

User Model User-Adap Inter
DOI 10.1007/s11257-007-9036-7

ORIGINAL PAPER

Entertainment capture through heart rate activity in physical interactive playgrounds

Georgios N. Yannakakis · John Hallam · Henrik Hautop Lund

Received: 29 June 2006 / Accepted in revised form: 18 May 2007
© Springer Science+Business Media B.V. 2007

Abstract An approach for capturing and modeling individual entertainment (“fun”) preferences is applied to users of the innovative Playware playground, an interactive physical playground inspired by computer games, in this study. The goal is to construct, using representative statistics computed from children’s physiological signals, an estimator of the degree to which games provided by the playground engage the players. For this purpose children’s heart rate (HR) signals, and their expressed preferences of how much “fun” particular game variants are, are obtained from experiments using games implemented on the Playware playground. A comprehensive statistical analysis shows that children’s reported entertainment preferences correlate well with specific features of the HR signal. Neuro-evolution techniques combined with feature set selection methods permit the construction of user models that predict reported entertainment preferences given HR features. These models are expressed as artificial neural networks and are demonstrated and evaluated on two Playware games and two control tasks requiring physical activity. The best network is able to correctly match expressed preferences in 64% of cases on previously unseen data (p -value $6 \cdot 10^{-5}$). The generality of the methodology, its limitations, its usability as a real-time feedback mechanism for entertainment augmentation and as a validation tool are discussed.

G. N. Yannakakis (✉) · J. Hallam · H. H. Lund
Maersk Mc-Kinney Moller Institute, University of Southern Denmark,
Campusvej 55, Odense M, 5230, Denmark
e-mail: georgios@mmmi.sdu.dk

J. Hallam
e-mail: john@mmmi.sdu.dk

H. H. Lund
e-mail: hhl@mmmi.sdu.dk

Keywords Entertainment modeling · Biosignals · Intelligent interactive playgrounds · Mixed-reality games · Artificial neural networks

1 Introduction

Cognitive modeling within human–computer interactive systems is a prominent area of research. Computer games, as examples of such systems, provide an ideal environment for research in artificial intelligence (AI), because they are based on simulations of highly complex and dynamic multi-agent worlds (Champanand 2004; Funge; Laird and van Lent 2000). Moreover, computer games offer a promising ground for cognitive modeling since they embed rich forms of interactivity between humans and non-player characters (NPCs). Being able to estimate the quantitative level of user (gamer) engagement or satisfaction in real-time can grant insights to the appropriate AI methodology for enhancing the quality of playing experience (Yannakakis and Hallam 2005) and furthermore be used to adjust digital entertainment environments according to individual user preferences.

Features of computer games that keep children (among others) engaged more than other digital media include 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 (Malone 1981). On the other hand, traditional playgrounds offer the advantage of physical play, which furthermore improves the child's health condition, augments children's ability to engage in social and fantasy play (Kline 1993; Postman 1983) and provides the freedom for children to generate their own rules for their own developed games. The 'Playware' intelligent interactive physical playground introduced in Lund et al. (2005) attempts to combine the aforementioned features of both worlds: computer games and traditional playgrounds. This innovative platform is described and experiments on developed Playware games with children will be presented in this paper.

Motivated by the lack of quantitative affective models of entertainment in physical play, an approach for estimating expressed player satisfaction in real-time through physiological signals measured during gameplay is introduced in the work presented here. We define "entertainment" primarily as the level of satisfaction generated by the real-time player-game opponent interaction—by 'opponent' we mean any controllable interactive feature of the game. This approach is based on the notion (Yannakakis and Hallam 2004; Koster 2005) that a game is primarily a learning process and the level of entertainment is kept high when game opponents enable new learning patterns ('*not too easy a game*') for the player that can be perceived and learned by the player ('*not too difficult a game*'). On the same basis, according to Kapoor et al. (2001)—within the axis of emotions varying from boredom to fascination—learning is highly correlated to interest, curiosity and intrigue perceived. We define the collection of these emotions as "entertainment" (or "fun," which is used synonymously in this paper). The word "fun" is used extensively hereafter since it captures best, in our view, children's notion of the term "entertainment" (Read et al. 2002) and is the term used by the children when making their experimental self-reports.

Our principal goal in the reported work is to construct a user model of the player of a game—in this case a child playing a Playware game—that can predict the answers to which variants of the game are more or less “fun.” In this work the model is constructed using machine learning techniques applied to statistical features derived from physiological signals measured during play. The output of the constructed model is a real number in the range $[0, 1]$ such that more enjoyable games receive higher numerical output. This basic approach, defined as *entertainment modeling*, is applicable to a variety of games, both computer (Yannakakis and Hallam 2006) and physical (Yannakakis et al. 2006a), using features derived from physiological data or from the interaction of player and opponent measured through game parameters.

Even though entertainment is a highly complicated mental state it is correlated with sympathetic arousal (Mandryk et al. 2006a,b) which can be captured through specific physiological signals such as heart rate and skin conductivity, as reported by researchers in the psychophysiological research field (Critchley et al. 2005; Zuckerman 2006). In this paper, we investigate the impact of entertainment on heart rate (HR) signals and attempt to capture HR signal features that correlate with children’s expressed entertainment preferences. HR signal data and children’s reported preferences between variants of Playware games are obtained through gaming experiments using the Playware playground. HR dynamics are represented by calculating several statistics and regression model parameters from gameplay experimental data, to serve as features for the construction of a user model as described above.

Statistical correlations between HR signal features and player reported entertainment preferences are first computed. Results demonstrate that the average HR, the maximum HR, the difference between the maximum and the minimum HR and the approximate entropy (Pincus 1991) are highly rank-correlated with children’s reported preferences between Playware game variants.

Using the statistical representation of HR dynamics as features, neuro-evolution models are trained and validated on those HR features and children’s self-report preference data to yield Artificial Neural Networks (ANNs) which function as efficient predictors of reported entertainment preferences given suitable specific HR signal features. Suitable input feature sets are constructed using two alternative feature selection schemes, the performances of which are compared.

The HR dynamics features found to correlate with self-reported entertainment preferences are ones that also correlate with physical activity. (This is unsurprising, since one would expect a more enjoyable game to induce greater physical effort from the player.) However, it is then unclear whether the constructed models distinguish more and less enjoyable games on the degree rather than the kind of physical activity they engender. To control for this, two kinds of non-entertaining physical activity were also tested.

The first exploratory experiment is focused on the distinction between HR signals corresponding to reported entertainment preferences in a gaming activity (entertaining or not) and HR signals corresponding to pure (non-game, and non-entertaining) physical activity. Obtained results show that HR dynamics can be used to construct models (of the kind described above) that discriminate well between entertaining game activities and physical exercise. However, the question of whether there is anything in the type of physical activity that is characteristic of an entertaining game remains

since the particular control physical activity employed (running in circles), while not entertaining, is also quite different in character to the more variable exertion of a typical Playware game.

A second experiment was therefore designed, following a suggestion from an anonymous reviewer, in which the control physical activity matched game activity more closely while nevertheless being non-entertaining, and in which children expressed preferences both between game variants and between game and control activity. ANN models trained on game-play data obtained from the first set of experimentation were then evaluated using unseen data from this second game-play and control experiment set. The results indicate that user models able to predict children's preferred game variants given suitable HR dynamics feature representations can indeed be constructed and that such models not only distinguish game-play from game-like non-entertaining physical activity but also generalize (to some extent) over children's individual preferences.

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.

2 Capturing entertainment

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 gameplay (Malone 1981), namely challenge (i.e. ‘provide a goal whose attainment is uncertain’), curiosity (i.e. ‘what will happen next in the game?’) and fantasy (i.e. ‘show or evoke images of physical objects or social situations not actually present’) as well as the well-known concepts of the theory of flow (‘flow is the mental state in which players are so involved in the game that nothing else matters’) (Csikszentmihalyi 1990) incorporated in computer games as a model for evaluating player enjoyment, namely *GameFlow* (Sweetser and Wyeth 2005) and captured through automatic speech recognition in a tutoring systems (D’Mello et al. 2008). Other studies include Lazzaro's “fun” clustering into four entertainment factors based on facial expressions and data obtained from game surveys on players (Lazzaro 2004). According to Lazzaro, the four components of entertainment are: hard fun (related to the challenge factor of Malone), easy fun (related to the curiosity factor of Malone), altered states (i.e. ‘the way in which perception, behavior, and thought combine in a collective context to produce emotions and other internal sensations’—closely related to Malone's fantasy factor) and socialization (the people factor). Koster's (2005) 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” measure is presented in Read et al. (2002) where “fun” is composed of three dimensions: endurance, engagement and expectations. Questionnaire tools and methodologies are proposed in order to empirically capture the level of “fun” for evaluating the usability of novel interfaces with children.

The current state-of-the-art in intelligent game design by the use of AI techniques is mainly focused on generating human-like (Laird and van Lent 2000) and intelligent

(Nareyek 2002) characters. Even though complex opponent behaviors emerge through various machine learning techniques, there is no further analysis of whether or what these behaviors contribute to player experience. The implicit hypothesis motivating this research is that intelligent opponent behaviors enable the player to gain more satisfaction from the game. According to Taatgen et al. (2003), believability of computer game opponents, which are generated through cognitive models, is strongly correlated with enjoyable games. These hypotheses may well be true; however, since no notion of entertainment or enjoyment has been explicitly defined, there is no evidence that a specific opponent behavior generates enjoyable games. This statement is the core of Iida's innovative work on quantitative entertainment metrics for variants of chess games (Iida et al. 2003).

Recognition of emotion, such as entertainment, is considered, in general, a hard problem mainly because understanding emotion is hard (Picard et al. 2001). Previous work in the field of *quantitative* estimation of entertainment in digital interactive entertainment systems 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 (Yannakakis and Hallam 2004). Based on this fundamental assumption, a quantitative metric for measuring the real time entertainment value of predator/prey games using statistics of the player–game interaction was proposed and was established as an efficient and reliable entertainment (‘interest’) metric by validation against human judgment (Yannakakis 2005). Further studies by Yannakakis and Hallam (2006) have shown that ANNs and fuzzy neural networks can extract a better estimator of player entertainment than a custom-designed one, given appropriate estimators of the challenge and curiosity of the game (Malone 1981) and data on human players' reported preferences. Such estimators constitute models of the player, in so far as they correctly predict the player's expressed preferences for games with respect to variations of game parameters that determine the player–opponent interaction.

Similar work addressing adjusting a game's difficulty to alter player satisfaction include endeavors through reinforcement learning (Andrade et al. 2005), genetic algorithms (Verma and McOwan 2005), probabilistic models (Hunicke and Chapman 2004) and dynamic scripting (Spronck et al. 2004). However, these attempts assume that challenge is the only factor contributing to enjoyable gaming experiences, and the results reported have not been cross-verified against human players. Moreover, within the edutainment and web design framework, heuristic evaluation methods have been proposed for usability testing of entertainment web sites (Wiberg 2003) and game design guidelines have been discussed for the enhancement of game experience and learnability in edutainment games (Wiberg and Jegers 2003). The level of student interest has also been captured through machine learning in educational tutoring systems (Porayska-Pomsta et al. 2008).

Following the theoretical principles reported from Malone (1981), Lazzarro (2004), Koster (2005) and Yannakakis (2005), this paper examines the contribution to entertainment of game opponents' behavior. We argue that, of the three dimensions of “fun” (endurability, engagement, expectations) defined in Read et al. (2002), it is only *engagement* that is affected by the opponent since both *endurability* and *expectations* are primarily based on the game design per se. Given a successful interactive game

design that yields expectations and endurance (such as the Playware platform—see Sect. 5) we only focus on the level of engagement that generates “fun” (entertainment).

Moreover, rather than being based on empirical or visual observations of children’s engagement, the work presented here attempts to model the relationship between physiological signals (i.e. HR signals) and the degree to which children are enjoying a game, measured by experimental data from surveys with children playing variants of two dissimilar games with the Playware playground (see Sect. 3). Previous work (Yannakakis et al. 2006b) has already shown a significant correlation between the average HR and reported entertainment preference in a Playware action game (Bug-Smasher). This paper expands this research in that an additional Playware game (Space-Invaders) is used for data collection and reported entertainment preferences are captured through an ANN modeling approach evolved on HR signal features.

