Utrecht University

Games and Media Technology
Master's Thesis

Modeling of fighting game players

Richard Konečný

richard.konecny@centrum.sk
student ID: 4279751

31. August 2016

First supervisor:

dr. dr. E.L. van den Broek
Department of Information and Computing Sciences, Utrecht University
E.L.vandenBroek@uu.nl

Second supervisor:

prof. dr. Georgios N. Yannakakis
Institute of Digital Games,
University of Malta
georgios.yannakakis@um.edu.mt

Second examiner:

prof. dr. R.C. Veltkamp
Department of Information and Computing Sciences, Utrecht University
R.C.Veltkamp@uu.nl

Associate supervisor:

dr. Antonios Liapis
Institute of Digital Games,
University of Malta
antonios.liapis@um.edu.mt

Acknowledgements

First of all, I would like to thank my first supervisor dr. dr. E.L. van den Broek for accepting the challenge and voluntarily supervise a project, in which the student is abroad for most of the project time frame. Also, for his professional and flexible guidance, a lot of valuable and insightful feedbacks, and support in general.

Next, I would like to express my gratitude to my supervisors in Malta, prof. dr. Georgios N. Yannakakis and dr. Antonios Liapis, for letting me do a part of my project in Malta and so gain very valuable international experience. Also, for their productive supervision, expert knowledge, and availability for meetings whenever it was needed.

Furthermore, I want to thank prof. dr. R.C. Veltkamp for finding some time to read and asses my thesis.

Also, I would like to thank to the Erasmus+ programme and the science international office at UU for making the study abroad possible.

The next part is meant for non English speakers, therefore I will continue in my native, Slovak language.

Chcel by som sa touto cestou poďakovať mojej rodine a hlavne rodičom za to že mi bolo umožnené študovať na univerzitách v zahraničí odbor, ktorému som sa chcel venovať už odmalička. Vďaka za ich finančnú, no hlavne duševnú a motivačnú podporu počas môjho pobytu v zahraničí, ale aj pred ním.

Abstract

A dynamic 7-dimensional skill-capturing game player model and accompanying novel real-time assessable metrics are introduced and validated by AI agents as well as human players. The model's dimensions quantify player's: 1) cognitive skills – distance estimate (DE), muscle memory (MM), reaction time (RT), space control (SC) and timing precision (TP); and 2) playing style – aggressiveness (AG) and decision making (DM). The games with AI agents indicate that methods proposed for metrics measurement are highly accurate – the anticipated outcome was achieved in 99.3 % of cases. Experiments with 16 human participants confirmed a significant correspondence between the methods' implementation and human perception of the metrics for AG, DM and SC. Moreover, the dimensions were used to estimate the challenge factor of our in-house developed fighting game. The estimated result indicate that DM, AG, RT and SC have the greatest effect on game's challenge; together constituting 70%. The final results of this study show that this model is very promising for applications requiring extensive behavioral and skill-capturing player characterization.

Keywords: Fighting Games, Player Modeling, Skill Capture, Skill Measure

1 Introduction

The field of adaptive video games (short: games) has grown significantly during past two decades and has turned out to be very useful in various application domains, including both educational games [1] and entertainment games [2]. The most common goals include: dynamic difficulty balancing [3]; experience optimization and predictability avoidance [4]; capturing skill [5]; emotionally adaptive content [6]; as well as training/serious games [7]. Many of these applications focus on creating an AI entity that adapts to player's gaming style [7], [8]. Besides many single-player games, there are also several multiplayer game approaches that use adaptive techniques – either adaptive AI agents, or dynamically generated content, such as quests in massively multiplayer online games [9].

Surprisingly, there has been done only little research in multiplayer games without computer-controlled characters; such as multiplayer one-on-one games (short: m1on1 games). This leaves an open question, whether it is possible to achieve results, similar to adapted AI agents in single-player games, with adaptive environment in multiplayer games. This is the main motivation for this research project. It is known that in order to provide the best enjoyment, the level of challenge should match player's skills. If a game is too hard, the player becomes frustrated and if a game is too easy, the player gets bored. Applying this knowledge to m1on1 games: if one player is much less skilled than the other, the game becomes hard for them, therefore frustrating. Conversely, after some time, the game may also become boring for the more skilled player (although, this may not always be the case when considering the social aspect). Therefore, relatively balanced skills of both players are usually necessary for a game to be neither frustrating nor boring for either of them. However, *"traditional game balancing fails to deal with the great diversity of players"* [3]. We believe that an explicit model, which describes diverse player characteristics individually and in more detailed manner, can solve this problem.

The main idea of this approach is following: Assume two players with a certain set of quantified skills. From these two sets, it is then possible to select the skills with mutually similar level. Then, by adapting the game environment, such that the weight of features requiring the selected skills would be increased, it might be possible to achieve more competitive gameplay. For example, if two players have similar reflexes, but one of them has significantly better aim, the gameplay requiring reflexes would be more competitive than the gameplay based on aim. Thus, increasing the impact of features that require reflex might increase competitiveness in the gameplay.

The first step towards this approach is to model player's skills, which could be compared and then used for environmental adaptations. The genre of fighting games has been chosen as a test bed for our study. This paper proposes a skill-based player model for fighting games. The model is evaluated both subjectively and objectively using an in-house developed fighting game. Furthermore, the challenge factor of the developed game is estimated, providing an rough idea of the distribution of weights of involved skills in the current, non-adapted gameplay. The developed game as well as the proposed model are described in more details in Section 2; experiments are explained in Section 3; the results are presented in Section 4; and the final discussion is provided in Section 5.

2 Related work

The structure of this section is following: First, a systematic review of player models is presented, followed a systematic review of adaptiveness in fighting games, then the dimensions for our player model are introduced, and the design of the developed game is presented at the end of this section.

2.1 Player modeling - systematic review

The following systematic review provides sources to answer the question: *"What are the current solutions to player modeling in dynamic adaptive games?"* . The search was performed in these three on-line libraries: *Google Scholar* [10], *The ACM Digital Library* [11] and *Web of Science* [12] on the 21.8.2016. The primary inclusion criteria were defined to improve relevance of the results. The sources should be:

- primarily concerned about player modeling in dynamic adaptive games
- published within last 10 years (since 2006)
- the most recently published material in case of duplicates

In order to meet the first primary criterion, specific search terms were defined. Note that with all nouns, both singular and plural variations were searched. To make sure the source is related to video games, the term "video game" or one of its synonyms (pc, computer, interactive game) had to be present. The next mandatory terms were: "adaptive" or one of its alternative forms (adaptable, adaptiveness, adapt, adaption, adaptivity) and "dynamic" or one of its alternative forms (dynamical, dynamically). The above mentioned terms must be present in the main body of the material, excluding the literature review and references sections. There are numerous applications of existing player models, but these are not the main focus of this review. This search is concerned about the materials that study player models as their primary interest. Thus, the term "player model" or one of its alternative forms (opponent m., cognitive m., user m.; model, modeling, modelling) had to be also present in the title.

Besides the primary criteria, there were also quality criteria. The studies should be:

- reproducible (several future proposals were excluded because of this criterion)
- published by an academic institution

As a result of this systematic review, 115 resources were positive, out of which 41 were identified as false positives. The next paragraphs provide a summary of the remaining 76. When multiple studies solve the same problem (for instance modeling players in the same game), or multiple materials are progressive parts of one study; then only the most relevant ones (with respect to their clarity, novelty and level of detail) are mentioned. A clear structured summary of player modeling and related studies published before 2011 is given by [13]. The summary presented here also includes more recent work. The next paragraphs give a general overview of player models; starting with explaining various application domains, presenting several potential applications in non-gameplay domains, discussing both general as well as game-specific approaches, giving examples of the most common utilizations of player models for various purposes. The last paragraph of this subsection discusses a few remarks of this systematic review.

Current application domains. In interactive storytelling (IS), case-based modeling addressed the problem of personalized player's experience in drama management [14]. A player model in PaSSAGE used originally dynamic learning to select prepared story fragments [15]. The PAST (Player-specific Automated StoryTelling) approach utilizes its extended version [15] with aim to solve the issue of increasing users' perception of IS's inner functioning.

In sport games, several approaches have been registered. [16] gives an overview of various player models for a 2D soccer simulation. It differentiates between approaches focused on individual and group performance. The paper also highlights that, in this case, a player model definition depends on the application and its purpose. *"Although, in theory opponent modeling can be very useful, in practice it is both difficult to accurately do and to effectively use to improve game play"* [16]. To address this issue, the approach presented in this paper provides a step towards practical use as it also presents concrete measurement methods used in a specific application.

Application domains also include serious games. Military uses cognitive models to improve simulation-based training [17]. Student modeling that takes into account identification, educational, affect and preferences information, and updates the knowledge level of students was used to motivate them and so improve their learning environment [1]. Several affective and engagement models have been analysed as potential player models for a platform to teach children pro-social skills [18]. Player models assessing user skills can be applied in rehabilitation processes. In a game called Nuts Catcher, the game parameters were modified using Q-learning algorithm, based on captured player skills [19]. Comparing to our proposal for skill capturing, this approach used one dimensional vector (the game score) to asses the user skill, whereas our approach proposes multi-dimensional vector.

Potential application domains. In some cases, using dynamically adaptive games, it has been shown that the use of player models has also a great potential to be beneficial outside of the gameplay domain. It could ameliorate group collaboration [20], enhance game performance using adaptive cinematographic experience [21], improve crowdsourcing by predicting effectiveness of workers' task completion [22], increase the quality cognitive realism of virtual characters [23] as well as sharpen argumentation strategies by adaptively choosing the best argument for a given situation [24].

General player models. Some studies attempted to define a general player model that would be applicable across all game genres. For instance, Yannakakis et al. [25] provide a high level taxonomy, which recognizes two different types of approaches: top-down and bottom-up. Various types of input and output are also characterized in more detail. The model we present in this material would be defined as top-down approach. ADAPTIMES [26] player model is another example of holistic player model. This one works with three dimensions: emotional (affective) state, performance and efficiency, and playing style.

Game-specific player models. Several game specific solutions have shown a successful use of methods that might also be applicable in broader field. The problem of user subjectivity being present while user testing was addressed by using persona-based player models with Monte Carlo Tree Search to construct personas for a puzzle game [27]. A 10 dimensional player model representing a combination of player strategies evolved using the NEAT neuroevolution method was used to predict arbitrary opponents in a simplified version of Texas Hold'em [28]. This demonstrated effectiveness of even low-dimensional models. The one proposed in this paper also belongs to this category.

Player model utilizations. This paragraph specifies most common utilizations of player models. The first one is clustering; i.e. player type clustering or player strategy clustering. In Tomb Raider, six gameplay features were extracted to cluster player into four player type groups: veterans, solvers, runners and pacifists [2]. In another example, five classes – Wary, Explorer, Winner, Impetuous and Neutral – were defined for the Maze-Ball game [29]. Strategy clustering for Real Tine Strategy (RTS) games was demonstrated using a behavioral hierarchical model, which records play style, building order and building units. Based on these data, seven different strategies were identified [30]. Player model can also be used to model and predict players' skills. An example of this use is a polynomial regression model for target-based games [31]. Our model also falls within this category.

Some models incorporate psychological studies, such as Behavlet model that uses a feature set of extracted dynamic gameplay behavioral traits of players [32]. A hierarchical clustering method by [33] attempts to predict players' emotional reactions. [34] gives an overview of behavioral models, identifying four different types of approaches. These are differentiated by the modeling target: modeling player a) actions, b) tactics, c) strategies or d) profiling.

