

頭部方向と非言語的発声を用いた VR 向け三次元オブジェクト操作手法

鈴木 皇志* 鈴木 悠真† 崔 明根‡ 志築 文太郎‡

概要. VR における発声による仮想オブジェクト移動操作（発声操作）は、三次元座標系の特定の軸方向へオブジェクトを移動させる。そのため、移動先の位置によっては多くの操作が必要となり、結果として多くの移動時間を要する。我々は、この課題を解決するために、発声操作に加えて頭部方向を用いる手法を開発した。本手法では、頭部の方向に、オブジェクトの移動に使われる座標系が連動して回転するため、ユーザは一度の操作によって目的の方向へオブジェクトを移動できる。ユーザ実験では、手法のより適切な設計を探るために、連続的な非言語的発声による操作および離散的な移動距離の発声による操作と、頭部による操作および視線による操作を組み合わせた計 4 つの設計を実装した。研究室内実験の結果、いずれの設計も発声のみの操作に比べて速いことが分かった。

1 はじめに

Virtual Reality (VR) において、インテリアシミュレーションでの家具、あるいはパズルゲームのコマやブロックなどの仮想オブジェクトを移動させる機会が多い。この操作をハンズフリーで行えれば、人または環境的な要因により手を使用できない場合においても操作できるので、便利である。

これまでに仮想オブジェクトのハンズフリーな操作手法の一つとして、発声による操作（発声操作）手法が検討されてきた。Aslam らの手法 [1] では、ユーザは移動方向（左右前後）と移動距離を発声する。Aslam らは、実験を通じて、この手法が手による手法よりも高速かつ高精度であることを示した。また、仮想オブジェクトではなくユーザの移動手法である Hombeck らの手法 [10] では、ユーザは移動方向（左右前後）を発声した上で停止コマンドを発声するまで移動し続ける。

ただし、既存の発声操作手法には、タスク完了時間または操作性の観点から以下の懸念がある。まず、ターゲット位置によってはタスク完了時間が大きくなる可能性がある。発声操作は特定の座標系の軸方向にのみオブジェクトを直線的に動かす操作であるため、複数回の操作が必要になる場合がある。例として、ワールド座標系を用いる操作では、斜め奥にオブジェクトを移動させるためには上下・左右・前後の 3 回の移動操作が必要となる。したがって、斜め奥に対しても高速にオブジェクトを移動させる手法が求められる。次に、オブジェクトの細かい位置調整が難しい可能性がある。移動距離を発声する手法は離散的な移動操作であるため、連続的な移動操作よりも細かい位置調整が困難であると考えられる。

ゆえに我々は、より高速にオブジェクト移動を達成するハンズフリーな発声操作手法を作成することを目的として、頭部方向を用いた非言語的発声による連続的なオブジェクト移動手法を提案する。本手法ではまずオブジェクトを発声によって選択する。その後、ユーザは頭部を動かすことにより、オブジェクトを選択時の頭部方向に垂直な平面に沿って移動できる。加えて、移動方向（上下・左右・前後）を発声した後に「あ〜」または「う〜」等の非言語的な発声 [2, 9, 12, 18] を行うことにより、発声している間オブジェクトを移動方向へ連続的に移動できる。本手法を用いることによって、移動方向ごとの移動操作を既存手法よりも少なくしつつ、かつ連続的な移動操作によって細かい位置調整も可能になると考えた。

本論文の貢献を以下に示す。

- 頭部方向および非言語的発声を用いたオブジェクト操作手法を設計し、さらに実装方法を示した。
- 研究室内実験にて、本手法のオブジェクト操作パフォーマンスを調査した。本手法が発声のみの手法よりも高速にオブジェクトを操作できること、および非言語的発声操作が移動距離の発声操作よりも高速な移動操作を可能とすることを確認した。

2 関連研究

本手法は、VR における発声および頭部操作を用いた仮想オブジェクト操作手法である。本節では、発声操作および頭部操作、視線操作に関する研究を述べる。

2.1 発声操作

大規模言語モデルの発展に伴い「ここに置いて」のように代名詞を用いた発声操作が発展している [4,

Copyright is held by the author(s).

