

チューリングテストによるゲームAIの客観的評価

安武 諒* 山口崇志**
マッキンケネス ジェームス** 永井保夫**

ゲームAIとはビデオゲームにおける人工知能技術のことであり、一般的にはプレイヤーに対して、知性を感じさせる存在のことを指す。ゲームAIでは、ゲームの挙動が違和感ないように見せることが要求される。特に、コンピューターが操作するキャラクターの制御を行うキャラクターAIは、ゲーム内におけるキャラクターの人間らしい思考・動作を表現するものであり、プレイヤーに直接認識される部分である。そのため、より知性の表現が重要となり、ゲームAIの人間らしさを担っている。

本研究は、ゲーム内におけるキャラクターの人間らしい思考・動作を表現することで、ゲームAIのリアリティを追求することを目標としている。ここで、ゲームAIの人間らしさを、AIに対しプレイヤーが人間であると感じる（錯覚する）ことであるとする。しかし、この人間らしさとは曖昧なものであり、その評価は個人の主観的体験によって判断される。そのため、客観的な評価に基づく具体的な判断基準が確立されていない。本研究では、ゲームAIの人間らしさを、客観的に評価する手法の確立を目的とし、ビデオゲームへのチューリングテスト適用方法の検討を行った。

実験には、FSM (Finite State Machine: 有限状態機械) を用いて作成したAIを使用した。これは、ゲーム制作において一般的に適用されているためである。また、AIは比較評価のために状態数の異なる2種類のFSMを作成した。一方は、相手を発見したら追跡するだけの単純なものである。もう一方は、状態数を増やし協調性を考慮して設計した。この2つを比較したとき、協調性を考慮したAIの方が単純追跡AIよりも状態数が多い複雑な動きをするため、評価が高くなると想定される。テスト結果において、協調性を考慮したAIの方が高い評価値になったならば、ゲームAIの比較評価が正しくできたと言え、チューリングテスト適用の有効性が認められる。

結果として、チューリングテストは、人間の主観に基づく評価を定量的に表わすことができ、正しく比較評価することができた。このことから、ビデオゲームへのチューリングテスト適用における、客観的評価の有効性が確認できた。しかしながら、AI同士の比較評価では相対評価に留まってしまうことがわかった。そこで、人間同士によるチューリングテストを行い、人間のゲーム操作から、被験者が感じる人間らしさを表す評価尺度を設定した。その結果、評価値は中央に収束することがわかり、中央値をゲームAIの人間らしさにおける評価の基準値として設定した。これにより、相対評価に留まっていたゲームAIの評価を、絶対評価で行うことができた。

キーワード：ゲームAI、チューリングテスト、客観的評価、有限状態機械、協調性

*東京情報大学 大学院 総合情報学研究科
Tokyo University of Information Sciences, Graduate School of Informatics (graduation in March, 2012)
2012年4月よりアイコムシステック株式会社所属

**東京情報大学 総合情報学部 情報システム学科
Tokyo University of Information Sciences, Faculty of Informatics, Department of Information Systems

Using the Turing Test for Objective Evaluation of Game AI

Ryo YASUTAKE *, Takashi YAMAGUCHI **,
Kenneth James MACKIN ** and Yasuo NAGAI **

Game AI is artificial intelligence technology applied to video games, and in broad terms refers to the presence of intelligence a human player feels in a video game. One requirement of Game AI is that the game behavior is natural and realistic. In particular, character AI, which is responsible for the human-like decision and behavior of the computer controlled character in the game, is directly recognized by the human player. Therefore, realizing an intelligent character AI becomes an important factor in creating a realistic and human-like Game AI.

The goal of this research is to pursue the reality of Game AI by realizing a character AI that makes human-like decisions and behavior. Here, the “human-likeness” of Game AI is defined to be the likelihood that a human player feels, or is eluded, that the controller of the character is human. However, this “human-likeness” is ambiguous and subjective evaluation is judged by individual experience. Hence, a concrete criterion based on objective evaluation has not yet been established. In this research, a method to objectively evaluate the “human-likeness” of Game AI by applying the Turing Test to video games is proposed.

In this paper, Finite State Machine (FSM) was used to create the tested AI. FSM is commonly used to model AI used in action video games. 2 types of the FSM models with different number of states were created for evaluation. The first model is a simple model only taking into account the location of the opponent. The second model has increased number of states and takes into account the location of friendly players as well, in order to realize cooperative behavior. The expectation is that the second, more complicated FSM model should get higher “human-likeness” marks. If the proposed method to objectively evaluate the “human-likeness” of Game AI can correctly rank the 2 different models, then the validity of the proposed method of applying the Turing Test to Game AI can be shown.

The experiment results showed that the proposed Turing Test method can be used to quantitatively compare evaluations based on subjective human evaluation. From this result, the validity of applying the Turing Test to video games for objective evaluation was confirmed. However, the quantitative results of compared Game AI could only be used for relative evaluation, since there is no absolute evaluation measure. Therefore, the same Turing Test was performed using human players only, in order to find the absolute evaluation measure for a true “human-like” character. The results showed that for true “human-likeness”, the evaluation results converged in the center of the proposed reference scale. From this result, the proposed method was shown to be useful for both relative and absolute evaluation of Game AI.

