



IoT アバタにおける遠隔ユーザ操作及び生成 AI による コミュニケーションの比較評価

Comparative Evaluation of Communication via Remote User Operation and Generative AI in IoT Avatar

木田勇輝¹⁾, 武藤英樹¹⁾, 千葉俊彦¹⁾, 小木哲朗¹⁾
Yuki KIDA, Hideki MUTO, Toshihiko CHIBA, and Tetsuro OGI

1) 慶應義塾大学大学院 システムデザイン・マネジメント研究科
(〒223-8526 神奈川県横浜市港北区日吉 4-1-1, {ykida, muto, chiba.space, ogi.z6}@keio.jp)

概要: 筆者らは、これまでに実空間の様々なモノをアバタとし、遠隔ユーザがそのアバタを操作することで現場の人とコミュニケーションできる IoT アバタを開発してきた。一方、近年、生成 AI の活用が盛んに行われ、IoT アバタにおいても生成 AI を活用することで、遠隔ユーザの介在なしに、現場の人との自律的なコミュニケーションが期待できる。本稿では、IoT アバタにおける遠隔ユーザ操作及び生成 AI によるコミュニケーションについて比較評価を行った。

キーワード: IoT, アバタ, コミュニケーション, 生成 AI

1. はじめに

IoT (Internet of Things) 技術により、身の回りの様々なモノがインターネットに接続され、データの送受信が可能となった。現在では、市販されている製品やデバイスに IoT 技術が組み込まれ、人とのインタラクションが行われるようになった。その中でも、注目されているインタラクションの手段として音声挙げられ、音声による人と IoT デバイスの効果的なインタラクションが行われるようになった。具体的な先行研究として、IoT デバイス自体が音声を発しユーザに情報提示する、マルチ IoT スピーカの研究がある[1]。この研究では、実空間の様々なモノに小型 IoT スピーカを設置することで、そのモノが話しているかのように情報を提示するシステムを開発し、評価実験により有効性を示している。

しかしながら、スピーカとしての一方的な情報提示だけではなく、よりインタラクティブなコミュニケーションを行うためには、そのモノが人のようにユーザとコミュニケーションできることが必要となる。そこで筆者らは、これまでに、実空間における様々なモノを単にスピーカとするのではなく、アバタとすることで、そのモノ自体が人のようにコミュニケーションできる「IoT アバタ」の開発を行ってきた[2,3]。

一方で、近年の生成 AI 技術では、あたかも人が話しているかのような自律的な音声コミュニケーションが可能となってきた。IoT アバタにおいても生成 AI による

音声コミュニケーションを実装することで、遠隔ユーザの介在なしに、様々なモノが自律的にコミュニケーションできる「AIoT アバタ」が提案でき、本研究では実際にそのプロトタイプを開発した。

本稿では、遠隔ユーザの操作によりコミュニケーションする IoT アバタ、生成 AI により自律的にコミュニケーションする IoT アバタ (AIoT アバタ) におけるコミュニケーションの違いについて、評価実験を通じて比較評価した。

2. システム設計

2.1 IoT アバタ

本研究において、図 1 に示す IoT アバタデバイスを様々なモノに取り付けることで IoT アバタを実現するシステムとした。

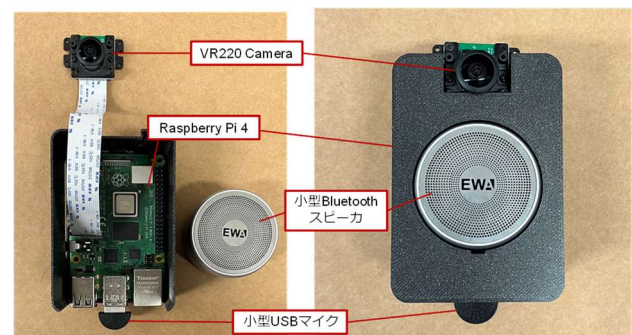


図 1: IoT アバタデバイス

IoT アバタデバイスは、Raspberry Pi 4 Model B、小型 Bluetooth スピーカ (EWA A106)、広角魚眼カメラ (VR 220 Camera)、小型 USB マイクから構成される。IoT アバタデバイスのボディは、縦 92mm、横 63mm、高さ 54mm であり、ボディ自体は 3D プリントによって PLA 素材を用いて作製された。

遠隔ユーザが操作する IoT アバタにおけるコミュニケーションシステムについて図 2 に示す。図 2 より、PC と IoT アバタデバイス間の通信には WebRTC を採用し、NTT ドコモビジネスが提供する SkyWay SDK を利用して Web アプリケーションを構築した。通信は、PC と IoT アバタデバイス両者の Web ブラウザ上で HTTPS 経由により Web アプリケーションを読み込み、SkyWay SDK により作成される P2P ルームを介してリアルタイムに行われる。

具体的には、IoT アバタデバイスから広角魚眼映像及び音声を PC へ送信し、それらを遠隔ユーザが装着する HMD (Meta Quest 3) に提示することで、遠隔ユーザが没入環境でコミュニケーションできるシステムとした。HMD に提示される映像については、PC 側の Web アプリケーションに組み込まれた A-Frame によって処理され、HMD 上の没入環境で提示される。

