A thesis submitted in partial satisfaction of the
requirements for the degree of
Master of Computer Science and Engineering
in the Graduate School of the
University of Aizu

" Playtesting and Believability Testing in Game AI Design"

by
m5191102
Akane Yamada

Reviewed by
Associate Professor Maxim Mozgovoy, Main Referee
Professor Vitaly Klyuev,
Senior Associate Professor Evgeny Pyshkin
March 2017

The thesis titled

"Playtesting and Believability Testing in Game AI Design"

by

m5191102
Akane Yamada

is approved:

Main referee
Associate Professor Maxim Mozgovoy

Referee 1

Professor Vitaly Klyuev

Referee 2

Senior Associate Professor Evgeny Pyshkin

Referee 3

University of Aizu
Spring 2017

Contents

Abstract

In this research, two types of tests were conducted in a commercial tennis application game. The first was a Turing test. The previous research of this game suggests that an AI character exhibits a similar playstyle to a human player who is a trainer for the AI character, which can be seen from automated comparison of the AI character and the human play style. However, it remains questionable whether the behavior of AI character really seems like human controlled. As a hypothesis it was suggested that AI- and human-controlled players may be perceived differently due to differences in timing of actions and elements of strategy. Present study conducted a Turing test to verify it and find elements that look like human controlled. Subjects watched two sets of replay video clips that contain matches of a character agent the same "AI coach" AI-controlled character. Subjects judged in which clips the character is controlled by a human. As a result of this test by nine subjects, more than half clips were correctly distinguished between human and AI controlled. It has suggested that merely similar playstyles do not seem to be human controlled. In addition, it seems that there are three elements that look like human controlled: movement, strategy, and mistake. Another test was playtesting, being conducted in order to find issues of game design and AI design. The states of subjects playing in the game with thinking aloud were recorded by two cameras. Thinking aloud is a usability test technique that manipulates evaluated objects while saying what they think. Subjects were given the task of winning one match in a mode exhibiting other users' ghost AI characters (characters trained on real user data). As a result out of eleven subjects, only two subjects were able to accomplish the task. The analysis of the observation data revealed that there are three problems: some subjects cannot learn the correct operation method from tutorial, the importance of indicating ongoing processes in the application, and often the coach AI behavior did not help user. This research on playtesting has presented the usefulness of playtesting and which problems should be addressed in games. It has suggested that both tests revealed issues of game AI design on evaluating games.

Acknowledgement

This work had not done without support by many people. First, my deepest appreciation goes to my supervisor, Professor Maxim Mozgovoy. He always encouraged and led me successful. I could not finish this work without his advices. Second, I would like to express my gratitude to Iskander Umarov who is a technical director of World of Tennis and all the members of World of Tennis team. My experiment could not materialize without this game which is the fruit of their labors. Third, I want to thank my test subjects. They worked hard at my experiments and gave me some feedback to improve World of Tennis. Finally, I would like to give special thanks to my family. They always took care of my health.

1. Introduction

1.1. Background

What is the role of game AI? The word “game AI” may refer to different concepts. In this paper, it refers not to board games such as Go and Shogi, but to digital games such as Role playing games and action games. Game AI has various roles. Dill said the purpose of Game AI is to support experience for player [1]. It greatly affects the quality of games. If enemy AI is too strong or too weak, do you have the will to continue that game? When walking along the way to your destination with an ally AI, if the AI runs ahead of you without considering the walking speed and actions you do, could you believe the AI?

Human likeness fulfills the believability of AI [2]. The human likeness AI has advantage of fun; Soni and Hingston proved that people enjoyed playing game with human-like AI based on a neural network than that of hand-coded [3]. In addition, human-like AI is necessary for simulation. Even simulations are diverse; there are cases of simulation at the time of disaster as well as military simulations. Biliss and Tidwell demonstrated the effectiveness of simulation, comparing three approaches: design drawing draft, Virtual Reality simulation and no training for improvement of rescue activities by firefighters [4]. Generally, disasters make so many people and things involved. Massaguer et al. experimented with multi agent simulation of disasters and suggested the importance of agents’ human-like behavior [5].

Several studies attempted to make human-like AI, evaluating the results [6][7]. Furthermore, there were some competitions for evaluating the AI; “2K BotPrize” [8] used a First person shooter game, and “Mario AI competition” [9] used Infinite Mario Bros, a clone of the classic platform game Super Mario Bros, Nintendo’s seminal platform game from 1985. Even through, every AI could not seem to behave like a real human player.

1.2. Objective

Previous research implemented self-learning AI for a mobile Tennis game [10], indicating experimental result that AI characters can exhibit

play styles resembling their human trainers [11]. However, this result still remains controversial of believability. There were two hypotheses why the result had lack of human-likeness; unnatural timing of actions and continuous non-adaptive behavior during each game run. Chapter 4 explained details of the hypotheses.

This present study puts its focus on believable game AI, conducting two experiments with a mobile tennis game. First, this study validated whether resembling human play styles exactly seems like human, exploring important elements for believable AI by using Turing test. Second, through playtesting it found issues of game design and AI design. In addition another short study has been done evaluating human-likeness of games which exists as sports in real; this case study gave new aspects.

1.3. Thesis structure

This paper is composed of six chapters, which were divided into three main parts. PART I (Chapter 2) is a review of existing literature around believability testing and playtesting. Part II (Chapters 3, 4 and 5) show detail of this research contents, at first explain testbed of this paper and formulate research questions of this paper. Moreover, in this part a result of two tests for game AI design is described. Part III (Chapter 6) presents discussion based on results, conclusion of this work and discussion of future work.

2. Literature review

This chapter provides the necessary prior knowledge to talk about the following chapters.

2.1. Believability

In the domain of digital games, believability means two side of a game character. One is character believability, "Someone believes that the character/bot itself is real, i.e. an actual living being (or actual autonomous robot etc.)". Another is player believability, "Someone believes that the player controlling the character/bot is real, i.e. that a human is playing as that character instead of the character being computer-controlled." [12]. In this study, player believability was focus on.

In order to verify this, Turing tests were actively conducted in the field of games. Turing test is an experimental method devised by Alan Turing and is performed to determine whether a machine has intelligence equally as a human being [13]. The machine is judged by human evaluators with having a chat session to a human and a computer program, then judge which is human. Several competitions to provide Turing test in the game were held.

2.1.1. 2kBotPrize

2kBotPrize is the first competition for accessing believability in computer games. Participants submit their agent to the server. Agents play a game called Unreal Tournament 2004. This is a first person shooter game. Evaluation method is observation tests. Observers also play the game with submitted agents. Then observers evaluate opponents to determine whether a game character is human controlled or computer controlled. This competition was held four times, the evaluation method was changed. Since 2012, submitted agents also can judge opponent is human controlled or not. Now this has two evaluation methods: one is judgement by only human observer, and another is by player characters included bots.

