第 28 回日本バーチャルリアリティ学会大会論文集（2023 年 9 月）

Conversation Echo: 会話の話題を反映した VR 空間のリアルタイム生成

Conversation Echo: Communication in virtual environments that reflects conversation contents

蜂須瞬，脇坂崇平，南澤孝太
Shun HACHISU, Sohei WAKISAKA, and Kouta MINAMIZAWA

慶應義塾大学大学院メディアデザイン研究科（〒 223-8526 横浜市港北区日吉 4-1-1,shun.hachisu@kmd.keio.ac.jp, wakisaka@kmd.keio.ac.jp, kouta@kmd.keio.ac.jp）

概要: 本研究では、会話の話題をリアルタイムで VR 環境に反映させるシステムである Conversation Echo を提案する。本手法では、AI による音声データのテキスト化、会話の話題抽出、パノラマ画像生成を利用して VR 環境を生成し、リアルタイムで環境を動的に変化させる。本手法により、会話の話題のきっかけやインスピレーションを生み出す体験の実現を目指す。

キーワード： コミュニケーション, インタラクションデザイン, 認知

1. はじめに

メタバースの普及に伴い、VR 空間におけるコミュニケーションの機会が増加している。これまでの研究では、コミュニケーション体験を向上させるための VR 技術の新たな応用可能性が模索されている。例えば、創造性を支援するために設計された VR 環境では、参加者により独創的なアイデアを生み出すことが報告されている [1]。具体的には、従来の作業環境とは異なる、海の中や森の中などのより創造的な環境によって、参加者の創造性課題における認知的柔軟性が高まることが示されている。また、オンラインでのコミュニケーションにおいて、会話の文脈に基づいたビジュアルを提示することで、複雑な概念や馴染みのない概念の理解が深まることが示されている [2]。ではもし、会話内容に応じてリアルタイムに VR 環境が変化するような状況に話者が身をおいた場合、その体験は会話そのものにどのような作用をもたらすだろうか。上述の先行研究を鑑みると、会話内容になんらかの肯定的な作用が生じると期待することは自然だろう。ただし、我々の知る限りでは参加者間のリアルタイムな会話に基づいて環境を提示することを試みた研究はこれまでのところまだ存在していない。しかしながら近年の人工知能技術の発展により、会話音声のテキスト変換、柔軟な会話キーワードの抽出、キーワードに応じた画像生成といった処理が、ある程度の精度でリアルタイムに実行することができるようになってきており、これらの技術を組み合わせることで、会話が環境に反映される体験は限定的ながらも現時点で実現可能なのである。

本研究では、そのようなシステム、すなわち会話内容をリアルタイムで VR 環境に反映させるシステムである Conversation Echo を提案する。Conversation Echo により従来のコミュニケーションでは見られないような、新たな話題のきっかけやインスピレーションを生み出す体験の実現が期待できる。

2. 提案手法

提案手法の概要を図 1 に示す。VR 空間上でオンラインコミュニケーションをする際に、体験者の会話から話題を抽出し、リアルタイムで VR 空間を生成し提示する。

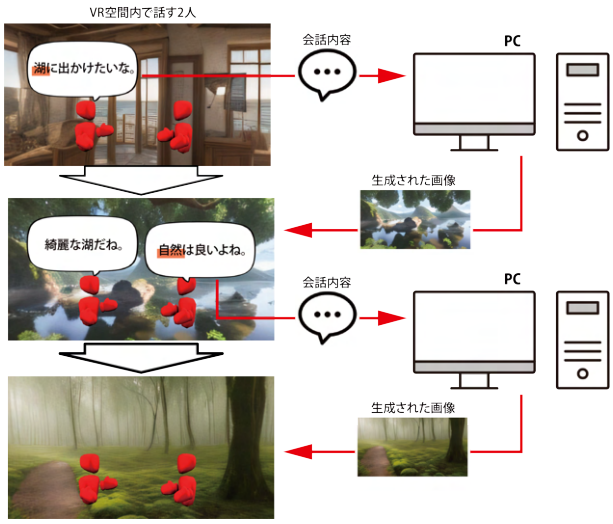

図 1: 提案手法概要図

2.1 システム構成

本研究では、Unity で構築されたマルチユーザ VR オンラインコミュニケーション環境を開発した（図 2）。参加者間での会話音声は PC に記録される。バッファリングされた会話音声は、音声テキスト化ツールである WhisperAPI[3]

