Towards Generic Models of Player Experience

Noor Shaker[1] **Mohammad Shaker**[2] **Mohamed Abou-Zleikha**[3]
[1]Center for Computer Games and Interaction Design at the IT University of Copenhagen, Copenhagen, Denmark
[2]Joseph Fourier University, Grenoble, France
[3]Audio Analysis Lab, AD:MT Aalborg University, Aalborg, Denmark
nosh@itu.dk,mohammadShakerGtr@gmail.com, moa@create.aau.dk

Abstract

Context personalisation is a flourishing area of research with many applications. Context personalisation systems usually employ a *user model* to predict the appeal of the context to a particular user given a history of interactions. Most of the models used are context-dependent and their applicability is usually limited to the system and the data used for model construction. Establishing models of user experience that are highly scalable while maintaing the performance constitutes an important research direction. In this paper, we propose generic models of user experience in the computer games domain. We employ two datasets collected from players interactions with two games from different genres where accurate models of players experience were previously built. We take the approach one step further by investigating the modelling mechanism ability to generalise over the two datasets. We further examine whether generic features of player behaviour can be defined and used to boost the modelling performance. The accuracies obtained in both experiments indicate a promise for the proposed approach and suggest that game-independent player experience models can be built.

Introduction

The field of predicting user affect is rich with many interesting studies focusing on estimating user's emotional states while interacting with the system. The methodologies followed usually include collecting informative indicators of user behaviour which are later employed by a machine learning technique to predict the appeal of a specific piece of software to a particular user or group of users. Several behavioural features could be employed such as subjective, objective or predefined selected features gathered from the interaction with the system. The ultimate goal of most of these systems is to construct accurate estimators of User Experience (UE) that could be later used to suggest content modifications or adjustments so that the system becomes aware of the user progression and needs and ultimately able to provide *personalised* content.

Most of the studies reported however focus on analysing and constructing models of UE in one setting. The models are usually built with one specific case in mind and do not usually scale; they can not be directly employed to predict or adapt in other contexts. For instance, the probabilistic model suggested in (Conati 2002; 2011) for affect detection in an educational software, although accurate, can not be directly used to predict affect in another educational system. Even if we are to use the same methodology, we would need to go through the whole process of data collection and model construction. This process is time and effort consuming and as a result, the applicability of the systems built will be limited to a specific case.

Each user has distinctive characteristics that could potentially be captured by a software. By investigating different systems and the way people interact with digital software, one could notice specific patterns and similarities of a certain user behaviour across multiple systems. This suggests the potential of constructing generic estimators of UE that could effectively predict user's affect regardless of the specific system the user is interacting with. One could ultimately envision a user-centric software that integrates different sources of information about the user (this could include personality traits, physiological data, demographics, preferences, desires and behaviour) (Cooper, Reimann, and Cronin 2007) and is capable of accurately infering user's affect and consequently suggest personalised content. The concept of context-independent users' models is not new (Kobsa 2001; Bartle 1996) yet very few examples can be found in the literature on providing a general representation of users (Alrifai et al. 2012; Heckmann et al. 2007) and this field has not been, to the best of our knowledge, researched within the games domain. This might be in part because of the difficulty in defining a shared model of UE that can be used by different systems and because inferring users' state from data collected about the interaction with multiple systems is not an obvious task.

Generic models of UE are particularly interesting in the computer games domain where players engage in a rich environment and where a shared model of user behaviour is relatively easier to construct. Given the playing style in one game, one could infer the player skill level in other games that are not necessarily from the same genre and ultimately be able to use this information to personalise the content (Bakkes, Tan, and Pisan 2012). It would therefore be highly useful for a player model constructed from players' data in one game to effectively predict the appeal of the content of another game to a specific playing style.

