

視覚と聴覚をシームレスに切り替え可能なMR読書インタフェース

松井 勇人¹ 藤田 和之¹ 高嶋 和毅² 北村 喜文¹

概要：Mixed Reality (MR) グラスの普及に伴い、場所や状況にかかわらずコンテンツ（例えばテキスト情報）とインタラクションする機会が増えている。しかし、コンテンツによる視界の遮蔽や現実空間へのアウェアネスの低下により、安全性の問題を生じる。そこで本研究では、テキスト情報を視覚（テキストの表示）だけでなく聴覚（テキストの読み上げ）で提示可能とし、それらを任意のタイミングでシームレスに切り替えられるMR読書インタフェースを提案する。具体的には、(1) 視線情報に基づいてユーザの読書位置を推定するとともに、(2) モダリティ切り替え時には現在の読書位置を視覚と聴覚の両方で提示することにより、読書位置のコンテキスト保持をサポートする。

1. 序論

近年、スマートフォンの普及やウェアラブル化が進展し、場所や状況を問わずインタラクションが可能となった。さらに最近では、Mixed Reality (MR) や Augmented Reality (AR) グラスの登場により、見たい情報を空中の任意の位置にバーチャルディスプレイとして浮かべることが可能となった。これに伴い、従来ディスプレイに目を向けることが難しかった移動中 [1][2] や作業中 [3] など様々な状況でのバーチャルディスプレイの利用が検討されてきた。

しかし、このようなAR/MRにおけるバーチャルディスプレイとのインタラクションにより、現実世界への注意が疎かになってしまう恐れがある。具体的には、空中に配置したバーチャルディスプレイにより現実世界の情報の一部が遮蔽されたり、ユーザの現実世界への注意が散漫になることで、認知負荷の増加や身の危険（例：歩行中に人やモノと衝突する）などの問題が生じる可能性がある。この問題に対し、先行研究では提示情報のレイアウト [1] や座標系 [4]、および、インタラクションのタイミングの工夫 [5] などにより改善が図られてきた。しかし、これらのいずれの手法においても、提示情報を取得するためには依然として対象を目で見る必要があるため、先ほど述べた問題の本質的な解決には至っていない。特に、長文のテキストを読むようなタスクでは、ユーザは対象を一定時間見続ける必要があるため、より深刻な問題へととなり得る。

この問題へのアプローチとして、本研究では、視覚以外

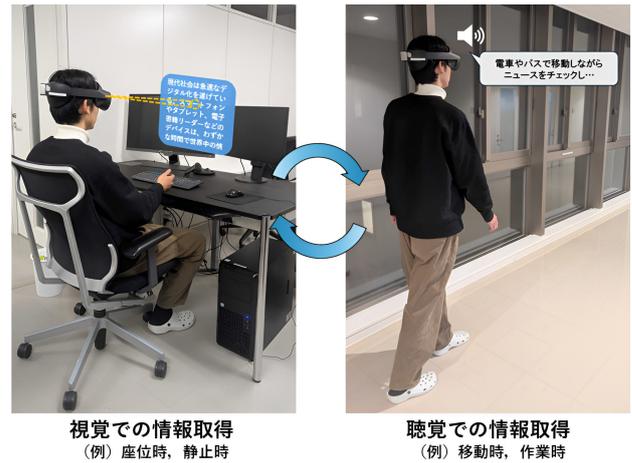


図 1 提案手法のコンセプト図

でテキスト情報を取得するための感覚器官（モダリティ）として聴覚に着目する。聴覚（音声によるテキスト読み上げ）は、視覚（文字の黙読）に比べ情報取得速度の点では劣るものの、歩行中や運転中など外部環境への注意が必要な場面では、認知負荷の低減という面で視覚より有効とされている [6]。このため、視覚と聴覚との間で情報提示モダリティを適応的に切り替えることで、時間とともに変化する状況下においても高パフォーマンスかつ低認知負荷のテキストインタラクションが期待される。視覚と聴覚を切り替えながら情報を取得する手法（切り替え読書）についてはスマートフォンを用いた研究例がある [7] が、ユーザはモダリティの切り替わりの際に読書位置を見失う可能性があり、切り替え読書におけるユーザインタフェースの検討は十分になされていない。