Discussion. While performing the systematic review, a few terms have been found that could be used in the future to improve the results. These are "learner model", "player involvement" and more sentential combinations of currently used variations of player model, for instance: "modeling decision player".

2.2 Adaptiveness in fighting games - systematic review

Since the systematic review focused on sources with primary interest of player models did not reveal any approach regarding the fighting games specifically, one more search has been conducted. This time, trying to answer the question *"What are the current approaches for adaptive gameplay in fighting games?"*. The libraries as well as the searched period remained the same. The studies should be primarily focused on fighting games, thus the term "fighting game" or its plural version should be present in the title. The body of the text should contain at least one of the above-mentioned alternatives for "player model" or one of the above-mentioned alternatives for "adaptive". Comparing to the previous systematic review section 2.1, the primary and quality criteria remained the same, except for the first one, which was adapted to:

- primarily concerned about adaptive gameplay in fighting games

In total, 19 materials were positive (14 true and 5 false). A summary of the main important ones is presented here. Adaptiveness in fighting games may serve several purposes. Mimicking human players attempts to solve the problem of the AI agents implemented withing most traditional fighting games being simplistic and predictable. The idea here is to let the AI agent to learn and mimic a human player's strategy. An overview of the most relevant mimicking techniques is given by [35].

Most of the approaches, however, focus on the adaptiveness of an AI agent. In 2013, an academic competition platform for fighting games, referred to as "FightingICE", was introduced [36]. This provided space for various AI agents to be compared against each other. Since then, several approaches have been developed. The list of recently implemented adaptive techniques include: Dynamic Scripting with reinforcement learning [8], which utilizes a set of rules created with expert knowledge; Massive play data technique [37], which trades the need of the expert knowledge for more demanding memory requirements; online learning using k-Nearest Neighbor (kNN) prediction [38]; fuzzy control with kNN prediction [39], which addresses the problem of its "cold start"; or a combination of rule-based design with online learning [40].

The main purpose of the above-mentioned techniques is to either overcome their opponent or adjust the performance level of the AI agent. In both cases, the performance is mostly assessed by a one-dimensional vector: the score, which is directly related to the character's given and taken damage. Furthermore, these techniques work mostly with situation-action pairs. The purpose of the model presented in this study differs in *a) purpose* – it characterizes a player for the purpose of environmental adaptations; as well as in *b) assessment* – it attempts to measure and asses player's individual skills and behavioral traits, rather then overall performance expressed by a scalar value.

2.3 Dimensions

The goal was to propose a player model, which would as accurately as possible capture player's features that characterize them while interacting with their opponent in fighting games. It is believed, that the scientific literature about relevant skills and player models in fighting games on its own, would not be sufficient to build accurate skill-capturing player model. This is because it might lack the perspective and experience of actual players. Therefore, our proposal is also inspired by non-academic resources from players' community, such as online interviews with professional fighting game players, their on-line videos, online tutorials and online articles. This way, the proposed player model should better reflect the actual situation, as it also takes into account the expert knowledge from professional players.

We propose the following characteristics related to skills and playing style to comprise our player model (from now on, these will be referred to as *"dimensions"*, since the player model can also be expressed as a multidimensional vector):

- *Aggressiveness* is a typical characteristic of game playing style: It is often a part of player models describing player's strategies; for instance in real-time strategy games [34]. Therefore it might be one of the essential components of player interaction in m1on1 games.

- *Decision making* is an important part of strategy and tactics. Most of the AIs developed for fighting games are based on the state-action principle – predicting the future action based on the current state, for example [40]). This is, in fact, a decision making process, thus it undauntedly contributes to the description of playing style.

- *Distance estimate* represents visuo-spatial ability [41] applied in fighting games. The correct estimation of ranges of kicks or punches is important. Various actions have different sizes of hit-boxes (zones in which the action gives damage), as shown on Figure 1. Furthermore, distance estimate is very important in the game phase called *footsies* [42], which is usually in the beginning. Here, both players try to attack their opponents using normal attacks from as far as possible.

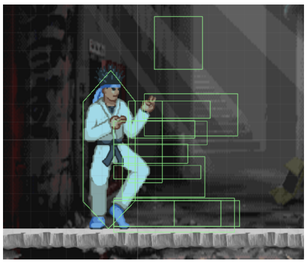


Figure 1: The collision volume (hexagon around the character) and the hit-boxes used in our implementation.

- *Muscle memory* represents execution of a sequence of button presses, which is a necessary part of fighting games [43]. This is because game mechanics are mostly based on different, quickly performed combinations of actions.

- *Reaction time* can have a significant effect on the fight results. The reason is the rock-paper-principle and actions' start-up phases (see Section 2.4). These two factors usually make every action defendable given that the right reaction (known as *counter hit* [42]) comes in time. The importance of this characteristic in fighting games interaction is also supported by [44], where it was extracted from a fighting game and proved to be a promising indicator of player's relaxation.

- *Space control* is the key to get and maintain an advantage over the opponent. The closer the character is to the wall behind them, the less spatial freedom they have [45]. In fighting games, both players usually try to stay as far from the wall as possible while maintaining a reasonable distance from their opponent. An example of spatial division, can be shown on Figure 2. Here, positions 2 and 4 are safest for both players.

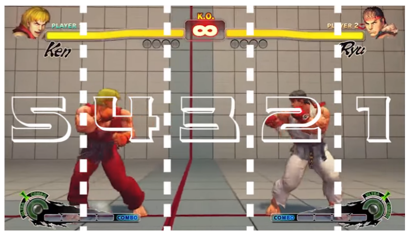


Figure 2: Space division in Street Fighter 4 [46]. (Image taken from [45]).

- *Timing precision* is very important when it comes to attack types, which the players' community denote as *frame traps*, *links* or *chains* [42]. They all have one common property: they include attacks that are performed before the opponent recovers from a previous hit. Consequently, the opponent has no time for any reaction.

The proposed model can be mapped to an existing concept of generalized player model introduced by Calleja [47]; and it is validated later in this paper. Table 1 displays the mapping of the proposed dimensions to the involvement dimensions of Calleja's [47] player model; which defines 6 types of player involvement: tactical, affective, spatial, performative, shared and narrative.

Table 1: Proposed dimensions mapped to Calleja's player model [47]

Dimension	Involvement
Aggressiveness	Tactical + Affective
Decision making	Tactical + Shared
Distance estimate	Spatial + Performative
Muscle memory	Performative
Reaction time	Shared + Performative
Space control	Spatial
Timing precision	Performative

2.4 Game Design Principles

In order to further work with the proposed dimensions, a simple fighting game has been developed from scratch in Unity. The game is set to always run in constant 60FPS. The main aim was to make it similar to commercial fighting games, including as most of the common concepts and mechanics present in these games, as possible. The most important ones are described below.

Figure 3: A screen-shot form the game.

The so-called "rock-paper-scissors" principle is often found in the most popular fighting games:

Figure 4: Health (green) and mana (blue) bars.

attack beats throw, throw beats block and block beats attack [48] The game also includes the standard mechanics for health and mana bars: each players starts with full amount of health points (HP) and 0 mana points (MP) (Figure 4). HP are subtracted when a player received some damage. MP are added over time and when a player is attacked and subtracted when a player performs more difficult action (referred to as "combo"). Activation of a combo requires a certain combination of actions to be performed in a sequence over short period of time. A combo indicator is displayed under the mana bar, showing the current state of combo activation in real-time (Figure 5).

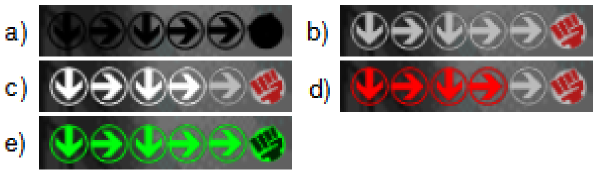


Figure 5: Combo indicator: a) not enough mana, b) enough mana = ready, c) activation in progress, d) mistake has been made, and e) combo successfully activated.

Depending on the type of action, each time a character receives some damage, the following three penalties may be applied: their HP may be decreased, they may be pushed further form the opponent and their ability to perform any action may be temporarily disabled for some time. Each non-movement action has three phases: start-up, active and recovery. These are defined in the number of frames and vary per each action. In the start-up and recovery phases, a player cannot do any other action, but is still exposed to the danger from the opponent; and in the active phase, the effect of the action is applied to the opponent [36].

The game mechanics include the following list of possible actions:

- **Movement:** crouch, walk backwards, walk forwards, jump up, jump forwards, jump backwards;

- **Attack:** short punch while: a) standing b) crouching c) jumping, long punch while: a) standing b) crouching c) jumping, short kick while: a) standing b) crouching c) jumping, long kick while: a) standing b) crouching c) jumping;

- **Throw:** one throw while standing;

- **Defense:** block while: a) standing b) crouching c) jumping;

- **Combo:** one special combo, consisting of five movement actions and one attack action (Figure 5)

The basic fighting mechanics (attack kicks and punches) were inspired by an open-source fighting game engine called MUGEN (by elecbyte [49], v1.0, available at [50]); the rest was developed in accordance with principles present in existing fighting games, described in section 2.4. Since the original MUGEN did not provide enough flexibility for the purpose of this study, the entire game was developed from scratch in Unity. Graphics for character and effects as well as sound effects were re-used from MUGEN.

3 Methods

The study consists of two phases. From now on, the first phase will be referred to as *PHASE-1* and the second one as *PHASE-2*. Although, both of them were conducted using the same hardware and with the same human participants, they are two completely different and independent experiments with different purposes, objectives, analyses and results.

PHASE-1 was conducted with primary purpose to evaluate metrics and measurement methods introduced later in Section 3.1, using both a simulation with AI participants – bots (objective validation), and a user experiment with human participants (subjective validation). On the contrary, in *PHASE-2*, only the human participants were needed. The main purpose of *PHASE-2* was to estimate the level of contribution of each dimension to the challenge factor of the game.

The structure of this section is following: firstly, the AI participants – bots (used in *PHASE-1*) are introduced, then human participants (used in both phases) are characterized, followed by the description of used apparatus (partially shared between both phases), and finally, procedures for both phases are defined separately.

3.1 Bots

To help to validate the proposed dimensions (Section 2.3), seven rule-based bots have been developed. Each bot has been designed to be as best as possible at exactly one dimension – their dimension of interest (DoI). Two random modules have been created, one for random movement and one for random fighting. These have been used with bots, whose primary purpose does not require implementation of movement or fighting aspects. Having developed bots that are supposed to be as best as possible in their DoI, the next step was to propose metrics to quantify the dimensions and methods to measure them. The bots and the metrics are described in Table 2.

3.2 Human Participants

The participants were 16 students of University of Malta, 10 males and 6 females, in the age group 19-30 years old (average: 22.2, standard deviation: 2.6), studying different disciplines with varying gaming experience. Seven of them were non-gamers, six stated to play games once a week and three of them were every-day players. Nine claimed to play fighting games occasionally, one regularly and six stated that they do not play fighting games at all. Two were left-handed and the rest of them were right-handed.