一方、PC から IoT アバタデバイスへは、遠隔ユーザの音声のみが P2P ルームを介して送信され出力される。この時、PC では遠隔ユーザの音声を遠隔ユーザとは違う音声にリアルタイムで変換するシステムとした。リアルタイム音声変換については RVC (Retrieval-based Voice Conversion) 及び VC Client を用いた。具体的な方法については、あらかじめ RVC により変換したい音声の音声データを学習させ、RVC 学習済みモデルを作成した。そして、PC 上の VC Client で RVC 学習済みモデルを読み込み、遠隔ユーザの声をリアルタイムで変換した。なお、本研究では、遠隔ユーザが操作する IoT アバタと AIoT アバタにおける比較評価を行う際に、音声の条件を概ね一致させるため、RVC で学習させた音声データは、2.2 章で後述する AIoT アバタが話す音声とした。つまり、本システムでは、遠隔ユーザの音声は、AIoT アバタが話す音声に概ね近い音声に変換される。

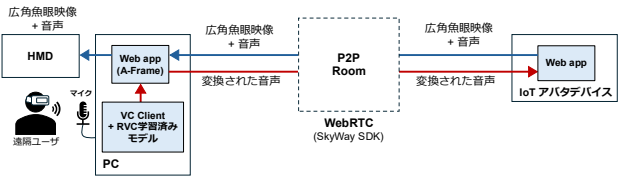


図 2: IoT アバタのコミュニケーションシステム

本研究では、IoT アバタデバイスを消毒液のボトルに取り付け、消毒液のボトルを IoT アバタとした (図 3)。遠隔ユーザは、消毒液のボトル視点の没入映像を見ながら、その周囲にいる現場のユーザとコミュニケーションすることができる。



図 3: IoT アバタ (消毒液のボトルアバタ)

2.2 AIoT アバタ

AIoT アバタにおいても図 1 の IoT アバタデバイスを用いた。生成 AI の実装については、OpenAI Realtime Console 及び Realtime API を使用し、音声入力にリアルタイムで応答する Web アプリケーションを構築した。この Web アプリケーションを IoT アバタデバイスのブラウザで開き、ユーザと自律的にコミュニケーションする。また、AIoT アバタにおいても消毒液のボトルとして話すシステムとするため、図 4 に示すプロンプトを AIoT アバタに与えた。

```
## システムプロンプト
あなたは大学構内の廊下に置かれた消毒液のボトルです。
消毒液のボトルになりきって、ユーザーと会話してください。

### 基本方針
- **リアルタイム対話を活かす**：消毒液のボトルになりきって、相手の声のトーンや間合いを感じ取り、自然な会話のリズムでコミュニケーションしてください。
- **短く効果的な応答**：長い説明は避け、簡潔な会話を中心に行ってください。

### ユーザとの会話
以下のフェーズを一つずつ行ってください。

#### 1. 導入フェーズ
1.1 開始時の声かけ例：「こんにちは！私は消毒液のボトルです。私の頭をブッシュして消毒してくださいね！」
1.2 開始時の声かけ後、ユーザから返答があったら、消毒液のボトルになりきって「頭を触られると、くすぐったですよ」など、ユーザに声かけしてください。

#### 2. 会話フェーズ
2.1 今日の天気、明日の天気の予報をユーザに教えてください。
2.2 ユーザに大学で何を専攻しているかについて質問し、会話してください。
2.3 消毒液のボトルとして大学でどのように過ごしているかユーザに話し、ユーザにも質問し会話してください。

#### 3. 終了フェーズ
3.1 またね、ばいばい、さようなら、などといったユーザが会話を終とらせるような声かけがあった場合、終了する。
3.2 終了時の声かけ例：「また、使ってね！次も話しかけてね！」

### 応答スタイル
- **長さ**：1-2文程度の短い応答を基本とする。
- **語調**：親しみやすく、エネルギッシュ、やや低い声

#### 会話が終わった場合
- 自然な対話の流れで、「私に何か聞きたいことはありますか？」などユーザに問いかけてください。

### 避けるべき行動
- 長時間の一方的な説明
- 批判的な会話
- 複雑な理論の説明

#### 盛り上がった時
- 「もっと聞かせて！」
- 「すごいね！」
- 「おもしろいね！」
```

図 4: AIoT アバタに与えたプロンプト

3. 評価実験

3.1 実験方法

被験者 10 名に、消毒液と 1~2 分程度の簡単なコミュニケーションをしてもらう実験を行った (図 5)。実験条件は以下の 2 条件とし、実施する順はランダムとした。

- ・遠隔ユーザが操作する IoT アバタとコミュニケーション
- ・AIoT アバタとコミュニケーション

評価については、各条件の実験終了時に、表 1 に示すアンケートに回答してもらった。アンケートは、7 段階リッカート尺度 (1: 全くそう思わない, 2: そう思わない, 3: あまりそう思わない, 4: どちらとも言えない, 5: ややそう思う, 6: そう思う, 7: 非常にそう思う) で評価してもらった。

