

なりきり雑談システムを評価するための キャラクターの評価者間一致に関する検証 - 4択クイズを用いた原作者とファンの比較 -

連慎治¹ 伊藤敏彦²

¹ 北海道大学 大学院情報科学院 ² 北海道大学 大学院情報科学研究院
{shinjimuraji,t-itoh}@ist.hokudai.ac.jp

概要

近年、架空の人物になりきって応答を生成する雑談対話システムの研究が盛んである。多くの研究が「キャラらしさ」の評価を人間による多数決で行っているが、この評価方法がなりきりの正確さを評価できているかは不明である。そこで、本稿では人間同士で「キャラらしさ」のイメージがどのくらい一致しているのかを確認する手法を提案する。具体的には、数名のファンがキャラらしいセリフとキャラらしくないセリフを書き、4択クイズを作成する。別の人間がクイズを解き、一致率などを分析することでキャラクターの違いを調べる。結果、原作者とファンと作品を知らない人が持つイメージに異なる特徴が見られた。

1 はじめに

近年、雑談を行う対話システムの研究が盛んである。中でも、創作上の架空の人物（以降、キャラと呼ぶ）になりきらせることで一貫した個性を持った応答を生成させる研究は、需要の高さと応用の広さから重要である。エンターテインメントの観点からはなじみ深いキャラと話せることが多くのユーザや研究者の注目を集めており [1, 2, 3, 4]、応用の観点からは対話システムに任意の個性を付加できることで実在人物の代役やユーザが好む個性の対話相手との会話が可能になる。これらの目的のためになりきり雑談システムはより目標とするキャラらしいことが望ましく、どの程度目標のキャラになりきれているかを評価する必要がある。

既存研究では、ほとんどが「より多くの人間が目標キャラに対して持っているイメージ」を優先した評価方法を用いている。しかし、作品内のキャラ発話と矛盾しないためには評価者が十分に作品を知

り、覚えておく必要がある。幾つかの研究 [4, 5, 6] では、評価者の質の保証を試みているが、目標キャラの何をどの程度覚えているかは詳しく確認されていない。また、作品を創作した作者（以下、原作者）はファンよりも目標キャラの考え方や目標キャラが何を知っているかについて詳しい可能性があるが、既存研究はファンによる評価を行っている。

そこで、本稿では作中のキャラ発話と数名のファンが共有するイメージを使って4択クイズを作成し、回答者が「どの程度正確に覚えているか」と、「他のファンと何を共有しているか」を測る手法を提案する。さらに、提案手法を用いてあるキャラに対して検証を行うことで、なりきり雑談システムを構築、評価するためのいくつかの洞察を提供する。

2 関連研究

最先端の研究では、キャラらしさの自動評価を目指す研究がいくつか見られる。Wang ら [7] は、Web上のキャラ性格投票サイトを使って不特定多数の性格イメージを正解にしている。また、Tu ら [8] は、目標キャラが持つ知識や振る舞いなどについて、12人が主観評価した結果を全て正解として用いている。しかし、アノテータの質について詳細を述べていない。Lu ら [9] は、LLMを用いた自動評価を試みており、なりきり全体に対して人間による評価を行っていない。

目標キャラらしさの要素に関しても様々な研究がある。口調に注目するもの [9, 10]、性格に注目するもの [7]、目標キャラが持つ知識に注目するもの [3, 9]、要素を調査するもの [6] などがある。

評価者の質を保証しようとする研究については、対象アニメを10話以上みた評価者が80%以上であるもの [5]、映像作品またはゲーム作品を一通り体験したことを条件にするもの [4] などがある。そこ

で、本稿では発売から10年以上経つ作品を用いて、作品を一通り体験することが目標キャラらしさを覚えていることを保証するかどうかを確認する。

3 提案するイメージ一致検証法

我々は、評価者の候補となる人がどの程度正確に覚えているかと他の人間と何を共有しているかについて興味がある。何を共有しているかを調べるためには誰かのイメージを基準に目標キャラらしい何かと目標キャラらしくない何かを用意する必要がある。そこで、我々はキャライメージの一致を確認する手段としてファン数名が4択クイズを作成し、実験参加者が回答することでクイズ作成者と回答者間の一致を確認する。また、回答者としては、原作者、ファン、作品を知らない人の3グループを集めグループごとの回答の傾向を分析することでイメージの違いを調べる。この章では、まず一致しているか調べたいキャラらしさの要素を検討し、次に要素を踏まえた4択クイズのデザインを提案し、最後に具体的な4択クイズの作成方法を述べる。

