

身体性を持つCG対話エージェントにおける カートゥーン調表現の方法論および比較評価

Cartoon-Style Expressions on Embodied Conversational Agent: Methodologies and Comparison

川又 朱莉¹ 上乃 聖¹ 李 晃伸¹
Akari Kawamata¹ Sei Ueno¹ Akinobu Lee¹

¹ 名古屋工業大学 大学院工学専攻

¹ Department of Engineering Nagoya Institute of Technology

Abstract: 擬人化CGキャラクターを用いた音声対話システムは、一般に人間らしさを重視した自然な表現やふるまいを設計し、違和感や話しづらさを解消する。しかし、CGキャラクターにはアニメ・漫画等で見られる特有の動作表現（カートゥーン調表現）が存在するため、人間らしさを追求することが最適な設計法ではない可能性がある。本研究では、アニメ・漫画のキャラクター表現技法を元にCGキャラクターらしさを重視した動作表現を行うCGエージェントを提案し、対話における印象評価を行った。その結果、CGキャラクターらしいふるまいはエージェントの生命感や存在感を高め、ユーザの対話意欲を促進することが確認された。一方、公共の場での情報提供タスクにおいて、エージェントが対話相手として認識されると、ユーザは知性や信頼性を感じやすい人間らしいふるまいを好む傾向があることが分かった。

1 はじめに

近年、駅や商業施設での案内業務や携帯端末でのアシスタントサービスなど、擬人化キャラクターをインターフェースとして用いて情報を提供するシステム（音声対話CGエージェント）を用いた様々なサービスの利用が広く検討されている [1][2]。CGエージェントは、スマートスピーカなど音声のみを用いた対話システムに比べて、より人間らしい自然な感情表現やふるまいが求められ、表情やジェスチャといったノンバーバルな情報を人間に似せることで、ユーザとの豊かなコミュニケーションの実現を目指す研究が多く進められている [3][4]。

しかし「不気味の谷現象」では、エージェントと人間の見た目や性質の類似度がある一定のラインを超えると、不気味さや強い違和感を感じるという現象が知られており、CGエージェントにおいても、その感情表現やふるまいを人間に近づけることが必ずしも適切なインタラクション設計とは限らない。また、ユーザはエージェントの見た目から類推される機能に応じて発話内容や態度を変える傾向があるため [5]、CGエージェントのキャラクター性の強さに矛盾しないインタラクションを設計する必要がある。

漫画やアニメーションにおいては、漫符や効果線、キャラクターの誇張された動作や感情表現といった独特

の表現様式（以下、カートゥーン調表現と呼ぶ）が使われている。これらの表現は、キャラクターの心情変化を強調し、視覚的に分かりやすく伝える効果がある。このことから、CGエージェントにおいてもカートゥーン調表現を用いることで、人間らしさを追求するよりも違和感の少ないインタラクションを実現できる可能性がある。一方、CGエージェントが人間同士の対話を模倣する存在と捉えるのであれば、そのふるまいは人間に近い方が適している可能性もある。しかし、既存の対話システムの多くは動作や表現が限定的であり、対話中のエージェントがキャラクターらしいカートゥーン調表現を使うべきか、人間らしいふるまいを重視すべきかについては、十分な知見が得られていない。

そこで本研究では、CGエージェントのインタラクションにおける人間らしさがユーザの印象に与える影響を明らかにすることを目的とする。具体的には、カートゥーン調表現を用いたキャラクターらしさを重視したインタラクションを設計し、エージェントに対する話しづらさや対話における違和感の改善にどのように寄与するか検討する。

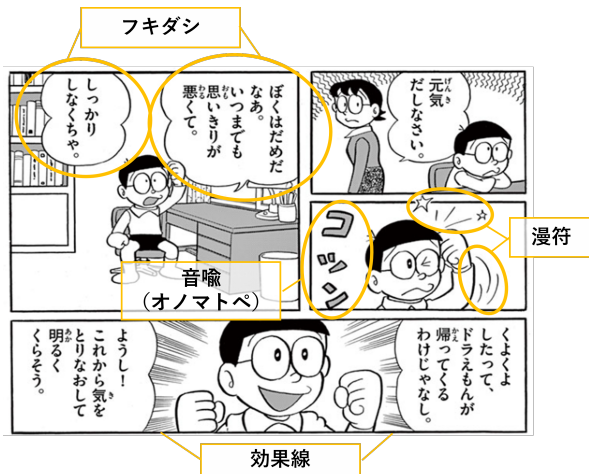


図 1: 漫画における記号的表現の例 [7]

