International Journal of Human–Computer Interaction

ISSN: 1044-7318 (Print) 1532-7590 (Online) Journal homepage: http://www.tandfonline.com/loi/hihc20

Approaches to Measuring the Difficulty of Games in Dynamic Difficulty Adjustment Systems

Dagmara Dziedzic & Wojciech Włodarczyk

To cite this article: Dagmara Dziedzic & Wojciech Włodarczyk (2018): Approaches to Measuring the Difficulty of Games in Dynamic Difficulty Adjustment Systems, International Journal of Human–Computer Interaction, DOI: 10.1080/10447318.2018.1461764

To link to this article: https://doi.org/10.1080/10447318.2018.1461764

Published online: 20 Apr 2018.

INTERNATIONAL JOURNAL OF HUMAN–COMPUTER INTERACTION
https://doi.org/10.1080/10447318.2018.1461764

Approaches to Measuring the Difficulty of Games in Dynamic Difficulty Adjustment Systems

Dagmara Dziedzic[a] and Wojciech Włodarczyk[b]

[a]Department of Logic and Cognitive Science, Institute of Psychology, Adam Mickiewicz University, Poznań, Poland; [b]Faculty of Mathematics and Computer Science, Information Systems Laboratory, Adam Mickiewicz University, Poznań, Poland

ABSTRACT

In this article, three approaches are proposed for measuring difficulty that can be useful in developing Dynamic Difficulty Adjustment (DDA) systems in different game genres. Our analysis of the existing DDA systems shows that there are three ways to measure the difficulty of the game: using the formal model of gameplay, using the features of the game, and direct examination of the player. These approaches are described in this article and supplemented by appropriate examples of DDA implementations. In addition, the article describes the distinction between task complexity and task difficulty in DDA systems. Separating task complexity (especially the structural one) is suggested, which is an objective characteristic of the task, and task difficulty, which is related to the interaction between the task and the task performer.

1. Introduction

For years, the difficulty was an integrated element of the game design process (Schweizer, 2006). First attempts to manipulate the difficulty of the game emerged with the launch of first gaming platforms and the coin-op arcade games in the 1970s. It was the time when games began to give the player an opportunity to choose the level of difficulty from one of the predefined options. As mentioned by Schweizer (2006), the first popular video game in which a player had a choice between two options (beginner's race and advanced player's race) was SpeedRace – an arcade game produced in 1974 by the Taito Corporation (1974). Soon after, game developers went one step further and created a first video game in which the difficulty was adjusted to player's doings. This game was the Heart of Africa (1985), in which the key locations were moved if the player had a problem with finding them (Brathwaite & Schreiber, 2008).

Today, the ability to adjust the game's difficulty to the needs and experience of a player is associated with the ability to choose one of several available options (e.g., easy, medium, hard). The choice between these options or their counterparts began to be commonly used in games in the 1990s (e.g., in Sid Meier's Civilization, 1991; Doom, 1993; Jazz Jackrabbit 2, 1998) and is still used today (we can find it in the latest productions, e.g., in Call of Duty, 2003; The Witcher, 2007; Dishonored, 2012; Sid Meier's Civilization VI, 2016). Regardless of which difficulty the player chooses, the difficulty of the game will usually be further increased with the progress – the further the player proceeds, the more difficult the game becomes.

In the examples described earlier, the player decides which difficulty of the game he/she wants to choose, but his/her choice is limited just to few options that were previously prepared by game designers. In addition, the difficulty of the game is always incremented at the fixed point in a game (e.g., when a player reaches a certain level). This approach is an example of the game-centered game design, in which human experts (i.e., designers) independently determine the parameters of the game (see Hristova, n.d.). In this type of game design, difficulty curve is static and is not influenced by the behavior of the player during the game. This approach treats every player equally and ignores differences in their abilities and preferences. This way, in some periods of the gameplay, difficulty may not be matched to the skills of the players (Yun, Shastri, Pavlidis, & Deng, 2009).

Recently, game designers tend to choose an alternative approach – the player-centered game design (or the user-centered game design), in which the games are designed, so that they can meet the expectations of players. One of the essential elements of player-centered game design is to provide an appropriate level of challenge, a properly matched learning curve and an enhancement of the gameplay experience for each player (Charles & Black, 2004). In this approach, developers design the so-called Dynamic Difficulty Adjustment (DDA) systems to ensure proper difficulty for each individual player.[1] Their purpose is to modify some elements of the gameplay, so that the general difficulty will be changed in the response to a player's performance, or his/her affective states. Measurement of difficulty in the DDA is a

CONTACT Dagmara Dziedzic dagmara.dziedzic@amu.edu.pl

[1]What is more adjusting the difficulty should be done without interrupting the game, so the player does not feel cheated by the game. Effective difficulty adjustment should take place without degrading player's experience and influencing his/her sense of control and accomplishment. For more about player's enjoyment and satisfaction, see Hunicke (2005) and Andrade, Ramalho, Gomes, and Corruble (2006).

case of measuring the difficulty in which the measurement directly affects the adjustment. The goal is not to measure the difficulty by itself, but rather focus on the issue of how it allows the system to adjust the difficulty of the game to the player.

