

非透過型 HMD における聴覚障がい者向け 実環境音声 HUD 字幕提示方式の提案

徳田 浩一[†] 岩田 拓也[†] 一宮 誠[†] 小松 慎太郎[†] 小林 彰夫[‡]

大阪府立だいせん聴覚高等支援学校 情報コミュニケーション科[†] 大和大学 情報学部 情報学科[‡]

1. はじめに

1.1 研究の背景

VR (Virtual Reality) 技術、特に、地理的・時間的・物理的な制約を超えて疑似体験を可能にする非透過型ヘッドマウントディスプレイ (以下、HMD) を活用した没入型教材は、学習者の動機づけを高めるツールとして学校教育現場で活用され始めている。しかし、聴覚障がい教育の現場において、非透過型 HMD の利用は実世界情報の完全な断絶、あるいは大きな遮蔽をもたらす。教員が生徒に指示を伝えるために、生徒の肩を叩いて注意を引き、一度 HMD を外させ、手話・口話等で指示を伝えた後、再度 HMD を装着させるという手順を踏まざるを得ない。これは授業のテンポを損なうだけでなく、VR 最大の利点である没入体験を分断し、生徒の学習効率や教員の導入動機を低下させる要因となる。

1.2 研究の目的と提案

本研究の目的は、聴覚障がい者が非透過型 HMD を装着したまま、実世界の音声指示をリアルタイムかつ非断絶的に受容できる環境を構築することである。本稿では、実環境の音声を自動音声認識 (ASR) によりテキスト化し、HMD 内のユーザー視界に追従するヘッドアップディスプレイ (HUD) として提示する方式を提案する。

2. 関連研究

2.1 既存の VR 音声認識機能

2026 年 1 月現在、Meta Quest シリーズに代表される非透過型 HMD では、聴覚障がい者に対するアクセシビリティ機能として、リアルタイム ASR によるライブキャプション機能が実装されている[1]。これにより、VR アプリ内のダイアログや、メタバース空間における他ユーザーの-avatar が発する音声のテキスト表示が一部可能となっている。しかし、これらは**仮想空間内の音声のみを対象**としており、実環境の音声は対象外である。

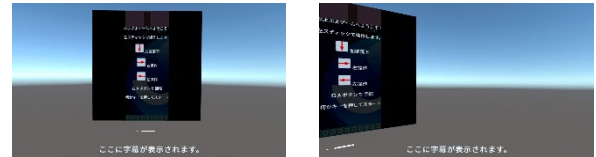


図 1 VR 空間における HUD 字幕提示

仮想空間・実環境の音声認識結果を、ウェブブラウザで提示する試みもある[2]。しかし、ユーザーが視界を動かした際に字幕表示ブラウザを見失う、あるいは本来の教材アプリと同時に起動できない、といった課題がある。

2.2 VR における臨場感

VR における臨場感 (presence) は、Slater が整理した「場所感 (Place Illusion: PI)」と「現実感 (Plausibility Illusion: Psi)」という二つの心理的錯覚によって説明されることが多い。PI は自分がその場所に存在しているという感覚を指し、主として視点移動や身体運動と感覚フィードバックの整合性によって生起する。実環境からの感覚的干渉の低減を図ることは、この錯覚を阻害しないための一要因となり得る。Slater は、没入型仮想環境において利用者が現実世界に近い行動反応を示すためには、これらの錯覚が維持されることが重要であると述べている[3]。

このような理論的枠組みと整合する形で、Meta 社の開発者向けドキュメントや公式ブログでは、仮想空間内の音響回折や反響を考慮した空間オーディオ表現により、仮想環境内の出来事に対する説得力 (plausibility) を高めることが重視されている。

3. 提案方式の設計指針

3.1 HUD 字幕の設計仕様

本方式では、アプリよりも前面に、顔の向きに追従するオーバーレイとして、音声字幕を表示する (図 1)。これにより、非透過型 HMD を利用する聴覚障がい者が VR アプリ内のどの方向を向いていても、実世界の指示に気づくことができる。

3.2 システム構成

本システムは、以下の 3 つのコンポーネントで構成される。

音声集音部 (発話者側): 発話者の付近のマイクから PC またはモバイル端末へ音声を入力する。

A HUD-Based Method for Presenting Real-World Speech Transcriptions for Deaf Users on Non-See-Through HMDs:

[†]Koichi Tokuda, Takuya Iwata, Makoto Ichimiya, and Shintaro Komatsu (Osaka Prefectural Daisen High School for the Deaf)

[‡]Akio Kobayashi (Yamato Univ.)