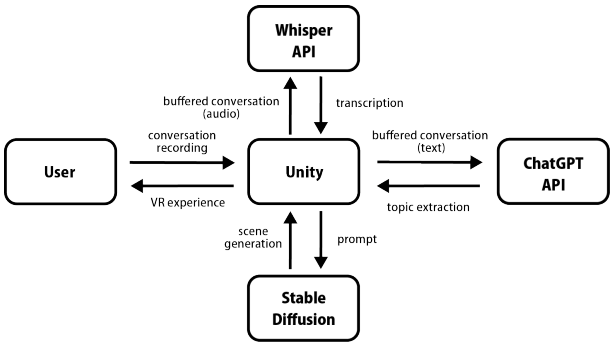

図 2: システム図

を用いてテキストデータとして送信される。テキスト化された会話は、ChatGPT によって処理され、会話の話題に関連する 3 つの単語が抽出される。この 3 つの単語は、パノラマ画像生成用に設計されたプロンプトとともに、テキストベースの画像生成ツールである StableDiffusion[4] に送られる。StableDiffusion はパノラマ画像を生成し、VR 環境に反映させる。これらの一連のサイクルを 8 秒ごとに繰り返し行なっている。

3. 体験例

ヘッドマウントディスプレイ (Meta Quest 2) を装着した参加者 2 人が、本システム上で「行きたい場所」をテーマに 5 分間会話する（図 3)。

図 3: 体験中の様子

3.1 体験の流れ

体験中に交わされた会話の中から一部の会話内容を抜粋する（表 1)。また、抜粋した会話時の VR 環境変化を（図 4）に示す。表 1 のシーン番号と図 4 の各シーン画像の左上の番号が紐付けられている。参加者らの会話の話題である海という話題が VR 環境に反映されたことで、参加者らの話題が海で具体的に何がしたいのかについての話題へと移った。そして、ダイビングをしたいという話題から VR 環境が水中へと変化し、参加者 B は VR 環境に反応を示している。次の VR 環境の変化では音声データのテキスト化に問題が発生し、参加者へ話題とは関係性のない森の VR 環境が提示された。しかし、参加者 A は森の VR 環境からキャンプを連想して話題が変化した。

参加者	会話	シーン番号
A	海に行きたいな。	1
B	海いいよね。	2
A	やっぱりダイビングとかしたいよね。	2
B	おおー、めっちゃ海綺麗じゃん。最高だね。	3
A	森で思い出したけど、キャンプとかも行きたいんだ。	4
B	キャンプいいよね。キャンプファイヤーしたいな。	4

表 1: 会話の内容

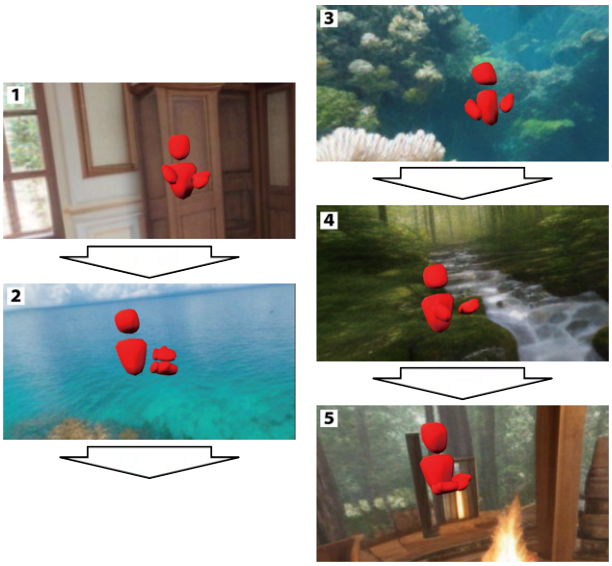

図 4: 参加者視点の環境変化

3.2 体験者からのフィードバック

多くの参加者から会話中に変化する VR 環境から会話の話題の着想を得ることができたという意見が述べられた。また、多くの参加者から会話の話題が途切れた時に変化する VR 環境から会話の話題の着想を得ることができたという意見が述べられた。

現在、本システムでは常に適切に話題を反映した VR 環境を提示することができるわけではない。そのため会話の文脈に合わない VR 環境が提示されることがある。一方で、多くの参加者から会話の話題とは関連性の低い VR 環境が提示されることがあるが、会話に変化を生むきっかけになったという意見が述べられた。

システムに関して多くの参加者が VR 環境が変化する間隔はちょうど良いと述べた。しかし、会話してから話題が VR 環境に反映されるまでにかかる時間が長い時があり、また、どのタイミングで VR 環境に反映されるのかに意識が取られる、という意見もあった。他には、参加者からアバターの髪形や服装などの見た目が話題に合わせて変化したらさらに会話が盛り上がるのではないかという意見も得られた。

4. 今後の展望

現在、本システムでは常に適切に話題を反映した VR 環境を提示することができるわけではないため、本システムのプロセスである、音声データのテキスト化、会話の話題抽出、パノラマ画像生成における精度の向上に取り組む。