This approach is very promising, because game design, where the difficulty is adjusted to the player's skills can bring many benefits. The problem of difficulty in the game is important for both practical and theoretical reasons. From a practical point of view, in this model, game developers are able to increase their profits and reach more customers. However, the design of DDA systems applies not only to commercial productions – this approach can be also easily implemented in games with a purpose, serious games, educational games, and rehabilitation games. Games with challenges adjusting to the skills of players can reach a larger audience, be more interesting for the players and educate them more effectively. For example, Hamari et al. (2016) report that educational games in which challenges increases with the growing skills and knowledge of students can efficiently engage them in learning activities.

From a theoretical point of view, the researchers of the difficulty in games (intentionally or not) are trying to answer questions that go beyond the realm of games, such as What is the difficulty? Can the difficulty be defined? How a person feels and perceives the difficulty? How the difficulty can affect the motivation and satisfaction?

Currently, there is no consistent definition of the difficulty that would be used by all designers of DDA systems. There is also no inclusive approach to the problem of difficulty and the possibility of adjusting it. In addition, in existing literature, concepts of difficulty and complexity are not distinguished and often used interchangeably.

This article is an attempt to present existing ways of measuring the difficulty of games for purpose of DDA systems. Within the article, we distinguish three types of difficulty measurement techniques. The DDA designer creating a new system can choose the technique which suits the characteristics of game he/she is working on. Moreover, the article attempts to address the issue of distinguishing the complexity and difficulty of games.

The article describes in Section 2 – purpose of DDA systems, in Section 3 – difficulty measurement techniques in DDA systems, and in Section 4 – the difference between the concepts of game's complexity and game's difficulty.

2. Purpose of DDA systems

As mentioned in Section 1, the implementation of DDA systems is the realization of player-centered game design. Its main goal is to adjust the gameplay to the skills of the individual player. To achieve this, the designers of DDA often rely on the flow theory, which suggests how well-designed automatic difficulty adjustment should work (regardless of the genre of games). Flow concept was introduced by Csikszentmihalyi (1975) who describes it as a "holistic sensation that people feel when they act with total involvement" (p. 36). Flow is something felt by people who are totally involved in a task they are performing, and this task is done for the pure enjoyment.

The state of flow can be successfully achieved by the players. However, to achieve it, appropriate conditions have to be met. According to Nacke (2012) in the case of games, the player must first perform a challenging activity that requires him/her to improve his/her skills. Second, the activity performed by him/her should have clear and easy to achieve goals (levels, missions, high scores, etc.) with an immediate feedback on his/her progression. Third, the final result of the activity the player is performing should be uncertain, but at the same time, he/she has to have a direct impact on its outcome.

One of the most important conditions for experiencing the flow is to provide the challenges that the player perceives as adequate to his/her abilities. Missura (2015) points out that an integral feature of any challenge (and the learning process required to master it) is its difficulty. The difficulty is a subjective factor, which changes (decreases) with time and effort that the player spends on learning the skill required to accomplish a certain task (Missura, 2015).

The flow theory suggests that the difficulty of the game should offer challenges that difficulty corresponds to the current skills of the player. If the player feels that the challenges in the game become too difficult for him/her too quickly – he/she experiences frustration (Falstein, 2005) or even an anxiety (Csikszentmihalyi, 1975). If the difficulty of the challenges don't change – and the skill of the player rises – the player experiences a boredom (Csikszentmihalyi, 1975). The path between these two extremes is referred to as the flow zone.

One of the solutions, which may be used to ensure that the player stays in the flow zone, regardless of his/her experience and skills is using DDA systems. Depending on the genre of the game, difficulty is handled in a different way, using different parameters of the game.

3. Difficulty measurement techniques in DDA systems

While creating a DDA system, a designer has to define a method for measuring the difficulty and conditions describing how it will be manipulated. After an analysis of existing DDA systems, one may realize that depending on various types of games, the difficulty can be manipulated using different elements of the game, also the intervention of a DDA system (which increases or reduces the overall difficulty of the game) may occur in different points during the game.[2] Therefore, these properties of DDA systems are usually imposed by the structure of the game itself (for more general discussion, see Dziedzic, 2016). Even though there is no single, generally accepted definition of the difficulty, designers are able to

[2]The analysis was carried out with four search engines for scientific articles: Google Scholar (https://scholar.google.com), EbscoHost (https://search.ebscohost.com), ACM Digital Library (http://dl.acm.org/), ScienceDirect (http://www.sciencedirect.com/), using a combination of keywords such as *dynamic, difficulty, adjustment, DDA, games*.

create a DDA system using simplified, informal metrics based on the characteristics of the game they are working on.

Examples of the DDA systems created in this way are listed in Tables 1–3. These are examples of measuring and/or adjusting the difficulty of games that allowed us to identify the three general approaches used to measure the difficulty. Each example includes the definition of difficulty proposed by the game authors, difficulty measurement (which describes in detail how the aforementioned definition was implemented), and the way in which difficulty was adjusted (if it was included in the research). Examples of DDA systems outlined in Tables 1–3 not only allowed us to distinguish three approaches to measure the difficulty but also showed that some DDA systems are based primarily on game complexity (see Section 4).

