

Title	仲間の立場から人間の行動を誘導するゲーム AI に向けての考察
Author(s)	山田, 直央
Citation	
Issue Date	2022-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/17640
Rights	
Description	Supervisor:池田 心, 先端科学技術研究科, 修士(情報科学)

修士論文
仲間の立場から人間の行動を誘導するゲーム AI に向けての考察

山田 直央

主指導教員 池田 心

北陸先端科学技術大学院大学
先端科学技術研究科
(情報科学)

令和4年3月

Abstract

In recent years, as the improvements in the communication techniques, the delay problem has been reduced. This enables people living remotely and a large number of people to play games even that lags of several frames may cause discomfort. Among games that have benefited from this fact, multi-player shooting games are a popular genre worldwide. It is also a popular genre for e-sports.

Many people want to start playing games belonging to this genre. However, when beginners with poor skills play these games, they are likely to be beaten easily by other players due to the difference in skills. Beginners may have no much opportunity to learn through trial and error while enjoying the games. Also, in each game, specific gestures are used to communicate intentions. It is often that beginners overlook such gestures, making them difficult to exchange information that leads to clearing the games. In this sense, players new to these games do not have time to learn how to play and may not enjoy the games because they cannot communicate smoothly with their teammates.

If machines can communicate in a way that is easy to understand and human-like, they can behave in such a way and serve as a starting point for humans to learn strategies for playing games. These machines will provide more opportunities for learning strategies. In order to provide humans with an appropriate environment for practice and learning, we consider it necessary to model intention communications that are conveyed without words.

To model intention communications that are conveyed without words, we need to analyze and reproduce the following topics: *how humans communicate intentions*, *how to utilize human-like intention communications*, and *how intention communications influence the opponents/teammates*. In this research, we aim at the second topic, analyzing and reproducing how to utilize human-like intention communications.

For this purpose, we discussed 3 conditions that human-like intention communications are likely to occur during gameplay: *human players have different information*, *it is easy to observe teammates' actions*, and *players have some rooms in terms of action space and some spare time*. Based on these conditions, we implemented such an experiment environment that intention communications are likely to occur. In more detail, each intention communication in this experiment environment is performed by a button push, making us easy to collect labels of the timings when the intention communications took place. In this research, we had 6 kinds of intention communications, coming from 3 circumstances: during stage selection, during item emergence, and during battle, where each has 2 choices (2 buttons).

We collected about 1,351 minutes of human behavior data from the implemented

experiment environment. We then used these data to do supervised learning to predict what kinds of intention communications will be done at which timings during gameplay. The learning network contained the architecture called C-LSTM, a combination of convolutional layers and LSTM (long short-term memory) layers. For the input data, we used map data of 86×43 and 56 data not in the same form as maps. Map data were input to the convolutional layers, and the obtained outputs were combined with the remaining input data, which were then input to LSTM layers. We set the output data of the network to 3 parameters, representing *whether the players are communicating intentions, whether the players press button 1 for intention communication, and whether the players press button 2 for intention communication.*

For the supervised learning, we tried different settings for comparison, including masking some information from the input data and varying the time length of the input data. We compared the prediction accuracy of these networks and confirmed the following results: the prediction on timings were strongly influenced by *which human players were put in the same team*, and the accuracy was higher when the input data contained 100 frames (10 seconds) than 500 frames (50 seconds). Among the predictions by the networks, the highest F-score was 0.306, which was not a high value. However, since it is known that there are differences between humans in *whether to press the buttons or not and how fast or slow they press the buttons*, the result alone does not mean poor performance.

In addition, we conducted an experiment to confirm whether the timings of doing intention communications by the proposed method were natural to human players. We made two sets of videos based on the collected human behavior data: *videos showing the timings that intention communications were actually done* and *videos showing the timings that intention communications were predicted to happen by the supervised learning model.* Participants of this experiment were asked to watch the videos and answer whether they would do intention communications at the next moment. 62.7% of the answers matched the data for making the former set of videos (those showing the timings that intention communications were actually done), and 55.0% matched the latter set (videos showing the timings that intention communications were predicted to happen by the supervised learning model). From the results, the naturalness of the timings predicted by the proposed method was slightly worse than the timings from human data. However, compared to the bad impression received from an F-score of 0.306, the difference from human data in this experiment was much more natural. From this point, we believe that the goal of this research has been achieved to some extent.

概要

近年では通信技術の向上により、通信遅延の問題などが緩和されたことで、数フレームの通信の遅れが人間にとって不快になるようなゲームを、今までより遠隔に住む人間で、しかもより大人数で遊べるようになってきた。その恩恵を受けたゲームの中でも、マルチプレイ・シューティングゲームは国内・国外問わず人気の高いゲームジャンルであり、e-sports のひとつのジャンルとしても人気が高い。

このジャンルで新たに遊び始めたいと考える人間は多いが、操作技術の拙い初心者がプレイする場合、他の人間との実力差によりすぐに打ち倒されてしまい、楽しみながら試行錯誤し学ぶ機会を得る時間が少なくなる。また、ゲーム内にはそのゲーム特有のジェスチャーを利用した意思伝達が存在するが、初心者はこの意思伝達を見逃してしまう場合も多く、意思伝達によってゲームのクリアに繋がる情報のやり取りがうまくいかない。このように、ゲーム初心者は学習の時間もなく、仲間との意思伝達が円滑にいかないためにゲームを楽しめない場合がある。

もし機械が人間に伝わりやすい、人間らしい意思伝達を行えるようになれば、人間に作戦を学ぶ取っ掛かりになるような行動をするように誘導することができ、作戦を学ぶ機会を多く提供できる。人間への適切な練習・教育の環境の提供のために、言葉を介さない意思伝達をモデル化することが求められる。

言葉を介さない意思伝達をモデル化するためには、「人間の意思伝達方法の分析と再現」と「人間らしい意思伝達の利用方法の分析と再現」・「意思伝達が相手に及ぼす効果の分析と再現」が必要である。本研究では、このうちの1つである「人間らしい意思伝達の利用方法の分析と再現」を目的として行っている。

この目的のために我々は、人間らしい意思伝達が発生しやすい条件として「人間プレイヤーの持つ情報に差があること」・「仲間の行動が見やすいこと」・「行動空間の幅と時間的余裕がある事」の3つを検討し、意思伝達の発生しやすい実験環境を自作した。また、この実験環境では、意思伝達を1ボタンで行えるようにすることで、いつ意思伝達をしたかというラベル付けを簡単にしている。なお、本研究で取り扱う意思伝達は、ステージ選択中・アイテム出現中・戦闘中の3つの状況についてそれぞれ2通り（2ボタン）の計6種類とした。

この実験環境から約1351分の人間の行動データを収集し、その行動データを使って、ゲーム中のどのタイミングでどの意思伝達が行われるかを出力とする教師あり学習を行った。学習のネットワークにはC-LSTMという、畳み込み層とLSTM層をつなげたものを使用した。入力データには86×43のマップデータと、56のマップデータ以外のデータを使用した。マップデータは畳み込み層から入力し、そこで得られた出力をマップデータ以外のデータと結合してLSTMに入力した。また、出力データは「意思伝達をしているかどうか」・「意思伝達ボタン1を押しているか」・「意思伝達ボタン2を押しているか」の3パラメータとした。

この教師あり学習を、入力データから一部の情報をマスクして行ったり、入力データの時間長さを変えたりして行った。そして、ここで得られたネットワークの予測精度を比較し、タイミングの予測が「どの人間が組んでいたか」の影響を強く

うけること、入力データが500フレーム分(50秒分)であるときより100フレーム分(10秒分)の方が精度が高いことを確かめた。予測を行った中で、最も高かったF値は0.306で、決して高い値とは言えないが、そもそも人間の間でも「押す押さない・早い遅い」には差があることもわかっており、これだけで性能が悪いということとはできない。

人間にとって提案手法のタイミングが自然であるかを確かめるための実験も行った。人間の行動データで、「実際に意思伝達が行われていたタイミング」と、「機械学習で予測された意思伝達が行われるタイミング」の動画を作成し、人間に対し、次の瞬間に意思伝達を行うかの被験者アンケートを行った。「実際に意思伝達が行われたタイミング」の正解率は62.7%、「機械学習で予測された意思伝達が行われるタイミング」の正解率は55.0%であった。結果として、提案手法の予測するタイミングの自然さは、人間のタイミングに比べて劣っていたが、0.306というF値の値から受ける印象よりは、人間にとってずっと自然な意思伝達をしていたことが明らかになった。このように、本研究の目的はある程度達成されたと考えている。

目次

第1章	はじめに	1
第2章	関連研究	4
2.1	機械と人の協力	4
2.2	機械と人間の違い	6
2.3	意思伝達の種類	7
2.4	ゲームにおけるゆとり	9
第3章	対象ゲームの考察と提案	10
3.1	対象とする意思伝達	10
3.2	意思伝達の発生する環境	11
3.2.1	人間プレイヤーの持つ情報に差があること	11
3.2.2	仲間の行動が見やすいこと	13
3.2.3	行動空間の幅と時間的余裕があること	14
3.3	実験環境の説明	15
3.3.1	操作キャラの動作	15
3.3.2	ステージの説明	16
3.3.3	敵キャラクタ	18
3.3.4	プレイ画面について	19
3.3.5	その他	21
3.4	分析する意思伝達	21
3.4.1	ステージ選択中	22
3.4.2	戦闘中	24
3.4.3	アイテム出現中	25
第4章	被験者実験の傾向と分析	26
4.1	意思伝達の各使用頻度	27
4.2	意思伝達の連続使用時間	29
4.3	主目的と意思伝達使用回数の関係	31
第5章	学習手法の提案	33
5.1	ニューラルネットワークの入出力	33
5.1.1	入力データ	33

5.1.2	出力データ	35
5.2	タイミングの予測を行うニューラルネットワーク	38
5.3	学習結果の評価方法	39
5.3.1	評価対象の処理	39
5.3.2	機械的評価方法	40
5.3.3	人間的評価方法	43
第6章	学習結果	44
6.1	機械的性能評価	44
6.1.1	教師データの長さの影響	46
6.1.2	ペア情報の重要性	47
6.2	人間的性能評価	48
第7章	おわりに	49

目次

2.1	ハンタータスクのクリア方法1	4
2.2	ハンタータスクのクリア方法2	4
2.3	目標物 T1 を狙うプレイヤー P1 の挙動	5
2.4	ジャンプ行動での挨拶	8
2.5	指さし行動による煽り	8
3.1	プレイヤーの持つ情報の差	12
3.2	死角の発生例	13
3.3	ゲームステージの構造	16
3.4	安全エリアと戦闘エリア間の扉の開閉	16
3.5	戦闘エリアの得点 (GP) と各プレイヤーの情報の差	17
3.6	移動しない敵キャラクター	18
3.7	移動する敵キャラクター	18
3.8	メインカメラとサブカメラ	19
3.9	プレイ画面の配置	20
3.10	ステージ選択中の回転運動の意味	23
3.11	ステージ選択中の前後運動の意味	23
4.1	意思伝達ボタン使用頻度 [回/分] のヒストグラム	27
4.2	送り手・受け手・ペアそれぞれでの1ゲーム中の意思伝達回数の平均	28
4.3	意思伝達ごとの連続使用時間のはこひげ図. はずれ値は第1四分位数と第3四分位数の差の1.5倍以上に, 第1四分位数または第3四分位数から離れているデータを示す. なお, このはずれ値も学習に利用した.	29
4.4	ペアごとの意思伝達ボタン連続使用時間の中央値のヒストグラム	30
4.5	送り手・受け手・ペアそれぞれでの平均得点	31
4.6	意思伝達の平均回数と総得点 (GP) の関係性	31
4.7	受け手・ペアごとの意思伝達の平均回数と総得点 (GP) の関係性	32
5.1	MapData へ変換する前の画像	34
5.2	MapData の画像	34
5.3	2 値データへの曖昧ラベルの適用	36
5.4	C-LSTM の概念図	38

5.5	ネットワークの出力値	39
5.6	予測値と閾値の関係の例（実際にボタンが押された瞬間より8フレーム早く予測してしまっている）	40
5.7	フレーム誤差の許容範囲	41
6.1	学習を行う区間	44
6.2	基準となるネットワークのF値・再現率・適合率	45
6.3	フレーム長によるF値の影響	46
6.4	ペア情報によるF値の影響	47

表 目 次

4.1	学習データの統計	26
4.2	1 試合中の各状況の平均時間	26
4.3	意思伝達ボタン使用頻度の中央値 [回/分]	28
5.1	BaseData の一部	35
6.1	基準となるネットワークの F 値最大となる時のパラメータ	45
6.2	500 フレームの教師データで学習させたネットワークの F 値最大となる時のパラメータと基準となるネットワークとの比較	46
6.3	ペア情報を入れていないネットワークの F 値最大となる時のパラメータと基準となるネットワークとの比較	47
6.4	実測値・予測値に対する人間的性能評価の割合	48

第1章 はじめに

対戦ゲームを他人と遊ぶ時、直接対面で遊ぶほかに、インターネットなどの通信手段を用いて遠隔で遊ぶこともある。オセロや将棋など、通信の遅れがゲーム進行の妨げになりづらいジャンルでは、通信を使ったゲーム（オンラインゲーム）においても、以前から広く普及している。近年では通信技術の向上により、通信遅延の問題などが緩和されたことで、数フレームの通信の遅れが人間にとって不快になるようなゲームを、今までより遠隔に住む人間で、しかもより大人数で遊べるようになってきた。その恩恵にあずかってきたゲームジャンルは多く、テトリスやぷよぷよに代表されるパズルゲームや、マリオカート等のレースゲームがその代表例だ。そういったゲームジャンルの中でも、マルチプレイ・シューティングゲームは国内・国外問わず人気の高いゲームジャンルであり、e-sportsのひとつのジャンルとしても人気が高い。

人気の高いゲームだけあって、このジャンルで新たに遊び始めたいと考える人間は多いが、操作技術の拙い初心者がプレイする場合、他の人間との実力差によりすぐに打ち倒されてしまい、楽しみながら試行錯誤し学ぶ機会を得る時間が少なくなる。また、ゲーム内にはそのゲーム特有のジェスチャーを利用した言葉を介さない意思伝達が存在するが、初心者はこの意思伝達を見逃してしまう場合も多く、意思伝達によってゲームのクリアに繋がる情報のやり取りがうまくいかない。このように、ゲーム初心者は学習の時間もなく、仲間との意思伝達が円滑にいかないためにゲームを楽しめない場合がある。

操作技術も意思伝達方法も拙い場合でも学ぶ機会を得るために、ランクマッチ制という、同じ技術力・意思伝達レベルをもつ者同士とのみゲームをする方法や、強い人間の実況動画を見ることで最適な操作方法を学ぶ方法が存在する。これらの方法にはそれぞれ問題点があり、前者は同じ実力ばかりの人間ばかりで自発的に他の作戦の発見が困難である点、後者は、人によってゲームを遊びながらでないで集中力が続かず、ゲームをやめてしまうという点である。実際のゲーム環境で遊びながら、実況動画を見ることと同じ学習速度で作戦を学べるような仕組みは一定の需要がある。

遊びながら学ぶために、ゲームの一場面を切り取って操作させる研究 [1] や、ゲームのバランスを変えて初心者と熟練者を対等に戦わせる研究 [2]、ゲーム AI を人間らしく振る舞わせる研究 [3, 4] などがある。しかし、“仲間”の立場のゲーム AI の行動が人間の作戦にどう影響を与えるかについてはほとんど研究されていない。

“仲間”のゲーム AI について研究されているものは、人間の好みに沿って作戦

を選択する仲間ゲーム AI の研究 [5] が主で、人間に対して能動的に作戦を提案できるようなゲーム AI の研究は稀である。人間と近い実力の“仲間”が、言葉を介さない意思伝達を行えるようになれば、人間に作戦を学ぶ取っ掛かりになるような行動をするように誘導することができ、作戦を学ぶ機会を多く提供できる。人間への適切な練習・教育の環境の提供のために、言葉を介さない意思伝達をモデル化することが求められる。

本研究では、言葉を介さない意思伝達をモデル化するために、問題を以下の3つに分類している。

- (A) 人間の意思伝達方法の分析と再現
- (B) 人間らしい意思伝達の利用方法の分析と再現
- (C) 意思伝達が相手に及ぼす効果の分析と再現

A は意思伝達の送り手の分析と再現で、人間の行う意思伝達に該当する行動の発見と、その行動をどういう意図で行ったかという紐づけを行い、それを AI に再現させる。B も意思伝達の送り手の分析と再現で、どういった状況で意思伝達が発生しやすくて、またどの意思伝達を選択されるのかを分析し再現させる。最後に、C は意思伝達の受け手の分析と再現で、送り手がおこした意思伝達に対して、受け手がその意思伝達に対して気付くか気付かないか、または気付いても反応を返すか返さないかを分析し再現する。本研究では、B の「人間らしい意思伝達の利用方法の分析と再現」を研究の主眼におき、A・C については本研究では触れない。以下にその理由を記す。

A に関して、国籍などの個々人の文化的背景や、そのゲームに無数に存在するコミュニティ独特の文化によって、同じ意思伝達方法でも意思伝達の意味が異なる場合が多くある。またゲーム中に観察できる意思伝達は、注意喚起や挨拶・挑発など、ゲームのクリアに直接かかわるものから直接かかわらないものまで多岐にわたる [6]。意思伝達の方法が膨大であるがために、本研究では意思伝達方法が定まっているゲーム環境を扱い、その意思伝達方法が用いられる利用方法の分析と再現の方に焦点をおく。

