情報処理学会論文誌 Vol.55 No.7 1655–1664 (July 2014)

生物学的制約の導入によるビデオゲームエージェントの「人間らしい」振舞いの自動獲得

藤井 叙人[1,2,a)] 佐藤 祐一[1] 若間 弘典[1] 風井 浩志[1,b)] 片寄 晴弘[1,c)]

受付日 2013年7月17日, 採録日 2014年4月4日

概要：ビデオゲームエージェント（ノンプレイヤキャラクタ：NPC）の振舞いの自動獲得において，「人間の熟達者に勝利する」という長年の目標を達成する日もそう遠くない．一方で，ユーザエクスペリエンスの向上策として，『人間らしい』NPC をどう構成するかが，ゲーム AI 領域の課題になりつつある．本研究では，人間らしい振舞いを表出する NPC を，開発者の経験に基づいて実現するのではなく，『人間の生物学的制約』を課した機械学習により，自動的に獲得することを目指す．人間の生物学的制約としては「身体的な制約：“ゆらぎ”，“遅れ”，“疲れ”」，「生き延びるために必要な欲求：“訓練と挑戦のバランス”」を定義する．人間の生物学的制約の導入対象として，アクションゲームの “Infinite Mario Bros.” を採用し，本研究で獲得された NPC が人間らしい振舞いを表出できているか検討する．最後に，獲得された NPC の振舞いが人間らしいかどうかを主観評価実験により検証する．

キーワード：ビデオゲーム，NPC，機械学習，スーパーマリオワールド，人間の生物学的制約

Autonomously Acquiring Video-Game Agent's Human-like behaviors with Biological Constraints

Nobuto Fujii[1,2,a)] Yuichi Sato[1] Hironori Wakama[1] Koji Kazai[1,b)] Haruhiro Katayose[1,c)]

Received: July 17, 2013, Accepted: April 4, 2014

Abstract: Designing the behavioral patterns of video game agents (Non Player Character: NPC) is a crucial aspect in developing video games. While various systems that have aimed at automatically acquiring behavioral patterns have been proposed and some have successfully obtained stronger patterns than human players, those patterns have looked mechanical. We propose the autonomous acquisition of video game agent behaviors, which emulate the behaviors of human players. Instead of implementing straightforward heuristics, the behaviors are acquired using techniques of reinforcement learning with Q-Learning, where *biological constraints* are imposed. Human-like behaviors that imply human cognitive processes were obtained by imposing *sensory error*, *perceptual and motion delay*, *physical fatigue*, and *balancing between repetition and novelty* as the biological constraints in computational simulations using “Infinite Mario Bros.”. We evaluated human-like behavioral patterns through subjective assessments, and discuss the possibility of implementing the proposed system.

Keywords: video game, NPC, machine learning, Mario Bros., biological constraints

[1] 関西学院大学大学院理工学研究科人間システム工学専攻
Department of Human System Interaction, Graduate School of Science and Technology, Kwansei Gakuin University, Sanda, Hyogo 669–1337, Japan
[2] 日本学術振興会特別研究員 DC2
Research Fellow of Japan Society for the Promotion of Science, Chiyoda, Tokyo 102–0083, Japan
[a)] nobuto@kwansei.ac.jp
[b)] kazai@kwansei.ac.jp
[c)] katayose@kwansei.ac.jp

1. はじめに

エンタテインメント系システムにおいて，プレイフィール（プレイ時の感覚や印象）の形成や，日常娯楽としての定着化は，ユーザ数の確保に大きな影響を与える．家庭用ゲームに代表されるビデオゲーム市場は，年々躍進を続けており，ユーザエクスペリエンス [1] の向上に余念がないこ

とがうかがえる（2012 年の国内市場規模：前年比 15.3%増の 9776.9 億円 [2]）．ビデオゲームにおけるプレイフィールを決定づける要因として，ゲーム内に登場するコンピュータ担当のエージェント（= NPC）の存在を無視することはできない．そのため，レベルデザイン（プレイヤのレベルにあわせた難易度の調整）を含めた，NPC の振舞いのデザインには，長らく，人手による煩多な作り込みが実施されてきた．

NPC の振舞いのデザインにおける作業負荷の軽減と，機械学習の応用領域としての学術的知見の発見を目的として，国内外で NPC の自動獲得に関する研究が執り行われている [3], [4], [5], [6]．この結果，人間の熟達者をも凌駕する “勝つための『強い』NPC” の自動獲得に至っているが，これらの振舞いは過度に最適化されており，人間にとっては機械的に映る．強い NPC を人間プレイヤの代替として扱った場合，エンタテインメント性が欠落するという問題が浮き彫りになっており，“人間プレイヤを楽しませるための『人間らしい』NPC” の自動獲得に興味が集まりつつある [7], [8], [9].

人間らしい NPC を試作検討する研究として，人間プレイヤの振舞いを記録し機械学習により模倣する手法 [7], [9]，強い NPC に対して恣意的にエラーを導入する手法 [8] などが発表されている．これらの研究は，開発者が意図した「強くない」NPC のデザインが可能となっており，レベルデザインの一アプローチとしては有効である．しかし，「どのような振舞いが人間らしいか」は，開発者の経験（ヒューリスティック）による煩多な作り込みにより実現されている．

本研究では，『人間の生物学的制約』の条件下での機械学習により，人間らしい NPC の振舞いを自動的に獲得する手法について提案する．人間の生物学的制約とは，人間が生得的に持つ性質から生じる制約や欲求を指す．人間の行動制御における制約 [10], [11] や自己実現理論 [12] から着想を得て，本稿では，人間の生物学的制約を「身体的な制約：“ゆらぎ”，“遅れ”，“疲れ”」「生き延びるために必要な欲求：“訓練と挑戦のバランス”」と定義する．人間の生物学的制約を機械学習の枠組みに導入することで，NPC の持つ「人間には実現不可能な（機械的な）」振舞いを抑制し，「人間らしさ」を与えることが可能となる．また，人間の生物学的制約は開発者のヒューリスティックに依存しないため，汎用的な振舞い獲得の手法として構築が可能である．振舞い獲得の対象としては，アクションゲームの “Infinite Mario Bros.” を採用し，自動的に獲得された振舞いが人間らしいかどうかを主観評価実験により検証する．

以下，2 章で，関連研究を紹介し，3 章で，人間の生物学的制約を導入する意義と，その定義を述べる．4 章で，“Infinite Mario Bros.” の仕様と，振舞い獲得の方法について説明する．5 章で，生物学的制約の導入による振舞いの変化を検証する．6 章で，獲得された振舞いが人間らしいかどうかを主観評価実験により検証する．