This paper presents the first step towards this goal by investigating the feasibility of constructing accurate *generic* estimators of players affect in the computer games domain. In particular, we investigate whether we can effectively capture players affect from generalised in-game behavioural features. We analyse two games from dissimilar genres where accurate game-dependent estimators of player Experience (PE) were previously constructed. We examine the gameplay behaviour in both games and we investigate the important features for predicting affect as selected by the modelling mechanism. We further discuss whether the selected features can be generalised and applied in both games and we consequently present experiments on constructing generic models that effectively capture PE in both games. The results suggest that accurate generic models of PE can indeed be accomplished. To the best of our knowledge this is the first attempt to construct generic models of UE from behavioural features in the games domain.

Related Work

In what follows we review the relevant literature. In particular, we introduce previous work on emotion recognition and on modelling UE with more emphasis on its application within the computer games domain.

Emotion Recognition

In computer science, affective computing is the study and development of methods that give the computers the ability to recognise and induce emotion and enable them to interact with humans in human-like ways (Picard 1995). The last three decades have witnessed increasing interest in automatic human affect analysis. Several authors conducted extensive surveys of work in the machine analysis of affective expressions (Sebe, Cohen, and Huang 2005; Pantic and Rothkrantz 2003; Zeng et al. 2009).

In the games domain, affect induction is an essential part, since most games can be tweaked in order to make the PE more expressive and, thus, produce multimodal data that can be analysed and classified (Scherer 2005). The literature defines three main modalities to capture affect (Yannakakis and Togelius 2011). Objective measures such as data collected from e.g. body movement (Isbister, Schwekendiek, and Frye 2011) and facial features (Hoque, McDuff, and Picard 2012) among many others. Most of these measures may be unsuitable for use in gaming context since they are highly intrusive. Subjective measures on the other hand rely on self-reports as the main indicator of affect. Despite their simplicity in inferring user emotions, these methods have been successfully implemented in a wide range of applications (Scherer 2005) including the work presented in this paper. Other studies relies on context-based features collected from the interaction. These features are usually statistical spatio-temporal features of interaction such as the amount of time or the frequency of doing a certain activity. We use such features in this work as the main method to capture behaviour

Modelling of User affects

Content personalisation is a crucial aspect in many applications where software are becoming more aware of their users and more capable of adapting according to individual needs (Viviani, Bennani, and Egyed-Zsigmond 2010). Rich information about users and their interactions with systems are collected to facilitate understanding users and their preferences for the purpose of providing "better" content. Such systems can be employed in many different applications: serious games, e-commerce, e-learning, entertaining games, etc.

User modelling (Kobsa 2001) plays an important role in this type of environments and forms the basis for context-independent personalisation (Niederée et al. 2004). Several studies in dissimilar domains suggest that providing customised content improves user's experience (Zakharov, Mitrovic, and Johnston 2008; Conati 2002)

There is an abundance of studies presented in the literature on constructing computational models of emotion (Wöllmer et al. 2013; Calvo and D'Mello 2010; Ortony, Clore, and Collins 1990; Conati 2002). There are also game-specific theories about player emotion (Malone 1981; Sweetser and Wyeth 2005; Koster 2004). Estimating affective and cognitive states in conditions of rich human-computer interaction, such as in games, is a field of growing academic and commercial interest. Several studies with varying success can be found on constructing models of PE in different game genres (Yannakakis and Hallam 2009; 2006; Pedersen, Togelius, and Yannakakis 2010).

While the Player Experience Models (PEMs) constructed in some of these studies achieved reasonable rates of accuracy (70-90%), they are still limited to predicting PE in the specific game used to collect the data. There is only one study we are aware of that investigate the construction of generic models of PE from physiological data (Martínez, Garbarino, and Yannakakis 2011). While this study suggests a promise for the generalisation approach, its application is limited to cases where physiological data is available or can be easily acquired. Such data is usually not available in the game domain where gameplay data is the most popular source of information about player behaviour. Moreover, the results obtained suggest that models with better accuracies might be constructed if other modalities are considered.