C に関して、意思伝達の方法が定まっていないようなゲーム環境であればこの分析は必須であるが、本研究では前述した通り、意思伝達方法が定まっているゲーム環境を扱う。これにより、C の意思伝達ที่相手に及ぼす効果の予想より、B の人間らしい意思伝達の利用方法の分析の方が優先度が高い。

以上から、本研究では B に焦点をおいて分析と再現を行う。手順としては、分析を行ったのちに、その分析が正しいかどうかを確かめるために、人間のプレイデータと比較して、意思伝達の選択が再現可能かどうかを確かめる。

各章の構成と内容は次のようになっている。第2章では、意思伝達の種類や発生する条件などに関する関連研究について述べる。第3章では、対象とする意思伝達の利用されやすい環境の提案とその仕様について述べる。第4章では、提案した

実験環境で収集したデータの分析を行い、どう意思伝達が利用されたかや、ゲームの主目的との関係などについて検証する。第5章では、意思伝達の発生するタイミングについてどう再現し評価したかについて述べる。第6章では、第5章で再現した結果について、機械的評価と人間的評価を行い考察する。第7章は本研究のまとめである。

第2章 関連研究

本章では、ゲーム内で機械同士の意思伝達と、機械と人間との協力に必要な工夫を関連文献を踏まえながら紹介したのち、機械に「人間の行動を誘導するような意思伝達」を獲得させるためにはなにが必要かを考察する。

まず機械同士の意思伝達について紹介する前に、そもそも機械がどの程度の実力を持っているかを説明する。昨今の探索・学習アルゴリズムの発展や計算機器の高速化によって、ゲームをクリアする条件（主目的）を、人間のプロプレイヤーレベルで達成できる人工知能（AI）の作成が可能になった。2015年には、囲碁において人間のプレイデータを利用し学習させたAlphaGoというAIが、プロの棋士にも勝利している。さらに、2017年には、人間のプレイデータすら使わないAlphaGo Zero[8]というAIがそのAlphaGoに勝利している。このように、1対1の対戦ゲームにおいて、人間のプレイデータを用いずとも、人間とまともに勝負できるような強さを持ったAIの作成が可能となってきた。

2.1 機械と人の協力

機械との協力を行うAIの題材に、ハンタータスクというゲームが使われることがある。ハンタータスクとは、操作可能キャラクター2体と目標物2体がいる環境で、プレイヤー2人が操作可能キャラクターをそれぞれ独立に操作し、操作キャラクターをそれぞれ異なる目標物に接触させることを目指すゲームである。なお、操作キャラクターは前後左右の移動のみが可能である。



図 2.1: ハンタータスクのクリア方法1 図 2.2: ハンタータスクのクリア方法2

図 2.1・図 2.2 でいうと青い丸で示した操作キャラクタ P1・P2 が、赤い丸で示した目標物 T1・T2 のいずれか異なる方に接触することでゲームクリアとなる。図 2.1 では P1 が T1 に、P2 が T2 に接触することでゲームクリアとなり、図 2.2 では P1 が T2 に、P2 が T1 に接触することでクリアしている。プレイヤーは、自分がどちらの目標物を狙うかをなんらかの方法で直接的または間接的に伝える、あるいは仲間がどちらの目標物を狙っているかを推測しながら行動する必要がある。

このハンタータスクを用いた研究 [10] で得られた協力プレーは、目標物までの進行ルートを工夫することで、自身がどちらの目標物を狙っているか仲間に伝えるというものであった。P1 が図 2.3 の初期配置を見て、自身の目標物を T1 にした方が、主目的の達成にもっとも効率的と考えた場合を考える。P1 は P2 に自身が思う効率的な方法をなんとかして伝えたい。P1 がなにも工夫せず、最短距離で T1 に近づく（図 2.3 の点線のルート）ことで、その狙いが P2 に伝わる可能性もある。しかし、このルート上の近くには T2 があるため、P2 には P1 が T2 を狙っていると誤って伝わる可能性がある。P2 が P1 の本当の狙いに気付くのは、P1 が T2 を通り過ぎたあたりだろう。そこで、P1 が最短距離のルートではなく、あえて遠回りした T2 を忌避するようなルート（図 2.3 の実線のルート）をとることで、最短距離のルートをたどるより早く P2 に P1 の狙いを気付かせることができる。このように、P1 は目標物までのルートを工夫することで、自身の定めた目標物を仲間に伝えるような協力プレーが可能である。この研究では、段階的に学習の方法をかえることで、こういった協力プレーを再現した。

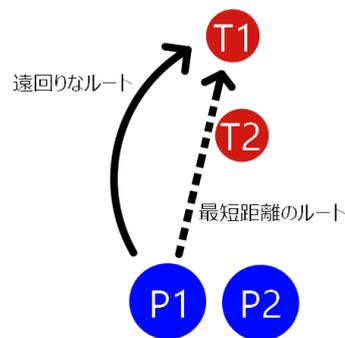


図 2.3: 目標物 T1 を狙うプレイヤー P1 の挙動

また、さらに複雑なゲーム環境で、AIに協力プレーを獲得させている研究がある。AlphaGo Zero を作製した人工知能開発会社 DeepMind は、Quake III Arena のキャプチャー・ザ・フラッグ (CTF) で人間に勝てる AI を作製した [9]。ここで使われた環境は、互いの陣地から旗などの目標物を奪い合う騎馬戦や棒倒しのようなゲームで、敵と射撃戦を繰り返しながら相手陣地にある旗を奪って自分の陣地に持ち帰るといった 2 対 2 のシューティングゲームである。このゲームでは、敵の動きや旗の位置・地形などの要素に加えて、2 対 2 特有の仲間の強さや性格などの多様性まで考慮しなくてはならない。ここで作製された AI は、単体のエージェントを使って強化学習をさせるのではなく、複数のエージェントで相互にプレイすることで多様性のある仲間を考慮した強化学習をしている。この AI で対戦相手としてもチームメイトとしても、人間と混合でゲームを行った結果、人間より高い勝率を得ることができた。また、この AI は人間の参加者より人間に協力的であるとの評価がなされた。

このように、主目的の達成という観点からみれば、複数人でのプレーであっても、人間と勝負できる強さの AI の作成が可能となってきている。しかし、主目的の達成のための手段が複雑であったり、そもそも人間が主目的のために行動していない場合、AI に協力させることはまだ難しく、現在も多くの研究者が研究している。例えば、ポケットモンスターやドラゴンクエストなどの RPG のゲームジャンルでは、特に主目的達成のための戦略の選択肢が多く複雑である。RPG の 1 戦闘でも「安全に」勝つ、「迅速に」勝つ、「アイテムを温存して」勝つなど、目的達成のための手段のうちどれを選ぶべきかという選択肢が生まれる。こういった、1 戦闘で複数の選択肢が生まれるよう、休憩所や敵キャラクタなどの配置を自動生成する研究 [11] もされている。また他にも、課題 A、課題 B、課題 C … という様に複数の課題が同時に与えられ、どの課題から取り組むかの選択が迫られるような場合もある。DeepMind 社は Overcooked というゲームでこの課題に取り組む AI を作る論文 [12] を 2021 年に発表している。

このように、主目的の達成手段が複雑である場合の AI はまだまだ発展途上である。

2.2 機械と人間の違い

第 2.1 節で説明した、人間のプレイデータを用いずに学習させた AI から生まれる協力プレーは、人間の行う全ての協力プレーを網羅しきれていない。再現できているのは、仲間が射撃行動で敵を牽制するのを見たら、自身は敵に気付かれないうようにこっそり敵に近づくとといった、仲間の行う主目的を達成するための行動をくみとり、それに合わせた行動を返すといった一部の協力プレーだけである。

機械だけで学習しても、人間同士のプレイで発生する、一見して主目的を達成するための行動だと分からない行動を再現したり、反応したりするのは困難である。例えば人間が RPG ゲームで、仲間の気付いていないアイテムを取らせるため

に、その近くでジャンプを行うというものがある。仲間にアイテムを取らせるという目的は、主目的達成のためにも有効となりえる目的だが、そのためのジャンプ行動を人間のデータを用いず学習した機械に再現させることは難しい。人間の行動データを用いず学習させ、人間のこういった行動を再現するAIを作成しようとしても、実際に使われている意思伝達行動とは違うものが創出される可能性が高い。

人間独自のコミュニケーションを獲得しようとしても、AI同士で会話させた場合にどんどん人間の理解できない言語を使い始めてしまう様に、機械だけの学習では機械独自の意思伝達方法が確立してしまい、一見して主目的と関係ないような人間の意思伝達の獲得は困難である。人間の意思伝達を再現するには、人間の行動データから学習させる必要がある。

2.3 意思伝達の種類

この節では、人間同士が行う意思伝達を分析するにあたり、意思伝達をいくつかの種類に分け、研究対象とする意思伝達を明確にする。

人間が行う意思伝達には、表情や会話・ハンドサインのようなゲームを通さない行動（ゲーム外行動）による意思伝達と、移動行動や射撃行動・ボイスチャットなどのゲーム内で設計されたゲーム内でおこせる行動（ゲーム内行動）を使った意思伝達と大別できる。オンラインゲームにおいては、知らない人と一緒に遊ぶ機会が多いため、ゲーム内行動のほうを用いての意思伝達が主流である。本研究ではこのゲーム内行動の意思伝達を扱う。また、ゲーム内行動の中でもボイスチャットなどのゲーム外行動でも可能な意思伝達方法は本研究では対象としない。

また、ゲーム内行動で行う意思伝達には、ゲームの主目的に関係性が高いものから、関係性の低いものまで様々である。先行研究 [6] では、意思伝達をいくつかに分類し、主目的との関係性の違いを紹介している。その中でも特に主目的と関係性が高い意思伝達に該当するのは、先行研究 [6] で「注意・警告」と分類される意思伝達である。このグループの代表例として、味方の人間または中立の人間が、自分にとって不利益になるような行動をしようとしたときに、その人間に対して攻撃を仕掛けたり、その人間の進路を妨げる動きをしたりなどが挙げられる。このグループは、相手の行動を注意することで、相手の行動を変更させることが可能で、こういった行動は主目的へと誘導する目的にも使える。

一方、主目的と関係性の低い意思伝達に該当するのは、「謝罪・挨拶」や「挑発」に分類されるグループである。「謝罪・挨拶」のグループは他の人間にアイテムを譲られたときにしゃがみ行動などの行動をすることで、相手への謝意を示したり、ゲーム開始時に互いにジャンプ行動をするなど同じ行動をする（図 2.4）ことで、「よろしく」の意を伝えたりする行動が代表的だ。そして、「挑発」のグループは、格闘ゲームで技が当たらない間合いで弱いパンチ等を繰り返して相手の神経を逆なでしたり、RPGゲームで倒した相手を指さすような行動をして煽ったりという行動（図 2.5）が、代表例として挙げられる。これらの意思伝達は、ゲームの主目的の達成のために別に行わなくてもいい行動であったり、むしろ主目的の達成を妨げる場合が多く、主目的との関係性は低い。

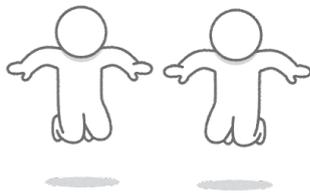


図 2.4: ジャンプ行動での挨拶



図 2.5: 指さし行動による煽り

このように、意思伝達にもゲームの主目的に関係性が高いものから、関係性の低いものまで多種多様である。仲間の行動を、ゲーム主目的の達成に近いものへと誘導するため、本研究ではゲームの主目的と関係性が高い意思伝達を対象とする。

2.4 ゲームにおけるゆとり

先行研究 [7] でもいわれている通り、プレイヤーがゲーム中で主目的以外の行動をするには、ある程度の「ゆとり」が必要になる。また、人間の意思伝達の方法や、意図の種類は、ゲーム内に設定されたゆとりによってその種類や複雑さが変わる。ここでいうゆとりとは、ゲーム内操作可能な行動（行動空間）の幅や、次の行動を考える時間的余裕のことである。

まず、行動空間の幅について説明する。例えば、オセロや将棋など、人間に可能な操作が1ターンに1度、駒を動かす事だけだった場合、伝えたい意図がどんなものであれ「駒を動かす」以外の手段では意思伝達を行うことができない。意思疎通の手段が絞られるために、相手に伝えることのできる意思も単純なものに限られてくる。一方、マイクラフトのような3Dゲームの場合、移動することの他に花などの特定のアイテムをゆずることや、弓による射撃攻撃、しゃがみ行動、ジャンプ行動など、人間が操作できる行動の種類が多い。このことにより、マイクラフトではオセロや将棋よりもはるかに意思伝達の種類が多くなり、必然的に複雑な意思でも伝達可能となる。このように行動空間の幅が大きいと、意思伝達で伝えられる意思が多くなる。

次に、時間的余裕について説明する。時間的余裕とは、意思伝達をすることが可能なゆとりのある時間のことである。この時間的余裕の多い少ないによって、それぞれ意思伝達の種類が異なったり、その発生頻度も異なる。

これを説明するための一例として、3DシューティングゲームであるAPEXを例に挙げて説明する。このゲームにおいて、時間的余裕が最も少ない時間が戦闘中で、最も多いのがセーフルーム内で待機する時間、その中間に位置するのが移動中やアイテムの探索をする時間と考えることができる。戦闘中であれば、基本的に味方と詳細な意図を伝える意思伝達が使いにくくなるとともに、意思伝達の方法は単純化され、主目的と関係性が深いものばかりになる。そして、拮抗した戦いであるほど、意思伝達自体の発生頻度が下がる。また、セーフルーム内で待機する時間であれば、アイテムを仲間の足元に落として射撃することでアイテムをゆずるといった意思伝達を行ったり、チャット機能などを使って人間の言葉で雑談を行ったりできる。このように、時間的余裕が大きければ、複雑な意思伝達を行えるようになるとともに、主目的とは関係が薄いものも増える場合が多い。一方、移動中やアイテムの探索中であれば、チャットなどは使わずに、探索をしながら、射撃やジャンプ運動によって探索するルートをコンパクトに伝えるなどの意思伝達が行われる。

以上のようにAPEXという一例では、時間的余裕が小さければ主目的に関わる意思伝達の割合が増えるかわりに、意思伝達自体の発生頻度が下がる。逆に大きければ、意思伝達自体の発生頻度が上がるかわりに、主目的に関わる意思伝達の割合が減る傾向がある。このように、時間的余裕は、意思伝達の種類と発生頻度に大きく関わり、こういった関係は多くのゲームにも当てはまる場合が多い。

第3章 対象ゲームの考察と提案

本研究では、人間に対しその行動を誘導出来るようなゲーム内の意思伝達の分析と再現を目的としている。それにあたり本章では、どのような意思伝達を対象とするかと、その意思伝達の発生しやすい実験環境について考察し、そこで得られた意思伝達をどう分析をするかを提案する。

3.1 対象とする意思伝達

人間はゲーム内で常にゲームの主目的のみを目的として遊んでいるわけではない。RPGゲームであえてゲーム内で弱いとされる最弱装備のみを使ってゲームのクリアを目指したり、レースゲームのゴール直前であえて停止して他の人間を煽ったりといったことをして人間は楽しむ。ゲームの主目的に関わる・関わらないを問わず、人間はこのようにゲームの主目的以外の遊び要素を見つけてゲームを楽しむ場合がある。意思伝達も、ゲームのクリアに関わるもの・関わらないものと分けて考える。

先行研究 [6] では、意思伝達を第 2.3 節でも触れた「謝罪・催促」・「挑発」を含む 7 つに分類し、ゲームの主目的と関係性が強いもの弱いものまで紹介している。特に、「注意・警告」のグループに属する意思伝達はゲームの主目的と関係性が強い。本研究では、ゲームの主目的とは関係性が低い、「謝罪・催促」・「挑発」などの意思伝達については対象としない。初心者ゲームのクリアに導くためという目的にとって重要度が高いのは、ゲームの主目的と関係性が強い「注意・警告」などの意思伝達の方である。よって、本研究では「注意・警告」に該当する、あるいは同程度にゲームの主目的に関わる意思伝達に絞って考察する。

3.2 意思伝達の発生する環境

現代では、複数人でのプレイが可能なゲームがたくさんある。それらのゲームは、主目的達成という面で意思伝達の重要度の高さが異なったり、発生頻度が異なったりする。本研究では、意思伝達が総じて主目的達成に関わっており、かつ発生頻度が高いような環境を選択または作成する。そのために本節では、ゲームの主目的に関わる意思伝達の発生する環境の条件を考察する。

「人間の行動を誘導する」ような、「ゲームの主目的に関わる意思伝達」が発生するのに必要な環境には、以下の条件が重要だと考えている。

- (a) 人間プレイヤーの持つ情報に差があること
- (b) 仲間の行動が見やすいこと
- (c) 行動空間の幅と時間的余裕があること

このそれぞれについて説明していく。

3.2.1 人間プレイヤーの持つ情報に差があること

本研究ではゲームの主目的と関係性が強い意思伝達に絞って考察する。こういった、ゲームの主目的と関係性が強い意思伝達は、プレイヤー間の持つ情報に差があるという状況で起こりやすい。シューティングゲームなどでよくあるのは、仲間が味方より先に敵を見つけた場合である。

