Procedural content generation based on a genetic algorithm in a serious game for obstructive sleep apnea

Konstantinos Mitsis
School of Electrical and Computer Engineering
National Technical University of Athens
Athens, Greece
kmhtshs@biosim.ntua.gr

Eleftherios Kalafatis
School of Electrical and Computer Engineering
National Technical University of Athens
Athens, Greece
leftkal@gmail.com

Konstantia Zarkogianni
School of Electrical and Computer Engineering
National Technical University of Athens
Athens, Greece
kzarkogianni@biosim.ntua.gr

George Mourkousis
Biomedial Simulations and Imaging Lab.,
National Technical University of Athens
Athens, Greece
mourkousis@biosim.ntua.gr

Konstantina S. Nikita
School of Electrical and Computer Engineering
National Technical University of Athens
Athens, Greece
knikita@ece.ntua.gr

***Abstract*— In this paper, a procedural content generation (PCG) technique that was incorporated in a novel serious game for obstructive sleep apnea (OSA) is presented. The technique is based on a genetic algorithm and aims to enhance user engagement and deliver educational material tailored to user needs. The genetic algorithm monitors user choices and game progress by means of two fitness functions that dictate suitable candidates to produce offspring in each generation. An initial validation in terms of user experience was conducted, by deploying three different versions of the serious game. Version A and B incorporated the genetic algorithm, along with mechanisms for automated game difficulty adjustment. Version A was designed to be difficult and frustrating while version B to adjust difficulty smoothly. Version C did not display any adaptive properties. 42 participants were recruited and split in two groups to play two versions of the game, A and C or B and C, without any prior knowledge of differences between them. After each session, two modules of the Game Experience Questionnaire (GEQ) were applied. The obtained results reveal statistically significant differences regarding user perception in terms of competence, challenge and negative experience for versions A and C respectively, and competence and negative experience for versions B and C respectively. Version B achieved better GEQ scores than version C while version A resulted in worse GEQ scores than version C.**

Keywords—procedural content generation, genetic algorithm, serious game, obstructive sleep apnea

I. Introduction

Procedural content generation (PCG) is a term used to describe techniques incorporated in games to empower user engagement and increase replay value by generating new content based on user choices and interaction with the game automatically [1]. Some modern examples of video games that employ such techniques are “Stardew Valley” (ConcernedApe and Chucklefish, London, England) incorporating map generation, reward and difficulty adjustment, “No Man's Sky” (Hello Games Ltd, Guilford, England) incorporating planet and landscape generation and the Borderlands series (Gearbox Software L.L.C, Frisco, Texas and 2K Australia Pty Ltd, Canberra, Australia) incorporating weapon generation. PCG techniques are based on rule based systems, random variables and artificial intelligence (AI).

AI describes the ability of a computer system to simulate characteristics corresponding to human intelligence in order to solve problems. AI methods often rely upon heuristic search functions to achieve their goals. Heuristics act as an evaluator, consulting the algorithm about which step is best to make. One example of such search heuristics is genetic algorithms (GA). Drawing inspiration from Darwin’s theory of evolution, GAs are search methods based on the principles of natural selection and genetics [2]. GAs have been used in PCG techniques due to their ability to produce highly customized content for a game, by “evolving” it according to the progress of the user.

Serious games constitute a widely recognized and effective means for educating, raising awareness and driving behavioral changes [3, 4]. In order to achieve these goals, serious games make use of heterogeneous and complex content and mechanics, which benefit greatly from PCG techniques [5].

In this paper, we present a PCG technique incorporated in a novel serious game that aims to raise awareness regarding Obstructive Sleep Apnea (OSA) while promoting effective behavior change. The proposed technique is based on a genetic algorithm and aims at increasing user engagement levels while delivering personalized educational content.

II. The Serious Game – “Wake Up for the Future”

“Wake Up for the Future!” is a serious game targeting adults, which aims to raise awareness and promote self-disease management regarding OSA. OSA constitutes the most prevalent sleep-related disordered breathing condition [6]. OSA manifests with recurrent episodes of upper airway collapse that result in a decline, or even interruption, of airflow with a duration of at least 10 seconds. Besides a variety of treatment options, certain behavioral changes can benefit patients suffering from OSA (e.g. weight loss, healthy diet options, limitation of alcohol and tobacco, proper sleeping routines and positions) [6]. To the best of our knowledge this is the first serious game for OSA.

