

# 異性アバタを用いる VTuber における 音声加工の有無の印象評価の検討

呉 健朗<sup>1</sup> 越後 宏紀<sup>2</sup> 新井 貴紘<sup>3</sup> 富永 詩音<sup>4</sup>

**概要:** YouTube やニコニコ動画, Twitch などの動画配信サイトで自身の作成したコンテンツを投稿することが容易になってきている。投稿される動画の配信スタイルも多様であり, 我々は特に VTuber として配信することに注目している。VTuber の配信を視聴すると, 配信者の声の出し方の違いに気づく。例えば, 男性が女性のアバタを用いる際には, ボイスチェンジャを利用して容姿だけでなく自身の声も女性に近づけるような配信者もいれば, 自身の地声をそのまま使用する配信者もいる。あるいは, VTuber としてデビューをした直後はボイスチェンジャを用いていたが, 途中から使用をしなくなるケースも存在する。しかし, このように多種多様な声の出し方がある一方で, 我々が調査した限りでは, VTuber として配信を行う際に, どのような声を用いればよいか検証を行っている事例は存在しなかった。この状況を鑑み, VTuber が配信を行う際に使用する声が視聴者に与える影響を調査を行ってきた。先行研究では男性が女性アバタを用いて配信する場合における視聴者からの配信者・コンテンツへの印象の評価を行ってきた。本研究では女性が男性アバタを用いて配信する場合の検証方法とその分析方法について検討を行ったことについて報告する。

KENRO GO<sup>1</sup> HIROKI ECHIGO<sup>2</sup> TAKAHIRO ARAI<sup>3</sup> SHION TOMINAGA<sup>4</sup>

## 1. はじめに

CG 技術の発達により, 配信者は顔出しをせずアバタを利用して配信を行うことが可能となっている。本論文ではこのように, 2D CG または 3D CG で描画されたアバタを利用して配信を行う配信者のことを VTuber (Virtual YouTuber) と総称する。VTuber は CG 技術の発達とともに活動の幅が広がっており, 顔を出して配信している YouTuber と遜色がなくなってきている。VTuber と YouTuber の違いについて数多くの調査も行われており, 我々は, 配信する動画カテゴリと配信スタイルの違いに対する視聴者の印象の変化の検証を行い, ゲーム配信時には配信者の反応が誇張されて伝わる配信スタイル, ニュース配信時には現実味が増すような配信スタイルが好まれる傾向にあること, また, 学習コンテンツ配信時には見た目よ

りも内容を重要視する結果が得られた [1]。

ユーザが適切な配信スタイルを選定する支援を行う研究を進める上で我々は VTuber が配信する際の声に注目している。現在 VTuber として配信する際の声の出し方として, 男性が女性のアバタを用いる際には, ボイスチェンジャを利用して容姿だけでなく自身の声も女性に近づけるような配信者もいれば, 自身の地声をそのまま使用する配信者もいる。あるいは, VTuber としてデビューをした直後はボイスチェンジャを用いていたが, 途中から使用をしなくなるケースも存在する。我々が調査した限りでは, VTuber として配信を行う際に, どのような声を用いればよいか検証を行っている事例は存在しなかった。この状況を鑑み, VTuber が配信を行う際に使用する声が視聴者に与える影響を調査を行ってきた [2]。実験の結果, 男性配信者が女性アバタを用いる場合, ボイスチェンジャを使用して声をアバタに合わせて加工することで, 視聴者からの配信者・コンテンツへの印象が, 地声のまま配信する場合よりも良くなることを確認した。本研究では女性が男性アバタを用いて配信する場合の視聴者からの印象変化と, 性差を明らかにするための検証方法について検討を行ったことについて報告する。

<sup>1</sup> 日本大学大学院総合基礎科学研究科  
Nihon University

<sup>2</sup> ソフトバンク株式会社  
SoftBank Co.

<sup>3</sup> 株式会社 NTT ドコモ  
NTT DOCOMO Inc.

<sup>4</sup> 株式会社ドワンゴ  
DWANGO Co., Ltd.