2.1 Physiology of entertainment

Philosophical discussion of the interplay between emotions and their respective bodily changes was initiated by Aristotle, who introduced the concept of physiology as an affect of emotional experience (see (Aristotle 2004) among others). As a variant of this view, James (1992) proposed the idea that emotions are feelings specifically caused by changes in physiological conditions relating to autonomic and motor functions.

Following from these fundamental concepts, heart physiological condition measures have been used extensively for emotion recognition in children and adults within the affective computing research area. More specifically, use of HR and HR variability (HRV) have resulted in effective discrimination between children’s exploration, problem-solving and play tasks (Hutt 1979). Experiments with two-year old children further showed suppression of HRV during exploration, and solution of a puzzle, suggesting that the task demands for these two activities were greater than those during play (Hughes and Hutt 1979).

Picard et al. (2001) recorded the physiological state of a single adult subject over six weeks during which period the subject was asked to express several emotions (*joy* included). Picard et al. observed day-dependence in the subject’s physiology, mainly due to day-to-day variations of caffeine and sugar concentrations in blood, sleep, hormones and mood. The main weakness of that approach lies in generalizing obtained affect recognition results to other subjects. Nevertheless, the study reveals that substantial changes in the relationship between affect and physiological state can occur from day to day.

Day-dependence and methodological conditions in capturing and classifying emotions when using physiological signal data raised by that study are satisfied in the work described in this paper. Specifically, the second control experiment described satisfies the day-dependence conditions and all experiments described meet the majority of the five factors for eliciting genuine emotion as presented in Picard et al. (2001): reported entertainment was *subject-elicited*, the experiments took place in a setup close to the *real-world* since children played in their school classroom, our emphasis was on internal *feelings*, subjects were implicitly aware that their physiology was being recorded (*open-recording*) and subjects were not aware of the purpose of the experi-

ment (*other-purpose*). The study presented here is not focused on the investigation of the long-term realistic physiology of children with regards to entertainment but rather the construction of a predictor of reported entertainment based on individual HR signal features. Moreover, children cannot be asked to express emotions in the same way as adults and therefore physiological effects on the emotion of, for instance, *joy* (Picard et al. 2001) cannot be compared straightforwardly with findings presented here.

Mandryk et al. (2006b) examine the correlations between physiological signals (galvanic skin response (GSR), electromyography (EMG) in jaw, respiration and cardiovascular measures) in reported adult user experiences (emotions) in computer games. The experiences examined include boredom, challenge, frustration and “fun.” Statistical analysis yields a significant correlation only between GSR and reported “fun” in video games. The study also demonstrates the advantage of social play versus playing against computer-designed opponents. In Mandryk et al. (2006a), a fuzzy model with rules grounded in psychophysiology theory shows that high arousal and positive valence (a combination corresponding to “fun” and excitement) is present when HR and GSR are high and the EMG corresponds to a smiling player. In that study, playing condition (vs. computer, friend or a stranger) demonstrates an additional effect on “fun” and excitement. The model is validated against subjects’ reported “fun” through correlation statistical tests; however, it provides an objective notion of “fun” based on the relations of the data obtained to subjects’ expressed preferences.

The preliminary experiments of Rani et al. (2005) for appropriately adjusting the level of challenge in the game of ‘Pong’ based on recorded physiological signals in real-time and subject’s self-reports of their emotional experiences during gameplay is closely related to our work. That study, however, is primarily focused on anxiety level detection in real-time and is limited by the number of human participants. Physiological state (HR, GSR) prediction models have also been proposed for potential entertainment augmentation in computer games (McQuiggan et al. 2006).

Working on the same basis as Mandryk et al. (2006a,b), Ravaja et al. (2006) examined whether the nature of the game opponent (computer vs. human and friend vs. stranger) influences the physiological state of players. In that study, self-report ratings, cardiac inter-beat intervals and facial EMG were measured to index physiological arousal and emotional valence. In addition, Hazlett’s (2006) work is focused on the use of facial EMG to distinguish positive and negative emotional valence during interaction with a racing video game. Within the edutainment framework, a community of researchers has been using physiological signals to model engagement of children playing educational games ((Conati 2002; Conati et al. 2003) among others).

All aforementioned studies are focused on the use of physiology for capturing user experiences (e.g. “fun,” engagement or excitement) applied within the computer and edutainment games framework. The work reported here is novel in that it examines the physiological state (HR signal) correlates of reported “fun” in *physical* activity games, isolates HR signal features attributed to reported entertainment in such physically demanding games and demonstrates a way of constructing a subjective model (a predictor of user preferences) of reported “fun” grounded in statistical features of HR signal dynamics. While in general one might expect entertaining games to increase arousal, the more specific effect investigated here is that entertaining *physical* games motivate one to play more energetically with consequent increase in HR activity.

3 Playware playground

Children's and youth's play has seen major changes during the last two decades. New emerging playing technologies, such as computer games, have been more attractive to children than traditional play partly because of the interactivity and fantasy enhancement capabilities they offer (Malone 1981). These technologies have contributed to transforming the way children spend their leisure time: from outdoor or street play to play sitting in front of a screen (Lindstrom and Seybold 1994; Turkle 1984). This sedentary style of play may have health implications (Lund and Jessen 2005).

A new generation of playgrounds that adopt technology met in computer games may address this issue. More specifically, intelligent interactive playgrounds with the ability to adapt the game according to each child's personal preferences provide properties that can keep children engaged in entertaining physical activity. Measuring the child's level of entertainment and adjusting the game in order to increase it, in such a situation, can only have positive effects on the child's physical condition. The Playware playground adopts these primary concepts.

The Playware (Lund and Jessen 2005) mixed-reality prototype playground consists of a set of identical tiles that allow the game designer (e.g. the child) to develop a significant number of different games within the same platform. For instance, tiles can be placed on the floor or on the wall in different topologies to create a new game (Lund et al. 2005). Tiles incorporate processing power, communication, input and output, and function as a distributed multi-agent system. They are 21 cm × 21 cm by 6 cm deep, with four Light Emitting Diodes (LEDs) and a Force Sensing Resistor (FSR) sensor through which the tile and child interact. A rubber shell, that covers and protects the electronics, includes a "bump" indicating the location of the FSR sensor (i.e. the interaction point) and a plexiglass window for the LEDs (see Figs. 1 and 2).



Fig. 1 A child playing Bug-Smasher



Fig. 2 A child playing Space-Invaders

3.1 Systems related to playware

The Smart Floor (Orr and Abowd 2000) and the KidsRoom (Bobick et al. 1996) are among the few systems that are conceptually similar to the Playware tiles. The first is developed for transparent user identification and tracking based on a person's footstep force features and the latter is a perceptually-based, multi-person, fully automated, interactive, narrative play room that adjusts its behavior (story-line) by analyzing the children's behavior through computer vision. As far as the concept of intelligent floors consisting of multiple building blocks is concerned, the Z tiles (Richardson et al. 2004) are closely related to Playware. However, the Z tiles are mainly used as input devices only whereas Playware incorporates both input and output devices. Finally, the *Scorpiodome* game system (Metaxas et al. 2005) is consistent with the Playware mixed (virtual and physical) reality and grid/tiles concept; however, it is primarily designed for investigating social, non-physical gaming—being a platform for deploying augmented reality games with remote controlled toy vehicles—and outside the principles of embedded AI within its *Active Landscape Grid* components/tiles.

Digiwall (Liljedahl and Lindberg 2006) and *Age Invaders* (<http://www.mixedreality.nus.edu.sg>) are mixed-reality systems closely related to the Playware playground. The first features a rock climbing wall with several hand grips that can be lit and sensors that indicate when each grip is grabbed. The latter is a mixed-reality game that uses a fusion of technology and art to make a new form from the traditional arcade game space invaders in a tile-based floor with which children and elderly can interact. *Age Invaders* is designed as a social physical game which promotes the elderly's mental and physical vitality. Similarities within the interactive physical game (*Exertion Interface*) field can also be found with the physically exhausting ball game by Mueller et al. (2003).

Other examples of digitally augmented platforms include the *STARS* platform for tabletop games (Magerkurth et al. 2003) and *PingPongplus* (Ishii et al. 1999) which

is a digitally enhanced ping-pong game using a table that incorporates sensing, sound and visual projection. However, the first lacks the physical activity context and the latter lacks the rich interactivity and adaptivity offered by Playware games.

Exergaming products such as *QMotions* and Konami's (<http://www.konami.com>) *Dance Dance Revolution* and *Kick & Kick* series of games constitute examples of entertainment media that mix physical activity with computer game playing. However, Playware offers the concept of building block game development to users as well as to game designers, providing a much higher degree of freedom and flexibility in game design. On that basis, the aforementioned games could be regarded as a sub-class of the games that can be designed with Playware.

4 Playware games

This section presents the two games designed with Playware tiles for our experiments and concludes with a comparison of their main features which collectively cover a large proportion of action interactive games designable within Playware.

4.1 Bug-Smasher

The first test-bed game used for the experiments presented here is called 'Bug-Smasher.' The game is implemented on a 6×6 square tile topology (see Fig. 1). During the game, different 'bugs'—colored lights—appear and disappear sequentially on the game surface, each 'bug' being visible for a short period of time. A bug's position is picked within a radius of three tiles of the previous bug's position. The new bug's position, among the candidate tiles of this radius, is selected according to a predefined level of the bugs' spatial diversity defined through a probability distribution of the bug visits on these tiles. Spatial diversity is measured by the entropy of the bug-visited tiles which is calculated and normalized into $[0, 1]$ via (1):

$$H = \left[-\frac{1}{\log 36} \sum_i \frac{v_i}{V} \log \left(\frac{v_i}{V} \right) \right] \quad (1)$$

where v_i is the number of bug-visits to tile i and V is the total number of visits to all visited tiles (i.e. $V = \sum_i v_i$). If the bug visits all tiles equally then $v_i = V/36$ for all 36 tiles and H will be 1; if the bug visits exactly one tile during the game, H is zero.

The child's goal is to smash as many bugs as possible by stepping on the lighted tiles. Different (animal/bug) sounds and colors represent different bugs appearing and being smashed in order to increase the fantasy entertainment factor (Malone 1981). Previous studies have shown that fantasy in Bug-Smasher has an objectively positive impact on reported "fun" (Yannakakis et al. 2006c). Feedback to the player, which is essential for a successful game design (Malone 1981), is provided through different characteristic sounds that represent good or bad performance. The Bug-Smasher game has already been used as a test-bed for obtaining HR signal correlates of reported "fun" in children (Yannakakis et al. 2006b), for investigating the impact of fantasy on

reported “fun” (Yannakakis et al. 2006c) and for modeling reported “fun” according to individual playing features obtained through the child–game interaction (Yannakakis et al. 2006a).

4.2 Space-Invaders

Given the available degrees of freedom when designing new games within the Playware playground, we designed a game that would differ from the Bug-Smasher game in its conceptual content while still being an action game. The second test-bed game used for the experiments presented here is called ‘Space-Invaders,’ its main concept being based on the classical arcade game released by Taito in 1978.

The game is implemented on a 5×10 tile topology (see Fig. 2). During the game, different alien spaceships (colored lights) appear on the top side of the game topology and move towards the bottom row of the game where the player’s loaded ‘guns’ are placed. The child’s goal is to shoot the alien spaceships down by firing at them. A shot is fired by pressing the lighted tiles indicating the guns. If the child is unsuccessful in that task and a spaceship reaches the bottom row of the game then a particular sound is heard indicating that the child’s “base” is hit.

A spaceship’s initial position is picked according to a predefined level of position spatial diversity, given by the entropy of the spaceship-visited tiles calculated and normalized into $[0, 1]$ via (1). Having 10 tiles as initial spaceship positions, the normalization parameter in (1) for this game is $1/\log 10$ instead of $1/\log 36$. The movement strategy of the spaceships is based on a probabilistic choice among the three neighbor tiles (forward/down, left, right) at each step, with fixed probabilities.

4.3 Properties: Bug-Smasher vs. Space-Invaders

These two Playware action games used differ in several respects. Based on the games’ main features, we believe that these two test-beds cover a large portion of the properties met in playground action games that can be designed within Playware or other interactive physical game systems.