2. 関連研究

2.1 強い NPC を追求した研究

振舞いを自動的に獲得する手法として，教師データを入力とする事例参照型の手法 [13] と，ゲーム木探索や試行錯誤による非事例参照型の手法 [3], [4], [14] がある．教師あり学習は前者に，経路探索や強化学習は後者に分類される．

教師あり学習は，事前に与えられた大量のデータセットを教師データ（入力データに対して出力されるべきデータの例）とし，有用なルールを学習する手法である．教師あり学習によるアプローチの代表的な研究として，保木は，コンピュータ将棋プログラムである Bonanza を構築している [13]．Bonanza は，プロ棋士の棋譜 6 万局のデータを教師とし，将棋の局面における評価関数を自動学習することで，従来手法よりも良い振舞いを得ることに成功している．この手法は Bonanza メソッドと呼ばれ，多くのコンピュータ将棋プログラムで採用されている画期的な手法である [5], [15]．将棋のように，強い人間プレイヤの膨大な棋譜データが用意できる場合には，教師あり学習による振舞い獲得は有効である．

経路探索は，ゲーム木におけるスタートからゴールまでの，最小コストとなる経路を探索する手法である．経路探索によるアプローチの代表的な研究として，Baumgarten は，2009 年の Mario AI Competition において，A*アルゴリズムに基づいた NPC を構築している [3]．Mario AI Competition とは，“Infinite Mario Bros.”（ランダムに生成されるステージを制限時間内に攻略する，「スーパーマリオワールド」のようなアクションゲーム）を対象とした NPC の評価コンテスト [16] である．Baumgarten の NPC はマリオや敵の動きを事前に解析し，A*アルゴリズムを用いた経路探索によって，ステージをほぼ最適解で攻略することが可能であり，評価コンテストで見事優勝を収めている．

強化学習は，自身の振舞いの試行錯誤を繰り返すことで最適な振舞いを獲得する手法である．強化学習によるアプローチの代表的な研究として，Tsay らは，2009 年の Mario AI Competition において，Q 学習に基づいた NPC を構築している [14]．Tsay らの NPC は，敵を倒すこと，アイテムをとること，穴を飛び越えること，敵を避けることに対して報酬を与えることで，マリオの振舞いの自動獲得に成功しており，評価コンテストで 4 位（強化学習 NPC では 1 位）の成績を収めている．Fujita らは，カードゲームの Hearts を題材とし，Q 学習に基づいた NPC を構築している [4]．巨大な状態空間となること，相手の所持するカードを観測できないこと，4 人対戦のゲームであること，の 3 つを Hearts における学習の困難性と考察している．その

上で，解決手法として，パーティクルフィルタによるサンプリング，相手の行動予測器，現在の戦局を評価する状態価値関数，ゲームの特徴に基づく次元圧縮を提案し，困難性の解決を図っている．実験の結果，人間の熟達者よりも優れた振舞いを得ることに成功している．

これらの手法を用いて獲得されたNPCは，きわめて最適であるがゆえに，人間にとっては機械的と感じる振舞いを表出してしまう．そのため，エンタテインメント性の向上という視点に立った場合，人間プレイヤの代替として扱うことははばかられる．ゲームAI領域では，人間プレイヤが強いNPCに勝てなくなる日もそう遠くないと考えられており，人間らしいNPCの構築が最重要課題となりつつある．

2.2 人間らしいNPCを実装した研究

人間らしいNPCを実装した関連研究として，Schrumらは，2012年のThe 2K BotPrizeにおいて，大会史上初となる，人間よりも人間らしいと評価されるNPCの構成に成功している[7]．The 2K BotPrizeとは，FPS（一人称視点シューティングゲーム）を対象とした，NPCの人間らしさを競う評価コンテストである．人間プレイヤの振舞いをトレースしたデータベースを基に，人間らしいと思われる振舞いを決定論的に定義し，ニューラルネットにおける制約として適用している．その結果，対戦相手の人間プレイヤから「人間らしい」と評価されるNPCの振舞いが獲得できている．

池田らは，コンピュータ囲碁を対象に，既存の強いNPCに意図的に人間らしいミスをさせることで，手加減と思われない程度の「強くなさ」を実現するための初期的検討を実施している[8]．現在の局面における予測勝率と候補手の選択確率を用いた形勢の制御，楽観派や悲観派といったプレイスタイルによる獲得戦略の分析をしており，ゲームのレベルデザインにおける一アプローチを提案している．

マリオの人間らしい振舞いを検討する研究として，Ortegaらは，人間プレイヤの操作ログを教師とし，人間の振舞いをトレースするよう学習するエージェントを，3つの教師あり学習手法で実現している[9]．トレース精度の評価の結果，ある程度，人間の振舞いを模倣できている．しかし，人間プレイヤの動画との主観評価実験を実施しているが，いずれのエージェントも人間プレイヤより人間らしいという評価には至っていない．

上記の手法は，人間らしいと思われる振舞いを，開発者が恣意的に定義したものである．そのため，振舞い獲得における作業負荷の軽減や汎用性は実現されていない．

3. 人間の生物学的制約

3.1 人間の生物学的制約を導入する意義

人間らしいNPCの振舞いに関わるプレイスタイルとそのパラメータは，従来，ゲームプログラマや開発者がヒューリスティックに基づいてアドホックに決定しており，特定のゲームタイトルや機械学習手法に限定的な作り込みを採用するほかなかった．また，従来研究では，人間が生得的に持つ特徴や性質から生じる生物学的制約の考慮はされていない．そのため，コントローラ操作の反応速度が速すぎる，コントローラのボタンの入力が正確すぎる，つねに一定の行動のみを正確に繰り返すといった，人間プレイヤでは実現不可能な振舞いが表出するケースも多い．また，レベルデザインを意識しすぎると，ゲームの途中から急に弱くなる，あからさまなコントローラ操作のミスをするといった，プレイスタイルの統一性が崩壊した振舞いが表出するケースもある．これらの振舞いは，「相手がいんちきをしているのではないか」，「本当に自分の力で勝ったのか」という疑念を生むため，エンタテインメント性を削ぐ要因となっている．