The difficulty measurement techniques described later define three approaches that can be used while developing a DDA system. The first two are directly related to gameplay, and the third is player-related. The first technique involves using a formal model, which can be defined as a detailed description of rules of the game (one of the consequences of having a formal model of the game is the ability to construct its game tree[3]). The second technique is used when the formal model of the game is unknown (or highly complicated). In such case, DDA systems use the features of the game to create an approximate measurement for the real difficulty. We distinguish two types of features: parameters that describe the selected elements of the game (e.g., number of opponent's pawns, number of enemies, number of lost lives, number of collected coins) and player's performance indicators that do not have direct representation in the game (e.g., the number of enemies killed per minute, average rate of gain of coins). In what follows, the terms "parameters" and "indicators" will be understood as described earlier. Apart from using gameplay-related techniques, creators of DDA systems can independently analyze the player himself/herself. In the third technique of measuring difficulty, DDA systems are evaluating the feeling of the difficulty perceived by the player. The measurement of difficulty involves querying the player about his/her perceived difficulty, or measuring emotions he/she felt during the game. This measurement can take place independently of the gameplay, so it can be mixed with the first two difficult measurement methods.

To sum up, one can define three distinct approaches to measure the difficulty: (1) the use of formal model of gameplay, (2) the use of selected features of the game, and (3) the direct examination of the player.

3.1. Using the formal model of gameplay

The use of this technique occurs in board games such as chess or connect four (see Table 1). These are games in which the player chooses his/her next move from a finite set of moves possible in his/her current state of the game and the gameplay is divided into units of time (turns).

Within this group, DDA systems are based on the analysis of the structure of the game tree. The number of possible moves at every moment is relatively small. Performance of the player (whether the player is going to win or lose) is evaluated based on the analysis of his/her movements – and more specifically possible consequences of his/her moves retrieved from the game tree. DDA systems working as described previously are most often based on search algorithms such as MiniMax or Alpha-beta Pruning (see Missura & G¨Artner, 2008). Algorithms of this type operate on the basis of the evaluation function of each player's movement. Typically, this function analyzes the consequences of player's move as a change in the game's parameters before and after player's action (e.g., change in the number of enemy pawns). For example, in chess, capturing opponent's pawn can be considered as a good move which leads to winning, and loss of pawn as a bad move that increases the chances that the player will lose. Using this knowledge, the system changes the current difficulty by selecting opponent's moves which difficulty matches those done by the player (the same measure is used to evaluate both player's and opponent's moves). One of the most important advantages of this approach is that even if game trees of different games can vary a lot, they can all be analyzed in the same manner. Regardless of whether a DDA system is designed for games like Tic-Tac-Toe or chess, it can use evaluation function based on the structure of the game tree such as the shortest path to success, or the average path length to failure. Intervention of the DDA system, created for this type of games, can be done in units of time naturally imposed by the game – after each round. Exemplary systems of the described type are presented in Table 1. In chess by Guid and Bratko (2013) and "Mastermind" by Gierasimczuk, Der Maas, and Raijmakers (2013), authors have proposed only a measure of difficulty, without including a DDA system in their research. However, both solutions show a good example of difficult measurment techniques for games with a formal model. Due to the fact that adjusting difficulty in various games with a game tree can be fairly similar, creating a measure of difficulty in this type of games is more important than creating an adjustment algorithm that could be transferred from one game to another.

Main assumption that one has to accept if he/she chooses to use this approach is that the structure of the game can reflect the performance of the player and influence the way he/she perceives the difficulty (features such as the length of the shortest path to success determines the difficulty of the game).

3.2. The use of selected features of the game

The second technique occurs in games, where the formal model is not known or it is too complicated to use it – due to a large

[3]Game tree refers to the concept of a tree known from the graph theory. In this tree, every possible state of the game is represented as a node in a graph. And if a player can move from a given state to another in exactly one move, then these states are connected with an edge. The final game states are represented as leaves of the tree. Each play of the game is represented as a path leading from the root to one of the leaves (for more details, see Gobe, De Voogt, & Retschitzki (2004)).

Table 1. The use of a formal model of the game in exemplary implementations.

The use of formal model of the game			
Game title and source	Definition of difficulty used in the research	Difficulty measurement	Adjustment
Connect Four by Missura and Gärtner (2008)	"The problem of an **automatic difficulty scaling** can be viewed then in a context of an interaction between a player and one (or more) in-game agent… It is natural to assume that at any given time the agent has a set of actions (strategies) available to it. The question of how to adjust the game difficulty automatically can then be formulated as what action (strategy) should an in-game agent choose as next." (Missura & Gärtner, 2008, p. 1)	Authors proposed a search-based algorithm called AdaptiveMiniMax (AMM). (1) Using a full game tree, MiniMax algorithm is creating a set of all possible moves together with their evaluation. (2) Score of each move depends on length of its path to winning or losing (it gets more points if player can win faster).	(1) Algorithm evaluates moves done by the player using this metric and creates a ranking of them. (2) AMM adjusts the difficulty of the game by selecting moves that matches an average score from the ranking.
Chess by Guid and Bratko (2013)	"We focus our investigation on problems in which the **difficulty** arises from the combinatorial **complexity of problems**. We propose a **measure of difficulty** that is based on modeling the problem solving effort as search among alternatives and the relations among alternative solutions." (Guid & Bratko, 2013, p. 860)	Authors proposed an search-based algorithm for measuring difficulty of chess tactical problem. Proposed metric was created using following principles: (1) If its solution requires many steps (solution lies deep in the game tree), the problem is considered difficult. (2) The problem is more difficult if the possible solutions are located far from each other in the game tree.	N/A
Mastermind (simplified) by Gierasimczuk et al. (2013)	"We associate the **difficulty** of Deductive Mastermind **game-items with the size** of the corresponding logical trees obtained by the tableaux method." (Gierasimczuk et al., 2013, p. 297) "Normatively speaking, the full tree generated by the tableau method for the set of formulas corresponding to a DMM-item, represents an adequate reasoning scheme. Therefore the **size of the tree** (e.g. the number of nodes) can be thought of as an **abstract complexity measure**." (Gierasimczuk et al., 2013, p. 307)	The authors developed a model for calculating the difficulty of the game, which is based on the logical reasoning, using analytic tableaux. (1) The algorithm contains a list of logic formulas that are used for reasoning. Formulas describe basic conclusions that can be made from the game's feedback. (2) The algorithm creates a tree (similar to game tree) with the states, which can be reached using the logical reasoning. (3) The difficulty of the game is computed as a size of the tree.	N/A

number of moves the player can make it is not possible to create a full game tree. Games of this type often include arcade elements, random events or time-related tasks. Examples of games using this technique belong to such genres as first-person shooters, tower defense, platformers, etc. (see Table 2).