2.1.2. Mario AI Championship

Mario AI Championship is a competition, were the testbed is Infinite Mario Bros, a clone of Super Mario Bros. Shaker et al created the Turing Test Track within the Mario AI Championship to spur and benchmark development of believable bots [14]. Three bots and two human players competed and there are 60 observers to vote whether Mario playing was controlled by a computer or a human or not decided. Each agent was shown at least twice, and the order of the agents was randomized to prevent guesses. As the result, highest human-likeness was achieved by a human player, and he got more 5 votes as human than the 2nd prize. Togelius et al. suggested that given this preliminary experimental protocol it appears that the 3rd person assessment approach is appropriate since believability can be successfully assessed [12]. The architecture and method of one of the submitted bots were summarized by Shaker et al. [14].

2.2. Playtesting

For game developers, there are two methods of reincorporating game design with play feedback, QA and playtesting. QA stands for Quality assurance, being performed to find bugs in the game. Some of the bugs in the game are complicatedly intertwined and diverse involved things that happen when you repeat certain operations a number of times, or because of a certain parameter exceeds the upper limit. In addition, sometimes bugs are occurred once more. Therefore, QA is performed closer to the end of the development of the product. In contrast, play test is to actually play the game in order to confirm the game balance and make sure the direction of the game system is correct. Hence playtesting can be executed even early in development. Actually, the playtesting method performed by publishers who make many games refers to a method of usability test [15].

2.2.1. Usability Test

Usability is defined by ISO 9241-11(1998): ”Extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use.” Abran et al. suggested that Usability refers to a set of multiple concepts such as time of execution, performance, user satisfaction, and ease of learning, taken

together [16].

The usability test is a generic name of a method for evaluating usability when users participate in order to improve the evaluated target of higher usability. It can be roughly divided into three phases, planning, execution and reporting for usability test. In the planning phase: objectives and budget for testing is set, target audiences are recruited, and the task of experiment is considered. In the subsequent execution phase, experiments are conducted with users. In some cases, developers and others observe in a manner that does not interfere with the experiment. The state of the experiment is recorded for the analysis. Finally, during reporting phase, the results of the experiments are summarized and the problems for the evaluated items are pointed out [17]. There are various methods of usability testing, among them having test users to Think Aloud during usability testing is generally believed to be an effective and successful technique [18].

Think aloud is a method that users operate evaluating items meanwhile they express their thoughts, for instance, what to do next, why they thought try to do something, what feelings they have. Nielsen pointed out that thinking aloud allows people to access the cognitive processes and mental behavior, just as it gives people insight into thinking [19].

2.3. Necessity of playtesting in games

Usability tests are conducted on everything such as home appliances and websites, but there is one major difference between them and games. Pagulayan et al. asserted that to be pleasurable to play and sufficiently challenging in order to provide a good gaming experience are intended on testing for games. In addition, learning the goals, strategies and tactics to succeed in a game is part of the fun [20]. Thus, unlike other products, the purpose of usability testing in games is not just the product's efficiency and ease of use.

The play test is effective even if it is expressed in terms of business. It is almost impossible for games to change the contents dramatically after being released to the world once. However, as a result of being published to the world, there is a possibility that the users consider the game boring.

The profit loss at that time is immeasurable arise from bloated development scale of games. It is generally three years or more, more than 100 people on developing one game called AAA (which is an informal classification used for video games with very high development and marketing budgets). Microsoft released many AAA games, they have own laboratory called GUR (Game User Research) for game user research, and they actually succeeded in making the game better by doing playtesting based on the think aloud method [21].

3. World of Tennis

In this chapter, a testbed game for two tests and research question on previous research is introduced. The name of the game is “World of Tennis”, and it is a tennis game for iOS and Android devices.

3.1. Contents of World of Tennis

User can play singles tennis match in this game. Users can play matches in two modes: league match and training match. They play against other user's *ghost* in a league match, or they play against several coach AI characters in a training match. A ghost means an AI character trained by each user. When a match is over, a user is asked whether or not let his ghost learn his playing of the match. If the user chooses "Yes", the ghost will learn based on his play data in that match. The more ghost learns, the more he can imitate his trainer’s play. Users can upgrade their characters, covering 11 skill types (Power, Precision, Speed, Stamina, Serve power, Serve precision, Lob, Smash, Spin, Backhand skill, Volley).

There are two actions which character can do during match. Figure 1 shows image of match in World of Tennis. Users always operate the bottom character. One action is called "SetHitPoint". If the user touches some point on the opponent side of the court, the point becomes a shot target. The target is drawn as a yellow circle on the screen. If the user touches the same place twice, the next shot turns into a lob shot. Lob shot is a high height ball in order to pass over front player. The power and precision of the shot are calculated according to player skill and time range of touching. The faster a touch is performed, the better the next shot will be. Another action is called "SetMovePoint". If the user touches some point on his side of the court, the player character moves there. The point is drawn as a red circle on the screen.

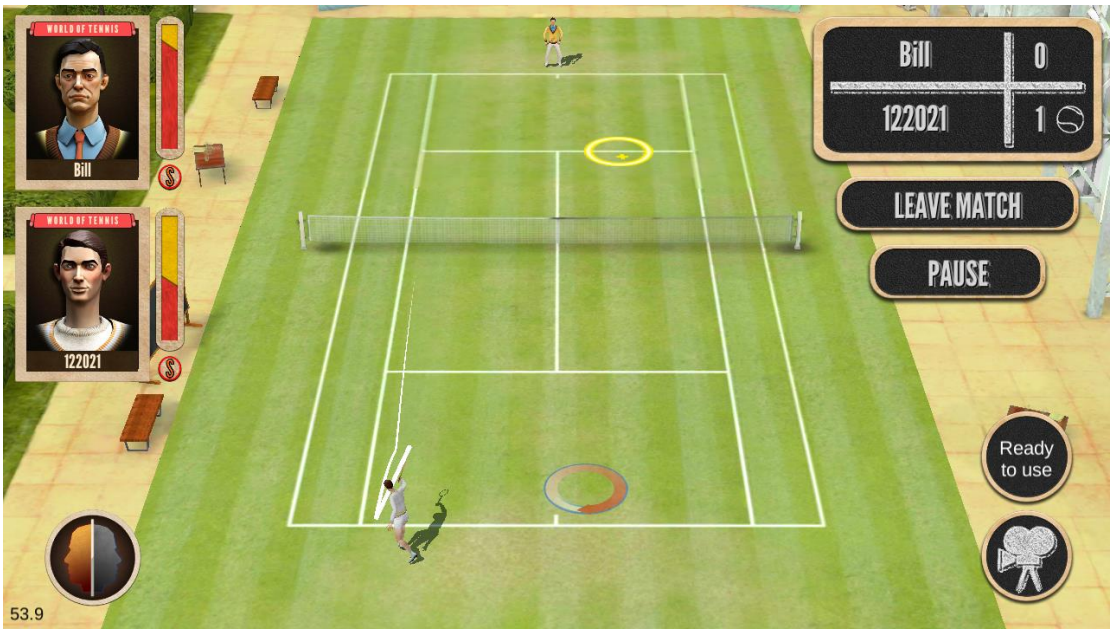

Figure 1. Screenshot of a match on World of Tennis

3.2. Research question

In previous research on World of Tennis, Mozgovoy et al. compared playstyles of total 16 characters consisting of 8 human players and 8 ghost AI characters of the corresponding players, indicating that ghost AI characters exhibit distinguishable play styles, similar to the styles of their human trainers [11]. However the subjective degree of human-likeness of these ghost characters still remained unaddressed. The question comes from two hypotheses: unnatural timing of actions and continuous non-adaptive behavior during each game run.