3.3 Apparatus

Hardware. Both *PHASE-1* and *PHASE-2* were run on a laptop Lenovo Y510P with Intel Core i7-4700MQ CPU 2.40GHz, 16GB RAM, Windows 8.1 64bit, GeForceGT 755M, and a 15.6" screen with the resolution of 1920 x 1080, 60Hz. The laptop keyboard with numpad and a wired optical USB

Table 2: Bots' descriptions together with definitions of metrics

Bot's DoI	Description of the bot	Metric
Aggressiveness	The bot keeps walking towards the opponent until it gets close enough to cause some damage, then it keeps attacking the opponent with attack actions randomly chosen from among the standing attack actions.	Average number of attacks per second while being near the enemy; the greater the better.
Decision making	This bot uses a simplified version of neural network, similar to [51],to choose the best possible action while taking into account several factors: the HP difference, the distance from the wall behind, the distance from the opponent as well as the opponent's current state.	Average difference between the chosen and the ideal decision scores; the lower the better.
Distance estimate	The random movement module is used in this bot. However, as soon as the bot gets into the position, from which it can attack the opponent, such that the range of the attack is just long enough to touch the opponent, it does so.	Average width of the overlapping area of the attack collision boxes and the opponent collision box, the lower the better.
Muscle memory	While having not enough mana to perform a combo, this bot relies completely on both the movement and the fighting random modules. However, as soon as the amount of MP is satisfactory to perform the combo, it does it as quickly as possible.	Average number of frames it takes a character to activate the combo, the lower the better.
Reaction time	The bot uses both movement and fighting modules up to the point, where it is under attack. Then it tries to either counter the attack, or avoid it as soon as possible.	Average reaction time a character needs to react to the enemies actions, the lower the better.
Space control	This bot is trying to push the opponent as close as possible back to the wall behind. It does it by all the means, using decision tree that takes into account the positions of both characters and the current status of the opponent. The bot uses counter attacks and blocks to force the opponent move backwards. Even if the opponent manages to push this bot backwards, the bot will attempt to throw the opponent on the other side, such that it will gain the advantage again.	Average distance from the wall behind the player, the greater the better.
Timing precision	The movement and fighting modules are used until the bot receives some damage. After the recovery, the bot attempts to perform two successive attacks, in between which, it gives the opponent (almost) no time (0 or 1 frame, depending on the arbitrary execution order in Unity) for any reaction.	Average number of frames between two consequent attacks, the lower the better.

gaming mouse with 2400 DPI and seven buttons were used as the input devices.

Software – PHASE-1. This phase consists of two parts: objective and subjective.

For the objective part, a simulation was prepared. This simulation ran fights between bots, one after each other and measured all their dimensions. One game rule was adjusted: a single game did not end unless enough data was recorded in order to measure both bots' DoI; the game continued even if one of the characters had already lost all their HP. However, some of the bots were not programmed to perform the combo at all (as it was not in accordance with their main priority), so no data fo Muscle Memory could be recorded. Therefore, the opponents of the Muscle Memory bot that did not perform the combo, were automatically assigned the worst possible value.

For the subjective part, an on-line survey was created using Google Forms. See a page-by-page transcript in Appendix A. Four versions of the survey were prepared. Each version was referring to a different group of videos.

The first four questions were identical in all versions: single choice questions about the basic demographics and experience with games (age, gender, frequency of playing games in general and fighting games specifically).

The subsequent questions differed per survey version. Each one of them was related to one video (provided in a YouTube embedded form). The video could be paused, rewound, and replayed as many times as necessary. Participants were asked to judge performance of two characters in the video, with respect to each dimension separately, using the options: Red $>$ Blue, Red $\approx$ Blue, or Red $<$ Blue.

In total, 21 videos of pre-recorded fights between all possible pairs of bot opponents were used. The measured values of all the dimensions form these fights had been stored for later comparison with human judgment. Each video showed a single fight between one pair of bots. The reasons for choosing these fights as a testing sample are following: firstly, bot vs bot fights lack the bias that might be introduced by a human player and secondly, since each bot was designed to be as best as possible in a different dimension, these videos should be diverse enough to provide users with sufficient information about each dimension.

In order to avoid participants getting bored or losing attention, the parts of videos, in which both players were statically standing for several seconds were cut out. This resulted in the average video length of 44 seconds, with the longest one being 70 seconds and the shortest one being 25 seconds long. For the same reason, the 21 videos were split into 4 groups (one group consisting of 6 videos and the other three of 5 videos). As mentioned above, every survey version referred to one of these video groups.

In total, three versions of the survey contained 9 questions, including 5 videos; and one version consisted of 10 questions, with 6 videos.

In all the versions, a brief explanation of all the dimensions was given: first explicitly between the questions 4 and 5 and then below every video question, so participants did not need to memorize it.

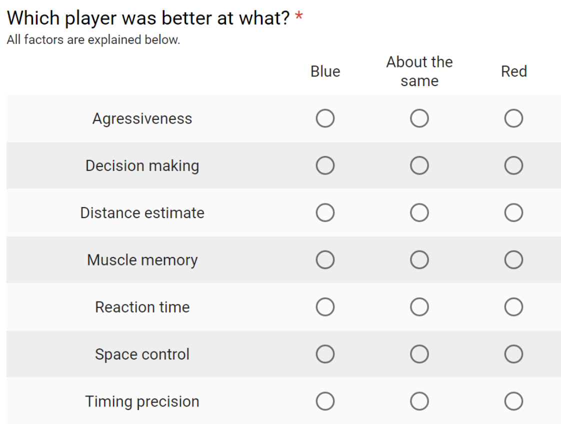


Figure 6: A question displayed below the video.

Software – PHASE-2. The survey was programmed into the existing game, so participants could stay within the same application throughout the whole process of testing. See a step-by-step screen-shots in Appendix B. By design, the controls as well as the set of moves required to perform a combo was always displayed in the middle of the screen (Figure 3) during the game.

The software started with two minutes of practice time, allowing the user to try various actions and to get used to the game mechanics. During the practice time, the opponent was in an idle state (statically standing and not reacting to the player's actions). Also, MP and HP of both players were constantly being recovered, making it impossible to beat one another.

The practice time was followed by seven bot-vs-user fights, where the user would face all the seven bots, one by one in random order. Each fight was considered to be finished when either of players' HP dropped to 0. Immediately after the end of each fight, a question asking to rate the level of experienced challenge appeared. A universally used [52] one-to-ten 1-dimensional scale was used; with 10 possible options, ranging from 1 labeled as "not challenging" to 10 labeled as "very challenging" (Figure 7).

With intention to get as accurate answers as possible, each time after selecting the latest answer, the software gave the user a chance to adjust the previous ones.

Despite the fact that by the time of taking part in *PHASE-2*, participants should have already seen 5 or 6 videos of the game play from *PHASE-1*, they might have, at first, estimated the extent of possible challenge inaccurately. Therefore, there is a chance that a possibility to change the previous answers could improve the accuracy of the final answers.

While rating the most recent bot, all the previous answers were shown. This should allow a player to rate the most recent bot more accurately as they can compare the current rating with each of the previous ones.

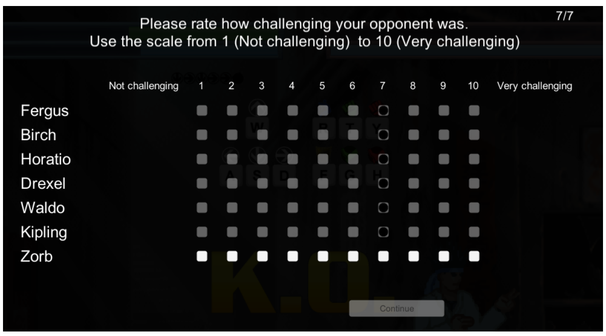


Figure 7: The survey inside the game, each row represents a different bot.

Additional materials – PHASE-2. A paper sheet (Appendix C) showing possible moves within the game mechanics was given to the participants before the start of *PHASE-2*. The sheet gives 13 examples, ranging from very basic ones, such as walking, to more difficult ones, like the combo.

3.4 Procedure – PHASE-1: Validation of Measurement Methods

The measurement methods proposed in Section 2.3 have been validated both objectively – using an off-line simulation with bots; as well as subjectively – by human judgment in a user experiment. Procedures of both parts are individually described below.

Bots. The off-line simulation consisted of 210 fights: all the seven bots played against each other ten times. For each game, two dimensions (DoI-s of both players) were measured using the proposed methods (Section 3.1). Expectedly, regardless of the opponent, each bot should always overcome their opponent in their DoI.

Human participants. Participants were split into four groups of 4 people. Each group was asked to fill in a different version of the on-line survey presented in Section 3.3-*Software*. The survey was opened in Google Chrome browser, in the full-screen mode, so participants could not be distracted by other pieces of software running on the laptop.

During this experiment, only a participant was present in the room. The instructor only came in to explain the procedure in the beginning, but spent the rest of the experiment in the room next door, giving participants enough space for concentration. However, the instructor was ready to answer any potential questions that might have arisen during the testing. The instructor entered the room once again after 2-3 minutes of testing to make sure the participant proceeds without any complications (some might be too shy to ask for help), but then he left again. Participants were instructed to notify the instructor when finished with *PHASE-1*. The instructor would then switch software to *PHASE-2*.

3.5 Procedure – PHASE-2: Challenge Factor Estimation

Upon the completion of *PHASE-1*, the instructor gave a participant new instructions for *PHASE-2* as well as the paper sheet (Appendix C) with example actions, and started corresponding software (Section 3.3-*Software*). Here, a participant was asked to play the game against each of the bots and rate their level of challenge immediately after every fight. The game was run in full-screen mode. The participant was asked to inform the instructor when the experiment was over. The instructor left the room as soon as the practice time was over and did only one check after 2-3 minutes of testing, similarly to the procedure of *PHASE-1*. The instructor was also ready to answer any potential questions at any time.

4 Results

The results section is split into two sub-sections, one focusing on the results of each phase individually. Different analyses were used for each of them, so the results from one phase should be considered as independent of and unrelated to the other.

4.1 PHASE-1: Validation of Measurement Methods

The results from the objective part of the phase (using bots) are presented first, followed by the results from the subjective part (with human participants).

Bots. In the 210 simulated fights, the expected result that a bot would beat their opponent in their DoI was achieved in 417 out of 420 cases (2 dimensions were measured for each fight), which is approximately 99.3%. The three cases that did not give the expected result share the same causes: a) the number of measured samples for one dimension during the fight was very low and thanks to this, when b) the random factor of the opponent's AI was extremely "lucky", it was possible to achieve even better results than the bot designed for this dimension. However, the differences, by which the opponents beat the original bots, were minor (namely 3.45%, 3.45% and 1.23%). These percentages are relative to *the possible ranges* of the dimensions, which were determined by the maximum and minimum values obtained from each dimension. For Muscle Memory, the range maximum was set to the maximum number of frames, in which it is possible to activate the combo.

Humans. First of all, the correspondence between the values measured in the fights that were shown in the videos (Section 3.3-*Software*) and participants' judgement (Section 3.4-*Human participants*) was analysed.

Every participant's answer was marked either as correct (C) when corresponding with the measured values or incorrect ($\neg C$) otherwise:

$$\begin{aligned} &\mathit{blue}\begin{cases} C & \text{if } R_{blue} > R_{red} \\ \neg C & \text{otherwise.} \end{cases} \\ &\mathit{same}\begin{cases} C & \text{if } |R_{blue} - R_{red}| \le t \\ \neg C & \text{otherwise.} \end{cases} \\ &\mathit{red}\begin{cases} C & \text{if } R_{blue} < R_{red} \\ \neg C & \text{otherwise.} \end{cases} \end{aligned} \tag{1}$$

R_{blue} and R_{red} denote performance of the blue and the red bots (with respect to the questioned dimension). The level of tolerance for the "*same*" option (t) was empirically determined as 15% of the dimension's *possible range*. Table 3 gives the correspondence expressed in percentage per each dimension. N denotes the number of cases. When no data was measured for either of the characters, the case was excluded from the analysis.