図 3.1 を使って説明する。図中の Player1 は人間プレイヤー 1，Player2 は人間プレイヤー 2 を示し，両者とも敵キャラクターを殲滅することをゲームの主目的として動いている。Player1 は壁の陰から周りをうかがっており，人間プレイヤー 1 より先に Enemy で示す敵キャラクターを発見している。敵キャラクターはプレイヤーの存在には気づいていない。この状況は同じ主目的を持つ人間プレイヤーの持つ情報に差が出ていると言える。もし，このまま人間プレイヤー 1 が敵キャラクターの存在を人間プレイヤー 2 に伝えずにいれば，プレイヤー 2 は敵キャラクターの存在に気付かずに壁の向こう側に行って，敵キャラクターに気付かれてしまうなどして，せっかくの殲滅のチャンスをふいにしてしまう可能性が出てくる。こういった，ゲーム内での人間プレイヤーの情報の差がゲームの主目的の達成を阻害してしまう可能性がある。こういった状況だと，プレイヤー 1 がそのゲームに許された射撃行動やしゃがみ行動などのゲーム内行動で，敵キャラクターの存在を警告し，人間プレイヤーの同士の持つ情報の差を埋めようとする事が知られている [6]。このように人間プレイヤーの持つ情報に差を作ることで，効果的にゲームの主目的と関係性が強い意思伝達の発生を促すことができる。

また，こういった図 3.1 のように壁を使う方法のほかに，特定のプレイヤーしか見ることのできないアイテムを用意するなどの方法でも，プレイヤーの情報に差を付けることができる。

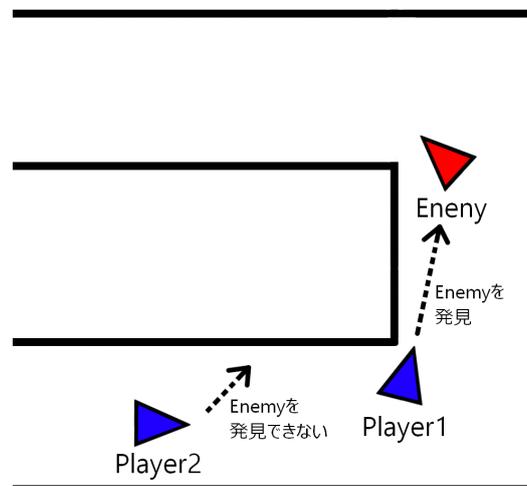


図 3.1: プレイヤの持つ情報の差

3.2.2 仲間の行動が見やすいこと

次に、(b) に関して説明する。ゲーム中には、操作している人間には認識できない“死角”が存在することが多くある。例えば、3Dシューティングゲームには多くの“死角”が存在する。図 3.2 を使って説明する。図 3.2 では、操作キャラクタ P1 と仲間である操作キャラクタ P2・P3 がゲーム上で遊んでいる。図 3.2 の右上の濃い灰色の部分には壁 (WALL) に阻まれて P1 には見えないものとする。P2・P3 が P1 に見えない重要なアイテムを見つけた時、そのアイテムを P1 に伝えられるのは P2 のみで、P2 は屈伸行動やジャンプ行動などのゲーム内で認められた行動で P1 に意思伝達を行うことができるが、P3 が同様の行動を行っても P1 に伝えることができない。このように、意思伝達を伝えるために、(b) の仲間の行動が見やすいことは必要な条件である。

仲間の行動を見やすくする方法の一つとして、単純にゲーム中の壁 (図 3.2 でいう WALL) をなくす手法があるが、これらの方法は第 3.2.1 節で説明した「人間プレイヤーのもつ情報」に差を付けづらくなる場合がある。そのため、情報の差を作りつつ仲間の状況を見やすくするために、特定プレイヤーしか見えないアイテムを作る方法と壁をなくす方法を併用するなどの工夫が必要である。また、レースゲームなどにある、ミニマップ (ゲーム全体の俯瞰図をサブカメラとして画面に写す方法) などの別カメラを使って仲間を映す方法もある。

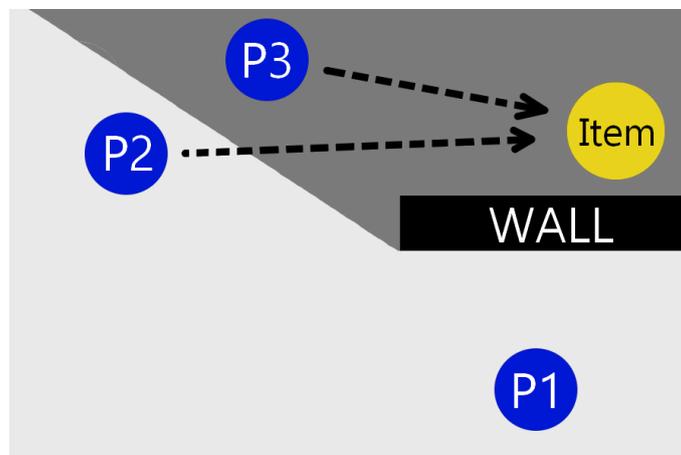


図 3.2: 死角の発生例

3.2.3 行動空間の幅と時間的余裕があること

最後に (c) に関して説明する。第 2.4 節でも触れたが、行動空間の幅と時間的余裕は、意思伝達の種類と発生頻度に大きく関わる。

一つ目の行動空間の幅は、意思伝達の方法を多く作成するために必要な要素である。この幅が大きいほど、様々な意思伝達方法が生まれる。もし研究の対象が、意思伝達方法がどう創発されるか、どんな行動が意思伝達方法として成立するかを分析するなら、この行動空間の幅を大きくすることが必要である。しかし、本研究は第 1 章で記した通り、「人間らしい意思伝達方法の分析と再現」は行わず、実験中の意思伝達の方法は指定するため、あえて行動空間の幅は過剰に大きくする必要はない。

一方、二つ目の時間的余裕の作り方については、ゲーム環境によって調整する必要がある。第 2.4 節で触れた 3D シューティングゲーム APEX においては、時間的余裕が小さければ主目的に関わる意思伝達の割合が増えるかわりに、意思伝達自体の発生頻度が下がる。逆に大きければ、意思伝達自体の発生頻度が上がるかわりに、主目的に関わる意思伝達の割合が減る傾向がある。この傾向のあるゲームでは、主目的に関わる意思伝達が十分発生するように、大きすぎず小さすぎず適度な時間的余裕を与えるよう設計すべきである。時間的余裕を調整する方法はいくつかあり、対戦ゲームであれば、敵キャラクタの AI の強さを変更することでこの時間的余裕を調整したり、ゲーム中にエイム補正などのハンデを調整して入れたりなどが有効である。

3.3 実験環境の説明

本研究では、第3.2節を踏まえて、APEXなどのシューティングゲームを簡略化した三人称視点の2人協力型2Dシューティングゲームを自作した。このゲームは、持ち時間内になるべく多くの得点（GP）を集めることを主目的としたゲームである。ゲーム内にはいくつかのエリアが配置され、その一部のエリアに敵キャラクターが配置されている。プレイヤーは、1つのエリアに存在する全ての敵キャラクターを倒すことでGPを加算することができる。本節では、この実験環境について詳しく説明する。

3.3.1 操作キャラの動作

本環境は二人の人間で行ってもらうゲームで、それぞれ割り当てられた操作キャラクター（P1またはP2）を操作する。P1とP2には、移動・方向転換・射撃・必殺技の行動が可能である。また、P2にのみ以下の行動が可能とした。

- 意思伝達ボタン1（回転運動）
- 意思伝達ボタン2（前後運動）

これらの意思伝達ボタンは、P2がP1に行う意思伝達として使ってもらう。ボタンそれぞれが何を意味するかについては、「戦闘中のボタン1は“助けて!”を意味する」などと実験前に被験者に伝え、それ以外の使い方を禁止する。これは、第1章で述べた通り、人間のどの行動が意思伝達行動で、どういった意図で使われたかという分析を省き、意思伝達を行うタイミングのみを観察するためである。また、P2のみに使用可能とした理由は、意思伝達の送り手・受け手を明確にすることで、「人間を誘導する」方の人間を固定できるからである。

また、操作キャラクターにはパラメータとして体力（HP）と強化値（SP）が設定されている。HPは敵キャラクターの射撃行動によって撃たれた弾に被弾した時に減少し、0になった時点で一定時間行動不能になる。SPはゲーム中に得られる強化アイテムを取得することで増加し、増加具合によって射撃の威力が変化する。また、SPを消費することで、必殺技を使うことができる。なお、このSPの値はHPが0になった時点で初期化される。

3.3.2 ステージの説明

次に、ゲームステージについて説明する。ゲームステージは正方形のエリアが3行に並んでおり、1・3行目に7つ、2行目に8つ並んでいる（図 3.3）。エリアには敵キャラクターのいない安全エリアと敵キャラクターのいる戦闘エリアが存在する。2列目全てが安全エリアで、1・3列目は戦闘エリアである。人間は、安全エリアでどの戦闘エリアに挑戦するかを選択し、各戦闘エリアの敵キャラクターを全て倒すことで、各戦闘エリアに設定されたGPを獲得できる。安全エリアと戦闘エリアの間にはそれぞれ扉が存在し、P1とP2の両方が近づくこと（P1とP2の同意）で扉が開き、戦闘エリアに侵入することができる（図 3.4）。

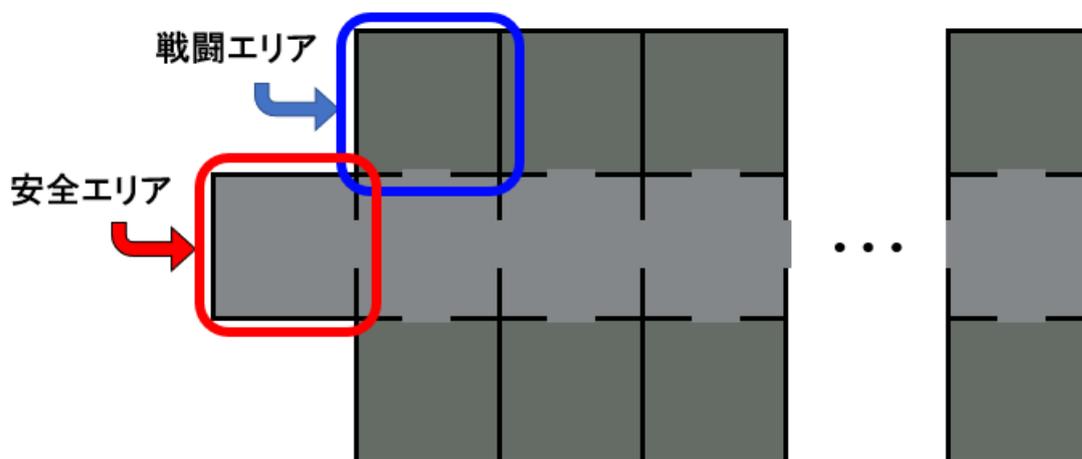


図 3.3: ゲームステージの構造

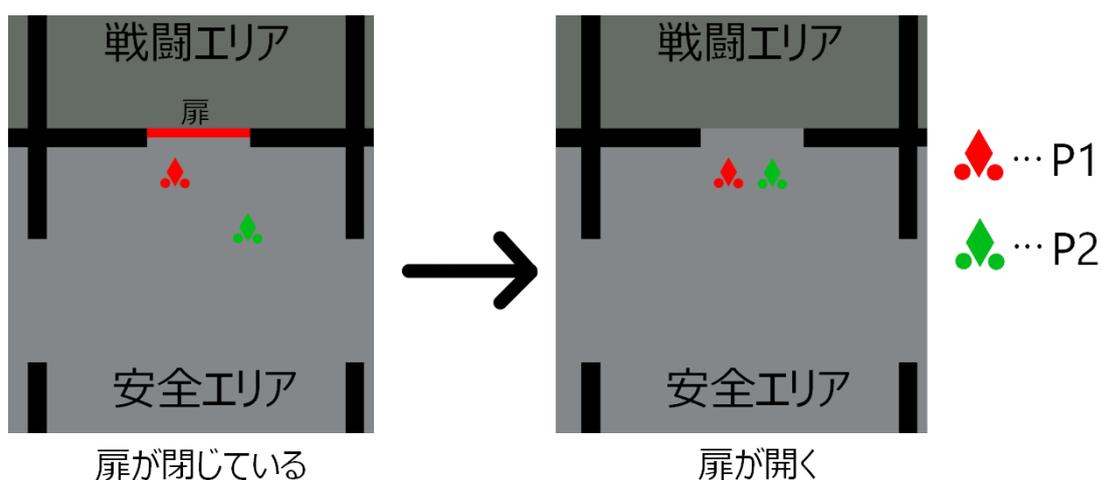


図 3.4: 安全エリアと戦闘エリア間の扉の開閉

また、各戦闘エリアに設定された GP は一定ではない。この GP の割り当てを複雑にした理由については、後の第 3.4.1 節で説明する。基本的に戦闘エリアで得られる GP は 1 点、2 点、...、7 点と右に行くほど高くなる。しかし、上下 1 列に並んでいる戦闘エリアには「はずれ」と「あたり」が 1 つずつあり、「はずれ」を攻略した場合に得られる GP は 0 点となる。そして、どちらが「あたり」かは、P1・P2 の協力によって始めて知ることができる。これは、P1 には誤りを含んだ「はずれ」・「あたり」の並びが伝えられ、P2 には P1 に伝えられた情報のうち、どの列の GP が間違っているかを教えることで実現している。例えば、図 3.5 の上図のように本当の GP が並んでおり、P1 の視点では図 3.5 の左下図のように見えているものとする。P1 の視点には、2 点の列と 7 点の列で誤った情報が与えられている。一方 P2 には 2 点と 7 点の列に誤りであることを示すマーク（図 3.5 の右下図の「R」）が描画される。このように、P1 だけ、P2 だけの情報では「あたり」のステージがどれかは分からず、意思伝達行動により情報を共有することで、はじめて「あたり」がどちらかを知ることができる。なお、これらの「あたり」・「はずれ」はいずれか 1 つの戦闘エリア攻略後に更新される。また、戦闘エリア攻略後は、一定時間後に敵キャラクターが出現し、再攻略可能になる

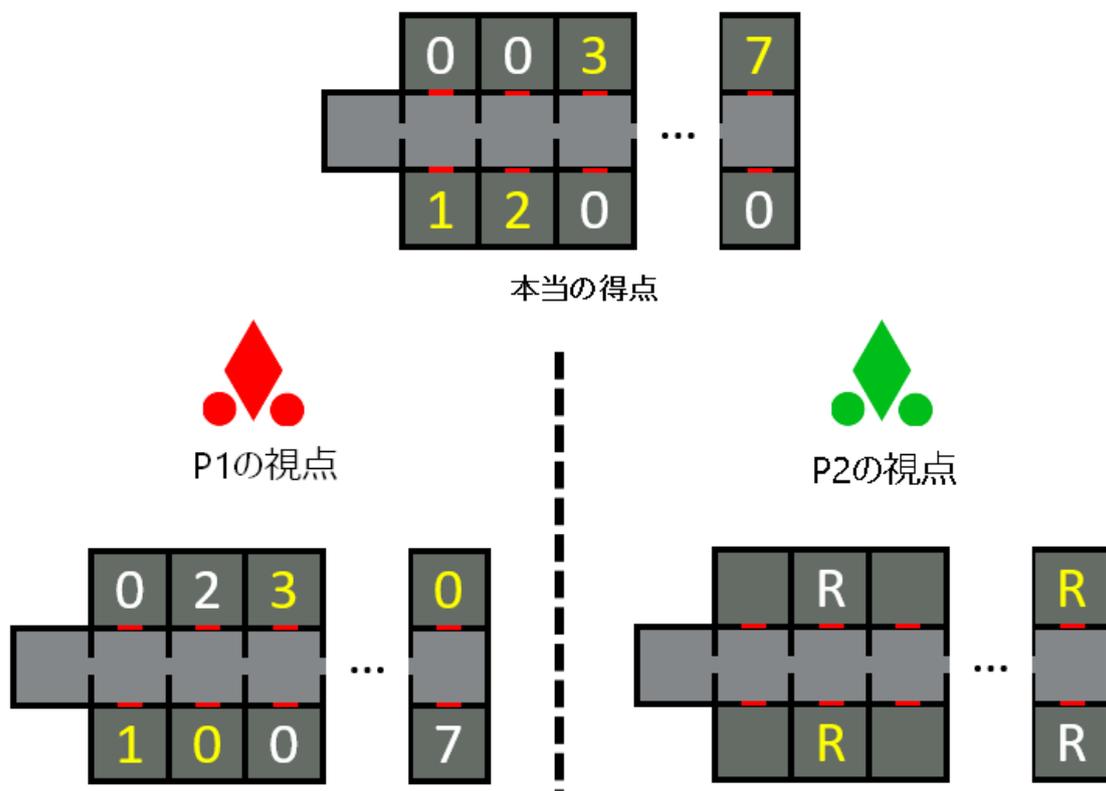


図 3.5: 戦闘エリアの得点 (GP) と各プレイヤーの情報の差

3.3.3 敵キャラクタ

戦闘エリアにいる敵キャラクタには2種類（E1 と E2）存在し、それぞれP1の射撃のみを受け付けるE1、P2の射撃のみを受け付けるE2がいる。また、それぞれの敵は2種類の銃弾をランダムに切り替えて射撃を行う。これら2種類の銃弾はそれぞれP1・P2の射撃でしか相殺できない。ただし、P1・P2の行う必殺技に関しては両方の敵・銃弾に対し有効である。

