experimental data obtained from a survey experiment with children playing with Playware playground (see Section III).

III. PLAYWARE PLAYGROUND

The Playware [2] prototype playground consists of several building blocks (i.e. tangible tiles) that allow for the game designer (e.g. the child) to develop a significant number of different games within the same platform. The overall technological concept of Playware is based on embodied AI [27] where intelligent physical identities (tiles) incorporate processing power, communication, input and output, focusing on the role of the morphology-intelligence interplay in developing game platforms. See [6], [2] for further details on Playware playground.

Fig. 1.   A child playing the Bug-Smasher game.

A. Bug-Smasher Game

The test-bed game used for the experiments presented here is called 'Bug-Smasher'. The game is developed on a 6 x 6 square tile topology (see Fig. 1). During the game, different 'bugs' (colored lights) appear on the game surface and disappear sequentially after a short period of time by turning a tile's light on and off respectively. A bug's position is picked randomly according to the predefined level of the bugs' spatial diversity. Spatial diversity is measured by the entropy (H) of the bug-visited tiles.

The child's goal is to smash as many bugs as possible by stepping on the lighted tiles. Bug-smasher has been used as a test-bed in previous work; further details can been found in [6], [20] and [28].

IV. EXPERIMENTAL DATA

The Bug-Smasher game has been used to acquire data of children's judgement on entertainment. Three states ('Low', 'Average', and 'High') are used for each of the two entertainment factors of challenge and curiosity summing up to 9 different game states. The fantasy factor is not investigated through this survey since the focus of this paper is on the opponent (bug) contribution to entertainment. See [28] for fantasy's positive impact on entertainment in Bug-Smasher.

We consider (as in [6]) the speed (S — in sec^{-1}) that the bugs appear and disappear from the game and their spatial diversity (H) on the game's plane as appropriate measures to represent the level of challenge and the level of curiosity (unpredictability) respectively [3] during gameplay. The former provides a notion for a goal whose attainment is uncertain and the latter effectively portrays a notion of unpredictability in the subsequent events of the game — the higher the H value the higher the bug appearance unpredictability and therefore the higher the curiosity.

Seventy two normal-weighted (based on their body mass index) children whose age covered a range between 8 and 10 years participated in an experiment. By experimental design, each subject plays against two of the selected game states in all permutations of pairs. The number of children participated in the experiment is derived from $2 \cdot C_2^9 = 72$ being twice the required number of all combinations of 2 out of 9 game states. In this experiment, each subject plays two games (A and B) for 90 seconds each; the two games differ in the levels of one or both entertainment factors of challenge and curiosity. Each time a pair of games is finished, the child is asked whether the first game was more 'fun' (see [12] for terminology used in experiments with children) than the second game i.e. whether A or B generated a more entertaining game. The 2-alternative forced choice (2-AFC) approach is used since it offers several advantages for a subjective entertainment capture: it minimizes the assumptions made about children's notions of "fun" and allows a fair comparison between the answers of different children. Since our focus is to construct a model relating reported entertainment preferences to game and player features that generalises over the reports of different children 2-AFC is preferred to a ranking approach [29]. Note also that, children are not interviewed but are asked to fill in a questionnaire, minimizing the interviewing effects reported in [29].

The child's answers are used to guide the training of an ANN model of reported entertainment (see Section V). In order to minimize any potential order effects we let each subject play the aforementioned games in the inverse order too. Statistical analysis of the subjects' answers shows that no significant order effect occurs ($r_c = -0.102$, p-value$= 0.224$). The reported insignificant order effect also, in part, demonstrate that effects such as a child's possible preference for the very first game played and the interplay between reported entertainment and familiarity with the game are statistically insignificant. The total number of game pairs played equals 144; however, data from 137 game pairs are used due to hardware (communication ports) failure during seven games.

Since, with the current implementation of the Playware platform, the only input to the system is through a Force Sensing Resistor (FSR) sensor, quantitative individual playing characteristics can only be based on three measurable features: the state (position and LEDs color) of a pressed tile, the time that a tile-press event took place and the pressure force on a pressed tile. Pressed tile events are recorded in real-time and a selection of nine personalized (individual) player features are calculated for each child. These include the number of smashed bugs over the total number of bugs appeared P (i.e. child's score); the number of interactions with the game environment N_I; the average and the variance of the response times ($E\{r_t\}, \sigma^2\{r_t\}$); the average and the variance of the distance between the pressed tile and the bugs appearing on the game ($E\{D_b\}, \sigma^2\{D_b\}$); the average and the variance of the pressure recorded from the FSR sensor $E\{p\}, \sigma^2\{p\}$); and the entropy of the tiles that the child visited H_C.

