# 視聴者反応をフィードバックする Web VR 環境を用いた 遠隔ポスター発表システム

宇都木契[1] 藤原貴之[1] Bhamgara, Hanoz Kaiwan[1] 小林美保[1]

**概要：** 我々は複数人数の視聴を前提としたポスター発表を Web 経由で行うシステムの試作を行なった．本発表ではこのような VR 型の閲覧において，特に並行型発表における閲覧者の行動を示すアバターの役割について論じ，この役割に特化した「聴衆型アバター」としてデザインする事例についての基礎検討を報告する．

# Web VR for Remote Poster Presentation with Interactive Emotion Feedback

KEI UTSUGI[1] TAKAYUKI FUJIWARA[1]
HANOZ KAIWAN BHAMGARA[1] MIHO KOBAYASHI[1]

## 1. 背景

感染症対策で社会活動のあり方が変化し，遠隔コミュニケーションが急速な普及した．その一方で，リモート通信では「直感的な会話ができない」との議論も依然として多い．この問題の一事例として，我々は複数人数の視聴を前提としたポスター発表を Web 経由で行うシステムの試作を行なった．本発表では，このシステムで検討された聴衆側からのフィードバックを主眼に置いたアバター表現手法についての考察を報告する．

## 2. 課題

本論では，不特定多数の人間が参画し，複数の話者の話を聴講にくるポスターセッション形態のコミュニケーションを取り扱う．

学会活動など発表の場では，発表を行う話し手の役割と，質疑に明示的に参画する聞き手の役割の間で「対話」が生まれる．対話では，関係性は必ずしも一方的なものではない．聞き手自身が，言葉で質疑することによって，あるいはジェスチャや表情などの非言語的なフィードバックによって，自身の理解状態を話し手に伝えることでより円滑になる．特に学会などの場では，熟練した聞き手が話を補佐し，話し手を育てる教育効果が主要な目的の一つである．

このような場をリモート環境で再現しようとすると，「聴衆の反応」が直感的にわかりにくいことが，話者にとっても，また聴衆自身にとってもストレスとなりやすい．

現在，リモート形態のコミュニケーションにおいて非言語情報を導入する方法としては，(i)ビデオ会議形式で顔動画を出す方法，(ii)アバターなどを用いる方法があるが，どちらも弱点がある．

前者(i)では，参加人数が増えるとネットワーク負荷が大きくなるため，話者側のみが顔画像を出す場合が多い．また聴衆側が不特定多数に顔を出すことに対してプライバシーの観点でリスクや心理的抵抗感が見られる．

後者(ii)では，意図的に入力した操作しか反映されない傾向があるため，顔表情や視線のような無意識のリアクションから得られる効果が与えられず，また操作に気を取られて肝心のコミュニケーションに集中できない場合も多い．

本論では，このような VR 型の閲覧において，特に並行型発表における閲覧者の行動を示すアバターの役割について論じ，この役割に特化した「聴衆型アバター」としてデザインする方法を検討した．

## 3. 先行研究

他者との協調行動が生むフィードバックについては，多くの議論が行われ，アバターを介しても共感を促進する効果についても近年詳しく検証されている[1]．特に会話行動においては，聞き手の反応が会話の促進に重要な地位を占めていること[2][3]は広く知られており，[4][5]ではそのフィードバックを，(A)聞き手の共感を示し，リズムをたもつ一般的なあいづち行動(generic listener response)と，(B)聞き手が内容に合わせて示す積極的な介入活動(specific listener response)に分類し，これらの活動が欠如すると話の質が下がることを報告している．このうち，(B) SLR のフィードバックはその機能や効果が明瞭であるため，リモート環境でも司会者やファシリテータなどが積極的に補完する傾向がある．その一方で，(A) GLR のような相槌活動は 1 対 1 の対話環境では自然に行われるものの，リモート会話では行いにくい．特に顔の出ない音声のみの会話では発話タイミングを見つけにくく，一方的に話すことが多くなってしまう．このような対話のフィードバックに着目したコミュ

1（株）日立製作所 研究開発グループ

© 2021 Information Processing Society of Japan

ニケーションの検討は，特にロボティクスでの対話再現の観点から検討が行われている[6][7].

また，この相槌行動に類似した問題として，動画視聴に対して視聴者間の一体感を検討する試みも，いくつものバリエーションが提示されている[8][9][10][11]

## 4. 設計コンセプト

本章では，聴衆型アバターのデザインに対する基本議論を記載する.