¹ 東北大学電気通信研究所

² 芝浦工業大学

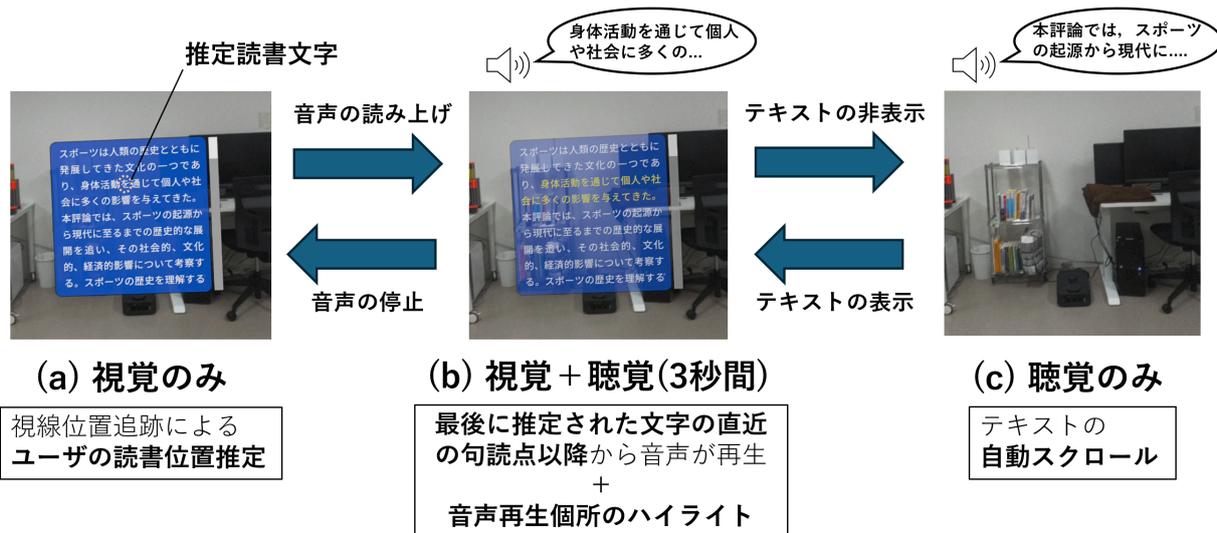


図 2 提案インターフェースのシステム概要

そこで本研究では、AR/MR 環境において視覚と聴覚のモダリティをシームレスに（すなわち、今どこを読んでいるかを見失わないように）切り替える読書ユーザインタフェースの実現を目的とする。このために、(1) 視線情報に基づいてユーザの読書位置を推定するとともに、(2) モダリティ切り替え時には現在の読書位置を視覚と聴覚の両方で提示することにより、読書位置のコンテキストを失わないようにするインタフェース（図 2）を提案する。このインタフェースは、視覚による情報提示が適さないタイミングを含む状況（例：歩行時、運動時、料理時、閉眼時）にも適用可能と考えられる。本稿では、提案インタフェースの設計および実装方法について述べる。

2. 提案手法

本研究では、AR/MR 環境において視覚と聴覚のモダリティをシームレスに切り替えながらの読書を実現するためのインタフェースを提案する。このインタフェースでは、視覚と聴覚の情報取得を切り替える際に、一定時間、表示されたテキスト（視覚）と音声読み上げ（聴覚）との対応付けをサポートするための「仲介インタラクション」を提供する。その際、視線情報に基づきユーザの読書位置を推定すること、その推定された読書位置から音声の読み上げを開始することにより、モダリティの切り替え時にも読書位置のコンテキストを把握したまま読み進められると期待される。以下では、提案インタフェースの実装方法について述べる。

2.1 ユーザの読書位置推定

本システムでは、ユーザが表示されたテキストに対する読書位置を推定するために、先行研究 [8] を参考に、視線情報を用いたアプローチを採用した。本実装では Meta