* 筑波大学 情報メディア創成学類

† 筑波大学 情報理工学位プログラム

‡ 筑波大学 システム情報系

11, 15, 16, 24]. 代名詞を用いて、音声アシスタントとコミュニケーションを行うことにより、意図の伝えやすさおよび表現力の向上に貢献することが知られているが、代名詞の対象がない場合および雑多な環境において代名詞の対象を特定することが困難な場合には、オブジェクト移動を行うことは難しい。

代名詞を用いない発声操作を用いたオブジェクト移動には、「あ〜」または「う〜」等の非言語的発声を用いた連続的な移動 [2, 9, 12, 18] がある。これを用いて、ユーザは、オブジェクトと目的位置間の距離を知らなくても、操作できる。また、他の代名詞を用いない移動として、10 mm のように移動距離を指定した離散的な移動（移動距離発声操作）[1, 18] がある。この手法において、オブジェクトと目的位置間の距離が不明な場合、ユーザは適切な移動をさせることが難しいが、移動距離を知っていた場合は正確な位置への操作が可能である。我々の知る限りでは、仮想空間における非言語的発声操作および移動距離発声操作の比較をしている研究が存在しないため、我々の手法では、非言語的発声と移動距離発声操作を比較することにより適切な発声移動設計を探索する。

2.2 頭部および視線による操作

頭部または視線のみを用いた入力是一定時間オブジェクトを注視することによる単一な入力しか実行できないため、オブジェクトの移動操作のような複雑な操作には適していない。複数回の注視動作により座標系の軸ごとに移動させる手法は提案されている [17] が、この手法も斜め奥への移動が困難である。そのため、頭部または視線を用いたオブジェクト操作手法の多くは手を用いることを前提としており [22]、ハンズフリーな頭部・視線オブジェクト操作は困難である。

いくつかの既存研究 [7, 19, 21] では、頭の動きのみによる操作（頭部操作）と視線による操作（視線操作）の比較において、頭部操作は視線操作よりも速いとされている。一方で、Blattgersteら [3] は視線操作の方が速いとしている。また、いくつかの既存研究 [13, 14, 20] では、頭部操作の方が精度の点で優れていると述べた。発声操作において頭部操作と視線操作のどちらが適しているかは不明であるため、本研究では頭部および視線を用いた発声手法を設計し、さらにそれらを比較することにより適切な本手法のデザインを探索する。

3 本手法の設計

ユーザは移動対象のオブジェクトを発声コマンドを用いて選択した後、手法毎に異なる移動方法を用いてオブジェクトの移動を行う。なお、いずれの手法においても、オブジェクトの移動は、オブジェクトが選択された際に定まる以下で定義する頭部水平

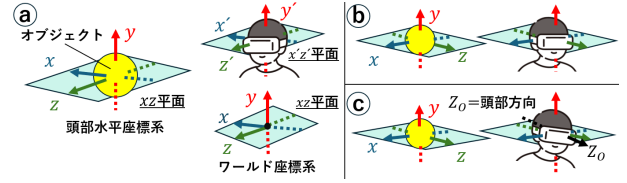


図 1. 頭部水平座標系. a: 初期状態. b: a の状態からユーザが左に向くと、 z 軸は左に向く. c: a の状態からユーザが左下を向いても、 z 軸は左に向く.

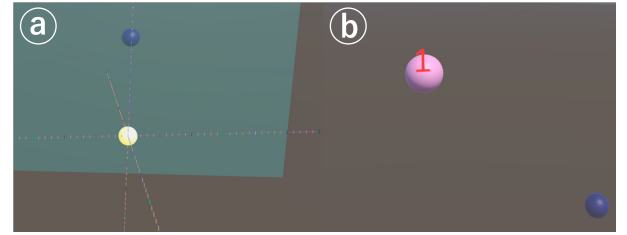


図 2. アプリケーションの画面. a: 頭部水平座標系および選択状態にあるオブジェクト（黄色の球）. b: 移動対象（桃色の球）および移動先（青色の球）.

座標系に基づく。

本研究では、発声に基づくオブジェクト移動のための座標系として頭部水平座標系（図 1, 図 2a）を設計した。この座標系の原点は選択されたオブジェクトの中心であり、さらに xz 平面はワールド座標系の xz 平面と常に平行、 z 軸は頭部座標系におけるユーザの頭部方向から y 成分をなくした xz 平面と平行なベクトルと常に平行、 y 軸はワールド座標系の y 軸と常に平行、 x 軸は z 軸と垂直を維持するように動く。またオブジェクトの移動方向はこの座標系に基づいて行われる。具体的には、「手前」と「奥」は z 軸の正方向と負方向、「上」と「下」は y 軸の正方向と負方向、「右」と「左」は x 軸の正方向と負方向に対応する。

この設計により、図 2a のように、ユーザはオブジェクト付近を俯瞰しながらオブジェクトの移動方向および現在位置から目的位置までの距離が把握しやすくなる。オブジェクトの高さと頭部の高さが異なるほど、頭部座標系と頭部水平座標系の違いは、大きくなると考えられる。さらに、この座標系の z 軸は頭部の xz ベクトル方向に同期しているため、視界の遮蔽を避けつつも xz 平面における斜め方向への移動も行いやすくなっている。