本研究でNPCに導入する『生物学的制約』は，人間が生得的に持っている制約や欲求である．人間プレイヤがゲームをするときにも生物学的制約が生じているはずであり，生物学的制約下で操作されたキャラクタの振舞いは，人間にとって最も一般的で見慣れた振舞いであると思われる．人間がキャラクタの振舞いの印象を評価をする際には，一般的で見慣れた振舞いを「人間らしい」と評価する可能性は高い．本研究では，人間は意識の有無にかかわらず『生物学的制約』を考慮しており，したがって，『生物学的制約』の条件下での機械学習によって「人間らしい」と評価されるキャラクタの振舞いが獲得できると考える．

生物学的制約としては「身体的な制約」と「生き延びるために必要な欲求」を機械学習の制約条件として課する．身体的な制約に関する研究例として，Cabreraらは人間の指先による倒立棒の制御実験を[10]，大平らは人間の直立姿勢の制御実験を実施している[11]．人間の行動制御には「ゆらぎ」「遅れ」「疲れ」といった制約が生じるが，人間は訓練によってこれらの制約を意識的もしくは無意識的に考慮し，安全性とパフォーマンスを両立させる行動制御が獲得できると提唱している．また，生き延びるために必要な欲求について，Maslowは人間の欲求を5段階の階層構造で理論化した「自己実現理論」を提唱している[12]．原始的な欲求に近い階層から順に，1) 生理的欲求，2) 安全の欲求，3) 所属と愛の欲求，4) 承認（尊重）の欲求，5) 自己実現の欲求，と人間の欲求を分類している．そして，「人間は自己実現に向かって絶えず成長する生きものである」という仮定の下，「訓練」による知識の定着や，「挑戦」による不満の解消といった行動の動機は，5) 自己実現の欲求に帰結すると考えられている．

3.2 人間の生物学的制約の定義

前節で述べた，Cabreraら[10]や大平ら[11]の「身体的

な制約」と，Maslow の自己実現理論 [12] で議論されている「生き延びるために必要な欲求」を考慮し，『人間の生物学的制約』を「身体的な制約：“ゆらぎ”，“遅れ”，“疲れ”」，「生き延びるために必要な欲求：“訓練と挑戦のバランス”」として，以下のように定義する．これらは，人間の生物学的制約の 1 例ではあるものの，NPC のリアルタイム制御を要求するゲーム全般においては必ず生じる制約である．

(1) **センサ系，運動系における「ゆらぎ」**
人間プレイヤは，操作対象や敵オブジェクトなどの位置（座標）を正確に観測し認識することは難しく，必ず誤差（ゆらぎ）が生じる（見間違い，操作ミスなど）．そこで，NPC が観測する操作対象の現在位置やゲームの局面情報に対し，ガウスノイズを付与することで再現する．

(2) **知覚から運動制御に至る「遅れ」**
人間プレイヤは，ゲームの局面を認識してから，実際に動作するまでに遅れが発生する（眼と手の協応動作における遅延など）．そこで，NPC が観測する操作対象の現在位置やゲームの局面情報を，数百ミリ秒過去の情報にすることで再現する．

(3) **キー操作の「疲れ」**
人間プレイヤは，ゲームのコントローラのキー操作を，きわめて短時間で何度も，または，長時間連続して実施すると疲れが生じる（ボタン連打，単調な操作の繰返しなど）．そこで，振舞いを学習する際に，NPC にキー操作変更による負の報酬を与えることで再現する．

(4) **「訓練と挑戦のバランス」**
人間プレイヤは，同じ行動を繰り返すことで「訓練」する一方で，同じ行動の結果に飽きたり，その行動で失敗を繰り返したりすると，飽きや失敗を解消するための新奇な行動に「挑戦」する．そこで，失敗を繰り返しているゲーム局面では，新奇な行動に挑戦する傾向を高め，逆に，失敗をほとんどしないゲーム局面では，同じ行動を繰り返して訓練する傾向を高めることで再現する．これは，強化学習における「探索と知識利用のジレンマ」と同等の考え方であるが，人間プレイヤの試行錯誤学習の中でも生じる現象であり，本研究では「訓練と挑戦のバランス」として定義する．

4. 振舞いの獲得

4.1 生物学的制約を課した Q 学習

ビデオゲームにおいては，教師となるプレイデータが大量に用意できないため，非事例参照型の手法である強化学習手法を用いることにする．強化学習手法の中でも，ゲーム内での形勢を報酬という形で直感的に設定できる Q 学習 [17] を用いる．Q 学習では，最適なルールの獲得として学習が進む点で，ゲームプログラマが利用しやすいというメリットもある．

Q 学習では，ゲームのある局面における最適な行動を以下の式で算出する．

$$argmax_{a_t} Q(s_t, a_t) \tag{1}$$

式 (1) において，t はゲーム開始からの時刻，s_t は時刻 t におけるゲーム局面，a_t は時刻 t において NPC が選択する行動，$Q(s_t, a_t)$ は局面 s_t と行動 a_t の組に対する，Q 値と呼ばれる評価値である．つまり，Q 学習では，局面 s_t において Q 値が最も高くなる行動が最適であると出力される．

また，NPC の行動に応じて，以下の式で Q 値を更新することにより学習が可能となる．

$$Q(s_t, a_t) = (1-\alpha)Q(s_t, a_t) + \alpha((r + \gamma max_p Q(s_{t+1}, p)) \tag{2}$$

式 (2) において，α は学習率と呼ばれる，Q 値の更新において新たな報酬 r をどれだけ重視するかを示す値，γ は割引率と呼ばれる，0 以上 1 以下の定数である．r は局面 s_t において行動 a_t を選択したことによって得られる報酬である．NPC の行動選択手法としては $\epsilon - greedy$ 法を用いる．$\epsilon - greedy$ 法は，$1-\epsilon$ の確率で Q 値が最大となる行動を選択し，ϵ の確率でランダムに行動を選択する．

リアルタイム性のあるビデオゲームにおける各項目の扱い方として，時刻 t はフレーム単位であり，局面 s_t や行動 a_t が無数に設定できる場合は，学習が実時間で収束するようゲーム特徴を考慮した状態圧縮が必要である．また，報酬 r として，操作対象の進んだ距離，経過時間，局面が遷移する際の評価値（形勢）の増減などを与えることで，NPC の振舞いの自動獲得が可能となる [4], [18].

