



THE VIRTUAL REALITY SOCIETY OF JAPAN

# ソーシャル VR コンテンツにおけるアバターへの吹き出し表現の付与による発話者同定手法の提案

YANG Guang<sup>1,a)</sup>, 松村 耕平<sup>1,b)</sup>, 大井 翔<sup>2)</sup>, 野間 春生<sup>1,c)</sup>,

1) 立命館大学 情報理工学研究科 (〒 525-8577 滋賀県草津市野路東 1-1-1, a) is0441pf@ed.ritsumei.ac.jp,

b) matsumur@fc.ritsumei.ac.jp, c) hanoma@fc.ritsumei.ac.jp)

2) 大阪工業大学 情報科学部 (〒 573-0196 大阪府枚方市北山 1-79-1, SHO.OOI@outlook.jp)

**概要:** ソーシャル VR コンテンツ環境における問題点として、発話中の参加者を同定や会話内容を理解することが困難である。本研究では、発話中のアバターに吹き出しを付与することでこれらの問題を解決することを目指す。吹き出しを用いることで、発話をしているアバターの区別が、吹き出しの付与が発話者や発話内容の同定にどのような効果を与えるのかについて検証するため実験を実施した。実験では複数のアバターが配置された VR 空間中で吹き出しの有無、形状、及び、吹き出しに表示する内容を変えたシーンを 9 パターン準備し、ある特定の内容について発話しているアバタを空間内から見つけ出すタスクを被験者に課した結果として、ソーシャル VR コンテンツ内において、吹き出しの形状や種類に関わらず、吹き出しを表示することで、特定の内容について発話しているアバタをより早く同定できることが分かった。

**キーワード:** コミュニケーション、提示、ユーザインタフェース

## 1. はじめに

VR 技術が普及し、VR 空間でのコミュニケーションが注目されている。特に、参加者が自身の VR アバタを介して空間に参加し、他のユーザとコミュニケーションをとることができる環境はソーシャル VR コンテンツと呼ばれる。参加者は、この空間内で自身のアバタを自由に移動し、また、近接する他の参加者のアバタと音声によるコミュニケーションをとることができる。

しかし、ソーシャル VR コンテンツ上での音声コミュニケーションには困難を伴う場合がある。例えば、ソーシャル VR コンテンツ空間において、誰かが発話している声が聞こえたとしても、アバタと声の結びつけができず、どの参加者(アバタ)が発話しているのかを特定することは難しい。また、複数の参加者が空間中で同時に発話している場合は、どの参加者(アバタ)がどのテーマについて話しているのか、発話内容とアバタの結びつけにも困難がある。「誰が話しているのか」、「誰がどの内容を話しているのか」が分からないことは、ソーシャル VR コンテンツでの体験を損なう。VR 空間中でのコミュニケーションをより円滑にするためには、これらの問題を解決する必要があると考える。

本研究では、これらの問題に対して発話をしているアバタの頭上に吹き出しを表示することで問題の解決を試みる。吹き出しを付与することによって、「誰が話しているのか」という問題について、吹き出しの有無で判別することができ、「誰がどの内容を話しているのか」という問題について、吹き出しの中に表示される内容を見ることで判別できる。

本論文では、アバタの頭上に吹き出しを表示することの

効果を実験的に検証する。実験空間は 3D ゲームエンジンである Unity を用いて作成する。作成した実験空間中に複数のアバタを配置し、特定の内容について発話しているアバタを探し出すタスクを実験として実施する。このタスクの達成度を調べることによって吹き出しの効果を調査する。実験では丸形、長方形、針形、看板形という 4 つの吹き出しの形状についてその効果の違いについても調査する。

## 2. 関連研究

小宮山らは、VR 空間内におけるアバタとの視線コミュニケーションについて、HMD を装着した状態での VR 空間内のアバタと対話する状況における視線検出の重要性を述べている [1]。その結果、アバタと対話する時に視線の計測により、アイトラッキングを用いて視線検出に空間共有感が感じられるような効果が見られることがわかった。