3.1 検証するキャラらしさの要素

イメージの一致を見るために、確認したい要素を決める。まず、キャラのイメージを大きく変える要素として口調が考えられる。一方で、口調が目標キャラのものであっても目標キャラが言わないであろう内容の発話はなりきりシステムに望ましくない。要素を増やしすぎるとクイズ作成や回答のコストが上がるため、我々はキャラらしさの要素として口調と内容を採用し、クイズの中にキャラらしい口調とキャラらしい内容のイメージを反映させる。

3.2 4択クイズのデザイン

クイズ作成のために、なりきり雑談システムが目指しているキャラらしさとは何かについて考える。ユーザは作品内のキャラの発話からキャラのイメージを作り出すため、作中の発話は十分に目標キャラらしいと言える。ここで、目標キャラの各発話は作品内の各状況と会話文脈で発話されたものである。したがって、作品内の対応する状況と会話文脈を条件として「実際に発話されたもの」が最も目標キャラらしいと言え、これに矛盾しないことを目指していると考えられる。そこで、本稿では「実際の発話」を最もキャラらしいとする。

ここで、作品内の対応する状況と会話文脈につい

1～4のうち、目標キャラの次の発話はどれでしょう？
以下は物語序盤の会話です（作品内のいつ行われた会話か）

他キャラ「こんにちは」
目標キャラ「こんにちは！」
目標キャラは元気にあいさつを返した（地の文）
他キャラ「もう夜だけだな」

1. 「そっちが言い出したんでしょ」（実際の発話）
2. 「なんで意地悪言うの」（ファンイメージ発話）
3. 「嫌な人ですね」（口調NG）
4. 「こんばんはって言った方が良かった？」（内容NG）

図1 4択クイズの例（選択肢の順はランダム）

て考える。物語の序盤と終盤ではキャラの言動や考え方が変わっている可能性がある。そこで、キャラに関する状況の説明として、作品内のいつ行われた会話かを4択クイズに付加する。また、キャラが実際の発話を行うのに必要な会話文脈を4択クイズに付加したい。作品中では動作の描写や主人公の心情描写（以下、地の文）、他キャラの発話等の一連の流れに続いて実際の発話が行われることから、これら全てを会話文脈と捉え、実際の発話から遡って人手で必要最低限な記述を決める。

次に、ファンが持つキャライメージによる目標キャラらしい発話と目標キャラらしくない発話について考える。前節で述べた通り、口調と内容のイメージが一致するか確認したいため、クイズの製作者は目標キャラが作品中で発話した状況と会話文脈で、「目標キャラが代わりに発話しそうなもの」を新しく作成する。作成された発話は、クイズ製作者にとって目標キャラらしい口調と内容であると言える。この発話を以降、ファンイメージ発話と呼ぶ。

加えて、目標キャラらしくない発話についても一致を確認したい。そのため、クイズ制作者は先ほどと同じ状況と会話文脈に対して、「言いそうな内容だが口調が目標キャラらしくない発話」と、「口調は合ってるが言いそうにない内容の発話」を作成する。以降、これをそれぞれ口調NGと内容NGと呼ぶ。これら4つを選択肢としてクイズを作成する。つまり、図1のように作品中の時間軸情報と会話文脈を参加者に与えて、実際の発話、ファンイメージ発話、口調NG、内容NGのどれが一番目標キャラのセリフとして適当かを選択する4択クイズを作成することで、回答者とクイズ製作者の持つ口調と内容のイメージが一致しているかを測る。

3.3 具体的な4択クイズの作成方法

クイズ作成は複数人で行う。まず、ひとりのファンが作品内から目標キャラらしさが反映されている発話を選択する。そして、必要な最低限の会話文脈と、作品内のいつ行われた会話かを書き出す。そして、実際の発話に代わるファンイメージ発話、口調 NG、内容 NG を作成する。