次に，Q 学習への人間の生物学的制約の導入に関して述べる．式 (1)，式 (2) のすべての s_t について，n フレーム過去（遅れ）の NPC の位置座標にガウスノイズ（ゆらぎ）を付与したゲーム局面 s_{t-n} とすることで，「ゆらぎ」と「遅れ」を実現する．「疲れ」は，式 (2) の Q 値の更新の際に，報酬 r にキー操作変更による負の報酬を与えることで実現する（報酬 r の詳細については 4.3 節で述べる）．「訓練と挑戦のバランス」は，ランダム行動選択確率 ϵ の設定において，失敗を繰り返しているゲーム局面 s_t では大きな値を設定することで，新奇な行動に挑戦する傾向を高め，逆に，失敗をほとんどしないゲーム局面 s_t では小さな値を設定し，同じ行動を繰り返して訓練する傾向を高めることで実現する．

4.2 “Infinite Mario Bros.” の仕様

本研究では，“Infinite Mario Bros.” を対象とし，人間らしい振舞いの自動獲得と，その比較検証，主観評価実験を実施する．“Infinite Mario Bros.” は，世界的に有名な 2D 横スクロール型アクションゲームである “スーパーマリオワールド” を模したゲームであり，そのゲーム画面を**図 1**

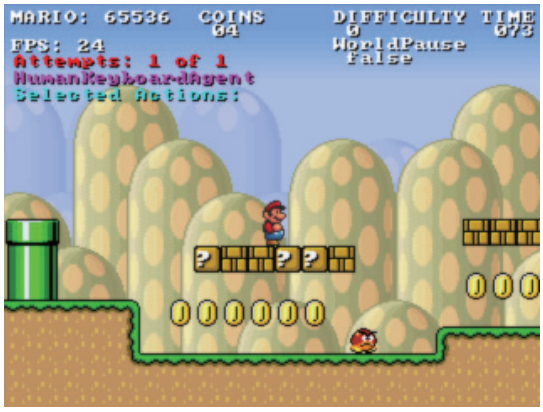

図 1 “Infinite Mario Bros.” のゲーム画面
Fig. 1 Screen capture of “Infinite Mario Bros.”

に示す．“Infinite Mario Bros.” を対象とした理由としては，以下の 4 つがあげられる．まず，1) ゲームの仕様やゲーム環境パラメータが公開されている，かつ，2)「敵や穴を避けてできる限り先に進む」という明確な目標が設定できるため，機械学習による NPC の振舞い獲得が可能である．次に，NPC の振舞いの人間らしさを評価するにあたり，3) 当該ゲームは人間型キャラクタを操作可能であるため，評価者は人間らしい振舞いを想起しやすい．最後に，キャラクタの人間らしさを判断する際の評価基準は，実際にそのゲームをプレイしたり，プレイ画面を見ているときに形成されると考えられるため，4) 世界的にきわめて有名なゲームジャンルである 2D 横スクロール型アクションは最適である．2D 横スクロール型アクションゲームは，複数人での協力プレイや敵対プレイが可能なタイトルも多数販売されており，人間の代替となる NPC の自動獲得は有用である（マリオシリーズのほかに，ソニックシリーズや星のカービィシリーズが世界的に有名）．

“Infinite Mario Bros.” における仕様は以下のとおりである．

- **ステージの自動生成**
  事前に与えた疑似乱数のシード値に従って無限にステージが生成される．
- **NPC の操作キャラクタ（マリオ）**
  NPC はマリオ（図 1 中央）を操作する．NPC によるマリオの操作はコントローラのキー入力（LEFT，RIGHT，DOWN，SPEED，JUMP）により行う．毎フレームのキーの押下状態により，マリオは対応した行動を行う（毎秒 24 フレームで動作）．また，マリオには「スーパーマリオ」「ちびマリオ」という状態が存在する．「スーパーマリオ」でダメージを受けた場合は「ちびマリオ」に変化し，「ちびマリオ」でダメージを受けた場合は死亡する．ダメージについては，後述の接触判定において説明する．穴に落ちた場合は「スーパーマリオ」「ちびマリオ」を問わず，死亡する．
- **敵キャラクタ**
  ステージには数種類の敵が登場し，敵はそれぞれ独自の動作をしている．NPC は，これらの敵を避けて進むか，倒して進むかを決定しなければならない．マリオは敵との接触判定によってダメージを受ける場合がある．踏むことができる敵は，踏む以外の行動で接触した場合にダメージを受ける．踏むことができない敵は，接触した場合にダメージを受ける．
- **スコアの獲得**
  マリオが死亡する，または，設定された制限時間に達するとプレイは終了し，スコアが表示される．スコアは Mario AI Competition [16] で規定されている評価関数で計算され，ステージを進んだ距離のみに応じてスコアが変動する（右に 1 ピクセル進むごとに 1.0 加算される）．
- **NPC の観測情報**
  NPC は，マリオの座標，マリオの状態（スーパーマリオかちびマリオか），敵の座標および種類，地形の座標および種類を観測する．ただし，観測可能な敵座標と地形座標は，マリオを中心に 22×22 ブロック（1 ブロックは 16×16 ピクセル）の範囲内のみである．NPC は毎フレーム観測情報を受け取り，マリオの行動制御を行うためのキー入力を返す必要がある．

4.3 Q 学習における “Infinite Mario Bros.”

“Infinite Mario Bros.” は，NPC が観測可能な 22×22 ブロックの範囲内に，4 種類のブロック，11 種類の敵キャラクタ，3 種類のアイテムが配置される可能性がある．そのため，全配置の組合せを考慮すると，現実的な時間で学習が収束しない可能性が高い．そこで，NPC が観測するゲーム局面 s の次元を，ゲームの攻略に重要でない情報を削減することで，以下のとおりに圧縮する．

- **マリオを中心に 7×7 ブロックの敵と地形の情報**
  1 フレームあたりのマリオの移動距離は小さく，画面内すべての敵座標や地形座標がマリオの行動に影響することはない．そこで，学習に使用する敵座標と地形座標は，マリオを中心に 7×7 ブロックの範囲内とする（**図 2**）．また，ブロックの種類は通行不可能ブロックと上に飛び乗れるブロックの 2 種類，敵の種類は踏める敵と踏めない敵の 2 種類，アイテムは考慮しないこととする．これにより，ゲーム局面 s の次元数は大幅に削減される．
- **マリオの進行方向**
  敵や地形との関係性を把握するための重要な要素であるため，NPC は 8 方向 + 停止の 9 次元としてマリオの進行方向を把握しておく必要がある．
- **「スーパーマリオ」か「ちびマリオ」か**
  「スーパーマリオ」でダメージを受けた場合は「ちびマ

リオ」に変化するだけでプレイを続行できるが，「ちびマリオ」でダメージを受けた場合は死亡となる．より長くプレイするうえで重要な要素であるため，NPC はマリオの状態を把握しておく必要がある．