ソーシャル VR コンテンツでのアバタの違いによる体験に関して、小柳らは、VR アバタをセルフアバタ、人型アバタ、アニメ調アバタに分類した [2]。ユーザが自身を表現するアバタとして一定期間以上使用したアバタをセルフアバタと定義し、他者の見た目を持つリアリスティックなアバタとアニメの基本要素を持つアバタを、それぞれ人型アバタとアニメ調アバタと定義した。

セルフアバタと人型アバタ、セルフアバタとアニメ調アバタの対照実験をそれぞれ行った結果、セルフアバタではその他のアバタよりも身体所有感と臨場感を有意に向上することを明らかにした。

### 3. 提案手法

本研究では、ソーシャル VR コンテンツにおいて、発話中の VR アバタに対して吹き出しを付加することを検討する。具体的には、ソーシャル VR コンテンツにおいて発話中のアバタに対して吹き出しを表示しながら、吹き出しの内容として、発話内容を反映させることを考える。図 1 に VR アバタの吹き出しを付加した例を示す。

発話の有無が吹き出しとして視覚化されることによってソーシャル VR コンテンツのユーザが、発話中のアバタを容易に判別できると考えられる。

また、吹き出し内に発話内容の要約、ないし、発話そのものを表示することで、ユーザはそのアバタがどんな内容の発話を行っているのかを判別することができる。同一の VR 空間内で同時に複数の会話が行われていてもどのアバタがどんな内容の発話をしているのかを知ることができる。

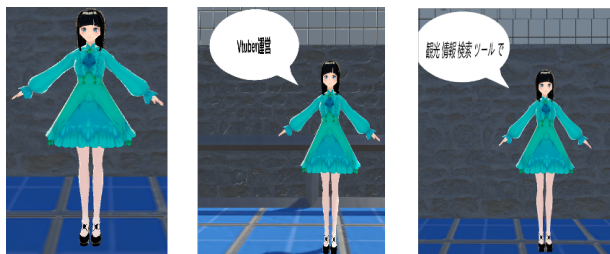


図 1: 吹き出しなし (左)、吹き出しあり (中)、STT (右)

#### 3.1 吹き出しの形状に関する検討

本研究では、ソーシャル VR コンテンツのシーンに対して、発話者の判別のために発話者に対して吹き出し表現を付与することを考える。ここで、吹き出しの形状が発話者の判別に影響を与える可能性がある。

この吹き出しについて、本研究ではマンガの表現手法を踏まえて図 2 に示すように、丸形、長方形、針形、看板形の 4 つの形状のプロトタイプを作成する。また、それらの吹き出しの注目効果について検討する。

#### 3.2 吹き出しの内容に関する検討

吹き出しの形状だけでなく、吹き出し内に表示される内容 (文字列/文章) についても、発話者判別への影響を及ぼす可能性が考えられる。

本研究では吹き出し内に表示する内容について、その抽象度に応じた二つの表示手法を検討する。

一つは、発話内容を要約したテーマを吹き出し内に表示するものである。発話内容を実験者が事前入力で要約したテーマを吹き出し内に表示することで、ユーザは吹き出しのつけられた発話者がどのようなテーマで会話をしているのかを判別することができる。

もう一つは、発話内容そのものを Speech to text (STT) によって被験者が聞こえた音声を認識し、吹き出し内に表示するものである。発話内容そのものを STT を用いて吹き出し内に示す事によって、ユーザは発話者がその時点で発話している内容を判別できることが期待できる。