また、戦闘エリア内にいる敵キャラクタは、得られるGPが高いエリアに行くほど（図3.3の左から右に行くほど）強くしている。例えば、GPが1点しかもらえない戦闘エリアでは、敵キャラクタがその場から動かない。この1点の戦闘エリアでは、図3.6のP1から伸びる点線のように、壁を利用して射撃を行えば、比較的簡単に敵を殲滅できる。また、GPが3点もらえる戦闘エリアでは、図3.7のE1・E2のように一定経路を動き回る。このように、敵キャラクタの強さは得られるGPが高いほど強くなる。また、これ以外にも、敵の数を増やしたり、射撃の頻度を増やしたり、射撃に追尾性能を付けたりと、得られるGPが高いほど戦闘エリアの攻略難易度が上がるよう設計されている。

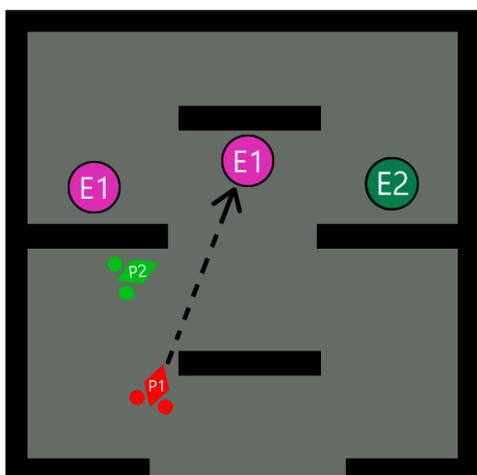


図 3.6: 移動しない敵キャラクタ

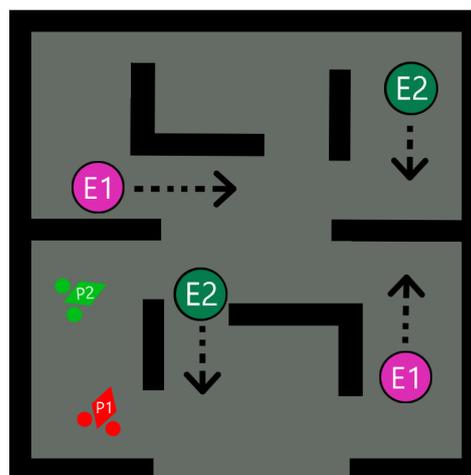


図 3.7: 移動する敵キャラクタ

3.3.4 プレイ画面について

本節では、人間へのプレイ画面の見せ方について説明する。プレイ画面は1000×1000画素で表示されている。また、プレイ画面中にはメインカメラとサブカメラによって描画されている箇所がある。まずこれらのカメラについて説明する。例えば、P1・P2のゲーム内での配置が図3.8の左図のような配置だとする。この時、P1を操作する人間のプレイ画面には、メインカメラ（図3.8の右上図）とサブカメラ（図3.8の右下図）が表示されている。メインカメラにはP1とその周囲の様子（E1・E2など）・P2のいる方向を描画し、サブカメラにはP2の挙動及びP1がP2のどちら方向にいるかを描画する。なお、メインカメラ・サブカメラはカメラ中心を常にP1・P2の重心に合わせているわけではなく、わずかに遅れをもってP1・P2の重心に追従している。この二つのカメラを実装することで、人間は仲間の操作キャラ（P2）の挙動、特に意思伝達ボタン1の回転運動と意思伝達ボタン2の前後運動を常に目にすることができる。つまり、仲間の意思伝達行動を見逃しづらくなる。また、メインカメラには仲間のいる方向（図3.8の左図のP1・P2間の棒）を描画することで、仲間の位置を把握しやすくしている。

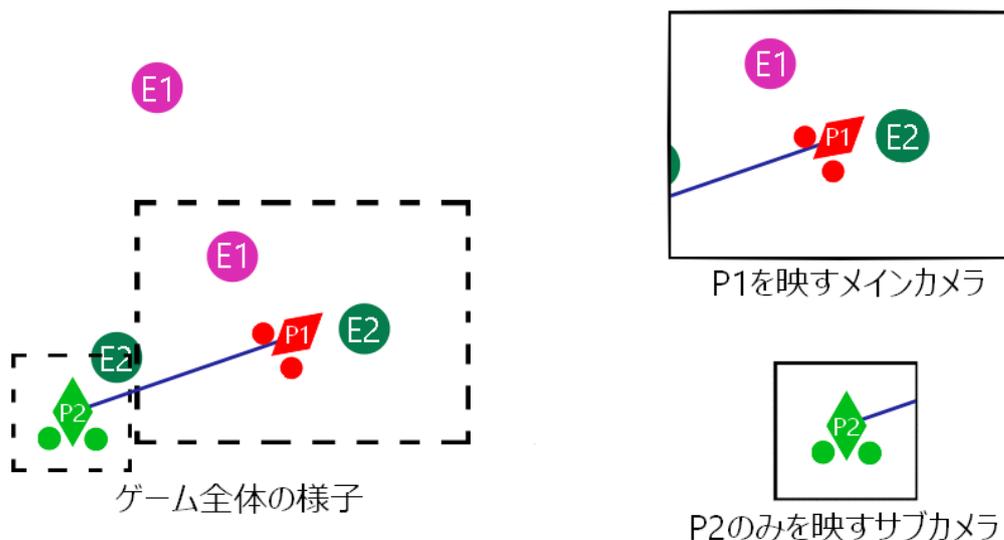


図 3.8: メインカメラとサブカメラ

人間に見せるプレイ画面は図 3.9 の赤枠で区切っているように、4つに分割されている。「メインカメラ」と「サブカメラ」・「意思伝達ボタン1の使い方」・「意思伝達ボタン2の使い方」の4つである。メインカメラには、前述した通り操作キャラクター（図 3.9 では P1）を中心とした周囲が表示される。また、サブカメラには仲間の操作キャラクターの行動が表示される。このうち、メインカメラには加えて、第 3.3.5 節で説明する持ち時間と、これまで得た得点数（GP）が表示される。一方、「意思伝達ボタン1の使い方」・「意思伝達ボタン2の使い方」には、ゲーム中の各状況での意思伝達ボタンに設定した意味を表示し、被験者実験中に人間が意思伝達ボタンの使い方を誤らないようにしてある。

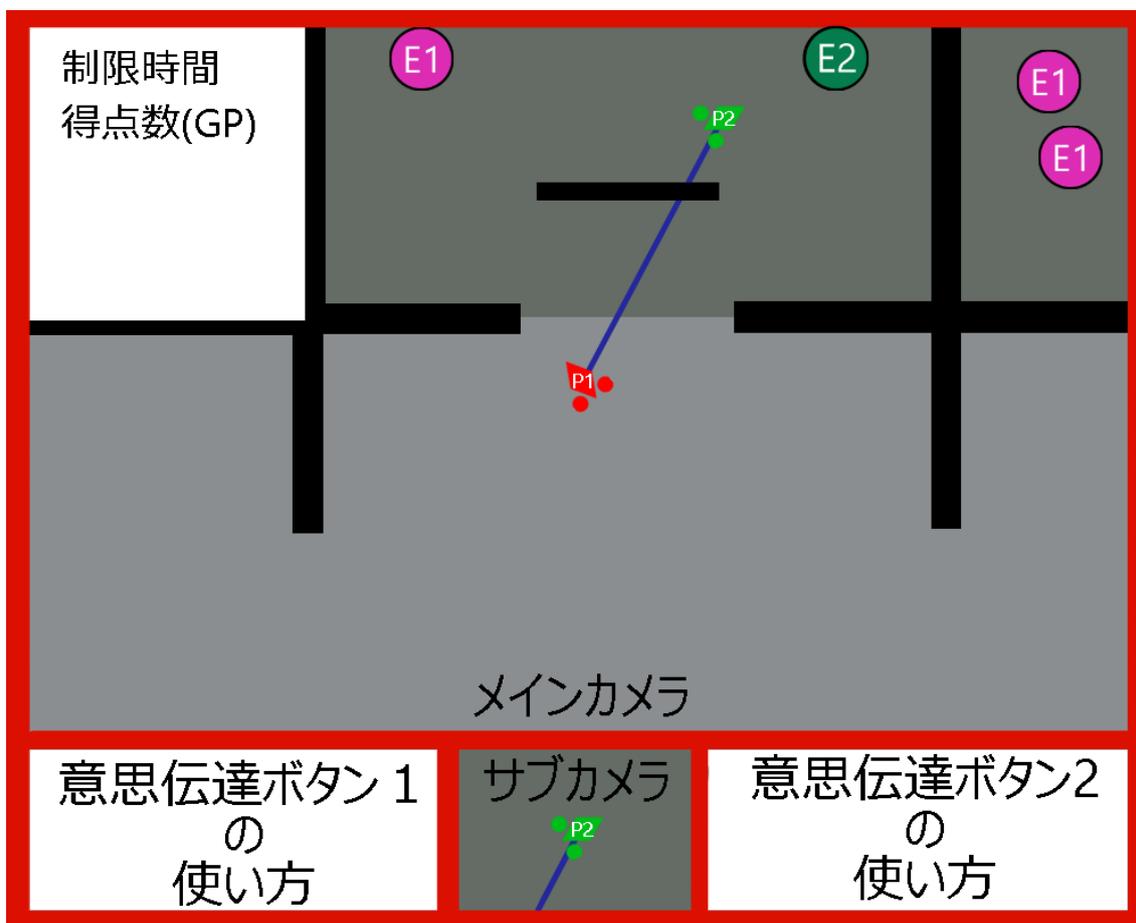


図 3.9: プレイ画面の配置

3.3.5 その他

ここでは、以下の2つについて説明する。

- 強化アイテム
- 持ち時間

まず1つ目に強化アイテムについて説明する。強化アイテムは、第3.3.1節でも少し触れたが、P1・P2のパラメータSPを強化することができ、射撃強化や必殺技に使うことができる。この強化アイテムはゲーム開始直後に5つ出現し、それとは別に戦闘エリア攻略後に当該エリアの出入口に1つ出現する。人間はこのアイテムを分け合いながらゲームを進行していく。

次に、持ち時間について説明する。プレイヤーには持ち時間（200秒）が与えられている。この持ち時間は、ゲーム中常に減り続けるわけではなく、戦闘エリアでP1・P2が戦闘を行っている時間でのみ減少する。人間は、この持ち時間が0になるまでの間に攻略エリアを攻略し、出来るだけ高い合計GPとすることを主目的としてゲームを行ってもらう。

3.4 分析する意思伝達

本節では、この環境を使ってどうやって「人間らしい意思伝達の利用方法の分析」のためのデータ収集を行うかを説明したのち、どういった意思伝達を分析するかについて説明する。

第3.3.1節でも説明したが、本実験ではP2のみ意思伝達ボタン1と意思伝達ボタン2を使用ができる。ボタンが使用されている間、P2は回転運動・前後運動を行う。また、本実験ではこれらのボタンを意思伝達行動として扱ってもらう。実験前にどういう意思伝達として意思伝達ボタンを使うかを指定することで、「人間の意思伝達方法」の分析を省き、「人間らしい意思伝達の利用方法」の分析を行う。また、意思伝達ボタンを使える場面ではなるべく使うよう指示を出した。

次に、どういった意思伝達を分析するかについて説明する。まず本実験に使用した環境には大きく分けて以下の3つの状況が存在する。

- ステージ選択中
- 戦闘中
- アイテム出現中

ゲーム中では主にこの順で状況が遷移する。このそれぞれについての説明と、各状況での意思伝達ボタンに割り当てた、意図について説明する。

3.4.1 ステージ選択中

1つ目の状況として「ステージ選択中」を説明する。これは、ゲーム中にP1とP2が次にどの戦闘エリアに挑戦するかを相談する時間である。ここでは、直前に挑戦した戦闘エリアでの消耗具合や、仲間の強さの様子を見て次に挑戦する戦闘エリアを決めたり、その後に第3.3.2節で説明した「はずれ」の戦闘エリアに誤って挑戦しないよう仲間と相談する。

戦闘中でない限り持ち時間も減少しないし、敵のいない安全エリアでの相談なので、この状況での時間的余裕はかなり多い。また、第3.3.2節で説明した通り、戦闘エリアと安全エリアの間には双方が近づかないと開かない扉がある。これにより、片方が仲間の意向を無視して挑戦エリアを決めることが出来ないし、逆に強制的に戦闘エリアに挑まされるといった、双方の同意なしで状況の遷移がおきにくい。

この状況での意思伝達ボタンの割り当ては以下の通りにした。

意思伝達ボタン1（回転運動）

この得点（GP）のエリアに挑戦しよう

意思伝達ボタン2（前後運動）

このエリアに移動しよう

前者は、安全エリア内でボタンを押した時に、その安全エリアに隣接する戦闘エリアのどちらかに挑戦しようという提案をする（次ページ図3.10）。一方後者は前後運動で指示した先のエリアに移動しようということを示す。例えば、次ページの図3.11左図では、下の戦闘エリアにむかって前後運動しているので、下の戦闘エリアに挑戦しようという提案をしている。また、図3.11右図では、左の安全エリアにむかって前後運動しているので、左の安全エリアに移動しようという提案をしている。以上のように前後運動は、戦闘エリアまたは安全エリアも指し示すことで、該当エリアへの移動を提案する。

「ステージ選択中」はお互いに相談し合いながら、どの得点（GP）の戦闘エリアに挑戦するかを決める。また、第3.3.2節で説明したとおり、エリアには「はずれ」・「あたり」があり、これらの情報はP1・P2の協力によってはじめて正確に知ることができる。このように各エリアで得られるGPを設計したのは、「あたり」の挑戦エリアを見つけるための意思伝達をする必要性を高めるためである。例えば、P1には「あたり」に見えるが本当は「はずれ」である戦闘エリアにしようとした時、P2にはその戦闘エリアが本当に「あたり」であるか本当は「はずれ」かをP1に教える必要がある。「あたり」の場合、P1により沿うだけでいいが、「はずれ」の場合、本当の「あたり」を意思伝達ボタン2を使って指し示す必要がある。これを怠るとGPを得られなくなる確率が高くなる。

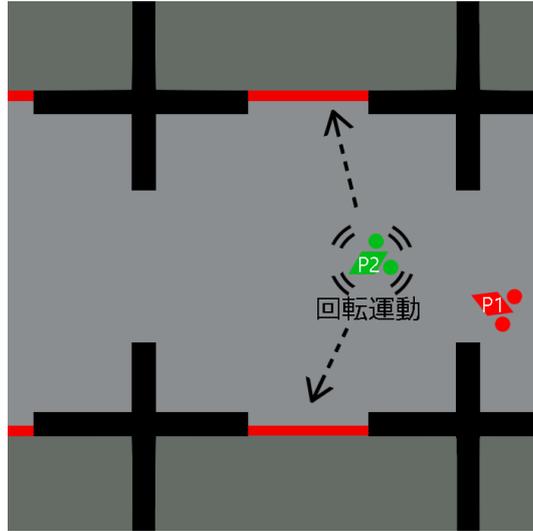


図 3.10: ステージ選択中の回転運動の意味

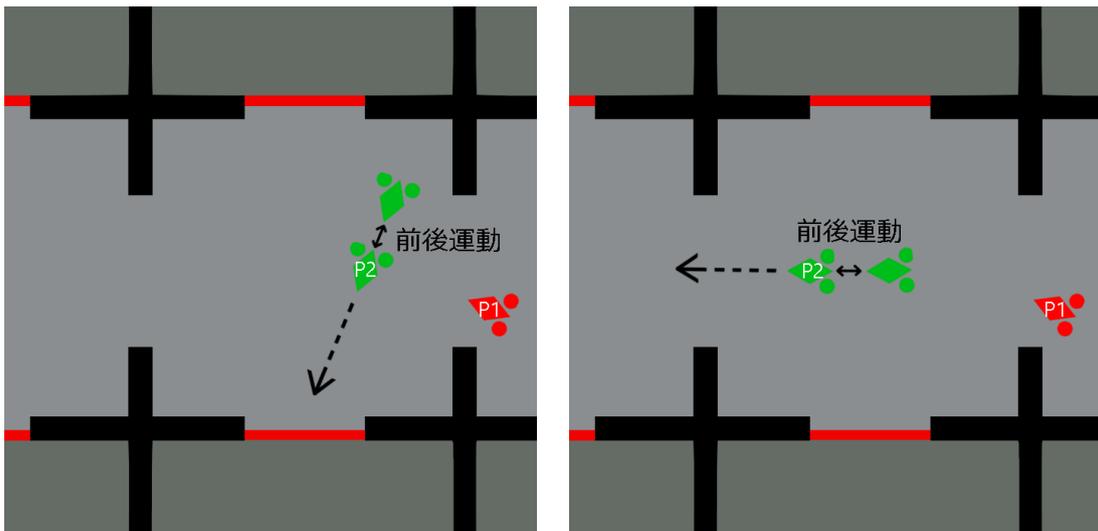


図 3.11: ステージ選択中の前後運動の意味

3.4.2 戦闘中

次の状況として、「戦闘中」を説明する。「戦闘中」は戦闘エリア内でP1・P2が敵キャラクターと戦闘をしている状況を示す。この状況においては「ステージ選択中」や「アイテム出現中」の状況とは違い、持ち時間が徐々に減少していくため、時間的余裕が少ない。また、戦闘エリア内には片方の操作キャラクターの方が倒しやすい敵が存在するので、お互いに守り合って攻略する必要がある。このことにより、ただ敵キャラクターを用意するよりもさらに仲間の動きを注視するようになっている。