- **マリオが地上にいるか**
  マリオは，地上にいる場合はダッシュやジャンプができるが，空中にいる場合はできない仕様である．マリオが地上にいるかどうかは，行動選択にあたって重要な要素であるため，NPC はマリオが地上にいるかどうかを把握しておく必要がある．

次に，Q 学習における選択可能な行動 a の設定方法について述べる．マリオの行動は，コントローラのキー入力によって決定される．マリオの行動制御に影響があるキー入力の組合せは 11 パターン存在する．そこで，選択可能な行動 a として**表 1** のとおり設定する．

最後に，Q 学習における報酬 r の設定方法について述べる．敵を可能な限り避け，ステージをより早く，より遠くまで進むためには，ステージを早く進むことに対して正の報酬を与え，逆にダメージを受ける，死亡するといった，攻略を阻害する要因に対して負の報酬を与えることが望ましい．また，キー操作変更による疲れを実現するため，キー操作を変更した場合は負の報酬を与える必要がある．そこで，報酬 r を以下のとおり設定する．

$$r = distance + damaged + death + keyPress \qquad (3)$$

式 (3) において，$distance$ は現フレーム t から次フレーム $t+1$ の間に行動 a_t によって右に進んだ距離（pixel）であり，正の報酬として与える（ただし，左に進んだ場合は負の報酬となる）．$damaged$ は行動によってダメージを受けた場合に与える負の報酬，$death$ は行動によって死亡した場合に与える負の報酬である．また，$keyPress$ は前フレームから行動を変更した場合に与える負の報酬である．予備実験の結果，$distance$ は右に進んだ距離（pixel）の 2 倍，$damaged$ は -50.0，$death$ は -100.0，$keyPress$ は -5.0 とした．

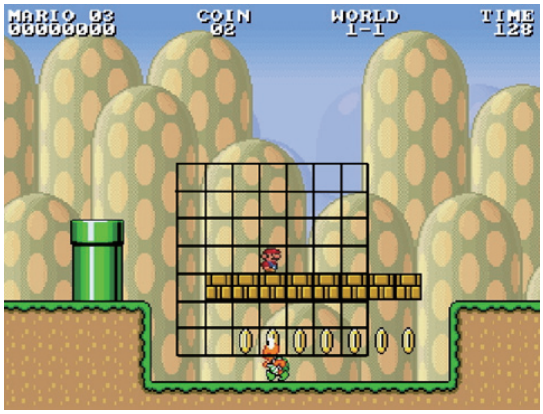

図 2 次元圧縮された観測情報
Fig. 2 Feature extractions from observable information.

表 1 行動の種類とキー入力の組合せ
Table 1 Action patterns of Mario.

行動の種類	(LEFT, RIGHT, DOWN, JUMP, SPEED)
右に歩く	(OFF, ON, OFF, OFF, OFF)
右に走る	(OFF, ON, OFF, OFF, ON)
右に歩きジャンプ	(OFF, ON, OFF, ON, OFF)
右に走りジャンプ	(OFF, ON, OFF, ON, ON)
左に歩く	(ON, OFF, OFF, OFF, OFF)
左に走る	(ON, OFF, OFF, OFF, ON)
左に歩きジャンプ	(ON, OFF, OFF, ON, OFF)
左に走りジャンプ	(ON, OFF, OFF, ON, ON)
真上にジャンプ	(OFF, OFF, OFF, ON, OFF)
しゃがむ	(OFF, OFF, ON, OFF, OFF)
静止	(OFF, OFF, OFF, OFF, OFF)

5. 獲得された振舞いの比較

人間の生物学的制約を導入した Q 学習エージェントの振舞いが，人間の生物学的制約を導入していない振舞いと比べ，どのように変化したかを検証する．用意した 2 つの Q 学習エージェント（**表 2**）について，Q 学習 1 は，ゆらぎ，遅れ，疲れを導入せず，また，失敗の有無にかかわらずランダム行動選択確率 ϵ が 0.0125 のままのエージェントである．Q 学習 2 は，ゆらぎ，遅れ，疲れを導入し，かつ，失敗をほとんどしないゲーム局面での ϵ は 0.0125，失敗を繰り返しているゲーム局面での ϵ は 0.2125 とすることで「訓練と挑戦のバランス」を実現したエージェントである．これらのエージェントにおいて，毎試行同じステージが生成されるようにした状態で 5 万プレイの学習試行を行い，最も高いスコアを獲得したプレイにおける振舞いを比較した．

予備実験として，20〜26 歳の男女 10 名（男性 8 名，女性 2 名）を対象に，人間の生物学的制約を導入して獲得された振舞いと，導入せずに獲得された振舞いとを比較させ，どういう点が人間らしい振舞いか，自由記述で回答させた．その結果の一部を以下に示す．

マリオの『人間らしい』振舞い

- コントローラの操作が不慣れであるかのような動き．
- 敵や穴に対して一瞬ためらう．

表 2 振舞いを比較する Q 学習エージェント
Table 2 Two video game agents for comparing behavioral patterns.

Q 学習エージェント	Q 学習 1	Q 学習 2
生物学的制約の導入	なし	あり
ゆらぎ（pixel）	0	4
遅れ（frame）(秒)	0 (0.0)	3 (0.125)
疲れ（$keyPress=$）	0.0	-5.0
学習率 α	0.2	0.2
割引率 γ	0.9	0.9
ランダム選択確率 ϵ	つねに 0.0125	0.0125（失敗繰返し時：0.2125）

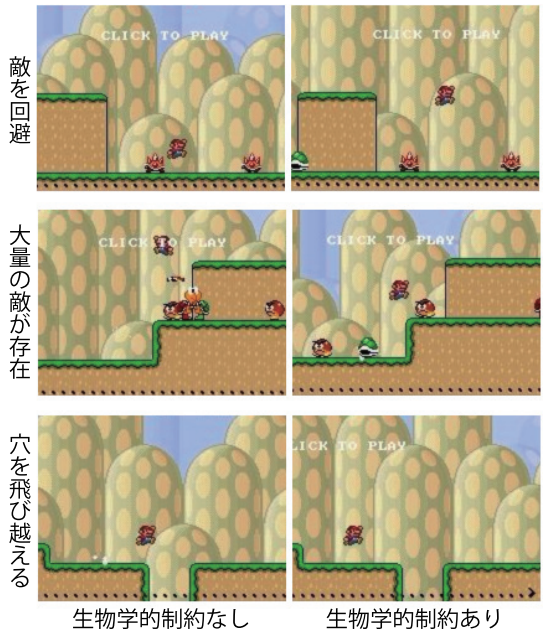

図 3 導入なし（左列）と導入あり（右列）での振舞いの比較
Fig. 3 Comparison of agent without biological constraints (left) and agent with them (right).