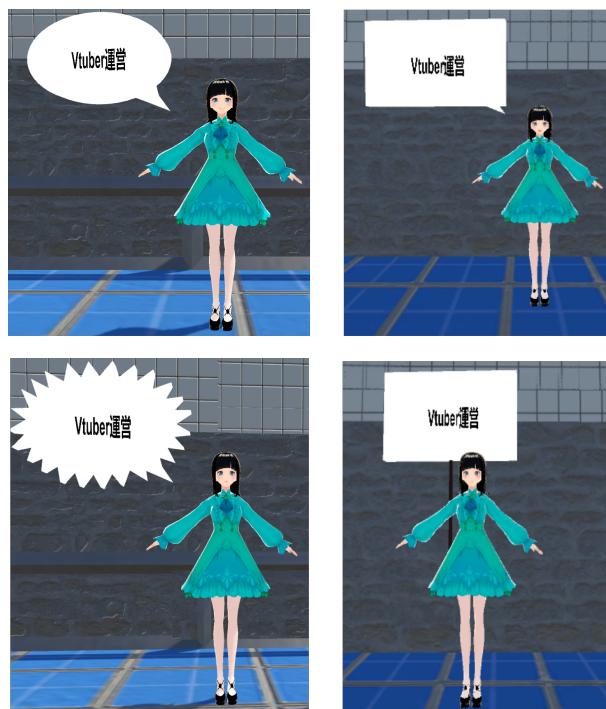


図 2: 吹き出しの種類 (丸形、長方形、針形、看板形)

### 4. 同定効果の検証

#### 4.1 検証実験概要

本研究では、各検討内容の同定効果を検証するため、実験を行う。実験では、ソーシャル VR コンテンツでの吹き出し表現がアバタへの注意喚起の効果に与える影響を検証する。吹き出しの内容及び吹き出しの形状の違いを含めて、実験環境として作成したシーンの組み合わせは図 3 のようである。

実験の条件は図 3 に示すように、吹き出しの種類が「なし、あり、STT」の 3 種類、吹き出しありの場合は更に形状別に 4 種類である。これらの組み合わせの条件で実験を実施する。

実験を行う空間について、ソーシャル VR コンテンツで実際の会場をシミュレーションするため、第 25 回 VR 学会大会におけるオープン・バーチャル・エキシビションの発表会場を模したシーンを作成した。

実験では、参加者に対して、VR 空間中に複数存在するアバタの中から、特定のテーマを話すアバタを既定時間内で見つけ出すタスクを課す。各条件下におけるタスクの正答率を比較することによって、ソーシャル VR コンテンツにおけるアバタへの吹き出し表現の付与による発話者同定の効果を調査する。

空間内アバタの配置について、図 4 で例を示す。

#### 4.2 実験目的

本実験の目的は、複数のアバタが参加し、それらが同時に発話をしているようなソーシャル VR コンテンツにおいて、特定のテーマについて発話するアバタを特定するために、吹き出しの付与がどのような影響を与えるかについて調査することである。ソーシャル VR コンテンツのユーザ

吹き出し	内容 (3種類)	吹き出しなし	吹き出しあり	Speech To Text (STT)	
	形状 (4種類)	丸形	長方形	針形	看板形

図 3: 実験したシーンの条件の組み合わせ

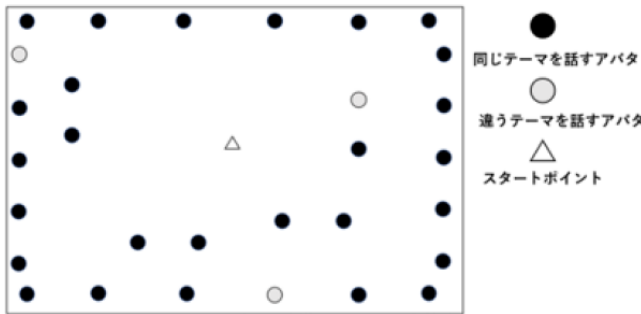


図 4: 実験会場内アバタの配置 (一例)

が学会などの環境で吹き出しを使用することで、多人数でコミュニケーションする場合でも、誰かが何の内容を話していることを把握することができるようになり、交流時間の節約や交流相手の選別機能が期待される。そのため、吹き出しの種類と内容により、提案手法が吹き出しの使用による発話者同定の効果を確認するために実験を実施した。

### 4.3 実験のルール

実験者は事前に 36 個のシーンを用意した。9 個のシーンは吹き出しの有無、吹き出しの内容についての 2 種類、吹き出しの形状について長方形、円形、針形、看板形との 4 種類の組み合わせから構成される。すなわち、吹き出し有りの条件については 4 つの形状と 2 種類の内容の 8 種類があり、また、吹き出しなしについては操作環境とアバタの種類別に 1 個のシーンが存在する。