この状況での意思伝達ボタンの割り当ては以下の通りにした。

意思伝達ボタン1（回転運動）

助けて

意思伝達ボタン2（前後運動）

逃げよう

前者は、自身が敵キャラクターに追い詰められたときに行う行為で仲間への救援要請を示す。後者は、現在挑戦している戦闘エリアへの攻略を諦め、他の戦闘エリアへ挑戦しようということを示す。

プレイヤーの実力以上の戦闘エリアに挑戦してしまった場合や、HPが0になったことによるSPの初期化がされた場合、該当の戦闘エリアの攻略は困難になる。こういった場合、どうしても撤退の指示が必要である。また、仲間が自身のピンチに気付いていない時など、気付かせる目的で「助けて」と伝えたい場合もある。このように、「戦闘中」でも意思伝達を行いたいと考える機会がある。

3.4.3 アイテム出現中

最後の状況として、「アイテム出現中」について説明する。この「アイテム出現中」は強化アイテムが出現している状況を示す。なお、強化アイテムは第3.3.5節で説明した通り、ゲーム開始直後に5つ出現し、また1つの戦闘エリア内の敵キャラクター殲滅後に1つ出現する。この状況は、持ち時間の現象が起きないため「戦闘中」より時間的余裕がある。この状況は「ステージ選択中」と違って、強化アイテムを片方の人間がもう片方の人間の同意なしに取得することができるといった、双方の同意なしでの状況の遷移が起こる。

この状況での意思伝達ボタンの割り当ては以下の通りにした。

意思伝達ボタン1（回転運動）

強化アイテムを頂戴

意思伝達ボタン2（前後運動）

強化アイテムをあげる

前者は、自身に強化アイテムを取得させてくれと仲間に示している。一方後者は、仲間に強化アイテムをゆずるということを相手に示す。

強化アイテムは、操作キャラのパラメータであるSPの値を増やす。このSPは、射撃の強化や必殺技に使われるため、強化アイテムはゲーム中でかなり重要な要素である。また、戦闘エリア内にはP1・P2片方の射撃しか受け付けない敵キャラクターが存在するため、プレイヤーの片方だけが強化アイテムを取得しても戦闘エリアをクリアしづらい。一方、人間間で実力に大きな差がある場合、弱い方の人間の操作キャラクターに、なるべく強化アイテムを取らせて強くするなどの戦略も存在する。このように、お互いの実力や現在のSPの値を見て、強化アイテムを均等に分けるか偏らせるか決めながら分け合う必要がある。この分け合いに、意思伝達ボタンを使用する。

第4章 被験者実験の傾向と分析

第3.3節で説明した実験環境を構築し、それを使ってデータ収集を行った。本章では、冒頭でどうデータ収集をしたのかを説明し、データの分析を行う。

まず、どうデータ収集をおこなったかについて説明する。4人5グループの20名で、第3.3節で説明した実験環境を用いデータ収集を行った。なお、1グループ内でペアを変えながら行ってもらった。1グループ4人の組み合わせは6通りあるのため、1グループからのべ6ペア分のデータが収集できた。収集できたペアの数は5グループ合わせて30ペアであった。また、1ペアの中でP1・P2を入れ替えてプレイしたので、その入れ替えた場合のペアを異なるペアと考えると、合計60ペアの行動データを収集できた。それぞれのペアで何回かゲームを行い、収集できた試合数は234試合となった。本章ではこのデータを使い、意思伝達の分析を行う。なお、学習データの統計を表4.1を示す。また、1試合での各状況の平均時間を表4.2に記す。

データセット	
被験者	20人
ペアの総数	60ペア
総試合数	234試合
総プレイ時間	約1351分

表 4.1: 学習データの統計

「ステージ選択中」	「アイテム出現中」	「戦闘中」	合計
121.50 秒	25.04 秒	200.00 秒	346.54 秒

表 4.2: 1 試合中の各状況の平均時間

4.1 意思伝達の各使用頻度

各意思伝達のボタンの使用頻度は図 4.1 のようになった。上段 3 つが意思伝達ボタン 1 の使用頻度で、下段 3 つが意思伝達ボタン 2 の使用頻度である。また、この 3 列のうち、一番左が「ステージ選択中」で、2 列目が「アイテム出現中」、3 列目が「戦闘中」の状況を示す。図 4.1 に示すように、左の列の「ステージ選択中」では、全く行わないものから 4 回以上行ったものまであり、ペアによって大きな違いがある。一方、右の列の「戦闘中」では、全く行わないペアが半数以上あり、「ステージ選択中」ほど大きな違いはなかった。このことから、状況により意思伝達の使用頻度に大きな違いがあることがわかる。このことを踏まえて、それぞれの状況でなぜここまで使用頻度が異なるのかを考察する。

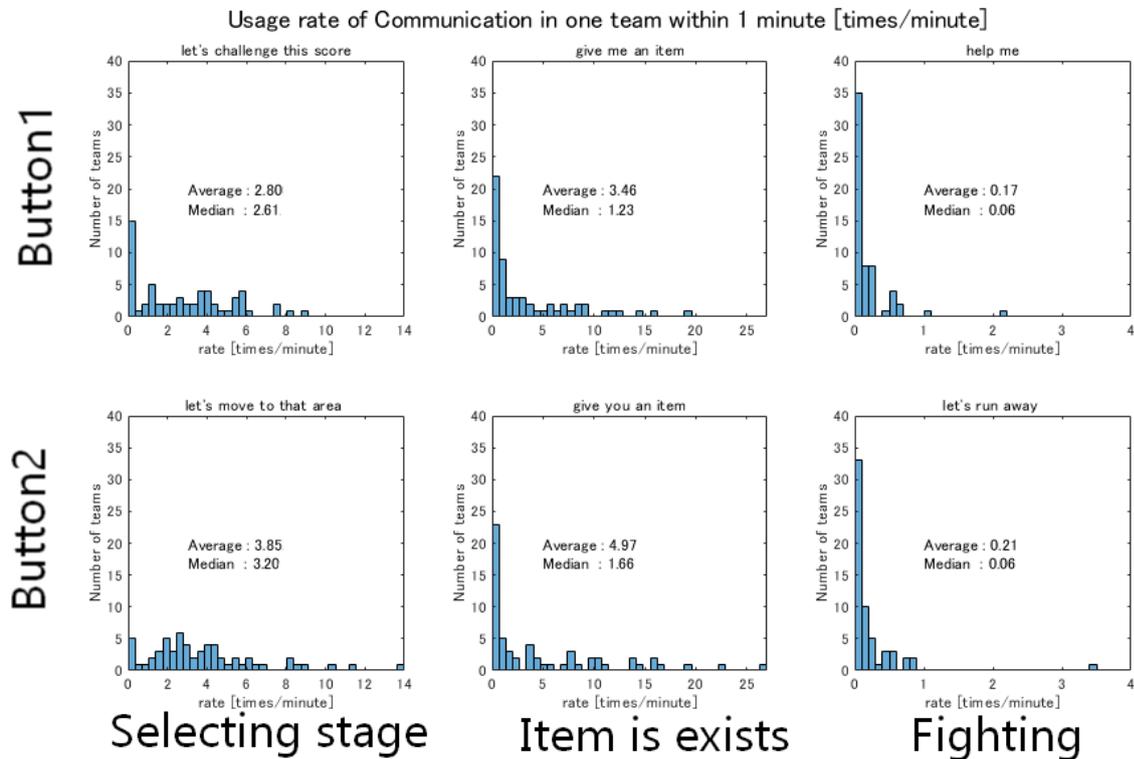


図 4.1: 意思伝達ボタン使用頻度 [回/分] のヒストグラム

	「ステージ選択中」	「アイテム出現中」	「戦闘中」
意思伝達ボタン1	2.61	1.23	0.06
意思伝達ボタン2	3.20	1.66	0.06

表 4.3: 意思伝達ボタン使用頻度の中央値 [回/分]

表 4.3 は、図 4.1 から得られた各使用頻度の中央値である。意思伝達ボタン 1・2 の使用頻度は「ステージ選択中」が一番多く、「戦闘中」が一番少なかった。この実験環境では、「ステージ選択中」の方が時間的余裕が多いため、こういった結果が得られたと考察できる。

次に、1 試合中の意思伝達回数について分析する。図 4.2 は左から順に、「送り手ごとの 1 試合中の意思伝達回数の平均」・「受け手ごとの 1 試合中の意思伝達回数の平均」・「ペアごとの 1 試合中の意思伝達回数の平均」である。いずれのヒストグラムでも意思伝達の回数のバラつきが大きく、人によって意思伝達をするしないの差が激しかった。このように、意思伝達を多く行う人間、あまり行わない人間とがいることがわかった。よって、「誰がゲームをしていたか」は意思伝達回数の予測に大きく関わる。

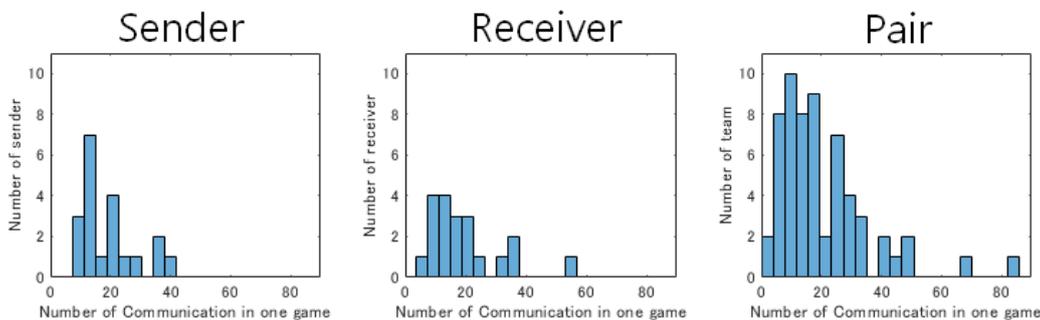


図 4.2: 送り手・受け手・ペアそれぞれでの 1 ゲーム中の意思伝達回数の平均

4.2 意思伝達の連続使用時間

本節では、意思伝達ボタンの連続使用時間に関して述べる。本実験中に観察できた、各状況での意思伝達ボタンそれぞれの使用時間を図 4.3 に示した。例えば、「ステージ選択中」の意思伝達ボタン1の使用時間は、中央値0.58で、第1四分位数が1.3および第3四分位数0.3を示す。図 4.3 は左から順に、意思伝達ボタン1の「ステージ選択中」・「アイテム出現中」・「戦闘中」、意思伝達ボタン2の「ステージ選択中」・「アイテム出現中」・「戦闘中」での使用時間を示す。

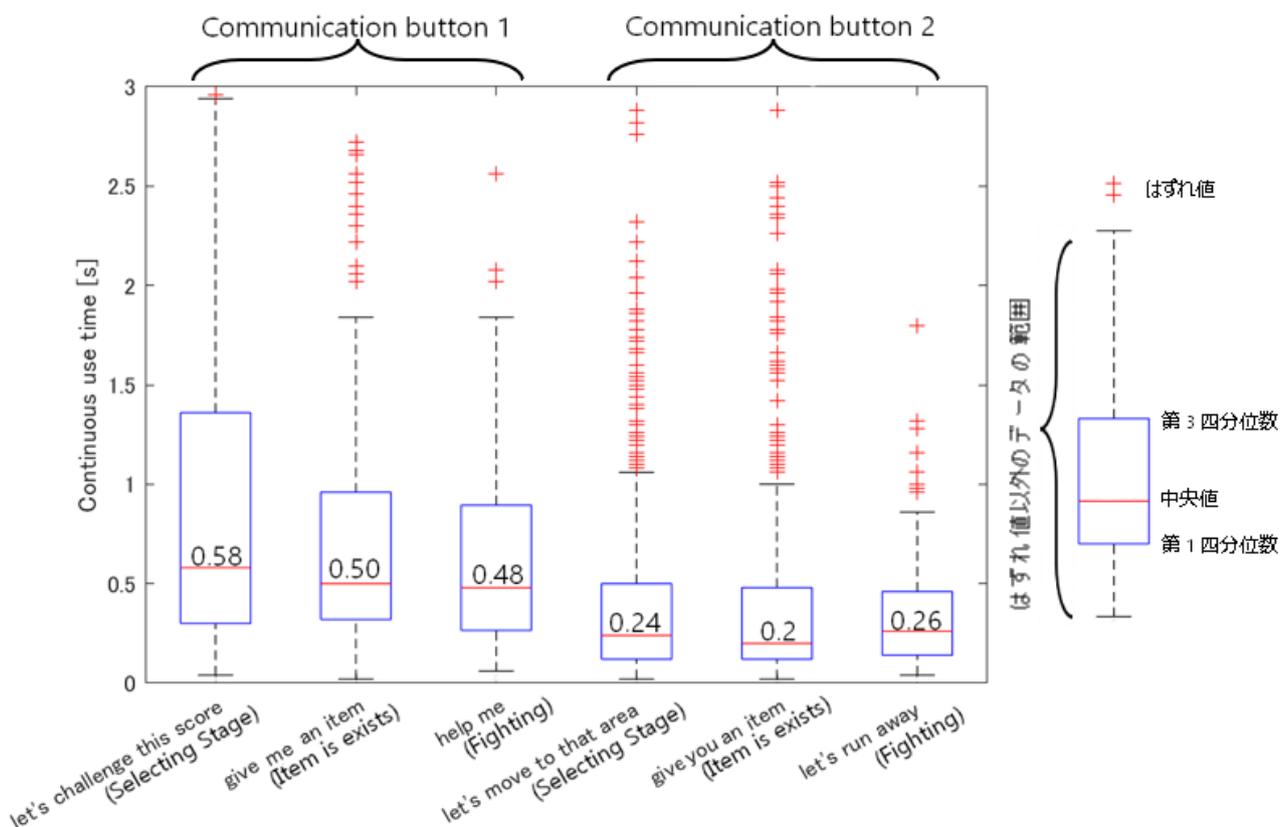


図 4.3: 意思伝達ごとの連続使用時間のはこひげ図。はずれ値は第1四分位数と第3四分位数の差の1.5倍以上に、第1四分位数または第3四分位数から離れているデータを示す。なお、このはずれ値も学習に利用した。

図 4.3 から、それぞれの中央値に注目すると、意思伝達ボタン1の連続使用時間は、3つの状況で差が0.10秒以下とほとんど変わらなかった。また、意思伝達ボタン2に関しても、3つの状況で差が0.06秒以下と小さかったうえに、第1四分位数および第3四分位数もほとんど差がなかった。一方、各状況で連続使用時間の中央値を比べてみると、意思伝達ボタン2より意思伝達ボタン1の方が倍近く長い傾向があった。これらのことから、意思伝達の連続使用時間は、時間的余裕による影響より、どちらのボタンが使われたかの影響の方が大きいと予想できる。

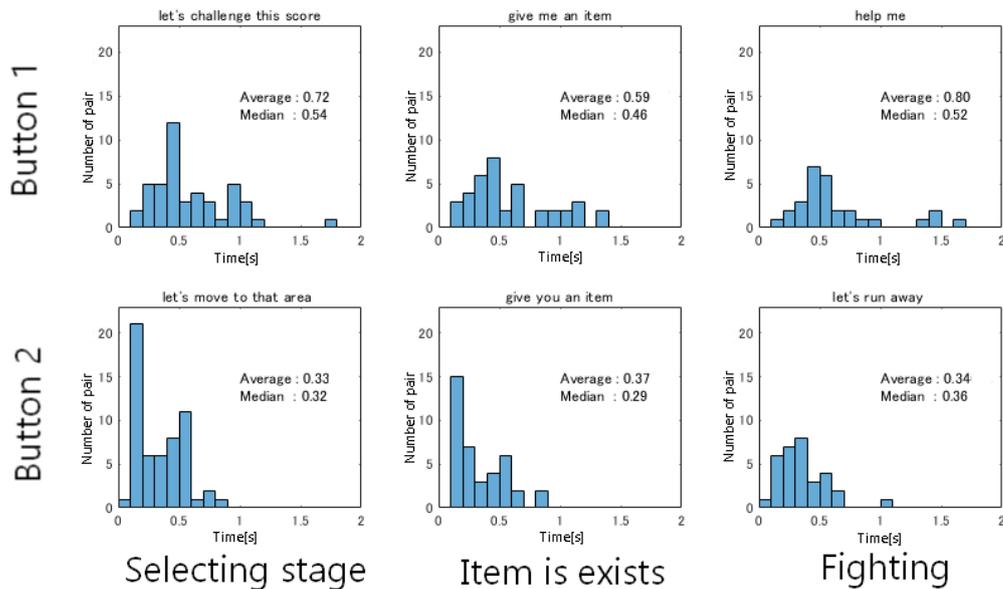


図 4.4: ペアごとの意思伝達ボタン連続使用時間の中央値のヒストグラム

図 4.4 は「ペアごとの、意思伝達ボタン連続使用時間」の中央値をヒストグラムで示したものである。どのヒストグラムを見ても、ペア間での連続使用時間の差が大きく現れていた。このように、意思伝達のそもそもの使い方がペアごとに異なる場合が多かった。一方、3つの状況それぞれのヒストグラムで、連続使用時間の中央値に注目すると、どの状況でもボタン2よりボタン1の方が大きくなったし、状況が異なっても、ボタン1・2それぞれで値の差が小さかった。やはり、連続使用時間は、各状況の時間的余裕より、どちらのボタンが使われたかの影響の方が大きいようだった。この理由として以下の2つが挙げられる。