Different studies on why people play games conducted in an independent cross-genre research revealed a number of key features of an optimal game experience that is generalised over game genres (Martínez, Garbarino, and Yannakakis 2011). These studies support our belief that player behaviour in one game carry information about her general behaviour in games and thereafter information about such behaviour can generalise. If we are to prove this hypothesis, there will be no longer a need to repeat the full process of designing experiments, collecting data, and construing models of PE. The process could be optimised so that previously construct models could be utilised for a new game and information coming from the user can be used to actively update the model so that they become more accurate while the game is being played.

Method

To test whether generic models could be constructed, two dataset from games from dissimilar genres are used.

Super Mario Bros

Super Mario Bros is a very popular 2D platform game published by Nintendo. An open source clone of the game, called *Infinite Mario Bros* (IMB) is used in our study. The clone is modified to permit control over level generation and thereafter provide different variations of content for players to experience and compare. The gameplay in Infinite Mario Bros consists of moving the player-controlled character, Mario, through two-dimensional levels. Mario can walk and run, duck, jump, and shoot fireballs. The main goal of each level is to get to the end of the level. Auxiliary goals include collecting as many coins as possible, and clearing the level as fast as possible. The player is given three lives to complete the game.

First-Person Shooter: Sauerbraten

A second dataset from a First-Person Shooter (FPS) game is used. The game is from a totally different game genre and is built on the game engine *Cube*. The game in our experiments is played in single player mode and the goal is to collect the highest score possible by traversing the arena for two minutes, killing as many of the enemies as possible and avoid being hit for a snapshot for one of the maps used). The player is equipped with a weapon and she can collect other types of weapons and resources. The weapons differ in their accuracy, damage caused and shooting range. Every time the player is killed, she looses one point and she is re-spawned again as long as she still has time left to play.

Study Procedure

Two different datasets were collected containing different number of participants. Game surveys were conducted to collect information about players' interaction with the games and their affective states. The same protocol is followed for data collection in both datasets. According to this protocol, players are presented with a pair of two sessions that differ along one or more aspects of game content. While playing, detailed information about player behaviour and actions were recorded. After playing each pair, players were asked to report their emotional/behavioural states following the four-alternative forced choice protocol that asks the players to express their preference of the three states: *engagement*, *frustration* and *challenge*. Pairwise preferences were adopted where the questionnaires presented are of the form: "Which game was more E?" where E is the state under investigation. The possible answers are: (1) game A [B] was more/less E than game B [A] (2) both equally or (3) neither. Full details about the procedure followed can be found in (Shaker, Yannakakis, and Togelius 2012; Shaker et al. 2013)

In the following sections, we describe the procedure followed to collect the data and the characteristics of each dataset.

Dataset 1: Super Mario Bros An executable version of the software was uploaded online and participants were invited to play the game and answer the questionnaire. The data was collected over a period of six months. Participants' age covers a range between 16 and 64 years (31.5% females). The final dataset consists of a total of 273 unique players who played 780 game pairs (1560 game levels). The data was preprocessed to remove the pairs in which players reported unclear preferences (their answers were neither or both equally), and after this step, the number of pairs remained were 597, 531 and 629 for engagement, frustration and challenge, respectively. Several representative features of player behaviour were extracted and Table 1 presents a subset of these features. The full set contains 30 features which can be found in (Shaker, Yannakakis, and Togelius 2012).

Dataset 2: First-Person Shooter A data collection event was organised and advertised over social networks were people are invited to participate. The event was held at the university over two days and a total number of 62 students participated. Participants' age covers the range between 18 and 24 years (12% females). Each player was asked to play at least one pair (two game sessions, each last for two minutes) and she can play as many pairs as she wants. The final dataset consists of 124 pairs and after removing the pairs with unclear preferences 115, 111 and 112 pairs remained for engagement, frustration and challenge, respectively. Players' behavioural features were extracted as indicators of players' style. Table 2 presents a subset of these features (the total set contains 23 unique features). The context features presented are the ones used to design the variations of the game content presented to the players (full details of the procedure can be found in (Shaker et al. 2013; 2015)).