### 4.1 聴衆の役割とアバター

仮想的な身体であるアバター(化身)という形で仮想空間に参加することの意義は，見る/作業するという主体役割(=I)であると共に，その行為自体が他者から見られる客体役割(=Me)という二面性を含んだ形で，いわば「社会性」を伴う参画行為が行えることにある．発表説明を VR 形態で実施する場において，聞き手が「見られる客体」として参加していることの社会的役割と目的を明確化することで，聴衆型アバターに必要な特徴要件が明確化できる.

見られる客体としての聞き手は，話し手から見られる客体，他の聞き手から見られる客体の二つに分別できる．話し手から見られることの意義としては，先に記したように相槌行動など会話への介入活動などのフィードバックがあげる．また他の聞き手から見られることの意義としては視聴者間の一体感や共感による他参加者の参画を活性化する引き込み効果などがあげられる.

また，必ずしも参加する聴衆者のすべてが見られる客体としての役割を望んでいるとは限らないことにも留意が必要である．話し手とのみ語りたいが，別の聞き手からは見られたくない場合もある．話し手のような自発的な参加者に比べて，その顕現性は抑制的にすることが好ましい.

### 4.2 聞き手型アバターの要件

これらの観点から，我々は，聞き手のアバターに必要な要件として，(1)非意図的な挙動の反映 と(2)匿名性の適切なコントロールの二つの要件に着目した.

**(1)非意図的な挙動の反映**

通常の対話空間では，非言語でのフィードバックが自然に行われるため，聞き手自身は見られるという状態を必ずしも意識していない．また話し手も視覚情報から聞き手の自然な身体挙動をあまり意識することなく判断しており，聞き手の反応から意図を汲み取って話し手が対応する一連のプロセスは，双方の意識を大きく妨げることはない.

だが，多くのアバター型参加システムでは，意識的に入力操作を行わない限りアバター動作へフィードバックすることができない．そのため，話し手は，聞き手が話に集中して聞いているのか，注意がそれているのか，場合によっては聞き手が存在していない状態の区別すらつかない．このため，聞き手が特に意図することなく，話し手へのフィードバックを頻繁に与え続けることで会話の円滑化を促すことが見込まれる．聞き手の理解に集中することを妨げないように，意図的な操作，意図のない操作の二種類を組み合わせた操作モデルが必要となる.

**(2)匿名性の適切なコントロール**

アバター型の VR 参画では個々のユーザが自由に形態をデザインできるが，一方で視覚的な多様性が高くなりすぎて，必要な情報が汲み取りにくくなる場合も散見される．混雑する場においては観察者にとって関係性の薄い聴衆型アバターがある程度一般的な様相をしていることで，視覚情報の整理に便宜を図れる.

また匿名性の提示においては，相手を絞って顕現性を調整する機構があることが好ましい．即ち，発表者や交友関係にあるユーザなど特定の相手に対しては顕現性の高い表現を行いつつ，見ず知らずの第三者に対しては匿名性を保ちたいと考える場合もある．このような条件に対して，観察者との関係性に合わせた表現が行えることが望ましい.

これらから，聴衆型アバターには，ユーザの望む匿名性の条件から以下の三つの段階を用意する.

匿名(PRIVATE)

このタイプのモードでは，ユーザアバターは群衆の一部として振る舞い，個別性を表現しない．役割としては集計用の視聴ログ情報に近く，どの個体が何の行動を行なったかの追跡はできないように表現が限定される．その一方で，「場の一体感」を構築する役割も担うため，入力された情報は匿名性を担保したままで，話し手または別の聞き手に伝達される.

部分隠蔽(PROTECTED)

このモードでは仮想空間内での個別性を表象する．つまり，文字コメントでの意見表示や相槌行動などの反応を行い，話しかけることのできる対象(=Me)としての個体性を持ったアバターとして認識される対象となる．この段階では，話し手へのフィードバックとして働きかけると同時に，別の聞き手に対してもまた共感反応を誘導する役割を担う.

公開(PUBLIC)

このモードでは匿名性は排除され，顔情報や参加所属などを含めて第三者に表明することを前提とした活動がなされる．アバター表現においても，現実世界と同様に顔画像などを含めた情報開示の許可が前提となっている．第三者との事後のコンタクト継続を可能とする場合には，連絡先の提示を含めてこのような状態が必要となる.

これらの要件検討を踏まえ，多人数コミュニケーションを行うためのプラットフォームを検討している．次章ではその試作検討について述べる.