Keywords: game AI, turing test, objective evaluation, finite state machine, cooperation

1. はじめに

現在、人間型のロボットの研究が進み、人間らしいロボットはどうあるべきかが重要な課題となっている。人間そっくりのアンドロイドを作って、人間と比較するチューリングテストを、トータルチューリングテストと呼ぶ[1][13]。このような動きから、チューリングテストを見直すことで、知能について考え直し理解を深める機会になっている[2]。

従来の大学におけるAI研究は、ゲーム機のような限定的なコンピューター環境を考慮して開発しておらず、「プレイステーション2」世代以前のハードウェアにおいて、最新のAI技術が導入されることはほとんどなかった。ゲームを実世界のシミュレーターとして扱うことで、学問として発展し、理論の蓄積をしてきたAI技術を実社会へ適用するシミュレーションが可能となる。以上のことから、ビデオゲームにおけるAI技術（ゲームAI）のリアリティを追求することで、ゲーム業界のみに留まらず、産学協同によるAI技術の相互発展が期待できる。

近年、ゲームAIへの関心が高まり重要視されている[3][4][18][19]。ゲームAIが未熟だと、ゲーム内のキャラクター動作や環境が不自然になり、プレイヤーは非常に不快感を覚えることになる。プレイヤーによるゲームAIのリアリティに対する要望から、ゲームAIのリアリティを追求することでビデオゲームの発展とおもしろさを向上させることが望まれている[3][10]。

本研究は、ゲーム内におけるキャラクターの人間らしい思考・動作を表現することで、ゲームAIのリアリティを追求することが目標である。ゲームAIの要件として、プレイヤーがゲームプレイ体験を通じ、知性を持っている、自然なものであると感じる存在でなければならない[4]。すなわち、ゲームAIとはプレイヤーを錯覚させるシステムである。このことから、本研究におけるゲームAIの人間らしさを、AIが制

御するキャラクターを人間が操作していると、プレイヤーが錯覚することであると定義する。

しかし、この人間らしさとは曖昧なものであり、その評価は個人の主観的体験によって判断される。そのため、客観的な評価に基づく具体的な判断基準が確立されていない。本研究では、ゲームAIの人間らしさを、客観的に評価する手法の確立を目的とする[14][15][16][17]。本論文では、その提案としてビデオゲームへのチューリングテスト適用方法について考える。

2. チューリングテスト

チューリングテストとは、ある機械に知能があるかどうか（人工知能であるかどうか）を判定するためのテストである[5][11][12]。その概要は、以下のとおりである：2台のディスプレイの前にテストをする人がいる。1台のディスプレイには隠れている別の人が、もう1台には人間を真似るように作られたコンピューターが受け答えした結果がそれぞれ表示される。テストを行う人はどんな質問をしてもよいとする。例えば、詩を作らせる、音楽の感想を聞くなどが挙げられる。また、コンピューターも人間を真似る努力をする。例えば、わざと計算に時間をかけたり、間違えたりする。こうしてテストをする人が、人間とコンピューターの明確な区別がつかなければ、このコンピューターは知能的であると判断するというものである。

ゲームAIの人間らしさにおける定義から、その評価はプレイヤーの主観に基づき判断される。つまり、ゲームAIの人間らしさにおける評価は、主観的な体験によって定義されているため、客観的な評価指標が確立されていない。本論文では、人間の主観的判断によるチューリングテストを対戦型アクションゲームに適用し、複数の被験者のテスト結果に対して統計データを求めることで、ゲームAIの人間らしさを客観的に評価することを提案する。

ゲームAIでは、AIが制御するキャラクター

を人間が操作していると錯覚させることが要件として挙げられる[4]。すなわち、人間が操作するキャラクターとAIが操作するキャラクターとの区別がつかなければ作成したAIは人間らしいと判断できるものとする。これをゲームAIにおけるチューリングテストの定義とする。チューリングテストが質問に対するコンピューターの答えを、人間の答えであると錯覚させることだと考えると、ゲームAIの人間らしさの評価に、チューリングテストを適用することの有効性が見出せると考える。

3. 実験環境

3.1 対戦型アクションゲーム

本研究では、対戦型アクションゲームにおける1ジャンルである、Third Person Shooting Game (TPS) を実験環境として取り扱う。TPSとは、三人称の視点で、ゲーム内のキャラクターを操作するアクションシューティングゲーム (STG、欧米ではshooterという表記が一般的) のことである。

本研究において、TPSを取り扱う理由は、アクションゲームがコンシューマーゲーム業界において、シェアの大半を占めており[6]、複数のプレイヤーで対戦されるビデオゲームをスポーツとして捉える electronic sports (eスポーツ) [7][8] 競技として確立しており、多くの競技人口を有しているからである。STGは多くのものが複数のキャラクター同士による団体戦である。ここでは、自分と味方、敵対する相手が必要となり、味方と協力しながら敵対相手と競い勝利することで達成感が得られる。このような、対人戦を想定しているゲームにおいて、味方や対戦相手のキャラクターを制御するAIはより現実性が求められる。そのため、対戦型アクションゲーム、特にTPSにおいてゲームAIが重要視される。

TPSが持つ機能は大きく索敵、移動、攻撃、回避の4つに分けられると考える[18][19][22]。本研究では、特にこの中の移動に着目し、キャ

ラクターがフィールド内を移動する部分を扱い、AIは追跡行動 (また、それに付随する行動) のみを取り扱う。

3.2 実験用ゲームの概要

実験環境として、対戦型アクションゲームを簡易モデル化した1対2の追跡ゲームを利用した。この追跡ゲームでは、壁が存在する限られたフィールド (縦:1,000ピクセル、横:1,200ピクセル) 内において、逃亡側1体と追跡側2体によるキャラクターが表現されている。終了条件は逃亡側が追跡側に捕まるか、制限時間30秒を経過した場合とした。検証では、被験者に逃亡側のキャラクターを操作してもらいチューリングテストを行った。追跡ゲームは、アクションゲームの基本機能である移動に着目し、キャラクターがフィールド内を移動する部分のみを取り扱った。