A. Overview

“Wake Up for the Future”, is an open world video game, featuring debate duels with the use of card decks (Fig. 1). It

Part of this research was supported within the framework of ENDORSE project, which is funded by the NSRF. Grant agreement: T1EΔK-03695

978-1-7281-4533-4/20/$31.00 @2020 IEEE

Fig. 1. "Wake Up for the Future"

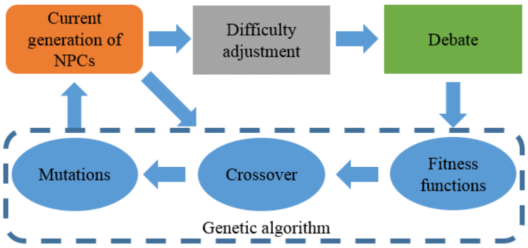

Fig. 2. Game content generation technique based on a genetic algorithm

has been implemented on the Unity platform and runs on Windows operating systems. Its design is based on a conceptual framework, linking the game mechanics with raising awareness and behavioral motivation, while targeting the adult population. It features a novel and suspenseful plot with the user traveling from a dystopian future where all knowledge regarding OSA is lost, to the present.

B. *Game description*

The user participates in debates with non-player characters (NPCs). These NPCs represent people with undiagnosed cases of OSA, unaware of their condition and lacking the necessary knowledge skills. The user's goal is to provide the NPCs with convincing arguments and contradict their false beliefs and unhealthy habits. In this way, the user takes up the role of a mentor, striving towards raising awareness and promoting self-disease management regarding OSA. The debates are simulated with a card game, where each card represents an argument, which is linked to a particular attribute.

Attributes are habits (smoking, alcohol consumption, sleeping position, etc.) and chronic conditions (obesity, hypertension, diabetes, etc.) that are relevant to the onset, diagnosis and progress of OSA. Each NPC is characterized by a number of the above attributes, forming a profile, which is presented to the user through a short biography before each debate. Each of these attributes is linked to a false argument the NPC presents during the debate, which strengthens his/her lack of motivation towards healthy lifestyle change. The short biography provides the user with insight about the upcoming debate, enabling him to prepare his arguments properly.

The debates are simulated by a card game system. There are two types of cards, NPC and user cards:

- NPC cards represent false arguments linked to an attribute. Each NPC possesses one NPC card for each attribute in his/her profile.
- User cards represent true arguments linked to an attribute. Before each debate, the user must form a deck of five user cards. There is a pool of available cards, containing two user cards per attribute.

The card game is played in rounds. The user has a maximum hand size of two cards. Before the first round, the user has the option to reshuffle his starting hand and draw a new one. At the start of each of the subsequent rounds, the user draws one new card, respecting the limit of maximum hand size. All of the NPC cards start the debate on the board facing down. At the start of every round, the NPC flips one of his/her cards open, if none of his/her cards are open.

On each round the user has the following available actions:

1) Play up to two cards from his/her hand. A user card destroys a face up NPC card of the corresponding attribute, by contradicting its false argument. If no such NPC card is facing up, then the user card reveals a face down NPC card of the corresponding attribute. If no such NPC card exists, the user card has no effect. When a card is played, it is discarded.
2) Pass the turn. In this case, the user has the option to reshuffle and draw up to two cards from his/her deck.
3) Surrender. The debate ends with a losing resolution.

The user wins the debate if he/she destroys all of the NPC cards within five rounds.

III. Game Content Generation

A PCG technique (Fig. 2) was designed and incorporated in the serious game, to empower the game's educational value and improve overall game experience. The proposed technique is based on a genetic algorithm and is responsible for the automated generation of new NPCs, based on user choices and game progress. The resulting adaptive serious game possesses the ability to automatically adjust difficulty levels and present educational content tailored to the user's needs.

