Human Perception of Human-Likeness in Fighting Game

Gai Nakazawa s1280209 Supervised by Prof. Maxim Mozgovoy

Abstract

AI to control NPCs is important in games of all genres, and the same is true for fighting games. If the behavior of NPCs is artificial and unnatural, it leads to a decrease in the player's interest in the game. We focus on the human perception of NPC behavior. The purpose of this study is to investigate how humans evaluate human-like behavior in response to human- or AI-generated behavior. To achieve this, we created an AI based on human play and evaluated it using a free-form questionnaire. As a result, we created an AI at a level indistinguishable from human-controlled play, and analysis of the evaluations revealed that there are differences in how people judge human and AI behavior.

1 Introduction

Video games, which have become a popular form of entertainment, have developed into a variety of genres, and fighting games are one of those categories. In most fighting games, players engage in one-on-one battles in a virtual space, using unique characters and techniques.

Artificial intelligence (AI) plays an important role in games of all genres, in fighting games AI plays an important role by controlling Non-Player Characters (NPCs) to become the player's opponents. Its main role is to provide challenging and engaging gameplay, adjusting the difficulty level, and employing a variety of tactics to keep the player engaged. To achieve this, developers are focusing on implementing the human-like behavior of NPCs. Realistic behavior and reactions increase immersion and make competitive play dynamic and unpredictable. Human-like behavior contributes not only to the overall game experience but also to the appeal of the fighting game itself.

There are several ways to evaluate the human likeness of AI, such as questionnaire-based evaluations and automated evaluations[1]. We focus on the behaviors and factors that provide clues to the judgments made by the evaluators in the Turing test, a questionnaire-based evaluation. The Turing test was called "Imitation game" at the time of its inception[2], in which human evaluators judged human-like conversational opportunities and human natural language conversations, but It is now widely applied, occasionally with modifications, to nonverbal tasks, such as robots in the real world and agents in the virtual world [3][4].

According to a study on the perception of AI and human behavior by the evaluator in a 3D video game navigation task[4], "jerky body movements", "movements unrelated to human play", and "goal indirect" were more frequently associated with AI, while "movements related to human play", "smooth body movements", and “goal indirect" were more frequently associated with human behavior. The 3D navigation task allows more flexibility in its operation than the environment we employ in this study. Therefore, it is expected that opinions on behavior will be more focused on a few specific elements and that there will be differences in the points of attention.

The purpose of this study is to investigate how humans evaluate human-like behaviors against human- and AI-generated behaviors. To achieve this, we create a human-like AI based on human play and discuss the factors that observers use as cues for their choices by evaluating the AI.

2 Method

2.1 Fighting Game

In this study, the environment for the experiment is the boxing game "Boxing Arena" developed by Maxim Mozgovoy et al. This game is considerably simpler than most fighting games in that it has no jumping or moving actions, no combos by performing a series of moves, and no special moves. The simplicity of the game, with few actions to control, is supposed to allow the evaluator to understand the factors more precisely to be focused on when distinguishing behavior in the Turing test. In addition, the lack of special moves and combos makes it less difficult for inexperienced fighting game players to make decisions.

The actions that the player can control are “Jab”, “Hook”, “Uppercut”, “Dodge”, “Body blow”, and “Block”, except for Block, which has two directions, left and right.

Figure 1: Screenshot from the game of Boxing Arena

2.2 AI System

2.2.1 Artificial Contender

Artificial Contender [3] is middleware for introducing AI into games. Instead of programming AI, developers can create agents that behave like humans by playing games and using the recorded files to train the AI. The AI created by AC is also referred to here as ACAI.

2.2.2 AC Game Viewer

AC game viewer is an important tool for AC, designing and tuning ACAI can be done through it. It also allows the user to review information about the current game situation by replaying a human-play recording of each game. Through the AC game viewer, users can load the AC knowledge file, which is the knowledge information generated by playing and training the game and used by ACAI to perform the game. This allows users to build and check the knowledge base in their environment.