The primary dissimilarities between Bug-Smasher and Space-Invaders are found in:

- *The child’s objectives and learned skills:* In Bug-Smasher the child has to react fast by killing (stepping on) bugs that appear in the game in unknown patterns. Therefore, pattern recognition and reaction time are the primary skills that are enhanced by this game. On the other hand, in Space-Invaders, the child has to develop both fast reaction and timing (prediction) skills for shooting the spaceships down.
- *The type of opponent (bug or spaceship) motion:* While bugs appear given a specific predefined pattern based on their spatial diversity, spaceships move based on a probability estimation of their neighbor tiles. Therefore, it is interaction points (game opponents) that appear and disappear (bugs) versus interaction points that move within the game platform (spaceships).

- *The topology*: Different tile topologies are used for game design: square against rectangle topology in Bug-Smasher and Space-Invaders respectively.
- *The playing style*: The different game designs generate diverse playing styles. More specifically, Bug-Smasher allows free movement within the game platform whereas Space-Invaders constrains the child’s play to one side of the game platform.
- *The concept*: Space-Invaders’ concept is arcade games inspired and based on shooting enemies in a virtual space world. On the other hand, Bug-Smasher projects a virtual animal world where children have to interact with its virtual bugs.

5 Experimental setup

Both Bug-Smasher and Space-Invaders test-beds have been used to acquire HR physiological data and data on children’s judgment of entertainment. Each game uses two different states (‘Low’ and ‘High’) of [Malone’s \(1981\)](#) three entertainment factors to generate a pool of 8 dissimilar games for children to play.

We consider the speed (S —in s^{-1}) that the bugs appear and disappear from the game and their spatial diversity (H) in the game field as appropriate measures of the level of challenge and the level of curiosity (unpredictability) respectively ([Malone 1981](#)) during gameplay for the Bug-Smasher game. Likewise, for the Space-Invaders game, challenge and curiosity are represented by the speed that the spaceships move from tile to tile and the probability p of moving to left or right from their current position, respectively. For both games, the former provides a notion of a goal whose attainment is uncertain—the higher the S value, the higher the goal uncertainty and furthermore the higher the challenge—and the latter effectively portrays a notion of unpredictability in the subsequent events of the game—the higher the H or p value the less predictable is the bug’s or spaceship’s behavior and therefore the higher the curiosity. Finally, the level of fantasy corresponds to the number of different types of bugs/spaceships (different colors and sounds) appearing in the game.

Fifty six normal-weighted (based on their body mass index) children whose ages covered a range between 8 and 10 years participated in an experiment. The 56 children were split into two groups of 28 children and each group was assigned to play either Bug-Smasher or Space-Invaders. The Bug-Smasher experiment has already been reported in previous work ([Yannakakis et al. 2006a,b,c](#)) while the Space-Invaders experiment is first reported here.

By experimental design, each subject plays against two of the selected game states of either Bug-Smasher or Space-Invaders in all permutations of pairs, thus, $C_2^8 = 28$ is the required number of children to cover all combinations of 2 out of 8 game variants for each game. More specifically, each subject plays two games (A and B) of either Bug-Smasher or Space-Invaders—differing in the levels of one or more entertainment factors of challenge, curiosity and fantasy—for 90s each. Each time a pair of games (‘game pair’) is finished, the child is asked whether the first game was more “fun” than the second game (see ([Read et al. 2002](#)) for the word terminology used for comparative “fun” experiments with children). Children are not interviewed but are asked to

fill in a questionnaire, minimizing the interviewing effects reported in [Mandryk et al. \(2006b\)](#).

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 effect of order of game playing on children's judgment of entertainment is presented in Sect. 5.1.

The playing time window chosen (90 s) is a compromise between effective data collection (long enough child–game interaction to support a relative fun judgment) and not overstretching children. Note that children's time is a valuable commodity and we attempt to exploit this to the maximum. Moreover, Playware games are physical activity games and children playing are asked to play four games in a row (6 min in total of energetic physical play with short rest breaks) which is a significant period of energetic physical play.

All subjects are given the same instructions by an experimenter who is unaware of the purpose of the experiment. No further oral or eye-contact communication during experiment tasks and questionnaire is present, minimizing experimenter expectancy effects ([Rosenthal 2003](#)). For the design of the children's self-reports we follow the principles of comparative “fun” analysis presented in [Read et al. \(2002\)](#) and [Yannakakis \(2005\)](#). The endurability and expectations for the majority of children who played both Playware games were very high, indicating that the game design used for both games was successful. Specifically, all children were excited to play with Playware as soon as they were informed about the rules of the games (determined using a *Funometer* tool application [[Read et al. 2002](#)]) and the majority of children were enthusiastic to play the game again (assessed using an Again-Again table [[Read et al. 2002](#)]).

To capture the children's preferences for the game variants played, we use a 2-alternative forced choice (2-AFC) approach 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 HR dynamics that generalizes over the reports of different children a pairwise preference approach (2-AFC) is preferred to a ranking approach ([Mandryk et al. 2006b](#)). Forcing the choice of children generates experimental noise, in that the child may have no significant preference for one or other of the game variants played yet must nevertheless express a preference; however, insignificant order effects provide evidence that the experimental noise generated in this way is random (see Sect. 5.1).

The HR of the children is recorded in real-time using a wireless ElectroCardioGram (ECG) device (POLAR s610i) consisting of pulse sensors placed on the chest of the child. The RR intervals (time duration between two consecutive R waves) recorded are automatically converted into HR with an accuracy of ± 1 heart beat per minute (bpm). Heart rates are calculated and stored every 5 s. Out of the total number of 224 games played in this experiment, in 82 games (41 game pairs) and 96 games (48 game pairs) for the Bug-Smasher and the Space-Invaders respectively the HR signal was properly recorded. In the remaining 46 games, HR data was lost because of hardware failure. Malfunction of the device's features (electrodes and data transmission) during the game was the main cause of the recording failure. While a loss of data from 20% of

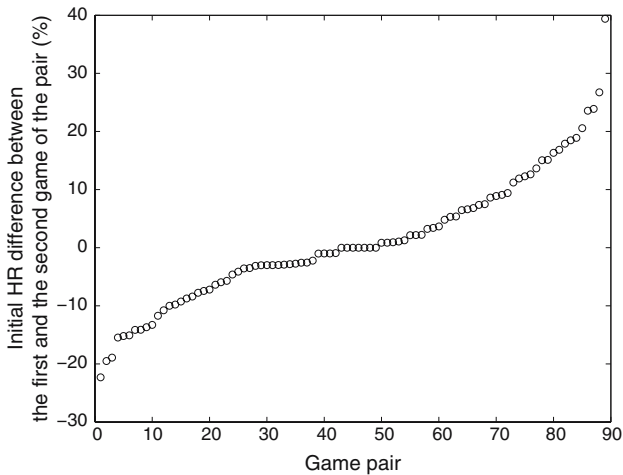


Fig. 3 Difference of the initial heart rates for the two games played by each child (178 games in total, 89 game pairs)

the experimental games is substantial, there is no reason to suppose that the hardware failure has any particular bias with respect to experimental hypothesis.

The set of HR time series collected from the 89 correctly recorded pair of games underlies the analysis presented in this paper (see Fig. 3). Note that children played all their assigned games on the same day, mitigating day-dependence effects on their physiology (Picard et al. 2001). Cultural differences in the impact of affect on physiology may also be present but are not examined here.

5.1 Order effects

To avoid any order effect of playing on the HR signal we allow each child 1 min to rest in between the two games. Figure 3 shows that the resting time has a positive impact on minimizing the percent difference between the initial heart rates in the two games played d_I . More specifically, the majority (60 out of 89) of d_I values lie between $\pm 10\%$. However there are 12 cases (out of 89) where $d_I < -10\%$, that is where child is less aroused at the start of the second game than the first, and 17 cases where the period of 1 min in between the games was not enough for the child's arousal to drop to the level it was just before the first game started (i.e. $d_I > 10\%$). A t -test for means of paired samples demonstrates no significant difference in the initial HR between the two games of the pair ($t = -0.5691$, p -value = 0.2853).

To check whether the order of playing Playware games affects the children's judgment of entertainment, we hypothesize that there is no order effect and proceed as follows. For each subject that played a pair of games in both orders, we count the times K and J that the subject chooses the first and the second game respectively as more entertaining in both pairs. In the case where the subject chooses the same game in both pairs played, we take no action. The test statistic used to assess the truth of

the hypothesis that there is no order effect is given by $z(K, J) = (K - J)/N$, where $N = 56$. The greater the absolute value of $z(K, J)$, the more the order of play appears to affect the subjects' reported entertainment preference. The obtained z value equals -0.1607 and its corresponding (Trinomial distribution) p -value equals 0.1058 . Therefore, the order effect null hypothesis is not rejected and we conclude that the order of play does not significantly affect the children's judgment of entertainment. (While a 10% probability is not especially high, this result is consistent with a number of other similar studies we have conducted that show no significant order effect (Yannakakis 2005; Yannakakis et al. 2006c; Yannakakis and Hallam 2007a); see also the second set of experiments reported in Sect. 9.1.2 where the order effect null hypothesis probability is approximately 20%.)

6 Entertainment within Playware games

In this section, the HR signals obtained from the experiment for both Playware games (see Sect. 5) are presented and analyzed. Specifically, an approach to processing the HR signal and a statistical analysis of the correlation between the HR signal's features and reported entertainment preferences is outlined.

6.1 HR signal processing

Given the HR time series over 90s for each game played, the following statistical parameters are computed in order to represent features of the HR dynamics that may correlate with expressed entertainment preferences of the child while playing a game.

- The average HR $E\{h\}$.
- The variance of the HR signal $\sigma^2\{h\}$.
- The maximum HR $\max\{h\}$.
- The minimum HR $\min\{h\}$.
- The difference D between the maximum and the minimum HR.
- The correlation coefficient R between HR recordings and the time t in which data were recorded. This parameter provides a notion of the linearity of the signal (HR data) over time.
- The autocorrelation ρ_1 (lag equals 1) of the signal, which is used to detect the level of non-randomness in the HR data.
- The approximate entropy ($ApEn^{m,r}$) (Pincus 1991) of the signal which quantifies the unpredictability of fluctuations in the HR time series.

$ApEn$ is a standard cardiovascular metric, which is closely related to the HRV metric since HRV represents the variability in the fluctuation between successive heart beats (Pincus and Goldberger 1994). A more regular heart beat will result in a lower HRV than an irregular heart beat. Likewise, high regularity in HR signals results in low $ApEn$ values while low HR signal regularity yields high $ApEn$ values. $ApEn$ is based on the natural logarithm of the conditional probability that two patterns that are similar for m points will still be similar within a tolerance r at the next point. On that basis, HR signals that demonstrate repetitive similar patterns tend to have relatively small

ApEn values whereas more complicated and less structured signals tend to have higher *ApEn* values.

According to Pincus (1991), to compute the approximate entropy, *ApEn*, of a HR time series $h(q)$, $q = 1, 2, \dots, N_q$, the series of vectors of length m , $v(q) = [h(q), h(q + 1), \dots, h(q + m - 1)]^T$ is derived. The distance $d(i, j)$ between two vectors $v(i)$ and $v(j)$ is defined as the maximum difference in the scalar components of $v(i)$ and $v(j)$. Then the number ($N_v^{m,r}(i)$) of vectors j such that the distance between vectors $v(j)$ and the vector $v(i)$ is lower than r ($d(i, j) < r$) is computed. The probability to find a vector which differs from $v(i)$ less than the distance (tolerance) r , $C^{m,r}(i)$ is defined as $C^{m,r}(i) = N_v^{m,r}(i)/(N_q - m + 1)$ and the logarithmic average over all vectors of the $C^{m,r}(i)$ probability is given by (2).

$$F^{m,r} = \frac{\sum_{i=1}^{(N_q - m + 1)} \ln(C^{m,r}(i))}{N_q - m + 1} \quad (2)$$

ApEn is given by (3).

$$ApEn^{m,r} = F^{m,r} - F^{m+1,r} \quad (3)$$

Thus *ApEn* of a time series $h(t)$ measures the logarithmic likelihood that runs of patterns of length m that are close to each other will remain close in the next incremental comparisons, $m + 1$. The *ApEn* fixed parameter values $m = 2$ and $r = 0.2\sigma\{h(t)\}$ used in this paper are recommended by the literature (Pincus and Goldberger 1994).