3.1 発声コマンド

本手法の音声コマンドは、Aslam および Zhao の研究 [1] と同じく、意図および特性（任意）から構成される。意図はコマンドの意味（例：「選択」）、特性は操作対象のラベルまたは移動量（例：「1」）である。意図には、オブジェクトを選択するための

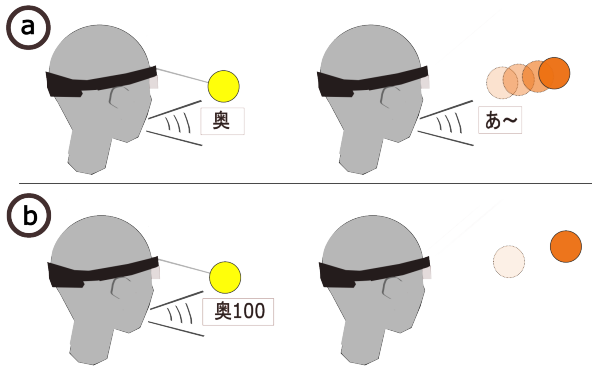


図 3. 各発声操作のイメージ. a: 非言語的発声操作のイメージ. ユーザは移動方向を発声した後に非言語的な発声を行うことにより, 発声を行っている間, オブジェクトをその方向へ連続的に移動できる. b: 移動距離発声操作のイメージ. ユーザは移動方向と移動距離を発声することにより, オブジェクトをその方向へその距離だけ瞬時に移動できる.

「選択」, 頭部操作を開始し移動平面を更新するための「視線」, オブジェクトの移動を停止するための「配置」, オブジェクトの移動方向を定義するための「奥」, 「手前」, 「上」, 「下」, 「右」, 「左」, オブジェクトを 0.01 m 移動させるための「少し奥」, …, 「少し左」がある.

3.2 オブジェクトの選択

ユーザは選択したいオブジェクトを視界内に入れる. 視界内のオブジェクトにはラベル(数字)が表示される(図 2b). ユーザは「選択 2」のように「選択」に続けて数字を発声すると, 対応するオブジェクトが選択される. また, ユーザは単に「選択」と発声することにより最も小さな数字のオブジェクトを選択できる. オブジェクトが選択されると, 視覚的フィードバックとしてオブジェクトの色が黄色に変わる.

3.3 オブジェクトの移動

本節では, オブジェクトの移動に関する各操作の説明を行う.

3.3.1 発声に基づく移動方法

発声に基づく移動方法の検討のために, 非言語的発声操作および移動距離発声操作という 2 種類の移動方法を設計した.

非言語的発声操作 ユーザは, まずオブジェクトの移動方向を定義するために, 「奥」, 「手前」, 「上」, 「下」, 「右」, 「左」のいずれかを発声し, その後に続けて「あ〜」のような非言語的発声を行う. この非言語的発声を続けている間, オブジェクトが連続

的に移動する(図 3a). この際, 発声の音量がオブジェクトの移動速度に対応付けられる(例: 音量が大きいと速い). この操作は, ユーザがオブジェクトの現在位置および目的位置を観察しながら移動速度を調節できるという特徴を持つ.

移動距離発声操作 ユーザは, 「手前 100」のように移動方向と移動距離(単位: mm)を続けて発声することによりオブジェクトを移動させる(図 3b). この操作は頭部水平座標系に基づいた正確かつ離散的な移動を可能とする.

3.3.2 視線および頭部動作に基づく移動操作

視線に基づく移動操作(視線操作)[17, 22]および頭部動作に基づく移動操作(頭部操作)を上述の移動操作と組み合わせることにより, 優れた操作を探る.

視線操作では, ユーザがオブジェクトを選択すると, オブジェクトの現在位置を含み, 頭部に正対し, かつ鉛直な平面(移動面)が生成される. ユーザが視線の向きを変えると, オブジェクトはこの移動面上の視線に正対する位置に移動する.

頭部操作では, ユーザがオブジェクトを選択すると, 視線操作と同じ移動面が生成される. ユーザが頭部の向きを変えると, オブジェクトは移動面上の頭部に正対する位置に移動する.

両方の操作において, 頭部水平座標系は頭部の動きにより回転する.

4 本手法の実装

今回の実験に用いたシステムの実装を述べる.

4.1 装置

今回行った実験では, 装置として Head Mounted Display (HMD) (Meta Quest Pro), ラップトップ PC (ASUS TUF Gaming A16, 2024), マイク付きヘッドセット (Logicool G433Bk) および WiFi ルータ (Archer AX23V) を使用した. WiFi ルータは HMD と PC の間の WebSocket 通信に用いる. ユーザは HMD を装着した上にマイク付きヘッドセットを装着した.