2 CG キャラクタにおけるカートゥーン調表現

メディアコンテンツには、作中のキャラクターの内面を視覚的に表現したり、場の臨場感やキャラクターの生命感を高めるためのキャラクター表現技法（カートゥーン調表現）が用いられる。例えば漫画では、キャラクターの感情や状況を視覚的かつ簡潔に伝えるために、効果線、フキダシ、漫符、音喩（オノマトペ）などの記号的表現が用いられる（図 1）。また、海外のカートゥーン作品においては、キャラクターの動作表現として“12 Basic Principle of Animation（アニメーションの 12 の原則）” [6] が適用されることが多い。これには、物理法則に従った動きのリアリティを追求する再現から、臨場感やキャラクターの感情を引き立たせるための現実を超えた誇張表現などが含まれる。

音声対話エージェントにおいても、このような特有の表現技法を取り入れることで、より親しみやすいエージェントや、対話中の臨場感を生み出すインタラク션을提案する。本研究では、音声対話エージェントのためのカートゥーン調表現として、表現上の制約を受けず、対話動作に関連する表現を選別した。その結果 (1) 漫符 (comic sign, manpu) と、動作の誇張表現である (2) 誇張ジェスチャ (gesture exaggeration) の 2 種類を、CG エージェントのためのカートゥーン調表現である「CG 特有表現」とし、これらを含んだインタラク션을行うエージェントを構築する。以下、採用した表現の詳細を示す。

(1) 漫符 (comic sign, manpu)

漫符とは、漫画でキャラクターの表情や動作に付与される記号やアイコンであり、感情や状況を視覚的に

表現する役割を持つ。複数の漫符を組み合わせることで、感情を強調したり、場面の状況をより明確に伝えることができる。音声対話 CG エージェントにおいても、対話の状況に応じて適切な漫符を表示することで、エージェントの心理状態を視覚的に表現し、ユーザーに分かりやすく伝えることができると考えられる。

(2) 誇張ジェスチャ (gesture exaggeration)

本研究では、対話相手としてのキャラクターを過度に変形させるのは望ましくないとの立場から、12 Basic Principle of Animation にある表現技法の中で、対話における基本動作にあたり、かつモデルを大きく変形させない以下の 4 規則を誇張ジェスチャとして採用する。

- i. 予備動作 (Anticipation)
副次アクションを行うまでのタメ動作
- ii. 両端詰め (Slow In and Slow Out)
動きははじめから徐々に速くなり、終わりに向けて徐々に遅くなる動作表現
- iii. 後追いの工夫 (Follow Through and Overlapping Action)
身体が動く際、動き始めた部位に近い順に動き出す動作表現
- iv. 副次アクション (Secondary Action)
感情を引き立たせる副次的な動作

3 提案システム

3.1 CG 特有表現を用いた対話の設計

本来は機械対話において、全ての動きに関して誇張ジェスチャをつけたり、場面に応じて漫符を表示するシステムが理想的であるが、あらゆる発話にこれらを適用することは難しい。そのため、本研究ではシチュエーションを限定し、それに沿った CG 特有表現を含むモーションを作成する。また、被験者は実際の対話を行わず、CG エージェントがユーザー役の演者と対話する様子を収録した映像を視聴する形で、対話中のエージェントの印象を評価する。対話タスクは、旅行案内エージェントによる情報提供とし、ユーザーが CG エージェントに卒業旅行の行き先を相談するものとする。ユーザーが行き先に悩んでいるという前提の元、エージェントは 2 種類の観光地を提案し、ユーザーの悩みに共感しながら観光地の魅力を伝える、というシナリオを採用する。