Because it is not possible to determine all the possible moves of the player, in this method, the model of the difficulty is reduced only to a closed group of features. In this model, DDA can use both parameters that describe the selected elements of the game (e.g., number of enemies) and indicators of player's performance (e.g., number of enemies killed per minute). Because these features are very dependent on the game's genre and the gameplay elements, each game has its own unique set of features, manually selected by the DDA designer.

Also, the intervention time of the system depends strictly on the characteristics of the game for which DDA is designed. If it is a game with distinct stages of gameplay (e.g., waves of enemies in tower defenses), the intervention occurs after each stage. Otherwise, if the game does not contain such stages, the intervention must occur in some fixed intervals (e.g., every 30 s).

DDA systems using this technique are based on various algorithms. Recently, one of the most popular approaches is to use machine learning algorithms (such as clustering algorithms – see Lora, Sánchez-Ruiz, González-Calero, & Gómez-Martín, 2016) to determine the relation between players performance and the difficulty of the game.[4] Exemplary systems, which use this technique of measuring the difficulty, are described in Table 2.

Main assumption that one has to accept if he/she wants to use this approach is that selected features are adequate and sufficient to measure and manipulate the difficulty felt by the player.

3.3. Direct examination of the player

In this method, designers of DDA systems analyze the difficulty perceived by the player instead of elements of the gameplay. Therefore, this approach can be applied to any game,

[4]Machine learning algorithms are designed so that for a set of input features algorithm assigns a result value (numerical for regression algorithms or class for classification algorithms). Machine learning works well in this approach because the difficulty model is also based on a set of features. A DDA creator using machine learning must prepare a training data with specific values of the features and a specific difficulty level assigned to them. Then the learning algorithm will try to find the relationship between the features and the level of difficulty. One of the most popular uses of machine learning in games is creating an artificial intelligence (AI) that plays with the player (e.g., AlphaGo created by DeepMind, https://deepmind.com/). While AI algorithms mentioned earlier is designed to simply win with the player, the purpose of the DDA is to adapt to his/her performance. For more details about machine learning, see Smola and Vishwanathan (2008).

Table 2. The use of selected parameters of the game in exemplary implementations.

	The use of selected parameters of the game		
Game title and source	Definition of difficulty used in the research	Difficulty measurement	Adjustment
Hamlet (first-person shooter) by Hunicke and Chapman (2004)	"Using techniques drawn from Inventory Theory and Operations Research, Hamlet **analyzes and adjust the supply and demand of game inventory** in order to control overall **game difficulty**." (Hunicke & Chapman, 2004, p. 96) "We propose a probabilistic technique that dynamically evaluates the **difficulty** of given obstacles based on user performance, as the game is running." (Hunicke & Chapman, 2004, p. 97)	Authors proposed a probabilistic model for measuring difficulty in the game using player's inventory. The algorithm uses the following rules to determine the current performance of a player. (1) Algorithm observe player's health and distribution of damages taken by player over time to predict potential health shortfalls. (2) Algorithm analyzes changes in player's inventory. How much he gains (e.g., ammunition found in creates) and losses during the game (e.g., used ammunition).	Authors created a number of adjustment policies with adjustment actions and cost estimations. (1) System support two types of actions: reactive (adjusting elements that are "on stage", e.g., entities that have notices the player) and proactive (adjusting elements that are "off state" e.g., spawning health). (2) System is determining the cost of a given action to make sure that player do not notice system's intervention.
Tower Defense by Sutoyo, Winata, Oliviani, and Supriyadi (2015)	"… we determine the **difficulties** based on players' **lives, enemies' health, and passive skills** … that are chosen by the player. With three of these factors, players will have varies experience of playing tower defence because different combination will give different results to the system and difficulties of the games will be different for each gameplay." (Sutoyo et al., 2015, p. 435)	Author selected three parameters that are monitored to determine difficulty of the game through player's performance. (1) Algorithm analyzes player's lives, enemies' health and passive skills (skill points). (2) Passive skills are three groups: offensive, defensive and support.	(1) DDA modifies the set of parameters (such as the number of gold or spawned opponents) and their multipliers (e.g., gold multiplier). (2) The DDA system adjusts the difficulty by modifying the aforementioned parameters after each round (e.g., decreases the number of opponents).
Tetris by Lora et al. (2016)	"When a player starts a new game we look at his first movements to find the most similar cluster. Then the system provides dynamic help to the player choosing "good" Tetris pieces from time to time. In our experiments, using DDA users obtain higher scores and report improvements in terms of their game experience." (Lora et al., 2016, p. 335) "The **difficulty** in Tetris depends mainly on two parameters: The **type of pieces** and the **falling speed**." (Lora et al., 2016, p. 337)	The authors created a statistical clustering model that is used to estimates player's skills. (1) Performance of the player is calculated based on his/her last 10 tactical decisions (placing the block on the map). (2) Each decision is described by a set of parameters (e.g., height of the blocks, scored points, maximum number of points to win). (3) Using collected data players were divided into three groups: newbie, average, expert.	(1) Algorithm is assigning player to a one of aforementioned groups (algorithm is updated after each move). (2) Depending on player's group, algorithm helps him/her frequently (for newbie) or infrequently (for experts). Algorithm helps the player by giving him block needed in current state of the game.

regardless of whether it has a formal model or not (see Table 3).