4.2 実験用アプリケーション

実験用アプリケーションは, HMD 用アプリケーションおよび PC 用アプリケーションから構成される.

HMD 用アプリケーションの実装には Unity 2022.3.59f1 を用いた. HMD 用アプリケーションは, 現在使用している手法および選択されているオブジェクトの情報を PC 用アプリケーションへ送信し, また PC 用アプリケーションから音声コマンドを受信する. さらに, オブジェクトを含めた仮想空間の構成および表示, および, 非言語的発声操作に必要な

発声音量をリアルタイムに得るために HMD のマイクから音声の取得を行う。

PC 用アプリケーションの実装には Python を用いた。本アプリケーションでは、まず既存の音声認識モデルである Vosk¹を用いてユーザの発声を文字に起こす。Vosk が出力する単語が音声コマンドに使われている場合はその単語を音声コマンドとして HMD 用アプリケーションに送信する。また、パイロットスタディにおいて、Vosk により誤認識された単語と音声コマンドを紐づける実装を行った。これにより、音声コマンドの誤認識率を減少させた。なお、Vosk にはローカル環境においても動作し、またこれに伴い低遅延であるという利点があることから我々は今回の実装に採用した。

4.3 表示の詳細

本手法では視野角 30 度の範囲内の各オブジェクトの表面にラベルを頭部座標（左目と右目の間）からのユークリッド距離の近い順に「1」から付与する（図 2b）。1 つでもオブジェクトが範囲から外れる際または範囲に入る際に、ラベルを更新する。また、ラベルを付与されているオブジェクトを 1 つを残して半透明（ $\alpha = 0.3$ ）にする。不透明のオブジェクトは 0.3 秒ごとにラベルの数字順に切り替わる。これらにより、雑多な環境においてもオブジェクトの選択を可能にする。

ラベルに加えて、選択されたオブジェクトを中心に図 2a のような頭部水平座標系を視覚フィードバックとして表示する。ユーザが移動方向および現在位置と目的位置の距離を把握しやすくするため、座標系の各軸には垂直な目盛りが 0.01 m 間隔で表示されている。その色は、オブジェクトの中心から 0.05 m, 0.10 m, 0.15 m, 0.20 m, 0.25 m, 0.30 m において、それぞれ赤色、紫色、緑色、桃色、深緑色、黒色である。その他の目盛りは橙色である。選択したオブジェクトのサイズの x , y , z のいずれかが 0.15 m, 0.30 m, 0.45 m を超えた際には、座標系の視覚フィードバックはそれぞれ 2 倍, 3 倍, 4 倍に拡大される。

オブジェクトの離散的な移動の際には、ユーザが移動を追従しやすくするために、オブジェクトを 0.1 秒かけて滑らかに移動させる。

4.4 発声操作の詳細

非言語的発声操作では、移動方向のフィードバックとして、オブジェクトの色を手前（緑）、奥（茶）、上（赤）、下（青）、右（マゼンダ）、左（シアン）に変化させる。また、移動方向決定後の非言語的発声を用いた連続的な移動の間、移動速度 V は、Bilmes らの研究 [2] を参考に、発声音量によって変化する。まず、現在のユーザの発声音量 a 、キャリブレーション

ンによって求めた発声しない状態の音量 a_{min} 、および基準となる発声音量 a_{max} を用いて、正規化された音量 $x = (a - a_{min}) / (a_{max} - a_{min})$ を求める。その上で移動速度 $V = 0.05e^x$ を求める。

4.5 視線操作および頭部操作の詳細

移動面を意識させるために、移動面の奥に移動面と平行な正方形を視覚フィードバックとして表示した。視線操作および頭部操作においてブレを抑えるために、1€ filter[6] を視線および頭部動作のトラッキングデータに適用した。

5 実験

本手法における空間的オブジェクト移動操作の性能比較のための研究室内実験を行った。

5.1 実験計画

本実験は参加者内計画にて行われた。独立変数は以下の手法の 6 条件と、移動させるオブジェクトおよび後述する目的位置であるターゲットの距離の 2 条件（0.15 m, 0.35 m）の 2 つである。

- eye-nonlinguistic (EN)：視線操作および非言語的発声操作を使用する。
- eye-coordinate (EC)：視線操作および移動距離発声操作を使用する。
- head-nonlinguistic (HN)：頭部操作および非言語的発声操作を使用する。
- head-coordinate (HC)：頭部操作および移動距離発声操作を使用する。
- voiceonly-nonlinguistic (VN)：非言語的発声操作のみを使用する。
- voiceonly-coordinate (VC)：移動距離発声操作のみを使用する。