A. Statistical Analysis

The aim of the statistical analysis presented here is to identify statistically significant correlations between children's

notion of entertainment and any of the aforementioned individual player features and/or the quantitative entertainment factors (game features): challenge and curiosity. For this purpose the following null hypothesis is formed: The correlation between observed children judgement of entertainment and recorded player and game features, as far as the different game states are concerned, is a result of randomness. The test statistic is obtained through $c(\overrightarrow{z}) = \sum_{i=1}^{N_s}\{z_i/N_s\}$, where N_s is the total number of game pairs played ($N_s = 137$) and $z_i = 1$, if the subject chooses as the more entertaining game the one with the larger value of the examined feature and $z_i = -1$, if the subject chooses the other game in the game pair i.

The obtained significant — significance equals 5%, high significance equals 1% in this paper — effects of the selected features on reported entertainment are: N_I ($c(\overrightarrow{z}) = 0.1678$, p-value $= 0.0298$), $E\{r_t\}$ ($c(\overrightarrow{z}) = -0.2262$, p-value $= 0.0050$), $\sigma^2\{r_t\}$ ($c(\overrightarrow{z}) = -0.1532$, p-value $= 0.0435$), $E\{p\}$ ($c(\overrightarrow{z}) = 0.1678$, p-value $= 0.0298$) and $\sigma^2\{p\}$ ($c(\overrightarrow{z}) = 0.1970$ p-value $= 0.0129$). These effects appear to be commonsensical since the Bug-Smasher game belongs to the genre of action physical games where the level of engagement of the user tends to have a significant effect on the number of interactions and the reaction time of the player [30]. In Bug-Smasher, the more a child is entertained the more (N_I) and harder ($E\{p\}$) she/he tends to interact with the game platform. This behavior generates lower average response time ($E\{r_t\}$) and higher average pressure on the tiles ($E\{p\}$). Moreover, it appears that the variability of the aforementioned individual characteristics ($\sigma^2\{r_t\}$, $\sigma^2\{p\}$) does have an effect on reported entertainment too. The obtained highly significant effect of $E\{r_t\}$ is consistent with previous experiments on the Bug-Smasher game [6].

On the other hand it appears that reported entertainment cannot be objectively modeled according to the levels of challenge ($c(\overrightarrow{z}) = 0.0$, p-value $= 0.5382$) and curiosity ($c(\overrightarrow{z}) = 0.0909$, p-value $= 0.1448$) since there exists a level of personalization which has to be included as a factor in entertainment modeling. The feature selection procedure presented in Section VI allows the designer to choose specific individual player features that can successfully map between children's behavior, game features and reported entertainment.

V. EVOLVING ANN

The proposed approach to entertainment modeling is based on selecting a minimal subset (see Section VI) of game and player features and constructing a quantitative user model that predicts the children's reported entertainment preferences. For this purpose, a fully-connected feedforward ANN for learning the relation between the selected game and player features (ANN inputs) and the "entertainment value" (ANN output) of a game is presented. The assumption is that the entertainment value y of a given game is an unknown function of player and game features which the ANN will learn. The children's expressed preferences constrain but do not specify the values of y for individual games but we assume that the child's expressed preferences are consistent. Since there are no prescribed target outputs for the learning problem (i.e. no differentiable output error function), ANN training algorithms such as back-propagation are inapplicable. Learning is achieved through artificial evolution [31] and is described in Section V-A.

The sigmoid function is employed at each neuron, the connection weights take values from -5 to 5 to match with input values normalized into [0, 1] before they are entered into the ANN. In an attempt to minimize the controller's size, it was determined that a single hidden-layered ANN architecture, containing 20 hidden neurons, is capable of successfully obtaining solutions of high fitness. This was determined by considering the performance of ANN architectures with up to two hidden layers containing up to 30 hidden neurons each.

A. Genetic Algorithm

A generational genetic algorithm (GA) [32] is implemented, which uses a fitness function that measures the difference between the children's reported preferences of entertainment and the model output value y. The ANN is itself evolved. In the algorithm presented here, the ANN topology is fixed and the GA chromosome is a vector of ANN connection weights. The algorithm is described briefly in this section since it has previously presented in [6].