- 敵や穴を安全に飛び越えようとする．
- ジャンプの高さやダッシュのスピードが一定でない．

予備実験の回答どおり，人間の生物学的制約の導入の有無によって，表出した振舞いの特徴に差異があった（**図 3**）．

触れることができない敵を回避する場面（図 3 上段）

- 導入なし（左）：最小限のジャンプ，かつ，ノンストップで攻略
- 導入あり（右）：大きくジャンプし，途中で一瞬止まるような行動をしつつ攻略

5 体の敵が段差の上に存在する場面（図 3 中段）

- 導入なし（左）：正確な行動制御で敵が大量に存在する区間を攻略
- 導入あり（右）：区間の手前で待機し，安全に進める状態に変化してから攻略

穴を飛び越える場面（図 3 下段）

- 導入なし（左）：穴に落ちる寸前のところから最小限のジャンプで攻略
- 導入あり（右）：穴の少し手前から大きくジャンプし余裕を持って攻略

以上の結果から，人間の生物学的制約の導入なしでは，パフォーマンスのみを重視しているが，導入ありでは，安全性も考慮した振舞いが獲得できているといえる．人間の生物学的制約を Q 学習に組み込むことで，獲得される振舞いの特徴が変化することが示された．

6. 主観評価実験

6.1 実験計画

人間の生物学的制約を導入した Q 学習エージェントによって獲得された振舞いが，本当に人間らしいかどうかを検証するため，20〜24 歳の男女 20 名（男性 13 名，女性 7 名）を対象に主観評価実験を実施した．実験参加者 20 名における，横スクロール型マリオのプレイ時間の累計は平均 $\mu = 34$ 時間，標準偏差 $\sigma = 29$ 時間であった．そこで，本実験においては，横スクロール型マリオの熟練度を 3 つのグループに分類した．横スクロール型マリオのプレイ時間が 5 時間（$\mu - \sigma$）未満の実験参加者 4 名を「初級者」（うち，3 名はプレイ時間が 0 時間の初心者），63 時間（$\mu + \sigma$）以上の実験参加者 2 名を「上級者」，5 時間以上 63 時間未満の実験参加者 14 名を「中級者」と定義した．

表 3 プレイ動画のラベルと内容
Table 3 Prepared videos and their labels.

ラベル	操作者	生物学的制約	再生時間	スコア
[強化, なし]	強化学習（NPC）	導入なし	10.62 秒	5,448
[強化, 導入]	強化学習（NPC）	導入あり	14.25 秒	4,069
[強化, 導入（挑戦のみ）]	強化学習（NPC）	導入あり（挑戦のみ）	15.57 秒	3,458
[中級者]	中級者（人間）	—	10.08 秒	6,031
[初級者]	初級者（人間）	—	14.25 秒	3,644
[上級者]	上級者（人間）	—	7.68 秒	7,371

実験手続きは以下のとおりである．まず，実験参加者に「ブロック，アイテム，コインなどは無視して，ステージの先に進め」と教示し，“Infinite Mario Bros.” を 10 回プレイ（1 プレイ 25 秒）させた．次に，プレイ動画を 2 つずつ比較させ「どちらのマリオが人間らしいプレイか」を 7 段階で評価させた．最後に，プレイ動画を 1 つずつ見せ「どのような振舞いが人間らしい（人間らしくない）と感じたか」を自由記述で回答させた．

実験に使用したプレイ動画を**表 3** に示す．本実験では，Q 学習エージェントによるプレイ動画を 3 つ，人間が操作したプレイ動画を上記熟練度を考慮して 3 つ用意した．Q 学習エージェントに関しては，5 章で振舞いを比較した表 2 の 2 つに加え，訓練をせず失敗に対する挑戦のみを実施するエージェントを 1 つ用意した．この Q 学習エージェントにおけるランダム行動選択確率 ϵ は 0.0，ただし，失敗を繰り返しているゲーム局面での ϵ は 0.2 と設定した．これにより，失敗をしないゲーム局面では新たな行動を選択しないため訓練ができず，失敗をしたゲーム局面でのみ新たな行動に挑戦するエージェントが生成される．人間の操作者に関しては，初級者動画は 2D 横スクロール型マリオのプレイ時間が 5 時間の人間プレイヤ，中級者動画は 50 時間の人間プレイヤ，上級者動画は 200 時間の人間プレイヤが操作したものである．また，敵，土管，穴といった障害物の有無や，マリオが敵に接触しダメージをうけるシーンが，人間らしさの評価に大きく影響を与えると考えられる．そ

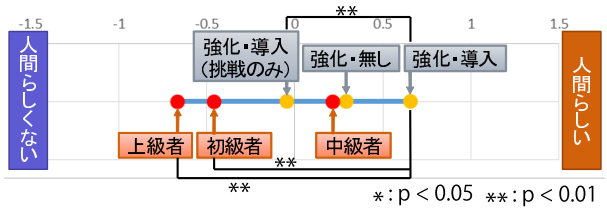

図 4 人間らしさに関する相対的嗜好度
Fig. 4 Relative preferences for human-like behaviors on 20 participants.

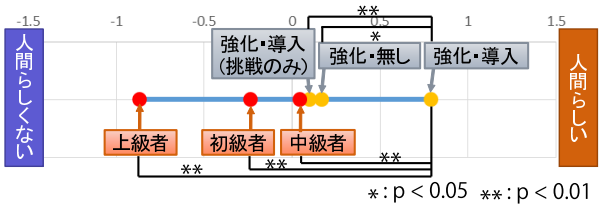

図 5 人間らしさに関する相対的嗜好度．速さと相関ありの実験参加者 5 名を除いた 15 名での分析結果．
Fig. 5 Relative preferences for human-like behaviors on 15 participants except for 5 in Speedy Group.