実験用ゲームは、PC上で実行し、図1はゲー

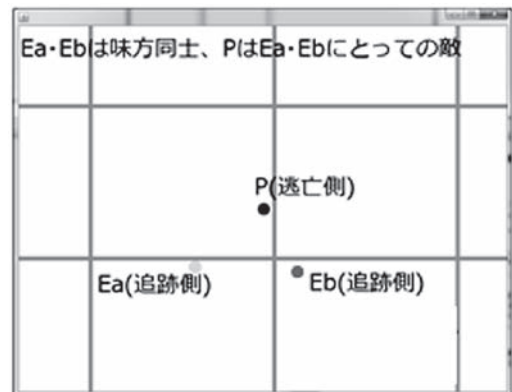


図1 実験用ゲーム画面



図2 PlayStation 2用コントローラー

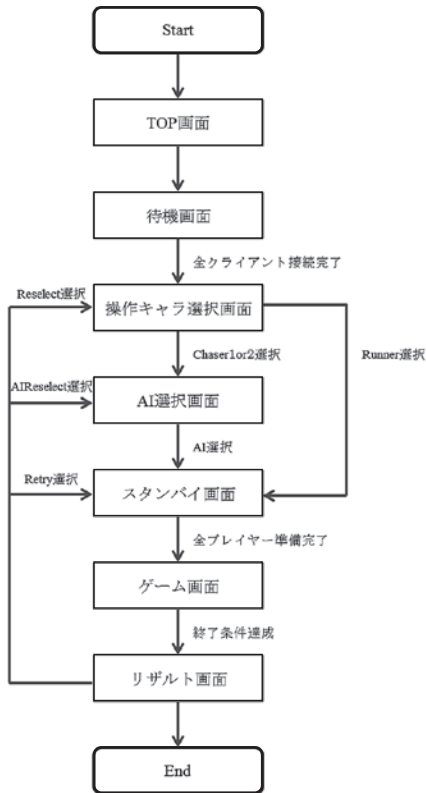


図3 実験用ゲームの画面遷移図

ム実行画面を示す。画面上のキャラクターPがプレイヤーによって操作される逃亡側キャラクターを、キャラクターEa・Ebが被験者以外の人間によって制御される追跡側キャラクターを示している。操作画面は常にキャラクターを中心に表示する。キャラクターの操作は、図2のようなPlayStation2 (PS2) のコントローラーを用いる。入力は、アナログパッド(矢印で示す部分)を用いて、上下左右及び斜めの8方向で行う。

図3は、ゲーム起動から終了までの画面遷移を表した図である。また、実験用ゲームは逃亡側1体、追跡側2体の複数プレイヤーで行われるため、サーバを通じた通信を行う。図4はクライアント・サーバ間の通信モデル図である。

4. 実験用AI

実験用に用いたAIは、ゲーム制作に一般的に適用されている有限状態機械(Finite State Machine: FSM)を用いて表現した。適用理由としては、状態の遷移を視覚的に確認でき、システムの挙動を追うことができるのでゲームの動作が分かりやすく、設計しやすくなる点が挙げられる。

4.1 オートマトン(有限状態機械: FSM)

オートマトンとは、離散的な入力と出力を持ち、自動的に動作する機械の数学的なモデルである[9][20][21]。このような機械は、内部にいくつかの状態を持ち、その状態を変化させながら動作する。オートマトンでは、入力、内部状態、出力が重要な要素となっている。状態が有限個であるオートマトンを有限オートマトンという。有限オートマトンを有限状態機械とも呼び、有限個の状態と遷移と動作の組み合わせからなる数学的に抽象化された「ふるまいのモデル」である。デジタル回路やプログラムの設計で使われる[9][20][21]。

有限オートマトンには、決定性有限オートマトン、非決定性オートマトンなどがある。ここでは、決定性有限オートマトンについて説明す

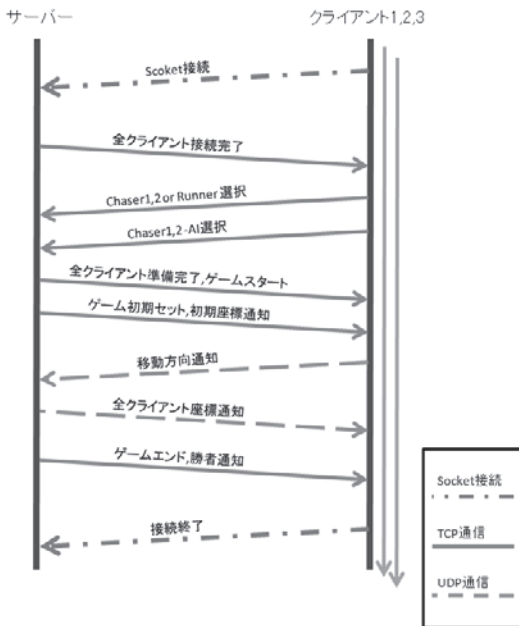


図4 実験用ゲームのクライアント・サーバ通信モデル図

る。決定性有限オートマトンMは、以下の5個組で定義される。

$$M = \langle Q, \Sigma, \delta, q_0, F \rangle$$

ここで、

Q : 状態の有限集合

Σ : 入力記号の有限集合

δ : 動作関数 ($\delta : Q \times \Sigma \rightarrow Q$)

q_0 : 初期状態 ($q_0 \in Q$)

F : 受理状態の有限集合 ($F \subseteq Q$)

である。

たとえば、次のような決定性有限オートマトン(有限状態機械)をMIを考える。

$$MI = \langle Q, \Sigma, \delta, q_0, F \rangle$$

$$Q = \{r, s\}$$

$$\Sigma = \{0, 1\}$$

$$\delta : Q \times \Sigma \rightarrow Q$$

$$\delta(r, 1) = r, \delta(r, 0) = s$$

$$\delta(s, 1) = s, \delta(s, 0) = r$$

$$q_0 = r$$

$$F = \{s\}$$

決定性有限オートマトンMの動作をわかりやすくするために、状態遷移図と呼ばれるラベル付きの有向グラフで表現できる。決定性有限オートマトンの状態遷移図は、以下のように構成される。

- ・各状態に対応して、頂点を作成し、状態の名前を頂点のラベルとする。
- ・初期状態に対応する頂点には、二重矢印⇒を付ける。
- ・受理状態に対応する頂点は、◎で囲う。
- ・各 $\delta(q, a) = p$ となる動作(状態遷移)に対応して、 q とラベル付けされた頂点へ向かう辺を作り、その辺のラベルを a とする。