A. *Genetic Algorithm*

An initial population, with NPCs as individuals that are generated randomly, is defined as the first generation. After every debate a new generation of NPCs is produced as offspring resulting from the fittest individuals of the previous generation. For each debate an NPC is chosen randomly from the current generation as an opponent. The available attributes, that characterize the NPCs, are joined in a string as genes to form chromosomes. Each chromosome has a number of genes that take binary values, "1" if the particular attribute exists in the NPC's profile and "0" if not. An example of a chromosome characterizing an NPC with 7 attributes is shown in Table I.

The number of expressed genes is proportional to the difficulty of the resulting debate, as every gene is translated in an NPC card the user must face. Furthermore, the genes are directly linked to knowledge domains regarding OSA in the form of attributes, constituting the core of the serious game's educational content. The genetic algorithm adjusts game difficulty dynamically and presents personalized educational content by selecting the fittest individuals to produce offspring, based on fitness scores calculated after each debate.

TABLE I. EXAMPLE CHROMOSOME WITH 7 ATTRIBUTES

Smoking	*Alcohol*	*Medication*	*Sleeping Position*	*Obesity*	*Hypertension*	*Depression*
S	A	M	SP	O	H	D
1	0	1	0	0	0	1

B. Fitness functions

The proposed technique employs two fitness functions, the winning fitness function (WFF) (1) and the losing fitness function (LFF) (2):

$$WFS = W_a * a + W_b * b + \ldots + W_x * x \quad (1)$$

$$LFS = L_a * a + L_a * b + \ldots + L_a * x \quad (2)$$

After each debate, the fitness scores WFS and LFS are calculated for every individual of the current population. The parameters (a, b,…, x) are the binary values of the genes in each individual. Parameters W_x and L_x represent weights with initial values of zero, which are trained during each debate according to the following rules:

1) Training rules for W_x:

a) $W_x = W_x - 1$, ∀ attribute of the opponent NPC.

b) $W_x = W_x - 1$, ∀ x linked to user card played correctly.

c) $W_x = W_x + 1$, ∀ x linked to user card played wrongly.

d) $W_x = W_x + 1$, ∀ x linked to user card left unplayed.

2) Training rules for L_x:

a) $L_x = L_x + 1$, ∀ attribute of the opponent NPC.

b) $L_x = L_x + 1$, ∀ x linked to user card played correctly.

c) $L_x = L_x + 1$, ∀ x linked to user card played wrongly.

d) $Lx = Lx - 1$, ∀ x linked to user card left unplayed.

If the player wins the current debate, the highest WFS scores are used to determine the fittest individuals of the current generation, benefiting mostly NPCs with new attributes and attributes that were involved in wrong user choices. If the player loses the current debate, the highest LFS scores are used to determine the fittest individuals among the current generation. This way, the player who suffers losses will face opponents with similar attributes so as to review educational content that he/she does not seem to comprehend.

C. Crossover, mutations and difficulty adjustment

The fittest NPCs form pairs, and each pair produces two offspring for the next generation, by exchanging genes according to a random crossover point. Number of selected fittest NPCs is selected so that population size of each generation remains the same. In the resulting generation, each NPC has a small chance to undergo a mutation that will change one gene from “0” to “1”. The mutation mechanism promotes some variability in the resulting population.

An NPC from the resulting generation is selected as an opponent for the next debate. The difficulty of the serious game is automatically adjusted by a rule based system that selects NPCs with a specific number of attributes. This number of attributes is defined according to user losses and victories in the previous debates.

IV. Material and Method

In order to conduct an initial validation of the content generation technique, in terms of user experience, a blind experiment was designed. Three versions of the serious game were developed, version A, B and C. In each version, the user participated in five iterations of the debate card game, versus different NPCs. A pool of seven possible attributes was available for NPC generation (Table I). An initial population of 20 NPCs, characterized by exactly 3 attributes, was generated randomly for all versions. After each debate, 5 fittest NPCs were selected to produce offspring paired in all possible ways. A basic tutorial was provided through in-game dialogues at the start of each version. All versions were visually indistinguishable to the user.

1) Version A, incorporated the genetic algorithm. Weights of WFF and WLF were not re-initialized after each debate. NPCs with one additional attribute were selected from the resulting generation with each win and one less for each loss (min 2 and max 5).