非言語的発声操作および移動距離発声操作は 3.3.1 節、頭部操作および視線操作は 3.3.2 節にて述べた手法である。また、発声操作のみの操作は頭部水平座標系において、頭部操作または視線操作を用いずに発声操作だけを使ってオブジェクトを動かす手法である。これにより頭部水平座標系における頭部操作または視線操作の有無による影響を調べる。

5.2 タスクおよび従属変数

本実験のタスクは、表示された桃色の球体オブジェクトを青色の球体ターゲットまで移動させることである（図 2b）。オブジェクトの中心がターゲットの中に収まっている状態において、タスクを終了させた場合にタスクを成功とした。参加者は仮想オブジェクトを選択した後、指定された手法を用いて、仮想オブジェクトの移動を行った。その後、「終了」と発声することによって、タスクを終了した。実験参加者は、可能な限り高速かつ正確にオブジェクト

¹ <https://alphacephei.com/vosk/>

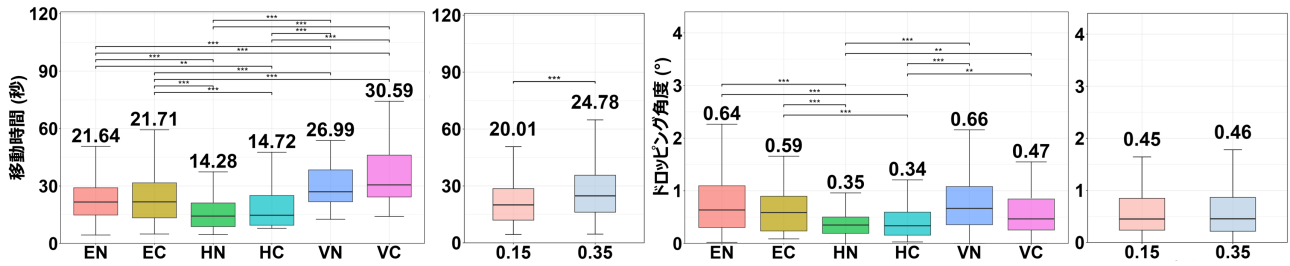


図 4. 実験結果. 左から移動時間 (手法条件別), 移動時間 (距離条件別), ドロッピング角度 (手法条件別), ドロッピング角度 (距離条件別). 有意差は * $p < .05$, ** $p < .01$, *** $p < .001$. ボックス上の数値は中央値.

をターゲットまで移動させるように指示された. タスクは座位にて行われた. タスクにおいて, 参加者はオブジェクトを選択した後, 最初に頭部動作を行うが, その後の頭部操作および発声操作はユーザの自由に行われた. 実験参加者は HMD を頭部に装着して, さらにその上にマイク付きヘッドセットを装着するように指示された.

実験参加者は, オブジェクトとターゲットが表示された時点においてタスクを開始した. また, 「終了」と発声した後, 3 秒後に次のタスクが開始された.

従属変数には, 移動時間 (オブジェクト選択後から, タスク完了までの時間), ドロッピング角度 (試行終了時の頭部とオブジェクトを結ぶ線と, 頭部とターゲットを結ぶ線のなす角度), ユーザビリティ (SUS: System Usability Scale [5]), およびワークロード (NASA-TLX: NASA Task Load Index [25]) を用いた.

5.3 参加者

実験には, 研究室内の男性の大学生および大学院生 12 名 (平均 22.83 歳, $SD = 1.11$ 歳, P1-P12) が参加した. 実験参加者のうち, VR を日常的に使用しているのは 3 名であり, 平均使用年数は 1.83 年 ($SD = 1.26$ 年) であった.

5.4 手順

実験概要の説明を行った後, 実験前アンケートを実施した. 実験参加者は, 年齢, 性別, 利き手, 利き目, VR の使用状況を回答した.

次に実験タスクを説明した後, 6 セッションから構成される実験を実施した. 実験条件はセッションごとに変更された. 実験条件の順番はラテン方格法を用いて, 実験参加者間のカウンタバランスをとることにより決定された. 各セッションにおいて, 実験参加者は発声操作および, 頭部操作または視線操作を確認した後, 練習タスクを行った. 本実験では, オブジェクトおよびターゲットのサイズの 2 条件 (0.0295 m, 0.0495 m) $\times$ オブジェクトとターゲットの距離の 2 条件 (0.15 m, 0.35 m) の 4 条件を行う. 本実験では, 参加者の目から前に 0.5 m の位置を中心とする, 幅 0.5 m, 高さ 0.3 m, 奥行き 0.5 m

表 1. 各手法条件における移動時間 (MT) およびドロッピング角度 (DA) の平均値 (左の値) と標準偏差 (右の値).