The difficulty measure of the game is based on the player's feelings. The specific feelings of a player are correlated with the various difficulty settings of the game. When a measuring tool recognizes that the game is too hard for the player (or too easy), the DDA system adjusts the difficulty.

Just like in the previous case, this technique does not impose the moment of intervening – as it depends directly on the characteristics of the game.

Examples of methods used for measuring the difficulty are questionnaires measuring affective states and physiological tests. Example systems, which use this technique of measuring the difficulty, are described in Table 3. All the examples cited in this table apply only to research carried out with the use of games, and not commercial games. Even though it is hard to apply this technique in commercial projects, it can provide the most reliable information about the perceived difficulty directly from the player. Furthermore, it is possible to use this technique to assign difficulty to appropriate gameplay parameters (e.g., the number of enemies) based on difficulty perceived by the player. At the end of the study, collected data can be used to create a statistical adjustment model that does not require polling (or testing) the player, and may already be used in commercial games.

Main assumption that one has to accept if he/she chooses to use this approach is that the player is able to determine the difficulty of the task and compare it with the difficulty of other, and that selected affective or physiological states are directly related to perceiving the difficulty of the game.

3.4. Conclusions

Due to the fact that it is difficult to find a game where it is possible to implement all three techniques of measuring difficulty in DDA systems and to test which gives better results, it is impossible to decide which of the above could be consider as a best one. Choosing the right technique highly depends on the structure of the gameplay and the purpose of the DDA system. Table 4 lists the practical suggestions that can help to select the right difficulty measurement technique for a new DDA system. It provides a summary of approaches described in this section, their advantages, disadvantages, and the suggested types of games for each of the techniques.

Table 3. Direct examination of the player in exemplary implementations.

Direct examination of the player			
Game title and source	Definition of difficulty used in the research	Difficulty measurement	Adjustment
Third-person shooting game by Yun et al. (2009)	"The measurements are based on the assumption that the players' performance during the game-playing session alters blood flow in the supraorbital region, which is an indirect measurement of increased mental activities." (Yun et al., 2009, p. 2195) "… we propose a system that adaptively and automatically adjusts the **game difficulty level** based on measurements of the **players' facial physiology**." (Yun et al., 2009, p. 2196)	Authors proposed a method for measuring difficulty by analyzing the player's changes in the skin temperature in the supraorbital region using the StressCam. (1) The change in temperature in this region is associated with the blood flow, which is an indirect measure of the mental activities. (2) When the game is hard and the player suffers stress or frustration – the temperature in the supraorbital region should grow.	(1) Authors created three game modes: easy, moderate, difficulty. Each mode is associated with different strength of enemies robots. (2) Game difficulty was altered during the game using the metric described earlier.
Pong by Liu, Agrawal, Sarkar, and Chen (2009)	"In this DDA mechanism, a **player's physiological signals** [e.g. skin conductance, skin temperature, heart sound] were analyzed to infer his or her probable **anxiety level**, which was chosen as the target affective state, and the **game difficulty level** was automatically adjusted in real time as a function of the player's affective state." (C. Liu et al., 2009, p. 506)	The authors created an affect based DDA, which recognize the affective state (anxiety) using psychophysiological analysis. (1) To develop affective model, the authors built mappings from physiological features to the intensity of anxiety. (2) Low anxiety means that the player perceives the game as easy and high anxiety level means that the game is to hard.	(1) Real-time player's anxiety level recognition was used to adjust the game's difficulty. (2) To manipulate the difficulty level of the game, authors selected a number of parameters (e.g., ball speed and size, paddle speed and size, sluggish or over-responsive keyboard, random keyboard) and adjusted them during the game.
Runner Game by Medeiros and Medeiros (2014)	"… we propose a way to automatically rank all combinations of features according to how fun that level is, and by doing so, choose the combination of features with the best flow for our runner game. After each 30 second playthrough, the **player is asked** to rate **how fun** and **how difficult** was that level on a scale from 1 to 5." (Medeiros & Medeiros, 2014, p. 797)	Authors proposed a system in which the difficulty measurement is done by the player himself after each round of the game. (1) Player plays the game for 30 seconds or until avatar's death. (2) After the game, he/she completes a survey, in which (on a scale from 1 to 5) he/she describes how fun and difficult the game was.	(1) The authors defined several features of the game (e.g., number of spikes per wave, distance between spikes) that are used to control the difficulty of the game. (2) The system selects a set of features for each game. Every time the player rated a particular configuration to be good/bad – the algorithm changes the probability of current configuration.

4. Complexity versus difficulty

Analyzing at how the difficulty is defined in games with a full game model (Table 1), one may notice that it is modeled using only game's parameters and actions related to them (e.g., capturing opponent's pawn as a decrease in the number of opponent's pawns on the map). In particular, the difficulty of these games is measured mainly on the basis of their game tree – core of DDA systems do not take into account the indicators of player's performance (how quickly he/she puts the pawns, how logically he/she reasons), nor his/her emotions and physiological states.

