ICIC Express Letters ICIC International ©2020 ISSN 1881-803X
Volume **14**, Number **11**, November **2020** pp. 1097–1104

DYNAMIC DIFFICULTY ADJUSTMENT WITH FACIAL EXPRESSION RECOGNITION FOR IMPROVING PLAYER SATISFACTION IN A SURVIVAL HORROR GAME

Andrew[1], Adithya Nugraha Tjokrosetio[1] and Andry Chowanda[2]

[1]Computer Science Department, BINUS Graduate Program – Master of Computer Science
[2]Computer Science Department, School of Computer Science
Bina Nusantara University
JL. K. H. Syahdan No. 9, Kemanggisan, Palmerah, Jakarta 11480, Indonesia
{ andrew014; adithya.tjokrosetio }@binus.ac.id; achowanda@binus.edu

Received February 2020; accepted May 2020

Abstract. *Dynamic Difficulty Adjustment (DDA) is a process of matching challenge to player skill by adjusting game parameters dynamically to improve player satisfaction. While most DDAs are performance based, mapping multiple performance metrics to satisfaction can be demanding and genre dependent. Such limitations do not apply to emotion based DDA as it attempts to address the main issue of satisfaction directly, rather than using heuristic scores. This study aims to apply emotion based DDA with Facial Expression Recognition (FER) in order to improve player satisfaction. A survival horror game was designed and developed in 2 versions: One implements DDA and the other does not. An empirical experiment ($N = 31$) was conducted in which both versions were played consecutively. A usability test was applied to evaluating the game and the DDA model in terms of satisfaction. The result indicated that the game, as a survival horror, performed statistically well in providing satisfaction and displayed statistically significant improvement in various performance metrics while applying the proposed DDA model.*
Keywords: Game balancing, Dynamic difficulty adjustment, Facial expression recognition, Player satisfaction

1. **Introduction.** Video game is considered one of the most popular forms of digital entertainment. To be profitable, a video game must offer high satisfaction for its players. Player satisfaction is influenced by many variables such as graphical interface, storyline, input device, and game balancing [1]. Game balancing refers to methods for adjusting game difficulty by modifying game parameters, scenarios, and behaviors in order to avoid undesired player emotions such as boredom or frustration [2]. Traditional approach such as static game balancing involves setting a pre-defined difficulty level that increases as players advance. However, static difficulty may be problematic due to player skill diversity. Alternatively, Dynamic Difficulty Adjustment (DDA) can be applied to adjusting game difficulty dynamically (i.e., as the game progresses). DDA successfully addresses static difficulty problems [1,3,4], and thus is deemed necessary for improving player satisfaction [1].

Most DDA implementations [1,5,6] are based on player performance. However, mapping performance to satisfaction while dealing with multiple performance metrics can be demanding and genre dependent. These limitations give rise to emotion based DDA as an attempt to address the main issue of player satisfaction directly [7]. Emotion based DDA is also less dependent of game genre; thus one model is applicable to large number of games. [8,9] have explored emotion based DDA using facial expression, as interacting with the game often provokes player emotion manifested by facial expression [8]. Additionally, facial expression may produce insight of player satisfaction at certain point of time. As

DOI: 10.24507/icicel.14.11.1097

facial expression correlates to player satisfaction and improving player satisfaction is the ultimate goal of DDA, facial expression should be a suitable input for DDA.

The idea of using Facial Expression Recognition (FER) to capture facial expression as the input for DDA has been proposed in some literature, but empirically evaluated only by few. Two studies [8,9] developed simple games using Unity game engine, implemented rule based DDA with FER, conducted experiments ($N >= 30$) where 2 game versions (with and without DDA) were played consecutively, evaluated the DDA model using certain questionnaires, and reported that the DDA model significantly increases player experience. This work attempts to validate previous studies and provide novel contributions by: 1) exploring DDA practice in an uncommon game genre (survival horror); 2) proposing a new DDA with FER model optimized not only for matching challenge to player skill, but also for improving the horror aspect in a survival horror game; 3) proposing a new method for DDA evaluation by incorporating both subjective and objective measurements. This work is organized as follows: Section 1 covers introduction of the study; Section 2 lists some literature and works related to the study; Section 3 elaborates design, development, and evaluation methodology used in the study; Section 4 discusses results of the study; Section 5 concludes the study and provides recommendation for future research.