Player Experience Modelling

In order to construct powerful estimators of PE derived from the in-game interaction we use Neuroevolution Preference Learning (NPL). This approach has been used in the literature for modelling PE from preference data with better accuracies than other approaches such as support victor machine and Bayesian learning (Yannakakis, Maragoudakis, and Hallam 2009). In the following we briefly describe the main steps followed by this approach to build Player Experience Models (PEMs), more details about the procedure can be found in (Shaker, Yannakakis, and Togelius 2012).

Feature Selection

The first step when constructing models of PE is to analyse the input space and select only the features that are relevant for accurate prediction. For this purpose, we use Sequential Forward Selection (SFS). This is achieved by training single-layer perceptrons (SLPs) and simple multi-layer perceptrons (MLPs) models through NPL.

Model Construction

SLPs and small MLPs can efficiently capture simple relationship between the features and reported affects. This re-

Table 1: Features extracted from Super Mario Bros data recorded.

Category	Feature	Description
Time	t_{comp}	Completion time
	t_{play}	Playing duration of last life over total time spent on the level
	t_{jump}	Time spent jumping (%)
	t_{left}	Time spent moving left (%)
	t_{right}	Time spent moving right (%)
	t_{run}	Time spent running (%)
	t_{small}	Time spent in Small Mario mode (%)
	t_{big}	Time spent in Big Mario mode (%)
Interaction with items	n_{coin}	Free coins collected (%)
	$n_{coinBlock}$	Coin blocks pressed or coin rocks destroyed (%)
Interaction with enemies	k_{goomba}	Times the player kills a goomba or a koopa (%)
	k_{stomp}	Opponents died from stomping (%)
Death	d_{total}	Total number of deaths
	d_{cause}	Cause of the last death
Miscellaneous	n_{mode}	Number of times the player shifted the mode (Small, Big, Fire)
	n_{jump}	Number of times the jump button was pressed
Context features	E	Number of enemies
	G	Number of gaps
	G_w	Average width of gaps
	E_p	Placement of enemies

Table 2: Extracted features from players data in the first-person shooter game.

Category	Feature	Description
Time	t_{life}	Duration of play
	t_{weapon}	Time spent using weapons (%)
	t_{shoot}	Time spent shooting (%)
	t_{still}	Time spent not moving (%)
	t_{jump}	Time spent jumping (%)
	t_{exp}	Time spent using explosive weapons(%)
Interaction with items	n_{health}	Health items collected (%)
	n_{armour}	Armours collected (%)
Interaction with enemies	e_{kill}	Number of times the player kills an enemy (%)
	p_{hit}	Number of times the player receives a hit from an enemy (%)
	e_{hit}	Number of times the player hits an enemy (%)
Miscellaneous	n_{death}	Number of times the player died
	s_{acc}	Shooting accuracy
Context features	E	Number of enemies
	E_{skill}	Skill level of enemies
	W_{type}	Type of weapons including explosive and non-explosive weapons
	H	Number of health items
	R	Number of resources such as bullets and armors

lationship however is most likely to be more complex and therefore after selecting the subset of relevant features, the topology of the MLP models is optimised. This process starts with small MLPs and the network topology is gradually complexified up to a certain predefined limit.

Experimental Setup

Parameter tuning tests were conducted to set up the parameters' values for neuroevolutionary. As a result, we use a population of 100 individuals and we run evolution for 20 generations. A probabilistic rank-based selection scheme is used, with higher ranked individuals having higher probability of being chosen as parents. Finally, reproduction is performed via uniform crossover, followed by Gaussian mutation of 1% probability. In all of our experiments, the data was first randomised to minimise the effect of player-specific data on training and testing. The performance is calculated as the average classification accuracy in three independent runs using 3-fold cross validation. Each experiment is repeated five times.

Results and Analysis

The above mentioned procedure was followed to construct models of PE for the the two datasets. In the following, we present the models constructed for each game and we discuss the generalisation procedure followed to obtain generic models.

Game-Specific Models

In previous studies, models of high accuracies were constructed from subsets of selected features in both datasets (Shaker, Yannakakis, and Togelius 2012; Shaker et al. 2015). Table 3 presents the final set of features selected following the PEM procedure discussed and the modelling accuracies obtained. Notice that a baseline model in our case would yield a prediction accuracy of 50%.