Table 3: Correspondence between the measured values and participants' judgement.

Dimension	Corresponding [%]	N
Aggressiveness	77.38	84
Decision making	67.86	84
Distance estimate	42.50	80
Muscle memory	47.50	40
Reaction time	51.19	84
Space control	58.33	84
Timing precision	40.00	60

To find out what correspondences are significant, the results were compared with the binomial distribution of the correct and incorrect answers (as described in [53]), calculated for each dimension separately as:

$$c(z) = z/N$$
$$z = \sum_{i}^{N} \{z_i\} \tag{2}$$

where z_i equals to:

$$z_i = \begin{cases} 1 & \text{if the anser is correct } (C) \\ -1 & \text{otherwise } (\neg C). \end{cases} \tag{3}$$

Since there are only two values possible for z_i, the distribution of probabilities for each $c(z)$ is binomial. However, in order to calculate it, the probability of success needs to be determined first. The presence of the "*same*" option in the survey questions and its tolerance (t=15%) affects the mean of the distribution in the following way: From the field of applied mathematics, we know that the probability that the absolute difference between two real numbers (a, b), randomly, uniformly and independently chosen from the range of 0%-100%, is lower or equal to 15%, is 27.75% (Equation (4)).

$$\begin{aligned}
&a, b \in \mathbb{R}, \quad b \in (0,1), \\
&\text{for } a \in (0,t): \\
&\quad p(|a-b| \le t) = \frac{t+2t}{2} = \frac{3}{2}t \\
&\text{for } a \in \langle t, 1-t \rangle: \\
&\quad p(|a-b| \le t) = 2t \\
&\text{for } a \in (1-t, 1): \\
&\quad p(|a-b| \le t) = \frac{t+2t}{2} = \frac{3}{2}t \\
&\text{for } a \in (0,1): \\
&\quad p(|a-b| \le t) = \\
&\quad = \frac{3}{2}t \cdot 2t + 2t \cdot (1-2t) = \\
&\quad = 2t - t^2 \\
&\quad p(|a-b| \le 15\%) = 27.75\%
\end{aligned} \tag{4}$$

With a large number of uniformly distributed samples of the measured value pairs, the options "*red*" and "*blue*" would both have 50% chance of being correct. The "*same*" option has tolerance of 15%, meaning that it is correct if the absolute difference between two measured values is less or equal to 15% of the dimension's *possible range*. As shown in Equation (4), the probability of this happening is 27.75%. Therefore, in 27.75 % of cases, two of the options would be correct (namely the option "*same*" together with either "*red*" or "*blue*"). Similarly, only one option would be correct in the rest 72.25% of cases. Thus, the probability of an option being correct (= probability of success) is equal to:

$$m = 27.75\% \cdot \frac{2}{3} + 72.25\% \cdot \frac{1}{3} \approx 42.58\% \tag{5}$$

Knowing the probability of success, the binomial distribution can now be drawn for each dimension. In order to find out, which correspondences are significant, the significance level of 1% was chosen: the correspondence is considered as significant if the p-value for given $c(z)$ is less than 1%. The p-value can be determined from the normal distribution because the binomially-distributed c(z) approximates the normal distribution when large samples are considered; the p-value is calculated as follows:

$$p = \begin{cases} P(C \le c) & \text{if } c(z) < m \\ P(C \ge c) & \text{if } c(z) \ge m \end{cases} \tag{6}$$

As an example, the binomial distribution of probability that the participants' judgment matched the measured values for muscle memory is shown on Figure 8; together with corresponding $c(z)$ and p values. In this case, the p-value is equal to the sum of all the probabilities of all c greater than or equal to $c(z)$. The rest of the dimensions have been calculated likewise.

Table 4 gives the results of the binomial distribution analysis for all dimensions, indicating that the significant correspondences are found in the following dimensions: Aggressiveness, Decision making and Space Control.

4.2 PHASE-2: Challenge Factor Estimation

In order to estimate the contribution of each dimension to the challenge factor of the game, the ratings recorded from the user survey (Section 3.5) have been analyzed using two different methods. The first one uses normalized values of the ratings whilst the second one converts the ratings into

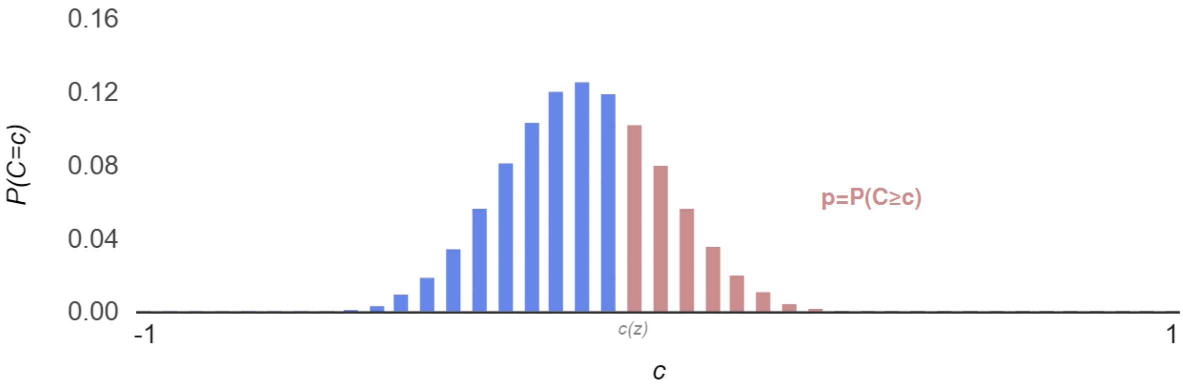


Figure 8: Binomial distribution of probability that the participants' judgment matches the measured values for muscle memory. N=40, $c(z)$=-0.05, p=0.3174 .

Table 4: Results from binomial distribution, significant dimensions and values (p less than 1%) are in bold. The + and − signs denote the number of cases, where z_i was positive or negative respectively.

Dimension	+	-	z	N	c(z)	p
Aggressiveness	65	19	46	84	0.55	$<$**0.0001**
Decision making	57	27	30	84	0.36	$<$**0.0001**
Distance estimate	34	46	-12	80	-0.15	0.5414
Muscle memory	19	21	-2	40	-0.05	0.3174
Reaction time	43	41	2	84	0.02	0.0693
Space control	49	35	14	84	0.17	**0.0026**
Timing precision	24	36	-12	60	-0.20	0.3947

rankings using preference pairs. Comparison of both methods and their results is given at the end of this section.

The results presented in the next two paragraphs are only indications taking into account only the main dimension of the bots. Although, it is expected that a dimension, for which a bot was designed, contributes to the human perception of challenge the most, it is not possible to separate dimensions from each other and therefore the minor influence of the other dimensions is always present. Thus, the results should be treated as estimations, rather than the exact values.

Method 1: Normalization. The first method uses normalized ratings. Normalization was done using a formula described in [54]. The equation takes into account maximum and minimum possible ratings (max and min) as well as maximum and minimum ratings given by a participant (max_p and min_p):

$$S_n = a \cdot S_p + b \tag{7}$$

S_n denotes the normalized score, S_p is a rating given by a participant, a and b are defined as follows:

$$\begin{aligned} a &= \frac{max - min}{max_p - min_p} \\ b &= max - a \cdot max_p \end{aligned} \tag{8}$$

This normalization converted all the ratings into real numbers between min and max. After all the values have been summed up for each direction, the contribution could be calculated (Table 5). The final estimated contribution to the game challenge factor is visualized on Figure 9.

Method 2: Preference pairs. Rating were first converted into rankings using pairwise preferences [55]. Then, ratings were converted into preference pairs (lines 6-18 in Algorithm 1), which were then used to calculate scores for each dimension (lines 19-21 in Algorithm 1). In the algorithm, P denotes the number of times a dimension is preferred in the preference pairs, N is the number of times a dimension appears in the preference pairs (either as preferred or non-preferred), d_X is the dimension to which rating X refers to and S is the dimension score. Table 6 gives the results.

Table 5: Results of the challenge factor phase; using normalization.

Dimension	total S_n	%
Aggressiveness	108.71	17.90
Decision making	113.80	18.73
Distance estimate	65.69	10.81
Muscle memory	54.71	9.01
Reaction time	96.64	15.91
Space control	95.37	15.70
Timing precision	72.51	11.94

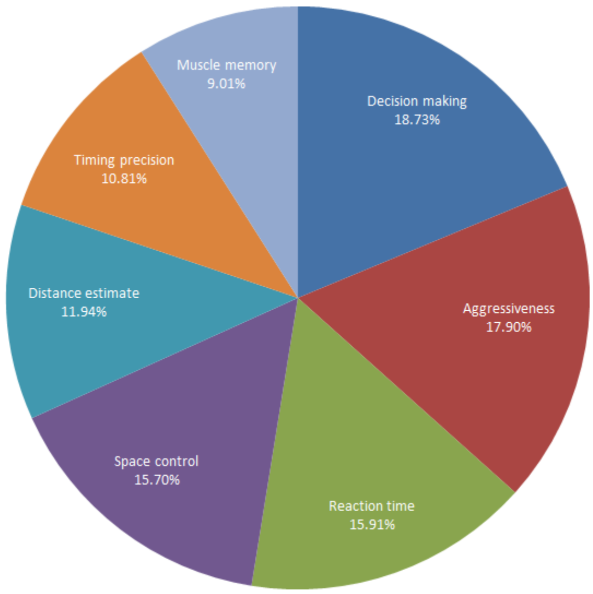


Figure 9: Estimated challenge factor using normalization.

Table 6: Results of the challenge factor phase; using preference pairs.

Dimension	P	N	S	%
Aggressiveness	58	87	0.67	18.98
Decision making	63	87	0.72	20.61
Distance estimate	31	90	0.34	9.81
Muscle memory	24	91	0.24	7.51
Reaction time	49	87	0.56	16.03
Space control	51	93	0.55	15.61
Timing precision	35	87	0.40	11.45

Algorithm 1 Conversion to preference pairs.

```
for all dimensions do
    d ← this dimension
    N[d] ← 0
    P[d] ← 0
    S[d] ← 0
for all participants do
    Take participant's ratings (7 in total)
    Make all possible pairs (21 in total)
    for all such pairs (A,B) do
        if A<>B then
            N[d_A] ← N[d_A] + 1
            N[d_B] ← N[d_B] + 1
            if A>B then
                Create a pref. pair (A ≻ B)
                P[d_A] ← P[d_A] + 1
            if A<B then
                Create a pref. pair (B ≻ A)
                P[d_B] ← P[d_B] + 1
for all dimensions do
    d ← this dimension
    S[d] ← P[d]/N[d]
```

Once the individual scores have been calculated for each dimension, they were summed up and the final percentages of their contribution to the challenge factor were calculated (the last column of Table 6 and Figure 10).

Table 7: Comparison of the results from both methods (short: Met.). All values are in %.