A population of N (N is 1000 in this paper) networks is initialized randomly. Initial real values that lie within [-5, 5] for their connection weights are picked randomly from a uniform distribution. Then, at each generation: (a) Each member (neural network) of the population is given two n_i-tuple (where n_i is the number of game or player features) values one for opponent/game A and one for opponent/game B for each pair j of games played in the survey experiment ($N_s = 137$) — see [6] for further details. In each case it returns two output values, representing the level of 'fun' in each game, namely $y_{j,A}$ and $y_{j,B}$. (b) Each member i of the population is evaluated via a fitness function f_i that promotes the matching between ANN outputs (y) and children's reported answers (see [6]). A high fitness results if the ranking of $y_{j,A}$ and $y_{j,B}$ matches the expressed preference of the children for each game pair j. (c) A fitness-proportional selection method is used. (d) Montana and Davis [33] crossover and Gaussian mutation are applied (see [6]).

The algorithm is terminated when either a good solution is found ($f > 0.95 f_{max}$; where f_{max} is the maximum fitness) or a large number of generations g is completed ($g = 10000$).

VI. FEATURE SELECTION

There are two different feature selection schemes applied and compared in this paper. Given both the individual player features and the game features presented in section IV the *n Best Features Selection* (nBest) and the *Sequential Forward Selection* (SFS) methods are applied. The nBest selection method picks the n individually best features (with regards to a performance function) from the feature subset. The SFS method, by contrast, is a bottom-up search procedure where one feature is added at a time to the current feature set.

The feature to be added is selected from the subset of the remaining features so that the new feature set generates the maximum value of the performance function over all candidate features for addition [7].

The SFS method is used since it has been successfully applied in a wide variety of feature selection problems yielding high performance values with minimal feature subsets: see [34], for example, for further discussion and application to the classification problem of process identification in resistance spot welding. On the other hand, the nBest method is used for comparative purposes being the most popular technique for feature selection. Features selected by each method constitute the input vector of the evolving ANN. The feature selection procedure followed here evaluates the usability of each one of the features available and obtains the minimal feature subset that performs best in the classification between games reported as entertaining and games reported as non-entertaining (see Section V).

To evaluate the performance of each feature subset the available data is randomly divided into training and validation data sets consisting of 2/3 and 1/3 of the data respectively. The performance of an ANN model is measured through the average classification accuracy of the ANN in three independent runs using the leave-one-out cross-validation technique on the training and validation data sets. Since we are interested in the minimal feature subset that yields the highest performance we terminate the feature selection procedure (nBest or SFS) when an added feature yields equal or lower validation performance than the performance obtained without it.

A. Single Feature Performance

The experiment presented here tests the validation performance of single individual player and game features. Given the selected feature (ANN input), ANNs are evolved by following the approach presented in Section V-A and evaluated through the leave-one-out cross-validation method (see Section VI). The training and validation performance of each of the individual player and game features are presented in Table I where features are ranked by validation performance.

The impact of the recorded response times (r_t) is demonstrated in Table I; both the average and the variance of these values generate the highest cross-validation performances (see also [6] for the impact of $E\{r_t\}$ on reported entertainment in the Bug-Smasher game). Results obtained show the incapability of a single feature to successfully model reported entertainment in Bug-Smasher. Given that the best performed feature ($E\{r_t\}$) yields a cross-validation performance of 62.22% it becomes apparent that more features are required to effectively model children's notion of entertainment. Moreover, it appears that results presented in Table I are consistent with the correlates of reported entertainment presented in Section IV-A. In fact, three out of four best features of Table I yield statistically significant effects on reported entertainment (see Section IV-A).