このシナリオに基づいて計 6 種類の CG 特有表現を用いたふるまいモーションを用いる。本シナリオにお

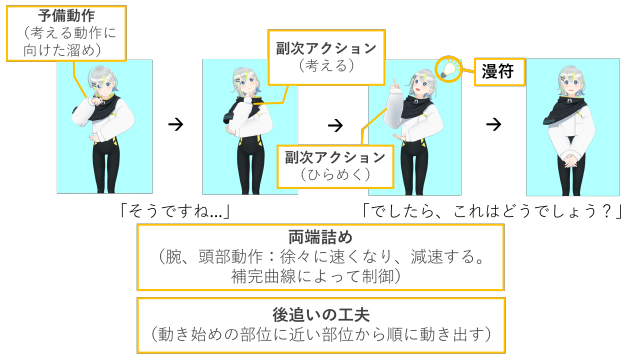


図 2: CG 特有のキャラクタ表現を用いたエージェントのふるまい例

けるふるまいの一例を図 2 に示す。ここでは、ユーザの質問に対してエージェントがしばらく考えた後、良い提案を思いついたという一連の動きを表現する。冒頭の「考える」動作では、大きく腕を広げ、膝を曲げながら腕を組む動作を行い、あごに手を当てる動作を行う。その後、良い案を思いついたことを示す「ひらめき」動作として、人差し指を立てるポーズをとる。このとき、「ひらめき」を示す「電球」の漫符を発言に合わせて表示する。一連の動作には両端詰めの表現を用いる。また全身の動きには後追いの工夫の表現を加え、動き出しに近い部位、例えば頭部から順に腕や手先を動かすよう設計し、「ひらめき」動作後には、停止前にわずかに戻る動きを加える。

各モーションは、3DCG ソフトウェアである Miku-MikuDance を使用して作成し、モデルへの漫符画像の追加は、PMXEditor にて漫符の画像をモデルのモーフとして追加する形で行った。参考として、作成した 6 種のふるまいモーションを図 A.1 に示す。

3.2 実験システム

システム構成について説明する。本システムで使用するモデルは、ムーンショット型研究開発「アバター共生社会」プロジェクトの一環として、名古屋工業大学で開発された対話エージェントの“ジェネ (gene)”である [8]。システムは MMDAgent-EX [9] を用いて構築する。シナリオ上の発言とモーション再生のタイミングは ボタンで順次再生できるように設計し、ユーザとシステムの会話の進行に合わせてオペレータがキーを押す形で対話を行った。音声合成は DNN-HMM TTS に基づく音声合成を用いた。また、作成した CG 特有表現のふるまいに加え、瞬きやうなずきなど対話における一般的なしぐさや、発言中のリップシンクおよび頭部動作を随時行う。

対話評価実験に先立ち、作成したモーションがシナ



図 3: Human-like agent と CG-like agent のふるまいの比較

リオの発言内容に即しているか、また CG 特有表現として違和感がないかを確認するため、本学の学生 23 名を対象に主観評価実験を行った。実験では、6つのモーションそれぞれについて、対話の状況を説明し、音声の代わりに発言内容の字幕をモーションと共に提示する方式でアンケート評価を実施した。その結果、各モーションと発言内容に違和感はなく、CG 特有表現としても問題なく認知されていることが確認された。

4 対話評価実験

4.1 実験内容・評価方法

提案手法である特有のインタラクションを行うエージェント (提案システム:CG-like agent) を、人間の動作を模倣したインタラクションを行うエージェント (従来システム:Human-like agent) と比較する。従来システムのふるまいモーションは、対話シナリオに沿って人間が演じた様子を Webcam Motion Capture でキャプチャしたものをベースに制作する。図 3 に、両システムのシナリオ上のふるまいの比較画像を示す。提案システムでは、感動や神秘性を示す星形の漫符と、震えながら伸び上がる誇張ジェスチャを特有表現として用いている。

対話映像の評価の際、被験者には事前説明として、図 4 に示す画像と共に、ユーザが大学の生協に設置された旅行相談エージェントに卒業旅行の行き先を相談する内容の映像であることを伝える。

被験者は、映像を視聴する前に、自身の音声対話エージェントに対する理解度や受容度、アニメや漫画を見る習慣などキャラクタに対する受容度を問う質問 全 11 項目に回答する。アンケートに回答後、大きく分けて 2 種類の評価を行う。



図 4: 対話の状況を示す写真

表 1: AB テスト 質問項目一覧

よりあてはまる方を選択して下さい。	
(1)	実際にこの場で旅行相談をしたら、どちらのエージェントとより話をしたいと思うか
(2)	どちらのふるまいや表現が、今回使用したエージェント (ジェネ) により適している (違和感が少ない) と感じるか
(3)	実際に話すとしたら、どちらのエージェントの方がストレスが少ないと感じるか
(4)	どちらのエージェントが好きか
(5)	どちらのエージェントのふるまいが、話をする相手としてふさわしいと思うか