被験者は、実験者の指示に従って指定したシーンの中から 60 秒以内に指定したテーマ (目標テーマ) を発話する 3 人のアバタを特定するタスクを行う。指定したテーマについて、吹き出しありのシーンでは、目標テーマが「Vtuber」であり、そのテーマについて 3 人のアバタが発話する。その他のアバタは「コロナ対策」をテーマに発話する。一方、STT のシーンでは、目標テーマ「いじめ」について 3 人のアバタが話し、その他のアバタは「学校問題」を設定する。

図 4 に示すように、各シーンには 30 人のアバタが配置され、そのうち 27 人のアバタは同じテーマについて話しており、残り 3 人のアバタがそれぞれ異なるテーマを話している。被験者は、その 3 人のアバタを見つけ出し、シーン中に示される、アバタの番号を実験者に伝える。実験者は被験者から伝えられた番号を記録し、実験終了後にその正誤を確かめる。正解であれば「○」とし、不正解であれば「×」として記録する。また、時間内にアバタを見つけた場

合、「1」とし、そうで無い場合、「0」として記録する。被験者がスペースキーを押したタイミングが実験開始となる。

実験開始と同時に、システムから音声再生される。実験者はストップウォッチを使って実験開始から終了までの時間を計測する。被験者が 3 人のアバタの番号を回答した後、ストップウォッチを押して実験を終了する。実験開始においては 60 秒を経過し、被験者がタスクを完了できない場合も実験を終了する。実験者は、測定した時間を秒単位で各シーンの結果に記録する。すべてのシーンの実験終了後、各シーンについて記録された「○」と「×」の数から各シーンの正答率を計算する。

### 4.4 実験手順

実験に際して、被験者に実験内容を 5~10 分で説明をし、その後同意書に署名を行ってもらおう。被験者が実験する時間がおおよそ一時間で、各シーンのテスト結果に応じて実験後に質問を聞くことがあり、その結果から実験するシーンを実験者から増減することがある。

次に、説明が終わってから実験を行うため、2D 操作環境の「吹き出しなし」のシーンに入り、1 分間被験者がシーンでの操作を体験し、操作方法に慣れてもらう。そして、被験者が吹き出しなし、吹き出しあり、STT の順番で実験を行い、さらにセルフアバタから人型アバタ、2D 操作環境から VR 操作環境の順番で実験を行う。図 5 に実験のシーンの順番に示す。2D 操作環境と VR 操作環境の実験をそれぞれ制限時間を 60 秒と 180 秒として実験を行う。

被験者がすべてのシーンの実験を終えると、ヘッドマウントディスプレイを外す。実験者から実験の終了を伝え、実験を終了とする。

## 5. 結果と考察

### 5.1 吹き出しの形の結果と考察

図 6 の実験結果を示すグラフにより、吹き出しの形については、丸形、長方形、針形、看板形それぞれのシーンの中の正答率が 0.8 から 0.9 の間で、大きな差が見えなかった。吹き出しありと吹き出しなしの 5 つのシーンの中では、

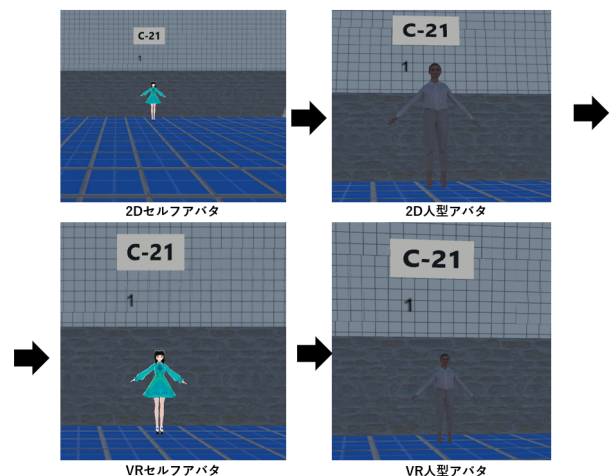


図 5: 実験進行の流れ

2D 正答率、VR 正答率の分散の結果から、2 つとも吹き出しの形が長方形のシーンの正答率の分散が 0.1 以下にあり、すべてのシーンの中で一番低いことがわかった。また、吹き出しの形について、実験の結果から吹き出しが長方形と看板形の正答率が 0.85、0.86 であり、すべてのシーンの正答率の結果の中で比較的に高いことがわかった。