## 2. 関連研究

コミュニケーション支援や動画配信など、ユーザがアバタを用いるシーンは数多く存在している。本章ではそれらの事例や、特に VTuber として配信にアバタを用いることに関する調査について紹介する。

渡辺ら [3] は対話者の化身であるバーチャルアクターを利用して対話するバーチャルコミュニケーションシステムを提案している。アバタに行わせる動作を頭部に限定した提案システムを用いてコミュニケーションを行う検証の結果、自己と他者のインクラクションが分かる場面が好まれること、対話者の呼吸の引き込み現象が存在することをそれぞれ確認している。下江ら [4] はアバタとピクトグラムを用いたウェブサイトを検討している。アバタとピクトグラムを組み合わせることで、非言語コミュニケーションが円滑に行えることを示している。松本ら [5] は、仮想空間でアバタを利用してコミュニケーションをとるシステムを提案しており、1対1でのコミュニケーションの場合、上半身の実写映像が表示されると、相手の雰囲気が高く伝わり、コミュニケーションしやすいということが示されている。Li ら [6] や黒田ら [7] の研究は、オンラインショッピングでアバタを利用する試みを行っている。顧客のアバタへの印象について、実際の商品の購入数を指標に検証を行った結果、外見と振る舞いの一致率が高い方が推薦効果があるという結果が得られている。

動画配信に自身の姿の代わりにアバタを用いるような、VTuber の数が増えてきている。VTuber の特徴として、YouTuber よりも視覚情報を気にする必要が無いことがあげられる。Mehrabian[8] は、人の判断に影響を与えるのは「外見などの視覚情報」が 55%、「声や話し方などの聴覚情報」が 38%、「話す言葉や内容などの言語情報」が 7%であると報告している。このため、実写で顔出しの配信を行う場合、メイクや髪色、服装、照明の当たり具合、カメラに映る周辺環境などの視覚情報に配慮する必要がある。一方で、VTuber のようにアバタを利用する場合はこれらの視覚情報に配慮する必要が無い。また、アバタは年月が経っても容姿が変化しないため、同じ容姿で長期間配信活動を行うことが可能である。このような特徴があることから、YouTuber やアナウンサーなど実写の顔出しで活動していた方が、VTuber としても活動するという事例も誕生している。

VTuber と YouTuber の違いについての調査も行われてきている。Lu ら [9] によると、VTuber の視聴者が配信をみるモチベーションは、YouTuber と類似しているものの、VTuber の方が YouTuber よりも配信者と視聴者の間に距離感を感じている傾向があることが示されている。この距離感とは、VTuber が 2D CG や 3D CG といった仮想世界の存在であるため、現実世界との距離があるものとする。

横田 [10] は、VTuber は視聴者にとって「自分の都合に応じて視聴できる」、「自分の安心できる世界観の中に浸ることができる」、「共感や自己肯定感を感じることができる」という特徴があることを示している。VTuber と YouTuber の大きな違いとして、VTuber はキャラクタ設定の自由度が高く、配信者に合った世界観を創り出すことができることがあげられている。

## 3. 研究課題

2章で述べた研究のように、オンラインでのコミュニケーションをとる手法としてアバタを利用することや、その効果の検証が行われている。我々はアバタの活用方法として特に、ユーザが動画の配信に利用することに注目している。2D CG または 3D CG で描画されたアバタを利用して配信を行う VTuber は、CG 技術の発達とともに活動の幅が広がっており、顔を出して配信している YouTuber と遜色がなくなってきている。近年では、VTuber として配信する方法についても様々な手法が増えてきている。例えば見た目について、吸血鬼や妖精などの非人間を用いる手法が存在する。あるいは、見た目は女性アバタにも関わらず、音声は男性という見た目と声異なる配信手法が存在する。このように多種多様に考えられる配信手法のうち、視聴者にとって印象が良いものを模索する上で、動画の視聴回数やチャンネル登録者数、アナリティクス機能、Twitter を中心とした SNS のトレンド、メディアによる話題性などの指標から判断することが多い。しかしながら、これらの指標は長期かつ持続的に配信を行うことで得られるデータであり、これから新たに配信者になろうと考えている人は得られない指標である。このため、これから配信者になろうとしているユーザにとっては、自身の考えている配信スタイルが視聴者にとって印象が良いかどうかについて判断することが難しい。我々は自身と異なる性別のアバタを用いて配信する際の声に着目している。現在 VTuber として配信する際の声の出し方として、男性が女性のアバタを用いる際には、ボイスチェンジャを利用して容姿だけでなく自身の声も女性に近づけるような配信者もいれば、自身の地声をそのまま使用する配信者もいる。あるいは、VTuber としてデビューをした直後はボイスチェンジャを用いていたが、途中から使用をしなくなるケースも存在する。我々が調査した限りでは、VTuber として配信を行う際に、どのような声を用いればよいか検証を行っている事例は存在しなかった。この状況を鑑み、男性配信者が女性アバタを用いて配信する際の声の違いが、視聴者からのコンテンツや配信者に対する印象に与える影響について検証を行ってきた [2]。しかし、先行研究は男性配信者が女性アバタを用いる場合の検証に留まっており、女性が男性アバタを用いた場合の視聴者からのコンテンツや配信者に対する印象や、男性配信者が女性アバタを用いて配信する際の声の印象と