The action timing refers to the way ghost operate two possible action types that characters can perform during a match. Timing of decision making of a ghost AI was defined for two actions: "SetHitPoint" is right after the opponent has hit back and when the ball crosses over the net, "SetMovePoint" is right after the player's own shot. However, changing the game situation in every moment makes it possible that the opponent runs to the place where ghost tried to hit when crossing the net. If people encounter such a situation, it is believed that they might try to hit other points where they will benefit.

AI ghost learns the behavior of playing after the match as mentioned earlier. However, human always learn throughout the match. For instance, suppose a situation of a certain game had taken place as illustrated by

Figure 2. Both players were the position where was described the figure when the cross player set his point to place drawn in circle. The ball was returned by triangle player since he runs fast and he was able to get there before the ball came. If the same situation occurred as in the figure again, human player will hit the same circle place? It seems that human can learn that the ball will be hit back arise from the previous situation and people will hit another place not shown with a circle. Although ghost AI is not adaptive, so it is suggested that it is not entirely human-like.

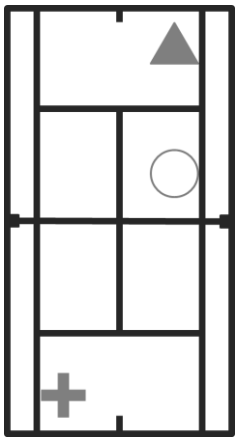

Figure 2. Example situation on a match

4. Method

For both Turing test and playtesting, the subjects were asked to answer a prior questionnaire before taking the test. The contents of the questionnaire are shown in Appendix 1 and were used for examining the background of the subjects. Both experiments were performed one by one in a closed room in order to keep contents of tests secret for the subjects who have not done them yet.

4.1. Turing test

The current investigation involved Turing test to validate whether resembling human play styles exactly seems like human, exploring important elements for believable AI.

4.1.1. Performance content

The basic procedure in this study was based on the process referenced in chapter 2. Experiments were conducted with two data sets for one subject. On the premise of carrying out this Turing test, subjects had played World of Tennis for at least 20 minutes. There was a laptop on a desk to watch video clips for test and there were writing tools to note during test; subjects sit down near the desk like shown in Figure 3. One file directory was displayed on the screen. There were two clips as one set of data inside one directory. Detail of clip contents are explained in the following section. Before starting Turing test, subjects got briefed as follows:

- The purpose of this test, which is to find out which elements of behavior seem like human.
- The structure of clips, that one clip shows a human player-controlled character, while another one shows an AI ghost-controlled character.
- Subject could repeat watching each video clip and note anything about any clips in anytime.

When a subject had no questions, he started watching clips. The subject filled a questionnaire after he decided which clip was controlled by a human player. The question items are “Which clip shows a human –

controlled character?" and "Why did you think so?" The former was a binary question, the later was free entry. When a subject provides the reason as a vague idea, he was asked more concretely like "What do you mean by it?" and he made to explain with his notes. In this case, the subject was told that this additional question was for extracting his thinking without missing out, and there is no relation with whether his answer is correct or not. Togelius et al. point out the problem that subjective mixing in the Turing test affects an answer whether a character is controlled by human or not and discuss ways to remove subjectivity as possible [22]. However, present research used Turing test to find elements that look like human playing. For that reason, it is essential to collect precisely what the subjects felt from what they watch, rejecting method to take away subjective elements.

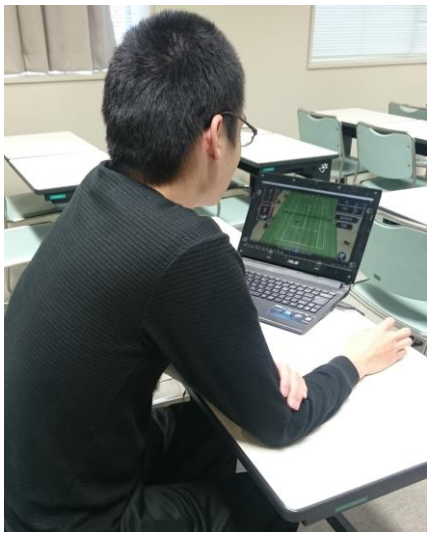

Figure 3. Actual experimental environment of Turing test

4.1.2. Data composition

A video clip is a record that shows a game between a character and a "coach AI" character. The coach exhibits a simple style: shooting to one of three areas (left service court, right service court and back court). He usually keeps his position near the end line side of the court. A data set consists of two 2 video clips. One is a replay of exhibition of a human player. Another is a replay of exhibition of a ghost AI of the human player. Ghost AI learns from human game record. Data sets were made by four different people. Each clip ends in 2-3 minutes. The total number of subjects for the Turing test was 9. All of them were men in their twenties.

4.2. Playtesting

The current investigation involved playtesting to found issues of game design including AI, which support game experiences for players.

4.2.1. Performance content

The basic procedure in this study was based on the process referenced in chapter 2. First, a subject was given explanations for the purpose of this playtesting which is to evaluate the performance of the game and confirmed consents to keep confidential and to take video recording during test. Second, the subject took a short lecture of think aloud by the interviewer. The interviewer described that "to do think aloud, you should say anything what you think about this game." showed demonstration of think aloud for subject with using Google Play Store application. In addition, to make nearly situation of playing game as in usual playing style even under surveillance, subject was advised that play this game with relaxing like ordinary you do in your home, and do not worry if you cannot play well since it is evaluating game that it is not evaluation of your playing. Third, subject was given a task which is to win at least one league match. The reason for having only one task was to confirm the quality of the tutorial. The tutorial starts when subject executes the game for the first time. During the tutorial, subjects learn how to move and hit on matches, how to operate the menu and how to upgrade the character. Because decomposing and giving many tasks makes it impossible to distinguish whether a subject wins the league match thanks to the tutorial. Typically, game users understand how to play games through tutorials. However, giving detailed tasks such as "hitting shots correctly", "playing offensive with being near the net", "strengthening a character" etc., required understanding of the basic game process as well. Therefore there was only one task that reveals the contribution of the tutorial in subjects' understanding. Finally, each subject played this game until he accomplished the task or spent over 30 minutes. The interviewer carefully observed a subject during test the, urged a subject to do think aloud when the subject gave a specific gesture, such as twisting his neck, raising a voice of surprise, suddenly stopping, and moving his face close to

the screen and so on.

Contents of World of Tennis in playtesting were as follows:

I. Input user name.
II. Tutorial for using the main menu (Change appearance of character and home ground, improve skill of character).
III. Tutorial for playing tennis game.
IV. Play a training match 1~3 times with the coach AI character*.
V. Play a league match*.

Subject can repeat *-marked actions during the test.

4.2.2. Experimental environment

The tools for playtesting were set up according to the environment based on previous researches projects in chapter 2. In this study, a tablet device was used because an evaluation object was a mobile game. One camera was used for recording subjects face and gesture; another camera was used for recording the tablet device, shooting from diagonally with using a tripod so as not to disturb play for a subject. Figure 4, 5, and 6 illustrate the situation when this test actually executed. Because there were no visitors, another display was not placed. A revision of a testbed game was selected that did not interfere with the operation for the test.