次に、4択クイズに反映されたキャライメージが偏らないように、作成されたクイズを他のクイズ作成者が確認する。具体的には、作成した問題のどの選択肢が「実際の発話、ファンイメージ発話、口調 NG、内容 NG」の想定で作られたかを知らせずに確認者に見せる。確認者は全ての選択肢に対して、口調がおかしいものと内容がおかしいものをマークする。全ての確認者が想定通りのマークをした問題を4択クイズとして使用する。さらに、作成者と確認者の偏りを減らすために全てのクイズ作成者がクイズを作成し、他人が作成した問題の確認を行う。

最後に、問題自体の厳選を行う。選択肢にはクイズ製作者全員が目標キャラらしくないと感じる発話が含まれているが、どの程度目標キャラらしくないかは不明であり、明らかに文脈に沿わないものも含まれている。目標キャラのことを知らなくても会話文脈から口調 NG と内容 NG であることが分かる問題はただの文章読解問題になっている可能性があり、キャライメージの一致を測る際にノイズになる。よって、キャラを知らない数名が問題を解き、誰も NG 二種を選択しない問題は削除する。

4 実験

本稿では、10年以上前に発売されたとあるビジュアルノベルゲームの1キャラについて提案手法で検証を行う。作品の制作会社には、シナリオを書き起こすことや実験を行うこと、結果を公表することの許可を頂いているが、作品名は伏せることを約束しているため詳細は述べない。元の作品に関して人間がキャラらしさをイメージするのに十分な量のシナリオであることを示すために目標キャラが登場しているシーンを全て書き起こし目標キャラに関する統計量を表1に示す。単語の分割には MeCab¹⁾の IPA 辞書を使った。

表1 目標キャラが登場しているシーンの統計量

	総発話数	総文字数	総単語数	単語種類
目標キャラ	3,799	44,731	28,320	1,841
他キャラ	4,258	64,671	41,152	3,086
地の文	3,549	69,753	44,636	4,225

表2 選択肢ごとの割合（グループ平均）

	実際の 発話	ファン イメージ発話	口調 NG	内容 NG
原作者	98.0%	2.0%	0.0%	0.0%
ファン	58.8%	33.6%	3.6%	4.0%
知らない人	31.3%	33.3%	20.0%	15.3%

4.1 実験設定

クイズ制作はファン3名が行った。会話文脈には、他キャラのセリフと目標キャラのセリフと地の文（動作や主人公の心情描写等）が1〜6文含まれた。作中いつの発話かの記述は人手で物語展開を5つに分け、どれに当たるかの記述を使用した。また、5つの時間幅で問題はそれぞれ同じくらいの数になるよう作成した。元発話が被った場合は三人の同意でより良い問題のみを残し、実際の発話とファンイメージ発話がほとんど同じになっている問題も削除した。各々が確認を行った結果、時間軸ごとで偏りができたため足りない期間の問題を追加し調整した。さらに作品を知らない3名がクイズを解き、簡単すぎる問題を削除した。今回、作成した4択クイズは50問であった。詳細は Appendix A に示す。

クイズの回答は原作者1名、ファン5名（20-30代男性）、作品を知らない人3名（20代男性）の計9名が行った。キャラのイメージを比較したいため、回答者は作品で確認することなく記憶だけを頼りに選択するよう指示した。また、クイズに回答する順番による影響を抑えるため、回答者にはどの順番で解いても良く、また戻って解きなおしても良いと指示した。このことから、作品を知らない人は厳密にはキャラらしさを全く知らないわけではなく会話文脈の方に含まれている目標キャラの発話（計62件）を目標キャラらしさの例として用いることができる。なお、クイズ製作者3名と厳選に参加したキャラを知らない3名は9名に含まれていない。クイズ製作者と厳選参加者の回答結果は Appendix B に示す。

4.2 実験結果

4択クイズの結果を表2に示す。まず、原作者は98.0%の割合で実際の発話を選択している。つまり、原作者が持つ目標キャライメージにより実際の発話

1) <https://taku910.github.io/mecab/>

表3 ファンごとの作品体験時期
最後に体験した時期

	最後に体験した時期
ファン A	直前
ファン B	3 か月前
ファン C	2 年前
ファン D	4～5 年前
ファン E	5 年前

表4 選択肢ごとの割合（ファン詳細）

	実際の 発話	ファン イメージ発話	口調 NG	内容 NG
ファン A	74.0%	24.0%	2.0%	0.0%
ファン B	84.0%	16.0%	0.0%	0.0%
ファン C	70.0%	28.0%	0.0%	2.0%
ファン D	22.0%	58.0%	6.0%	14.0%
ファン E	44.0%	42.0%	10.0%	4.0%