Dimension	Met. 1	Met. 2	Δ
Aggressiveness	17.90	18.98	1.08
Decision making	18.73	20.61	1.88
Distance estimate	10.81	9.81	1.01
Muscle memory	9.01	7.51	1.50
Reaction time	15.91	16.03	0.12
Space control	15.70	15.61	0.09
Timing precision	11.94	11.45	0.49

Comparison. The second method (conversion to rankings using preference pairs) only takes into account a boolean: whether rating A is greater or less then rating B, while ignoring by how much. Conversely, the first method (normalization) takes this difference into account too. This is the reason why the results are different. Although, it might be debatable what method is more accurate, the differences for all the dimensions are less than 5% (Table 7), therefore not significant. Furthermore, the average difference between the results from both methods is only 0.88%. This means that, in this case, both methods return very similar results and the difference by how much rating A is greater than rating B does not make any significant effect on the final results.

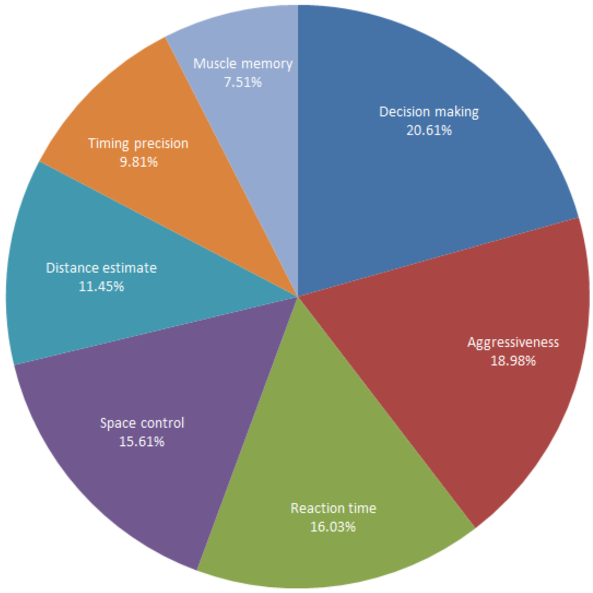


Figure 10: Estimated challenge factor using preference pairs.

5 Discussion

In this section, the interpretation of the results is provided, followed by reflection of the work done. The outcomes are then put into perspective of player modeling, explaining most important pros and cons of the model. Next, several possibilities for the future work are mentioned. Finally, the conclusion is presented.

Interpretation. The objective analysis using bots showed that in 99.3% of the cases, the results were in accordance with the expectations, which indicates a strong confirmation of correctness of the measurement methods. The subjective analyses suggests that the human perception of the dimensions is in significant correspondence with measurement methods in the following dimensions: aggressiveness, decision making and space control. The lowest correspondence has been identified in timing precision and distance estimate.

However, this does not necessary imply incorrectness of the last mentioned. This is because some of the dimensions can be classified as behavioral (e.g. aggressiveness) or more visually visible (e.g. space control). Both of these categories might be more evident and recognizable to human participants. Others are more of a technical manner (e.g. timing precision), which are undoubtedly harder for humans to compare and evaluate without using any other electronic devices. The results of the subjective analysis support this assumption, as the behavioral and more visually visible dimensions reached higher correspondence with humans' judgment than the ones defined more technically (requiring higher extent of technical precision in measurement).

The challenge factor of the developed fighting game has also been analyzed and individual dimensions' contributions to it have been estimated. The final outcome has been estimated by two different methods and their results differed in 0.88%. Both results indicate that the following dimensions contribute to the challenge factor most: decision making, aggressiveness, reaction time and space control. Only these four together take approximately 70%.

Reflection. Although, the results from *PHASE-2* may be very useful when it comes to implementation of this model, it is important to note the limitations of this phase: A) The game, in which the user testing is performed may significantly affect the results. Therefore, different results might be obtained from different games. B) Since the main objective was to evaluate contributions of the dimensions to the challenge factor, its a simplified version was considered; containing exclusively only the dimensions of the proposed player model. Although, this research tried to map player dimensions as accurately as possible, it cannot be excluded that there could be some other dimensions, besides the ones mentioned in this research, which may also contribute to the actual challenge factor of the game. C) The minor effects of the non-dominant dimensions interfering in the dominant one while playing against bots were ignored (as described in Section 4.2).

One drawback of the developed game mechanics was discovered when testing the bots. The game contained a winning strategy; the unbeatable combinations of moves were possible. This means that a player could perform such a combination of moves, which would not leave any time for opponent's reaction. For obvious reasons, it is very important in fighting games to make sure that a game does not contain such combinations. However, it usually takes exhaustive user tests to be certain that they are not present. Although, this problem was underestimated during the game development, the recorded fights from human experiments show that it did not have a significant effect on the results as participants discovered it very rarely.

Perspective. This work is a contribution to the class of skill-capturing player models. As demonstrated in the literature review, most of the player modeling approaches either cluster players into groups based on playing style or strategy, or attempt to predict the future actions based on the current state. However, player modeling in our case serves a purpose different than adaptive AI opponent, thus, the behavioral skill-capturing model was chosen. In the field of fighting games, this model provides more descriptive way to asses players. In the majority of present approaches, the performance of players is assessed and compared by some kind of score that takes into account either damage or health data. Our model proposes seven additional metrics that could be used to gain more accurate assessment.

The main advantage of this model is the identification the most important measurable and quantifiable player characteristics in fighting games. It provides a very specific representation of them, allowing for more descriptive multi-dimensional assessment of player's performance; as well as mutual comparison of players' skills. Furthermore, the dimensions are measured and updated in the real-time, which makes all necessary data available at any time. Also, the dimensions are being measured from the very first moment. It may take some time in the beginning when a new user starts playing the game to capture enough data for dimensions to be accurately quantified, but this process runs simultaneously as the user is playing the game. Thus, no explicit preliminary learning stage is required.

The limitation of this model is that it requires a game-specific tailoring when it comes to implementation of measuring methods, since the contribution of each dimension to the game challenge factor highly depends on the implemented game mechanics. As shown by the results, there is also room for improvement of the proposed dissension and their measuring methods, especially with regards to the human perception of more technically defined ones. Our current implementation does not take the learning process into account, which means that it does not capture the speed of player's improvement. This is very important to be included for any future applications.

Future work. As mentioned above, the potential for improvement lies in the implementation for measuring the learning process, since the current solution ignores the fact that the player's skills might be improved over time.

The proposed model was created with the aim of being as general as possible and thus potentially applicable to as many fighting games as possible. Although, some of the measuring methods may need to be adjusted for each specific game, the proposed dimensions provide a solid base for several potential application domains.

First of all, in multi-player game balancing, the player model represented by 7-dimensional vector provides more descriptive and more accurate characterization of the player. Therefore, it provides a potential for more game-balancing techniques than commonly used 1 dimension represented by handicap.

Next, the adaptive AI opponents in fighting games may be improved. Their behavior might become more accurate when taking into account more characteristics of the player in real-time. It could also boost their performance against a human player, since the proposed model could be used to identify the player's weaknesses.

Lastly, having this player model designed opens new possibilities for applications in yet unexplored area of adaptive environment in m1on1 games. Dynamically adjusted weights of various game mechanics could result in more accurate balancing, which could in turn lead to more competitive gameplay and higher player's enjoyment.

Conclusion. This paper proposed and analyzed a new behavioral skill-capturing player model for fighting games (consisting of: aggressiveness, decision making, distance estimate, muscle memory, reaction time, space control and timing precision). To validate the model, a new fighting game has been developed from scratch. For each dimension, a metric and a real-time measurement method was created and validated both subjectively and objectively. A rough estimation of the dimensions' contribution to the challenge factor the game was quantified. Although, the player model was validated for the game created specifically for this research, it is believed to be applicable for most of the traditional fighting games too. This is because the developed game was designed to be as similar to the most popular fighting games as possible; implementing the most common gameplay principles present in today's fighting games.

The proposed model provides more descriptive characterization of player's skills and their playing style than current approaches. It has a great potential to deal with various diversity of players in fighting games. Thus, is can certainly improve their assessment as well as provide more data for potential game adaptations. It is very promising for applications, where a complex real-time assessment of player's skills and playing style is needed.

References

[1] S. Shabani, F. Lin, and S. Graf, "A framework for user modeling in quizmaster," *Journal of e-Learning and Knowledge Society*, vol. 8, no. 3, 2012.

[2] A. Drachen, A. Canossa, and G. N. Yannakakis, "Player modeling using self-organization in tomb raider: underworld," in *2009 IEEE symposium on computational intelligence and games*, IEEE, 2009, pp. 1–8.

[3] G. Andrade, G. Ramalho, A. S. Gomes, and V. Corruble, "Dynamic game balancing: an evaluation of user satisfaction.," *AIIDE*, vol. 1, pp. 3–8, 2006.

[4] G. N. Yannakakis and J. Hallam, "Towards optimizing entertainment in computer games," *Applied Artificial Intelligence*, vol. 21, no. 10, pp. 933–971, 2007.

[5] D. Buckley, K. Chen, and J. Knowles, "Rapid skill capture in a first-person shooter," 2014.

[6] P. A. Nogueira, R. A. Rodrigues, E. C. Oliveira, and L. E. Nacke, "Guided emotional state regulation: understanding and shaping players' affective experiences in digital games.," in *AIIDE*, 2013.

[7] W. Doesburg, A. Heuvelink, and E. L. van den Broek, "TACOP: a cognitive agent for a naval training simulation environment," in *Proceedings of the Industry Track of the Fourth International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS-05)*, M. Pechoucek, D. Steiner, and S. Thompson, Eds., Utrecht, The Netherlands: New York, NY, USA: ACM, 2005, pp. 34–41.

[8] K. Majchrzak, J. Quadflieg, and G. Rudolph, "Advanced dynamic scripting for fighting game ai," in *International Conference on Entertainment Computing*, Springer, 2015, pp. 86–99.

[9] S. Natkin, C. Yan, S Jumpertz, and B Market, "Creating multiplayer ubiquitous fames using an adaptive narration model based on a user's model," in *Digital Games Research Association International Conference (DiGRA 2007)*, 2007.

[10] (2016). Google scholar, [Online]. Available: `scholar . google . com/` (visited on 02/30/2016).

[11] (2016). The acm digital library, [Online]. Available: `dl . acm . org` (visited on 02/30/2016).

[12] (2016). Web of science, [Online]. Available: `webofknowledge . com` (visited on 02/30/2016).

[13] M Machado, E Fantini, and L. Chaimowicz, "Player modeling: what is it? how to do it?" *Proceedings of SBGames*, 2011.

[14] M. Sharma, S. Ontañón, M. Mehta, and A. Ram, "Drama management and player modeling for interactive fiction games," *Computational Intelligence*, vol. 26, no. 2, pp. 183–211, 2010.

[15] A. Ramirez and V. Bulitko, "Automated planning and player modeling for interactive storytelling," *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 7, no. 4, pp. 375–386, 2015.

[16] S. Pourmehr and C. Dadkhah, "An overview on opponent modeling in robocup soccer simulation 2d," in *Robot Soccer World Cup*, Springer, 2011, pp. 402–414.

[17] B. Emond, H. Fournier, and J.-F. Lapointe, "Applying advanced user models and input technologies to augment military simulation-based training," in *Proceedings of the 2010 spring military modeling and simulation symposium*, 2010, pp. 1–7.

[18] L. Middleton, S. Crowle, and K. Meacham, "D3. 2 1st prosocial affect fusion and player modelling," 2015.

