

ゲーム AI におけるチューリングテストの適用評価 Application and Evaluation of the Turing Test to Game AI

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†
Ryo Yasutake Shinya Iwasaki Takashi Yamaguchi Kenneth J. Mackin Yasuo Nagai

1. はじめに

近年、ゲーム AI(ビデオゲームにおける人工知能技術)への関心が高まり重要視されている[1]。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。プレイヤーのゲーム AI のリアリティに対する意識が高まりつつある背景から、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。

1.1 ゲーム AI

ゲーム AI とは、プレイヤーに対して知性を感じさせる存在のことを指し、主にキャラクターAIとメタAIにわけられる。ゲームAIでは、それぞれの持つ役割がゲームの挙動を違和感がないよう自然に見せることが要求される[2]。特に、キャラクターAIはプレイヤーに直接認識される部分であるため、より知性の表現が重要となり、ゲームにおける人間らしさを担っている。以降、本稿におけるゲームAIはキャラクターAIのことを指す。

ここで、ゲームAIの人間らしさを、AIが操作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と定義する。本稿では、過去に提案した対戦型アクションゲームへの、チューリングテスト適用方法を改良したものについて考える。実験環境として、1対2の対戦型アクションゲームを基にした追跡ゲームを、追跡問題として取り扱った。過去の論文において、チューリングテストに基づいたテストを用意し、検証と考察を行った[3]。考察からテスト内容の修正点が明らかになったため、改良したチューリングテストの適用方法を提案し、その有効性について検証・考察を行う。

2. チューリングテスト

ゲームAIの人間らしさの定義から、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の人間らしさの評価は、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。本稿では、人間の主観的判断によるチューリングテストをゲームAIに適用し、複数の被験者による統計を出すことで、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。

チューリングテストは、質問をいくつか繰り返しそれに対する人間とコンピューターの答えがどちらだか区別がつかなければそのコンピューターは知能的であると判断する[4]。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[2]。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製したAIは人間らしいといえる。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だ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込まれる。

3. 実験環境

実験環境として、1対2の対戦型アクションゲームを基にした追跡ゲームを、追跡問題として扱った。今回取り扱う追跡問題は、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で表現した。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(Finite State Machine:有限状態機械)[5]を用いて作製した。今回、比較評価のために状態数の異なる2種類のFSMを作製した。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よりも複雑な動きをするため、評価が高くなると予想される。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は実験を行ったゲーム実行画面である。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。



図1 ゲームの実行画面

4. 検証

本稿では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の人間らしさの評価を行えるか検証する。そのために、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。テストは目的が異なる2つのパターンを用意した。同時に、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たずねた。

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

4.1 テスト内容

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を、外れていれば被験者を錯覚させたとして高い評価値を与える。各テストは AI ごとに 5 回ずつ行い、評価を統計的に分析した。

4.1.1 テスト 1: 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。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 2 の評価尺度を用いてゲームプレイ 1 回ごと被験者に判断してもらった。



図 2 テスト 1 における評価尺度

4.1.2 テスト 2: 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 3 の評価尺度を用いて被験者に判断してもらった。



図 3 テスト 2 における評価尺度

4.2 テスト結果

被験者ごとの評価値として、最後に行ったテストの評価を最終評価値、テスト 5 回分の平均を平均評価値として求めた。図 4 は、単純追跡 AI と協調性を考慮した AI の評価値を比較し、値が高かったのがどちらであったかをグラフ化したものである。結果として、協調性を考慮した AI の評価値が高くなる被験者の方が多かった。

表 1 は、被験者による評価値の合計から求めた AI の人間らしさの度合いを表したものである。表から、協調性を考慮した AI の数値が高く、単純追跡よりも高い評価であることが分かる。この度合いを比較することで、ゲーム AI の人間らしさを相対評価することができた。

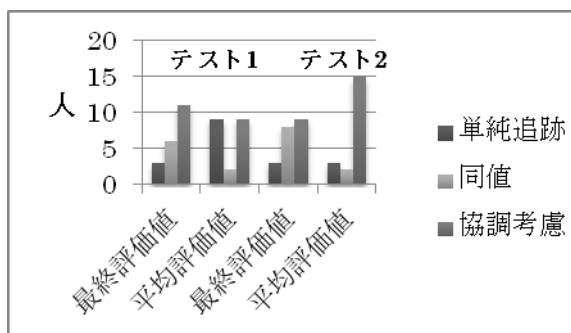


図 4 評価値による AI の比較

表 1 評価値から求めた人間らしさの度合い

	テスト 1		テスト 2	
	単純追跡	協調考慮	単純追跡	協調考慮
最終評価	23.75%	43.75%	21.25%	37.50%
平均評価	36.50%	38.00%	20.75%	34.75%

5. 考察

今回行ったテストから、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、純粋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。このようなゲームプレイを通じ、回り込むような動きは人間らしいさがみられ、一定行動の繰り返しは機械的などの理由から、人間か AI かを判断していた。このことから、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 を作製することができると考えられる。

今回は比較評価の検証のために、評価値を 0~4 と仮定し実験を行った。しかし、このような評価値の設定については統計学の観点から議論し、順序尺度について改めて定義する必要がある。また、ゲームプレイが早く終了してしまった時の評価値について、評価そのものを無効とし切り捨てるか、AI に一定以上の強さは必要と判断し中央値を与えるか考慮する必要がある。

6. おわりに

本稿では、ゲーム AI の客観的な人間らしさの判断基準を確立することを目的とし、チューリングテストに評価を提案、その適用方法について議論した。実験は、FSM を用いて作製した 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。今後の予定として、統計学を用いた評価値の設定や、被験者の主観的モデルの抽出方法について考えていく。

参考文献

- [1] 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)
- [2] 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, pp73-136, 財団法人デジタルコンテンツ協会(DCAJ) (2008)
- [3] 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(2010)
- [4] Stuart Shieber (ed.), The Turing Test, The MIT Press (2004)
- [5] 小淵洋一, 離散情報処理とオートマトン, 朝倉書店 (1999)
- [6] Foost Raessens, Jeffrey Goldstein, Handbook of Computer Game Studies, The MIT Press(2005)
- [7] 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)
- [8] Daniel Livingstone, Turing's Test and Believable AI in Games, ACM Computers in Entertainment (CIE), Vol.4, No.1 (2006)
- [9] 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)
- [10] 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)