If so, the difficulty measurement for games with a full game model corresponds to struturalist viewpoint of task complexity proposed by Liu and Li (2012). These authors provided a comprehensive review and conceptualization of task complexity and tried to determine the differences between task complexity and task difficulty.

Further in this section we rely primarily on their work because it is a holistic overview based on the analysis of several dozen research studies in which the task complexity appeared. Their approach has not been previously applied to games, but in our opinion the complexity of the game can be considered as a special case of much more general task complexity – and more specifically task complexity in structuralist viewpoint.

According to Liu and Li (2012), in structuralist approach, task complexity is defined as the structure of the tasks and it may be computed as a function of the number of elements that form the task and the relations between these elements (e.g., number of elements, number of targets, number of solutions, number of tracks, number of alternatives).

In the case of the complexity and the difficulty, there are no definitions which would clearly explain these concepts, as a consequence, they are often confused. In accordance with Liu and Li (2012), we claim that in the case of games:

> To avoid the misuse and misunderstanding of these related terms, we claim that task complexity can be distinguished from task difficulty in that task complexity involves the objective characteristics of a task, whereas task difficulty involves the interaction among task, task performer, and context characteristics. Subjective task complexity is a perceived complexity by the task performer. In general, task difficulty refers to the extent to which task performers feel difficulty in performing a task. (p. 559)

The term difficulty includes not only the complexity of the task but also the performer who performs it and the context of the task. If so, then the complexity of the game is only a subset of the difficulty, and to create a more complete game difficulty model for games with a full model, DDA designers should consider incorporating the aspect of the task performer (player) into their systems.

DDA systems created for games with a full model operate on the basis of objective gameplay characteristics and they depend only on actions of a player (and not his/her skills or emotions). Liu and Li (2012) suggested that objective task complexity is independent of task performer and it is closely related to task characteristics. By looking at the definitions in Table 1 (compared to Tables 2 and 3) only in the first one, the authors define the difficulty based solely on the task complexity. The definitions imposed by the authors are determined only by the structure of the problem that player has to solve (as in the structuralist task complexity by Liu & Li, 2012). In this approach, the role of the performer is limited or does not appear at all. In addition, taking the performer into account can be a challenge, because in logic games it is difficult to capture concepts such as logical reasoning (which affects the next move of the player) in a measurable way. In the game with a full game model, DDA system analyzes only the final effect of the player's actions (e.g., the fact of capturing the pawn on the board in chess is more important than the number of pawns captured per minute).

In the games where the formal model of the game is unknown (or highly complicated), complexity is also present but it cannot be described in a simple way (e.g., due to the limited computing power of current computers). Hence, it is necessary to simplify the game to a closed set of measurable features. In games of this type, DDA system analyzes the sequence of player's actions spread over time (e.g., the time needed for the player to aim at the enemy). The task performer's aspect is measured indirectly with the player's performance indicators, which directly reflect the player's skill (such as hand–eye coordination). The task difficulty seems to be more appropriate for describing a game in which the task complexity is not determined by the formal model and the aspect of the task performer is taken into account.

In the case where the task performer's aspect is measured directly with the player's examination, the presence of the performer's task is evident – regardless of whether the game contains a full model or not, this technique is based on task difficulty. A summary of what constitutes the base of the difficulty measure for each difficulty measurement technique is given in Table 4.

The conclusion that complexity is only a subset of difficulty can also be reached by considering situations in which the same complexity of tasks may have different difficulties for different performers. For example, for most children learning the multiplication tables, multiplying "six times eight" is far more difficult than the multiplication of other numbers. Multiplying "six times ten" and "six times eight" have the same structural complexity, but in the latter case, the children respond poorly in 62.5% of cases (Magazine Monitor, 2014).

Another case, in which the same structural complexity of the task gives a different feeling of the difficulty is chess. In studies conducted by Hristova, Guid, and Bratko (2014), chess experts evaluated the difficulty of chess tactical problems. For this purpose, authors selected 12 positions – chess tactical problems (selected randomly from Chess Tempo[5]), which were divided into three classes of difficulty (easy, medium, hard). This division was made on the basis of the Chess Tempo rating – ranking, which is based on the success and failures of the players (if the player solves the problem correctly, its rating goes down, and if incorrectly –

Table 4. Comparison of difficulty measure techniques in DDA systems.

Techique name	The use of formal model of gameplay	The use of selected features of the game	The direct examination of the player
Difficulty measure	Based on the game tree	Based on the game parameters and player's performance indicators	Based on the difficulty perceived by the player
Base of difficulty measure	Task complexity (task performer's aspect is limited or does not appear)	Task difficulty (task performer's aspect is measured indirectly with player's performance indicators)	Task difficulty (task performer's aspect is measured directly with the player's examination)
Difficulty adjustment method	Changes the difficulty by selecting moves of the opponent using a measure based on the game tree	Changes the parameters of the game affecting the difficulty	Changes the parameters of the game affecting the difficulty
Advantages	DDA systems created for different games shares similar algorithms (e.g., MiniMax)	The possibility to create DDA systems for complex games by narrowing them to a set of measurable features	The player directly reports his or her feelings of difficulty rather than indirectly, through the values of the selected metrics
Disadvantages	Requires a full game model	Large range of parameters and indicators that are the basis of DDA systems for different games (there can be multiple, different DDA systems created for the same game)	Creators of DDA system need an access to the player (large interference in the player's environment) and to have the tools to measure the player's, e.g., declared difficulty, affective states or physiological states
Typical moment of DDA system's intervention	After each game's turn	Depends on the structure of the game (the moment is chosen by the designer or after each turn in the game)	Depends on the moment when the player is questioned and the game structure (the moment is chosen by the designer or after each turn in the game)
Suggested game types	Board games, logic games	Arcade games, shooter games, racing games, platformers	Arcade games, shooter games, racing games, platformers