As mentioned in Sect. 2.1, the use of HRV as a cardiovascular metric has been extensive in affective computing. The use of HRV in this paper might have revealed changes in the frequency bands of the RR interval time series which would correspond to changes attributed to entertainment versus changes attributed to physical activity. However, HRV could not be calculated given the data provided by the available recording apparatus since the detected RR intervals are opaquely converted into HR estimates by the wireless POLAR s610i ECG device used.

In addition, three different regression models were used to fit (in a least square sense) the HR signal: linear, quadratic and exponential. The regression model parameters under investigation are:

- The slope s of the linear regression ($h_L(t) = st + \alpha$) on the signal. Linear least squares fitting is used.
- The parameters β and γ of the quadratic regression ($h_Q(t) = \beta t^2 + \gamma t + \delta$) on the signal which respectively quantify the curvature and the rotation angle with respect to the x -axis of the quadratic curve. QR factorization is used to fit the data.
- The parameters A , B and b of the exponential regression $h_E(t) = A(1 - e^{bt}) + B$. A quantifies a notion of difference between initial and maximum HR; b portrays the slope of this difference over time and B corresponds to an estimate of the initial HR. A trust-region reflective Newton optimization algorithm is used to fit the data.

The disadvantage of the aforementioned regression model parameters against the HR signal statistics is that they cannot be computed in real-time. The choice of the specific

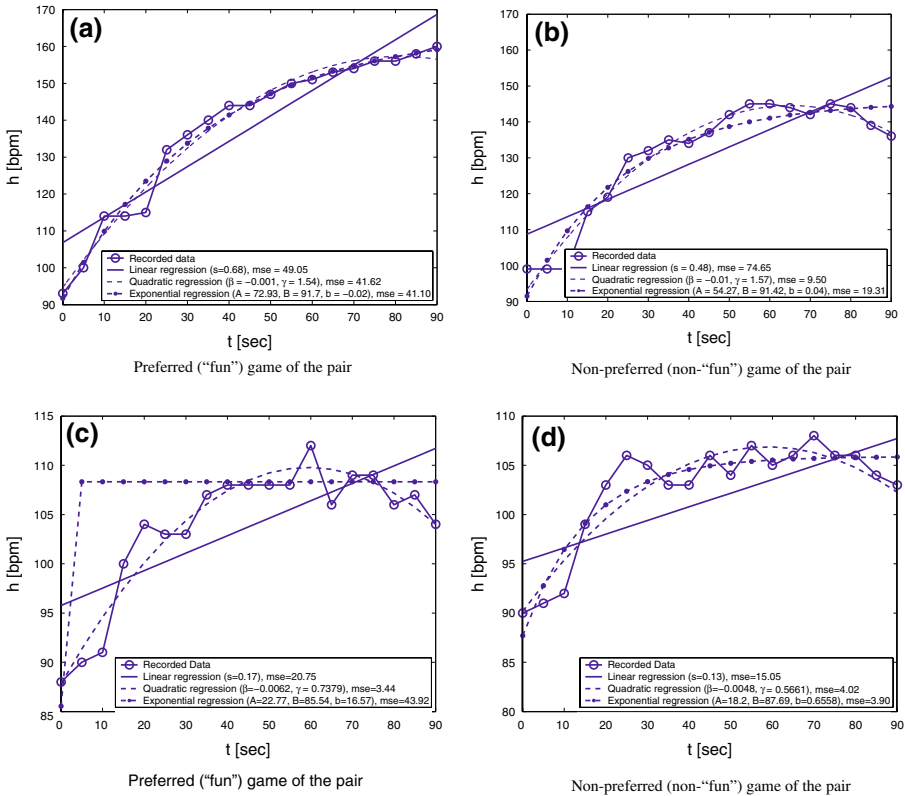


Fig. 4 Example of the HR signals and their corresponding regression lines in two pairs of game: A subject (subject no. 18) plays a game of *High* challenge and curiosity (a) and then a game of *Low* challenge and curiosity (b). Another subject (subject no. 21) plays a game of *High* challenge and *Low* curiosity (c) and then a game of *Low* challenge and *High* curiosity (d)

regression models was made in order to cover a decent amount of the HR dynamics which, in their majority, seem to follow the exponential function. The $h_E(t)$ function gave the smallest standard error of fitting in 95.55% of HR signal samples (see Fig. 4(a) and (d)). On the other hand Fig. 4(b) and (c) correspond to examples of HR signals where exponential regression failed compared to the other regression models. Linear and quadratic regression were applied so as to obtain additional features such as the average slope and the curvature of the signal.

Figure 4 illustrates an example of the HR data recorded from two different children playing a pair of Bug-Smasher games and the lines produced by the three regression models. The left graphs of Fig. 4 correspond to the game selected by the child as more entertaining of the two. For reasons of space, we present the HR dynamics of only two game pairs. Note that the qualitative features of the signals are similar for all children that played both Bug-Smasher and Space-Invaders: an initial rapid increase of HR during the first seconds of the game followed by a stable, but noisy, condition of high HR values.

6.2 Statistical analysis

The aim of the statistical analysis presented here is to identify statistically significant correlations between children's notions of entertainment and any of the aforementioned HR signal features. For this purpose the following null hypothesis is formed: the correlation between children's reported judgment of entertainment and the examined parameters obtained from the recorded HR signal of the child, as far as the different games played are concerned, is a result of randomness. The test statistic is obtained through

$$c(\vec{z}) = \frac{1}{N} \sum_{i=1}^N z_i \quad (4)$$

where N (89, in this paper) is the total number of game pairs where HR signals were properly recorded, $z_i = 1$ if the subject chooses as the more entertaining game the one with the larger value of the examined HR signal feature and $z_i = -1$ if the subject chooses the other game in game pair i .

Table 1 presents the $c(\vec{z})$ values and their corresponding p -values for all above-mentioned HR signal features for each of the two games independently and combined. Average HR ($E\{h\}$), maximum HR ($\max\{h\}$), D and $ApEn$ appear to be the features examined that are significantly—significance equals 5%, high significance equals 1% in this paper—correlated to reported entertainment preferences. The obtained effects of $E\{h\}$, $\max\{h\}$, D appear to be commonsensical since both Playware games belong to the genre of action physical games where the level of engagement of the user tends to have a significant effect on physical activity. The significant correlation of $E\{h\}$ with entertainment preferences for the Bug-Smasher experiment alone has already been reported in previous work (Yannakakis et al. 2006b).

In addition, Table 1 shows that the only feature correlated with entertainment preferences in both games and in their combined HR data is $E\{h\}$. Space-Invaders seems to have a greater impact on the combined significant effects of $\max\{h\}$ and D which are due to the different type of physical effort demanded by the dissimilar games. Bug-Smasher appears to be a demanding physical game that may in some variants be too hard for some children at the end of 90-s time window. This may account for why children's maximum HR ($\max\{h\}$) and the difference between their minimum and maximum HR (D) appear less correlated with entertainment preferences in Bug-Smasher than the respective HR signal features, generated through the Space-Invaders game.

On the other hand, Space-Invaders demands lower physical effort because of the longer time interval in which the game's main goal is to be accomplished. For instance, if a bug appears for t seconds in Bug-Smasher, a spaceship appears for $4t$ seconds in Space-Invaders while randomly moving from the top row of the game down to its base (4 tiles of translation). As previously mentioned in Sect. 4.3, in the first case the child has to learn to react fast (greater physical effort) whereas in the latter case, the child has to learn to plan and predict fast (greater mental effort).

Table 1 Correlation coefficients and their corresponding p -values between reported entertainment preferences and HR signal features

	Feature	Bug-Smasher		Space-Invaders		Total	
		$c(\vec{z})$	p -value	$c(\vec{z})$	p -value	$c(\vec{z})$	p -value
Statistics	$E\{h\}$	0.41	<u>0.0057</u>	0.29	<u>0.0297</u>	0.34	<u>0.0006</u>
	$\sigma^2\{h\}$	-0.02	0.5000	0.20	0.0967	0.10	0.1982
	$\max\{h\}$	0.17	0.1744	0.37	<u>0.0066</u>	0.28	<u>0.0052</u>
	$\min\{h\}$	-0.02	0.5000	-0.12	0.2354	-0.07	0.2625
	D	-0.02	0.5000	0.50	<u>0.0003</u>	0.28	<u>0.0052</u>
	R	-0.02	0.5000	-0.08	0.3327	-0.05	0.3359
	ρ_1	0.12	0.2663	0.20	0.0967	0.16	0.0686
	$ApEn$	0.12	0.2663	0.25	0.0557	0.19	<u>0.0446</u>
$h_L(h)$	s	0.07	0.3776	0.16	0.1561	0.12	0.1445
$h_Q(h)$	β	0.07	0.3776	-0.25	0.0557	-0.16	0.0686
	γ	-0.12	0.2663	0.33	<u>0.0146</u>	0.12	0.1445
	$h_E(h)$	A	0.07	0.3776	0.20	0.0967	0.14
	B	-0.07	0.3776	-0.25	0.0557	-0.16	0.0686
	b	0.02	0.5000	0.00	0.5572	0.01	0.5000

Significant p -values (at 5% or better) are tabulated in underlined bold, thus **0.01**. The column named ‘Total’ corresponds to the combined data of both games

If we hypothesize that in such games the level of engagement has a linear monotonic relation with average HR ($E\{h\}$) and/or the maximum HR ($\max\{h\}$) and/or the difference between the maximum and the minimum HR (D) it appears that the higher these features’ values during a game the higher the child’s perceived entertainment and the higher his/her physical activity. Moreover, the significant effect of the $ApEn$ value demonstrates that games that generate more complicated and less structured HR signals tend to generate higher entertainment for the child.

The reported significant correlation between the average response time of children interacting with Playware games and entertainment preferences (Yannakakis et al. 2006a) is consistent with the interplay between engagement, physical activity and reported entertainment preferences demonstrated here. Moreover, it has been shown (see Yannakakis et al. 2006a) that the challenge and curiosity entertainment factors are not rank correlated with reported entertainment and only ‘average’ levels of game speed and opponents’ spatial diversity appear to generate entertaining games. This projects the personalization of “fun” for children and the individual requirements (neither too small nor too large) they have on entertainment features such as challenge and curiosity for a game to be appealing.

A further analysis of the only statistically significant correlation between HR signal features and reported entertainment obtained from both games and in total—that is $E\{h\}, c(\vec{z}) = 0.3483$ —is presented here. A t -test for means of paired samples between $E\{h\}$ values corresponding to preferred (*entertaining*) games and non-preferred (*non-entertaining*) games demonstrates significant difference ($t = -2.6417, P(T \leq t) = \mathbf{0.0048}$).

7 Learning entertainment preferences

The statistical analysis presented in Sect. 6.2 demonstrated a significant rank correlation only between $E\{h\}$ and reported entertainment preferences for both Bug-Smasher and Space-Invaders games. This section examines the hypothesis that a non-linear approach (ANN) can yield a better predictor of children's entertainment preferences.

The proposed approach to entertainment modeling is based on selecting a minimal subset (see Sect. 7.1) of HR signal features and constructing a quantitative user model that predicts the children's entertainment preferences. For this purpose, a fully-connected feedforward ANN for learning the relation between the selected HR signal 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, which models the child's internal response to playing the game, that is, how much fun it is, is an unknown function of HR signal 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 is no differentiable output error function for the learning problem, ANN training algorithms such as back-propagation are inapplicable. Learning is achieved through artificial evolution (Yao 1999) and is described in Sect. 7.2.

The sigmoid function is employed at each neuron, the connection weights take values from -5 to 5 and all input values are normalized into $[0, 1]$ before they are entered into the ANN. In an attempt to minimize the controller's size, it was determined that ANN architectures with two hidden layers, containing five hidden neurons each, are 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.

7.1 Feature selection

Two different input feature set selection schemes are used and compared in this paper. Given the HR signal features presented in Sect. 6.1, 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 (Devijver and Kittler 1982).

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 Haapalainen et al. (2005), 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 approximation to the feature subset that performs best in the classification between preferred games, non-preferred games and physical (non-game) exercise control tasks (see Sect. 7.2).

In order 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.