Figure 4. Actual experimental setup of playtesting

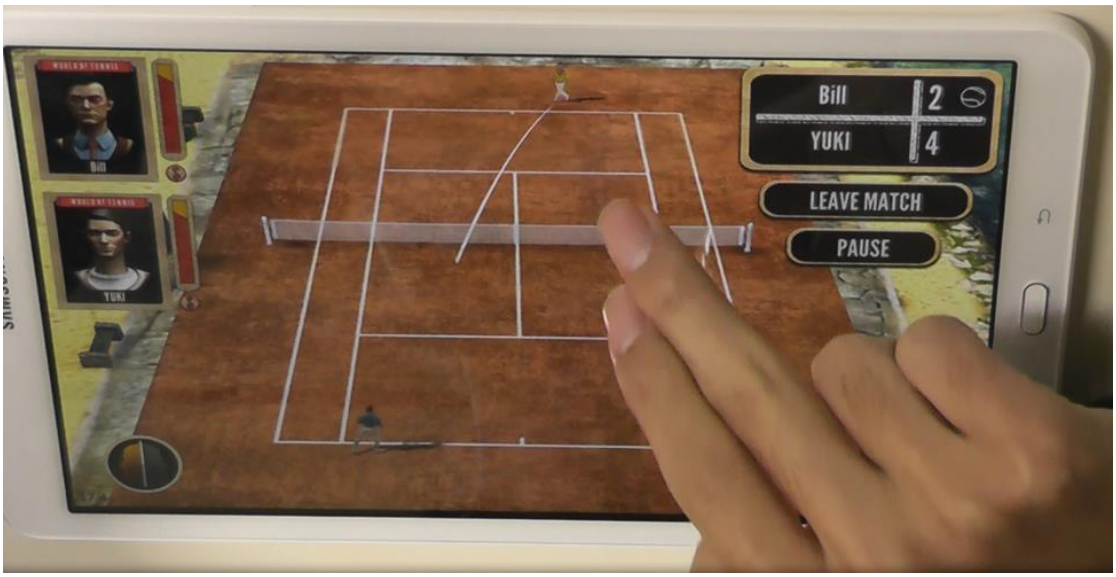

Figure 5. Camera view for tablet recording

Figure 6. Camera view for face and gesture recording

5. Result and Consideration

In this study, two experiments were conducted with the test bed game to confirm that designed AIs were supporting game users to have better gaming experience. The Turing test was executed in order to determine if that movement of AI looks human and to discover human-like elements. The playtesting was performed to verify that the user understands the contents of the game correctly as the user advances the game, and whether AIs helped it. The number of subjects was 11, and all were men in their twenties. Attributes of them were gathered using questionnaire of appendix 1.

Id	Skill of action games	Tennis Video Game experience	Time of game playing per a week	Real Tennis experience
1	Intermediate	Mobile and Console	20 hours	A class at junior high school or high school
2	Beginner	Console	7 hours	A class at junior high school or high school
3	Beginner	Nothing	20 hours	None
4	Intermediate	Console	22 hours	None
5	Intermediate	Console	20 hours	None
6	Intermediate	Mobile	4 hours	3 years and likes watching tennis on TV
7	Beginner	Nothing	1 hour	None
8	Beginner	Mobile	3 hours	None
9	Beginner	Nothing	No	A class at junior high school or high school
10	Intermediate	Console	10 hours	6 years
11	Gave up	Console	No	7 years and likes watching tennis on TV

Table 1. Subjects' background

5.1. Turing test

On this Turing test, subjects were showed two video clips and given a question which one was operated by human player. Both of the two clips were replay movies of a bout against a coach AI, one was that character was controlled by human player, another was that character was controlled by ghost AI which was learned by the human player's. It took about 15 minutes to get answer for two clips. This process was performed twice per one subject.

5.1.1. Result

This Turing test was conducted by nine subjects except the subjects No. 9 and No. 10 in Table 1. The result of the test is summarized on Table 2. The "Correct" means subject thought AI is playing and actually an AI was playing. The "Incorrect" means subject though AI is playing while actually a human was playing. The subjects could distinguish people and AI in over 70% of cases. The more ghost AI performs like human, the less distinguishable is the difference between two clips for subjects. It became clear that the ghost AI still could not be seen like human controlled.

Table 2. Final results of the Turing test (Unit: People)

	Case1	Case2	Case3	Case4	Total
Correct	3	4	3	3	13
Incorrect	2	1	1	1	5

The background information of subjects was scored as shown in Table 3. Table 4 shows that people who played real tennis or a tennis video game tend to get more correct answers than who don't have experience of tennis. However the information was insufficient to judge whether the tennis video game experience had the same effect on the accuracy rate as in the real tennis experience. Regarding tennis video game experience, the score is not determined by the duration of playing time, and if the subject played tennis video games even a little, it is counted as having experience. Although the minimal time period during which subjects could attend a school tennis club was one year. It seems that the question remains in the accuracy of the information which how many hours have subjects played

tennis arise from most of people do not mind how many hours they played in the daily gameplay. Judging of the basis of time, it is impossible to rank the subjects by skill. Furthermore, homogenized skills cannot be determined on account of any existing tennis games experience, since they have unique concepts and objects that can be operated.

Table 3. Score list on each attributes of individual subjects

Score	Skill of action games	Tennis video Game experience	Time of video game playing per week	Real Tennis experience
0	Gave up	Nothing	Less than 1 hour	Nothing
1	Beginner	Consumer	1 hour or more and less than 10 hours	Only taking class at any school
2	Intermediate	Mobile	10 hour or more and less than 20 hours	Less than 3 years
3	Advanced	Consumer and Mobile	20 hours or more	Over 3 years

Table 4. Relation between the number of correct answers and tennis experiences

Tennis Game experience	Real Tennis experience	Total experience score	First result	Second result
1	3	4	○	○
2	2	4	○	○
3	1	4	×	○
2	1	3	○	○
2	0	2	○	○
1	0	1	○	×
1	0	1	×	○
0	0	0	×	×
0	0	0	○	○

When classifying the answers obtained by the Turing test, it seems that elements that look like human playing were roughly divided into these three categories. They are movement, strategy and misses.

The first element is character movement. The hypothesis regarding

movements is related to timing, there were five subjects mentioned about it. Below are the excerpts of the comments related to movement as reasons for a character being judged as looking like human.

- The feeling of impatience was transmitted since he moved around here
- It was human-like to be not able to reach the ball at the last minutes
- The character movement looked like that of mine.
- The movement was briskly big. His move was wasteful.
- I felt his reaction was slow when the opponent shot.
- He developed a wiggle.
- Recovery movement is early.
- It was not catching up when opponent shot reverse side.

The second element is about strategy. The hypothesis regarding strategy is to change it by each rally along with opponent's reaction, it is suggested that one subject focus on it. Below are the excerpts of the comments related to strategy as reasons for a character being judged as looking like human.