TABLE I

TRAINING AND VALIDATION PERFORMANCE OF INDIVIDUAL PLAYER AND GAME FEATURES. $E\{r_t\}$ AND $\sigma^2\{r_t\}$ IS THE AVERAGE AND THE VARIANCE OF THE RESPONSE TIME RESPECTIVELY; $\sigma^2\{D_b\}$ IS THE VARIANCE OF THE DISTANCES BETWEEN THE PRESSED TILE AND THE BUGS APPEARING ON THE GAME; N_I IS THE TOTAL NUMBER OF INTERACTIONS; H IS THE QUANTITATIVE MEANS FOR THE GAME CONTROLLABLE FEATURE OF CURIOSITY; $E\{p\}$ IS THE AVERAGE PRESSURE FORCE RECORDED FROM THE FSR SENSOR; H_C IS THE ENTROPY OF THE TILES THAT THE CHILD VISITED; $\sigma^2\{p\}$ IS THE VARIANCE OF THE PRESSURE FORCES RECORDED FROM THE FSR SENSOR; $E\{D_b\}$ IS THE AVERAGE DISTANCE BETWEEN THE PRESSED TILE AND THE BUGS APPEARING ON THE GAME; S IS THE QUANTITATIVE MEANS FOR THE GAME CONTROLLABLE FEATURE OF CHALLENGE AND P IS THE TOTAL NUMBER OF SMASHED BUGS.

Feature	Training Performance (%)	Validation Performance (%)
$E\{r_t\}$	69.47	62.22
$\sigma^2\{r_t\}$	67.91	61.11
$\sigma^2\{D_b\}$	68.54	56.67
N_I	66.36	56.67
H	66.04	55.56
$E\{p\}$	64.17	53.33
H_C	59.81	53.33
$\sigma^2\{p\}$	66.73	51.11
$E\{D_b\}$	65.42.	51.11
S	43.93	46.67
P	65.42	43.33

B. More Features: Selection Method Comparison

This section presents experiments for finding the minimal feature subset that yields the highest classification performance in matching the ANNs output with children's reported answers on entertainment in unknown data (validation data set). For this purpose, the two feature selection methods described in Section VI are applied and compared. The initial subset (ANN input) for both methods includes the feature that performs best in the single feature experiment: $E\{r_t\}$. ANNs are evolved by following the approach presented in Section V-A. The data is partitioned in training (2/3 of total data) and validation (1/3 of total data) portions and the leave-one-out cross-validation technique is used to obtain the classification performance of the ANNs.

Table II presents the above-mentioned comparative study between nBest and SFS. SFS appears to generate feature subsets that yield higher validation performance than feature subsets generated by nBest. The best cross-validation performance (77.77%; average of 70%, 73.33% and 90%) is achieved when the ANN input contains $E\{r_t\}$, $\sigma^2\{p\}$, H and N_I. The binomial-distributed probability of this performance to occur at random is 0.0019 demonstrating statistical significance and providing evidence for this solution's robustness. Note that, challenge is absent from the obtained feature subset indicating that the spatial diversity of the bugs (curiosity) has a higher impact on children's reported

entertainment than the speed of the game (challenge).

Difficulties in obtaining higher classification accuracy are found in experimental noise in both the recorded features and the children's answers on self reports. Even though comparative fun analysis is a reliable and established method for capturing reported entertainment in computer [18] and mixed-reality [6] games, it generates a significant amount of uncertainty in subjects' reported answers. Uncertainty appears when the two games played are not significantly different with regards to the entertainment value they generate for the player and therefore cannot be distinguished.

<div align="center">TABLE II</div>

VALIDATION PERFORMANCE OF INDIVIDUAL PLAYER AND GAME FEATURES AND THEIR RESPECTIVE BINOMIAL DISTRIBUTED P-VALUES. $E\{r_t\}$ AND $\sigma^2\{r_t\}$ IS THE AVERAGE AND THE VARIANCE OF THE RESPONSE TIME RESPECTIVELY; $\sigma^2\{p\}$ IS THE VARIANCE OF THE PRESSURE FORCES RECORDED FROM THE FSR SENSOR; $\sigma^2\{D_b\}$ IS THE VARIANCE OF THE DISTANCES BETWEEN THE PRESSED TILE AND THE BUGS APPEARING ON THE GAME; H IS THE QUANTITATIVE METRIC OF CURIOSITY AND N_I IS THE TOTAL NUMBER OF INTERACTIONS.

| Features	nBest	p_{nBest}	Features	SFS	p_{SFS}
$E\{r_t\}$	62.22	0.1270	$E\{r_t\}$	62.22	0.1270
$\sigma^2\{r_t\}$	58.88	0.2179	$\sigma^2\{p\}$	67.77	0.0400
$\sigma^2\{D_b\}$	44.44	0.2551	H	68.88	0.0307
N_I	46.67	0.4277	N_I	**77.77**	**0.0019**
H	52.22	0.4759	$\sigma^2\{r_t\}$	63.33	0.1002

For reasons of space, only the feature subset $\{E\{r_t\}, \sigma^2\{p\}, N_I, H\}$ with the highest validation performance (90.00%, in one of the three learning attempts) is presented in this paper. Note that, the qualitative features of the surfaces plotted in Fig. 2 appeared in all three different learning attempts of the cross-validation procedure for this feature subset.