7.2 Genetic algorithm

A generational genetic algorithm (GA) (Holland 1975) is implemented, which uses an evaluation function that measures the difference between the children's reported preferences of entertainment (assuming that pure physical exercise generates lower levels of entertainment than the games classified by children as entertaining) and the model output value y . In the algorithm presented here, the ANN topology is fixed and the GA chromosome is a vector of ANN connection weights.

The evolutionary procedure used can be described as follows. A population of N (1000, in this paper) networks is initialized randomly. Initial real values that lie within $[-5, 5]$ are picked randomly, from a uniform distribution, for their connection weights. Then, at each generation:

Step 1 Each member (neural network) of the population is given three n_i -tuple (where n_i is the number of selected HR signal features) values one for A , one for B and one for E and returns three output values, namely $y_{j,A}$ (entertainment value of the preferred game), $y_{j,B}$ (entertainment value of the non-preferred game) and $y_{j,E}$ (entertainment value of the physical activity control task—see Sect. 8) for each triple j (two games played and one exercise task) available. When the $y_{j,A}$, $y_{j,B}$ ANN output values are consistent with the reported preference of subject j (i.e. $y_{j,A} > y_{j,B}$) or when the $y_{j,A}$ and $y_{j,E}$ values are consistent with our assumption that the physical activity control task yields lower levels of entertainment than the preferred game (i.e. $y_{j,A} > y_{j,E}$) then we state that: 'the values agree with the subject' or that there is 'agreement' with the subject. In the opposite case, we state that: 'the values disagree with the subject' or there is 'disagreement.' Note that no ordering is assumed between the less-preferred game and the physical control task.

Step 2 Each member i of the population is evaluated via the fitness function f_i :

$$f_i = \frac{1}{2N_s} \sum_{j=1}^{N_s} \left\{ g(d_j^{AB}, \epsilon) + g(d_j^{AE}, \epsilon) \right\} \quad (5)$$

where $d_j^{AB} = y_{j,A} - y_{j,B}$, $d_j^{AE} = y_{j,A} - y_{j,E}$, N_s is the number of training game pairs, $g(d_j, p) = 1/(1 + e^{-pd_j})$ is the sigmoid function and $\epsilon = 30$ if there is agreement and $\epsilon = 5$ if there is disagreement. Both the sigmoidal shape of the fitness function and its selected ϵ values are inspired by its successful application in previous classification attempts of reported entertainment preferences of children playing with Playware (Yannakakis et al. 2006a).

Step 3 A fitness-proportional scheme is used as the selection method.

Step 4 Selected parents clone an equal number of offspring or reproduce offspring by crossover so that the total population reaches N members. The Montana and Davis (1989) crossover operator is applied with a probability 0.4.

Step 5 Gaussian mutation occurs in each gene (connection weight) of each offspring's genome with a small probability $p_m = 1/n$, where n is the number of genes.

The algorithm is terminated when either a good solution ($f > 0.9$), a large number of generations g has been completed ($g = 10,000$) or when the classification accuracy on a randomly selected data set consisting of 10% of the training data begins to drop. The last termination condition is equivalent to the early stopping technique used in conjunction with backpropagation training for over-fitting avoidance.

8 Preliminary physical activity experiment

Results presented in Sect. 6.2 have shown that $E\{h\}$, $\max\{h\}$ and D are HR signal features that correlate with reported entertainment. Indeed, these are features that also correspond to physical activity and therefore the main conclusion to be drawn is that the more engaging the gameplay, the higher the physical activity (through the aforementioned HR features) and the higher the perceived entertainment for a child.

This raises the question of whether the statistical effects observed genuinely reflect *entertainment* value or merely the tendency of more engaging games to elicit more physical activity. That is, is there anything in the type of physical activity that is characteristic of an enjoyable game or is the analysis just comparing amount of physical activity?

This section presents a preliminary set of physical control experiments and results obtained through the proposed learning methodology in brief. Detailed analysis of this set of experiments can be found in Yannakakis and Hallam (2007b).

8.1 Experiment description

To investigate the interplay between entertainment and physical activity we asked all children that played Playware games to participate in an additional experiment: each child was asked to run around a $3\text{ m} \times 3\text{ m}$ space for 90 s. HR signal recordings were obtained during this exercise task. The assumption here is that this exercise (physical activity control) task is a non-entertaining activity for the child. This assumption was supported by most children asking for the time remaining during the task, suggesting a certain level of boredom for the activity. However, children were not asked whether the running task was “fun” or not and it was not compared to any physical game task.

Further details on the experimental protocol used can be found in (Yannakakis and Hallam 2007b).

Children conscientiously followed the rules set by the experimenter and ran continuously for 90 s. Running generated higher $E\{h\}$, $\max\{h\}$, and D values than the corresponding values for Playware gameplay since it required more demanding physical activity from the child (Yannakakis and Hallam 2007b).

8.2 Feature selection

Given the selected features (ANN input) and the 150 pairs of comparisons (75 pairs of preferred/non-preferred game and 75 pairs of preferred game/exercise task) ANNs are evolved by following the approach presented in Sect. 7.2 and evaluated using leave-one-out cross-validation (see Sect. 7.1). The two feature selection methods described in Sect. 7.1 are applied and compared. The initial subset (ANN input) for both methods includes the feature that performs best in the single feature experiment: $ApEn$ (Yannakakis and Hallam 2007b).

The best cross-validation performance (80.66%; average of 88%, 78% and 76%) is achieved when the ANN input contains $ApEn$ and $E\{h\}$. More HR signal features added in the feature subset do not yield significantly higher classification accuracy (Yannakakis and Hallam 2007b).

The obtained classification accuracy demonstrates the existence of an ANN model that successfully predicts the children's reported entertainment preferences given a child's individual HR signal features: $ApEn$ and $E\{h\}$. However, difficulties in obtaining higher classification accuracy are found because of experimental noise in both the recorded features and the children's answers in self-reports. Even though comparative "fun" analysis is a reliable and established method for capturing entertainment preferences in computer (Yannakakis and Hallam 2007a) and mixed-reality (Yannakakis et al. 2006a) games, it generates a certain 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. In this circumstance, players appear to express a random preference. This 'dilutes' the data in which genuine preferences are expressed from the point of view of the machine learning algorithm.

HR signals obtained through this preliminary control experiment show that the running task appears to involve much more physical effort (high $E\{h\}$ values) than the physical effort required in a Playware game, and further that the physical effort involved is different in kind (low $ApEn$ values; high regularity of the HR signal). It follows that the exercise control experiment may generate HR dynamics rather easy to separate from game-play HR dynamics, and allows one to distinguish entertaining game-play from exercise purely on the artificial basis of the kind of physical activity taking place. It is therefore, in retrospect, not a good control for physical activity effects. Detailed experiment analysis can be found in Yannakakis and Hallam (2007b).

Consider, though, that one cannot control completely for physical activity effects, since the games being played are *physical* games whose differing enjoyability may naturally be expected to result in different degrees of physical engagement by the player.

A better control, therefore, would provide game-like—but non-entertaining—physical activity. Section 9 proposes a more appropriate experiment, based on a suggestion from one of the anonymous reviewers, for controlling and isolating the elements of physical activity from an HR signal so that features of HR signal corresponding to entertainment become more apparent.

9 Controlled physical activity experiment

The question of whether there is anything in the type of physical activity that is characteristic of an entertaining game remains since the first (exploratory) experiment for controlling for physical activity effects through recorded HR signals suffered from the fact that the required physical activity in game and non-game tasks was artificially different. Therefore, we designed a new experiment where the physical activity control is achieved through a non-entertaining variant of the Bug-Smasher game named the “Stomping game.”

This section describes the experimental protocol used and presents a neuro-evolution model for classifying between entertaining and not entertaining game physical activity through HR signal features. The new ANN model is trained on data obtained from the previously described experiment (data set 1)—excluding HR signal data obtained from the running task—exactly as described in the foregoing text. The model is then evaluated using unseen data from the physical activity control experiment described here, to determine the extent to which the constructed ANN user model generalizes. Data from this experiment is referred to as “data set 2” in the sequel.

9.1 Experiment description

For the new experimental protocol, we asked 18 naïve normal-weighted children (9 boys and 9 girls) aged 8–10 years to play five games each on the Playware platform. The games were played on the same day, satisfying the day-dependence condition of the physiological state (Picard et al. 2001). The set of five games played comprised four games of Bug-Smasher, in two pairs as in the experiment described in Sect. 5, and a physical activity control game. As in the previous experiment, two game variants of different levels of entertainment features (challenge and curiosity) were played in both orders, giving four Bug-Smasher variant games plus the control game.

Bug-Smasher was used for two reasons. First, the new control game is derived (following Malone’s methodology) by eliminating putative entertaining features from Bug-Smasher which makes the latter an appropriate comparison for the new control. Second, while an analogous control for the Space Invaders game could have been devised, making a between-game comparison possible in addition to a within-game comparison, available experimental time limited the complexity of the protocol that could be completed.

The control game played is the “Stomping game” mentioned above. For this game children were asked to stomp on a different 1 of 4 constantly lighted tiles of different color each time they heard a sound coming from the platform, rather than to run around a closed square space. The four tiles are placed at the corners of a 3×3 square in

the center of the 6×6 platform; two tiles equals the average distance between bugs appearing in the Bug-Smasher game. The sound representing the frequency of child–game interaction was emitted at a rate equal to the average of the bug appearance rates of the two different levels of challenge used in the Bug-Smasher game.

The devised control game is designed on the basis of Malone’s studies where features of the game are subtracted and the effect of such changes on children’s entertainment is investigated (Malone 1981). The “Stomping” control game lacks an essential element for successful game design, which is the provision of an apparent goal for the user (Malone 1981). Moreover, the curiosity entertainment feature is minimal since there is complete information of the game’s future states and game interaction is absent given that the playground does not react to the child’s actions. That is, bugs are not smashed (turn red and disappear) when pressed. Given these changes, the proposed control game should not be entertaining for children; and children’s self-reports confirm that in the majority of cases this is so.

All details regarding the protocol of the experiment follow the principles of the previous experiment described in Sect. 5. The only difference lies in that, unlike the previous experiment, children are asked for their preference between the control Stomping game and the variants of Bug-Smasher played and they were given additional instructions for it. To minimize order effects of the control game, the Stomping game is placed either first in the sequence of five games played, after the first pair of Bug-Smasher games played (third in the sequence), or after both pairs of Bug-Smasher games played (fifth in the sequence) with equal probability. Having five tasks (games) in total for each child leaves four fun comparisons (expressed preferences) for each child to report: 24 “fun” comparisons between the Bug-Smasher game variants and the Stomping game and 38 “fun” comparisons between different variants of the Bug-Smasher game.

In the “fun” comparison between the Bug-Smasher game and the Stomping game 22 out of 24 preferences expressed by the children preferred the Bug-Smasher game—confirming the assumption that the Stomping game is a comparatively non-entertaining physical game. Some indicative graphs from children’s real-time HR recordings during control experiment 2 are presented in Fig. 5. The clear observation derived from Fig. 5 is that children appear to be engaged more in the preferred (reported as more entertaining) game than either the Stomping game or the non-preferred (reported as less entertaining) game. This is reflected in their higher $E\{h\}$ values, among other things. These observed effects are valid for the majority of children that participated in this experiment. Please note that $E\{h\}$ values generated from the preferred game are greater than the respective $E\{h\}$ values generated from the Stomping and the non-preferred game in 44 out of 62 instances in total.