2. **Literature Review and Related Works.**

2.1. **Game balancing and dynamic difficulty adjustment.** As a part of game balancing, difficulty adjustment has always been a classic challenge. Koster [2] incorporated the concept of *flow channel* proposed by Csikszentmihalyi [10] into gaming scenario and suggested that players must be kept inside the flow channel to ensure optimal experience. In that case, game developers need to adjust difficulty level in a way that satisfies every player. Static game balancing, the traditional approach of game balancing, involves setting a pre-defined difficulty level that gradually increases as player advance. However, this approach creates a static difficulty curve which can lead to mismatches between player skill and game challenges [11]. To address static game balancing weaknesses, researchers have proposed another approach called dynamic game balancing. It involves Dynamic Difficulty Adjustment (DDA), process of dynamically tweaking game parameters to adjust game difficulty based on inputs, for instance, player performance. In some cases, game with DDA can determine player strategy quality to either increase or decrease the game difficulty [3]. DDA provides a more challenging gaming experience for everyone with various skill levels, thus retaining player interest throughout the game [11].

Several studies have demonstrated the superiority of DDA over static game balancing. Klimmt et al. [4] highlighted DDA ability to adjust the game with player performance and produce successful experience in every stage of the game. Andrade et al. [1] evaluated multiple dynamic game balancing approaches and stated that agents with DDA performed the closest to player level, thus providing the highest player satisfaction. Although there is no "general rule" on how to implement DDA [9], multiple approaches do exist. A simple, yet powerful approach would be building Artificial Intelligence (AI) script for game agents [9]. More complex approaches use dynamic computational intelligence where difficulty adjustment is based on player pattern evaluation [8]. Examples of this category include Hamlet system [5], multi-layered perceptron, and reinforcement learning [7]. These approaches will make DDA even more dynamic but require more computing resources.

2.2. **Evaluating player satisfaction.** Yannakakis and Hallam [6] categorized approaches for capturing player satisfaction into qualitative and quantitative. Qualitative approaches are based on empirical observations or linear correlations between user data and Likert scale questionnaire results, while quantitative approaches attempt to quantify entertainment using performance metrics. Andrade et al. [1] proposed a usability test which adopts Maguire's evaluation methods for human-centered design [12]. The test includes controlled user testing, satisfaction questionnaire, and post-experience interview. Controlled user testing is done by measuring player performance and feedback; satisfaction questionnaire is applied to collecting subjective data using a certain Likert scale; post-experience interview is conducted to obtain data not covered by the questionnaire.

2.3. **Related works.** Being the only few studies which empirically evaluated the use of DDA with FER, Moniaga et al. [8] and Akbar et al. [9] developed simple games using Unity game engine. They applied rule based DDA with AFFDEX SDK [13] as the FER system and conducted experiments ($N >= 30$) in which 2 game versions (with and without DDA) were played consecutively. Moniaga et al. [8] evaluated the DDA model using an immersion questionnaire [14], while Akbar et al. [9] utilized game experience questionnaire [15], a questionnaire designed to measure 7 components of gaming experience: Competence, Sensory and Imaginative Immersion, Flow, Tension/Annoyance, Challenge, Positive Affect, and Negative Affect. Both studies concluded that their DDA model significantly increases player experience. Such findings present a compelling notion that facial expression can be utilized to produce successful difficulty adaptation. While the studies applied generally standard methodologies and produced satisfying results, further validation is required as questionnaire-based evaluations only incorporate subjective measurements.

3. **Methods.** This study consists of 3 main phases: design, development, and evaluation. In design phase, a simple survival horror game was designed. Survival horror is a video game genre in which players try to survive the game while being frightened with horror graphics and ambience. Two scaring techniques often used in survival horror games are surprise and suspense, where successful combination between those two may present a terrifying yet entertaining experience, promoting player satisfaction in horror-themed video games. Survival horror was selected as the game genre mainly due to its ability to provoke unique and colorful emotions (e.g., surprise and fear). Such ability may enable DDA to not only match challenge to player skill, but also improve horror aspect of the game. Furthermore, survival horror has been gaining popularity; thus empirical study of DDA practice in this genre might contribute for survival horror game development in general.