とファンイメージ発話を見分けることが可能であった。さらに、他の回答者と比較するために各ファンが最後に作品をやった時期と結果の内訳を表3と表4に示す。事前のアンケートから原作者が最後に作品を体験したのは5年前であったが、回答に参加したどのファンよりも正確に覚えていた。このことから、原作者はファンの持つイメージよりも正確なキャラのイメージを持っており、正確に目標キャラらしさを評価ができる可能性が高い。

次に、ファン同士を比較する。ファンは、平均92.4%の割合でクイズ製作者が目標キャラらしいと想定している選択肢を選んでいる。一方で、実際のセリフのみに限定すると割合は58.8%まで落ちる。つまり、ファン同士ではイメージを共有しているものの実際の発話を見分ける能力は原作者ほどなかった。さらに、ファンごとの結果を分析する。表3と表4から、作品を体験してから時間が経っている人ほど他のファンが目標キャラらしくないと感じる発話を目標キャラらしいと感じるようになっていた。

最後に、原作者とファン群と知らない人群中、実際の発話を選択する回数に対しフィッシャーの正確確率検定（両側）を行ったところ、原作者>ファン群>知らない人群中で有意差が見られた（ $p<.01$ ）。実際の発話かファンイメージ発話のどちらかが選択される回数に対しては、原作者とファン群は有意傾向（ $.05<p<.10$ ）に留まり、知らない人群はどちらも有意差が見られた（ $p<.01$ ）。これらから、原作者は実際の発話を選択できる独自のキャライメージを持っており、ファンは原作者ほど明確なイメージを持っていないが、知らない人群よりもクイズ製作者が目標キャラらしくないと想定した発話を見分けるのに十分なキャライメージを持っていると考えられる。

表5 原作者とクイズ製作者のイメージ一致率

	実際の 発話	ファン イメージ発話	口調 NG	内容 NG
一致率	100.0%	88.0%	88.0%	82.0%

4.3 4 択クイズの各選択肢分析

実験の結果から、原作者は最も目標キャラの発話を正確に覚えており、作品内の実際の発話に即した評価ができることが分かった。そこで、今回作成した4択クイズの全ての選択肢に原作者がコメントを付けることでさらなる分析を行う。具体的には、原作者はそれぞれの選択肢が目標キャラらしければ○、目標キャラらしくなければその理由を書く。書かれた内容を人手で確認し、クイズ製作者の想定と一致するかを確かめる。

表5に結果を示す。口調 NG の88%が、内容 NG の82%がクイズ製作者の想定と一致していた。また、ファンイメージ発話の88%が原作者にとっても十分に目標キャラらしいとされた。これらから、4択クイズの質は十分であったと言える。さらなる詳細な分析はAppendix Cに示す。

5 まとめと今後の展望

なりきり雑談システムを実現させるためには「キャラらしさ」を評価する必要がある。我々は、ファン数名が共有しているキャラのイメージを元に人間同士が持つキャライメージの一致を4択クイズで確認する手法を提案した。提案手法で一人のキャラに分析を行ったところ、原作者はファンより正確に目標キャラの発話を覚えていること、ファン同士ではイメージを共有しているものの、作品を体験してから時間が経つと他のファンが目標キャラらしくないと感じる発話を目標キャラらしいと感じるようになること、作品をあまり知らない人はファンと目標キャラらしさのイメージを十分に共有していないことが分かった。これらから、作品を知っているか、覚えているかどうかで目標キャラのイメージが変わってしまうため、評価者の選択時にどの程度覚えているかを確認する必要があることが示された。

今後の展望としては、なりきり雑談システムを実装し今回提案した手法を用いてより正確に覚えている評価者を選択することで目標キャラの設定と一貫した評価ができるかどうかを確認する予定である。