[19] K. d. O. Andrade, G. Fernandes, G. A. Caurin, A. A. Siqueira, R. A. Romero, and R. d. L. Pereira, "Dynamic player modelling in serious games applied to rehabilitation robotics," in *Robotics: SBR-LARS Robotics Symposium and Robocontrol (SBR LARS Robocontrol), 2014 Joint Conference on*, IEEE, 2014, pp. 211–216.

[20] J. R. Octavia, A. Beznosyk, K. Coninx, P. Quax, and K. Luyten, "User modeling approaches towards adaptation of users' roles to improve group interaction in collaborative 3d games," in *International Conference on Human-Computer Interaction*, Springer, 2011, pp. 668–677.

[21] P. Burelli and G. N. Yannakakis, "Adapting virtual camera behaviour through player modelling," *User Modeling and User-Adapted Interaction*, vol. 25, no. 2, pp. 155–183, 2015.

[22] C. P. Santos, V.-J. Khan, and P. Markopoulos, "On utilizing player models to predict behavior in crowdsourcing tasks," in *International Conference on Social Informatics*, Springer, 2014, pp. 448–451.

[23] J. C. C. Ramírez, A. S. López, and A. S. Flores, "An architecture for cognitive modeling to support real-time adaptation and motivational responses in video games," in *Mexican International Conference on Artificial Intelligence*, Springer, 2013, pp. 144–156.

[24] N. Oren and T. J. Norman, "Arguing using opponent models," in *International Workshop on Argumentation in Multi-Agent Systems*, Springer, 2009, pp. 160–174.

[25] G. N. Yannakakis, P. Spronck, D. Loiacono, and E. André, "Player modeling," *Dagstuhl Follow-Ups*, vol. 6, 2013.

[26] B. Bontchev, "Holistic player modeling for controling adaptation in video games," *e-Society 2016*, p. 11, 2016.

[27] C. Holmgård, A. Liapis, J. Togelius, and G. N. Yannakakis, "Monte-carlo tree search for persona based player modeling," in *Eleventh Artificial Intelligence and Interactive Digital Entertainment Conference*, 2015.

[28] A. J. Lockett and R. Miikkulainen, "Evolving opponent models for texas hold'em," in *2008 IEEE Symposium On Computational Intelligence and Games*, IEEE, 2008, pp. 31–38.

[29] H. P. Martínez, K. Hullett, and G. N. Yannakakis, "Extending neuro-evolutionary preference learning through player modeling," in *Proceedings of the 2010 IEEE conference on computational intelligence and games*, IEEE, 2010, pp. 313–320.

[30] A. Shantia, "Dynamic formation and opponent modeling in real time strategy games," *SC@ RUG 2011 proceedings*, p. 97, 2011.

[31] Y. Mutneja, "Player modelling in target-based games," *International Journal of Computer Applications*, vol. 130, no. 15, 2015.

[32] B. Cowley and D. Charles, "Behavlets: a method for practical player modelling using psychology-based player traits and domain specific features," *User Modeling and User-Adapted Interaction*, pp. 1–50, 2016.

[33] P. A. Nogueira, R. Aguiar, R. A. Rodrigues, E. C. Oliveira, and L. Nacke, "Fuzzy affective player models: a physiology-based hierarchical clustering method.," in *AIIDE*, 2014.

[34] S. C. Bakkes, P. H. Spronck, and G. van Lankveld, "Player behavioural modelling for video games," *Entertainment Computing*, vol. 3, no. 3, pp. 71–79, 2012.

[35] S. S. Saini, "Mimicking human player strategies in fighting games using game artificial intelligence techniques," PhD thesis, © Simardeep Singh Saini, 2014.

[36] F. Lu, K. Yamamoto, L. H. Nomura, S. Mizuno, Y. Lee, and R. Thawonmas, "Fighting game artificial intelligence competition platform," in *2013 IEEE 2nd Global Conference on Consumer Electronics (GCCE)*, IEEE, 2013, pp. 320–323.

[37] H. Park and K.-J. Kim, "Learning to play fighting game using massive play data," in *2014 IEEE Conference on Computational Intelligence and Games*, IEEE, 2014, pp. 1–2.

[38] Y. Nakagawa, K. Yamamoto, and R. Thawonmas, "Online adjustment of the ai's strength in a fighting game using the k-nearest neighbor algorithm and a game simulator," in *2014 IEEE 3rd Global Conference on Consumer Electronics (GCCE)*, IEEE, 2014, pp. 494–495.

[39] C. Y. Chu and R. Thawonmas, "Applying fuzzy control in fighting game ai," *Information Processing Society 77th Annual Conference*, vol. 4, p. 02, 2015.

[40] N. Sato, S. Temsiririrkkul, S. Sone, and K. Ikeda, "Adaptive fighting game computer player by switching multiple rule-based controllers," in *Applied Computing and Information Technology/2nd International Conference on Computational Science and Intelligence (ACIT-CSI), 2015 3rd International Conference on*, IEEE, 2015, pp. 52–59.

[41] F. Meijer and E. L. van den Broek, "Representing 3d virtual objects: interaction between visuo-spatial ability and type of exploration," *Vision research*, vol. 50, no. 6, pp. 630–635, 2010.

[42] (2015). Concepts every 2d fighting game player should know, [Online]. Available: `https://www.youtube.com/watch?v=nd9sEB6ku14` (visited on 02/25/2016).

[43] N. D. Sorenson, "The evolution of fun: a generic model of video game challenge for automatic level design," PhD thesis, Communication, Art & Technology: School of Interactive Arts and Technology, 2010.

[44] P. Jarnfelt, S. Selvig, and D. Dimovska, "Towards tailoring player experience in physical wii games: a case study on relaxation," in *Proceedings of the International Conference on Advances in Computer Enterntainment Technology*, ACM, 2009, pp. 328–331.

[45] (2015). Concepts every 2d fighting game player should know, [Online]. Available: `https://www.youtube.com/watch?v=3y8GyMDt2fU` (visited on 02/25/2016).

[46] (2008). Streetfighter, [Online]. Available: `http://store.steampowered.com/app/21660/` (visited on 02/26/2016).

[47] G. Calleja, "Revising immersion: a conceptual model for the analysis of digital game involvement," *Situated Play*, pp. 24–28, 2007.

[48] Y. I. Gingold, "From rock, paper, scissors to street fighter ii: proof by construction," in *Proceedings of the 2006 ACM SIGGRAPH symposium on Videogames*, ACM, 2006, pp. 155–158.

[49] (2013). Elecbyte, [Online]. Available: `http://www.elecbyte.com/mugendocs-11b1/mugen.html` (visited on 07/30/2016).

[50] (2016). M.u.g.e.n, [Online]. Available: `http://mugen.en.softonic.com/` (visited on 07/30/2016).

[51] C. Holmgård, A. Liapis, J. Togelius, and G. N. Yannakakis, "Evolving personas for player decision modeling," in *2014 IEEE Conference on Computational Intelligence and Games*, 2014, pp. 1–8.

[52] R. D. Wimmer and J. R. Dominick, *Mass Media Research: An Introduction*, 10th ed. Cengage learning, 2014, ch. 2, pp. 42–63.

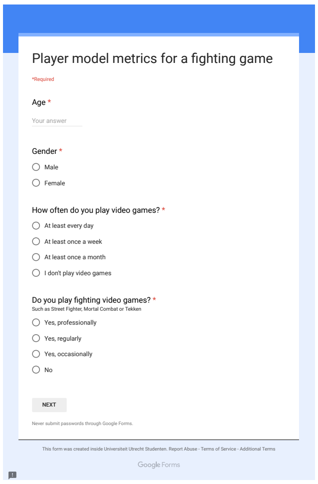
[53] G. N. Yannakakis and J. Hallam, "Towards optimizing entertainment in computer games," *Applied Artificial Intelligence*, vol. 21, no. 10, pp. 933–971, 2007.

[54] E. L. Van Den Broek, P. M. Kisters, and L. G. Vuurpijl, "Content-based image retrieval benchmarking: utilizing color categories and color distributions," *Journal of imaging science and technology*, vol. 49, no. 3, pp. 293–301, 2005.

[55] H. P. Martinez, G. N. Yannakakis, and J. Hallam, "Don't classify ratings of affect; rank them!" *IEEE Transactions on Affective Computing*, vol. 5, no. 3, 2014.

Appendix A Google Form

The transcript of the Google form used in the experiment with human participants in *PHASE-1*. Each one of the nine columns represents one step in the form. Steps 1; 2; and 3:



Player model metrics for a fighting game

*Required

Age *

Your answer

Gender *

- Male
- Female

How often do you play video games? *

- At least every day
- At least once a week
- At least once a month
- I don't play video games

Do you play fighting video games? *

Such as Street Fighter, Mortal Combat or Tekken

- Yes, professionally
- Yes, regularly
- Yes, occasionally
- No

NEXT

Never submit passwords through Google Forms.

This form was created inside Universiteit Utrecht Studenten. Report Abuse - Terms of Service - Additional Terms

Google Forms

Player model metrics for a fighting game

You will watch several videos. Pay attention to the following factors:

The notes below will be available at the end of each section, so you don't need to memorise any of it.

Agressiveness

Which character plays more aggressively, regardless of who won and who lost the game.

Decision making

Which character do you think makes decision based on a current situation?

Distance estimate

Which character is able to estimate the range of their attacks better? It is best to hit the opponent when they are just inside of the range of a punch or kick.

Muscle memory

Which character is more successful in execution of a special combo attack? The following indicators show when a character is trying to execute a combo and its current status.

Reaction time

Which character has better reaction time? Reaction time is important when it comes to blocking or avoiding incoming attacks.

Space control

Which character is able to keep better position throughout the game? It is more advantageous to keep as far as possible from the wall behind you since it is hard to escape from the opponent while being stuck in the corner.

Timing precision

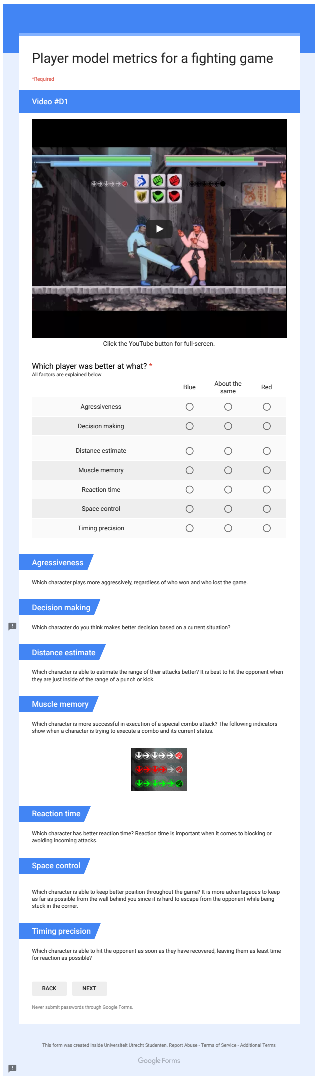
Which character is able to hit the opponent as soon as they have recovered, leaving them as least time for reaction as possible?

BACK NEXT

Never submit passwords through Google Forms.

This form was created inside Universiteit Utrecht Studenten. Report Abuse - Terms of Service - Additional Terms

Google Forms



Player model metrics for a fighting game

*Required

Video #D1

Click the YouTube button for full-screen.

Which player was better at what? *

All factors are explained below.