前述の決定性有限オートマトンMIを状態遷移図で表現したものが図5である。

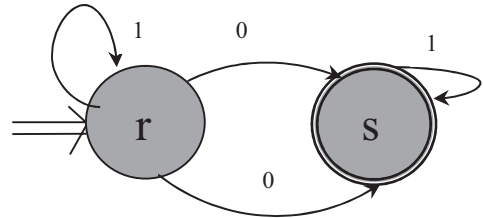


図5 状態遷移図

ゲームAIは、一般的に決定論的と非決定論的に分けられる。決定論的AIでは、その振る舞いまたは性能は、あらかじめ規定され、予測可能である。その一例は、単純な追跡アルゴリズムが該当する。決定論的AIは、ゲームAIの基本となり、その手法は予測可能で、高速で、実装、テスト、デバッグが簡単におこなえることが特徴である。これに対して、非決定論的AIでは、その振る舞いにある程度不確定要素があり、予測不可能な部分がある。そうした一例には、学習機能を利用したアルゴリズムが該当する。非決定論的AIは、学習や予測不可能なゲームプレイを可能にし、開発者はすべてのシナリオを予測して、すべての振る舞いを明示的にコーディングする必要がないという特徴を持つ[22]。

ゲームAIの自然な表現として、以下を挙げることができる。

- ・拡張性がある
- ・カスタマイズ可能である
- ・多様性を内包している

本研究では、決定論的AIを対象として、決定性有限オートマトン(有限状態機械:FSM)による表現を用いて、拡張規模が大きくなった時でも、状態を増やしていくことで対応でき、拡張性をもったAIを作成していく。

4.2 実験用AIの詳細

実験用に用いたAIは、ゲーム制作に一般的に適用されているFSMを用いて作成した。今

回、比較評価のために状態数の異なる2種類のFSMを作成した。一方は、相手を発見したら追跡するだけの単純なものである(単純AI)。もう一方は、状態数を増やし協調性を考慮して設計した(協調AI)。ここでの協調性は、学問的に定義されているものではなく、追跡行動に最低限必要だと個人的主観に基づいて作成したものであり、状態数が増えすぎないように比較的単純になるように作成した。この2つを比較したとき、協調AIの方が単純AIよりも状態数が多い分複雑な動きをするため、評価が高くなる

と想定される。テスト結果において、協調AIの方が高い評価値になったならば、ゲームAIの比較評価が正しくできたと言え、チューリングテスト適用の有効性が認められる。

AIの各状態は境界内に敵・味方・壁が存在するかしないか、自分を中心とした時の敵と味方との角度、敵を中心とした時の自分と味方との角度、敵の進行方向から定義され、一定時間ごとにこれらを確認し状態遷移を行う。図6は実験を行ったゲームの実行画面における、AIの認識を表しており、図7及び図8は単純AIと協調AIのFSM図を示す。画面上のキャラクターPがプレイヤーによって操作される逃亡側キャラクター、キャラクターEa・EbがAI(または被験者以外の人間)によって制御される追