参考文献

- [1] Cheng Li, Ziang Leng, Chenxi Yan, Junyi Shen, Hao Wang, Weishi MI, Yaying Fei, Xiaoyang Feng, Song Yan, HaoSheng Wang, Linkang Zhan, Yaokai Jia, Pingyu Wu, and Haozhen Sun. Chatharuhi: Reviving anime character in reality via large language model, 2023.
- [2] Seungju Han, Beomsu Kim, Jin Yong Yoo, Seokjun Seo, Sangbum Kim, Enkhbayar Erdenee, and Buru Chang. Meet your favorite character: Open-domain chatbot mimicking fictional characters with only a few utterances. In Marine Carpuat, Marie-Catherine de Marneffe, and Ivan Vladimir Meza Ruiz, editors, **Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies**, pp. 5114–5132, Seattle, United States, July 2022. Association for Computational Linguistics.
- [3] Yunfan Shao, Linyang Li, Junqi Dai, and Xipeng Qiu. Character-LLM: A trainable agent for role-playing. In Houda Bouamor, Juan Pino, and Kalika Bali, editors, **Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing**, pp. 13153–13187, Singapore, December 2023. Association for Computational Linguistics.
- [4] 光田航, 東中竜一郎, 李廷軒, 杉山弘晃, 水上雅博, 中村竜太, 安達敬武, 川端秀寿, 吉田仙, 杵渕哲也. なりきり質問応答と大規模言語モデルに基づくなりきり AI の構築. 自然言語処理, Vol. 30, No. 4, pp. 1172–1205, 2023.
- [5] Chiaki Miyazaki, Saya Kanno, Makoto Yoda, Junya Ono, and Hiromi Wakaki. Fundamental exploration of evaluation metrics for persona characteristics of text utterances. In Haizhou Li, Gina-Anne Levow, Zhou Yu, Chitralakha Gupta, Berrak Sisman, Siqi Cai, David Vandyke, Nina Dethlefs, Yan Wu, and Junyi Jessy Li, editors, **Proceedings of the 22nd Annual Meeting of the Special Interest Group on Discourse and Dialogue**, pp. 178–189, Singapore and Online, July 2021. Association for Computational Linguistics.
- [6] 岩田伸治, 伊原滉也, 佐藤志貴, 馬場淳, 邊土名朝飛, 眞洋, 塩塚勇気, 吉本暁文. ゲームの台詞を題材としたキャラクターらしさに寄与する要素の選定. 人工知能学会研究会資料 言語・音声理解と対話処理研究会, Vol. 102, No. 0, pp. 48–52, 11 2024.
- [7] Xintao Wang, Yunze Xiao, Jen-tse Huang, Siyu Yuan, Rui Xu, Haoran Guo, Quan Tu, Yaying Fei, Ziang Leng, Wei Wang, Jiangjie Chen, Cheng Li, and Yanghua Xiao. In-Character: Evaluating personality fidelity in role-playing agents through psychological interviews. In Lun-Wei Ku, Andre Martins, and Vivek Srikumar, editors, **Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 1840–1873, Bangkok, Thailand, August 2024. Association for Computational Linguistics.
- [8] Quan Tu, Shilong Fan, Zihang Tian, Tianhao Shen, Shuo Shang, Xin Gao, and Rui Yan. CharacterEval: A Chinese benchmark for role-playing conversational agent evaluation. In Lun-Wei Ku, Andre Martins, and Vivek Srikumar, editors, **Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 11836–11850, Bangkok, Thailand, August 2024. Association for Computational Linguistics.
- [9] Keming Lu, Bowen Yu, Chang Zhou, and Jingren Zhou. Large language models are superpositions of all characters: Attaining arbitrary role-play via self-alignment. In Lun-Wei Ku, Andre Martins, and Vivek Srikumar, editors, **Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 7828–7840, Bangkok, Thailand, August 2024. Association for Computational Linguistics.
- [10] 馬ブン, 岸野望叶, 古宮嘉那子, 新納浩幸. T5 によるデータ拡張を利用した特定キャラクタ風の日本語言語モデルの構築. 自然言語処理, Vol. 31, No. 3, pp. 1076–1106, 2024.