Table 3: Features selected from the set of extracted parameters for predicting engagement, frustration and challenge. The table also presents the corresponding average performance ($\bar{P}_{avg}$) and the maximum (P_{max}) values obtained. Context features appear in bold.

	Sauerbraten			Infinite Mario Bros		
	Engagement	Frustration	Challenge	Engagement	Frustration	Challenge
$Selected$ $features$	p_{hit} t_{still} $\mathbf{E_{skill}}$ $\mathbf{E}$ $\mathbf{W_{type}}$ t_{exp} n_{armour}	p_{hit} e_{hit} e_{kill} t_{still}	t_{life} n_{death} $\mathbf{E}$ $\mathbf{E_{skill}}$ $\mathbf{W_{type}}$ t_{weapon}	t_{comp} n_{coin} d_{cause} t_{small} $\mathbf{E}$ t_{jump} $n_{coinBlock}$ t_{big} t_{run} n_{jump}	t_{right} d_{total} d_{cause} k_{goomba} t_{play} $\mathbf{\bar{G}_w}$ $\mathbf{G}$ n_{jump}	t_{play} n_{jump} d_{total} n_{coin} t_{right} $\mathbf{\bar{G}_w}$ $\mathbf{E_p}$ t_{left} k_{stomp}
P_{avg}	71.38%	80.91%	96.25%	67.18%	76.50%	74.03%
P_{max}	77.19%	89.18%	99.09%	73.50%	83.00%	79.10%

Model Generalisation

Our first experiment aimed at investigating whether models built on one dataset can be generalised to predict players' affect on another dataset. To examine this, the accuracies of the PEMs constructed from the IMB dataset are calculated for the FPS data and vice versa. In order to allow such setup, a transformation from one feature space (which is the input space to our PEMs presented in Table 3) to another is needed. The features extracted fall in different ranges on the game-dependency dimension. Some of the features can be easily generalised such as the FPS features: the number of enemies E, the number of enemies hit e_{hit} and the average amount of time spent in each life t_{life}. These features carry more or less the same information as the IMB features: the total number of enemies E, the number of enemies killed k_{goompa} and the time spent playing the game t_{comp}, respectively. Other features can be interpreted according to the game such as the number of coins collected in IMB which, to some extents, can be viewed as the number of health items n_{health} collected in the FPS game since they can be considered as different types of rewarding schemes. Some of the features, however, such as the time spent using a specific type of weapon in FPS, t_{weapon}, the time spent in a big mode in IMB, t_{big}, and the average width of gaps, $\bar{G}_w$, are game-dependent and cannot be intuitively generalised. There are a small number of these features and for testing purposes we choose to fix their values to the average when evaluating the models. After feature generalisation, The accuracies of the models presented in Table 3 are recalculated on the data from the other dataset and presented in Table 4.

The results indicate a wide verity in the performance. The FPS models for predicting engagement for instance, although relatively accurate on the FPS data, performed poorly when evaluated on the data from players playing SMB. On the other hand, the FPS models for predicting frustration performed very well on both datasets. It is interesting to note that some of the models achieved better results when evaluated on another dataset than the one used for model construction. For instance, the IMB models for predicting

Table 4: Average accuracies of the models (M) when tested on players' data (D) from a different dataset. The models built on the IMB dataset are tested on the FPS dataset and vice versa.

	Engagement	Frustration	Challenge
FPS_M/SMB_D	31.47%	99.43%	72.01%
SMB_M/FPS_D	76.80%	55.61%	58.22%

engagement yield better results when evaluated on the FPS data than those obtained from the IMB data. In general, it appears that the differences in the performance obtained are not contingent on predicted affective state nor on the prediction models. This suggests that specific models have better generalisation ability than others. This is linked in part to the set of the input features, their accurate generalisation and the efficiency of the mapping process.

As discussed, although high accuracies are obtained is some cases, poor performance is observed in others which motivate conducting our second experiment.