- His tennis is very aggressive.
- He returned to the area where it was easy to take points.
- He went back to the middle which is the easiest to return.
- His strategy was stubborn.
- He is a man of principle because his movement is repetitive.
- He shot into an opposite side in order to make opponent run much.
- Following the previous results, he moved into center when he will shoot.
- His movement is routine. I think people tend to behave decided movement.
- He aimed an edge of service area because he got a point as service ace before.
- He moved in anticipation.
- He had predicted where the ball would come.
- There was no hesitancy in his movements.
- He returned into the place where opponent had missed.
- I also do something outrageous when I'm bored on playing tennis games
- He selected different strategy by each rally.

- It seemed like he had knowledge to shoot the place where the opponent was hard to get back.
- I felt it was reasonable movement returning to the original position after he shot.

The third element is about misses. It was indicated by a half of the subjects. Below are the excerpts of the comments related to ball misses as reasons for a character being judged as looking like human.

- Service miss.
- The character swings, but it was late.
- Overlooked the ball.
- He sometimes lost points because of net.
- Service precision was inconsistent.

5.1.2. Consideration

It was evident from Turing test that there were two ways to make agent human-like. One way is to add noise for timing of decision making and precision of serve. People tend to move late when something unexpected happens. For instance, a player thought his shot is sure to become point but the shot was returned. In such case, people forget to prepare the next shot and they often look hard for the next action. Another way is to make agents learn from data obtained during the current exhibition. Our ghost AI is not adaptive: it does not understand what kind of behavior leads to a victory or a defeat. If this feature is integrated, a ghost would be able to select behaviors that tend to get point against a general player. Moreover, to make provision for unpopular player strategy, our ghost AI should reflect each rally result since find out what strategy is effective for the current opponent.

5.2. Playtesting

In this study, subjects played a tennis game called World of Tennis on a tablet device. Subjects did “think aloud” that is action to speak what they think about the game during tests in order to collect their concerns, confusion and dissatisfaction etc. of this game. There were two cameras to record the experiment, one shooting the screen of the tablet device, and another shooting the facial expression and gestures of subjects. The target

task is to win one exhibition in a mode of fighting other users called league game.

In advance anticipation, all subjects could understand the game system and operation owing to information which was included in World of tennis, and most of subjects could accomplish the task. Subjects could play against characters of various strengths in the league. However, they had some gaps and weak points, so it is not difficult to win in the league if subjects play properly.

5.2.1. Result

Although the method was as described above, only one exceptional process had to be performed. It is originally suggested, even if the subject is doing the wrong operation, the interviewer does not point out to the subject. The interviewer has to find out a trigger which makes the subject do correct operation through watching for a while. However, one person tried the same operation which prevents the progress of the test more than three times, so it was pointed out as an exception.

Recorded data were used for analysis. At the beginning, recorded data by two cameras were played back with careful observation. The behavior was written down into a table, if the subject said something or he performed unnatural actions such like twisting his neck, touch the screen many times. Then related information i.e. what the situation and expected reason why the behavior was happened, what is the problem on this game, and improvement plan for the found problem were added on the table. Figure 7 lists one problem that occurred in one subject as a sample. Analyzed data of all subjects were summarized by problem identified in the Appendix 2. Three main problems revealed through analyzing recordings of playtesting.

Subject's ID	Situation	Problem	Speech or reaction	Expected reason or idea to improve
2	After result of league match	Subject did not notice the text [Return to my club] is labbel of button	"Return to my clab..." "Ah, this is that I have to tap"	Subject did not recognize it as button. Increase the size or make it more like a button

Figure 7. Sample of analyzed data on playtesting

First, the subjects did not learn the correct operation method of a playable character during the tutorial or a training match. They could not

understand how to control the character well because our tutorial was very simple and cut off some explanation. Then they could not see back how to control since the tutorial forced the subjects to go next.

Second, it is difficult to understand ongoing processes in the application. For example, most subjects tapped too much times “apply” button which is on menu of improving skills for character since they did not make sure the button was pressed. In order to clearly indicate that the button was pressed, it was necessary to play a sound effect, expand or contract the size of the button, issue a message indicating loading, and so on when pressed it. Another example, two subjects played a recorded replay of the game they did before in the league match. Even if you tap inside the court, the characters that appear on the screen do not move as expected because of it was a replay. However, they did not notice it and thought it was a real game, not a replay. It was not enough to change only button label. A label of button to start one league match was “Play Match” before playing the match. After they played, the label was changed “Watch Replay”. However, some subjects did not notice the change and they tapped display forward the replay.

Third, is about AI design during training matches. In training, if a certain condition occurs, ~~the game is forced to a pause, and~~ a tutorial message window containing advices appears on the screen. The behavior of the AI coach in training matches is an important factor because it may be included in that condition. The game is forced to a pause while the window is displayed, and in order to resume the game, it is necessary to press a skip button on the window or follow the advice. For example, a window which describe message "Now try to attack by running at the net!" appears when subjects return a ball which was a first shot by the opponent. This made subjects learn and practice the offensive strategy near the net. However, AI coach shot to the opposite side from the subjects, and it went near the end line of the court with five out of eleven subjects. In addition, the subjects had to tap earlier if their characters went forward since the distance between their playable character and a ball become shorter. Because of these reason, some subjects thought net play was difficult for them and gave up behaving the play style. The biggest problem of AI

design is that one coach AI hardly ever made certain movement to cause specific conditions to make a tutorial message window how to hit a lob shot appear. The lob shot is a high ball that crosses over the head of the opponent. The condition under which this tutorial message window appears was that the coach AI was near the net in the state where playable character was not near the net. However, coach AI did not run near the net himself due to his strategy is to keep his position to end of the court. Subjects had to repeat to return near the net in order to make coach run there. Subjects finished test without knowing about lob shot result form they tended to return balls to middle or end of the court.

5.2.2. Consideration

The tablet device was hardly ever moved by the subjects during this test caused the position of the camera had to be moved accordingly it. When some subjects were concentrating on the game, there were cases that they will tap vigorously by all means created tablet has moved even if subjects took a caution from interviewer that "You try not to move tablet as much as possible because it was recorded" in advance of test. It took a couple of seconds to move the correct position. If in the meantime the subjects had some sort of interesting action, they could have missed it. When the tablet moved, the interviewer could say "Stop playing", but there were behaviors that cannot be paused during the game, and it is suggested that the flow was cut off. Therefore, it seems that the tablet should be fixed with instruments such as antislip mat so that the tablet will not move during test.

If a subject did not speak for about 2 minutes or the interviewer could not understand the reason why subject took some reactions or a target of think aloud, the interviewer asked question like "What are you seeing?" and "What's wrong?" to the subject. There was a stop button on the screen during exhibitions, and subjects could pause the game by pressing it. Interviewer did not dare mention it in that interviewer wanted subjects to notice the existence of the button without the reference. However, especially during exhibitions, subjects tried to answer without stopping their hands caused two harmful effects. First, if subjects were focused on the progressive game, they gave an obscure answer on some

occasions. On the other hand, if subjects concentrate on answering questions, they lost sight of the progress of the game and responding immediately, it rarely drops points. It is suggested that interviewer should mention beforehand that subject can pause at any time during exhibitions and answer as clearly as possible when interviewer asked something to subject.