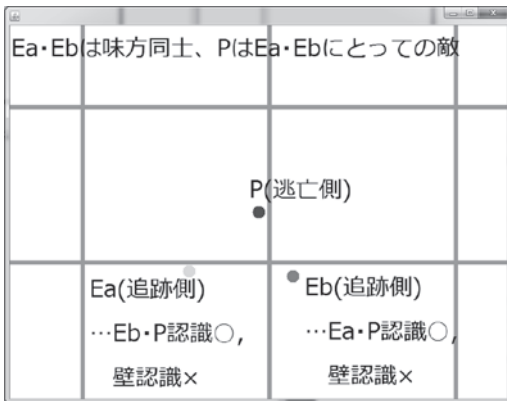


図6 ゲーム実行画面におけるAIの認識

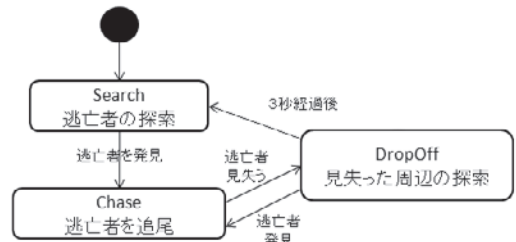


図7 単純AIのFSM図

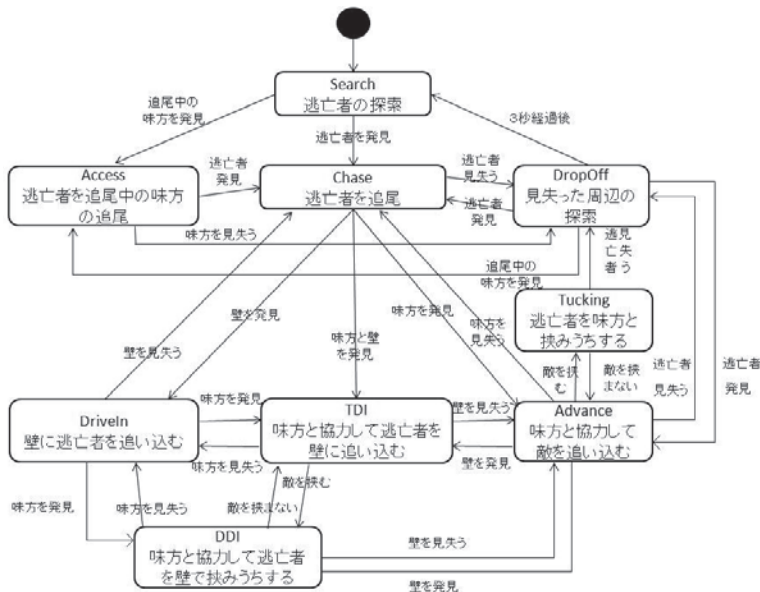


図8 協調AIのFSM図

跡側キャラクターである。各キャラクターは視界を持っている。その視界は、表示される操作画面で表しており、追跡側は縦：500ピクセル、横：700ピクセルに対し、逃亡側は縦：600ピクセル、横：800ピクセルとなっており、逃亡側の方が追跡側の視界に比べ100ピクセルずつ広い。追跡側は味方であるもう一方のキャラクターと重ならないようにするために、一定距離まで近づくとその距離を保つ。スピードも、追跡側が12ピクセル／フレームに対し、逃亡側が14ピクセル／フレームと2ピクセル／フレーム速く設定している。単純AIは、視界内に捕らえた相手（逃亡側キャラクター）に対して、最短距離を選び追跡を行う。視界から外れた時は、直前まで相手を認識していた方向に向かい一定時間移動する。相手を視界に捉えたならば、追跡状態に移り、見失ったまま一定時間が経過したならば、探索状態に移る。しかし、自分よりも速い相手を追いかけるため、相手の後を単純に追っているだけではなかなか捕まえることができない。そこで、協調AIのほうは視界内にいる敵・味方・壁を認識し、それぞれの位置関係から相手をフィールドの角へ追い込んでいく。このときに、追跡側は味方の状態と敵の進行方向を確認し、敵の先回りをしようとする状態や味方と挟み込むような状態をとる。図6においてEa・Ebは視界内に味方であるお互いと敵であるPを認識している。このときEa・Ebの視界内に壁は存在していないため認識しておらず、Pを中心としたEaとEbの角度からP・Ea・Ebの位置関係を確認し、フィールド中央でPをEa・Ebで挟み込みを行っている状態である。

5. チューリングテスト適用の有効性検証

本実験の目的は、対戦型アクションゲームにチューリングテストを適用することにより、ゲームAIにおける人間らしさの評価が可能かを検証し、その有効性を示すことである。そのために、2パターンのチューリングテストを用

意した。パターンを分けた理由は、状況を変えることで、テスト結果がかわるかを調査するためである。その際に、個人の主観に基づいた被験者の評価から、どのような傾向がみられるかを調査する目的で、判断理由を尋ねた。

5.1 検証

実験環境として、対戦型アクションゲームを簡易モデル化した1対2の追跡ゲームを利用した。この追跡ゲームでは、壁が存在する限られた空間内において逃亡側1体、追跡側2体によるキャラクターが表現されている。終了条件は逃亡側が追跡側に捕まるか、制限時間30秒を経過した場合とした。検証では被験者に逃亡側のキャラクターを操作してもらいテストを行った。本稿における追跡ゲームでは、アクションゲームが持つ機能の移動に着目し、キャラクターがフィールド内を移動する部分を取り扱い、AIは追跡行動（それに付随する行動）のみを行う。実験用に用いたAIは、ゲーム制作に一般的に適用されているFSMを用いた。図1より、画面上のキャラクターPがプレイヤーによって操作される逃亡側キャラクター、キャラクターEa・EbがAI（または被験者以外の人間）によって制御される追跡側キャラクターである。

被験者をテストごとに10人ずつ用意し正解率を調べた。また、判断理由も記述してもらった。このときの正解とは、被験者の回答と、実際に人間が操作していたキャラクターが一致することを指す。

5.1.1 テストS (Single)：キャラクター単体の評価

このテストは、どのキャラクターを人間が操作しているか当てる、キャラクター単体として見たときの評価が目的である。図1のようなゲーム画面において、人間が操作する追跡側キャラクターと、AIが制御する追跡側キャラクターが混在しており、2体の追跡側キャラクターEa・Ebが、逃亡側キャラクターを追跡する。被験者には逃亡側キャラクターを操作して

もらう。ゲームプレイ1回終了ごとに、人間が操作しているキャラクターはEa・Ebのどちらかを、被験者に当ててもらう。追跡側キャラクターEa・Ebの内、人間が操作するキャラクターは1体だけとは限らず、Ea・Eb両方AIが制御している可能性もあれば、両方人間が操作している可能性もある。

5.1.2 テストP (Pair) : チームとして見た時の評価

ゲームを2回プレイしてもらう。その際、追跡側キャラクターは人間のみが操作する場合とAIのみが制御する場合に分ける。ゲームプレイ2回終了ごとに、人間が操作していたのは1回目と2回目のどちらかを被験者に当ててもらう。テストPは、人間が操作している組み合わせはどちらかを当てる、チームとして見た時の評価が目的である。

5.2 結果及び考察

テスト結果として、テストSでは10人中6人、テストPでは10人中8人の被験者に、人間が操作していたキャラクターを当てられた。よって、今回定義したFSMでは、AIが制御するキャラクターを人間が操作していると錯覚させることはできなかったと考えられる。被験者から得られた判断理由は以下の通りである。

●テストS

被験者の操作に対する瞬時的な反応や急激な方向転換、一定行動の繰り返しが見られた。被験者はそれらを機械らしい行動だと評価し、AIが制御しているキャラクターだと判断していた。それに対し、被験者の急激な方向転換に即座に対応できず、無駄な行動やランダムな行動を取っていたキャラクターを被験者は人間らしいと評価し、人間が操作しているキャラクターだと判断していた。

●テストP

テストSと違い、1回目と2回目のどちらかが必ずAIだとわかっているため、2回のゲームプレイを比較して判断していた。判断理由としては、テストSと同様の意見が多くみられた

が、より明確な評価をしている傾向が見られた。また、AI同士が類似した行動を取っている組み合わせを機械らしいと評価していた。

以上のことから、被験者は個人の持つ人間らしい行動のモデルを基に、ゲーム内のキャラクターを比較・評価し、どのキャラクター（の組み合わせ）を人間が操作していたのか判断していることが分かった。すなわちチューリングテストは、ゲームAIにおける人間らしさを、被験者の主観的判断に基づき、比較評価できることを示している。よって、ゲームAIにおける人間らしさの評価を客観的に行う上で、チューリングテストを用いることは有効であると言える。

しかし、今回のテストではどのような結果が得られれば人間らしいと言えるかを明確にしておらず、また、AIを一種類しか用意していなかったため、比較評価ができなかった。さらに、被験者には1回のみでのテストで評価を行ってもらったため、曖昧な判断を下していた節があり、評価にばらつきが見られた。以上のことから、テストの内容を見直し、明確な評価基準を定めた適用方法を提案する。

6. チューリングテストの適用手法決定

チューリングテストを適用する際、テスト内容そのものについて深く考える必要があった。具体的には、何を評価し、どのような結果が得られれば人間らしいとするかの定義をはっきりと明言する必要がある。また、テストの有意性を示すためのAIの比較評価がなかった点、被験者の評価におけるばらつきが考慮されていなかった点が、反省点として挙げられる。