音声認識部：クラウドベースの AmiVoice API を介し、実時間でテキスト化する。

字幕表示部（聴覚障がい者側）：テキストデータを HMD へ転送し、Unity 等の開発環境で構築したオーバーレイ・キャンバス上に、視界固定の HUD として字幕を描画する。

本方式は現時点では、字幕表示用のオーバーレイを対象 VR アプリに組み込む形（Unity 実装）で実現している。したがって、アプリ改変なしに任意の VR コンテンツへ適用できるシステムレベルの汎用オーバーレイとしては未だ実装していない。

4. モックタスクによる提案方式の評価

4.1 モックタスクに用いたゲーム

高い視覚的注意と同時処理を要するタスクとして、落下型パズルゲーム「ぷよぷよ」を採用し、取得点数で評価を行った。本実験では参考文献[4]により html, JavaScript, CSS を用いてウェブブラウザで動作するゲームアプリを作成し、それを用いた。

4.2 被験者

聴覚障がい成人 9 名。全員学校教育関係者であり、相手や場面に応じて、手話と口話を使い分けてコミュニケーションを取っている。

4.3 熟練者

モックタスクの高得点の鍵となる、ぷよの消去の連鎖を得意とし、音声で操作の指示が可能な人材 1 名。

4.4 手続き

本実験では、条件 1: 自力プレイと、条件 2: 熟練者の音声指示を HUD 字幕として提示し、それに従ってプレイする条件の 2 条件を設定し、同一被験者が両条件を試行した。順序効果を低減するため、被験者 9 名を 2 群に分け、試行順序をカウンタバランスさせた。具体的には、5 名の被験者は条件 1→条件 2 の順で実施し、残り 4 名は条件 2→条件 1 の順で実施した。条件 2 では、熟練者 1 名が被験者の HMD 表示内容をミラーリング画面で確認しながら、音声によりリアルタイムで操作指示を行った。これらの音声はクラウド型自動音声認識（ASR）を介してテキスト化され、被験者の HMD 内に視界固定の HUD 字幕として提示された。音声認識結果に誤認識が含まれる場合でも、テキストの手動修正は行わず、熟練者がミラーリング画面上で誤認識を確認した際には、即座に言い換えや再発話を行うことで補完する運用とした。各条件の試行時間は、被験者の視覚疲労および集中力低下を考慮し、1 試行あたり 180 秒とした。

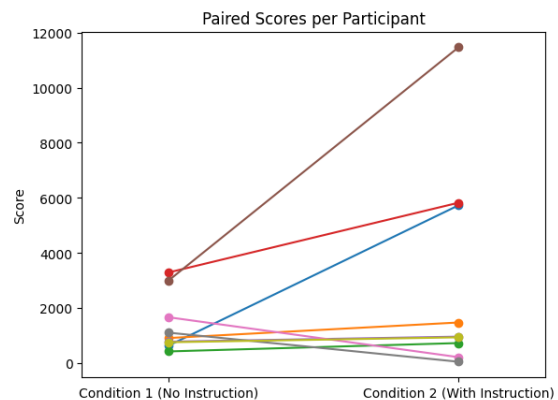


図 2 HUD 字幕による教示有無によるモックタスクの成績

4.5 結果

本モックタスクによる実験の結果を図 2 に示す。条件間の得点差に対して符号付順位検定を行った結果、検定統計量は $T=11$ となり、有意水準 5% の下では帰無仮説は棄却されなかった。一方で、9 名中 7 名の被験者において教示あり条件で得点向上が観察された。

5. 考察

モックタスクの結果から、実環境にいる熟練者の音声認識結果を非透過型 HMD に提示した場合、取得点数が向上する被験者が多数観察された一方で、統計的に有意な差は確認されなかった。これは、音声認識に伴う遅延や誤認識、あるいは指示情報が視覚的負荷として作用し、一部の被験者において集中を阻害した可能性を示唆していると考えられる。

また、手話で意思を表示するろう者は、音声認識系に依らない情報保障が必要である。すなわち、実環境に動画カメラ、非透過型 HMD にピクチャー・イン・ピクチャー機能を実装することにより、実環境の手話を見ることができると考える。

謝辞 本研究の一部は、大阪府学校経営推進費、文科省 DX ハイスクールの支援を受けた。

参考文献

- [1] Meta (2024). Meta Quest のアクセシビリティ機能について説明します。 <https://www.meta.com/ja-jp/help/quest/674999931400954/>
- [2] 大阪府立だいせん聴覚高等支援学校 (2024). いつでも VR 聴覚障がい生徒のための授業の視覚化、学習モチベーション向上～実践報告書。
- [3] Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364 (1535), 3549–3557.
- [4] 紀平拓男. (2025) すぐわかる！ぷよぷよプログラミング SEGA 公式ガイドブック