Fig. 2 illustrates the trained ANN output with regards to $\sigma^2\{p\}$ and N_I for six points in the $(E\{r_t\}, H)$ search space. These values constitute the combinations of two $E\{r_t\}$ states (0 and 1 named *Fast* and *Slow* respectively), and the three states used for H (0.33, 0.66 and 1 named *Low*, *Average* and *High* respectively). The above presentation helps towards interpreting the mapping between $\sigma^2\{p\}$, N_I and reported entertainment according to how fast children react with the playground and the level of curiosity.

As seen from Fig. 2, fast children ($E\{r_t\} = 0$) appear to enjoy average and high curiosity values except when high N_I values are combined with low values of $\sigma^2\{p\}$ (see Fig. 2(e) and Fig. 2(f)). Fast children's preference for low levels of curiosity is met only when their behavior combines low values of N_I and high values of $\sigma^2\{p\}$ (see Fig. 2(d)). On the other hand, slow children appear to prefer low curiosity levels except when the N_I value they generate is low and combined with either very high or very low $\sigma^2\{p\}$ values (see Fig. 2(a)). Average curiosity levels are preferred by slow children in many fewer cases; that is when their N_I value
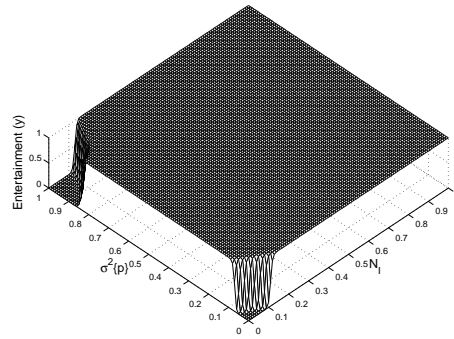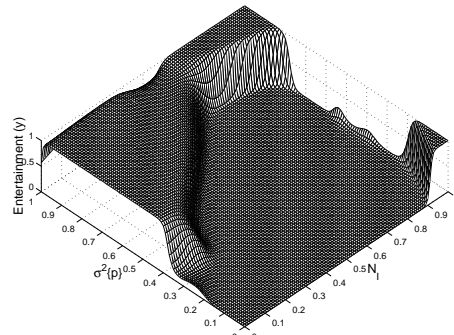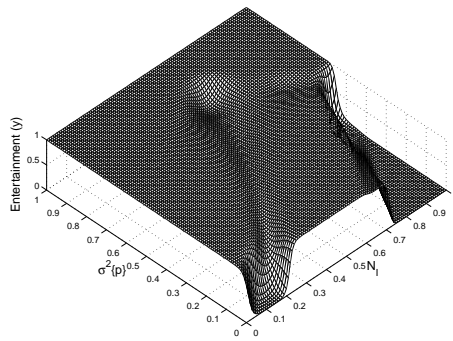

(a) $E\{r_t\} = 1.0$ (Slow), $H = 0.33$ (Low)
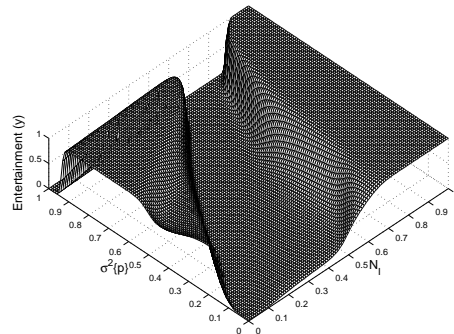
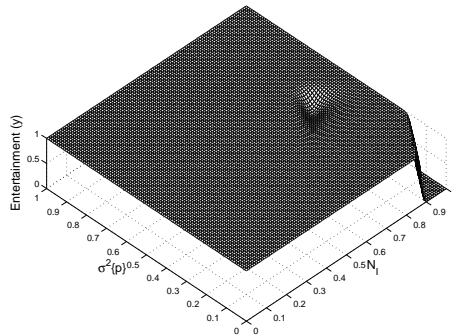(b) $E\{r_t\} = 1.0$ (Slow), $H = 0.66$ (Average)
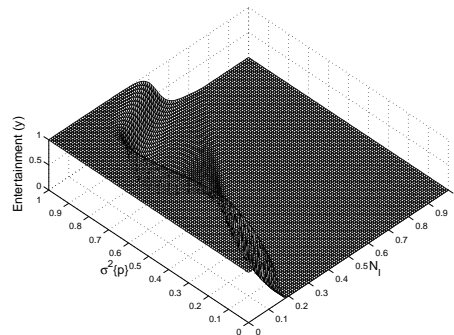
(c) $E\{r_t\} = 1.0$ (Slow), $H = 1.0$ (High)