	Blue	About the same	Red
Agressiveness	○	○	○
Decision making	○	○	○
Distance estimate	○	○	○
Muscle memory	○	○	○
Reaction time	○	○	○
Space control	○	○	○
Timing precision	○	○	○

Agressiveness

Which character plays more aggressively, regardless of who won and who lost the game.

Decision making

Which character do you think makes better decision based on a current situation?

Distance estimate

Which character is able to estimate the range of their attacks better? It is best to hit the opponent when they are just inside of the range of a punch or kick.

Muscle memory

Which character is more successful in execution of a special combo attack? The following indicators show when a character is trying to execute a combo and its current status.

Reaction time

Which character has better reaction time? Reaction time is important when it comes to blocking or avoiding incoming attacks.

Space control

Which character is able to keep better position throughout the game? It is more advantageous to keep as far as possible from the wall behind you since it is hard to escape from the opponent while being stuck in the corner.

Timing precision

Which character is able to hit the opponent as soon as they have recovered, leaving them as least time for reaction as possible?

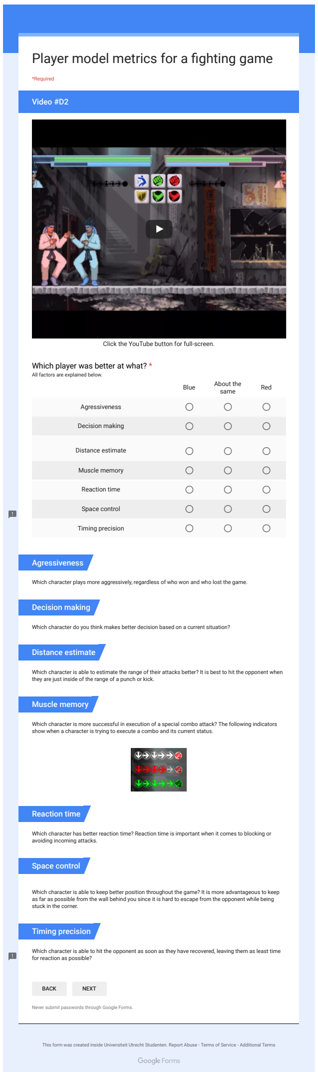
BACK NEXT

Never submit passwords through Google Forms.

This form was created inside Universiteit Utrecht Studenten. Report Abuse - Terms of Service - Additional Terms

Google Forms

Steps 4; 5; and 6:



Player model metrics for a fighting game

*Required

Video #D2

Click the YouTube button for full-screen.

Which player was better at what? *

All factors are explained below.

	Blue	About the same	Red
Agressiveness	○	○	○
Decision making	○	○	○
Distance estimate	○	○	○
Muscle memory	○	○	○
Reaction time	○	○	○
Space control	○	○	○
Timing precision	○	○	○

Agressiveness

Which character plays more aggressively, regardless of who won and who lost the game.

Decision making

Which character do you think makes better decision based on a current situation?

Distance estimate

Which character is able to estimate the range of their attacks better? It is best to hit the opponent when they are just inside of the range of a punch or kick.

Muscle memory

Which character is more successful in execution of a special combo attack? The following indicators show when a character is trying to execute a combo and its current status.

Reaction time

Which character has better reaction time? Reaction time is important when it comes to blocking or avoiding incoming attacks.

Space control

Which character is able to keep better position throughout the game? It is more advantageous to keep as far as possible from the wall behind you since it is hard to escape from the opponent while being stuck in the corner.

Timing precision

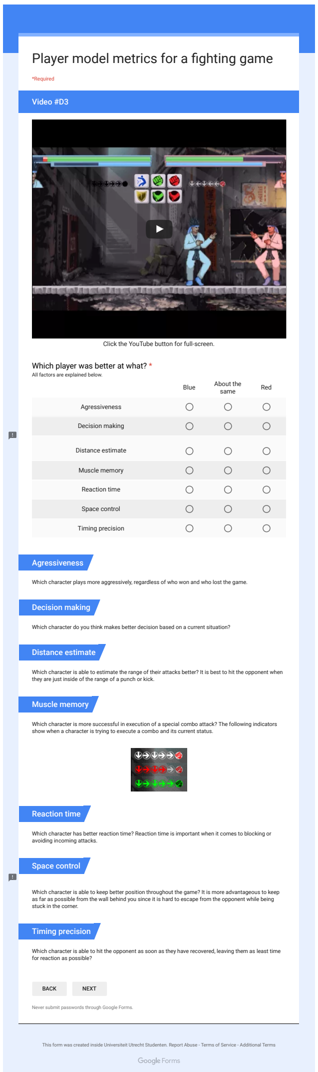
Which character is able to hit the opponent as soon as they have recovered, leaving them as least time for reaction as possible?

BACK NEXT

Never submit passwords through Google Forms.

This form was created inside Universiteit Utrecht Studenten. Report Abuse - Terms of Service - Additional Terms

Google Forms



Player model metrics for a fighting game

*Required

Video #D3

Click the YouTube button for full-screen.

Which player was better at what? *

All factors are explained below.

	Blue	About the same	Red
Agressiveness	○	○	○
Decision making	○	○	○
Distance estimate	○	○	○
Muscle memory	○	○	○
Reaction time	○	○	○
Space control	○	○	○
Timing precision	○	○	○

Agressiveness

Which character plays more aggressively, regardless of who won and who lost the game.

Decision making

Which character do you think makes better decision based on a current situation?

Distance estimate

Which character is able to estimate the range of their attacks better? It is best to hit the opponent when they are just inside of the range of a punch or kick.

Muscle memory

Which character is more successful in execution of a special combo attack? The following indicators show when a character is trying to execute a combo and its current status.

Reaction time

Which character has better reaction time? Reaction time is important when it comes to blocking or avoiding incoming attacks.

Space control

Which character is able to keep better position throughout the game? It is more advantageous to keep as far as possible from the wall behind you since it is hard to escape from the opponent while being stuck in the corner.

Timing precision

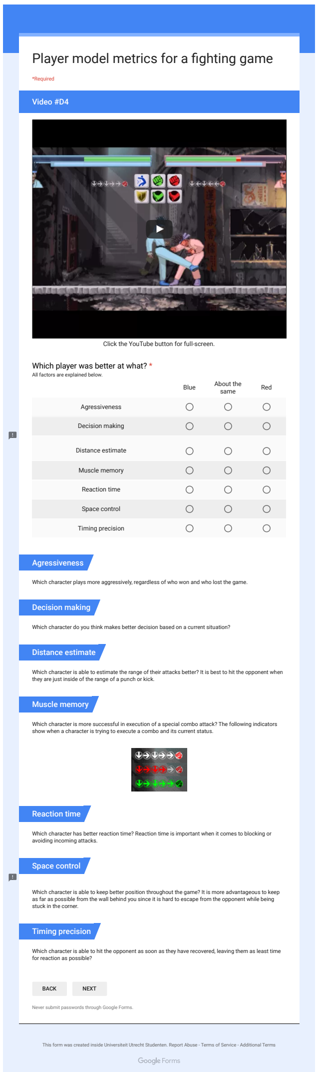
Which character is able to hit the opponent as soon as they have recovered, leaving them as least time for reaction as possible?

BACK NEXT

Never submit passwords through Google Forms.

This form was created inside Universiteit Utrecht Studenten. Report Abuse - Terms of Service - Additional Terms

Google Forms



Player model metrics for a fighting game

*Required

Video #D4

Click the YouTube button for full-screen.

Which player was better at what? *

All factors are explained below.

	Blue	About the same	Red
Agressiveness	○	○	○
Decision making	○	○	○
Distance estimate	○	○	○
Muscle memory	○	○	○
Reaction time	○	○	○
Space control	○	○	○
Timing precision	○	○	○

Agressiveness

Which character plays more aggressively, regardless of who won and who lost the game.

Decision making

Which character do you think makes better decision based on a current situation?

Distance estimate

Which character is able to estimate the range of their attacks better? It is best to hit the opponent when they are just inside of the range of a punch or kick.

Muscle memory

Which character is more successful in execution of a special combo attack? The following indicators show when a character is trying to execute a combo and its current status.

Reaction time

Which character has better reaction time? Reaction time is important when it comes to blocking or avoiding incoming attacks.

Space control

Which character is able to keep better position throughout the game? It is more advantageous to keep as far as possible from the wall behind you since it is hard to escape from the opponent while being stuck in the corner.

Timing precision

Which character is able to hit the opponent as soon as they have recovered, leaving them as least time for reaction as possible?

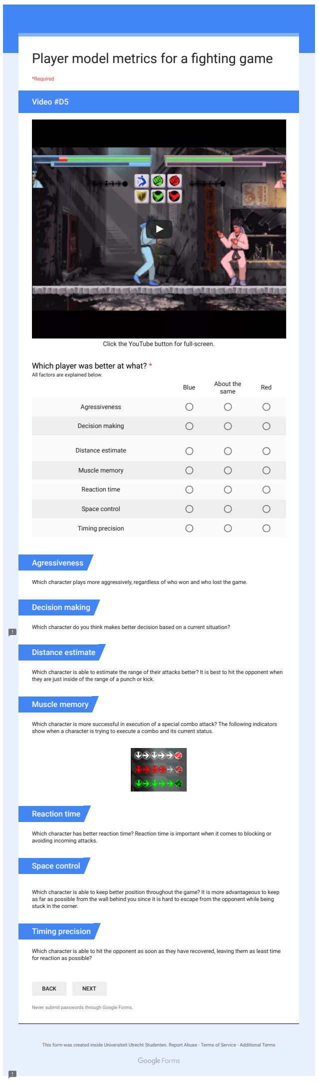
BACK NEXT

Never submit passwords through Google Forms.

This form was created inside Universiteit Utrecht Studenten. Report Abuse - Terms of Service - Additional Terms

Google Forms

Steps 7; 8; and 9:



Player model metrics for a fighting game

*Required

Video #D5

Click the YouTube button for full-screen.

Which player was better at what? *

All factors are explained below.

	Blue	About the same	Red
Agressiveness	○	○	○
Decision making	○	○	○
Distance estimate	○	○	○
Muscle memory	○	○	○
Reaction time	○	○	○
Space control	○	○	○
Timing precision	○	○	○

Agressiveness

Which character plays more aggressively, regardless of who won and who lost the game.

Decision making

Which character do you think makes better decision based on a current situation?

Distance estimate

Which character is able to estimate the range of their attacks better? It is best to hit the opponent when they are just inside of the range of a punch or kick.

Muscle memory

Which character is more successful in execution of a special combo attack? The following indicators show when a character is trying to execute a combo and its current status.

Reaction time

Which character has better reaction time? Reaction time is important when it comes to blocking or avoiding incoming attacks.

Space control

Which character is able to keep better position throughout the game? It is more advantageous to keep as far as possible from the wall behind you since it is hard to escape from the opponent while being stuck in the corner.

Timing precision

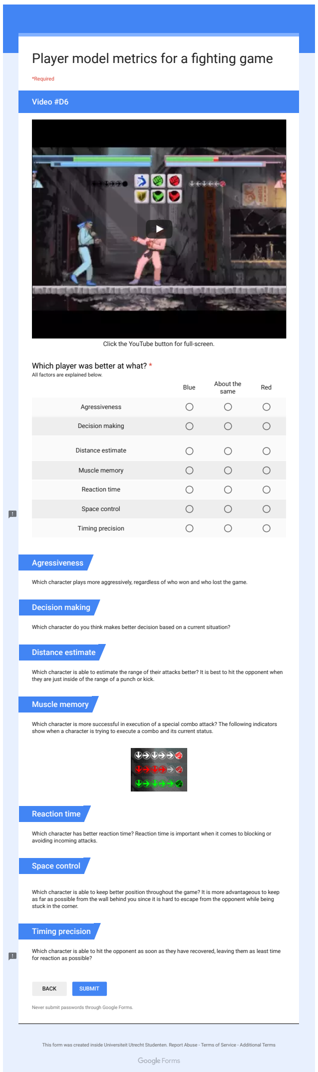
Which character is able to hit the opponent as soon as they have recovered, leaving them as least time for reaction as possible?