5章で行ったテストにおいて判明したこれらの要素を補足するため、新たに見直したチューリングテストの適用方法を提案し、検証を行う。ここでは、新たに提案するチューリングテストをゲームに適用することにより、ゲームAIにおける人間らしさの評価を行えるか検証する。反省点より、単純追跡AIと協調性を考慮したAIの2種類による比較評価を行う。ま

た、評価基準を明確にするため、5段階の評価尺度を定義した。テストは目的が異なる2つのパターンを用意した。テスト過程において、個人の主観に基づいた被験者の評価から、どのような傾向がみられるかを調査する目的で判断理由を尋ねた。

6.1 検証

実験環境は、5章と同様の環境を利用した。また、本節では比較評価のために状態数の異なる2種類のFSMを用意した。一方は、相手を発見したら追跡するだけの単純なものである。もう一方は、状態数を増やし協調性を考慮して設計した。この2つを比較したとき、協調性を考慮したAIの方が単純追跡AIよりも状態数が多い分複雑な動きをするため、評価が高くなると想定される。テスト結果において、協調性を考慮したAIの方が高い評価値になったならば、ゲームAIの比較評価が正しくできたと言え、チューリングテスト適用の有効性が認められる。図1より、画面上のキャラクターPがプレイヤーによって操作される逃亡側キャラクター、キャラクターEa・EbがAI（または被験者以外の人間）によって制御される追跡側キャラクターである。

今回は被験者20人に対しテストを行った。評価については5段階評価とし、0～4の評価値を設定した。被験者の判断が追跡側の操作と一致していれば低い評価値0を、外れていれば被験者を錯覚させたとして高い評価値4を与える。各テストはAIごとに5回ずつ行い、評価の統計を出した。

テストS (Single) : キャラクター単体の評価

このテストは、キャラクター単体における人間らしさの評価を目的としている。図1のようなゲーム画面において、追跡側キャラクターEa・Ebの片方を人間、もう片方をAIが操作する。被験者には逃亡側キャラクターを操作してもらう。Ea・Eb、どちらのキャラクターを人間が操作していたのか、図9及び図10の評価尺度を用いてゲームプレイ1回終了ごと被験者に

判断してもらった。評価値は、被験者の判断と実際の操作とのギャップになるので、Eaを人間が操作していた場合を考えると、図9のように「Eaが人間」という項目から0、1、2、3、4となり、Ebを人間が操作していた場合、評価値は逆になる（図10）。テストは被験者1人に対して、AIごとに5回ずつ、計10回行う。

テストP (Pair) : チームとして見た時の評価

このテストは、EaとEbをチームとしてみなした時の評価を目的としている。こちらのテストでは、ゲームを2回プレイした後に評価を行ってもらう。このとき、1回を人間同士、もう1回をAI同士によってEa・Ebを操作する。1回目と2回目のゲームプレイを比較し、どちらが人間同士の組み合わせであったのか、図11及び図12の評価尺度を用いて被験者に判断してもらった。テストPの場合、1回目が人間同士



図9 テストS (Single) における評価尺度 (Eaを人間が操作していた場合)



図10 テストS (Single) における評価尺度 (Ebを人間が操作していた場合)



図11 テストP (Pair) における評価尺度 (1回目が人間同士の組み合わせだった場合)



図12 テストP (Pair) における評価尺度 (2回目が人間同士の組み合わせだった場合)

の組み合わせだった場合を考えると、図11のように「1回目が人間」という項目から0、1、2、3、4となり、2回目が人間同士の組み合わせだった場合、評価値は逆になる（図12）。テストは被験者1人に対して、AIごとに5回ずつ、計10回行う。しかし、ゲームを2回プレイした後に評価を行うので、テスト1回はゲームプレイ2回分となる。そのため、実際のゲームプレイは20回行うことになる。

6.2 結果及び考察

被験者ごとの評価値を、各AIの最後に行ったテストの評価を最終評価値として求めた。図13及び図14は、それぞれテストS、テストPにおける単純追跡AI（単純AI）と協調性を考慮したAI（協調AI）の評価値を比較したグラフである。また、統計から中央値、平均値、最頻

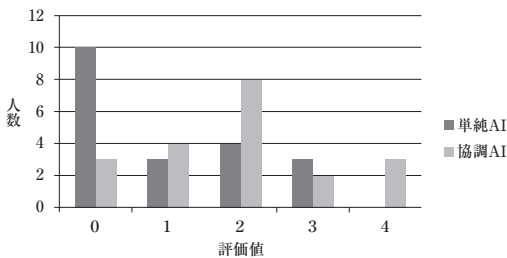


図13 テストS (Single) における最終評価値

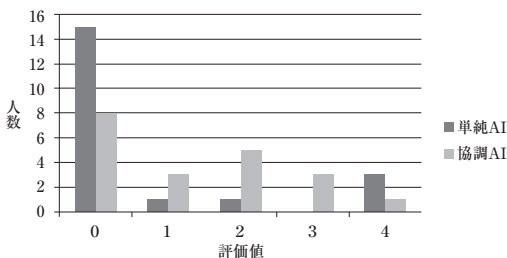


図14 テストP (Pair) における最終評価値

表1 テスト結果の基本統計量

		中央値	平均値	最頻値
テストS	単純AI	0.5	1	0
	協調AI	2	1.9	2
テストP	単純AI	0	0.75	0
	協調AI	1	1.3	0

値を求めた。（表1）

結果として、グラフより協調AIのほうが、単純AIより評価値が高いことが読み取れる。基本統計量も、協調AIのほうが、単純AIよりも大きい値を示している。よって、テスト結果において、協調性を考慮したAIの方が高い評価値になったので、ゲームAIの比較評価が正しくできたと言え、チューリングテスト適用の有効性が認められる。