まず、従来システムおよび提案システムそれぞれについて、単独で評価を行う。各映像を視聴後、形容詞対項目 (例えば「不活発な — 対話的な」) で与えられる相反する印象のどちらにより近いかを 7 段階で評定する 23 項目と、CG エージェントのふるまいや対話全体の印象、エージェントに対する対話意欲などを 7 段階で評価する 7 項目に回答する。

次に、従来システムおよび提案システムの映像を同時に視聴することによる AB テスト形式での比較評価を行う。両システムの対話の様子を同時に視聴し、5 項目 (表 1) に対しどちらのエージェントがより当てはまるか、2 択で回答する。

4.2 実験結果

被験者は本学の学生 36 名 (男性 26 名, 女性 10 名) である。そのうち、情報系の学生は 24 名である。ここでは、主要な評価結果を抜粋して示す。

形容詞対を用いたエージェントの印象評価の結果を図 5 に示す。対応のある t 検定 (両側検定, 有意水準 5%) を行った結果, エージェントの生命感, 愉快さ,

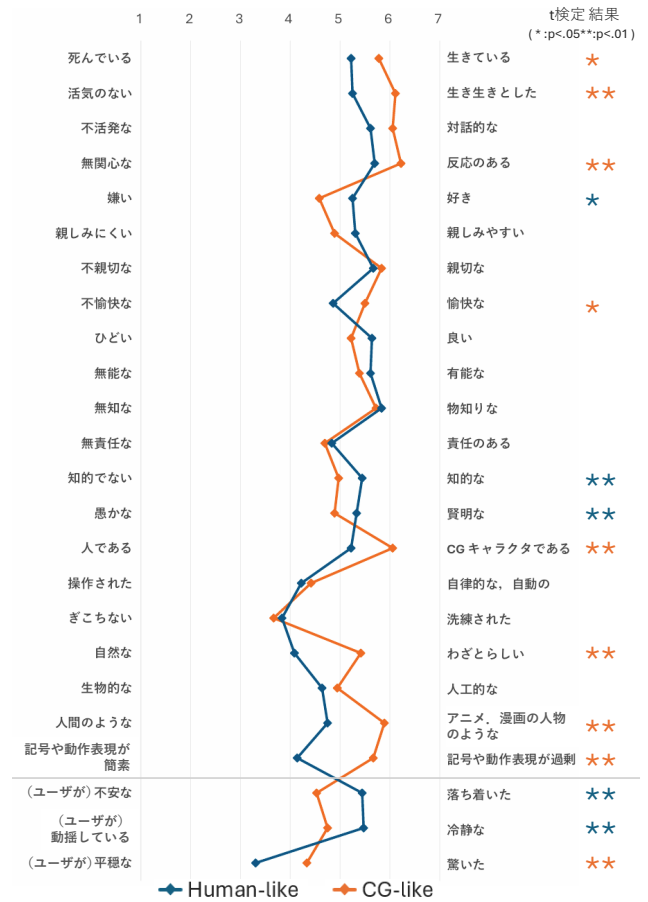


図 5: 形容詞対によるエージェントの印象評価

動きのキャラクターらしさに関する 8 項目, およびユーザの驚きを測る項目で, CG キャラクターらしいふるまいが有意に高い値を示した。一方, 人間らしいふるまいの方が知的さや賢明さの 2 項目, およびユーザの落ち着き, 冷静さの 2 項目で有意に高い値を示した。これらから, CG キャラクターらしいふるまいの方が生命感や存在感が高く, ユーザが興奮を感じるが, 人間らしいふるまいの方が優れた知能を持ち, ユーザが落ち着いて対話しやすいと感じる傾向があるといえる。

続いて, AB テストの結果のグラフを図 6 に示す。「実際に話す際にストレスが少ない」については 36 名中 26 名, 「対話相手にふさわしい」については 23 名が Human-like agent を選択しており, 対話相手のふるまいとして全体的に人間らしいふるまいを好む傾向が見られた。一方, 「実際に話をしたい」「好き」と感じるエージェントについては結果がほぼ均等に分かれた。このことから, CG 特有表現を用いたふるまいは, 対話中の話しにくさや違和感につながる場合があると考えられる。また, 本実験は公共の場に設置されたシステムとの対話を想定したが, Human-like agent を選択した理由として, CG-like agent は「人前で話すのが恥ず