9.1.1 Statistical analysis

The HR signal features presented in Sect. 6.1 are used and the statistical analysis described in Sect. 6.2 is followed for data set 2. The only features that demonstrate highly significant effects on reported preferences of entertainment are the $E\{h\}$ ($c(\vec{z}) = 0.392$, p -value = **0.003**) and $\max\{h\}(c(\vec{z})) = 0.357$, p -value = **0.007**). The obtained effect of $E\{h\}$ values are consistent with the ones derived from data set 1

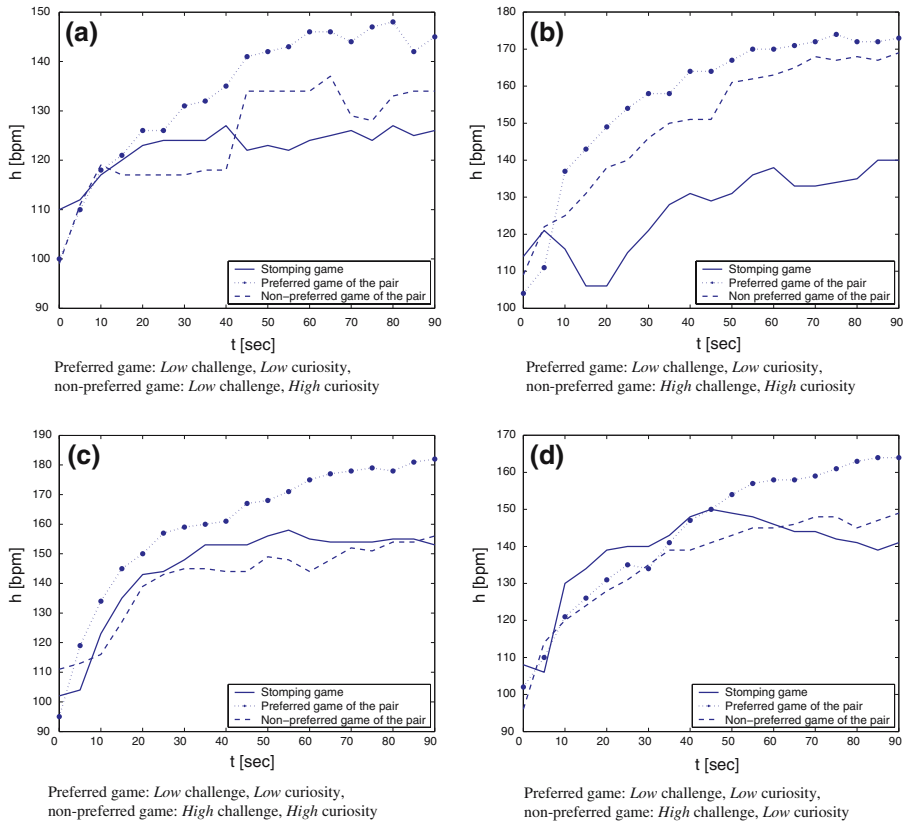


Fig. 5 Comparative HR signal graphs of four children: Stomping, preferred and non-preferred Bug-Smasher games. All four children preferred the game (preferred or not) to the stomping game. In subfigures (a), (b) and (d) even the non-preferred game is preferred to the stomping game. In subfigure (c) there is no information for the child's preference between the game reported as non-preferred and the stomping game since the first is played two games after the second

verifying the rank-correlation relationship of average HR with entertainment preferences of the Bug-Smasher game.

9.1.2 Order effects

As in the first experiment, presented in Sect. 5, we allow each child one minute to rest in between two games in order to avoid any order effect of playing on the HR signal. Given the initial HR of data set 2, the majority (51 out of 62) of the percentage differences between the initial heart rates in the two games played (d_I) lie within $\pm 10\%$. There are seven cases where $d_I < -10\%$ and 4 where $d_I > 10\%$. A t -test for means of paired samples demonstrates no significant difference in the initial HR between the two games of the pair ($t = 0.6896$, p -value = 0.4926).

To check whether the order of playing the games of experiment 2 affects the children's expressed entertainment preferences, we used the test statistic z presented in

Sect. 5.1. The obtained z value is -0.1666 with corresponding p -value is 0.2025 , demonstrating that the order of play does not significantly affect children's reports.

9.2 Feature selection

Training attempts presented in Sect. 8.2 included HR signal data obtained from the exercise (running) task since the latter was used there as a control experiment for physical activity. The purpose of the second experiment is to determine whether game-like physical activity can be discriminated effectively by an ANN user model constructed using the obtained HR signals. Thus, this time, HR signal data from experiment 1 (data set 1) are used for training, but HR data obtained from the running task is excluded. Feature selection is again used as in Sect. 8.2 to find a new minimal feature set that yields the highest measured consistency between the ANN output and the children's reported entertainment preferences on the validation data. Then, data set 2 is used unseen for evaluating the highest performing ANNs evolved on data set 1 (see Sect. 9.4).

As in Sect. 8.2, nBest and SFS described in Sect. 7.1 are applied and compared. The only difference between the GA described in Sect. 7.2 and the one used here is the omission of the $g(d_j^{AE}, \epsilon)$ component in Eq. 5 since there is no need for comparison between exercise E and game A physical activities. Given the 75 pairs of preferred/non-preferred game comparisons, ANNs are evolved by following the approach presented in Sect. 7.2. As in the previous experiment, the data is partitioned into 2/3 training and 1/3 validation data subsets and the leave-one-out cross-validation technique is used to obtain the average classification performance of the ANNs.

The initial subset (ANN input) for both methods includes the feature that performs best in the single feature experiment. This feature is the correlation coefficient R between HR recordings and the time t in which data were recorded with a cross-validation performance of 72.00%. The difference in the best feature obtained ($ApEn$ was the best feature in the previous experiment) is evidently due to the different nature of the running HR signal data included in the data set from that experiment. The success of $ApEn$ in that experiment appears to be due to its usefulness for discriminating between game and non-game activities rather than preferred (entertaining) and non-preferred (less-entertaining) games.

Applying feature selection for more than a single feature, the SFS method generates feature subsets that yield higher validation performance than feature subsets generated by nBest, as presented in Table 2. The best cross-validation performance (76.00%; average of 80.00%, 76.00% and 72.00%) is achieved with the feature subset $\{R, E\{h\}\}$ while adding more HR signal features to the subset do not yield significantly higher performance (see bottom row of Table 2).

Results show the difficulty in distinguishing HR signals between games in terms of the reported preferences of entertainment. This is demonstrated through the lower performance obtained without HR signals from the running task (76%) compared to the respective performance on training including those HR signals (80.66%). Since HR signals obtained from the running task apparently have quite different dynamics from game activity HR signals, the overall classification task appears to be easier when running HR signal data are included in the data set. The binomial-distributed probability

Table 2 Classification accuracy (%) of random network, nBest and SFS feature selection methods

Random network P	nBest		SFS	
	Feature subset	P	Feature subset	P
45.86	$\{R\}$	72.00	$\{R\}$	72.00
	$\{R, s\}$	70.66	$\{R, E\{h\}\}$	76.00
	$\{R, s, E\{h\}\}$	72.00	$\{R, E\{h\}, D\}$	74.66

Random network performance is the average performance of ten random weight value initializations of the network. The random network's input vector consists of the best feature subset that generates the highest cross-validation performance (i.e. $\{R, E\{h\}\}$)

of the 76.00% cross-validation performance occurring at random is **0.0014**, demonstrating statistical significance and providing evidence for this solution's robustness.

9.3 Evolved ANN: $\{R, E\{h\}\}$ feature subset

A more detailed analysis of the best classifier obtained from experiment 2 is presented here. Given the $\{R, E\{h\}\}$ feature subset as inputs, the evolved ANNs correctly match 70.66 % (average of the three training trials; $\sigma = 1.15\%$) of children's answers on entertainment while achieving a classification performance of 76.00% ($\sigma = 6.42\%$) on unknown validation data. Low training performances are due to the early stopping mechanism applied as described in section 7.2. The function between $R, E\{h\}$ and the game's predicted entertainment value (y) given by the highest performing ANN found is illustrated in Fig. 6. As in results presented in Yannakakis and Hallam (2007b), all three fittest ANNs generated, each trained on different sets comprising 2/3 of total data, exhibit similar qualitative features of the surface illustrated in Fig. 6.

The general trend appearing in Fig. 6 is that higher $E\{h\}$ values, independently of the HR signal's linearity (R), tend to correspond to games of higher entertainment, which is the observation made early in the statistical analysis of the data (see Sect. 6.2). Another clear observation is that the combination of low R and $E\{h\}$ values ($R < 0.3, E\{h\} < 0.3$) corresponds to games of low entertainment values. Furthermore, high degree of linearity in the HR signal ($R > 0.8$) combined with low average HR ($E\{h\} < 0.3$) results in high entertainment preference ($y > 0.9$). This suggests that even though $E\{h\}$ is kept low, the game is still entertaining if HR increases linearly with respect to time. In this case, the child is highly engaged in the game throughout the 90-s time window and reports the game as entertaining. Non-preferred games appear to generate a stable (or even decreased) HR in particular at the end of the game which reflects low R values. This corresponds to a rather static playing behavior demonstrating the low level of the child's engagement. Thus, in a sense, conclusions derived from Fig. 6 are consistent with previous studies (Yannakakis et al. 2006a) on the correlates between child-game interaction features and entertainment preferences in the Bug-Smasher game where a significant correlation between the number of interactions with the game platform and reported entertainment is demonstrated.

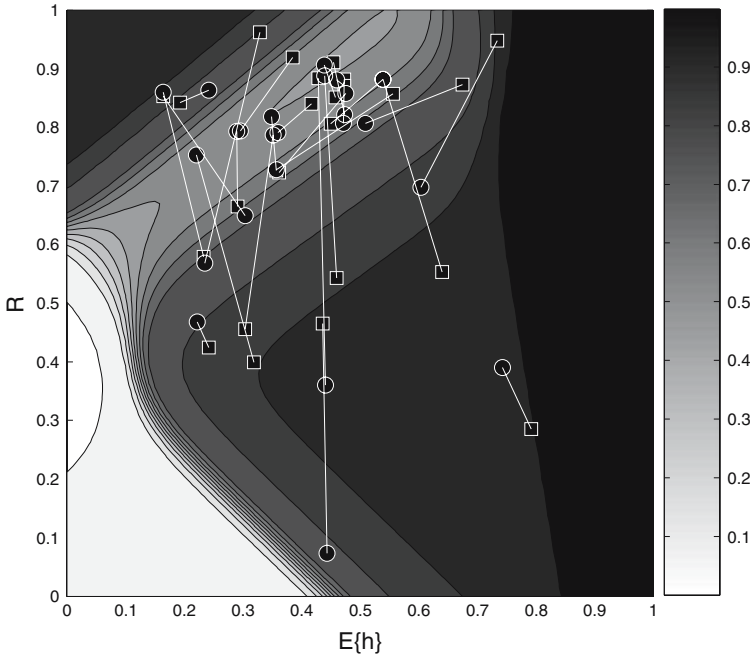


Fig. 6 Evolved ANN that yields the best classification accuracy on unknown data (80.00%): ANN output y (entertainment; the darker the higher) with regards to R and $E\{h\}$. Points plotted correspond to the 50 data of the validation set including 25 preferred games (*squares*) and 25 non-preferred games (*circles*). Game pairs are connected with lines

A t -test for means of paired samples of the $E\{h\}$ values between the games chosen by the subjects as more entertaining and the games chosen as less entertaining shows a significant difference ($t = 2.71$, p -value = **0.0041**) while the respective t -test for the R values shows that the examined HR signal feature values are not significantly different ($t = 0.25$, p -value = 0.7962). The non-correlation with R values combined with the statistically significant binomial-distributed probability (**0.0014**) indicate that the non-linear combination of R and $E\{h\}$ is necessary for a successful predictor of reported entertainment for these Playware games.

9.4 Validation on data set 2

To investigate the extent to which the predictive model of entertainment preference computed using the data from set 1 (excluding the running task) generalizes to new experimental data, the best performing evolved ANNs presented in Sect. 9.3 are presented with and evaluated on the unseen game and control task data of experiment 2.

Table 3 shows the average total classification accuracy (fourth column) and the sub-classification performance for the comparisons between the Bug-Smasher game played and the Stomping game (second column) as well between the Bug-Smasher game chosen as more entertaining and the Bug-Smasher game chosen as less entertaining (third column) of all three evolved ANN.

Table 3 Evolved ANNs (feature subset: $\{R, E\{h\}\}$) trained on data set 1: classification accuracy (%) on unseen data set 2

	Bug-Smasher game vs. Stomping game	Preferred game vs. non-preferred game	Total
Random network	39.13	33.33	35.59
$\{r, E\{h\}\}$	69.56	61.11	64.40

The average classification performance of 10 networks with random weights is given for comparison

The performance obtained equals 64.40%, which appears rather low compared to 76% of correct matching on the validation data of set 1. However, the reported complexity of classifying emotions through physiological state (Picard et al. 2001) and the binomial-distributed probability of this performance to occur at random ($6.03 \cdot 10^{-5}$) suggest that the evolved ANNs are successful predictors of children's reported entertainment preferences based on their physiological state. The average performance of 10 ANNs identical in structure to the evolved ones, but with random weights, is given for comparison.