表 6 作られたクイズの内訳			
	作成数	確認後	厳選後
製作者 A	123	44	21
製作者 B	65	48	23
製作者 C	99	8	6
三者合計	287	100	50

表 7 厳選時の知らない人ごとの割合				
	実際の 発話	ファン イメージ発話	口調 NG	内容 NG
知らない人 A	39.0%	35.0%	14.0%	12.0%
知らない人 B	39.0%	32.0%	14.0%	15.0%
知らない人 C	43.0%	32.0%	16.0%	9.0%
平均	40.3%	33.0%	14.7%	12.0%

A 4 択クイズ作成の詳細

クイズ作成者はキャラらしいと感じた作中セリフを元にクイズを作成する。そのため、作成者ごとに作成数が異なる。また、作成者ごとにファンイメージ発話の作成能力も異なるため、確認時に没となる割合も異なる。作成者ごとの詳細を表 6 に示す。

厳選時は、作品を知らない 3 名にクイズを解いてもらい、全員が口調 NG と内容 NG を選ばなかった問題を削除した。この 3 名の回答結果を表 7 に示す。今回は 3 名がこの作業を行っているため偶然削除された問題も存在する。この工程の作業人数を増やしたり、選択ではなくそれぞれの選択肢に対する評価を問うことで、より正確にクイズの厳選を行うことができる可能性がある。

B 4 択クイズ作成者の回答結果

クイズ作成者は自分が作成していないクイズの確認のために、口調がおかしいものと内容がおかしいものにマークを付ける。そして、全ての製作者が同じ想定になったものをクイズとして使っている。一方で、確認を行う際にどれが実際の発話かについては知らされておらず、クイズ製作者自身のイメージを元に確認を行っている。よって、クイズ製作者がクイズを解いた際に、全ての問題で実際の発話を選択できるとも限らない。そこで、クイズ製作者がどの程度実際の発話を見分けることができたかを表 8 に示す。なお、全てのクイズ製作者がクイズ制作直前に作品を体験している。最後に体験した時期は同じであるものの結果に差が出ていることは作品を繰り返しした回数 (A>B>C) の影響と考えられる。

表 8 選択肢ごとの割合 (製作者詳細)				
	実際の 発話	ファン イメージ発話	口調 NG	内容 NG
製作者 A	96.0%	4.0%	0.0%	0.0%
製作者 B	90.0%	10.0%	0.0%	0.0%
製作者 C	76.0%	22.0%	2.0%	0.0%

C 原作者コメントの詳細な分析

原作者が各選択肢に書いたコメントの内容を手手で分類した結果を報告する。表 5 ではクイズ製作者の想定との一致を測ったが、これは口調 NG が口調のみ変であることを指摘された場合に一致とカウントしている。一方で、原作者にとって目標キャラらしいと感じる選択肢は全 50 問中の口調 NG が 0 件、内容 NG が 1 件、ファンイメージ発話が 44 件、実際の発話が 50 件だった。よって、ファンが目標キャラらしくないと感じる発話は理由を問わなければ原作者と概ね同じイメージだと言える。また、原作者が目標キャラらしくないと感じたファンイメージ発話 6 件の内、5 件は「発話自体はキャラらしいが文脈に合わない」という指摘であった。このことから、共有されたファンイメージを用いても一部の状況や会話文脈では元作品に即したなりきりが難しいことが分かった。

D 考察と提案手法の限界

本研究では、より作品内発話に沿った「キャラらしさ」を目指す場合に評価者の選択が必要であることを示した。しかし、目標キャラの作中の発話を覚えていて人ことは作中の発話に矛盾しない評価ができることに留まる。なりきり雑談システムでは作品中で明示されていない情報が話題になる可能性があり、作品をよく覚えている人を集めても作品中から分からない情報に関しては評価が一貫しない可能性がある。また、今回の実験の原作者は作品のことを正確に覚えていることが分かったが、原作者のみが評価を行った場合は作中で表現されていない原作者しか知らない情報を元に評価が行われる可能性がある。原作者のイメージが作品をよく覚えているファンのイメージとも異なる可能性があることには注意が必要である。一方で、よく覚えているファンの多数決より原作者のイメージを優先することで一貫性が上がる可能性がある。また、今回は 1 キャラに対してのみ実験を行った。全てのキャラで同じことが言えるかどうかは追加の検証が必要である。