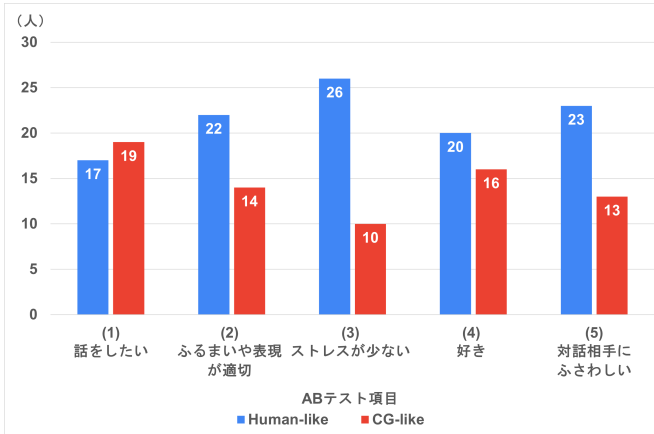


図 6: AB テスト結果

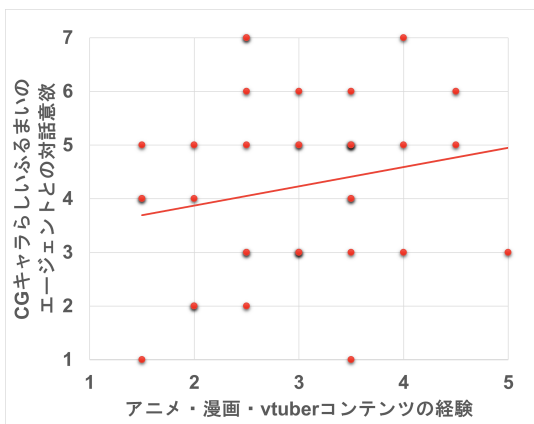


図 7: アニメや漫画, VTuber コンテンツの視聴頻度と CG-like agent に対する対話意欲の関連性

かしい」「落ち着かない」などと回答した被験者がいた (4名)。このことから、オープンスペースでは、CG キャラクターらしい動作が好ましくないと捉えられた可能性がある。

被験者自身に関するアンケート結果と、CG エージェントとの対話意欲に関する評価値の関連性を図 7 に示す。日常的なメディアコンテンツの視聴頻度と CG-like agent に対する対話意欲には、強い相関は見られなかった (相関係数 0.202)。またメディアコンテンツの視聴頻度と特有表現に対する興味の相関係数は 0.183 であったことから、本実験ではアニメや漫画の視聴頻度と特有表現の受容度には強い関連性はなかった。

また、エージェントのふるまいや対話意欲に関する質問項目と評価の平均値を表 2 に示す。エージェントの心情変化や反応のわかりやすさ、エージェントへの興味は CG-like agent の方が平均値が高かった。一方で、ストレスや違和感の項目も高く、CG 特有表現に抵抗感を抱く被験者が一定数存在することが示された。この傾向は図 5 や図 6 の結果と一致している。

表 2: エージェントのふるまいや対話意欲に関する質問項目と平均値一覧

評価項目 (1: 全くそう思わない 7: 非常にそう思う)	平均値	
	Human	CG
(1) エージェントの動作や表情から、エージェントの心情の変化が読み取れる。	4.50	6.083
(2) エージェントが、人の発言に反応したように見える。	5.111	5.806
(3) (あなた自身が) エージェントと もっと話がしたいと思う。	4.444	4.222
(4) エージェントが次にどんな動きをするのか興味がある。	4.056	4.722
(5) 映像中のエージェントと話すことに抵抗感・ストレスを感じるだろう。	3.444	3.917
(6) エージェントの見た目に対し、対話中のエージェントの行動や表現、ふるまいに違和感がある。	3.083	3.75

以下、CG エージェントのインタラクション設計論、および対話タスクやユーザの好みの観点から本実験を論じる。

CG エージェントのインタラクション設計論

t 検定の結果、CG エージェントの客観評価において、CG-like agent の方が、Human-like agent よりも生命感や存在感が高く、エージェントへの関心を高めることが示された。一方、AB テストでは、エージェントを画面内のキャラクターではなく対話相手として捉えた場合、被験者の 6 割以上が Human-like agent の方が対話中の負担を感じにくく、対話相手としてふさわしいと評価した。これにより、ユーザがエージェントとの対話に主体的に関わる場面では、感情を誇張した表現が過剰となり、対話を妨げる可能性があることが分かった。一方、人間らしさや知性を強調したふるまいの方が、対話においてより自然で違和感が少ないことが分かった。