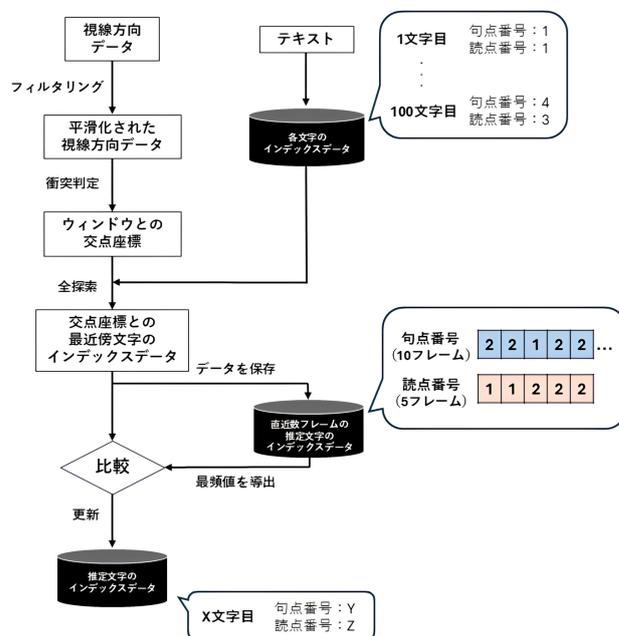


図 3 ユーザの読書位置推定の処理フロー

Quest Pro^{*1}内蔵の視線トラッカーを使用した。

図 3 は、ユーザの読書位置推定の全体的なワークフローである。まず、システムは事前に、提示される文章の各文字が属する句点番号（全体の中でこの文字が何番目の文にあるか）と読点番号（その文の中で、読点によっていくつ目の区切りの場所にあるか）を示すインデックスデータを作成する。次に、システムは各フレームにおいてユーザの視線方向を 3 次元ベクトル形式で得る。この際、生の視線方向データにはノイズが多く含まれるため、先行研究 [9] のサッケード検出と平滑化アルゴリズムを使用して視線方向データをフィルタリングする。このフィルタリング後の視線方向データとテキストが表示されるウィンドウが交差

*1 <https://www.meta.com/jp/quest/quest-pro/>.

する場合、交点座標に最も近い文字を全探索で特定する。そして、視線データのノイズの影響をさらに低減させ、文字の推定をロバストにするために、以下の手順を実施する。(1) 事前に作成されたインデックスデータをもとに、推定された読書文字が属する句点番号と読点番号を特定する。(2) これらの番号を直近数フレームにおける推定読書文字の番号の最頻値と比較する。(3) 両者が一致している場合のみ、推定読書位置を更新する。以上の方法により、視線方向データと各文字の表示位置を正確にマッチングさせ、各フレームでユーザが読んでいる文字を詳細に推定することが可能となる。

2.2 仲介インタラクション

提案インタフェースでは、テキストの表示（視覚）と音声でのテキスト読み上げ（聴覚）を任意のタイミングで切り替え可能であり、切り替わり際には視覚と聴覚の両方で情報することでモダリティ間の対応付けの把握を促す「仲介インタラクション」を数秒間設ける（図 2b）。以下では、一方のモダリティからもう一方のモダリティへと切り替わる際の具体的な動作について説明する（なお、本研究では提示すべきモダリティの決定方法やその切り替えタイミングについては検討対象外とし、本実装では予めシステムが決めたタイミングでモダリティの切り替えを実行する）。また、モダリティ切り替え時間等のインタフェースの挙動に関する細かなパラメータは、パイロットスタディ（N=8）において得られたユーザの好みに基づいて決定している。

2.2.1 視覚から聴覚への切り替わり

視覚から聴覚へと情報提示モダリティが切り替わる際には、テキストが不透明で表示されている状態（図 2a）から、テキストが半透明になり音声再生される状態（図 2b）に移行する。音声は、2.1 節で述べた読書位置推定により最後に推定された文字の直近の句読点以降の文字から読み上げられる。さらに、ユーザが再生中の文章を視認できるように、音声読み上げ位置のテキストは黄色くハイライトされる。視覚・聴覚両方での情報提示が 3 秒間続いた後、テキストが完全に見えなくなり、音声のみが再生される状態（図 2c）に移行する。

2.2.2 聴覚から視覚への切り替わり

聴覚での情報提示中（図 2c）、再生されている文章がウィンドウの最上部に表示されるように、自動的にスクロールされる（ただし、この時点ではテキストおよびウィンドウは非表示状態であるため、ユーザには提示されない）。聴覚から視覚への切り替わりが発生し、視覚・聴覚両方での情報提示（図 2b）に移行すると、再生箇所がハイライトされたテキストがウィンドウの最上部にある状態で半透明表示される。そして、3 秒経過すると音声停止し、テキストが不透明に表示される状態（図 2a）に移行する。