BACK NEXT

Never submit passwords through Google Forms.

This form was created inside Universiteit Utrecht Studenten. Report Abuse - Terms of Service - Additional Terms

Google Forms



Player model metrics for a fighting game

*Required

Video #D6

Click the YouTube button for full-screen.

Which player was better at what? *

All factors are explained below.

	Blue	About the same	Red
Agressiveness	○	○	○
Decision making	○	○	○
Distance estimate	○	○	○
Muscle memory	○	○	○
Reaction time	○	○	○
Space control	○	○	○
Timing precision	○	○	○

Agressiveness

Which character plays more aggressively, regardless of who won and who lost the game.

Decision making

Which character do you think makes better decision based on a current situation?

Distance estimate

Which character is able to estimate the range of their attacks better? It is best to hit the opponent when they are just inside of the range of a punch or kick.

Muscle memory

Which character is more successful in execution of a special combo attack? The following indicators show when a character is trying to execute a combo and its current status.

Reaction time

Which character has better reaction time? Reaction time is important when it comes to blocking or avoiding incoming attacks.

Space control

Which character is able to keep better position throughout the game? It is more advantageous to keep as far as possible from the wall behind you since it is hard to escape from the opponent while being stuck in the corner.

Timing precision

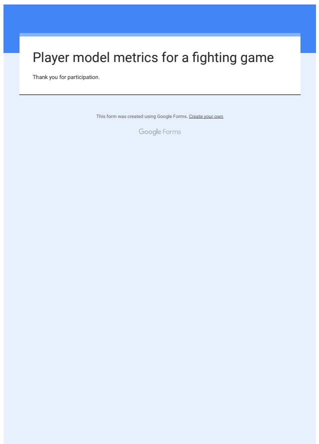
Which character is able to hit the opponent as soon as they have recovered, leaving them as least time for reaction as possible?

BACK SUBMIT

Never submit passwords through Google Forms.

This form was created inside Universiteit Utrecht Studenten. Report Abuse - Terms of Service - Additional Terms

Google Forms



Player model metrics for a fighting game

Thank you for participation.

This form was created using Google Forms. Create your own

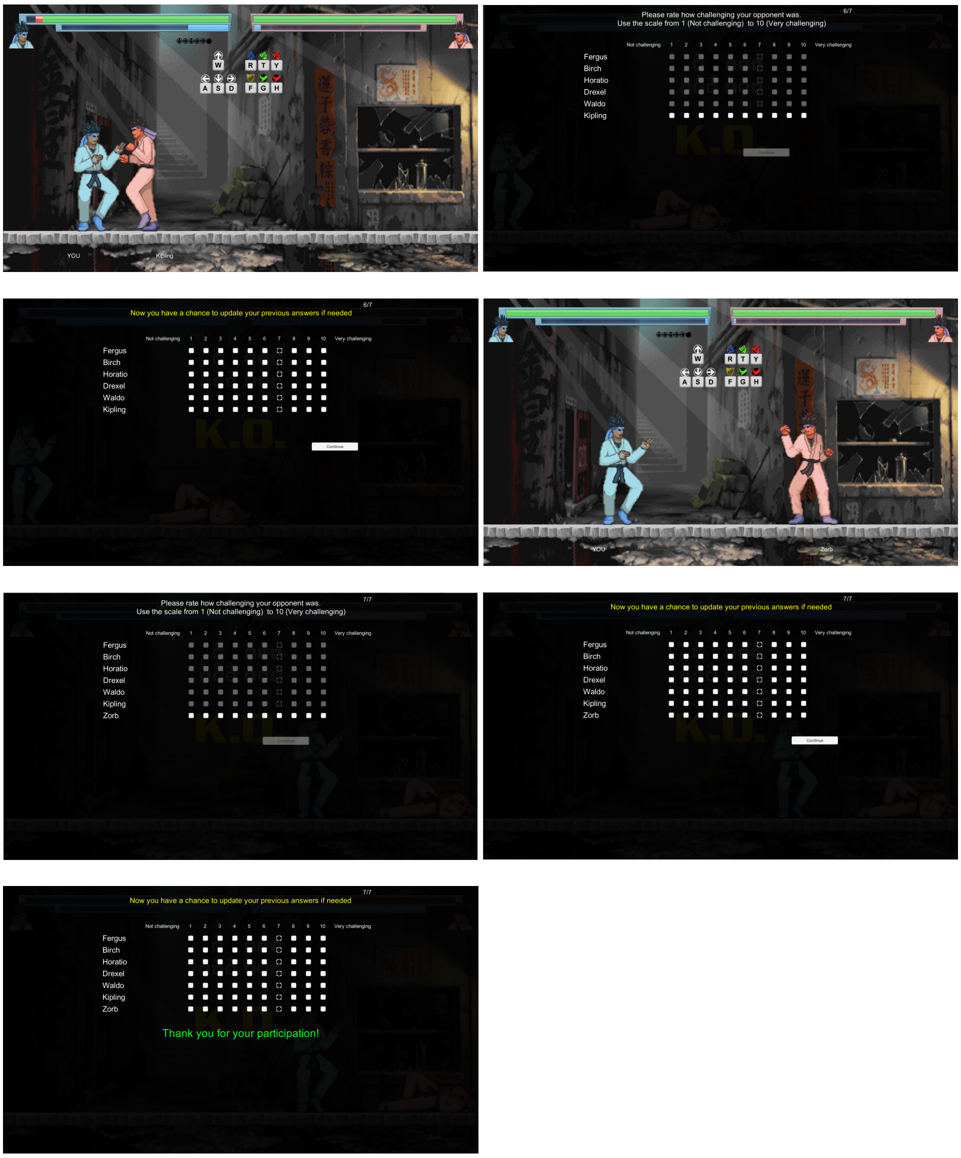
Google Forms

Appendix B Challenge factor survey

The following screenshots show the step-by-step process of *PHASE-2* experiment. From left-to-right top-to-bottom: 1) Practice time; 2) Demographics question; 3) Fight#1; 4) Evaluation of Fight#1; 5) Fight#2; 6) Evaluation of Fight#2; 7) A chance to adjust the previous answers; 8) Fight#3;

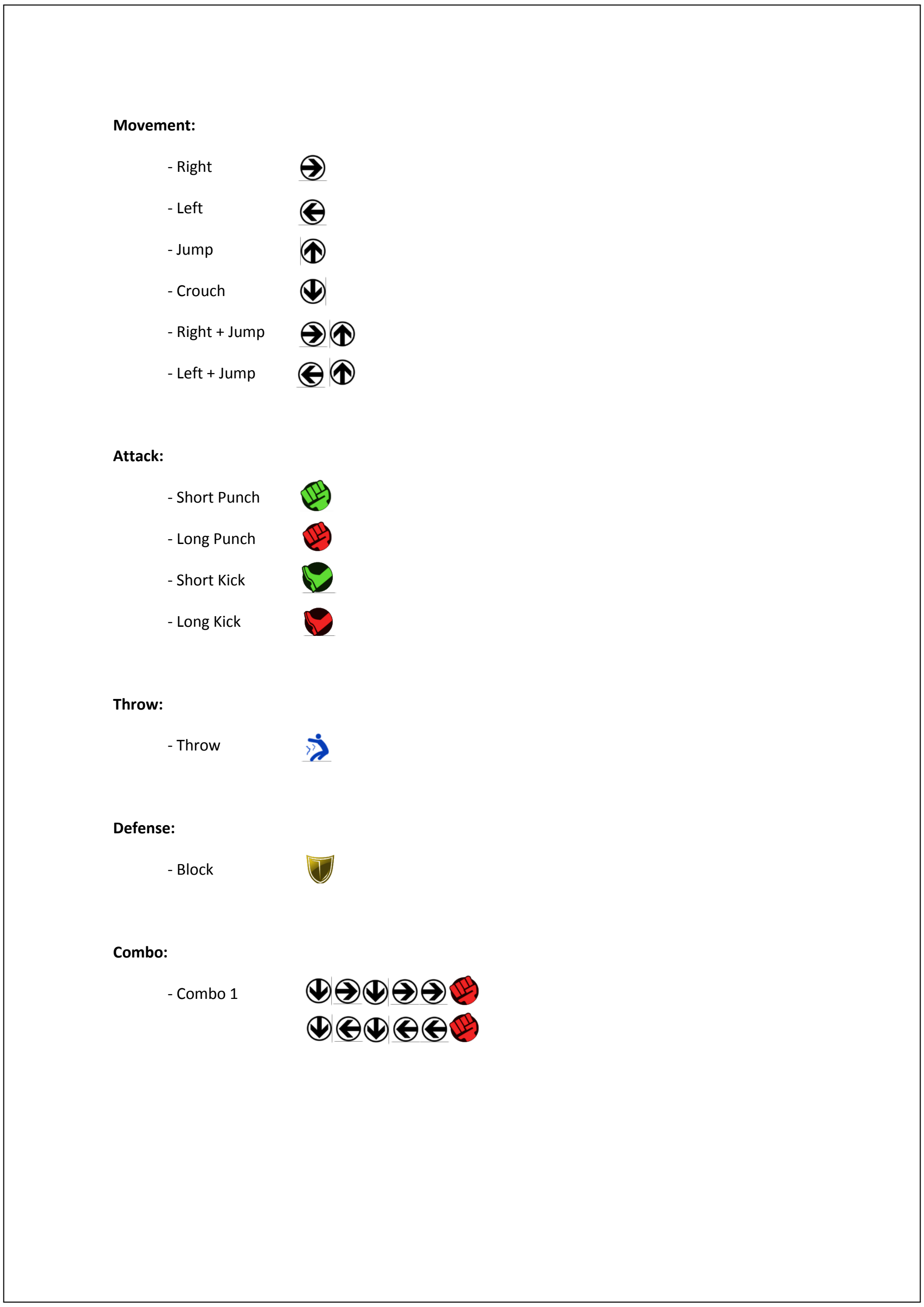
From left-to-right top-to-bottom: 9) Evaluation of Fight#3; 10) A chance to adjust the previous answers; 11) Fight#4; 12) Evaluation of Fight#4; 13) A chance to adjust the previous answers; 14) Fight#5; 15) Evaluation of Fight#5; 16) A chance to adjust the previous answers;

From left-to-right top-to-bottom: 17) Fight#6; 18) Evaluation of Fight#6; 19) A chance to adjust the previous answers; 20) Fight#7; 21) Evaluation of Fight#7; 22) A chance to adjust the previous answers; 23) Final Page.



Appendix C Examples of possible moves

This paper sheet was given to the participants at the beginning of *PHASE-2* in a printed form.



Movement:

- Right
- Left
- Jump
- Crouch
- Right + Jump
- Left + Jump

Attack:

- Short Punch
- Long Punch
- Short Kick
- Long Kick

Throw:

- Throw

Defense:

- Block

Combo:

- Combo 1