(d) $E\{r_t\} = 0.0$ (Fast), $H = 0.33$ (Low)

<div align="center">**249**</div>

(e) $E\{r_t\} = 0.0$ (Fast), $H = 0.66$ (Average)

(f) $E\{r_t\} = 0.0$ (Fast), $H = 1.0$ (High)

Fig. 2. The trained ANN ($f = 95.85$) that yields the highest validation performance (90.00%): ANN output y (entertainment value) with regards to $\sigma^2\{p\}$ and N_I for all six combinations of two $E\{r_t\}$ states (Slow, Fast) and three H states (Low, Average, High).

is low and $\sigma^2\{p\}$ value is high or when their N_I value is high and $\sigma^2\{p\}$ value is low (see Fig. 2(b)). Finally, high curiosity is rarely preferred by slow children and this occurs only when their N_I values are low independently of their $\sigma^2\{p\}$ value (Fig. 2(c)).

The obtained effects of curiosity in reported entertainment are consistent, in part, with previous studies on the Bug-Smasher game [6]. In that study the relation between challenge, curiosity and average response time was reported through a lower scale experiment of 28 children. It was found that fast children liked games independently of curiosity whereas children reacting slowly with the playground preferred games of low curiosity levels.

VII. CONCLUSIONS & DISCUSSION

This paper introduced feature selection methods for obtaining minimal feature subsets that successfully model children's notion of entertainment through the Bug-Smasher game played on the Playware playground. More specifically, the nBest and the SFS feature selection methods were applied and compared demonstrating the ability of SFS in finding feature subsets that yield higher validation performance.

The fittest ANN solution presented derives from a feature subset of four features: $\{E\{r_t\}, \sigma^2\{p\}, N_I, H\}$. Experi-

ments with additional features (inputs of the ANN) could not improve the model's validation performance. This model manages to map between children's average response time, the variance of their force pressure on the tiles, the number of interactions with the playground, the game feature of curiosity and the children's notion of gameplay entertainment with a cross-validation accuracy of 77.77% (binomial-distributed p-value = 0.0019). The main reason for not obtaining a higher cross-validation performance appears to be the experimental noise existent in the self-reports designed for comparative entertainment (fun) analysis. Moreover, the learned mapping between $\{E\{r_t\}, \sigma^2\{p\}, N_I, H\}$ and children's notion of entertainment showed that, in general, fast responding children show a preference for high curiosity games whereas slow responding children tend to prefer games of low curiosity. The obtained results are consistent with previous work on the impact of the factors of challenge and curiosity and the average response time in Playware games [6].

The main limitation of the proposed approach lies within the complexity of entertainment as a mental state. The generated y value cannot be regarded as a mental affective state approximator but as a correlate of expressed children's preferences on entertainment. However, this correlate serves the purposes of this work well as far as entertainment modeling is concerned. In addition, Malone's entertainment factor of fantasy is omitted from the results in this paper since the focus is on the contribution of the opponent behaviors to the generation of entertainment; however, fantasy's positive impact on reported entertainment has been reported in a previous study [28].

Even though the comparative fun protocol (2-AFC) used serves well the purpose of this work, a 4-alternative forced choice (4-AFC) approach is considered for future experiment protocol design. Children will be able to choose among the following alternatives: one game is more "fun" than the other (2-AFC), both games are equally "fun", neither game was "fun". This protocol would provide more information for the machine learning process and eliminate the noise generated by 2-AFC.