条件	MT (秒)	DA (度)
EN	23.96 (13.1)	0.759 (0.581)
EC	25.23 (15.7)	0.627 (0.441)
HN	16.22 (9.13)	0.412 (0.323)
HC	18.77 (10.8)	0.423 (0.354)
VN	30.00 (11.0)	0.763 (0.513)
VC	35.89 (16.2)	0.580 (0.412)

の空間の範囲内にオブジェクトおよびターゲットを生成する. この空間は, 手の可動域および参加者からの見やすさを考慮して決定した. オブジェクトおよびターゲットは, 与えられた距離およびサイズの条件下において, 空間の中心を基準とする (1, 1, 1), (1, 1, -1), (1, -1, 1), (-1, 1, 1) のベクトル配置から 0 度から 15 度の範囲にて回転させた配置関係にて生成される. 練習タスクでは, オブジェクトおよびターゲットのサイズおよび距離の 4 条件をランダムな順において 1 回ずつ行った. 続いて, 本番タスクを行った. 本番タスクでは, オブジェクトおよびターゲットのサイズおよび距離の 4 条件をランダムな順において 2 回ずつ行った. 本番タスク終了後, 実験参加者は, SUS, NASA-TLX, 自由記述から構成されるアンケートに回答した. 全セッション終了後, 実験後アンケートとして, 自由記述を行った.

6 実験結果

実験において, タスクの失敗が EC において 2 回, HN において 1 回, VC において 1 回見られた. また, 頭部操作または視線操作のみを用いてオブジェクトをターゲットまで動かすことができたため, 発声操作が不要だった試行が 2 回あった.

6.1 移動時間およびドロッピング角度

移動時間 (低いほど良い) およびドロッピング角度 (低いほど良い) を, それぞれ図 4 および表 1 に示す. 移動時間およびドロッピング角度にシャピロ・

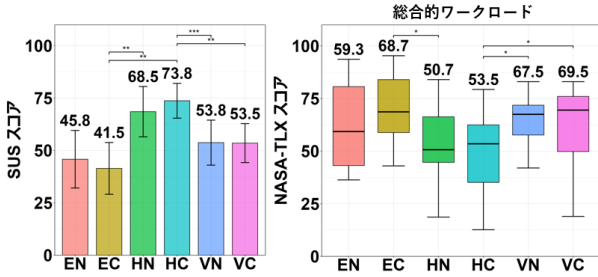


図 5. 実験結果. 左: 手法条件ごとの SUS スコア. エラーバーは 95%信頼区間. バー上の数値は平均値. 右: 手法条件ごとの NASA-TLX 総合的ワークロードスコア. ボックス上の数値は中央値. 有意差はそれぞれ * $p < .05$, ** $p < .01$, *** $p < .001$.

ウィルク検定を行った結果, 正規性は確認されなかった. そのため, 手法条件および距離条件を独立変数として整列ランク変換 [23] を行った後, 二元配置分散分析を行った. さらに, 両条件に下位検定として, ART-C [8] を実施した.

移動時間について, 手法条件 ($F_{5,501.09} = 56.66$, $p < .001$) および距離条件 ($F_{1,501.07} = 50.28$, $p < .001$) に主効果が確認された. ドロッピング角度について, 手法条件 ($F_{5,501.09} = 15.72$, $p < .001$) に主効果が確認された.

6.2 SUS および NASA-TLX

SUS スコア (高いほど良い) および NASA-TLX スコア (低いほど良い) を図 5 に示す. SUS および NASA-TLX は手法条件のみを独立変数にする. SUS にシャピロ・ウィルク検定を行った結果, 正規性が確認された. そのため, 一元配置分散分析を行った. その後, 下位検定として t 検定を実施した後, Holm 補正を行った. SUS について, 手法条件 ($F_{5,55} = 8.65$, $p < .001$) に主効果が確認された.

NASA-TLX にシャピロ・ウィルク検定を行った結果, 正規性が確認されなかった. そのため, フリードマン検定を行った. その後, 下位検定として, ウィルコクソンの符号順位検定を実施した後, Holm 補正を行った. 総合的ワークロードについて, 手法条件 ($\chi^2(5) = 19.50$, $p < .01$) に主効果が確認された.

7 考察

実験結果より, 視線操作と発声操作を用いる手法 (EN, EC) および頭部操作と発声操作を用いる手法 (HN, HC) は, 発声のみの手法 (VN, VC) よりも有意に高速な移動時間を達成した. また頭部操作を用いる手法は発声操作のみを用いる手法よりもドロッピング角度が有意に低いうえ, 特に HC は VN と VC よりも SUS スコアおよび総合的ワークロードスコアが有意に優れていた. これらの結果は頭部操作または視線操作と発声操作を組み合わせること

によってパフォーマンスを向上できることを示している.