[5]The Chess Tempo is an online chess platform available at www.chesstempo.com.

its rating goes up). Chess experts were asked to solve these 12 tactical problems, then rate them from 1 to 12 (they did not know that the problems had previously been divided into three classes of difficulty using the Chess Tempo rating). The study included a retrospective reports, which would help to understand the approach that experts used while solving a certain position. The results showed a large disproportion between Chess Tempo rating, and the estimations of experts. Sometimes, the difficult chess problem was considered easy by the experts and easy problem as hard. The same arrangement of pieces on a chessboard (structural complexity) produced a different result in the assessment of the difficulty made by the experts.

The same complexity of gameplay can result in different difficulty for different players. Therefore, DDA systems that do not take into account aspects of the player will not be able to fully adapt to the specific player. The problem of dependency between complexity and difficulty in games is still open and there is currently no research showing how these two concepts affect each other.

5. Conclusions

For years, designers have created games that give gamers a choice of gameplay difficulty. In the most basic approach, players can choose between easy, medium, and hard mode of game. However, this may be insufficient because the players' skills and experience are so different that it is difficult to grasp them just with the use of a finite number of categories. In addition, the player learns during the game, so his/her skills and expectations of the difficulty may change over time. The solution to this problem is to design DDA systems that are based on the flow theory. According to it, the challenge must be adjusted to the player's skill, so he/she can constantly enjoy the game. DDA systems are designed to measure difficulty (based on their difficulty conception), and adapt it.

However, existing DDA systems are missing a coherent approach to measure the difficulty. After an analysis of existing implementations of DDA systems, we have proposed a difficulty measurement techniques, which can be applied in different game genres. The approaches presented in this article distinguish three types of difficulty measures: using the formal model of gameplay, using the features of the game, and direct examination of the player. The first two are separable and relate to gameplay itself, while the third one is related only to the player and can be used independently of the first two (together with them or completely separate).

The designer creates a DDA system for full model games, and it should definitely use it to adjust the difficulty. But taking into account conclusions from Section 4, he/she may consider extending the difficulty model and including the performer's performance (e.g., by incorporating the techniques of measuring difficulty from direct examination of the player). If the designer creates a DDA for a game without a full model, then he/she should use a closed group of game parameters along with indicators of player's performance to adjust the difficulty. The selection of features depends on the game genre and gameplay elements. As basic features, it is important to choose the ones that define the key elements of the gameplay: e.g., in the shooter games – the number of killed enemies, in tower defence – the number of lives, and in racing games – speed of the car and player's ranking position.

In addition, this article describes the distinction between task complexity and task difficulty in DDA systems. We suggest separating task complexity (especially the structural one), which is an objective characteristic of the task, and task difficulty, which is related to the interaction between the task and the task performer. In DDA systems, which use the full model of gameplay (e.g., chess), the adjustment mechanism is based on the complexity of the gameplay (usually, the structure of the game tree, which is an objective characteristic of gameplay). Since this issue (to our best knowledge) has not been noticed before, there is currently no research that would show the influence of task complexity on the perceived (subjective) difficulty or enjoyment felt by player.

References

Andrade, G., Ramalho, G., Gomes, A. S., & Corruble, V. (2006). Dynamic game balancing: an evaluation of user satisfaction. In J. E. Laird & J. Schaeffer (Eds.), *AIIDE* (pp. 3–8). Menlo Park, CA: The AAAI Press.

Arkane Studios. (2012). *Dishonored [PC game]*. Bethesda Softworks.

Brathwaite, B., & Schreiber, I. (2008). *Challenges for game designers* (1st ed.). Rockland, MA, USA: Charles River Media, Inc.

CD Projekt RED. (2007). *The Witcher [PC game]*. Atari, CD Projekt.

Charles, D., & Black, M. (2004). *Dynamic player modelling: A framework for player-centered digital games*. Proceedings of the international conference on computer games: Artificial intelligence, design and education (p. 29–35). Reading: Microsoft Campus.

Csikszentmihalyi, M. (1975). *Beyond boredom and anxiety*. San Francisco, CA: Jossey-Bass Publishers.

Dziedzic, D. (2016). Dynamic difficulty adjustment systems for various game genres. *Homo Ludens*, *9*(1), 35–51.

Epic MegaGames. (1998). *Jazz Jackrabbit 2 [PC game]*. Gathering of Developers, Logicware, Project Two Interactive.

Falstein, N. (2005). Understanding fun - the theory of natural funativity. *In* S. Rabin (Ed.), *Introduction to game development*. Hingham, MA: Charles River Media.

Firaxis Games. (2016). *Sid Meier's Civilization VI [PC game]*. 2K Games.