2) Version B, incorporated the genetic algorithm. Weights of WFF and WLF were re-initialized after each debate. For difficulty adjustment, a rating value with an initial value of “1” was assigned to the user. This value increased by 1 with each win and reduced by 1 with each loss. NPCs were chosen randomly from the current generation according to the following rules:

a) Rating between 1 and 3: 3 attributes

b) Rating above 3 (max 5): 4 attributes

c) Rating below 1 (min -2): 2 attributes

3) Version C, generated the attributes characterizing the NPCs in a random manner. Number of attributes of the NPCs was also chosen randomly, between 2 and 4, with the exception of the NPC of the first debate, who was characterized by 3 attributes.

A total number of 42 participants (Table II) enrolled for this initial validation process. Participants were split in two groups. Group 1 played versions A and C, while group 2 played versions B and C. Participants in both groups played their corresponding versions, one on each of two consequent days. The participants were not informed of any differences between the two versions, and the order in which they played each version was random. In order to evaluate user experience, two modules of the Game Experience Questionnaire [7] were deployed after each session. The core module provides insight about competence, sensory and imaginative immersion, flow, tension, challenge, negative affect and positive affect. The Post-game module asserts positive experience, negative experience, tiredness and returning to reality. Paired sample t-test were applied on the results to investigate significant differences between scores obtained from the two versions.

V. Results

Scores of the GEQ are presented in Table III, for groups 1 and 2, respectively. Results indicate differences in terms of user experience between adaptive and non-adaptive versions of the serious game. Analysis of the results revealed a statistically significant decrease in competence ($t(22)=2.46$, $p=0.021$), and an increase in challenge ($t(22)=-2.48$, $p=0.020$) and negative experience ($t(22)=-3.42$, $p=0.002$), between versions C and A respectively, played by group 1. On the other hand, the feeling of competence increased significantly ($t(18)=-2.30$, $p=0.033$) and negative experience dropped ($t(18)=2.53$, $p=0.020$) between versions C and B respectively. All other dimensions of the GEQ did not show statistically

TABLE II. PARTICIPANTS OF THE STUDY

	All (N=42)	*Group 1 (N=23)*	*Group 2 (N=19)*
Gender	male (25), female (17)	male (12), female (11)	male (13), female (6)
Age	27.90 ± 4.93	29.56 ± 4.77	25.89 ± 4.35

significant differences ($p>0.05$) between game versions for both groups.

A small set of participants (N=10) were chosen at random from both groups, to participate in semi-constructed interviews after both sessions had ended. Participants were asked if they understood the difference between the two versions and to provide general feedback regarding the serious game. All of the participants stated that while they understood differences in game difficulty between versions they were not certain of the mechanics governing each version. Additional comments involved problems with UI elements, the tutorial and game bugs that occurred during the sessions.

VI. DISCUSSION

Initial validation of the proposed content generation technique indicated that participants experienced differences between the game versions deployed. Version C, lacking any adaptive features, was used as a benchmark to test the effect of the other two versions on user experience. Versions A and B were designed to test the potential of the proposed technique in two opposite directions. Version A punished successful progress by increasing difficulty to near impossible levels. Weights in the fitness functions were not re-initialized, hence NPC attributes linked to user errors were promoted by the genetic algorithm perpetually to future generations. Additionally, the number of NPC attributes would increase after each winning. The effects of this harsh difficulty adjustment system were evident in GEQ score results observed in group 1. Sense of competence was significantly lower in version A, while negative experience and challenge was higher than those observed in version C.

In contrast, version B was designed to provide a smooth adaptive experience. The weights of the fitness functions were re-initialized after each debate. This feature allowed for a wider variety of attributes to be promoted to next generations, resulting in more heterogeneous populations of NPCs. In this way, the user would confront NPCs that would share attributes linked to past mistakes, along with novel ones, resulting in more engaging game content. Furthermore, the number of NPC attributes scaled in a slow and controlled way, avoiding the sudden spikes in game difficulty. GEQ score results obtained from group 2 are consistent with design considerations of version B, as negative experience was significantly lower in comparison to version C, and sense of competence significantly higher. Interestingly, sense of challenge did not display significant differences between versions B and C. This fact indicates that the proposed technique has the potential, if applied properly, to promote the perception of user competence without effecting perception of difficulty.