1. 偶然、ボタン1に割り当てた意図の方が長い時間を必要とするものだったから、
2. ボタンに割り当てた運動（回転運動・前後運動）それぞれがはっきりと行われたと認識できる時間が異なるから。

このことを確かめるためには、それぞれのボタンで行われる運動を逆にして実験を行う必要がある。しかし、これは第1章で説明したAの「人間の意思伝達方法の分析と再現」の領分に入るため、今回は行わなかった。

4.3 主目的と意思伝達使用回数の関係

本節では、意思伝達ボタンの使用回数が、主目的（高いGPを得ること）に対し影響を与えたかどうかを確認する。まずは、使用回数とは関係なく、得点（GP）の分布についてみてみることにする。図 4.5は左からそれぞれ「送り手ごとの平均得点のヒストグラム」・「受け手ごとの平均得点のヒストグラム」・「ペアごとの平均得点のヒストグラム」を示す。どのヒストグラム見ても GP のばらつきが大きく、一番バラつきの少なかった送り手の点数も最大 10 点以上の差があった。主目的の遂行という視点で、個人差が大きく現れている。このような個人差が大きく出るゲームであると、人間の行動の推定が難しい。

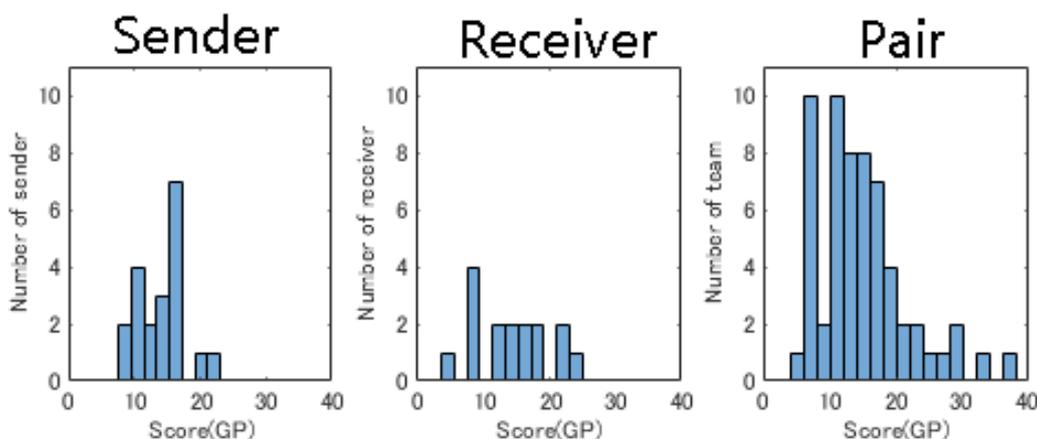


図 4.5: 送り手・受け手・ペアそれぞれでの平均得点

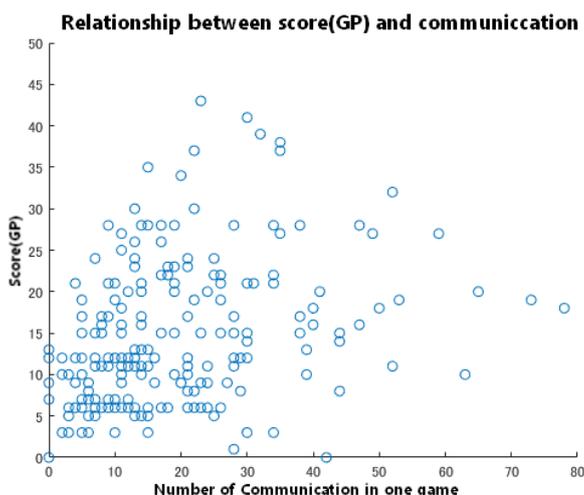


図 4.6: 意思伝達の平均回数と総得点（GP）の関係性

図 4.6 に収集した 234 試合分の「意思伝達の平均回数と総得点 (GP) の関係性」を図示する。この意思伝達ボタンの総使用回数と、総得点との相関係数を計算すると 0.356 となり、弱い正の相関があった。このことから、意思伝達を行った方がやや総得点数が高くなる傾向があることが確認できた。

また、本実験で行った 60 ペアごとのでグループ化し、意思伝達の平均回数と総得点の関係性 (図 4.7 の左図) の相関係数を計算すると 0.460 となり、1 試合ごとで見るとよりも意思伝達の平均回数と総得点の間に相関があり、意思伝達の盛んなペアの方がゲームを優位に進められていることが分かる。次に、受け手ごと・送り手ごとでグループ化し、意思伝達の平均回数と総得点の関係性 (図 4.7 の中心の図・右図) の相関係数を計算すると 0.620, 0.489 となった。このように、受け手の方の相関がやや強い相関を持っていて、受け手にとって意思伝達を多く使用する相手と組んだ方が点数を取りやすいことがわかる。

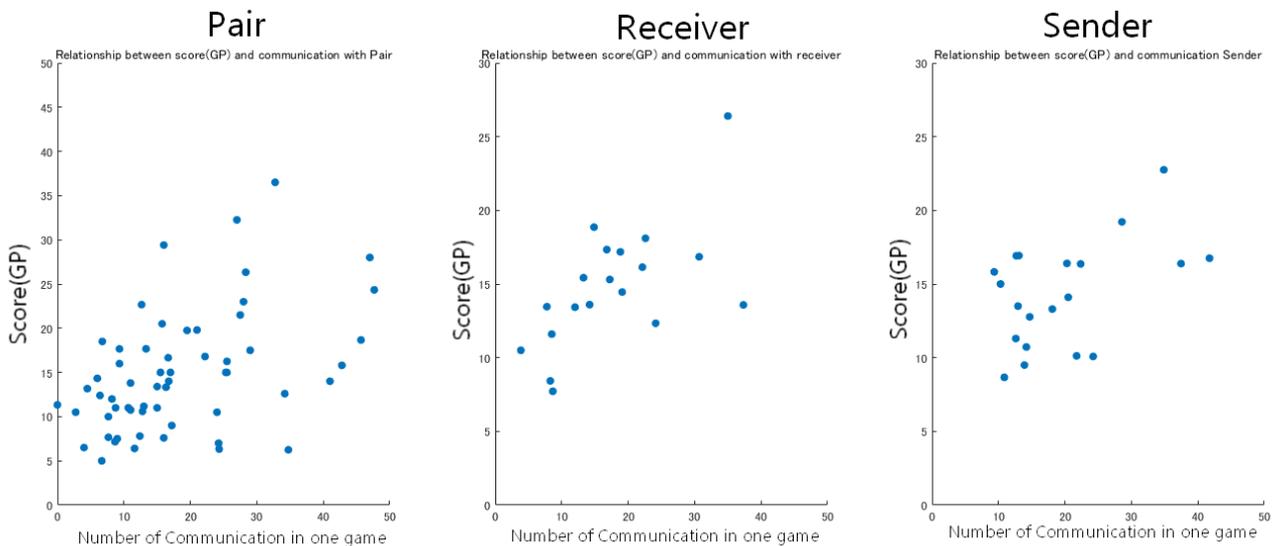


図 4.7: 受け手・ペアごとの意思伝達の平均回数と総得点 (GP) の関係性

第5章 学習手法の提案

本研究では、「人間らしい意思伝達の利用方法の再現」が研究目的の一つであった。そこで、本研究では人間の実際の行動データを用いて教師あり学習を行い、人間プレイヤーなら意思伝達を行うだろうタイミングを推定する。本節ではそのための学習法について述べていく。なお人間プレイヤーなら意思伝達を行うだろうタイミングとは、第3.3.1節で説明した意思伝達ボタン1と意思伝達ボタン2の押されているタイミングとする。

5.1 ニューラルネットワークの入出力

本節では、教師あり学習で人間の行動データを学習するにあたり、こういった入出力データを扱ったかを説明していく。

5.1.1 入力データ

本研究で行う学習は、P2が意思伝達ボタンを行うタイミングなので、P2側のプレイ画面から得られる情報を入力とする。入力データには、以下の2種類の情報を用いる。

MapData メインカメラに映るゲームステージ上の情報

BaseData P1・P2の位置関係や角度などの情報¹

まず、MapDataについて説明していく。MapDataは第3.3.4節で説明したメインカメラに映るゲームステージ上の情報を記したもので、P2を中心とした周囲のP1・E1・E2の位置情報や、エリア内にある壁、銃弾などの画像情報が入れられている。ただし、ゲーム画面中のメインカメラの画像情報をそのまま入れるわけではない。そのまま入れてしまうとRGB情報（色情報）が大きいことや、画素数が大きすぎることにより、学習時間が長くなってしまうためである。そこで、元画像（次ページ図5.1）から、操作キャラクターや敵キャラクターなどの位置情報を取得し、それぞれを 86×43 の離散値データとして変換したのち、この離散値データを 43×43 のデータ2枚に分けたものをMapDataとしている。わざわざ2枚に分け

¹サブカメラから得られる情報もここに含む

た理由は、 86×43 のままだと用意した機材のメモリが足りなくなるからである。なお、 86×43 の離散値データを白黒画像にしたものを次ページ図 5.2 に示す。離散値に加工するだけではなく、P1・P2・E1・E2 やそれぞれの弾のレイヤーマップに分けた方が性能がよくなるが、今回の実験ではそれを行っていない。この理由は、機材の関係上 86×43 の離散値データでも学習時間が約 600 分程になり、いくつものネットワークの性能を比較しているためには時間がかかりすぎると考えたためである。

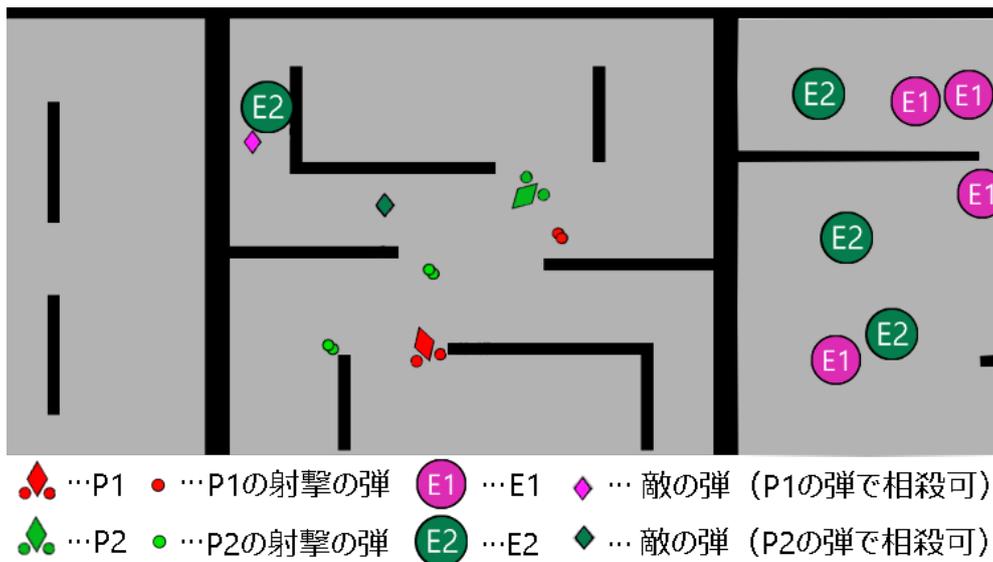


図 5.1: MapData へ変換する前の画像



図 5.2: MapData の画像

次に BaseData について説明する。プレイヤーに与えられる MapData 以外の情報であり、表 5.1 に示す 16 の情報と、後述する one-hot 法で与えられたペア情報で構成されている。なお、表 5.1 のデータ番号 13 の「P2 のいるエリア番号」とは、P2 が現在どの安全エリア・戦闘エリアにいるかを示すものである。また、データ番号 14,15,16 の「P1 のもつ戦闘エリアの「あたり」の情報間違っているかどうか」とは、第 3.3.2 節で説明した P2 に与えられる「P1 のもつ戦闘エリアの「あたり」・「はずれ」のヒント」が間違っているかどうかを 2 値で示している。現状 P2 のいるエリアの列のヒントと、その列に隣接するエリア列のヒントで 3 パラメータを取っている。

ペア情報に関してだが、本実験では、合計 20 名で操作キャラクター (P1・P2) を入れ替えながら行った。よって one-hot 法で、P1 を操作した人間はどの人間であったか、P2 を操作した人間はどの人間だったかをそれぞれ使うため、それぞれ 20 ずつのパラメータで合計 40 のパラメータでこの情報を扱っている。以上のように、5.1 に示す情報とペア情報で合計 56 のパラメータで BaseData は構成されている。

これらの入力データの中に、攻略済のエリアかどうかの情報を入れなかった理由は、MapData 内に含まれる情報を見れば、敵キャラクターの有無が簡単に確認できるからである。

データ番号	データ内容
1	残りの持ち時間
2	総得点数 (GP)
3,4,5	P1 の X,Y 座標 , 向き
6,7	P1 の HP,SP
8,9,10	P2 の X,Y 座標 , 向き
11,12	P2 の HP,SP
13	P2 のいるエリア番号
14,15,16	P1 のもつ戦闘エリアの「あたり」の情報 が間違っているかどうか

表 5.1: BaseData の一部

5.1.2 出力データ

人間の行動データから得られたデータから、以下 3 つのパラメータを抽出する。

- 意思伝達ボタンがいずれも押されていない。
- 意思伝達ボタン 1 が押されている。
- 意思伝達ボタン 2 が押されている。

これらのパラメータは0と1の2値データで構成されており、例えば意思伝達ボタン1が押されている時のデータは(0,1,0)、いずれの意思伝達ボタンも押されていない時は(1,0,0)となる。なお、この環境では、意思伝達ボタン1・2が同時に押された場合、意思伝達ボタン1が適用される仕様にしてあるので、(0,1,1)というデータはない。

この抽出されたデータをそのまま出力データとして学習させるのには問題がある。1分につき、意思伝達ボタン1・2が使われる回数は0~3回程度で(表4.3)、出力データが「(0,1,0)や(0,0,1)となる時間」が「(1,0,0)となる時間」より短い。データに偏りがある学習データでの学習は一般的に困難で、偏りを少なくするために、多いほうのデータを間引いたり、逆に少ないほうのデータを焼き増したりといった処理が行われることもある。今回の学習では、データを焼き増しするために、曖昧ラベル[13]を使う方法を用いた。

曖昧ラベルとは、2値データのような急激な変化を伴うラベルを、なめらかに変化するラベルへと変える方法である。例えば、図5.3の実線のように2箇所だけ1で他が0と、2値データでラベル付けがされたデータがあるとする。100fps(1秒に100フレーム出力される)のデータだった場合、人間にとって1フレームのずれはほとんど認識できない。よって、1フレームのずれでラベルの予測がずれていた場合、これは単純に不正解ではなく、「惜しくも不正解」または「ほとんど正解」と評価すべきである。評価時にはこれらを「正解」とみなすが、これを明示的に表現するために、曖昧ラベルを使用し、図5.3のように0と1の間の変化を滑らかにする。

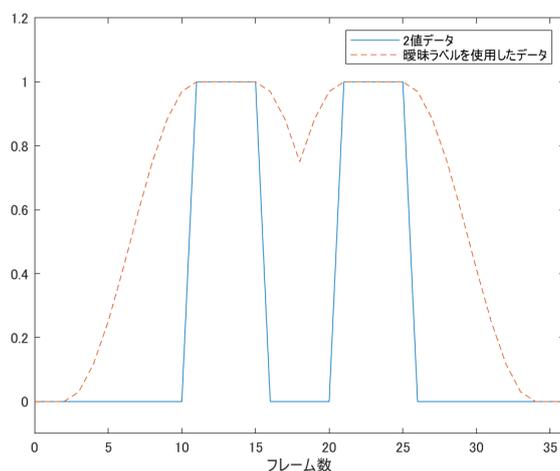


図 5.3: 2 値データへの曖昧ラベルの適用

本研究で、曖昧ラベルをどう使ったのかを説明する。まず、曖昧ラベルを適用する前の2値データと適用後のデータを以下のように定義する。

$$\begin{aligned} OriginData_{(t)} &:= \text{「曖昧ラベルを適用する前の2値データ」} \\ OutputData_{(t)} &:= \text{「曖昧ラベルを適用した後の2値データ」} \end{aligned} \quad (5.1)$$

ただし、

$$t \geq 1 \quad (5.2)$$

とする。なお、 t はゲームが始まってからのフレーム数とする。また、 $OutputData_{(t)}$ の計算に必要なため次式を定義した。

$$\begin{aligned} FrameError_{(t)} \\ := \text{「}t\text{に最も近い、}OriginData_{(t_2)} = 1\text{となる}t_2\text{の値」} \end{aligned} \quad (5.3)$$

ただし、 $OriginData_{(t_2)} = 1$ となる t_2 が存在しない場合は、

$$FrameError_{(t)} = \infty \quad (5.4)$$

とする。 $OutputData_{(t)}$ は、この $FrameError_{(t)}$ を利用して

$$OutputData_{(t)} = \begin{cases} \frac{\cos((|t-FrameError_{(t)}|/10)\cdot\pi)+1}{2} & (-10 \leq FrameError_{(t)} \leq 10) \\ 0 & (otherwise) \end{cases} \quad (5.5)$$