© 2021 Information Processing Society of Japan

## 5. 試作

ポスターセッション型の発表環境に対する試作開発を行った．図１にブラウザベースで表示可能な仮想的なポスターセッション提示環境を構築した事例を示す．

ユーザはドラッグによる見回しと，アイコン操作による構内の移動を行い，発表風景を見回す散策が可能となっている．興味のあるポスターに対して，発表者アイコンの操作により説明映像の再生やコメント書き込みの操作を行う．

各機能はウェブアプリケーションとして実装され，PC ブラウザやタブレットからの閲覧が可能となっている．全周囲カメラで構内の複数個所を撮影した映像を背景として設置し，その風景内にバーチャルポスターと発表者へのアクセスのアイコンが設置されている．

図１：VR ポスター視聴風景

### 5.1 Panorama Viewer による空間構成

本実装では，3D 空間モデルではなく Panorama Viewer 方式による擬似空間構築を用いている．実空間を撮影した全周囲画像の見回し表示(Panorama Viewer)[12]をマップ上の各地点に設置して，擬似的な移動と散策体験を行う手法は，Google Street View [13]以降広く用いられている技法である．写真のクラウド共有サービスや GIS データなどの Web 技術と相性が良く，動作時の計算負荷も低いため，特にブラウザ利用を前提とした JavaScript ベースのライブラリが各種実装されている．本試作では PanoLens[14]をベースとして実装を行った．この Panorama viewer 型の表示の特性として，以下の二点がある．

- 3D モデルの作り込みが不要で，現地の撮影だけで作れるため低コストとなる．
- 移動を許さない分，利用者は行動の自由度が低いが，その分移動操作に気を取られないため UI 負荷が少ない．

今回のような発表閲覧を主眼とした仮装空間構築においては，コスト面のメリットに加え，移動操作がないことで UI に関する心理負荷が少なく，発表内容の視聴に集中しやすいメリットがある．

本施策の実装では，この Panorama Viewer 型の一つの見回し視点当たり最大４種のポスターを表示するようにしている．各ポスターのリンク(説明者アイコン)にアクセスすることで，説明動画再生や詳細発表へのリンク移動を行う．この時，同じ見回し視点をリアルタイムで共有している聴衆ユーザは，お互いの行動のフィードバック情報を直接やり取りする対象となる．

### 5.2 聴衆型アバターの配置

この疑似仮想空間内に聴衆型アバターを配置し，各ポスターを視聴する視聴行動を，ポスター前の人だかりのアバターとして表現する．360 度の見回し画像の視点位置を複数人数で共有する際に，実際の視点位置とアバター表示の位置を一致させてしまうと，全アバターが同じ位置を占めてしまい，自然な光景とならない．そこで，視点位置とは別に，視聴対象に基づいた仮想的なアバター表示位置を別途設定し，第三者からはこの位置にアバターが存在するかのように表示する(図２)．

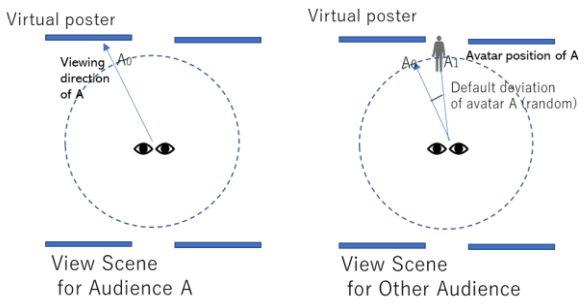


図２：アバターの配置

ユーザ A のアバターの表示位置で，ユーザ A が何を見ているかを伝わるように配置する．観察者 A の視点位置を O，観察者の視線方向（見回し式の場合表示画面の中央）に向けて半径分移動し，アバター表示位置を $A_0$ として設定する．この場合，観察者 B の画面では，アバターA の視線方向にアバターA が描画される．

だが，この条件で複数人数のアバター位置を設定した場合，参加者の視線はポスターの中央に集まる可能性が高いため，聴衆型アバターの配置が密になりすぎる．この状態を避けるため，各参加者に対して表示補正項の情報をあらかじめ割り振り，アバターの描画位置を図２内の $A_1$ 相当の位置に修正する．

### 5.3 ユーザ表情によるフィードバック情報の入力

アバターの操作として，画面ドラッグによる視線方向の意図的操作とともに，視聴者の表情をカメラ撮影した情報を用いる．対話者の感情分類を行った結果をアバターの色や位置として表現，また多数者の反応分布を可視化する．

視聴者の顔画像から特徴点の抽出と感情区分の分類を行う処理には，Web ブラウザのみで動作できる OSS の Clmtrackr [15][16]を用いた．画像からの顔特徴点抽出と，表情からの感情分類の評価をクライアントサイドのみで行

© 2021 Information Processing Society of Japan

えるため，ユーザがプライバシー観点の理由で撮影画像外部に共有したくない場合にも対応が可能である．

抽出された情報は WebSocket 通信を介してフィードバック情報の処理サーバに定期的に送信される．同サーバからは，同じ表示グループに属するユーザに対して定期的に各アバターの情報を配信する．

図３に，視聴者側の UI を示す．視聴者側での撮影状況と判別情報を提示し，この情報の送信可否についての同意を行ったうえでシステムへのフィードバック連結を行う．

図３：顔画像からの情報取得と送信

### 5.4 聴衆型アバターでのユーザ反応の表出

表示される聴衆型アバターには二種類が混在する．

リアルタイムアバター: リアルタイムで接続されているユーザによって動作が制御されているもの．

NPC アバター： リアルタイムユーザ以外のデータに基づいて表示されているもの (NPC．図内の薄い人型表示).