体験者からのフィードバックで得られた、会話の話題とは関連性の低い VR 環境が提示されることがあるが、会話に変化を生むきっかけになったという意見から、意図的に抽出した話題に変化を加えて VR 環境を提示することで、会話における話題のためのインスピレーションへと繋げることができるのではないかと考える。今後、抽出した話題に関連するキーワードと会話の話題と異なるあらかじめ用意されたランダムなキーワードを組み合わせて StableDiffusion のプロンプトに入れることで、会話の話題に変化を与える VR 環境を生成する機能の実装に取り組む。

現在、図 2 のシステムのように音声データのテキスト化、会話の話題抽出、パノラマ画像生成、生成画像の VR 空間への反映の一連のサイクルを 8 秒ごとに行っている。そのため、VR 環境が切り替わる際に進行中の会話を中断させてしまうことがある。そのため、画像生成から動画生成にすることで連続的で自然な環境変化の実現を目指す。

現在、VR 環境変化のためのパノラマ画像の自動生成のみを行っているが、VR 空間上でのアバターの髪形や服装などの見た目も話題を反映して変化させる機能を実装する。話題に関連する単語と予め用意した見た目に関連する部位の特定の単語を組み合わせて画像生成することで各部位のテクスチャーを変化させる。

また、Conversation Echo の今回の体験とは異なる目的での応用可能性も検討している。プレゼンテーションや講義などの一方向的なコミュニケーションにおいて、伝達したい内容に合わせてリアルタイムで環境を生成し複数人で共有することにより理解を深める補助的効果が期待できると考える。また、映像配信などのエンターテインメントコンテンツにおいて、配信者と視聴者間でやり取りされるコミュニケーションの内容に合わせて同じ VR 環境をリアルタイムで共有することにより、場の共有による一体感を生み出す新たなコミュニケーション体験を実現できるのではないかと考える。

さらに、本システムは複数人でのコミュニケーションだけでなく、個人での体験も可能である。個人がアイディエーションの際に声に出して思考することにより、発した内容の話題に合わせたビジュアル提示が行われ、環境としての視覚的情報を得ながら思考を行うことができ、より効果的なアイディエーションを可能にするのではないかと考える。

また、Conversation Echo は、会話を通してではあるが思考が環境に反映されるという体験であり、これは SF 等のフィクションの世界では古くから取り扱われてきた、ある意味馴染み深いプロットであるといえる。そういったフィクションにおいては、内的思考と外界の境界が曖昧になるような特異な認知現象の体験として描かれることが多い。思考内容の視覚化については、これまで脳情報デコーディング技術を用いて様々な先行研究の中で試みられてきたが [5]、それらは現時点で fMRI 装置といった大掛かりな装置が必要であり、日常において実現する段階には程遠い。今後 Conversation Echo の改良を重ね、体験の質を向上すれば、そのような特異な体験を日常において実際に実現することも可能だと考えている。そのとき人の思考がどう影響を受けるかは、本稿の範疇を超えるものの、極めて興味深いテーマであり、将来的な研究の展開可能性の一つとして触れておきたい。

5. まとめ

本研究では、会話の話題をリアルタイムで環境に反映する Conversation Echo を提案した。今後は、本システムの改良を重ねて体験のクオリティを向上させていく。また、本体験がもたらす新たな会話へのインスピレーションやアイデア創出については、ユーザーテストを通して検証を行っていく予定である。

謝辞 本研究は JST ムーンショット型研究開発 Cybernetic being プロジェクト (JPMJMS2013) の支援を受けて行われた。

参考文献

[1] Guegan, Jérôme and Nelson, Julien and Lubart, Todd, The Relationship Between Contextual Cues in Virtual Environments and Creative Processes, Cyberpsychology, Behavior, and Social Networking, vol.20, num.3, pp.202–206, 2017.

[2] Xingyu“Bruce”Liu and Vladimir Kirilyuk and Xiuxiu Yuan and Peggy Chi and Xiang‘Anthony’Chen and Alex Olwal and Ruofei Du, Visual Captions: Augmenting Verbal Communication with On-the-fly Visuals, Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI), 2023.

[3] Alec Radford and Jong Wook Kim and Tao Xu and Greg Brockman and Christine McLeavey and Ilya Sutskever, Robust Speech Recognition via Large-Scale Weak Supervision, Proceedings of the 40th International Conference on Machine Learning, vol.202, pp.28492–28518, 2023.

[4] Robin Rombach and Andreas Blattmann and Dominik Lorenz and Patrick Esser and Björn Ommer, High-Resolution Image Synthesis with Latent Diffusion Models, 2022.

[5] Shen G, Horikawa T, Majima K, Kamitani Y, Deep image reconstruction from human brain activity, 2019.