The entertainment modeling approach presented here demonstrates generality over the majority of action games created with Playware since the quantitative measures of challenge and curiosity are estimated through the generic features of speed and spatial diversity of the opponent on the game's surface. Thus, these or similar measures could be used to adjust player satisfaction in any future game development on the Playware tiles. However, each game demonstrates individual entertainment features that might need to be extracted and added on the proposed measures and therefore, more games of the same and/or other genres need to be tested to cross-validate this hypothesis.

The proposed approach can be used for adaptation of the game opponents (e.g. bugs) according to the player's individual playing style, based on reaction time, recorded pressure on tiles and amount of interactions, and as far as the curiosity factor of entertainment is concerned. Given the

real-time average response time of a child, the variance of his/her pressure forces on the tiles and the number of times he/she interacts with the environment, the partial derivative of the model output $\vartheta y/\vartheta H$ can be used to appropriately adjust the level of entropy (curiosity) of the opponent for the entertainment value y to be augmented. Such a direction constitutes an example of future work on Playware and computer games. The level of engagement or motivation of the user/player/gamer of such interactive environments may be identified and augmented by the use of the presented approaches.

REFERENCES

[1] G. N. Yannakakis and J. Hallam, "A scheme for creating digital entertainment with substance," in *Proceedings of the Workshop on Reasoning, Representation, and Learning in Computer Games, 19th International Joint Conference on Artificial Intelligence (IJCAI)*, August 2005, pp. 119–124.

[2] H. H. Lund, T. Klitbo, and C. Jessen, "Playware technology for physically activating play," *Artifical Life and Robotics Journal*, vol. 9, no. 4, pp. 165–174, 2005.

[3] T. W. Malone, "What makes computer games fun?" *Byte*, vol. 6, pp. 258–277, 1981.

[4] N. Postman, *The Disappearance of Childhood*. London: Allen, 1983.

[5] S. Kline, *Out of the Garden: Toys and Children's Culture in the Age of Marketing*. Verso, 1993.

[6] G. N. Yannakakis, H. H. Lund, and J. Hallam, "Modeling Children's Entertainment in the Playware Playground," in *Proceedings of the IEEE Symposium on Computational Intelligence and Games*. Reno, USA: IEEE, May 2006, pp. 134–141.

[7] P. Devijver and J. Kittler, *Pattern Recognition - A Statistical Approach*. Engelwood cliffs, NJ: Prentice-Hall, 1982.

[8] M. Csikszentmihalyi, *Flow: The Psychology of Optimal Experience*. New York: Harper & Row, 1990.

[9] P. Sweetser and P. Wyeth, "GameFlow: A Model for Evaluating Player Enjoyment in Games," *ACM Computers in Entertainment*, vol. 3, no. 3, July 2005.

[10] N. Lazzaro, "Why we play games: Four keys to more emotion without story," XEO Design Inc.," Technical Report, 2004.

[11] R. Koster, *A Theory of Fun for Game Design*. Paraglyph Press, 2005.

[12] J. Read, S. MacFarlane, and C. Cassey, "Endurability, engagement and expectations," in *Proceedings of International Conference for Interaction Design and Children*, 2002.

[13] P. Vorderer, T. Hartmann, and C. Klimmt, "Explaining the enjoyment of playing video games: the role of competition," in *ICEC conference proceedings 2003: Essays on the future of interactive entertainment*, D. Marinelli, Ed. Pittsburgh: Carnegie Mellon University Press, pp. 107–120.

[14] D. Choi, H. Kim, and J. Kim, "Toward the construction of fun computer games: Differences in the views of developers and players," *Personal Technologies*, vol. 3, no. 3, pp. 92–104, September 1999.

[15] H. Iida, N. Takeshita, and J. Yoshimura, "A metric for entertainment of boardgames: its implication for evolution of chess variants," in *IWEC2002 Proceedings*, R. Nakatsu and J. Hoshino, Eds. Kluwer, 2003, pp. 65–72.

[16] G. N. Yannakakis and J. Hallam, "Evolving Opponents for Interesting Interactive Computer Games," in *From Animals to Animats 8: Proceedings of the 8th International Conference on Simulation of Adaptive Behavior (SAB-04)*, S. Schaal, A. Ijspeert, A. Billard, S. Vijayakumar, J. Hallam, and J.-A. Meyer, Eds. Santa Monica, LA, CA: The MIT Press, July 2004, pp. 499–508.