こで，すべての動画でプレイ区間を統一し，マリオが敵に接触しダメージをうけたプレイ区間は不採用とした．表 3 の再生時間とは，統一されたプレイ区間におけるリプレイ時間を意味する．これ以降，プレイ動画を表 3 のラベル名で表記する．

6.2 分析手法と結果

本実験では，ランダムな組合せで表示される 2 つのプレイ動画を比較し，人間らしさについて 7 段階で評価する．統計的分析手法としてシェッフェの一対比較法（中屋の変法）[19] を使用し，分散分析で主効果に対する有意差の有無を確認する．その後，ヤードスティック法によりプレイ動画の嗜好度を一本の直線上にプロットし，動画どうしの相対的な関係性と，信頼区間について検討する．

図 4 は，人間らしさに関する相対的嗜好度を直線上にプロットしたものである．まず，Q 学習エージェントどうしの比較結果を述べる．人間の生物学的制約を導入した [強化, 導入]（相対的嗜好度：0.66）は，人間の生物学的制約を導入していない [強化, なし]（相対的嗜好度：0.29）と比較して，人間らしいという結果が得られた．しかしながら，相対的嗜好度の差（$0.66 - 0.29 = 0.37$）が 95%信頼区間である 0.48 より小さいため，5%水準の有意差は認められなかった．この結果を，以降（差：0.37 < 95%信頼区間：0.48）と表記する．

次に，Q 学習エージェントと人間プレイヤの比較結果を述べる．人間の生物学的制約を導入した [強化, 導入] は，人間プレイヤの [初級者][中級者][上級者] より人間らしいという結果が得られた．ただし，有意差が認められたのは，[強化, 導入] と [初級者]（差：1.12 > 99%信頼区間：0.58），[強化, 導入] と [上級者]（差：1.33 > 99%信頼区間：0.58）であり，[強化, 導入] と [中級者]（差：0.44 < 95%信頼区間：0.48）では有意差が認められなかった．

図 4 において，[強化, 導入] と [強化, なし]，[強化, 導入] と [中級者] の間に有意差が認められなかった理由として，実験参加者間で人間らしさの評価基準が異なっている可能性が高い．マリオの振舞いの包括的な指標として表 3 の再生時間があり，それは，統一されたプレイ区間におけるマリオの平均スピードを意味している．そこで，各動画のマリオの平均スピード（表 3 の再生時間）と，その動画に対する実験参加者の人間らしさの評価点数との相関係数 r を算出したところ，相関の有無が認められたため，相関係数 r によって 20 名の実験参加者を以下の 3 つのグループに分類する．

速さと相関あり（$0.4 < r$） マリオのスピードが速い方が人間らしいと評価する傾向にある実験参加者 5 名が該当．マリオの累計プレイ時間は平均 46 時間．

速さと逆相関あり（$r < -0.4$） マリオのスピードが遅い方が人間らしいと評価する傾向にある実験参加者 9 名が該当．マリオの累計プレイ時間は平均 25 時間．

スピードと相関なし（$-0.4 \leq r \leq 0.4$） マリオのスピードと人間らしさの評価に相関がない実験参加者 6 名が該当．マリオの累計プレイ時間は平均 36 時間．

図 5 は，速さと相関ありの実験参加者 5 名を除いた 15 名での分析結果である．図 4 とは異なり，[強化, 導入] は [強化, なし] より有意に人間らしく（差：0.62 > 95%信頼区間：0.57），かつ，[中級者] より有意に人間らしい（差：0.74 > 99%信頼区間：0.67）という結果が得られた．

7. 考察

主観評価実験の結果から，人間の生物学的制約を Q 学習に導入することによって，人間らしさの評価基準は実験参加者間で異なっているものの，多数の実験参加者にとってより『人間らしい』と解される NPC を自動的に構成できることが示された．では，「どのような振舞いが人間らしいのか」について，主観評価実験の結果（図 4，図 5 と自由記述質問の回答）から考察していく．

[強化, 導入] は全動画中で最も人間らしいと評価されている．自由記述質問では，人間らしいと感じる理由として「敵や穴を飛び越えるときに躊躇する」，「敵や穴を大きく飛び越える」，「ときどき不必要な行動をとる」という回答があった．この結果から，「ためらい」や「余裕」，「熟慮（試行錯誤）」が人間らしさを感じさせる要素であり，人間らしい振舞いの獲得に『人間の生物学的制約』の導入が妥当であることが示された．

[上級者] は人間プレイヤの操作であるにもかかわらず，人間らしくないという評価を得ている．自由記述質問では，人間らしくないと感じる理由として「敵や穴をギリギ

リまで避けない」,「無駄な行動がいっさいない」,「動きが一定である」という回答があった．この動画は，全動画中で最も高いスコアであることからも，「過度に最適化された振舞いは人間らしくない」ことを意味している．2.1 節で述べたとおり，強い NPC は人間プレイヤの代替として扱うことはできないといえる．

[強化, 導入（挑戦のみ)] は，人間の生物学的制約を導入しているにもかかわらず，人間らしくないという評価を得ている．自由記述質問では「段差や土管にぶつかってからジャンプする振舞いが人間らしくない」という回答があった．この動画は，全動画中で最も低いスコアであり，その「たどたどしい」振舞いは，コントローラの操作やゲームのルールに慣れていない，あたかもゲーム初心者の操作のようであった．この結果は,「初心者相当の下手すぎる振舞いは人間らしくない」ことを意味している．3.1 節で述べたとおり，急に弱くなったりあからさまな操作ミスをする NPC もまた，人間プレイヤの代替として扱うことはできないといえる．しかし,「訓練と挑戦のバランス」を変化させることで，人間の熟達過程を再現できる可能性があるともいえる．

図 4 と図 5 の結果の差異から，マリオのスピードは人間らしさの評価基準の 1 つであり，実験参加者によって人間らしいと感じるスピードが異なることが示唆された．速さと相関ありの実験参加者 5 名を除いた 15 名にとっては，生物学的制約の導入が NPC の振舞いの人間らしさを向上させる要因の 1 つであると考えられる．また，速さと相関ありの実験参加者のマリオの累計プレイ時間が平均 46 時間，スピードと相関なしの実験参加者は平均 36 時間，速さと逆相関ありの実験参加者は平均 25 時間であることから，実験参加者のゲームへの熟練度が人間らしさの評価基準を決定づけている可能性も示唆された．

8. おわりに

ビデオゲームにおけるユーザエクスペリエンスの向上には，人間らしい NPC の実装が必要不可欠であり，その自動獲得がゲーム AI 領域の最重要課題となりつつある．人間らしい NPC の自動獲得には，従来，ゲームジャンルやゲームタイトルに合った人間らしさの解析が必要であり，ゲームプログラマの作業負荷（開発コスト）は多大であった．