の違いについて検証を行っていない。

以上のことから、自身の性別とは異なる性別のアバタを用いて配信を行う際に使用する声の違いが、視聴者に与える影響を性差の観点から明らかにすることを本稿の研究課題とする。

## 4. 実験計画

3章で述べたように、本研究はVTuberが配信を行う際に使用する声が視聴者に与える影響を明らかにすることを目的としている。研究を段階的に進めるうえで、本実験では性別が異なるアバタを使用する際に、地声を用いて配信した場合と、性別に合わせてボイスチェンジャを用いた場合で、配信に対する視聴者の印象が異なるかどうか明らかにすることを実験目的とする。

### 4.1 実験条件・手順

本研究では、自身の性別とは異なる性別のアバタを用いて配信を行う際に使用する声の違いが、視聴者に与える影響を性差の観点から明らかにすることを目的としている。このため、実験では男性が女性アバタを用いる際の声の異なる3条件と、女性が男性アバタを用いる際の声の異なる3条件をそれぞれ用意する。

**M-C1 男性-地声条件：**配信者は男性である。声は地声を用いる。アバタは女性のキャラクタを用いる。

**M-C2 男性-異性変換条件：**配信者は男性である。声はボイスチェンジャを用いて女性の加工された音声を用いる。アバタは女性のキャラクタを用いる。

**M-C3 男性-同性変換条件：**配信者は男性である。声はボイスチェンジャを用いて男性の加工された音声を用いる。アバタは女性のキャラクタを用いる。

**F-C1 女性-地声条件：**配信者は女性である。声は地声を用いる。アバタは男性のキャラクタを用いる。

**F-C2 女性-異性変換条件：**配信者は女性である。声はボイスチェンジャを用いて男性の加工された音声を用いる。アバタは男性のキャラクタを用いる。

**F-C3 女性-同性変換条件：**配信者は女性である。声はボイスチェンジャを用いて女性の加工された音声を用いる。アバタは男性のキャラクタを用いる。

実験参加者には、上述の条件それぞれによる動画を1度ずつ視聴してもらい、配信者・コンテンツへの印象に関するアンケートにそれぞれ回答してもらった。M-C1~C3, F-C1~F-C3条件間で比較を行う際、配信者が用いるアバタの服装の違いなどが実験に与える影響を少なくするためなるべく男性アバタと女性アバタの見た目を近づけるように自作した。配信者が配信する動画のコンテンツについては、先行研究[1]と同様に、ニュースカテゴリの動画として、

我々が過去に学会で発表した内容や最先端技術についてまとめたスライドを発表する内容を収録して用いる。ニュースカテゴリの動画コンテンツは6種類用意し、M-C1~C3, F-C1~F-C3条件でそれぞれ収録する。この際、同一コンテンツを異なる条件で撮り直しする際におこる実験へのノイズをなくすために、M-C1, F-C1の地声条件で収録した動画に対して、Retrieval-based-Voice-Conversion(RVC)[11]を用いて配信者の声を加工することでM-C2, C3, F-C2, F-C3のボイスチェンジャ条件の動画を作成する。動画中のアバタは、多くのVTuber配信者が典型的に行っている配信スタイルに合わせ、主要コンテンツの右下に上半身が見えるように配置する。以上の条件で作成された動画を図1, 図2に示す。



図1 女性アバタを用いた動画

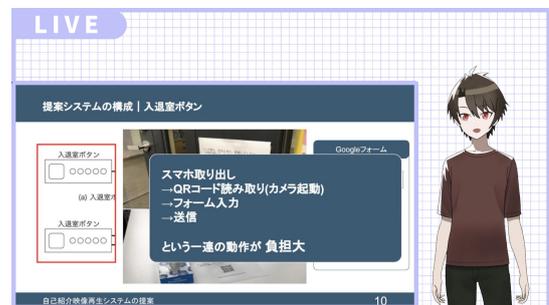


図2 男性アバタを用いた場合の動画

実験手順を下記に示す。

**Step 1:** 実験者がM-C1~C3, F-C1~F-C3条件間からランダムに選択する。

**Step 2:** 実験参加者は選択された条件の動画を視聴する。

**Step 3:** 実験参加者はStep2で視聴した動画についてアンケートに回答する。

**Step 4:** Step1に戻り、選択されなかった条件をランダムに選択した後に再度動画の視聴とアンケートへの回答を繰り返す。

アンケートの内容は下記の通りである。各アンケート、7段階のリッカート尺度(7:とても感じた~1:全く感じなかった)で回答を行わせる。Q1~Q8が配信者に対する印象、Q9~Q11が動画に対する印象を問う内容になって