[17] ——, "A Generic Approach for Obtaining Higher Entertainment in Predator/Prey Computer Games," *Journal of Game Development*, vol. 1, no. 3, pp. 23–50, December 2005.

[18] G. N. Yannakakis, "AI in Computer Games: Generating Interesting Interactive Opponents by the use of Evolutionary Computation," Ph.D. thesis, University of Edinburgh, November 2005.

[19] G. N. Yannakakis and J. Hallam, "Towards Capturing and Enhancing Entertainment in Computer Games," in *Proceedings of the 4th Hellenic Conference on Artificial Intelligence, Lecture Notes in Artificial Intelligence*, vol. 3955. Heraklion, Greece: Springer-Verlag, May 2006, pp. 432–442.

[20] G. N. Yannakakis, J. Hallam, and H. H. Lund, "Capturing Entertainment through Heart-rate Dynamics in the Playware Playground," in *Proceedings of the 5th International Conference on Entertainment Computing, Lecture Notes in Computer Science*, vol. 4161. Cambridge, UK: Springer-Verlag, 2006, pp. 314–317.

[21] P. Rani, N. Sarkar, and C. Liu, "Maintaining optimal challenge in computer games through real-time physiological feedback," in *Proceedings of the 11th International Conference on Human Computer Interaction*, 2005.

[22] S. McQuiggan, S. Lee, and J. Lester, "Predicting User Physiological Response for Interactive Environments: An Inductive Approach," in *Proceedings of the 2nd Artificial Intelligence for Interactive Digital Entertainment Conference*, 2006, pp. 60–65.

[23] G. Andrade, G. Ramalho, H. Santana, and V. Corruble, "Extending reinforcement learning to provide dynamic game balancing," in *Proceedings of the Workshop on Reasoning, Representation, and Learning in Computer Games, 19th International Joint Conference on Artificial Intelligence (IJCAI)*, August 2005, pp. 7–12.

[24] M. A. Verma and P. W. McOwan, "An adaptive methodology for synthesising Mobile Phone Games using Genetic Algorithms," in *Congress on Evolutionary Computation (CEC-05)*, Edinburgh, UK, September 2005, pp. 528–535.

[25] R. Hunicke and V. Chapman, "AI for Dynamic Difficulty Adjustment in Games," in *Proceedings of the Challenges in Game AI Workshop, 19th Nineteenth National Conference on Artificial Intelligence (AAAI'04)*, 2004.

[26] P. Spronck, I. Sprinkhuizen-Kuyper, and E. Postma, "Difficulty Scaling of Game AI," in *Proceedings of the 5th International Conference on Intelligent Games and Simulation (GAME-ON 2004)*, 2004, pp. 33–37.

[27] R. Pfeifer and C. Scheier, *Understanding Intelligence*. Cambridge, MIT Press, 1999.

[28] G. N. Yannakakis, J. Hallam, and H. H. Lund, "Comparative Fun Analysis in the Innovative Playware Game Platform," in *Proceedings of the 1st World Conference for Fun 'n Games*, 2006, pp. 64–70.

[29] R. L. Mandryk, K. M. Inkpen, and T. W. Calvert, "Using Psychophysiological Techniques to Measure User Experience with Entertainment Technologies," *Behaviour and Information Technology (Special Issue on User Experience)*, vol. 25, no. 2, pp. 141–158, 2006.

[30] C. Beal, J. Beck, D. Westbrook, M. Atkin, and P. Cohen, "Intelligent modelling of the user in interactive entertainment," in *Proceedings of the AAAI Spring Symposium on Artificial Intelligence and Interactive Entertainment*, Stanford, 2002, pp. 8–12.

[31] X. Yao, "Evolving artificial neural networks," in *Proceedings of the IEEE*, vol. 87, no. 9, 1999, pp. 1423–1447.

[32] J. H. Holland, *Adaptation in Natural and Artificial Systems*. Ann Arbor, MI: University of Michigan Press, 1975.

[33] D. J. Montana and L. D. Davis, "Training feedforward neural networks using genetic algorithms," in *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence (IJCAI-89)*. San Mateo, CA: Morgan Kauffman, 1989, pp. 762–767.

[34] E. Haapalainen, P. Laurinen, P. Junno, H. Tuovinen, and J. Roening, "Methods for classifying spot welding process: A comparative study of performance," in *IEA/AIE*, 2005, pp. 412–421.