これらの結果から、公共の場に設置されるエージェントはキャラクターと人間の中間的存在として位置づけられるべきであり、対話前と対話前後で対話のシチュエーションによってふるまいを切り替えるインタラクションデザインが適している可能性がある。

対話タスクやユーザの好み

本実験は公共の場で旅行先の情報を提供するタスク志向型の対話を対象としたが、AB テストの結果、CG エージェントの誇張された感情表現が人前での対話に

おける恥ずかしさや違和感を引き起こす可能性があることが分かった。対話はキャンパス内のオープンスペースで収録され、他者が頻繁に通る環境だったため、その傾向が強く現れたと考えられる。一方で、よりプライベートな環境や雑談タスクでは、話しやすさが向上する可能性がある。また本実験では、CG エージェントの合成音声は両システムで共通のものを使用した。CG 特有表現と音声の不一致を指摘した被験者が一定数存在した (5名)。これは、使用した合成音声は自然な読み上げ調のものであったため、CG 特有の誇張表現と調和せず、違和感を引き起こしたと推察される。

また、AB テストの選択理由を見ると、CG-like agent を選んだ被験者は、主に反応の興味深さや対話の臨場感を理由に挙げた。一方、Human-like agent を選んだ理由の 1 つに、エージェントに的確なアドバイスや情報収集を求め、ふるまいに対しより知性や情報の信憑性が感じられたことが挙げられた (3名)。このことから、システムに対しユーザが何を求めるか、というユーザ側の欲求が、対話中に好まれるインタラクションに影響を与えることが示唆された。今後は、ユーザの好む対話戦略やスタイルを動的に読み取り、それに応じて柔軟に感情表現を変化させるユーザモデルの構築が、最適なインタラクションの設計において重要であるといえる。

5 まとめ

本研究では、漫画、およびアニメーションにおけるカートゥーン調表現のうち (1) 漫符、動作の誇張表現である (2) 誇張ジェスチャの 2 種類を再現したふるまいモーションを制作し、CG キャラクターらしさを重視した動作表現を行うエージェントを提案した。その結果、CG キャラクターらしいふるまいはエージェントの生命感や存在感を高め、ユーザの対話意欲を促進する効果があることが確認された。一方、新たな知見として、公共の場での情報提供タスクでは特有表現がかえって不自然に感じられる場合があり、エージェントにサービス提供者としての知性や信頼性を感じられる人間らしいふるまいを求める傾向があることが分かった。

今後は、様々なタスクや実際の対話による評価の違いを検証し、CG エージェントに最適なインタラクション設計法を検討する予定である。

謝辞

本研究は、JST ムーンショット型研究開発事業、JP-MJMS2011 の支援を受けたものである。

参考文献

- [1] TIFANA. 受付無人化で業務効率・利便性アップ! 北陸名鉄開発株式会社に「ai さくらさん」が受付として導入されます. <https://www.tifana.ai/news/2024111111>, 2024. Accessed: 2025-1-7.
- [2] INC. NTT DOCOMO. my daiz (マイデイズ). <https://www.docomo.ne.jp/service/mydaiz/>, 2024. Accessed: 2024-12-31.
- [3] Isaac Wang and Jaime Ruiz. Examining the use of nonverbal communication in virtual agents. *International Journal of Human-Computer Interaction*, Vol. 37, No. 17, pp. 1648-1673, 2021.
- [4] 北岡教英, 西村良太, 太田健吾. フォトリアル CG エージェントとのマルチモーダル対話. 日本音響学会誌, Vol. 78, No. 5, pp. 257-264, 2022.
- [5] 山田誠二, 角所考, 小松孝徳. 人間とエージェントの相互適応と適応ギャップ (<特集> HAI: ヒューマンエージェントインタラクションの最先端). 人工知能, Vol. 21, No. 6, pp. 648-653, 2006.
- [6] Frank Thomas and Ollie Johnston. *The Illusion of Life: Disney Animation*. Hyperion, New York, 1981. Contains the "12 Basic Principles of Animation".
- [7] 藤子・F・不二雄. ドラえもん, 第 7 巻. 小学館, 1975.
- [8] Lee Akinobu. CG Cybernetic Avatar model "Gene". <https://github.com/mmdagent-ex/gene>, December 2023.
- [9] Lee Akinobu. MMDAgent-EX. <https://github.com/mmdagent-ex/MMDAgent-EX>, December 2023.

A 付録



図 A.1: CG 特有表現を含んだエージェントのふるまいモーション 6 種