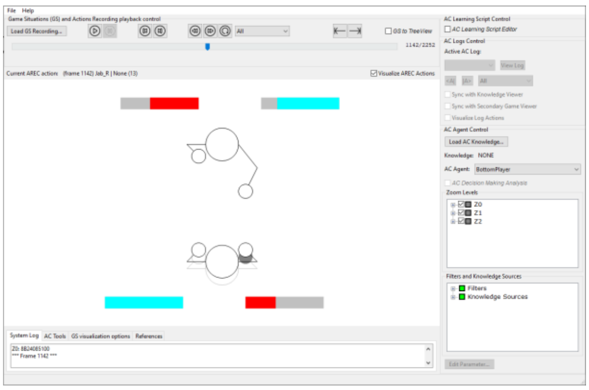

Figure 2: AC Game Viewer

2.2.3 Acting Graph

The acting graph is a type of finite state machine and is the primary data structure for storing and representing knowledge in ACAI. By playing the game in the learning mode of the game development environment, the knowledge base is built and updated based on the actions performed by the player and the situation at that time (the character's strength and stamina, opponent's actions, etc.).

The nodes of the acting graph represent the game state, and the edges represent actions that change the game state. Each edge (action) is weighted according to the frequency of the action, so not all options are selected with equal probability. This means that different action patterns will be generated depending on the supervised player's play style, even in the context of the same set of actions.

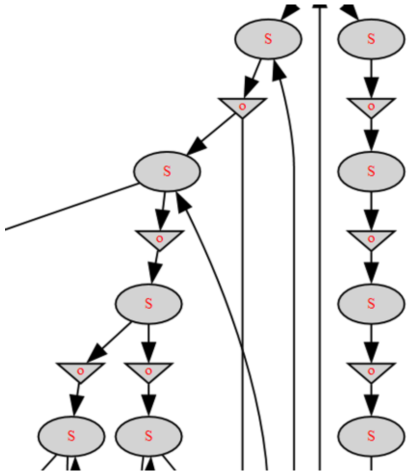

Figure 3: A part of the acting graph

2.2.4 Acting AI

ACAI determines the next action for the current game situation based on the constructed knowledge base. Specifically, ACAI looks for nodes in the action graph that are the same as the current situation and selects one of the edges to go out from that node. (case-based inference). Even if the same situation cannot be found, the action is determined by searching for similar situations. The level of abstraction (called zoom level) is used to find similar situations and can be checked in the AC game viewer.

As Table 1 shows, the zoom level zero is the most detailed, with the greatest number of elements used in the decision, and as the zoom level decreases, the number of elements used in the decision decreases. It can be said that the matching condition is relaxed.

Therefore, if there is no identical game situation, the game situation is checked to see if it matches at zoom level zero, and if it still does not match, the zoom level is lowered further to see if the game situation matches again.

Zoom level	factors
Zoom level 0	Player's next action Player's previous action Opponent's next action Opponent's previous action Player's Health Player's Stamina Opponent's Health Opponent's Stamina Player is Blocking(bool) Opponent is Blocking(bool)
Zoom level 0	Player's next action Player's previous action Opponent's next action Opponent's previous action Player is Blocking(bool) Opponent is Blocking(bool)

Table 1: The factors for each zoom level

2.3 AI Evaluation

Several papers have shown that the behavior created by ACAI does not differ significantly enough from human behavior to deceive humans [5][6]. In addition, our goal is to investigate what factors evaluators base their judgments on, so to focus more on this aspect, we conducted a questionnaire in the following way. Judges compare parts of two game videos to determine which video's agents can be said to be more human-controlled. By having the evaluators compare the two videos, it is supposed to increase the number of descriptions of AI-like and human-like behaviors, rather than describing their thoughts on a single video. For comparison with ACAI, we also test with a built-in AI controlled by a handcrafted finite state machine, which is implemented in the boxing arena. The built-in AI is also used as the opponent in all tests. The test is conducted on all combined sets of human play, ACAI, and built-in AI.