2.2.3 テキストの表示インタフェース

バーチャルウィンドウのサイズや色、テキストの色や表示行数などのパラメータは、先行研究やガイドライン [10][11][12][13] を参考に設計した。バーチャルウィンドウはユーザの前方 2.0m、視線の高さよりも 5° 下方にウィンドウの中心が位置するように配置した。ウィンドウの大きさは 1.0m x 1.0m で、不透明度は 92% である。ウィンドウには約 9 行のテキストが表示され、各行は約 15 文字で構成された。また、テキストの表示量と読書位置推定の精度とのトレードオフを考慮し、テキストの行間は 1.2 に設定した。テキストへのインタラクション方法として Meta Quest Pro のコントローラを用い、コントローラのサムスティックでテキストのスクロールを可能にした。

また、音声の読み上げには Azure AI Speech の Text-to-Speech*² による音声合成 API を使用した。リアルタイムに音声を生成することも可能であるが、生成から再生までにやや時間がかかる（約 1 秒）ことを考慮し、本プロトタイプでは、事前に音声を取得しておき、ユーザの読書位置に対応した音声ファイルを再生するようにした。

3. 結論

本研究では、視覚と聴覚の情報取得をシームレスに切り替えるための MR 読書インタフェースを設計・実装した。今後の展望として、ユーザスタディを通じて提案手法の有効性を評価することや、モダリティの切り替えタイミングを検討し、歩行中などの実利用場面でも使用可能なインタフェースの設計に取り組んでいきたい。

参考文献

- [1] Rufat Rzayev, Paweł W. Woźniak, Tilman Dingler, and Niels Henze. Reading on smart glasses: The effect of text position, presentation type and walking. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, page 1–9, New York, NY, USA, 2018. Association for Computing Machinery.
- [2] Guanghan Zhao, Jason Orlosky, Joseph Gabbard, and Kiyoshi Kiyokawa. Hazardsnap: Gazed-based augmentation delivery for safe information access while cycling. *IEEE Transactions on Visualization and Computer Graphics*, pages 1–10, 2023.
- [3] Chen Chen, Cuong Nguyen, Jane Hoffswell, Jennifer Healey, Trung Bui, and Nadir Weibel. Papertoplace: Transforming instruction documents into spatialized and context-aware mixed reality experiences. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, pages 1–21, 2023.
- [4] Shogo Fukushima, Takeo Hamada, and Ari Hautasaari. Comparing world and screen coordinate systems in optical see-through head-mounted displays for text readability while walking. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 649–658. IEEE, 2020.

*² <https://learn.microsoft.com/en-us/azure/ai-services/speech-service/text-to-speech>

- [5] Feiyu Lu, Shakiba Davari, Lee Lisle, Yuan Li, and Doug A Bowman. Glanceable ar: Evaluating information access methods for head-worn augmented reality. In *2020 IEEE conference on virtual reality and 3D user interfaces (VR)*, pages 930–939. IEEE, 2020.
- [6] Kent Lyons Thad Starner Julie Jacko Kristin Vadas, Nir-mal Patel. Reading on-the-go: a comparison of audio and hand-held displays. In *Proceedings of the 8th conference on Human-computer interaction with mobile devices and services*, MobileHCI '06, pages 219–226, 2006.
- [7] Chen-Hsiang Yu. Mobile continuous reading. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems*, pages 1405–1410, 2012.
- [8] Ru Wang, Zach Potter, Yun Ho, Daniel Killough, Linxiu Zeng, Sanbrita Mondal, and Yuhang Zhao. Gazeprompt: Enhancing low vision people’s reading experience with gaze-aware augmentations. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI '24, New York, NY, USA, 2024. Association for Computing Machinery.
- [9] Rohan Puranik Terry Winograd Andreas Paepcke Manu Kumar, Jeff Klingner. Improving the accuracy of gaze input for interaction. In *Proceedings of the 2008 symposium on Eye tracking research applications*, ETRA '08, pages 65–68, 2008.
- [10] Tilman Dingler, Kai Kunze, and Benjamin Outram. Vr reading uis: Assessing text parameters for reading in vr. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI EA '18, page 1–6, New York, NY, USA, 2018. Association for Computing Machinery.
- [11] Saverio Debernardis, Michele Fiorentino, Michele Gattullo, Giuseppe Monno, and Antonio Emmanuele Uva. Text readability in head-worn displays: Color and style optimization in video versus optical see-through devices. *IEEE Transactions on Visualization and Computer Graphics*, 20(1):125–139, 2014.
- [12] Chunxue Wei, Difeng Yu, and Tilman Dingler. Reading on 3d surfaces in virtual environments. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 721–728, 2020.
- [13] Microsoft. Microsoft hololens—mixed reality technology for business, 2019. <https://www.microsoft.com/en-us/hololens>.