このテストから、以下のような被験者の判断傾向が見られた。最初はそれぞれの動きを確認するために、純粹に逃げきろうとする。しかし、ある程度動きの確認ができると周回行動や不規則な行動などを行い、その行動に対する反応を確認していた。その後、「回り込むような動きは人間らしい」、「一定行動の繰り返しは機械的」などの理由を基に、人間の操作かAI制御かを判断していた。以上のことから、被験者はゲームプレイを通じ、それぞれの行動をモデル化し比較評価していることがわかった。被験者の主観的モデルを抽出することができれば、それを基により人間らしいゲームAIを作成することができると考えられる。

結果として、チューリングテストを適用することにより、人間の主観に基づく評価を、定義した評価尺度から数値として表わすことができた。そして、評価尺度を基に比較評価することで、ゲームAIにおける人間らしさの客観的評価を行うことができた。しかしながら、AI同士の比較評価では相対評価に留まってしまうことがわかった。次章では、明確な判断基準値を定め、絶対評価を可能とする手法を提案する。

7. 絶対評価基準値の決定

本章では、6章で明らかになったことから、人間同士によるチューリングテストを行うことにより、人間のゲーム操作における人間らしさを抽出し基準値を設定することで、相対評価に留まっていたゲームAIの評価を、絶対評価で行うことができると想定し、検証を行う。その

際に、主観に基づいた被験者の評価から、どのような傾向がみられるかを調査する目的で判断理由をたずねた。

7.1 検証：人間同士によるテスト

実験環境は5、6章と同様の環境を利用した。本実験では、追跡側キャラクターEa・Ebの内、一方のキャラクターを正解データとして固定した人間が操作し、もう一方のキャラクターを対象データとして別の人間（今回は6人）が操作する。ここで、対象データを人間と同等の思考をするAIであると仮定する。Ea・Ebの内、被験者に一方は人間が操作しており、もう一方はAIが制御していると伝え、人間が操作しているキャラクターはどちらであるかを、図15および図16の評価尺度を用いてゲームプレイ1回ごとに被験者に判断してもらう。

仮に、被験者に対し追跡側を両方人間が操作していると、事前に伝えたとする。この場合、正解データの操作パターンを見せ、操作の癖などを見てもらったうえで評価を行う必要がある。さらに、判断内容もどちらが正解データの操作であったかというものになり、趣旨と異なる評価になってしまう。しかし、今回のような状況を設定することで、6章で行ったチューリングテストと同様の認識で検証を行える。その結果、人間の操作における人間らしさから、ゲームAIの人間らしさにおける評価の基準値が求められ、絶対評価を行うことができると想定される。

今回、被験者20人に対しテストを行った。評

価については5段階評価とし、0～4の評価値を設定した。被験者の判断が正解データと一致していれば低い評価値0を、外れていれば被験者を錯覚させたとして高い評価値4を与える。評価値は、正解データと被験者の選択とのギャップになるので、Eaが正解データだった場合を考えると、図15のように「Eaが人間」という項目から0、1、2、3、4となり、Ebが正解データの場合、評価値は逆になる（図16）。テストは被験者1人に対し、対象データごと5回ずつ計30回行った。

7.2 結果及び考察

図17は、人間同士によるチューリングテストの結果を表わすグラフである。横軸が図15および図16の評価尺度に基づき算出した評価値を示しており、縦軸が評価値の頻度を示している。600回分のテスト結果から評価値の平均を求めたところ、1.93という値が得られた。

本実験において、被験者は追跡側キャラクターを操作しているのは人間とAIだと認識して評価している。しかし、実際にはどちらも人間が操作しているため、どちらが人間の操作しているキャラクターであるか、という質問に対して被験者は明確に判断することができない。これは、2章で示したゲームAIにおけるチューリングテストの定義と合致している。結果として、人間同士でチューリングテストを行うと、その評価は中央に収束することになった。ゲームAIの人間らしさにおける評価の基準値を中央値である2（評価尺度における



図15 評価尺度 (Eaを人間が操作していた場合)



図16 評価尺度 (Ebを人間が操作していた場合)

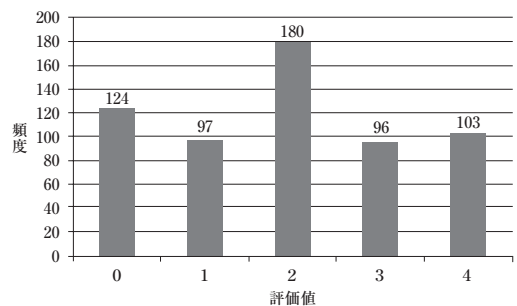


図17 人間同士によるチューリングテストの結果

「わからない」と設定することで、相対評価に留まっていたゲームAIの評価を絶対評価で行うことができる。すなわち、本研究におけるチューリングテストによるゲームAIの評価は、中央値2に近づくほど人間らしいゲームAIであると言える。

テスト過程では、「最初はそれぞれの動きを確認するために、純粹に逃げきろうとするが、ある程度動きの確認ができると周回行動や不規則な行動などを行い、その行動に対する反応を確認していた」というような被験者の判断傾向が見られた。このような傾向から、ゲームプレイにおいて、被験者が主観に基づき、機械らしい行動と人間らしい行動をそれぞれモデル化し、比較・評価していることが明らかになった。人間が無意識のうちにやっている、主観に基づいた行動のモデル化を定義することができれば、より人間らしいゲームAIを作成できると想定される。

8. おわりに

本研究は、ゲーム内におけるキャラクターの人間らしい思考・動作を表現することで、ゲームAIのリアリティを追求することを目標としている。ゲームAIの人間らしさを、AIに対しプレイヤーが人間であると感じる（錯覚する）ことであると定義したとき、その評価は個人の主観的体験によって判断されるため、客観的な評価に基づく具体的な判断基準が確立されていないという問題点があった。そこで本研究では、ゲームAIの人間らしさを、客観的に評価する手法の確立を目的とし、ビデオゲームへのチューリングテスト適用方法の検討を行ってきた。

5章では、ゲームAIにおける人間らしさの評価において、対戦型アクションゲームにチューリングテストを適用することの有効性を示す目的で、検証を行った。テスト結果から、被験者は個人の持つ人間らしい行動のモデルを基に、ゲーム内のキャラクターを比較・評価し、どのキャラクター（の組み合わせ）を人間が操作し