とする。本研究では、このように曖昧ラベルを適用した。

5.2 タイミングの予測を行うニューラルネットワーク

学習には論文 [13, 14] で述べられている C-LSTM モデルのニューラルネットワークを使用する。本節ではこの C-LSTM の概要について述べる。

C-LSTM とは CNN (Convolutional neural network) と LSTM (a long short-term memory recurrent neural network model) の二つを用いたモデルのことを指す。CNN は画像や動画認識に広く使われるネットワークで、2D 画像の特徴量の抽出などを得意とする。一方、LSTM は入力データの長期記憶を可能にした再帰型ニューラルネットワークの一つであり、動画などの時系列データを含む入力データでの学習を得意とする。C-LSTM はこれらのネットワークを組み合わせて作成されている。

本研究で使った C-LSTM の概念図を図 5.4 に示した。入力データは第 5.1.1 節で述べた通り、MapData と BaseData の二つとした。MapData は $43 \times 43 \times 2$ の大きさのデータで、BaseData は 56 の大きさのデータである。C-LSTM では、まず MapData を畳み込み層に入れる。畳み込み層はカーネル $7 \times 7 \times 1$ である。次に、この畳み込み層の出力と BaseData を合わせて、200 ユニットの LSTM 層に入れる。この LSTM の出力を 512 ユニットの全結合層に用いる。最後は Sigmoid 出力により、出力データの 3 パラメータを出力する。なお、畳み込み層と全結合層の活性化関数は ReLU、LSTM 層は tanh を用いる。

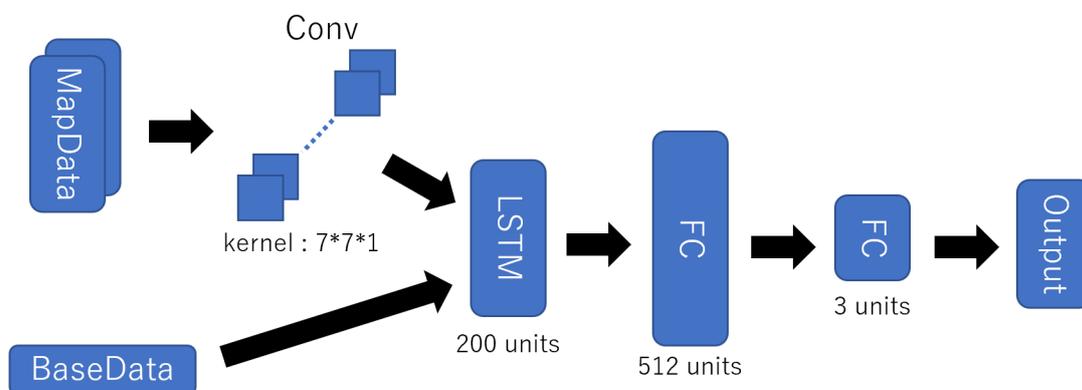


図 5.4: C-LSTM の概念図

5.3 学習結果の評価方法

本節では、ネットワークから得られた予測の精度をどう評価するかについて説明する。

5.3.1 評価対象の処理

まず、評価するにあたり、ネットワークから得られた予測を評価しやすい形に処理する必要がある。この処理について本節で説明する。

ネットワークから出力される値は「意思伝達ボタンがいずれも押されていない」・「意思伝達ボタン1が押されている」・「意思伝達ボタン2が押されている」の3つの値である。「意思伝達ボタン1が押されている」・「意思伝達ボタン2が押されている」の2つの出力値がわかれば、おのずと「意思伝達ボタンがいずれも押されていない」かについても推測できるので、この2つのみを使って評価する。この2つの出力値は図 5.5 のように 0~1 の連続値になる。図 5.5 は上から順に、「意思伝達ボタン1が押されている」・「意思伝達ボタン2が押されている」を示す。例えば、上図の「意思伝達ボタン1が押されている」のグラフに関して、30 フレーム目から値が上がりはじめ、50 フレーム目にはほとんど 1 となっている。つまり、この 30 フレーム目から 50 フレームまでの間に意思伝達ボタン1が押される可能性が高いことを示す。

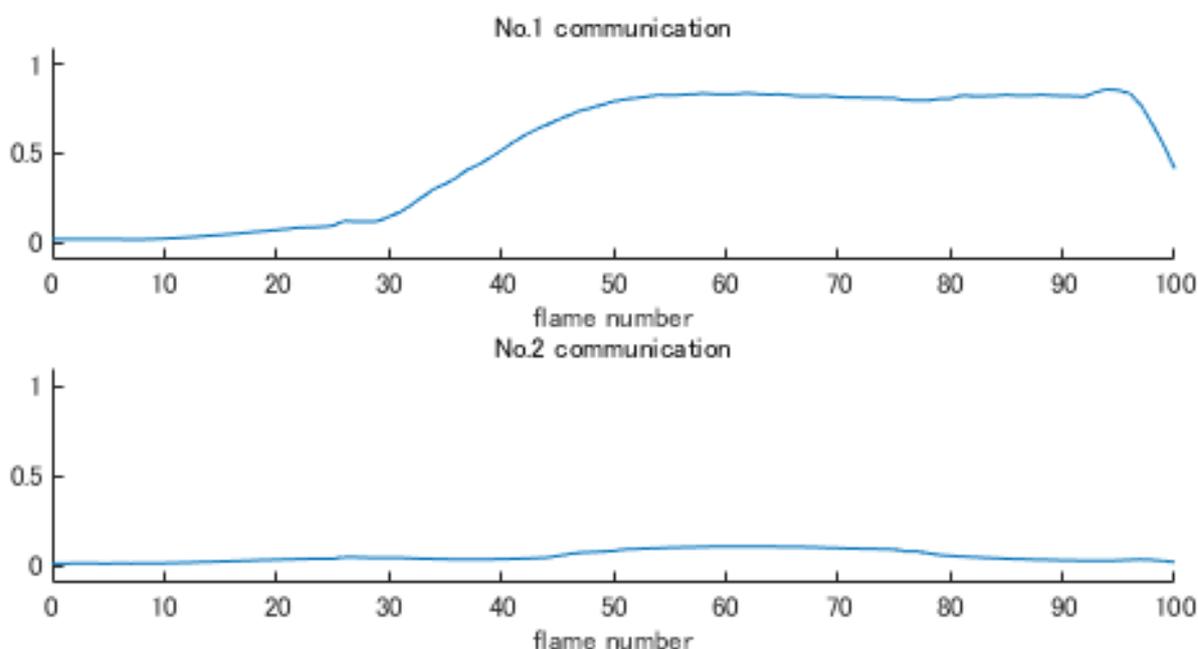


図 5.5: ネットワークの出力値

評価対象は、ネットワークから得られた連続値をそれぞれ閾値を使って2値化し、その値が0から1に変化したタイミングとする。例えば、図 5.6 の青色の実線のように、「意思伝達ボタン1が押されている」の出力値があったとする。これを閾値 0.6 で2値化し、その値が0から1に変わった瞬間を取り出す（図 5.6 の黄色の実線）。この場合は、42 フレーム目で0から1に変化している。この予測値から計算した閾値を超えた瞬間のフレームと、実際にボタンが押された瞬間のフレームとで比較し評価する。

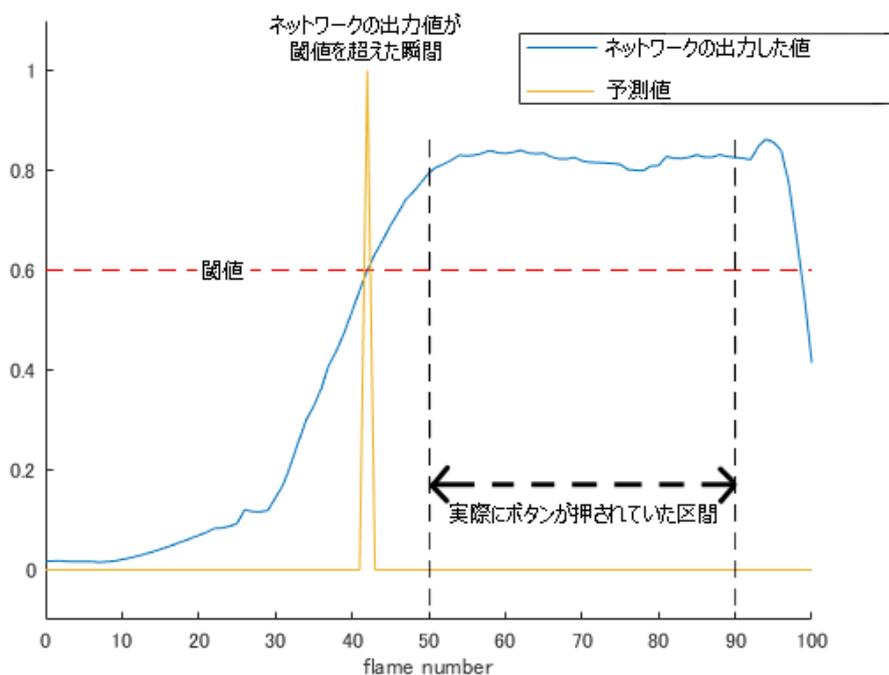


図 5.6: 予測値と閾値の関係の例（実際にボタンが押された瞬間より 8 フレーム早く予測してしまっている）

5.3.2 機械的評価方法

本節では、第 5.3.1 節で述べた評価対象に対し、どう評価をするのかについて説明する。

まずフレーム誤差の許容範囲について説明する。フレーム誤差とは予測値（ネットワークの予測する、ボタンが押された瞬間）と実測値（実際にボタンの押された瞬間）のずれのことである。本人の疲れやゲームに対する慣れ・仲間となった相手との相性などの関係で、同じ人間が同じ状況で行う意思伝達でも、数フレームのずれが生じる。異なる人間ならばこのずれはさらに大きなものになる。よって、予測値が実測値からずれたとしても、そのずれが小さいものならば不正解と

評価するのではなく、正解と評価してもよいと考える。例えば、図 5.6 に示した予測値は、実際にボタンが押されていた区間より前に「ボタンが押される」と予測している。予測値は本来より早めのタイミングでの推測となっているが、これは概ね適切な推測であると考えられる。このように、予測値について評価を行うために、このフレーム誤差をどの程度許容して「正解」と評価するかについて考えなくてはならない。

本実験に使った評価方法では、フレーム誤差の許容範囲を以下のように設定している。例えば図 5.7 の上図のように 0 フレーム目に実測値があった場合、その直前の 20 フレーム（ゲーム内で 2 秒）から直後の 1 フレームまでに、予測値が 1 となるフレームが存在すれば、その実測値は正しく再現されたものとする。また、予測値が、図 5.7 の下図のように 0 フレーム目に存在した場合、直前の 1 フレームから直後の 20 フレームまでに実測値が 1 となるフレームが存在すれば、その予測値は正しく予測をしたものとする。一見、実測値の許容範囲と予測値の許容範囲が異なっているように見える。これについて説明する。例えば、実測値にとって直前の許容範囲の 20 フレームは、予測値にとって直後 20 フレームの範囲であると認識できる。一方、実測値にとって直後の許容範囲 1 フレームは、予測値にとって直前 1 フレームであると認識できる。このように、一見実測値と予測値の許容範囲が異なるように見えるが、それぞれの目線が異なっているだけで同じ許容範囲である。

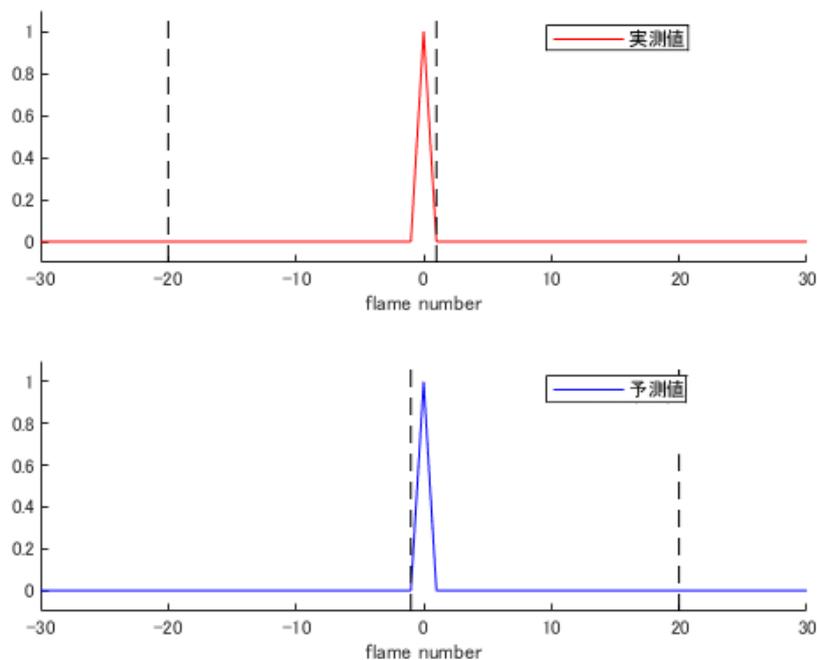


図 5.7: フレーム誤差の許容範囲

次に、なぜフレーム誤差の許容範囲が、実測値の直前 20 フレームなのに対し、直後は 1 フレームだけなのかについて説明する。本実験に使用した環境で意思伝達ボタン 1・2 を使用すると、それぞれ回転運動・前後運動が行われる。この情報は、第 5.1.1 節で説明した入力データの中に、P2 の位置情報や向き情報として含まれている。つまり、この意思伝達ボタンが押されている間は、入力データの中に答えが含まれている状態になる。この押されている間の予測は、もはや予測ではなく答えをそのまま伝えていただけになり、予測であるとは言えない。よって、フレーム誤差の許容範囲を直後 1 フレームに限定している。1 フレームならば、意思伝達ボタンが押されたという正解は認識しきれないと考えた。

最後に、この許容範囲に対し、再現率と適合率・F 値を以下のように定義する。

再現率

「許容範囲内に予測値があるような実測値の数」 / 「実測値の総数」

適合率

「許容範囲内に実測値があるような予測値の数」 / 「予測値の総数」

F 値

再現率と適合率の調和平均

本実験での評価は、この 3 つの値を使って評価する。基本的にこの評価方法では、二つのネットワークを比べて F 値が高い方が性能が高いと評価することにする。

5.3.3 人間的評価方法

次に、人間的評価方法について述べる。この人間的評価方法は第5.3.2節の機械的評価方法とは異なり、ネットワークの出力を人間の目で評価させるものである。評価には以下の方法を用いた。

1. 実測値のある場所 100 個と、予測値のある場所 100 個をランダムに選択し、その時の意思伝達ボタンの行動（回転運動または前後運動）と対応する意味（「助けて」など）のラベルをつける。
2. 選択した実測値および予測値が出現する 8 秒前から直前までの動画を合計 200 個作成する。
3. 人間に対し、動画と 1 でつけたラベルを見せ、以下の 3 値で評価する。
 - 「私も、指定したラベル（意図）で意思伝達行動をする」
 - 「私なら、指定したラベル（意図）と異なる意図で意思伝達行動をする」
 - 「私なら、意思伝達行動をしない」

この方法によって得られた結果から、「私も、指定したラベル（意図）で意思伝達行動をする」を選んだ場合は「正解」とし、それ以外を選んだ場合は「不正解」とした。そして、実測値に対しての正解率と、予測値に対しての正解率を計算し、その正解率の値を比較した。この正解率が高いほど、被験者にとってより自然な意思伝達をしていたといえる。

なお、ここでは実測値・予測値のない場所での評価を行わない。これは、実測値・予測値のない部分の動画の作成が困難だからである。動画の中から実測値・予測値の無いところは広く存在するが、その大半の部分で押さない理由は「押す必要がない状況」だからである。判定がしたいのは、「押す価値がある状況」で、人間なら押すのかどうかであるから、この実測値・予測値の無いところのうち、「押す価値がある状況」を検出するのは難しい。よって本実験では実測値・予測値のない場所自体の評価を行わなかった。

第6章 学習結果

本章では第5章で説明したネットワークを学習させた結果について、機械的性能評価と人間的性能評価を行い考察を行う。

6.1 機械的性能評価

本節では、ネットワークの学習を入力データの長さや、その内容を変更しながら行い、第5.3.2で説明した機械的性能評価で評価することで、予測精度を向上させるのに必要な要素を確かめる。テストデータは、60ある各ペアから1試合ずつをランダムに選び、合計60試合分のデータとしている。教師データをテストデータを除いて、合計174試合としている。