Building Generic Models

In this experiment, we attempt to construct generic models from game-independent features as we believe that such models will be more accurate, and ultimately more generic, than those in the previous experiment. For this purpose, we generalise the features in the two datasets and use the full resultant set to construct new models. Practically speaking, the features in each dataset are transformed into a more abstract level so that they represent boarder notion of players' actions and interactions with the games' objects. The definitions are chosen in a way that preserve the meaning of the features while permitting their applicability to a wider range of games. Feature generalisation is performed by going through all features from both games one-by-one and trying to find a possible generalisation for it that can be scaled to other games. This process resulted in a new set of features that can be applied in both games. As not all fea-

Table 5: Features selected from the set of generic features for the prediction of engagement, frustration and challenge for the combined dataset of IMB and FPS. The table also presents the corresponding average performance ($\bar{P}_{avg}$) and the maximum (P_{max}) values obtained over five runs.

	Combined Dataset (FPS + SMB)		
	Engagement	Frustration	Challenge
$Selected$ $features$	n_{jump} E_{skill} E t_{still}	t_{life} t_{still} n_{death} e_{kill}	t_{life} t_{still} E
P_{avg}	70.88%	79.81%	76.89%
P_{max}	72.71%	80.99%	79.62%

tures can be scaled, some of the them are removed resulting in a total of 14 generic features taken from both games. Such features include for instance, the amount of time spent playing, moving or standing, the number of enemies killed, the number of times the player died, etc.

The dataset used to construct the models is the result of combining the data in the IMB and the FPS. The process of feature selection and model construction is followed to buid generic models and the experimental protocol presented in Section is followed. The results obtained are presented in Table 5.

Generic models with high prediction accuracies are obtained in all cases. The best models obtained are for predicting frustration with accuracy up to 81% while engagement is the hardest to predict (72.71%). The results are very promising and suggest that generic feature of player behaviour can indeed be built and used to infer affect across multiple games.

Notice that the preprocessing step of feature generalisation was necessary in the current study because we had the features in both datasets already collected. However, the purpose of this experiment is to demonstrate that one could benefit from designing data collection experiments with generalised features in mind, as those can be used almost directly to study behaviours in other similar games and the feature generalisation stage will no longer be necessery.

Conclusions

In this paper, we presented and evaluated a method for building generic models of PE. For this purpose, we used two datasets of player behaviour while interacting with games from different genres. To test whether generic PEM can be constructed, we employed previously constructed accurate PE models and we examined their input space in an attempt to generalise them. These models were then tested on another dataset than the one initially used to construct the models. The results obtained from this experiment suggest a potential for the method and a possibility for better prediction performance if the input space is optimised. As our modelling method is heavily dependent on the input features selected, starting from generic features and allowing the method to choose the relevant ones has the potential of improving the performance. To further investigate this, we ran another experiment where the feature space is generalised prior to model construction. PEM were then built with the generic features as input and the two dataset combined for training and testing. The experiment resulted in generic models of high accuracies for the prediction of engagement, frustration and challenge.

The experiments presented in this paper aimed at constructing models from players' data in two different games. Although the games are taken from dissimilar genres, they share some characteristics that allowed feature generalisation. However, we might not be able to model PE from data taken from players playing Tetris and SMB as there will be very few features in common. The ultimate goal of this line of research, is to construct models that could effectively work across many dissimilar game genres. This could be achieved if richer information about the players is considered. Generic models can be constructed that incorporate information about culture, knowledge, traits, demographics, preferences, and many other objective and subjective modalities. Although more powerful machine learning techniques might then become necessary, as the feature space will be huge and since the mapping is not straightforward, the potential of such systems is undeniably significant.

Another potential application for the experiments presented in this paper is the possibility of providing a better understanding of player behaviour in general that could infer and improve the design of user experiments. The generic models constructed can be used as indicators of informative features that impact affects and that designer should consider when designing a new game or when examining PE.