非言語的発声操作を用いた操作 (EN, HN, VN) と移動距離発声操作を用いた操作 (EC, HC, VC) を比べた際に, 発声操作ごとに移動時間, ドロッピング角度, ワークロード, ユーザビリティに有意差は確認されなかった. 一方で, 中央値およびグラフから見える傾向としては, 移動時間は非言語的発声操作の方が低い傾向があった. 非言語的発声操作に関するコメントとしては, 「微調整が行いやすい」 (P1, P5, P8) および 「長距離の移動が疲れる」 (P6) との回答があった. また移動距離発声操作に関しては, 「精度がよい」 (P6, P12) および 「大きな移動には便利」 (P8) との回答があった. ゆえに, 非言語的発声操作は移動時間, 移動距離発声操作は精度における有用性が示唆されている. このため, 両者の使い分けが重要であると言える.

頭部操作を用いる手法 (HN, HC) は, 視線操作を用いる手法 (EN, EC) よりも有意に高速かつ高精度であることが確認された. これは, HMD の視線推定精度の低さおよび視線の揺れから, 視線操作によるオブジェクト操作は困難であったためだと考えられる. ゆえに, 発声操作と組み合わせる操作手法として, 視線よりも頭部の方が優れていることが明らかになった.

8 まとめ

本論文では, 頭部水平座標系を用いた仮想オブジェクト操作手法を示した. 研究室実験を通じて, 視線または頭部操作を組み合わせる手法および発声操作のみを用いた手法の性能を調査した. 実験の結果, 特に, 頭部操作および非言語的発声操作を組み合わせる条件の平均移動時間は 16.22 秒, 平均ドロッピング角度は 0.412 度であった. これは発声のみの手法と比べて有意に高速かつ高精度であった. この結果より, 本手法は仮想オブジェクト操作手法として使用可能であることが示唆された.

今後は, 研究室実験を通じて本手法の正確な性能を調査するとともに, Mihara らの研究 [18] のように, 非言語的発声および移動距離発声操作の双方を用いた実装および母音による操作の切り替え [2, 9] の実装についても検討を進める.

参考文献

- [1] F. Aslam and R. Zhao. Voice-Augmented Virtual Reality Interface for Serious Games. In *2024 IEEE Conference on Games (CoG)*, pp. 1–8, 2024.
- [2] J. A. Bilmes, X. Li, J. Malkin, K. Kilanski, R. Wright, K. Kirchhoff, A. Subramanya, S. Harada, J. A. Landay, P. Dowden, and