ここから、注意喚起に最も影響を与える条件は、吹き出しありかつ形状が長方形の条件であると考えられる。以上の実験結果から、吹き出しの形が長方形であるシーンではソーシャル VR コンテンツで最も発話するアバタを同定しやすい条件であることがわかった。

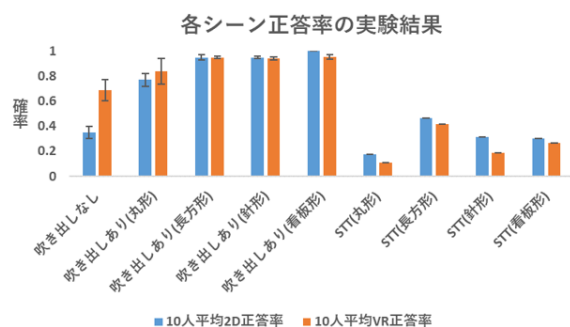


図 6: 各シーン正答率の実験結果

## 5.2 吹き出しの内容の結果と考察

図 6 の実験結果を示すグラフにより、STT の四つのシーンの正答率の結果では、吹き出しありの 4 つのシーンの結果により正答率が低かった。吹き出しなしのシーンでは、2D 環境の達成度の分散が低いですが、2D 環境の正答率と VR 環境の正答率の分散は吹き出しありのシーンとほぼ同じであった。

以上の実験の結果から、ソーシャル VR コンテンツで吹き出しがあるシーンがないシーンより、使用者の注意力を喚起したことがわかった。

吹き出しの内容について、まとめたテーマの方（吹き出しあり）が被験者が聞こえた音声から随時変換されたテキスト（STT）より、注意喚起の効果に与える影響が大きいことがわかった。

## 5.3 結論まとめ

これまでの実験結果に対する考察から、本実験では、ソーシャル VR コンテンツにおけるアバタの吹き出し表現の付与によって、ソーシャル VR コンテンツでの複数のアバタが同時発話する場合、特定したテーマを発話するアバタをより早く同定できることが分かった。

## 6. 終わりに

本研究では、ソーシャル VR コンテンツでの吹き出しの使用を検討し、指定した代表的な吹き出しの内容と種類を用意して、吹き出しの注意喚起効果について判別実験を行った。その中、アバタを作成段階から用意し、実験で使われるソーシャル VR コンテンツをシミュレーションする 3D モデルを作成し、吹き出しの有無と形状からシーンをそれぞれ 9 つ用意して、さらにそれらのシーンを利用してアバタ別で注意喚起の効果を検証した。

結果として、ソーシャル VR コンテンツで使用した VR アバタと操作環境を問わず、長方形の吹き出しを使って、内容が事前に用意した発話内容のまとめである場合 VR アバタへの注意喚起の効果が最も大きいことが明らかになった。それによって、吹き出しの使用からアバタへの注意向上効果により、ソーシャル VR コンテンツでの複数のアバタが同時発話する場合、特定したテーマを発話するアバタをより早く同定できることが分かった。今後とも、本研究の実験内容を用いて、更なる研究成果を上げるために内容を修正したいと考える。

## 参考文献

- [1] 小宮山 撰, 亀川 真吾, 柿沼 育, and 盛川 浩志. VR 空間内におけるアバタとの視線コミュニケーション (ヒューマンインフォメーション人工現実感, エンタテイメント, メディアエクスペリエンスおよび一般). *映像情報メディア学会技術報告 = ITE technical report*, 42(17):21-26, jun 2018.
- [2] 小柳 陽光, 鳴海 拓志, and 大村 廉. ソーシャル VR コンテンツにおける普段使いのアバタによる身体所有感と体験の質の向上. *日本バーチャルリアリティ学会論文誌*, 25(1):50-59, 2020.