Results from this small scale preliminary validation of the proposed method indicate its potential to enhance user experience while delivering personalized educational content in the context of a serious game. This observation was possible, despite the fact that the deployed versions of the serious game were limited in content and duration. Chromosomes in the proposed genetic algorithm consisted of a small set of 7 genes. Additionally, only 5 generations of NPCs were produced in each play through. Future validation will deploy the main version of the serious game, incorporating larger gene sets and more generations of NPCs.

Insight gained from this validation needs to be applied in the design and development of future versions, incorporating the content generation technique. The effect of the PCG technique in the serious game's educational value needs to be investigated by means of a randomized controlled trial. Finally, the generalization capabilities of the proposed method need to be investigated in other card-based serious games.

TABLE III. RESULTS FOR GROUPS 1 AND 2

GEQ dimension (scale 0-4)	Mean ± Standard Deviation			
	Group 1		*Group 2*	
	Version A	*Version C*	*Version B*	*Version C*
Core Module				
Competence	**2.03** ± 0.87	**2.46** ± 0.64	**2.45** ± 0.71	**2.13** ± 0.84
Immersion	2.21 ± 0.84	2.14 ± 0.66	2.28 ± 0.72	2.25 ± 0.78
Flow	1.88 ± 0.95	1.69 ± 0.86	1.60 ± 0.84	1.67 ± 0.95
Tension	0.47 ± 0.57	0.31 ± 0.45	0.15 ± 0.27	0.21 ± 0.43
Challenge	**1.31** ± 0.50	**0.95** ± 0.42	0.84 ± 0.36	1.04 ± 0.44
Negative Affect	0.84 ± 0.89	0.59 ± 0.50	0.39 ± 0.35	0.47 ± 0.42
Positive Affect	2.34 ± 0.67	2.37 ± 0.75	2.45 ± 0.76	2.53 ± 0.77
Post-game Module				
Positive experience	1.29 ± 0.86	1.50 ± 0.71	1.53 ± 0.86	1.49 ± 0.78
Negative experience	**0.27** ± 0.30	**0.12** ± 0.24	**0.02** ± 0.08	**0.11** ± 0.17
Tiredness	0.06 ± 0.22	0.04 ± 0.20	0.13 ± 0.35	0.13 ± 0.31
Returning to reality	0.55 ± 0.56	0.46 ± 0.40	0.36 ± 0.56	0.47 ± 0.53

VII. CONCLUSION

This study summarized findings from the initial evaluation of a PCG technique, based on a genetic algorithm, in terms of user experience. The proposed technique was incorporated in a novel serious game for OSA, and aimed to enhance the serious game's educational value, by generating educational content tailored to user needs, and empower user engagement. Preliminary results demonstrated the technique's potential, while future research will focus on measuring its effect on the serious game's educational value.

REFERENCES

[1] A. B. Moghadam and M. K. Rafsanjani, "A genetic approach in procedural content generation for platformer games level creation," 2017 2nd Conference on Swarm Intelligence and Evolutionary Computation (CSIEC), 2017.

[2] Sastry K., Goldberg D.E., Kendall G. (2014) Genetic Algorithms. In: Burke E., Kendall G. (eds) Search Methodologies. Springer, Boston, MA

[3] K. Mitsis, K. Zarkogianni, K. Dalakleidi, G. Mourkousis and K. S. Nikita, "Evaluation of a serious game promoting nutrition and food literacy: experiment design and preliminary results," 2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE), 2019.

[4] K. Mitsis, K. Zarkogianni, N. Bountouni, M. Athanasiou and K. S. Nikita, "An ontology-based serious game design for the development of nutrition and food literacy skills," *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (EMBC), 2019.

[5] R. Dey and J. Konert, "Content Generation for Serious Games," *Entertainment Computing and Serious Games Lecture Notes in Computer Science*, pp. 174–188, 2016.

[6] E. Cortés-Reyes, K. Parrado-Bermúdez, and F. Escobar-Córdoba, "New perspectives in the treatment of obstructive sleep apnea–hypopnea syndrome," *Colomb. J. Anesthesiol.*, vol. 45, no. 1, pp. 62–71, 2017.

[7] IJsselsteijn, W. A., de Kort, Y. A. W., & Poels, K. (2013). The Game Experience Questionnaire. Eindhoven: Technische Universiteit Eindhoven.