The experiments presented in this paper are still preliminary indicators of the potential of the approach. The method followed and the experimental setup used are motivated by previous research. However, since we are aiming at a different goal, it would be interesting to try other modelling approaches for preference learning such as random forests (Abou-Zleikha and Shaker 2015), multivariate adaptive regression spline (Abou-Zleikha, Shaker, and Christensen) models or active learning approaches (Shaker, Abou-Zleikha, and Shaker) as well as experimenting with different parameters tuned for the new data.

Also, the models constructed predict players' affect across three different affective states. With personalising PE as our ultimate goal, one need to investigate whether previous approaches for personalising PE are still valid (Shaker, Togelius, and Yannakakis 2010; Bakkes et al. 2014) and analyse new approaches to select the best combinations of content features that optimises player's experience across multiple dimensions.

Finally, we would like to point out that this paper draws the outlines of this line of research, and thereof the word *towards* in the title, and it illustrate that there are indeed common gameplay behavioural patterns that can be captured and that are scalable across games. The paper shed the light on an important field of research in the game domain and motivates further investigations.

Acknowledgments. This research was supported by the Danish Research Agency, Ministry of Science, Technology and Innovation project *PLayGAIe*; project number: 274-09-0083.

References

Abou-Zleikha, M., and Shaker, N. 2015. Evolving random forest for preference learning. In *Applications of Evolutionary Computation*. Springer. 318–330.

Abou-Zleikha, M.; Shaker, N.; and Christensen, M. G. Preference learning with evolutionary-based multivariate adaptive regression spline model.

Alrifai, M.; Gennari, R.; Tifrea, O.; and Vittorini, P. 2012. The user and domain models of the terence adaptive learning system. In *International Workshop on Evidence-Based Technology Enhanced Learning*, 83–90. Springer.

Bakkes, S.; Whiteson, S.; Li, G.; Visniuc, G. V.; Charitos, E.; Heijne, N.; and Swellengrebel, A. 2014. Challenge balancing for personalised game spaces. In *Games Media Entertainment (GEM), 2014 IEEE*, 1–8. IEEE.

Bakkes, S.; Tan, C. T.; and Pisan, Y. 2012. Personalised gaming: a motivation and overview of literature. In *Proceedings of the 8th Australasian Conference on Interactive Entertainment: Playing the System*, 4. ACM.

Bartle, R. 1996. Hearts, clubs, diamonds, spades: Players who suit muds. *Journal of MUD research* 1(1):19.

Calvo, R. A., and D'Mello, S. 2010. Affect detection: An interdisciplinary review of models, methods, and their applications. *Affective Computing, IEEE Transactions on* 1(1):18–37.

Conati, C. 2002. Probabilistic assessment of user's emotions in educational games. *Applied Artificial Intelligence* 16(7-8):555–575.

Conati, C. 2011. Combining cognitive appraisal and sensors for affect detection in a framework for modeling user affect. In *New Perspectives on Affect and Learning Technologies*. Springer. 71–84.

Cooper, A.; Reimann, R.; and Cronin, D. 2007. *About face 3: the essentials of interaction design*. John Wiley & Sons.

Heckmann, D.; Schwarzkopf, E.; Mori, J.; Dengler, D.; and Kröner, A. 2007. The user model and context ontology gumo revisited for future web 2.0 extensions. *Contexts and Ontologies: Representation and Reasoning* 37–46.

Hoque, M.; McDuff, D.; and Picard, R. 2012. Exploring temporal patterns in classifying frustrated and delighted smiles.

Isbister, K.; Schwekendiek, U.; and Frye, J. 2011. Wriggle: An exploration of emotional and social effects of movement. In *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems*, 1885–1890. ACM.

Kobsa, A. 2001. Generic user modeling systems. *User modeling and user-adapted interaction* 11(1-2):49–63.

Koster, R. 2004. *A theory of fun for game design*. Paraglyph press.

Malone, T. 1981. What makes computer games fun? New York, NY, USA: ACM.

Martínez, H.; Garbarino, M.; and Yannakakis, G. 2011. Generic physiological features as predictors of player experience. *Affective Computing and Intelligent Interaction* 267–276.