本研究では，人間の生物学的制約を強化学習に導入することで，多数の人間プレイヤにとってより『人間らしい』と解される NPC を自動的に構成できることが示された．主観評価実験では，人間の生物学的制約を導入した Q 学習エージェントが，導入していないエージェントや人間プレイヤよりも人間らしい振舞いを獲得できていることを示した．また，マリオのスピードは人間らしさの評価基準の 1 つであり，かつ，実験参加者間で人間らしさの評価基準が異なっていることが示唆された．人間の生物学的制約は，開発者のヒューリスティックや，ゲームタイトルごとの人間らしさの解析に依拠しない要素である．したがって，あるゲーム状況を入力とし，そのゲーム状況で最適な行動をリアルタイムに出力する必要があるゲームであれば，ゲームジャンルや振舞い獲得の手法を問わず，人間の生物学的制約の導入によって人間らしい NPC の振舞いを獲得できると考えられる．

本研究の振舞い獲得手法を使用することで，人間らしい NPC を実装したいゲームプログラマにとって，以下の 3 つのメリットがある．1) ヒューリスティックの導入に係る煩多な作業負荷（開発コスト）を削減できる．2) 人間が持つ生物学的制約であるため生理学的・心理学的知見に基づいて設定できる．3) リアルタイムな操作を要求する様々なゲームジャンル，様々な機械学習手法に対しても，汎用的に導入できる．また，人間らしい NPC が実現することで，ゲームをプレイする人間プレイヤにとって，満足感の確保やエンタテインメント性の持続といった，ユーザエクスペリエンスの向上につながると考えられる．

今後の展望として，実験参加者をより精密に群分けし人間らしさの評価基準を特定する，Q 学習以外の非事例参照型手法にも人間の生物学的制約の導入を試みる，アクションゲーム以外のジャンルへの生物学的制約の導入を目指す．

参考文献

[1] Norman, D.: *Invisible Computer: Why Good Products Can Fail, the Personal Computer Is So Complex and Information Appliances Are the Solution*, MIT Press (1999).

[2] エンターブレイングローバルマーケティング局：ファミ通ゲーム白書 2013，エンターブレイン (2013).

[3] Togelius, J., Karakovskiy, S. and Baumgarten, R.: The 2009 Mario AI Competition, *Evolutionary Computation (CEC) 2010 IEEE*, pp.1–8 (2010).

[4] Fujita, H. and Ishii, S.: Model-based reinforcement learning for partially observable games with sampling-based state estimation, *Neural Computation*, Vol.19, pp.3051–3087 (2007).

[5] Hoki, K. and Kaneko, T.: The Global Landscape of Objective Functions for the Optimization of Shogi Piece Values with a Game-Tree Search, *Advances in Computer Games 2012, Lecture Notes in Computer Science*, Vol.7168, pp.184–195 (2012).

[6] Fujii, N., Hashida, M. and Katayose, H.: Strategy-acquisition System for Video Trading Card Game, *International Conference on Advances in Computer Entertainment Technology 2008 (ACE 2008)*, pp.175–182 (2008).

[7] Schrum, J., Karpov, I.V. and Miikkulainen, R.: Human-like Behavior via Neuroevolution of Combat Behavior and Replay of Human Traces, *2011 IEEE Conference CIG'11*, pp.329–336 (2011).

[8] 池田　心，Viennot, S.：モンテカルロ碁における多様な戦略の演出と形勢の制御～接待碁 AI に向けて～, *GPW2012*, pp.47–54 (2012).

[9] Ortega, J., Shaker, N., Togelius, J. and Yannakakis, G.N.: Imitating human playing styles in Super Mario

Bros., Vol.4, pp.93–104 (2013).
[10] Cabrera, J.L. and Milton, J.G.: On-Off Intermittency in a Human Balancing Task, *Physical Review Letters*, Vol.89, No.15 (2002).
[11] 大平 徹，保坂忠明：不安定な状況でのノイズと遅れの役割と制御への考察，交通流のシミュレーションシンポジウム，pp.19–22 (2004).
[12] Maslow, A.H.: A Theory of Human Motivation, *Psychological Review*, Vol.50, pp.370–396 (1943).
[13] 保木邦仁：局面評価の学習を目指した探索結果の最適制御，*GPW2006*, pp.78–83 (2006).
[14] Tsay, J.-J., Chen, C.-C. and Hsu, J.-J.: Evolving Intelligent Mario Controller by Reinforcement Learning, pp.266–272 (2011).
[15] Sugiyama, T., Obata, T., Hoki, K. and Ito, T.: Optimistic Selection Rule Better Than Majority Voting System, *Computers and Games, Lecture Notes in Computer Science*, Vol.6515, pp.166–175 (2011).
[16] Togelius, J., Karakovskiy, S., Koutnik, J. and Schmidhuber, J.: Super Mario Evolution, *2009 IEEE Conference CIG'09*, pp.156–161 (2009).
[17] Watkins, C.: *Learning from Delayed Rewards*, PhD thesis, Cambridge University, Cambridge, England (1989).
[18] Patel, P.G., Carver, N. and Rahimi, S.: Tuning Computer Gaming Agents using Q-Learning, pp.581–588 (2011).
[19] Scheffe, H.: An Analysis of Variance for Paired Comparisons, *Journal of the American Statistical Association*, Vol.47, No.259, pp.381–400 (1952).

藤井 叙人 （学生会員）

2009 年関西学院大学大学院理工学研究科情報科学専攻修士課程修了．2009〜2012 年株式会社野村総合研究所勤務を経て，現在，関西学院大学理工学研究科人間システム工学専攻博士課程．日本学術振興会特別研究員（DC2）．研究テーマ「ビデオゲームエージェントにおける人間らしい行動戦略の自動獲得」．

佐藤 祐一

2010 年関西学院大学理工学部情報科学科卒業．2012 年同大学大学院理工学研究科情報科学専攻修士課程修了．

若間 弘典

2010 年関西学院大学理工学部情報科学科卒業．2012 年同大学大学院理工学研究科情報科学専攻修士課程修了．

風井 浩志 （正会員）

1998 年関西学院大学大学院文学研究科博士課程単位取得退学．現在，関西学院大学大学院理工学研究科専門技術員．日本心理学会，日本生理心理学会各会員．博士（心理学）．

片寄 晴弘 （正会員）

1991 年大阪大学大学院基礎工学研究科博士課程修了．博士（工学）．イメージ情報科学研究所，和歌山大学を経て，現在，関西学院大学理工学部教授．音楽情報処理，感性情報処理，HCI の研究に従事．科学技術振興機構さきがけ研究 21「協調と制御」領域研究者．科学技術振興機構 CREST「デジタルメディア（略称）」領域 CrestMuse プロジェクト研究代表者．電子情報通信学会，人工知能学会各会員．