The game itself was designed using the famous video game *Five Nights at Freddy's* (*FNaF*) as its reference for the following reasons: 1) *FNaF* is popular enough to represent survival horror genre; 2) *FNaF* is a point-and-click game, which is relatively easy to master and therefore suitable even for novices; 3) *FNaF* offers unique gameplay that emphasizes on the horror aspect rather than the action. The game adopts *FNaF* gameplay and objective with some modifications to make it simpler and more suitable for experiment purpose. Player acts as a character in first-person trying to survive the night in a security room of a haunted building using 2 lamps and 2 electric doors. Player should carefully manage power consumption by turning on lamps and shutting doors only if necessary. Player is also provided with a camera control panel, which can be used to observe frightening figure wandering around the building and occasionally toward the security room. Lamps are used to reveal whether the figure has reached the door, and when it does, player must immediately close that door. Failure to do so might cause figure to successfully enter the security room and jump scare the player, causing game over. However, if player survives the night (5 minutes of real time) without ever receiving a jump scare, the game ends in victory. If player runs out of power, a blackout will occur, forcing doors to stay open and ultimately leading to game over, unless the player makes it to the end of the night.

In development phase, the game was developed using Unity game engine. The development implemented Scrum as software development framework. The game was developed in 2 versions: one implements DDA (Dynamic Version) and one does not (Static Version). Both versions utilize the same FER system: AFFDEX SDK from Affectiva [13]. The SDK processes 3D data, which is better at capturing and preserving facial features compared to 2D data [16], to recognize up to 7 emotions and 15 facial expressions. In both versions, their values are recorded as part of performance metrics. In Dynamic Version exclusively, their values are also interpreted and passed into game mechanics in real time to serve as input for DDA. Meanwhile, the proposed DDA model revolves around the adjustment of one game parameter: *Figure Movement Interval* (*FMI*). *FMI* represents the amount of seconds before the next figure movement. Hence, smaller *FMI* will result in a more aggressive figure movement, causing figure to threaten the player more often. Formula (1) and Formula (2) are used for calculating *FMI* for Static Version (FMI_{STATIC}) and *FMI* for Dynamic Version ($FMI_{DYNAMIC}$), respectively. At the beginning of the game, *FMI* is calculated, and the countdown starts. When the countdown reaches 0, the figure moves and *FMI* is recalculated for the next countdown. This iteration continues until the game ends.

$$FMI_{STATIC} = (30 - 5 * \beta) + R \tag{1}$$

$$FMI_{DYNAMIC} = (30 - 5 * \beta - \alpha) + R \tag{2}$$

$$\alpha = \min(\alpha_{TEMP}, 6) \tag{3}$$

In all versions, *FMI* is affected by survival time: For every minute passed (β), *FMI* gradually becomes smaller. To keep the game less predictable, *FMI* is also affected by random value (R). R is a decimal number between 0 and 4 generated using pseudo-random number generator. In Dynamic Version exclusively, *FMI* is further affected by emotion/expression value (α). Despite having the ability to capture surprise and fear, neither emotions are used to define α. A validation study by Stöckli et al. [17] reveals that AFFDEX performs poorly in detecting fear and often confuses fear with surprise. Alternatively, valence and attention value are selected. AFFDEX SDK assigns valence value $[-100, 100]$ by calculating degree of emotions based on iMotions emotion classification [18]. AFFDEX SDK assigns attention value $[0, 100]$ by calculating face orientation towards the webcam. Positive valence and low attention are not desired during the game, as they suggest either mismatch between player skill and challenge or inadequate horror aspect delivered. Therefore, they both will influence *FMI* value. The actual value of α is then calculated as follows. When the game begins, both α and temporary variable α_{TEMP} are set to 0. At the same time, FER system starts capturing and interpreting player face every second. If an interpretation suggests positive valence (valence $>=$ 10) or low attention (attention $<$ 10), α_{TEMP} is incremented by 1. If FER system cannot detect any face, player is considered not paying attention, setting valence and attention value to 0 and by previous rule, incrementing α_{TEMP} by 1. Every 60 seconds, the accumulated α_{TEMP} is used to define α using Formula (3) before being reset back to 0. In other words, valence and attention values accumulated in the first minute determine the $FMI_{DYNAMIC}$ for the next minute.