ていたのか判断していることが分かった。すなわちチューリングテストは、ゲームAIにおける人間らしさを、被験者の主観的判断に基づき、比較評価できることを示している。よって、ゲームAIにおける人間らしさの評価を客観的に行う上で、チューリングテストを用いることは有効であると言えた。しかし、チューリングテストを適用する際、どのような結果が得られれば人間らしいと言えるかを明確にしていなかった。また、テストの有意性を示すためのAIの比較評価がなかった点、被験者の評価におけるばらつきが考慮されていなかった点が、反省点として挙げられた。

6章では、5章で行ったテストにおいて判明した反省点を考慮し、新たに見直したチューリングテストの適用方法を提案し、検証を行った。実験は、FSMを用いて作成した2種類のAIによる比較評価を行った。テストは目的が異なる2つのパターンを用意した。その結果、ゲームAIの人間らしさを数値化し、比較評価することができた。被験者は人間とAIの各操作を比較して判断する場合、個人の主観によりそれぞれのモデルを定義し評価していることがわかった。しかしながら、AI同士の比較評価では相対評価に留まってしまうことがわかった。

7章では、人間同士によるチューリングテストを行うことにより、人間のゲーム操作における人間らしさを抽出し基準値を設定することで、相対評価に留まっていたゲームAIの評価を、絶対評価で行うことができると想定し、検証を行った。実験では、人間同士によるチューリングテストを行い、人間の操作における人間らしさから、被験者が感じる人間らしさを表す評価尺度を設定した。その結果、評価値は中央に収束することがわかり、中央値をゲームAIの人間らしさにおける評価の基準値として設定した。これにより、相対評価に留まっていたゲームAIの評価を、絶対評価で行うことができた。また被験者は、人間とAIの各操作を比較して判断する場合、個人の主観によりそれぞれ

れの行動をモデル化し、評価していることがわかった。このことから、人間が無意識のうちにやっている、主観に基づいた行動のモデル化を定義することができれば、より人間らしいゲームAIを作成できると想定される。今後の課題として、実際のゲームにおけるチューリングテストの適用方法と、被験者の主観的モデルの定義方法について考えていくことが挙げられる。

【参考文献】

- [1] 石黒浩, アンドロイドによるトータルチューリングテストの可能性, 人工知能学会誌, Vol.26, No.1, pp.50-54 (2011)
- [2] 松原仁, チューリングテストとは何か, 人工知能学会誌, Vol.26, No.1, pp.42-44 (2011)
- [3] 三宅陽一郎, デジタルゲームにおける人工知能技術の応用, 人工知能学会誌, Vol.23, No.1, pp.44-51 (2008)
- [4] 三宅陽一郎, プログラミングAI, デジタルコンテンツ制作の先端技術応用に関する調査報告書2008年度版, pp.73-136, 財団法人デジタルコンテンツ協会 (DCAJ) (2008)
- [5] Stuart Shieber(ed.), The Turing Test, The MIT Press (2004)
- [6] 2011 Essential Facts About the Computer and Video Game Industry, The Entertainment Software Association(esa), 公式HP, <http://www.theesa.com/>
- [7] INTERNATIONAL E-SPORTS FEDERATION (IeSF), 公式HP, <http://ie-sf.org/eng/index.asp>
- [8] eスポーツ学生連盟 (e-Sports Promote Association : eSPA), 公式HP, <http://senshu-e-sports.com/index.html>
- [9] 小渕洋一, 離散情報処理とオートマトン, 朝倉書店 (1999)
- [10] John E. Laird, Michael van Lent, THE ROLE OF ARTIFICIAL INTELLIGENCE IN COMPUTER GAME GENRES, Handbook of Computer Game Studies, pp.205-215, The MIT Press (2005)
- [11] J. Barnes, J. Hutchens, S. Rabin (ed.), Testing Undefined Behavior as a Result of Learning, AI Game Programming Wisdom, pp.615-623, Charles River Media Inc. (2002)
- [12] Daniel Livingstone, Turing's Test and Believable AI in Games, ACM Computers in Entertainment (CIE), Vol.4, No.1 (2006)
- [13] 石黒浩, アンドロイドは人になれるか?, 電子情報通信学会, 第2回ネットワーク・ロボット時限研究会 (NR-TG) (2005)
- [14] 安武諒, 山口崇志, マッキンケネスジェームス, 永井保夫, 対戦型ビデオゲーム用ゲームAIにおけるチューリングテストの有効性検証, 第9回情報科学技術フォーラム (FIT2010) (2010)
- [15] 安武諒, 山口崇志, マッキンケネスジェームス, 永井保夫, ゲームAIにおけるチューリングテストの適用方法についての提案, 第37回ファジィ・ワークショップ (2011)
- [16] 安武諒, 岩崎信也, 山口崇志, マッキンケネスジェームス, 永井保夫, ゲームAIにおけるチューリングテストの適用評価, 第10回情報科学技術フォーラム (FIT2011) (2011)
- [17] 安武諒, 岩崎信也, 山口崇志, マッキンケネスジェームス, 永井保夫, チューリングテストによるゲームAIの客観的評価方法, 情報処理学会第74回全国大会 (2012)
- [18] 松井悠, デジタルゲームの技術, ソフトバンククリエイティブ (2011)
- [19] デジタルゲームの教科書制作委員会, デジタルゲームの教科書, ソフトバンククリエイティブ (2010)
- [20] 大川知, 広瀬貞樹, 山本博章, オートマトン・言語理論入門, 共立出版 (2012)
- [21] Michael Sipser (太田和夫, 田中圭介監訳, 阿部正幸, 植田広樹, 藤岡淳, 渡辺治訳), 計算理論の基礎 原著第2版1 オートマトンと言語, 共立出版 (2008)
- [22] David M. Bourg, Glenn Seemann, 株式会社クイープ訳, ゲーム開発者のためのAI入門, オライリー・ジャパン (2005)