いる。

- Q1: 配信者に親しみを感じましたか？
- Q2: 配信者にユーモアを感じましたか？
- Q3: 配信者への印象は良いと感じましたか？
- Q4: 配信者の見た目に関心を感じましたか？
- Q5: 配信者の見た目への印象は良いと感じましたか？
- Q6: 配信者の声に関心を感じましたか？
- Q7: 配信者の見た目と声があると感じましたか？
- Q8: 配信者の声への印象は良いと感じましたか？
- Q9: 視聴した動画が興味深いと感じましたか？
- Q10: 視聴した動画が面白い(ユーモア)と感じましたか？
- Q11: 同一配信者の別の動画を見たいと感じましたか？

## 5. おわりに

本稿は、自身と異なる性別のアバタを用いて VTuber として配信する際、自身の生の声で配信した場合と、ボイスチェンジャーにより性別にあった声で配信した場合の、視聴者の配信への印象を比較するための実験計画を述べたものである。今後は検討した内容を元に、実験を行うことで女性が男性アバタを使用した際の声の違いと視聴者のコンテンツ・配信者への影響の検証や、先行研究の男性が女性アバタを用いた際の声の違いが与える影響との比較を行っていく。

## 参考文献

- [1] 越後宏紀, 呉 健朗, 新井貴紘, 富永詩音, 小林 稔: カテゴリ別における VTuber と YouTuber の配信スタイルによる印象評価, 情報処理学会論文誌, Vol. 64, No. 1, pp. 86–95 (2023).
- [2] 新井 貴紘, 富永 詩音, 呉 健朗: VTuber の変声と地声の違いによる印象評価, 情報処理学会シンポジウム論文集, マルチメディア, 分散, 協調とモバイル (DICOMO2023), Vol. 2023, pp. 1643–1648 (2023).
- [3] 富夫渡辺, 雅史大久保: 身体的コミュニケーション解析のためのバーチャルコミュニケーションシステム, 情報処理学会論文誌, Vol. 40, No. 2, pp. 670–676 (1999).
- [4] 下江優太, 濱本和彦, 野須 潔, 小河幸次: アバターとピクトグラムを用いた非言語コミュニケーションの検討, 電気学会論文誌C (電子・情報・システム部門誌), Vol. 130, No. 3, pp. 476–482 (2010).
- [5] 敏宏松本, 宣彦松浦, 昌平菅原: デスクトップ VR 会議におけるアバタ表現方式の評価, 情報処理学会研究報告, Vol. 99, No. 88, pp. 13–18 (1999).
- [6] Li, F.-L., Zhao, Z., Lu, Q., Lin, X., Chen, H., Chen, B., Pu, L., Zhang, J., Sun, F., Liu, X., Xie, L., Huang, Q., Zhang, J. and Chen, H.: AliMe Avatar: Multi-Modal Content Production and Presentation for Live-Streaming E-Commerce, *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, p. 2635–2636 (2021).
- [7] 黒田拓也, 山田誠二, 寺田和憲: オンラインショッピングにおける商品推薦エージェントの外見と振る舞いの関係が購買意欲に与える影響, 人工知能学会論文誌, Vol. 31, No. 2, pp. G-F78 1–11 (2016).
- [8] Mehrabian, A., Hayden, R., Evans, P., Broad, D. and

Mickelson, V.: *Silent messages : implicit communication of emotions and attitudes*, Wadsworth Pub. Co. (1981).

- [9] Lu, Z., Shen, C., Li, J., Shen, H. and Wigdor, D.: More Kawaii than a Real-Person Live Streamer: Understanding How the Otaku Community Engages with and Perceives Virtual YouTubers, *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21 (2021).
- [10] 横田健治: バーチャル YouTuber の提供価値の分析, 電子情報通信学会誌, Vol. 102, No. 7, pp. 654–659 (2019).
- [11] Retrieval-based-Voice-Conversion: <https://huggingface.co/lj1995/VoiceConversionWebUI/tree/main>.