In evaluation phase, both the game and the DDA model were evaluated. An empirical experiment was conducted in which participants played both versions back-to-back. For preserving validity and avoiding biases, these procedures were applied for every participant: 1) Participant filled Demography Form; 2) Participant played the Tutorial Version while being informed about the basic gameplays; 3) Participant rolled random number $[1, 100]$ to determine the order of the games. Participant plays either Dynamic Version first or Static Version first if even number was rolled; 4) Participant played the first game; 5) Participant played the second game; 6) Participant filled satisfaction questionnaire; 7) Participant answered interview questions.

The experiment goal is to evaluate player satisfaction for both the game and the DDA model by utilizing a usability test proposed by Andrade et al. [1]. The test was selected in the hope of producing more accurate representation of player satisfaction, as it combines quantitative and qualitative approaches. In controlled user testing, several performance metrics (Table 1) were recorded using game scripts. Positive valence rate and low attention rate are the core metrics for DDA model evaluation. Lower value of these metrics implies better player satisfaction. The rest of the metrics will provide insights regarding difficulty level and player competence. Then, satisfaction questionnaire derived from the Game Experience Questionnaire (GEQ) [15] was filled by participants to evaluate the game in 6 components. Competence is not measured, as objective measurement of competence (based on performance metrics) should yield more accurate result. Twelve questions (Table 2) were extracted from GEQ – Core Module and 5-point Likert scale scoring between 0 and 4 was applied. Lastly, post-experience interview was conducted to collect supporting data. Three questions were inquired regarding the horror aspect of the game, version difference, and version preference. Responses were grouped into different categories that may provide subjective feedback towards the game and the DDA model.

Table 1. List of performance metrics used in the game

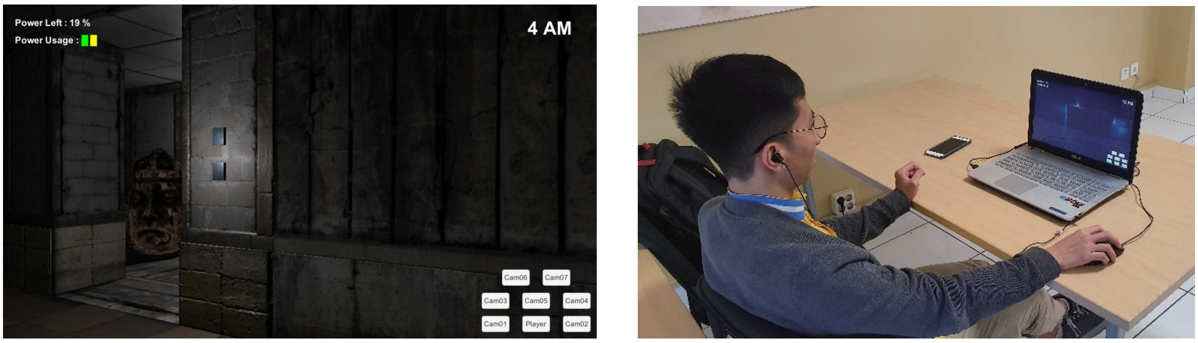
Performance metric	**Definition**	**Purpose**
Positive valence rate	The amount of time FER system captures positive valence (valence >= 10) divided by overall game time	DDA evaluation
Low attention rate	The amount of time FER system captures low attention (attention < 10) divided by overall game time	DDA evaluation
Win rate	Number of wins divided by number of losses	Difficulty level evaluation
Average game time	The average time players spend in the game in seconds (maximum is 300)	Player competence evaluation

Table 2. List of questions used in the satisfaction questionnaire

Game component	**Question 1**	**Question 2**
C1: Sensory and Imaginative Immersion	I found it impressive	It felt like rich experience
C2: Flow	I lost track of time	I was deeply concentrated in the game
C3: Tension/Annoyance	I felt annoyed	I felt frustrated
C4: Challenge	I felt pressured	I felt challenged
C5: Negative Affect	I found it tiresome	I felt bored
C6: Positive Affect	I thought it was fun	I enjoyed it

4. **Result and Discussion.** Figure 1 shows the result of design/development phase and the process of evaluation phase.