後者の NPC 型の聴衆型アバターには, 過去の視聴ログデータや，VR 以外の形態で同一発表を視聴した人数などを反映し可視化したものである．聴衆型アバターが，各発表の人気投票を表現する側面がある．リアルタイムで参加する利用者以外の「人の賑わい」を調整して表示することで，シーンの雰囲気管理が可能になる．

前者のリアルタイム制御されるアバターについては，や 5.2 の視点位置（画面スクロール位置）に情報と，5.3 のユーザ表情に基づく情報が色情報として反映される(図４).

図４：フィードバック情報の反映．感情分類の結果に応じてアバターの色が動的に変化する．

また，この際，相互の関係性に応じて表示匿名性の調整を行う．動作表現の更新頻度や，参加名の提示有無，感情提示の ON/OFF は，同一グループに属しているかどうかで提示条件の変更を行う．

### 5.5 検証

本ポスター発表システムの初期段階の実装を社内において運用テストを行った．だが 2021 年 5 月現在，感情表現と匿名機能についてはユーザテストが行えておらず，ユーザテストの実施と効果検証の結果については将来課題となっている．直接面識のない視聴者同士の間で，フィードバック反応を共有し，話者・視聴者間の共感を促進する効果を期待している

## 6. 結言

聴衆型アバターを伴うポスター型発表環境の試作から得られた知見について報告した．

## 参考文献

[1] Numata, T., Sato, H., Asa, Y. et al. "Achieving affective human-virtual agent communication by enabling virtual agents to imitate positive expressions". Sci Rep 10, 5977 (2020).
[2] 串田秀也, "聴き手による語りの進行促進 — 継続支持・継続催促・継続試行 —", 認知科学, 16(1), pp.12-23,(2009).
[3] 城綾実,細馬宏通, "多人数会話における自発的ジェスチャの同期",認知科学, 16(1), pp.103-119. (2009).
[4] Bavelas,et.al. Listener as co-narrators. Unpublished manuscript, Department of Psychology, University of Victoria (2000).
[5] 若島孔文,"一般的聞き手反応と特殊聞き手反応の分類",立正大学心理学部研究紀要 (1), 113-123,(2003).
[6] 折重将人,Samratul Fuady,李浩襲,三武裕玄長谷川晶一, "アバターロボットによる存在感通信について",VR 学会大会論文集(2016).
[7] 西村優,佑中村泰,石黒浩,"敵対的生成ネットワークを用いた対話中の人間の振る舞いのモデル化",日本ロボット学会誌 Vol.37.No7.pp632-638,(2019).
[8] 高野祐太郎,田島孝治,大島浩太,寺田松昭"投稿型動画視聴におけるユーザ間リアルタイムコミュニケーション支援システム", 電子情報通信学会論文誌 D,vol.J93-D,No.10,pp2302-2316,(2010).
[9] 吉田有花,宮下芳明:身体動作の重畳表示による動画上での一体感共 有 , 情報処理学会インタラクション 2012 論文集,pp.527-532(2012).
[10] Yuichi Bannai,Yusuke Wakatsuki,"WakWak Tube:An Asynchronous YouTube Co-Viewing System", Proc. IWIN2017, pp.159-164, (2017).
[11] 松長雄也,谷中俊介,坂内祐一, "動画視聴における他視聴者の音声情報の再生方式",Rroc.DICOMO2018,pp530-533,(2018)
[12] E. Chen. “QuickTime VR—An image-based approach to virtual environment navigation,” Proc. of SIGGRAPH, 1995.
[13] Google Street View, https://www.google.com/streetview/
[14] PanoLens.js,https://pchen66.github.io/Panolens/
[15] “clmtrackr | Javascript library for precise tracking of facial features via Constrained Local Models (auduno.com)
[16] Jason M. Saragih, Simon Lucey, and Jeffrey F. Cohn. 2011. "Deformable Model Fitting by Regularized Landmark Mean-Shif"t. Int. J. Comput. Vision ,91(2) , pp200–215,(2011)

© 2021 Information Processing Society of Japan