Contrary to expectations, the number of subject who clears the task was only 2 (the subjects 3 and 4 in Table 1). Other than contents of the testbed game, it seems that there were several reasons why the prediction which was that most subjects achieve task so far out. First, the time limit of 30 minutes might have been short. Most of the subjects got used to the operation of the game gradually, but the time when they become playing frequently was not enough. Some subjects were accustomed to operation at the time of a few matches; some subjects became familiar with only simple operations until the end of the experiment. Second, FPS (it is unit indicates the frequency rate at which an imaging device displays consecutive images called frames) of the tablet which was used on this test was between 30 and 40, However, in commercially available game machines, 60 FPS is common. Because of that, it seems that subjects were required to operate earlier and earlier, which might have increased their difficulty. Third, it could be inferred that subjects had to do think aloud in the mean time they played the game and that state was seen by an interviewer may give some influence on their playing.

6. Discussion and Conclusion

6.1. Turing test

As mentioned on chapter 3, previous research implemented self-learning AI, indicating experimental result that AI-controlled characters exhibit resembling play styles of their human trainers. This result, however, still remains controversial regarding believability. There were two hypotheses why the result had lack of human likeness: continuous non-adaptive behavior during each game run and unnatural timing of actions. This research performed Turing test in order to verify the AI ghost characters really looked human-like to human subjects, and to find elements that look like human playing.

As the result, subjects could distinguish people and AI in 70% of cases. It became clear that ghost AI still could not be seen like human controlled. Mainly, movement, strategy, and misses were seemed that elements that look like human playing. Similarity among these is that breadth of flexibility and behavior that can occur that is why it is a human being. It could be inferred that continuous learning and adaptation is also required for making human like behavior.

Although there are some studies on human-like behavior of game AI, the difference between them and this research lies in the feature of the game. This study used tennis games and tennis is a reality sports. To the knowledge of the author, Turing test of games on existing ones have not performed before. It was clear that people who have done that game up to now can tell them apart. Fujii et al. showed that playing experience of the evaluated game has some influence on the results of the Turing test [23]. In addition to this, this study revealed that the actual experience related to the evaluated game also affects the result.

Even if elements of movement, strategy, and error are put in the future to make a more human like AI, it is difficult to determine which of them should be incorporated them into program. Taking player mistakes, which were mentioned by half of the subjects as an example; if some mistakes are made intentionally, some users may be dissatisfied by looking as if the opponent deliberately loses. It is necessary to avoid causing unnatural

mistakes as much as possible in this case. It depends on the playing environment as well, i.e. we should consider hardware and network used for playing. In the case of this game, the hardware is portable; the mistake may occur when the opponent drops one's cell phone accidently. Even if a person makes mistakes at unnatural places, it may not be bothered much because there is no way for the others to know it. Moreover, unnaturalness will change depending on how big the chance of mistake is. In case of this game, if users aim for the end line or side line at the last minute, the possibility of "out" which means losing point on tennis become higher. The precision of shots relies on the timing of taps. The longer the time from tapping to hitting, the more accurate are the shots. Considering that, it seems that mistakes such as "ball out" look more natural.

6.2. Playtesting

Playtesting was conducted to find issues in game design and AI design. The results revealed that some subjects cannot learn the correct operation method of playable characters during the tutorial or a training match, the importance of indicating ongoing processes in the application, and the coach AI behavior did not always help the user experience.

As for the case where the some subjects cannot understand the operation method properly, it is believed that it is necessary not only to increase the amount of information, but also to explicitly indicate to the user that their operation is incorrect. As an example of problem actually occurring in the test, in order to hit the ball, subjects have to tap the screen, but some subjects swiped the screen. It is inferred that the reason why some subjects behave like this depended on the most famous tennis app game in the subjects' country because the method for hitting a ball is swipe in the game. Tap and swipe detection is programmable, and if it is detected within the only tutorial phase, it does not influence CPU loads during league match and training match. When a swipe operation is detected, if a message saying like "Please TAP the display NOT swipe" is appeared, the user will straightforwardly notice that that operation was wrong.

As the importance of indicating processes in the application, a small experiment was conducted when a minor game updated after the playtesting. The contents of the minor update had improvements to the

Apply button and the replay screen that pointed out the problem in chapter 5. The improvement of Apply button was simple; when the user taps this button, the color of a button turns from yellow into gray. After this improvement, subjects stopped tapping the button again and again. For the replay screen, when the user watches replay of league match, a text “Replay” is displayed in the bottom of the scene. A subject could notice that he not playing match but is watching the replay. It is conceivable that this playtesting was able to contribute to improve this game are proved from the results.

About the matter that some elements of the coach AI behavior did not help the user experience, although it was found that while some movements of the coach AI obstructs the user's learning, it should be noted that the problem it is not only to fix it as it is, but to do it with reasonably limited efforts. If a large scale change is required, programmers will have to spend more time to accomplish it. In addition, it is basically a task of a game designer to think about the purpose of the AI, and to confirm the work to be done. The communication cost can be reduced if only this can be done solely by the game designer. For example, with regard to the problems found this time, it can be tuned solely by the game designer because coach AI merely learns some extent repeatedly sometimes going near the net and hitting a low ball when the opponent comes near the net. This has benefited both programmers and game designers.

Nowadays, many people use smartphone on a daily basis, which extends the possibility of playing for people who have not played games until now. In addition, mobile games can be basically played free of charge. These things bring out that if the user feels like "bored", "hard to operate" or "I don’t make sense" to the game, they will uninstall it immediately. However, there is not much to our knowledge about papers on play tests related to mobile games. As a case study, it is believed that this study could contribute by presenting the usefulness of playtesting and what we mainly should pay attention to.

As a conclusion, both tests have demonstrated that they revealed issues of game AI design in the evaluated game. However, in mobile games it is impossible to increase the capabilities of AI imprudently due to the

limitations of hardware, network and data capacity. In addition, monetization of mobile games is typically made on the basis of item payment, and basically it can be free for playing. Because of this business issue, it is important to expand the frontiers of people who can play the game as much as possible. There are various specifications of the smart devices in the world, if we apply high-load processing for AI, most people will not even be playing. As a commercial mobile game, how to estimate and keep a balance between functional conditions and game design conditions is the future task. It is hoped that the outcome of the present study would be of some use to study of game AI.

7. References

[1] K. Dill, "What is game AI?," *Game AI Pro Collect. Wisdom Game AI Prof.*, p. 3, 2013.

[2] S. McGlinchey and D. Livingstone, "What believability testing can tell us," *Proc. Int. Conf. Comput. Games Artif. Intell. Des. Educ.*, no. January 2004, pp. 273–277, 2004.

[3] B. Soni and P. Hingston, "Bots trained to play like a human are more fun," *Proc. Int. Jt. Conf. Neural Networks*, no. 2008, pp. 363–369, 2008.

[4] J. P. Bliss, P. D. Tidwell, and M. A. Guest, "The effectiveness of virtual reality for administering spatial navigation training to firefighters," *Presence Teleoperators Virtual Environ.*, vol. 6, no. 1, pp. 73–86, 1997.