The questions ask the following information:

- About their fighting game experience.
- About video. (As mentioned above).
- About the reason for the decision.
- About the sureness of the decision.

3 Result and Discussion

Here we show the results of the survey by 30 evaluators. The results of the selection of human-like videos by the evaluators are shown in Table 2, which shows that the Built-in AI is not more human-like than the other two, but the ACAI is so human-like that it cannot be compared to actual human play.

Q. Which is more human-like?

Set of videos (A and B)	A (%)	B (%)
Human and Build-in AI	73.3	26.7
Human and ACAI	46.7	53.3
ACAI and Build-in AI	80	20

Table 2: Result of test

The main reasons given by the evaluators for their judgment are shown in Figure 4. People mainly consider the following factors to explain human-like behavior (in descending order of frequency): repeat certain actions, speed of response, accuracy of action, and result of the game. On the other hand, people consider the following factors to explain AI behavior (in descending order of frequency): various actions, inaccuracy of action, strategy throughout the game, and accuracy of action. This shows that there are differences in how people judge human-like and AI behavior. In addition, the main factors mentioned as reasons are 84 related to AI and 64 related to human-like. Thus, it can be seen that people think less about how human-like a behavior is when judging whether it is human-like and more about how much it is not caused by AI.

Furthermore, we discuss the differences in the reasons for each decision when the evaluator made the right decision and when the evaluator made the wrong decision. The factors frequently considered in each case are listed below.

When the evaluator makes the correct judgment on human behavior (Figure 5) (in descending order): various actions, strategy, inaccuracy of action, and accuracy of action.
In contrast, when the evaluator makes the incorrect decision, (in descending order): repeats certain actions and the result of the game.

Then, when the evaluator made the correct judgment on AI behavior (Figure 6), (in descending order): repeat certain actions, accuracy of action, speed of response, and result of the game.
In contrast, when the evaluator makes incorrect decisions, (in descending order): accuracy of action, repeat certain actions, strategy, and various actions.

From these results, we can see that the factors considered when people misperceived human behavior were different from those when they made correct decisions, but the factors considered when they misperceived AI behavior were the same as those when they made correct decisions. In other words, the focus of people's perceptions of AI behavior is more stable than for humans. In addition, the kinds of factors that can be reasons for judging a behavior to be human are more numerous than in the case of AI, while the factors that can be reasons for judging a behavior to be AI are some of the typical AI behaviors that people imagine, which can be different for each person.

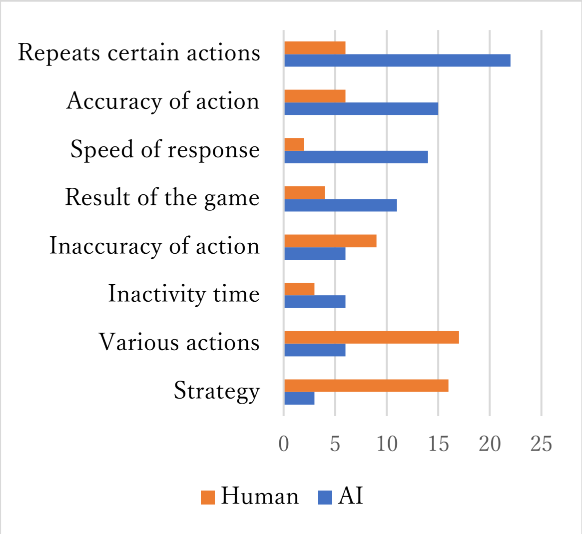

Figure 4: Main factors in judging each behavior.