10 Discussion and conclusions

This paper explored the interplay between physiological signals and children's entertainment preferences in physical playgrounds. More specifically, the quantitative impact of HR signal statistics and regression model parameters on children's reported entertainment was investigated through two dissimilar action games (Bug-Smasher and Space-Invaders) played on the Playware playground. The principal focus of this paper was on the construction of a user model of a child playing a Playware game that can predict the child's answers to which variants of the game are more or less "fun." The model is constructed using neuro-evolution applied to statistical HR signal (measured during play) features. The output of the constructed model is a real number in the range $[0, 1]$ such that more enjoyable games receive higher numerical output. This work is novel in that it examines the physiological state (HR signal) correlates of reported "fun" in physical activity games, isolates HR signal features attributed to entertainment preferences in such physically demanding games and provides a subjective model (predictor) of reported "fun" grounded in HR signal features obtained through the comparative "fun" experiments proposed.

Among the HR signal features examined, the average HR ($E\{h\}$), the maximum HR, the difference between maximum and minimum HR and the approximate entropy ($ApEn$) of the HR signal of the child were found to be the statistics that correlate significantly with the child's expressed entertainment preference. However, only the average HR was correlated significantly with children's entertainment preferences in both Bug-Smasher and Space-Invaders games played. While it is perhaps technically inappropriate to combine data from both Playware games for statistical analysis we, nevertheless, argue that it is appropriate to combine these data for training user models because we explicitly want to examine a model that performs well on both games. Building a separate model for each game would result in two easier machine

learning problems. Thus, the classification performances presented in this paper are underestimates of what might be possible with separate training.

HR signal features provide a means for distinguishing between entertaining games and non-entertaining games as well as between gaming activities and both game-like and non-game-like non-entertaining physical activities (stomping and running). This paper examined the hypothesis that physical activity through games reported as entertaining has a quantitatively dissimilar impact on HR dynamics than a non-entertaining form of physical activity. The first experiment designed to control physical activity was based on children's HR signals during a running task. HR signal feature selection on these HR signal data extracted the feature subset $\{E\{h\}, ApEn\}$. These inputs feed ANNs which correctly predict the reported entertainment preferences of children with a cross-validation accuracy of 80.66% on unseen data. Even though the ANN model is an effective discriminator of physical game activity and running, its good performance is partly due to running's being a qualitatively different type of physical activity.

A more suitable experiment for controlling and isolating the elements of physical activity from an HR signal was therefore designed. The hypothesis under investigation here is whether there is some kind of physical activity that an entertaining game elicits and a non-entertaining game does not. The second control experiment designed is based on a comparative "fun" analysis between variants of the Bug-Smasher game and the Stomping game which is objectively reported by children as less entertaining than the Bug-Smasher game. A new ANN model trained on data obtained from the first experiment—excluding HR signal data obtained from the running task—and validated on unknown data of the same set yield a cross-validation performance of 76%. Feature selection is repeated generating a new feature subset including the average HR ($E\{h\}$) and the correlation coefficient R between HR samples and the time in which samples were recorded. A cross-validation performance of 64.40% is obtained when evaluating the ANN trained on data from the first experiment using unseen data from the second experiment. Even though the obtained performance appears low one has to consider the difficulty of classifying accurately emotions through physiological state (Picard et al. 2001) and the binomial-distributed probability of this performance to occur at random ($6.03 \cdot 10^{-5}$). These provide evidence that the evolved ANNs are successful predictors of children's reported entertainment grounded on their physiological state and validate the hypothesis that there are HR signal features ($E\{h\}$ and R) corresponding to physical activity that can capture entertainment in physical games.

The statistics and regression models (linear, quadratic, exponential) used cover a large portion of the HR signal's qualitative features. However, more HR signal statistics could have been used (e.g. HRV, HR gradient statistics proposed by Picard et al. 2001) if the equipment used stored RR intervals as well as the HR signals obtained. Specifically, the use of HRV measures would allow observing the frequency bands attributed to physical activity separately from those attributed to entertainment. Even though the approximate entropy of HR signals is somewhat correlated with HRV, the latter is of great importance for this work for the above-mentioned reasons. Note also that one cannot control completely for the features of physical activity because the phenomenon measured is physical activity. The Stomping game used is a game-like control task which is objectively less entertaining than the Bug-Smasher game; thus, being a suitable control of physical activity. The use of HRV measures combined with

suitable control tasks like the Stomping game could lead to a more reliable control of the relationship between entertainment, arousal and physical activity. Thus, future experiment plans include the use of more advanced ECG apparatus for HRV calculation as well measurements of other biosignals (e.g. GSR). Alternatively, the use of accelerometers, or the computation of game action measures (e.g. number of interactions/stomps on the playground) and their embedding in the model would further enhance the control of physical activity.

However, the enhanced model is still unlikely to recognize positive or negative valence of the affect—e.g. pleasurable excitement from anger. Children irritated by playing Playware games in general may show signs of arousal without increased physical activity, while children motivated to play harder by anger with competitors will exhibit arousal and physical activity but may not be having “fun.” To the extent that their physical activity is typical of the Playware game, the ANN model presented here will nevertheless assert that they are enjoying themselves.

The proposed entertainment model, in its current state, demonstrates strong evidence that when children are having “fun” during physical play they are engaged more which is reflected through increased physical activity. On the other hand, there is little evidence about the inverse relation between affect and physical activity in this study due to the aforementioned limitations. Measures obtained through future experiments (e.g. using HRV) and additional valence control experiments will throw more light on the relation in physical games between arousal, physiology, physical activity and excitement; whether that is “fun,” anger or other.

As far as the experiment protocol is concerned, a 4-alternative forced choice (4-AFC) approach will be adopted for future 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 provides similar information for the machine learning process while eliminating the noise generated by 2-AFC. The idea of implementing the 4-AFC approach was suggested by an anonymous reviewer and was considered during the design of the second reported experiment; however, the 2-AFC approach was kept for consistency with data obtained from the initial exploratory experiment.

We believe that the entertainment modeling approach through HR dynamics presented here is general over the majority of action games that could be created with Playware since Bug-Smasher and Space-Invaders between them represent a large proportion of the features met in Playware action games. Moreover, it is our belief that the entertainment models proposed here may very well be applied to other interactive entertainment systems that include physical activity. However, each game demonstrates individual entertainment features which may have an impact on the child’s HR signal and therefore, new games need to be tested to confirm the generality of the approach.

Individual differences in children’s physiology, preferences and playing behavior generate difficulties in generalization over subjects. This is a fundamental limitation of attempting to construct a model based on combined data from multiple subjects: a game that to one child is exciting and fun may to another be too fast, or too slow. Nevertheless the results presented show that some generalization across individuals is possible, in that the evolved ANNs do predict children’s preferences with

reasonable performance. With further work it may be possible to improve performance, for example by clustering individual players into classes depending on observed playing style. Each class could then have its own model, simplifying the machine learning problem.

The proposed approach can be used for adapting the game's entertainment features according to the player's individual HR signal dynamics in "real-time" (i.e. at regular intervals or using a sliding window). The key to this is the observation that the models (ANNs) relate HR signal features to an estimated value (y) of entertainment preference which is a continuous differentiable function of the feature signals. It is therefore possible in principle to infer what changes to game features and furthermore to HR signal features will cause an increase in the interestingness of the game, and to adjust game parameters to make those changes. (This may seem surprising since the model inputs consist of rank-preferences at particular points; however, the machine learning process *interpolates* a consistent function between these data.)

Thus, the partial derivatives of $\partial y / \partial E\{h\}$ and $\partial y / \partial R$ indicate the change in entertainment preference for a small change in an individual HR signal feature. One of the advantages of the feature subset $\{E\{h\}, R\}$ found is that both its feature values can be easily computed from windowed data. Given knowledge of the relationship between game features (i.e. challenge, curiosity and fantasy) and HR signals, one might then determine which game features need to be adjusted and to which levels in order for the HR signals to change in the directions required by the aforementioned partial derivatives.

Physiological signals are of great importance for the investigation of the interplay between physiology and reported entertainment and could be combined with player-game interaction data for the effective construction of user models. However, the long-term goal of this study is not on the use of physiology, but rather on the use of data obtained from the player-game interaction for the construction of entertainment models within the Playware playground. Given that the current state of commercial bio-sensing technology is too intrusive, in the real-world scenario, any Playware physical game could not easily incorporate such biosignal equipment—though of course future and emerging technologies in this area may make access to biosignals natural in future versions of the system.

Acknowledgements The authors thank Henrik Jørgensen and the children of Henriette Hørlücks School, Odense, Denmark, who participated in the experiments. The authors also thank the anonymous reviewers for their constructive and thought-provoking criticism of earlier drafts of the paper. The tiles were designed by C. Isaksen from Isaksen Design and parts of their hardware and software implementation were collectively done by A. Derakhshan, F. Hammer, T. Klitbo and J. Nielsen. KOMPAN, Mads Clausen Institute, and Danfoss Universe also participated in the development of the tiles. This work was supported in part by the Danish Research Agency, Ministry of Science, Technology and Innovation (project no: 274-05-0511).

References

- Andrade, G., Ramalho, G., Santana, H., Corruble, V.: 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), pp. 7–12 (2005)
- Aristotle: *On the Soul*, Book III. Kessinger Publishing (2004)

- Bobick, A., Intille, S., Davis, J., Baird, F., Pinhanez, C., Campbell, L., Ivanov, Y., Schutte, A., Wilson, A.: The kidsroom: A perceptually-based interactive and immersive story environment. Technical Report 398, MIT Media Laboratory (1996)
- Champanand, A.J.: *AI Game Development*. New Riders Publishing (2004)
- Conati, C.: Probabilistic assessment of user's emotions in educational games. *J. Appl. Artif. Intell.* special issue on "Merging Cognition and Affect in HCI" **16**, 555–575 (2002)
- Conati, C., Chabbal, R., Maclaren, H.: A study on using biometric sensors for detecting user emotions in educational games. In: Proceedings of the Workshop "Assessing and Adapting to User Attitude and Affects: Why, When and How?". In conjunction with UM'03, 9th International Conference on User Modeling, Pittsburgh, PA, U.S.A., pp. 60–65 (2003)
- Critchley, H.D., Rotshtein, P., Nagal, Y., O'Doherty, J., Mathias, C.J., Dolan, R.J.: Activity in the human brain predicting differential heart rate responses to emotional facial expressions. *Neuroimage* **24**, 751–762 (2005)
- Csikszentmihalyi, M.: *Flow: The Psychology of Optimal Experience*. Harper & Row, New York (1990)
- Devijver, P., Kittler, J.: *Pattern Recognition—A Statistical Approach*. Prentice-Hall, Engelwood Cliffs, NJ (1982)
- D'Mello, S., Craig, S.D., Whitherspoon, A., McDaniel, B., Graesser, A.: Automatic detection of learner's affect from conversational cues. *User Model. User-Adap. Inter., Special Issue: User Model. Affect. Comput.* (2008) (to appear)
- Funge, J.D.: *Artificial Intelligence for Computer Games*. A. K. Peters Ltd, Wellesley, Massachusetts, USA
- Haapalainen, E., Laurinen, P., Junno, P., Tuovinen, H., Roening, J.: Methods for classifying spot welding process: A comparative study of performance. In: IEA/AIE, pp. 412–421 (2005)
- Hazlett, R.L.: Measuring emotional valence during interactive experiences: boys at video game play. In: CHI '06: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1023–1026. ACM Press, New York, NY, USA (2006)
- Holland, J.H.: *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor, MI (1975)
- Hughes, M., Hutt, C.: Heart-rate correlates of childhood activities: play, exploration, problem-solving and day-dreaming. *Am. Psychol.* **8**, 253–263 (1979)
- Hunicke, R., Chapman, V.: 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), AAAI Press (2004)
- Hutt, C. In: *Exploration and Play, in Play and Learning*, pp. 175–194. Gardner Press, New York (1979)
- Iida, H., Takeshita, N., Yoshimura, J.: A metric for entertainment of boardgames: its implication for evolution of chess variants. In: Nakatsu R., Hoshino J. (eds.) IWEC2002 Proceedings, pp. 65–72. Kluwer (2003)
- Ishii, H., Wisneski, C., Orbanes, J., Chun, B., Paradiso, J.: PingPongPlus: design of an athletic-tangible interface for computer-supported cooperative play. In: CHI '99: Proceedings of the SIGCHI Conference on Human factors in Computing Systems, pp. 394–401. ACM Press, New York, NY, USA (1999)
- James, W. In: *William James: Writings 1878–1899*. The Library of America, pp. 350–365 (1992). Originally published in 1890
- Kapoor, A., Mota, S., Picard, R.: Towards a learning companion that recognizes affect. In: Proceedings of Emotional and Intelligent II: The Tangled Knot of Social Cognition, AAAI Fall Symposium, AAAI Press (2001)
- Kline, S.: *Out of the Garden: Toys and Children's Culture in the Age of Marketing*. Verso (1993)
- Konami Digital Entertainment B.V.: (DDR and Kick & Kick) <http://www.konami.com>
- Koster, R.: *A Theory of Fun for Game Design*. Paraglyph Press (2005)
- Laird, J.E., van Lent, M.: Human-level AI's killer application: Interactive computer games. In: Proceedings of the Seventh National Conference on Artificial Intelligence (AAAI), pp. 1171–1178, AAAI Press (2000)
- Lazzaro, N.: Why we play games: Four keys to more emotion without story. Technical Report, XEO Design Inc. (2004)
- Liljedahl, M., Lindberg, S.: DigiWall—an audio mostly game. In: International Conference on Auditory Display, London (2006)
- Lindstrom, M., Seybold, P.: BRANDchild: Insights into the Minds of Today's Global Kids: Understanding Their Relationship with Brands. Kogan Page (1994)