The experiment involved 31 participants (age 18-30, 74.19% male, 16.13% liked horror entertainment, 58.06% have played or watched other people play *FNaF* before). After conducting the experiment, core metrics for DDA model evaluation were obtained. Kolmogorov-Smirnov and Shapiro-Wilk test were applied and revealed that both metrics most likely do not come from a normal distribution data ($p < 0.05$). Hence, Wilcoxon signed-rank test was selected as a non-parametric statistical hypothesis test to statistically compare the metrics in both versions. The test result reveals a statistically significant difference ($p < 0.05$) for positive valence rate ($p = 0.037$) and low attention rate ($p = 0.043$),



(a) (b)

FIGURE 1. (a) Screenshot of the game; (b) a participant playing the game

		Static Version	Dynamic Version
Positive Valence Rate	Mean	0.14	0.0977
	Standard Deviation	0.1414	0.0973
Low Attention Rate	Mean	0.2471	0.16
	Standard Deviation	0.2655	0.1642

(a)

	Static Version	Dynamic Version
Win Rate	61.29%	41.94%
Average Game Time	90.49% (271.48 seconds)	81.08% (243.24 seconds)

(b)

FIGURE 2. (a) DDA metrics comparison; (b) other metrics comparison

where the metrics values in Dynamic Version are lower than those in Static Version (Figure 2(a)). Therefore, the DDA model statistically provides higher player satisfaction by matching challenge to player skill and improving the horror aspect of the game.

The rest of the metrics were compared. The result (Figure 2(b)) shows that player wins more often and survives longer in Static Version, corresponding to the fact that the Dynamic Version with its DDA model may present higher difficulty. Both versions, however, produce balanced (i.e., near 50%) win rate, indicating that the difficulty levels of both versions are well-adjusted. Above 80% average game time suggests that the player competence in the game is relatively well since the average players can survive at least 4 out of 5 minutes game time, preventing multiple game over scenarios in the process.

Satisfaction questionnaire results were collected. Six game components were measured with negative affect (C5) being the only undesired component in survival horror game. Composite scores for every game component were obtained by averaging the two questions measuring the component. Central tendency and variability were acquired by calculating mean and standard deviation, respectively. The result (Figure 3) indicates that all desired game components (C1, C2, C3, C4, and C6) average above 2.00, while the undesired game component (C5) averages below 2.00, concluding that the game statistically satisfies the players in every desired and undesired game component.

Post-experience interview results were compiled. According to the responses, 77.42% participants agreed that the game is scary/thrilling, validating the game capacity as a survival horror. 70.97% participants noticed valid difference between the 2 versions and 81.82% of them agreed that the Dynamic Version is more challenging, showing that the DDA model truly works as intended. When asked to play the game once more, 61.29% participants preferred the Dynamic Version, mostly due to the more challenge it provides.

	C1	C2	C3	C4	C5	C6
Mean	2.8548	2.7419	2.0968	2.8387	1.5161	2.8871
Standard Deviation	0.721	0.8744	1.2545	0.7894	0.747	0.8237

FIGURE 3. Central tendency and variability for every game component

5. **Conclusion and Future Works.** This study demonstrates on how FER can be used to provide input for emotion based DDA. A survival horror game based on the famous video game *Five Nights at Freddy's* was designed and developed in 2 versions: Dynamic Version implements DDA and Static Version does not. Both versions were played consecutively in an empirical experiment ($N = 31$). The game and the DDA model were evaluated by utilizing a usability test. Positive valence rate and low attention rate were the performance metrics selected for DDA model evaluation. Wilcoxon signed-rank test was applied and revealed a statistically significant improvement ($p < 0.05$) for both metrics in the Dynamic Version. Such improvement, supported with results from other performance metrics, satisfaction questionnaire, and post-experience interview, conclude that the game, as a survival horror, performed statistically well in providing player satisfaction and can be further enhanced by applying the proposed DDA model.