[5] D. Massaguer, V. Balasubramanian, S. Mehrotra, and N. Venkatasubramanian, "Multi-agent simulation of disaster response," *Atdm Work. Aamas 2006*, 2006.

[6] D. Livingstone, "Turing's test and believable AI in games," *Comput. Entertain.*, vol. 4, no. 1, p. 6, 2006.

[7] B. Gorman, C. Thurau, C. Bauckhage, and M. Humphrys, "Believability Testing and Bayesian Imitation in Interactive Computer Games," *From Anim. to Animat. 9*, vol. 1, pp. 655–666, 2006.

[8] "The 2K BotPrize." [Online]. Available: http://botprize.org/. [Accessed: 24-Jan-2017].

[9] "Mario AI competition." [Online]. Available: http://www.marioai.org/. [Accessed: 24-Jan-2017].

[10] A. Yamada and M. Mozgovoy, "Towards Self-Learning AI for the Videogame of Tennis," *Int. Work. Appl. Inf. Technol.*, pp. 79–80, 2015.

[11] M. Mozgovoy, M. Purgina, and I. Umarov, "Believable Self-Learning AI for World of Tennis," in *Computational Intelligence and Games, 2016 IEEE Symposium on*, 2016, pp. 247–253.

[12] J. Togelius, G. N. Yannakakis, S. Karakovskiy, and N. Shaker, "Assessing believability," in *Believable bots*, Springer, 2013, pp. 215–

230.

[13] A. M. Turing, “Computing machinery and intelligence,” *Mind*, vol. 59, no. 236, pp. 433–460, 1950.

[14] N. Shaker *et al.*, “The Turing test track of the 2012 Mario AI championship: entries and evaluation,” in *Computational Intelligence in Games (CIG), 2013 IEEE Conference on*, 2013, pp. 1–8.

[15] G. Amaya *et al.*, “Games User Research (GUR): Our Experience with and Evolution of Four Methods,” *Game usability: advice from the experts for advancing the player experience*. pp. 35–64, 2008.

[16] A. Abran, A. Khelifi, and W. Suryn, “Usability Meanings and Interpretations in ISO Standards,” pp. 323–336, 2003.

[17] C. M. Barnum, *Usability Testing and Research*, 1st ed. Needham Heights, MA, USA: Allyn & Bacon, Inc., 2001.

[18] B. Shneiderman and C. Plaisant, “Designing the User Interface: Strategies for Effective Human-Computer Interaction,” *Pearson*, vol. 5th, no. 2. 2010.

[19] J. Nielsen, T. Clemmensen, and C. Yssing, “Getting access to what goes on in people’s heads?: reflections on the think-aloud technique,” in *Proceedings of the second Nordic conference on Human-computer interaction*, 2002, pp. 101–110.

[20] R. J. Pagulayan, K. Keeker, D. Wixon, R. L. Romero, and T. Fuller, *User-centered design in games*. CRC Press Boca Raton, FL, 2002.

[21] R. J. Pagulayan, K. R. Steury, B. Fulton, and R. L. Romero, “Designing for fun: User-testing case studies,” in *Funology*, Springer, 2003, pp. 137–150.

[22] J. Togelius, G. N. Yannakakis, S. Karakovskiy, and N. Shaker, “Assessing believability,” in *Believable Bots: Can Computers Play Like People?*, 2012, pp. 215–230.

[23] N. Fujii, R. Fukushima, and H. Katayose, “Study of user control and experimental techniques in subjective evaluation for entertainment systems,” in *Entertainment Computing Symposium*, 2015, pp. 429–434.

8. Appendices

Appendix 1: Preliminary questionnaire on tests

Appendix 2: Analyzed data from recordings on playtesting

Appendix 1: Preliminary questionnaire on tests

Preliminary Questionnaire on Test

Thanks for your cooperation. Please honestly answer the following questions.

* Required

Please enter your name. *

_____

Q1. How is your proficient of playing action-game skill? *

Mark only one oval.

- ◯ Give up
- ◯ Beginner
- ◯ Intermediate
- ◯ Advanced

Q2. Have you played tennis game on smartphone or tablet devices? *

Mark only one oval.

- ◯ Yes
- ◯ No

Q3. Have you played tennis game on except smartphone or tablet devices? *

Mark only one oval.

- ◯ Yes
- ◯ No

Q4. If you remember the name of the game on who answered "Yes" with Q2 or Q3, please enter the title name of the game.

_____

Q5. How many hours do you play games total in a week? (There are no restrictions on hardware, including smartphone) I think that it will change depending on busy, so fill in the approximate average value. If you do not do it at all, please write "None". *

_____

Q6. Please check all items that apply to your tennis experience. *

Check all that apply.

- ☐ I have never played any real tennis.
- ☐ I have played tennis in a class at a school (but not including a university)
- ☐ I belonged to the tennis club on school when I was a student.
- ☐ I went to a tennis club specializing in tennis.
- ☐ I often watch tennis exhibition and enjoy it.
- ☐ Other: _____

Q7. For those who belonged to the tennis club, please fill in the duration of it.

_____

Appendix 2: Analyzed data from recordings on playtesting

Order	Situation	Problem	Count	Expected reason or idea to improve
1	Tutorial window which has the text "To control the game, tap onscreen objects..."	Subject did not know where is game icon	1	Only game logo icon was not highlighted
2	Character select	The appearance of the character will not change by only scroll	2	If the process is heavy we should write text "tap it".
3	Style select	Subject tap [Apply button] without changing [bottom]	8	In the sentences, both the top and bottom are written, but we do not specify where users change the bottom. After tapping on TOP's clothes, the cursor will go to Apply so user have not noticed [bottom]?
4	Camera Change	Subject do not understand the intention of changing the viewpoint	1	Since it does not feel that the point of view has changed much if it is only one time, if it is better to switch a lot
5	Shoot 5 times	Subject did swipe display not tap	4	We have to explain how to shoot FIRST. Subject could not read text on second tutorial window because the window was closed by swipe. Subject noticed that it was a tap instead of a swipe in the second tutorial window. Animation of cursor looks like swipe. It might be good if a circle ripple comes out.
6	Shoot 5 times	Subject did swipe display not tap. Then character shot there because the end edge of swipe was opponent side.	1	We have to explain how to shoot FIRST. Then subject could not read tutorial text because tutorial window was closed by swipe

7	Shoot 5 times	Subject did not make sense what should do	3	We have to explain how to shoot FIRST. It is not good for the user to know how to shoot until the ball comes to hand.
8	Shoot 5 times	Subject did not make sense how to shoot	1	Additional Controls Tutorial is needed. Subject misunderstood timing is important.
9	Shoot 5 times	Subject was surprised that the ball flies to circle not where he tapped right now.	3	We did not explain about shoot circle
10	Recovery 5 times	Subject tries to shoot too.	3	
11	Recovery 5 times	Subject was surprised because the screen got dark suddenly. He have not noticed that he have finished five times	2	He cannot see the check mark hidden by hand. Hope there is sound.
12	Recovery 5 times	Subject swipe display for recovery	2	Animation of cursor looks like swipe.
13	Recovery 5 times	Subject does not understand the intent of recovery	5	I wonder if the words are bad. It does not convey that it is moving to protect defense
14	Recovery 5 times	Tap red circle	2	It is bad that it is not written in the tutorial that the red circle is the destination.
15	Recovery 5 times	Subject did not know how to do recovery. To make matters worse, second tutorial window was closed because subject did swipe for recovery.	1	He advanced without reading the explanation. In order to prevent erroneous taps, I wonder if there is a next button during the tutorial
16	Recovery 5 times	Subject misunderstood that he needs to tap for moving to get closer to the returning ball	1	We do not explain auto adjustment in the tutorial!
17	Recovery 5 times	Subject did not know if he could recover well so he want to do it again but he can not	1	Rather than automatically transitioning to the next after 5 times, it is better to display a window like "Do you want to proceed to the next step?" and user can select weather go next or not.