- Lund, H.H., Jessen, C.: Playware—intelligent technology for children's play. Technical Report TR-2005-1, Maersk Institute For Production Technology, University of Southern Denmark (2005)
- Lund, H.H., Klitbo, T., Jessen, C.: Playware technology for physically activating play. *Artif. Life Robot. J.* **9**, 165–174 (2005)
- Magerkurth, C., Stenzel, R., Prante, T.: STARS – a ubiquitous computing platform for computer augmented tabletop games. In: *Extended Abstracts of UbiComp '03*, pp. 267–268. Springer (2003)
- Malone, T.W.: What makes computer games fun? *Byte* **6**, 258–277 (1981)
- Mandryk, R.L., Atkins, M.S., Inkpen, K.M.: A continuous and objective evaluation of emotional experience with interactive play environments. In: *Proceedings of the Conference on Human Factors in Computing Systems (CHI 2006)*, pp. 1027–1036. ACM Press (2006a)
- Mandryk, R.L., Inkpen, K.M., Calvert, T.W.: Using psychophysiological techniques to measure user experience with entertainment technologies. *Behav. Inform. Technol.* (Special Issue on User Experience) **25**, 141–158 (2006b)
- McQuiggan, S., Lee, S., Lester, J.: Predicting user physiological response for interactive environments: an inductive approach. In: *Proceedings of the 2nd Artificial Intelligence for Interactive Digital Entertainment Conference*, pp. 60–65. AAAI Press (2006)
- Metaxas, G., Metin, B., Schneider, J., Shapiro, G., Zhou, W., Markopoulos, P.: SCORPIODROME: An Exploration in Mixed Reality Social Gaming for Children. In: *Proceedings of ACM SIGCHI International Conference on Advances in Computer Entertainment Technology*, ACM Press (2005)
- Mixed Reality Lab: (Singapore) <http://www.mixedreality.nus.edu.sg/>
- Montana, D.J., Davis, L.D.: Training feedforward neural networks using genetic algorithms. In: *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence (IJCAI-89)*, pp. 762–767, Morgan Kaufman, San Mateo, CA (1989)
- Mueller, F., Agamanolis, S., Picard, R.: Exertion interfaces: sports over a distance for social bonding and fun. In: *CHI '03: Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pp. 561–568. ACM Press, NY, USA (2003)
- Nareyek, A.: Intelligent agents for computer games. In: Marsland, T., Frank, I. (eds.) *Computers and Games, Second International Conference, CG 2002*, pp. 414–422, LNCS, Springer (2002)
- Orr, R.J., Abowd, G.D.: The smart floor: a mechanism for natural user identification and tracking. In: *CHI '00: CHI '00 Extended Abstracts on Human Factors in Computing Systems*, pp. 275–276. ACM Press, NY, USA (2000)
- Picard, R.W., Vyzas, E., Healey, J.: *IEEE Trans. Pattern Anal. Mach. Intell.* **23**, 1175–1191 (2001)
- Pincus, S.M., Goldberger, A.L.: Physiological time-series analysis: what does regularity quantify? *Am. J. Physiol.* **266**, H1643–56 (1994)
- Pincus, S.M.: Approximate entropy as a measure of system complexity. *Proc. Natl. Acad. Sci.* **88**, 2297–2301 (1991)
- Porayska-Pomsta, K., Mavrikis, M., Pain, H.: Towards predictive modelling of tutorial interactions: student affect from tutor perspective. *User Model. User-Adap. Inter., Special Issue: User Model. Affect. Comput.* (2008) (to appear)
- Postman, N.: *The Disappearance of Childhood*. Allen, London (1983)
- Rani, P., Sarkar, N., Liu, C.: Maintaining optimal challenge in computer games through real-time physiological feedback. In: *Proceedings of the 11th International Conference on Human Computer Interaction*, Lawrence Erlbaum (2005)
- Ravaja, N., Saari, T., Turpeinen, M., Laarni, J., Salminen, M., Kivikangas, M.: Spatial presence and emotions during video game playing: does it matter with whom you play? *Presence Teleoperators & Virtual Environments* **15**, 381–392 (2006)
- Read, J., MacFarlane, S., Cassey, C.: Endurability, engagement and expectations. In: *Proceedings of International Conference for Interaction Design and Children* (2002)
- Richardson, B., Leydon, K., Fernström, M., Paradiso, J.A.: Z-Tiles: Building Blocks for Modular, Pressure-Sensing Floorspaces. In: *Proceedings of CHI 2004*, pp. 1529–1532. ACM Press, NY, USA (2004)
- Rosenthal, R.: Covert communication in laboratories, classrooms, and the truly real world. *Curr. Direct. Psychol. Sci.* **12**, 151–154 (2003)
- Spronck, P., Sprinkhuizen-Kuyper, I., Postma, E.: Difficulty Scaling of Game AI. In: *Proceedings of the 5th International Conference on Intelligent Games and Simulation (GAME-ON 2004)*, pp. 33–37 (2004)
- Sweetser P., Wyeth P.: GameFlow: a model for evaluating player enjoyment in games. *ACM Comput. Entertain.* **3**, (2005)

- Taatgen, N.A., van Oploo, M., Braaksmā, J., Niemantsverdriet, J.: How to construct a believable opponent using cognitive modeling in the game of set. In: Proceedings of the Fifth International Conference on Cognitive Modeling, pp. 201–206 (2003)
- Turkle, S.: *The Second Self: Computers and the Human Spirit*. Simon and Schuster, New York (1984)
- Verma, M.A., McOwan, P.W.: An adaptive methodology for synthesising Mobile Phone Games using Genetic Algorithms. In: Congress on Evolutionary Computation (CEC-05), Edinburgh, UK, pp. 528–535, IEEE (2005)
- Wiberg, C.: *A Measure of Fun. Extending the Scope of Web Usability*. PhD Thesis, Department of Informatics. Umea University (2003)
- Wiberg, C., Jegers, K.: Satisfaction and learnability in edutainment: a usability study of the knowledge game ‘Laser Challenge’ at the Nobel e-museum. In: Proceedings of HCI International—10th International Conference on Human Computer Interaction, Crete, Greece (2003)
- Yannakakis, G.N.: *AI in Computer Games: Generating Interesting Interactive Opponents by the use of Evolutionary Computation*. Ph.D. Thesis, University of Edinburgh (2005)
- Yannakakis, G.N., Hallam, J.: Evolving opponents for interesting interactive computer games. In: Schaal, S., Ijspeert, A., Billard, A., Vijayakumar, S., Hallam, J., Meyer, J.A. (eds.) *From Animals to Animals 8: Proceedings of the 8th International Conference on Simulation of Adaptive Behavior (SAB-04)*, pp. 499–508. The MIT Press, Santa Monica, LA, CA (2004)
- Yannakakis, G.N., Hallam, J.: 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), pp. 119–124 (2005)
- Yannakakis, G.N., Hallam, J.: 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, pp. 432–442. Springer-Verlag, Heraklion, Greece (2006)
- Yannakakis, G.N., Hallam, J.: Towards optimizing entertainment in computer games. *Appl. Artif. Intell.* to appear (2007a)
- Yannakakis, G.N., Hallam, J.: Preliminary studies for capturing entertainment through physiology in physical play. Technical Report TR-2007-5, Maersk Institute, University of Southern Denmark (2007b)
- Yannakakis, G.N., Lund, H.H., Hallam, J.: Modeling children’s entertainment in the playware playground. In: Proceedings of the IEEE Symposium on Computational Intelligence and Games, Reno, USA, pp. 134–141. IEEE (2006a)
- Yannakakis, G.N., Hallam, J., Lund, H.H.: 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, pp. 314–317. Springer-Verlag, Cambridge, UK (2006b)
- Yannakakis, G.N., Hallam, J., Lund, H.H.: Comparative Fun Analysis in the Innovative Playware Game Platform. In: Proceedings of the 1st World Conference for Fun’n Games, pp. 64–70 (2006c)
- Yao, X.: Evolving artificial neural networks. In: Proceedings of the IEEE, vol. 87, pp. 1423–1447, IEEE (1999)
- Zuckerman M.: Sensation seeking in entertainment. In: *Psychology of Entertainment*, pp. 367–387. Lawrence Erlbaum Associates Publishers (2006)

Authors’ vitae

Dr. Georgios N. Yannakakis received a 5-year Diploma (1999) in Production Engineering and Management and a M.Sc. (2001) degree in Financial Engineering from the Technical University of Crete, and a Ph.D. degree (2005) in Informatics from the University of Edinburgh. Since 2006 he has been a postdoctoral fellow and principal investigator of the CEPEG project, funded by the Danish Research Agency, at the Maersk Mc-Kinney Moller Institute, University of Southern Denmark, where he works in the fields of affective computing, entertainment modeling, game AI and neuro-evolution. His contribution is based on experiences gained both from his Ph.D. work as well as his current research.

Dr. John Hallam is Professor at the Maersk Mc-Kinney Moller Institute (MMMI), University of Southern Denmark. He graduated with First Class Honours in Mathematics from the University of Oxford in 1979, completed a Ph.D. in the Department of Artificial Intelligence at the University of Edinburgh in 1984 and joined the teaching Faculty in that Department in 1985. He established the Edinburgh Mobile Robotics Research Group, having been active in mobile robotics research for almost 25 years, and moved to MMMI in 2003. The current foci of his catholic research interests in artificial intelligence are biological modeling

using robotic techniques, evolutionary robotics, and entertainment modeling. He has published around 100 journal and international conference papers on various robotic, non-symbolic computing and entertainment modeling topics, and has designed electronic hardware both for research and teaching.

Dr. Henrik Hautop Lund, Professor, is founder and director of the Adaptronics Group working in robotics and modern artificial intelligence. His team developed the ATRON shape-shifting robots, developed playware with the novel intelligent playgrounds, and has won the RoboCup Humanoids Free Style World Championship 2002. Professor Henrik Hautop Lund is engaged also in the development of intelligent artifacts, I-BLOCKS, for the teaching of creativity in Africa, where his team were one of the driving forces behind the first science park in East Africa in Iringa, Tanzania. He is founder and board member of the industrial cluster, RoboCluster, comprising more than 100 industrial interested parties. He has developed RoboMusic together with the remix musician and World Music Award winner Funkstar De Luxe. Henrik Hautop Lund is director of the company Entertainment Robotics. He is member of the Danish national research council, and numerous international committees. He has produced more than 100 scientific publications in robotics and modern artificial intelligence. He recently founded the Center for Playware at the University of Southern Denmark.