Niederée, C.; Stewart, A.; Mehta, B.; and Hemmje, M. 2004. A multi-dimensional, unified user model for cross-system personalization. In *Proceedings of the AVI 2004 Workshop on Environments for Personalized Information Access*, 34–54.

Ortony, A.; Clore, G.; and Collins, A. 1990. *The cognitive structure of emotions*. Cambridge university press.

Pantic, M., and Rothkrantz, L. 2003. Toward an affect-sensitive multimodal human-computer interaction. *Proceedings of the IEEE* 91(9):1370–1390.

Pedersen, C.; Togelius, J.; and Yannakakis, G. N. 2010. Modeling player experience for content creation. *IEEE Transactions on Computational Intelligence and AI in Games* 2(1):54–67.

Picard, R. 1995. *Affective computing*. Perceptual Computing Section, Media Laboratory, Massachusetts Institute of Technology.

Scherer, K. 2005. What are emotions? and how can they be measured? *Social science information* 44(4):695–729.

Sebe, N.; Cohen, I.; and Huang, T. 2005. Multimodal emotion recognition. *Handbook of Pattern Recognition and Computer Vision* 4:387–419.

Shaker, N.; Abou-Zleikha, M.; and Shaker, M. Active learning for player modeling.

Shaker, N.; Shaker, M.; Abu-Abdallah, I.; Al-Zengi, M.; and Sarhan, M. H. 2013. A quantitative approach for modelling and personalizing player experience in first-person shooter games. In *UMAP Workshops*.

Shaker, M.; Al-Zengi, M.; Abu-Abdallah, I.; Sarhan, M. H.; and Shaker, N. 2015. Personalizing content generation in first person shooter games through player modeling. In *to appear*.

Shaker, N.; Togelius, J.; and Yannakakis, G. N. 2010. Towards automatic personalized content generation for platform games. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE). AAAI Press*.

Shaker, N.; Yannakakis, G. N.; and Togelius, J. 2012. Crowdsourcing the aesthetics of platform games. *IEEE Transactions on Computational Intelligence and Games, Special Issue on Computational Aesthetics in Games*.

Sweetser, P., and Wyeth, P. 2005. Gameflow: a model for evaluating player enjoyment in games. *Computers in Entertainment (CIE)* 3(3):3–3.

Viviani, M.; Bennani, N.; and Egyed-Zsigmond, E. 2010. A survey on user modeling in multi-application environments. In *Advances in Human-Oriented and Personalized Mechanisms, Technologies and Services (CENTRIC), 2010 Third International Conference On*, 111–116. IEEE.

Wöllmer, M.; Kaiser, M.; Eyben, F.; Schuller, B.; and Rigoll, G. 2013. Lstm-modeling of continuous emotions in an audiovisual affect recognition framework. *Image and Vision Computing* 31(2):153–163.

Yannakakis, G., and Hallam, J. 2006. Towards Capturing and Enhancing Entertainment in Computer Games. In *Advances in Artificial Intelligence*, volume 3955 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg. 432–442.

Yannakakis, G. N., and Hallam, J. 2009. Real-time Game Adaptation for Optimizing Player Satisfaction. *IEEE Transactions on Computational Intelligence and AI in Games* 1(2):121–133.

Yannakakis, G. N., and Togelius, J. 2011. Experience-Driven Procedural Content Generation. *IEEE Transactions on Affective Computing*.

Yannakakis, G. N.; Maragoudakis, M.; and Hallam, J. 2009. Preference learning for cognitive modeling: a case study on entertainment preferences. *IEEE Transactions on Systems, Man, and Cybernetics. Part A* 39:1165–1175.

Zakharov, K.; Mitrovic, A.; and Johnston, L. 2008. Towards emotionally-intelligent pedagogical agents. In *Intelligent Tutoring Systems*, 19–28. Springer.

Zeng, Z.; Pantic, M.; Roisman, G.; and Huang, T. 2009. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31(1):39–58.