18	Serve 5 times	Tap many places outside the service area	5	We have to explain that user have to tap suitable area of serve. In the case of most famous Japanese tennis game, user can tap anyplace for serve.
19	Serve 5 times	Subject did hold display. Mini game was hidden by his hand.	1	I think that it is necessary to change sentences like this; Once you tap the area, a serve gauge comes out. When the icon comes in a round area, tap again to serve.
20	Serve 5 times	Subject have not noticed that I did 5 times	1	SE or some popup text is needed.
21	Serve 5 times	Subject is not sure he could serve properly	1	Rather than automatically transitioning to the next after 5 times, it is better to display a window like "Do you want to proceed to the next step?" and user can select weather go next or not.
22	Serve 5 times	The coach did serve after subject did serve five times	3	
23	Tiebreak	Subject did not know about tiebreaks	2	If you can write a tie break (seven points preemption)
24	Tiebreak	Subject slipped backwards	1	Only this sentence has a very large font
25	Tutorial match	Subject was not aware that the game started	1	We need text like let's start the game
26	Tutorial match	The coach is weak towards the front	1	Rather, the back cover rate is too high!
27	Tutorial match	[Leave match button] was pressed	1	I feel polite when user press the [LEAVE MATCH button], Message window which is like league match is displayed during the tutorial game.
28	Improve Skill	Tap [Apply button] many times	1	Subject might think that it will disappear if he press [Apply button] ...?
29	Improve Skill	The skill change is fixed [accuracy] and subjects cannot choose freely	1	

30	Club view	Tap the coat many times.	1	Subject did not understand or forgot the explanation of tap menu or object
31	Club view	Hit judgment of each tab on [menu bar] is small	3	Extend the area of collider downward
32	Training match	When volley tutorial window is appeared, subject cannot approach the net as the ball hits behind	6	When the volley tutorial is happened, let coach shoot near the net.
33	League match	Subject has not noticed that his operation method is wrong. Subject taps his character tighter when he wants to shoot.	1	The tutorial that goes on without understanding is bad. After each step, the window likes "Did you understand this step properly?" "Yes" "No" may be enough to display.
34	League match(Replay)	Subject has not noticed that he was watching a replay even when it starts.	3	
35	general match	Subject was surprised that a serve was arbitrarily struck	6	We have not explained that a serve has time limit
36	general match	Subject was surprised that he can move left and right before he starts serve	2	We do not tell it on tutorial
37	general match	For serve, Subject noticed that he can tap after the slider rises to the top	1	We have to explain it or show how it works on tutorial
38	general match	Subject want to go to the next screen and press [OK Button] many times, but it takes much time	3	It would be better to display "please wait" again after pressing the [OK button]
39	general match	Subject misunderstood that he have to hit according to the timing	4	We did not explain relation of shot circle and accuracy on tutorial. We should tell it on basic tutorial because user shoots according to the timing in the case of most famous Japanese mobile tennis game.

40	general match	Subject should have tapped but character could not hit it	1	It might be happened because of delay. We should tell user have to do SetHitpoint (or if the circle is out) before the ball arrives. And also we have to explain shoot circle and precision.
41	general match	Subject did not notice that he got points	2	I think that it is hard to understand it that texts of [Opponent Serve] and [Score Bill] are exactly the same character decoration.
42	general match	Character did not run when the character is serve-receiving	1	Is auto adjustment fine if the distance is far? Or because it was a loose ball?
43	general match	Even though the player character is in the position of the ball, it gestures because the hit determination and the timing at which the shot circle disappears become the same	5	This was happened when subjects tap many times for shoot. It is polite to give message for notice this thing
44	general match	Subject did not understand about auto adjustment	1	We did not explain about auto adjustment in tutorial. We should tell user that character run automatically toward comming ball.
45	general match	Subject did not know he cannot miss to serve because fault becomes loosing point	3	Fixed
46	general match	Subject took a power shot button after taking a point or losing a point	1	We did not explain power shot button in tutorial
47	general match	Subject does not understand difference between orange circle and yellow circle	1	We have to explain about rob shot on tutorial
48	general match	Subject did not know whether there is the concept of deuce	1	We should not write 6 to 6 but DEUCE on score UI
49	general match	Subject did not watch the gauge of serve	1	I think he did not understand why the gauge here is.
50	general match	Subject thought that the ball would enter by tapping in the court. But it went out	2	We did not explain the shoot will go anywhere in the circle on tutorial! And if it was out, some SE is needed, I think.

51	general match	Understanding did not catch up as the tap which was meant to return ball behaved SKIP	2	If the ball becomes net, some SE is needed.
52	general match	Subject did not notice that now is match point	1	Need some SE!
53	general match	Subject think he cannot do rob shot	2	We have to explain about rob shot on tutorial
54	general match	Subject did not know about side-out	1	We should explain for beginner of tennis
55	general match	He have not noticed that it is a opponent serve, the reaction is delayed and he could not return	1	It is bad that opponent hit it suddenly, I want opponent to wait ...
56	general match	He do not know exactly what the stamina means	1	There are no explaining for stamina on tutorial
57	League board	Subject want to play league game but he pressed [Watch Replay button]	2	Subject did not notice that the label of button has changed
58	League board	Tap the club picture to start	1	I feel that subject think that anyone can play against. Why should you add label of number with listed items?
59	League board	Subject wants to freely choose the order to fight	1	
60	League board	Subject does not know what the number next to the name (like NAME-XXX) means	1	
61	After result of league match	Subject did not notice the text [Return to my club] is label of button	2	You can increase the size or make it more like a button
62	Result of match	Subject did not make sense wheatear the [OK button] was pressed. Then he tap the button many times	5	It would be better to display "please wait" again after pressing the [OK button]
63	Skill Menu	Subject tap [Apply button] many times because he was not sure the change is applied	7	Fixed
64	Skill Menu	Subject was surprised that items of	1	If user can scroll items, we should display

		skills are below		scroll bar
65	Training menu	Subject did not have make sense what is in basic training	1	When there is a cursor in Basic Training, there is no message (no matter how you slide it and see other coaches ...)
66	Equipment	He don't make sense what is difference each rackets	1	User do not know that question Mark is an explanation
67	Others - tutorial	Tutorial was ended without reading the last text because subject tapped continuously	1	It is good to prevent continuous tap
68	Others - tutorial	Subject want to read log text	2	If user can switch back, we don't have to prevent continuous tap