- H. Chizeck. The vocal joystick: a voice-based human-computer interface for individuals with motor impairments. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, HLT '05, pp. 995–1002, USA, 2005. Association for Computational Linguistics.
- [3] J. Blattgerste, P. Renner, and T. Pfeiffer. Advantages of eye-gaze over head-gaze-based selection in virtual and augmented reality under varying field of views. In *Proceedings of the Workshop on Communication by Gaze Interaction*, COGAIN '18, pp. 1–9, New York, NY, USA, 2018. Association for Computing Machinery.
- [4] R. A. Bolt. “Put-that-there”: Voice and gesture at the graphics interface. In *Proceedings of the 7th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '80, pp. 262–270, New York, NY, USA, 1980. Association for Computing Machinery.
- [5] J. Brooke. SUS: A Quick and Dirty Usability Scale. *Usability Evaluation in Industry*, pp. 189–194, 1996.
- [6] G. Casiez, N. Roussel, and D. Vogel. 1 € filter: a simple speed-based low-pass filter for noisy input in interactive systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pp. 2527–2530, New York, NY, USA, 2012. Association for Computing Machinery.
- [7] C.-L. Deng, C.-Y. Tian, and S.-G. Kuai. A combination of eye-gaze and head-gaze interactions improves efficiency and user experience in an object positioning task in virtual environments. *Applied Ergonomics*, 103:103785, 2022.
- [8] L. A. Elkin, M. Kay, J. J. Higgins, and J. O. Wobbrock. An Aligned Rank Transform Procedure for Multifactor Contrast Tests. In *The 34th Annual ACM Symposium on User Interface Software and Technology*, UIST '21, pp. 754–768, 2021.
- [9] S. Harada, J. A. Landay, J. Malkin, X. Li, and J. A. Bilmes. The vocal joystick: evaluation of voice-based cursor control techniques. In *Proceedings of the 8th International ACM SIGACCESS Conference on Computers and Accessibility*, Assets '06, pp. 197–204, New York, NY, USA, 2006. Association for Computing Machinery.
- [10] J. Hombeck, H. Voigt, T. Heggemann, R. R. Datta, and K. Lawonn. Tell Me Where To Go: Voice-Controlled Hands-Free Locomotion for Virtual Reality Systems. In *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pp. 123–134, 2023.
- [11] X. Hu, D. Ma, F. He, Z. Zhu, S.-K. Hsia, C. Zhu, Z. Liu, and K. Ramani. GesPrompt: Leveraging Co-Speech Gestures to Augment LLM-Based Interaction in Virtual Reality. In *Proceedings of the 2025 ACM Designing Interactive Systems Conference*, DIS '25, pp. 59–80, New York, NY, USA, 2025. Association for Computing Machinery.
- [12] T. Igarashi and J. F. Hughes. Voice as sound: using non-verbal voice input for interactive control. In *Proceedings of the 14th Annual ACM Symposium on User Interface Software and Technology*, UIST '01, pp. 155–156, New York, NY, USA, 2001. Association for Computing Machinery.
- [13] S. Jalaliniya, D. Mardanbeigi, T. Pederson, and D. W. Hansen. Head and Eye Movement as Pointing Modalities for Eyewear Computers. In *2014 11th International Conference on Wearable and Implantable Body Sensor Networks Workshops*, pp. 50–53, 2014.
- [14] M. Kytö, B. Ens, T. Piumsomboon, G. A. Lee, and M. Billingham. Pinpointing: Precise Head- and Eye-Based Target Selection for Augmented Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pp. 1–14, New York, NY, USA, 2018. Association for Computing Machinery.
- [15] J. Lee, S. S. Rodriguez, R. Natarrajan, J. Chen, H. Deep, and A. Kirlik. What 's This? A Voice and Touch Multimodal Approach for Ambiguity Resolution in Voice Assistants. In *Proceedings of the 2021 International Conference on Multimodal Interaction*, ICMI '21, pp. 512–520, New York, NY, USA, 2021. Association for Computing Machinery.
- [16] J. Lee, J. Wang, E. Brown, L. Chu, S. S. Rodriguez, and J. E. Froehlich. GazePointAR: A Context-Aware Multimodal Voice Assistant for Pronoun Disambiguation in Wearable Augmented Reality. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI '24, pp. 1–20, New York, NY, USA, 2024. Association for Computing Machinery.
- [17] C. Liu, A. Plopski, and J. Orlosky. OrthoGaze: Gaze-based three-dimensional object manipulation using orthogonal planes. *Computers Graphics*, 89:1–10, 2020.
- [18] Y. Mihara, E. Shibayama, and S. Takahashi. The migratory cursor: accurate speech-based cursor movement by moving multiple ghost cursors using non-verbal vocalizations. In *Proceedings of the 7th International ACM SIGACCESS Conference on Computers and Accessibility*, Assets '05, pp. 76–83, New York, NY, USA, 2005. Association for Computing Machinery.
- [19] K. Minakata, J. P. Hansen, I. S. MacKenzie, P. Bækgaard, and V. Rajanna. Pointing by gaze, head, and foot in a head-mounted display. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, ETRA '19, pp. 1–9, New York, NY, USA, 2019. Association for Computing Machinery.
- [20] P. Monteiro, G. Gonçalves, H. Coelho, M. Melo, and M. Bessa. Hands-free interaction in immersive virtual reality: A systematic review. *IEEE Transactions on Visualization and Computer Graphics*, 27(5):2702–2713, 2021.
- [21] Y. Y. Qian and R. J. Teather. The eyes don't have it: an empirical comparison of head-based

- and eye-based selection in virtual reality. In *Proceedings of the 5th Symposium on Spatial User Interaction*, SUI '17, p. 91–98, New York, NY, USA, 2017. Association for Computing Machinery.
- [22] U. Wagner, A. Asferg Jacobsen, T. Feuchter, H. Gellersen, and K. Pfeuffer. Eye-Hand Movement of Objects in Near Space Extended Reality. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*, UIST '24, pp. 1–13, New York, NY, USA, 2024. Association for Computing Machinery.
 - [23] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins. The Aligned Rank Transform for Nonparametric Factorial Analyses using Only Anova Procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pp. 143–146, 2011.
 - [24] L. Zhang, J. Pan, J. Gettig, S. Oney, and A. Guo. VRCopilot: Authoring 3D Layouts with Generative AI Models in VR. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*, UIST '24, pp. 1–13, New York, NY, USA, 2024. Association for Computing Machinery.
 - [25] 芳賀 繁, 水上 直樹. 日本語版 NASA-TLX によるメンタルワークロード測定—各種室内実験課題の困難度に対するワークロード得点の感度. 日本人間工学会誌, 32(2):71–79, 1996.