Gierasimczuk, N., der Maas, H. L. J. V., & Raijmakers, M. E. J. (2013). An analytic tableaux model for deductive mastermind empirically tested with a massively used online learning system. *Journal of Logic, Language and Information*, *22*(3), 297–314. doi:10.1007/s10849-013-9177-5

Gobe, F., de Voogt, A., & Retschitzki, J. (2004). *Moves in mind: The psychology of board games*. Hove, East Sussex: Psychology Press.

Guid, M., & Bratko, I. (2013). *Search-based estimation of problem difficulty for humans*. Artificial intelligence in education - 16th international conference, AIED (pp. 860–863). Memphis, TN. doi: 10.1007/978-3-642-39112-5_131

Hamari, J., Shernoff, D. J., Rowe, E., Coller, B., Asbell-Clarke, J., & Edwards, T. (2016). Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning. *Computers in Human Behaviour*, *54*, 170–179. doi:10.1016/j.chb.2015.07.045

Hristova, D. (n.d.). *Dynamic difficulty adjustment (DDA) in first person shooter (FPS) games*. Retrieved April 05, 2017 from https://www.academia.edu/4921794/Dynamic_difficulty_adjustment_DDA_in_First_person_shooter_FPS_games

Hristova, D., Guid, M., & Bratko, I. (2014). Assessing the difficulty of chess tactical problems. *International Journal on Advances in Intelligent Systems*, *7*(3), 728–738.

Hunicke, R. (2005). *The case for dynamic difficulty adjustment in games*. Proceedings of the 2005 ACM SIGCHI international conference on advances in computer entertainment technology (pp. 429–433). New York, NY: ACM.

Hunicke, R., & Chapman, V. (2004). AI for dynamic difficulty adjustment in games. *Challenges in Game Artificial Intelligence AAAI Workshop* (pp. 91–96). San Jose, CA.

id Software. (1993). *Doom [PC game]*. GT Interactive.

Infinity Ward. (2003). *Call of Duty [PC game]*. Activision, Aspyr.

Liu, C., Agrawal, P., Sarkar, N., & Chen, S. (2009). Dynamic difficulty adjustment in computer games through real-time anxiety-based affective feedback. *International Journal of Human-Computer Interaction*, *25*(6), 506–529. doi:10.1080/10447310902963944

Liu, P., & Li, Z. (2012). Task complexity: A review and conceptualization framework. *International Journal of Industrial Ergonomics*, *42*(6), 553–568. doi:10.1016/j.ergon.2012.09.001

Lora, D., Sánchez-Ruiz, A. A., González-Calero, P. A., & Gómez-Martín, M. A. (2016). *Dynamic difficulty adjustment in Tetris*. Proceedings of the twenty-ninth international florida artificial intelligence research society conference, FLAIRS (pp. 335–339). Key Largo, FL.

Magazine Monitor. (2014). *Who, what, why: Why does the sum 7 × 8 catch people out?* Retrieved April 05, 2017, from http://www.bbc.com/news/blogs-magazine-monitor-28143553

Medeiros, R. J. V. D., & Medeiros, T. F. V. D. (2014). *Procedural level balancing in runner games*. Proceedings of the 2014 Brazilian symposium on computer games and digital entertainment (pp. 109–114). Washington, DC: IEEE Computer Society.

Missura, O. (2015). *Dynamic difficulty adjustment* (Doctoral dissertation). University of Bonn. Retrieved April 05, 2017, from http://hss.ulb.uni-bonn.de/2015/4144/4144.pdf

Missura, O., & Gärtner, T. (2008). *Online adaptive agent for connect four*. Proceedings of the fourth international conference on games research and development cybergames (pp. 1–8).

MPS Labs. (1991). *Sid Meier's Civilization [PC game]*. MicroProse.

Nacke, L. E. (2012). *Flow in games: Proposing a flow experience model*. Proceedings of the workshop on conceptualising, operationalising and measuring the player experience in videogames at fun and games (p. 104–108). Toulouse, France: ACM.

Ozark Softscape. (1985). *Heart of Africa [Commodore 64 game]*. Electronic Arts.

Schweizer, B. (2006). Difficulty. *In* H. Lowood & R. Guins (Eds.), *Debugging game history: A critical lexicon* (pp. 109–117). Cambridge, MA: MIT Press.

Smola, A. J., & Vishwanathan, S. (2008). *Introduction to machine learning*. Cambridge, MA: Cambridge University Press. Retrieved from http://alex.smola.org/drafts/thebook.pdf,/bib/smola/smola2008ml/thebook.pdf

Sutoyo, R., Winata, D., Oliviani, K., & Supriyadi, D. M. (2015). Dynamic difficulty adjustment in tower defence. *Procedia Computer Science*, *59* (1), 435–444. doi:10.1016/j.procs.2015.07.563

The Taito Corporation. (1974). *Speed Race [Arcade game]*. Author.

Yun, C., Shastri, D., Pavlidis, I., & Deng, Z. (2009). *O' game, can you feel my frustration?: Improving user's gaming experience via StressCam*. Proceedings of the SIGCHI conference on human factors in computing systems (pp. 2195–2204). New York, NY: ACM. doi: 10.1145/1518701.1519036

About the Authors

Dagmara Dziedzic's research is associated with the possibility to make games used in the study, which were a useful tool for researchers and an attractive entertainment for the player. She is also working on the use of user experience techniques to improve gameplay enjoyment.

Wojciech Włodarczyk focuses on designing and creating crowdsourcing systems used for acquiring linguistic resources. His scientific interests are also connected with creating artificial intelligence agents in computer games, as well as applying machine learning mechanisms in order to improve them.