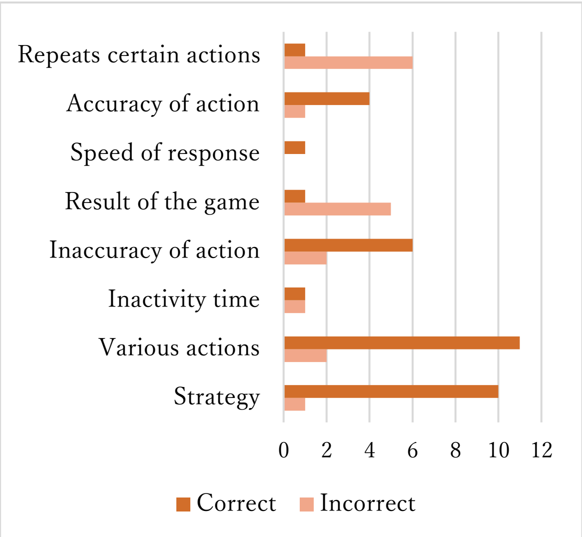

Figure 5: Main factor in judging human behavior correctly or incorrectly.

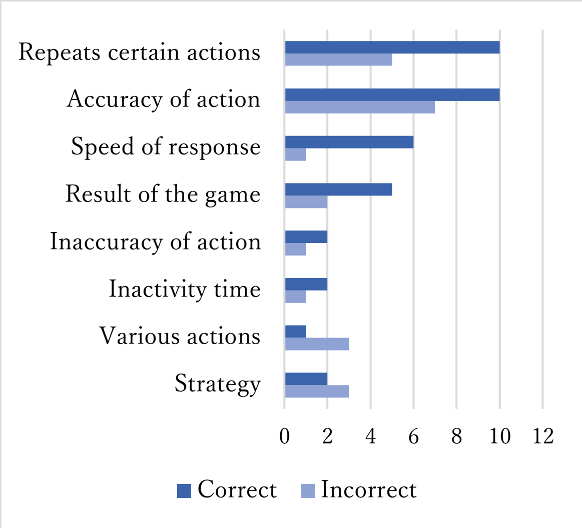

Figure 6: Main factors in judging AI behavior correctly or incorrectly

4 Conclusion

In this study, we introduced Artificial Contender in a boxing game and created a human-like AI using a behavior capture approach. By evaluating it, we also investigated how people make comparative judgments between behaviors by AI and those by humans. The results showed that there are differences in how human evaluators judge AI and human behavior. In addition, evaluators focused on similar factors when they judged behavior by AI incorrectly as they did when they judged it correctly. This indicates that people's focus when recognizing AI behavior is more stable than when recognizing human behavior. It also suggests that there may be potential biases in how evaluators perceive AI behavior. Understanding these biases may help reduce the chance that players perceive their opponents' behavior as artificial. We hope that this result will be of some help in the development of games that generate more human-like behavior and entertain players.

References

[1] Yuda, Kaori, et al. "Identification of play styles in Universal Fighting Engine." arXiv preprint arXiv:2108.03599 (2021).

[2] Turing, A. Computing machinery and intelligence. Mind, 59(236):433–460, 1950.

[3] Ciardo, Francesca, Davide De Tommaso, and Agnieszka Wykowska. "Human-like behavioral variability blurs the distinction between a human and a machine in a nonverbal Turing test." Science robotics 7.68 (2022): eabo1241.

[4] Zuniga, Evelyn, et al. "How humans perceive human-like behavior in video game navigation." CHI Conference on Human Factors in Computing Systems Extended Abstracts. 2022.

[5] I. Umarov, M. Mozgovoy. Creating Believable and Effective AI Agents for Games and Simulations: Reviews and Case Study. Contemporary Advancements in Information Technology Development in Dynamic Environments, IGI Global, 2014, pp. 33-57.

[6] Ryoya Ito, Developing Learnable AI From Human Player's Behavior in Universal Fighting Engine, The University of Aizu Graduation thesis, 2022.