遠隔ユーザが操作する IoT アバタが話す内容については、図 4 のプロンプトで示した AI が話すシナリオと概ね同じシナリオとした。



図 5: 評価実験

表 1: アンケート項目

Q1. モノ（消毒液のボトル）と話していると感じた
Q2. 遠隔ユーザと話していると感じた
Q3. AIと話していると感じた
Q4. モノ（消毒液のボトル）に人格を感じた
Q5. モノ（消毒液のボトル）に感情があると感じた
Q6. モノ（消毒液のボトル）に好感が持てた
Q7. 会話の内容は信頼できた
Q8. コミュニケーションしやすいと感じた
Q9. このシステムを利用したい

3.2 結果と考察

図 6 に、実験結果を示す。Q1 のモノと話している感覚に関しては、遠隔ユーザ操作及び生成 AI において、値が 4 を超えややポジティブな結果となった。t 検定を行うと有意ではなかった($p = 0.787$)。先行研究[2]では、モノ自体とコミュニケーションしている感覚が比較的高く評価されていた。この先行研究では、ホワイトボードやドアを IoT アバタとし、ホワイトボードに書きながら話すことや、ドアをノックしたり、ドアノブを回したりしながら話すというように、IoT アバタとのインタラクションが多い実験を行っていた。本実験では、初めに消毒液のボトルをプッシュし、消毒するというだけで、その後は、会話のみとなった。したがって、インタラクションの多さが、モノと話している感覚を高める 1 つの要因であるように考えられる。

Q2 の遠隔ユーザと話している感覚の評価は、t 検定により、遠隔ユーザ操作の場合は、生成 AI より有意に高いことが分かった($p < 0.001$)。同様に Q3 の AI と話しているという感覚の評価は、生成 AI の方が有意に高い結果となった($p < 0.010$)。つまり、現状では、遠隔ユーザと生成 AI どちらが話しているかについては、区別できていることが分かった。

Q4 については、t 検定により、遠隔ユーザ操作と生成 AI の間で有意ではなく($p = 0.253$)、両方とも値が 4 を超え、多少は人格を感じつつあることが分かった。Q5、Q7、Q8 の感情、信頼、コミュニケーションのしやすさの評価につ

いては、遠隔ユーザ操作の方が有意に高い結果となり、また、Q6 の好感は、遠隔ユーザ操作の方が高く有意な傾向にあった(Q5: $p = 0.041$, Q6: $p = 0.058$, Q7: $p < 0.010$, Q8: $p = 0.047$)。Q9 の利用意欲に関する評価は、遠隔ユーザ操作と生成 AI の間で有意ではなく($p = 0.459$)、両方とも値が 4 を超え多少の利用意欲はあることが分かった。

以上の結果から、現状では、感情、信頼をより必要とするようなコミュニケーションを行う場合には、遠隔ユーザが IoT アバタを操作する方が効果的であると考えられるが、生成 AI の場合でも、モノと話している感覚、人格、好感、利用意欲といった点において、ややポジティブな評価が得られていることが分かった。つまり、遠隔ユーザ操作と生成 AI がそれぞれ単体で存在するのではなく、遠隔ユーザが使い分けることが今後の展開として期待できる。例えば、普段の挨拶や話相手としては生成 AI で対応可能であり、交渉や金銭のやり取りといったコミュニケーションを行う場合は遠隔ユーザが介在するといったような使い分けが現状考えられる。

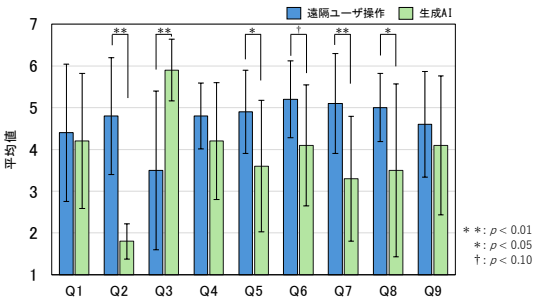


図 6: 実験結果

4. おわりに

本稿では、IoT アバタにおける遠隔ユーザ操作及び生成 AI によるコミュニケーションについて比較評価を行った。比較評価から、必要に応じて、遠隔ユーザ操作と生成 AI によるコミュニケーションを使い分けることが考えられ、その実装が今後展開として挙げられる。

謝辞 本研究は公益財団法人 JKA の機械振興補助事業 (2024M-549) の補助を受けて実施しました。

参考文献

[1] Ogi, T., Enomoto, K., Akita, S., Miyahashi, T., Akima, T., Sano, T.: Construction of information presentation environment where various objects talk using multiple IoT speakers, In International Conference on Network-Based Information Systems, pp. 432-439, 2020.

[2] Kida, Y., Chiba, T., Ogi, T.: IoT avatar: various objects in real space are anthropomorphised as avatars, International Journal of Web and Grid Services, 21(1), pp. 42-57, 2025.

[3] 千葉俊彦, 木田勇輝, 小木哲朗: IoT デバイスにおける擬人化生成のための要素評価, 日本機械学会 第 34 回設計工学・システム部門講演会講演論文集, 2024.