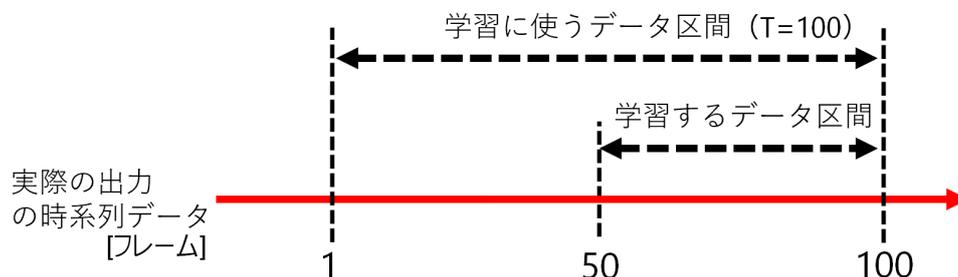


図 6.1: 学習を行う区間

また、機材のメモリの関係上、1試合約6分のデータをまとめて学習させることは難しく、数秒単位に区切って学習を行う必要があった。この学習に使うデータ区間を T フレームとする。なお、この T フレームのネットワークへの影響については第6.1.1節で述べる。この T フレーム分を1試合を取り出してくるが、これは25フレームずつずらして取り出す。例えば $T = 100$ の時、学習に使うデータ区間は、 $-74 \sim 25$ フレーム、 $-49 \sim 50$ フレーム、 $-24 \sim 75$ フレーム...のように25フレームずつずらして取り出した。なお、この負のフレームはゲームの初期状態(1フレーム目の状態)とする。また、ネットワークに学習させるのは T フレーム全ての区間ではない。図6.1のように、 $T = 100$ フレームのデータが実際の出力データの1~100フレームから取り出されたとする。このうち、ネットワークで学習するデータ区間は50~100フレームの区間のみとし、1~50フレームの区間は、

ネットワークには入れるが学習には使わない。この区間を学習に使わないのは理由がある。この学習で使うネットワークは、直前までの入力されたデータも使って出力を計算している。この直前までのデータが少ない場合、データが足りず正確な予測がしにくい。こういった区間は学習性能を著しく落とす危険性があるため、この学習しない区間を設けた。なお、この学習しない50フレームの区間は、元のデータから取り出した学習に使うデータ区間（ T フレーム分のデータ）の最初の50フレームとする。例えば、25～125フレームで取り出した場合は最初の25～75フレームの区間、 $T = 500$ として1000フレームから1500フレームで取り出した場合は1000フレームから1050フレームの区間が学習しない区間となる。

次に基準となるネットワークについて述べる。基準となるネットワークは、第5.1.1節で述べたMapDataとBaseDataを用い、100フレーム（10秒）ずつで区切って学習させた。また、第5.3.1節で説明したネットワークの出力を2値化するための閾値を変化させながら、最も値の高いF値に注目して基準とすることにした。図6.2にこの基準となるネットワークを記した。左図が「意思伝達ボタン1が押されるタイミングの評価」、右図が「意思伝達ボタン2が押されるタイミングの評価」である。F値が最も高くなる閾値はそれぞれ、0.20と0.28（図6.2の赤い丸）であった。この閾値でのF値・再現率・適合率を表6.1に示す。この値を基準に、以降ほかのネットワークを評価する。なお、このネットワークの出力が人間の目でみて自然なものかどうかは第6.2節で考察する。

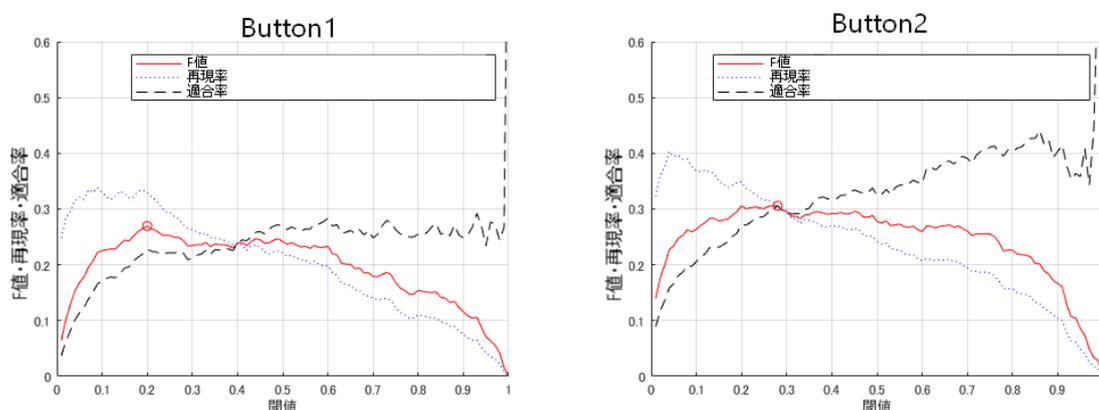


図 6.2: 基準となるネットワークの F 値・再現率・適合率

	閾値	F 値	再現率	適合率
意思伝達ボタン 1	0.20	0.269	0.330	0.228
意思伝達ボタン 2	0.28	0.306	0.305	0.306

表 6.1: 基準となるネットワークの F 値最大となる時のパラメータ

6.1.1 教師データの長さの影響

この節では、教師データの長さによる影響を述べる基準となるネットワークに使った教師データは100フレーム（10秒）ずつに区切って行っていたが、この長さを500フレーム（50秒）に変化させて学習を行った。図6.3には意思伝達ボタン1・2それぞれに対して、「入力データを100フレーム（10秒）ずつ区切って学習を行ったネットワーク」（赤）と「500（50秒）フレームずつで区切って学習を行ったネットワーク（青）」を記した。また、表6.2にこの「500フレームずつで区切って学習を行ったネットワーク」のF値が最大となるときの閾値・F値・再現率・適合率を示す。この図6.3から、広範囲の閾値で「入力データを100フレームずつ分けて学習を行ったネットワーク」の方が、F値が大きく性能がいい。しかし、この二つのネットワークのF値の最大値の差は0.02未満でほとんど同じ値で、一概に100フレームで区切った方が性能が高いとは言えない。

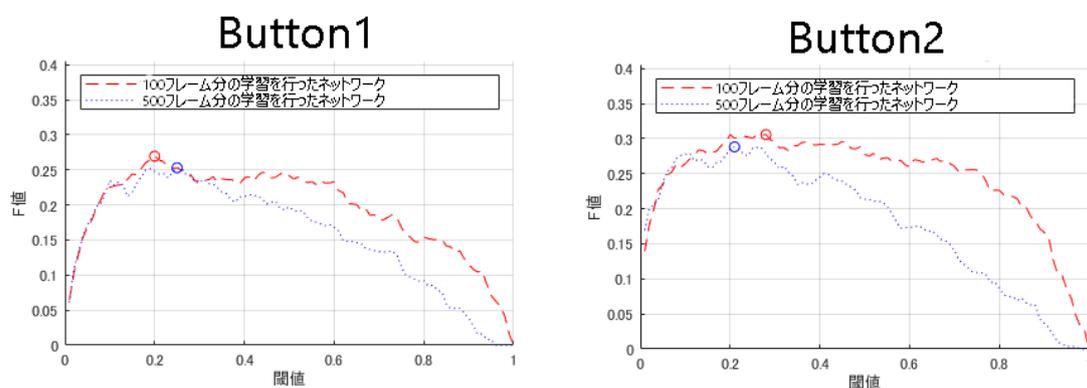


図 6.3: フレーム長による F 値の影響

	閾値	F 値	再現率	適合率
意思伝達ボタン 1	0.25(+0.05)	0.253(-0.016)	0.261(-0.069)	0.245(+0.017)
意思伝達ボタン 2	0.21(-0.07)	0.288(-0.018)	0.264(-0.041)	0.318(+0.012)

表 6.2: 500 フレームの教師データで学習させたネットワークの F 値最大となる時のパラメータと基準となるネットワークとの比較

6.1.2 ペア情報の重要性

次に、「ペア情報を入れていないネットワーク」と「ペア情報を入れたネットワーク（基準のネットワーク）」を比較する．図 6.4 に「ペア情報を入れていないネットワーク」と「ペア情報を入れたネットワーク」それぞれの F 値を示す．また，表 6.3 に F 値が最大となる時の閾値・F 値・再現率・適合率を示す．左図が意思伝達

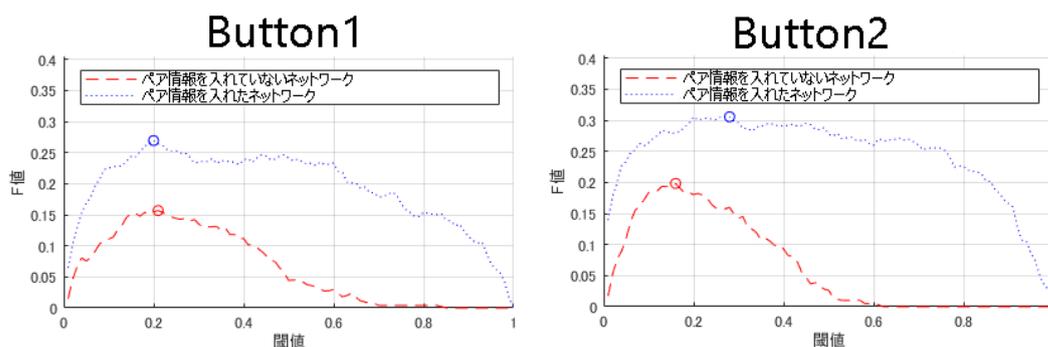


図 6.4: ペア情報による F 値の影響

	閾値	F 値	再現率	適合率
意思伝達ボタン 1	0.21(+0.01)	0.157(-0.112)	0.317(-0.013)	0.104(-0.124)
意思伝達ボタン 2	0.16(-0.12)	0.198(-0.108)	0.262(-0.043)	0.16(-0.146)

表 6.3: ペア情報を入れていないネットワークの F 値最大となる時のパラメータと基準となるネットワークとの比較

ボタン 1 の F 値，右図が意思伝達ボタン 2 の F 値を示す．この図から「ペア情報を入れていないネットワーク」の F 値が第 6.1.1 節の時以上に F 値が下がったことが見て取れる．そして，それぞれのネットワークの F 値の最大値を比べると差が 0.1 以上あり，「ペア情報を入れたネットワーク」の方が性能がいいといえる．このことから，ペアによって意思伝達の方法が大きく異なっていることがわかる．

6.2 人間的性能評価

最後になるが、第 6.1 節で扱った「基準となるネットワーク」を、第 5.3.3 節で述べた人間的性能評価で評価する。この性能評価は 3 名の被験者に行った。実測値・予測値に対する人間的性能評価の割合を表 6.4 に示す。第 5.3.3 節で述べたように「私も、指定したラベル（意図）で意思伝達行動をする」を「正解」として扱えば、実測値の正解率は 62.7% で、予測値の正解率は 55.0% であった。このように人間が他人のプレイ動画を見た時の正解率が 62.7% しかなく、人間の意思伝達行動であっても、他人から見れば予測が容易なものではないということが分かった。このことを考えると、予測値の正解率が 55.0% というのは実測値の正解率には及ばないが、決して低い値というわけではない。このネットワークの機械的性能評価による適合率は、意思伝達ボタン 1・意思伝達ボタン 2 それぞれで 0.228, 0.306 であった（表 6.1）。この人間的評価による適合率は 0.550（予測値の正解率）となっていた。このように基準となるネットワークの 0.306 という F 値の値から受ける印象よりは、人間にとってずっと自然な意思伝達をしていたといえることが明らかになった。

	実測値に対する 評価割合	予測値に対する 評価割合
私も、指定したラベル（意図）で 意思伝達行動をする	62.7%	55.0%
私なら、指定したラベル（意図）と 異なる意図で意思伝達行動をする	13.3%	13.3%
私なら、意思伝達行動をしない	24.0%	31.7%

表 6.4: 実測値・予測値に対する人間的性能評価の割合

第7章 おわりに

本論文では、人間らしい意思伝達の利用方法の分析と再現をおこなった。人間が意思伝達を行いやすい条件として、「人間プレイヤーの持つ情報に差があること」・「仲間の行動が見やすいこと」・「行動空間の幅と時間的余裕があること」を挙げ考察を行った。その3つの条件を考慮しながら、人間の意思伝達が発生しやすい三人称視点の2人協力型2Dシューティングゲームを製作した。

製作した環境で、6つの意思伝達を観察した。この意思伝達のうち、「ステージ選択中」の「このエリアに移動しよう」などの時間的余裕の大きい状況では意思伝達が多く使われた一方、「戦闘中」の「逃げよう」などの時間的余裕の小さい状況だと意思伝達はあまり使われないことが確認できた。また、意思伝達の使用頻度には個人差が大きく影響すること、意思伝達の頻度が高いペアほど高得点を取る傾向があることが確かめられた。この環境で得られた人間のデータを使い、C-LSTMを用いた教師あり学習での再現を試み、人間の誰と誰が組んだかという情報が予測性能に与える影響が大きいことが確かめられた。

また、人間の目でネットワークの予測が自然であるかを確かめるための実験を行った。この実験で、人間の目で他の人間が意思伝達をするかどうかの正解率は62.7%と低かったことを確認した。このように人間の意思伝達行動であっても、他人から見れば予測が容易なものではないことが分かった。一方、ネットワークの予測した、意思伝達をするタイミングの正解率は55.0%となった。ネットワークのF値0.306という値から受ける印象よりは、人間にとってずっと自然な意思伝達をしていたと言えることが明らかになった。

今後は、ゲームの主目的達成という観点で強いゲームAIを実装し、そこに本研究で再現した意思伝達の利用方法を組み込んで、人間の行動に対し、どのように影響を与えるかについて観察したいと考えている。

参考文献

- [1] 宗藤大貴, 長尾智晴 進化計算法を用いた詰将棋の自動生成, ゲームプログラミングワークショップ 2019 論文集, pp.1-6,2019-11-01
- [2] Rodrigo Vicencio-Moreira, Regan L. Mandryk, and Carl Gutwin, Now You Can Compete With Anyone: Balancing Players of Different Skill Levels in a First-Person Shooter Game, Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems April 2015 Pages 2255–2264
- [3] 藤井叙人, 佐藤祐一, 中畠洋輔, 若間弘典, 風井浩志, 片寄晴弘 生物学的制約の導入による「人間らしい」振る舞いを伴う ゲーム AI の自律的獲得, ゲームプログラミングワークショップ 2013 論文集, pp.73-80,2013-11-01
- [4] 池田祐太郎, 池田心 アクションゲームにおける特定のプレイヤーの特徴を模倣する AI プレイヤーの作成, 情報処理学会研究報告. GI, ゲーム情報学, 2020-GI-43(9):1-8
- [5] Naoyuki Sato, Kokoro Ikeda and Takayuki Wada, Estimation of player’s preference for cooperative RPGs using multi-strategy Monte-Carlo method, IEEE Conference on Computational Intelligence and Games (CIG2015), pp.51-59, 2015-08.
- [6] 中川絢太, ゲームの主目的達成を意図しない人間らしい行動の分類と模倣, 北陸先端科学技術大学院大学 修士論文 (情報科学) 2016
- [7] 栗原一貴, Toolification of Game: 既存ゲームの余剰自由度の中で非ゲーム目的を達成するゲーミフィケーションの考察, エンタテインメントコンピューティングシンポジウム 2015 論文集 (2015):8-17
- [8] David Silver, Thomas Hubert, Julian Schrittwieser, Ioannis Antonoglou, Matthew Lai, Arthur Guez, Marc Lanctot, Laurent Sifre, Dhharshan Kumaran, Thore Graepel, Timothy Lillicrap, Karen Simonyan, Demis Hassabis, “A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play, SCIENCE Vol.362, No.6419 (2018)

- [9] Max Jaderberg, Wojciech M. Czarnecki, Iain Dunning, Luke Marris, Guy Lever, Antonio Garcia Castañeda, Charles Beattie, Neil C. Rabinowitz, Ari S. Morcos, Avraham Ruderman, Nicolas Sonnerat, Tim Green, Louise Deason, Joel Z. Leibo, David Silver, Demis Hassabis, Koray Kavukcuoglu, Thore Graepel, “” Human-level performance in 3D multiplayer games with population-based reinforcement learning, SCIENCE Vol.364, No.6443 (2019)
- [10] 横山絢美, 大森隆司 協調課題における意図推定に基づく行動決定過程のモデル的解析, 電子情報通信学会論文誌 A vol.J92-A No.11 pp.734-742,2009-11-01
- [11] ナムサンギユ, 池田心, 強化学習を用いたターン制 RPG のステージ自動生成, 第 23 回ゲームプログラミングワークショップ (GPW-18), 2018-11
- [12] DJ Strouse, Kevin R. McKee, Matt Botvinick, Edward Hughes, Richard Everett, Collaborating with Humans without Human Data, NeurIPS 2021
- [13] Liang Yubin, 池田心, リズムゲームの上達を支援するコンテンツ自動生成法, 情報処理学会 第 39 回ゲーム情報学 (GI) 研究発表会, 2018-3
- [14] Chris Donahue, Zachary C. Lipton, Julian McAuley, Dance Dance Convolution, Proceedings of the 34th International Conference on Machine Learning, PMLR 70:1039-1048, 2017.

謝辞

本論文は、筆者が北陸先端科学技術大学院大学 先端科学技術研究科 池田研究室に在籍していた期間に行った研究をまとめたものです。本研究を遂行するにあたり、多くの方のご支援とご指導を賜りました。特に、筆者の指導教官である情報科学研究科の池田心教授には多忙な中熱心に見守ってくださり、非常に感謝申し上げます。これからも、この経験を糧として生かしていきたいと考えています。

他にも、実験環境の製作についてアドバイスをくださった研究室同期の酒見真氏や藤平啓汰氏、ネットワークの構成についてご指導いただいた Li Wanxiang 氏には誠に感謝いたします。

また、筆者の生活を支えてくれた家族、筆者と一緒に励ましあいながら論文を執筆した友人たちに、深く感謝申し上げます。

本論文は皆様のご協力なくして完成することはなく、本論文の執筆にあたり多大なご支援とご協力を頂いた皆様にこの場を借りて感謝申し上げます。