Future works for this study would be improving the FER aspect by applying a FER system capable of interpreting a broader range of emotions/expressions with higher accuracy and passing them into the game mechanics. Additionally, future experiments may apply multimodal emotion recognition to addressing FER weaknesses. Combining FER with other biometric sensors such as microphone or electrocardiogram may enable system to interpret player emotion with higher confidence level. Study of emotion based DDA can be further extended by exploring its utilization in other horror-themed games or other game genres.

REFERENCES

[1] G. Andrade, G. Ramalho, A. S. Gomes and V. Corruble, Dynamic game balancing: An evaluation of user satisfaction, *AIIDE*, vol.6, pp.3-8, 2006.
[2] R. Koster, *Theory of Fun for Game Design*, O'Reilly Media, Inc., Sebastopol, California, 2013.
[3] R. Sutoyo, D. Winata, K. Oliviani and D. M. Supriyadi, Dynamic difficulty adjustment in tower defence, *Procedia Comput. Sci.*, vol.59, pp.435-444, 2015.
[4] C. Klimmt, C. Blake, D. Hefner, P. Vorderer and C. Roth, Player performance, satisfaction, and video game enjoyment, *International Conference on Entertainment Computing*, pp.1-12, 2009.
[5] R. Hunicke and V. Chapman, AI for dynamic difficulty adjustment in games, *AAAI Work. – Tech. Rep.*, vol.WS-04-04, pp.91-96, 2004.
[6] G. N. Yannakakis and J. Hallam, Real-time game adaptation for optimizing player satisfaction, *IEEE Trans. Comput. Intell. AI Games*, vol.1, no.2, pp.121-133, 2009.
[7] M. Zohaib, Dynamic difficulty adjustment (DDA) in computer games: A review, *Adv. Human-Computer Interact.*, pp.1-12, 2018.
[8] J. V. Moniaga, A. Chowanda, A. Prima, Oscar and M. D. T. Rizqi, Facial expression recognition as dynamic game balancing system, *Procedia Comput. Sci.*, vol.135, pp.361-368, 2018.
[9] M. T. Akbar, M. N. Ilmi, I. V. Rumayar, J. Moniaga, T.-K. Chen and A. Chowanda, Enhancing game experience with facial expression recognition as dynamic balancing, *Procedia Comput. Sci.*, vol.157, pp.388-395, 2019.
[10] M. Csikszentmihalyi, *Flow: The Psychology of Optimal Experience*, Harper & Row, New York, 2009.
[11] R. Hunicke, The case for dynamic difficulty adjustment in games, *Proc. of the 2005 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology*, pp.429-433, 2005.
[12] M. Maguire, Methods to support human-centred design, *Int. J. Hum. Comput. Stud.*, vol.55, no.4, pp.587-634, 2001.
[13] D. McDuff, A. Mahmoud, M. Mavadati, M. Amr, J. Turcot and R. el Kaliouby, AFFDEX SDK, *Proc. of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pp.3723-3726, 2016.

[14] C. Jennett et al., Measuring and defining the experience of immersion in games, *Int. J. Hum. Comput. Stud.*, vol.66, no.9, pp.641-661, 2008.

[15] W. A. IJsselsteijn, Y. A. W. de Kort and K. Poels, *The Game Experience Questionnaire*, https://research.tue.nl/en/publications/the-game-experience-questionnaire, 2013.

[16] W. Wei and Q. Jia, 3D facial expression recognition based on Kinect, *International Journal of Innovative Computing, Information and Control*, vol.13, no.6, pp.1843-1854, 2017.

[17] S. Stöckli, M. Schulte-Mecklenbeck, S. Borer and A. C. Samson, Facial expression analysis with AFFDEX and FACET: A validation study, *Behav. Res. Methods*, vol.50, no.4, pp.1446-1460, 2018.

[18] IMotions, *Facial Expression Analysis: The Complete Pocket Guide*, 2017.