

HMD環境における両手動作を用いた日本語文字入力手法

古野 祐貴¹ 岨野 太一² 高田 秀志²

概要：VR/MR 空間ではコミュニケーションや情報検索に伴いテキスト入力が必要となるが、外部デバイスの追加や精密な接触操作に依存する方式では、入力姿勢や設置条件の制約を受けやすいという問題が存在する。そこで本研究では、HMD 内蔵カメラによるハンドトラッキングを用いた両手動作型の日本語文字入力手法を提案する。提案手法は、右手の手形と手首ロールで子音行を選択し、左手の親指と人差し指によるピンチとスワイプ動作で母音を選択して文字を確定する。比較手法（フリック入力）との実験では、1 正解文字あたりの手の運動量が提案手法で低い傾向を示し、NASA-TLX の身体的要求も低下した一方で、入力速度は低下し、精神的要求と努力が増加した。

1. はじめに

ヘッドマウントディスプレイ（HMD）は普及傾向にあり、メタバース関連サービスの利用も拡大している。それに伴い、ユーザがバーチャルリアリティ（VR）空間やミックスドリリアリティ（MR）空間内で情報検索やコミュニケーションを行う機会が増加している。VR 空間内では音声を紹介したコミュニケーションが中心となりやすい一方で、メッセージやメールへの返信、ユーザネーム入力、パスワード入力など、テキスト入力が求められる場面も存在する。これまで HMD 環境での文字入力手法として、物理キーボードを用いる方式、HMD 環境に仮想的なキーボードを表示する方式、音声認識を用いる方式、ジェスチャーを用いる方式などが提案されてきた。

物理キーボード方式は、現実空間の手元を視認しにくい没入型 VR ではタッチタイピングへの依存が大きく、入力姿勢や設置位置の制約も受けやすい。仮想キーボード方式は、学習コストが低い一方で、キー選択が逐次的になりやすく、物理的フィードバックの不足から指先位置の確認が必要となる場合がある。音声認識方式は高速入力が期待できるが、音声入力に対する忌避感や誤認識の修正負担や使用環境に依存する課題が残る。これらに対し、近年は HMD に内蔵されたカメラにより外部デバイスなしでハンドトラッキングが利用可能となり、現実空間内の姿勢や位置にとらわれずに入力できるジェスチャー方式は有力な選択肢となり得る。

一方で、ジェスチャー入力は、学習コストに加え、入力中に腕や手指を広く動かす設計では身体負担が増加しや

すいという課題を持つ。そこで本研究では、ハンドジェスチャーを用いた日本語文字入力に着目し、両手動作を活用して入力操作を分担することで、入力に必要な腕の動作量を低減しつつ、HMD 環境における文字入力効率の向上を図る。

2. 関連研究

2.1 HMD 環境の文字入力

2.1.1 HMD 前面にタッチデバイスを取り付けた手法

山口らは、HMD 前面に取り付けたタッチデバイスを物理的な入力面として用い、触覚フィードバックを得ながら日本語フリック入力を行う手法を提案した [1]。本手法では、従来の VR フリック入力手法と比較して入力速度が向上し、主観的ワークロードが低下した一方で、トータルエラー率は増加したと報告している。また、入力はトラッキングセンサ範囲外の HMD 前面で行われるため、手指を VR 内に表示できず、固有感覚に依存した操作となる点、タッチデバイスの追加による HMD 重量増加や、HMD 前面での接触操作自体が腕負担になり得る点が課題として挙げられている。

2.1.2 手掌上に仮想フリックキーボードを配置する手法

Ogitani らは、手掌上に仮想フリックキーボードを配置する文字入力手法を提案した [2]。また、事前実験として、画面中央に配置した仮想フリックキーボードと仮想 QWERTY キーボードを用いて日本語の入力速度を比較し、HMD 環境における日本語入力では、フリックキーボードが QWERTY キーボードより高い入力速度を得られる可能性を示した。

提案手法である手掌上のキーボード配置は、テキスト入

¹ 立命館大学情報理工学研究科

² 立命館大学情報理工学部

力時の安全性や没入感を損なうことなくキーボードを任意の位置に提示できる点で有望である。しかし、両手の重なり起因する追跡誤りが誤入力を誘発し、画面中央に配置する方式と比較して速度および正確性が低下する点が課題として報告されている。

2.1.3 片手ジェスチャによる文字入力手法

Chuらは、ウェアラブル環境での文字入力において、従来の手認識における背景の複雑さに起因する頑健性不足や、文字をジェスチャへ符号化する際の学習コストが高くなりやすい点を課題として整理し、少数の静的ジェスチャで文字入力を行う方式を提案した[3]。同手法ではアルファベットを2次元配置し、行と列を指定する2段階の入力として、7種類の手形ジェスチャで文字を選択する。評価結果としては入力速度5.6 wpm、平均誤り率1.79%を報告している。しかし、1文字あたり複数回のジェスチャが必要になる設計そのものが速度向上の上限になり得る。実際にChuらは、将来的な改善として両手を用いた入力や、より効率的な文字符号化の検討を挙げている。また、同手法では文字を2段階で選択するという構造上、1段階目で誤った行や領域を選択した場合でも、途中での訂正が難しく、2段階目の操作を経て一度文字を確定した後に削除して再入力する手順になる。このような誤選択の訂正コストは、学習途上のユーザほど顕在化しやすく、速度低下や修正負担の増大につながり得る。

2.1.4 仮想キーボードとジェスチャーを組み合わせた入力手法

Rahimらは、仮想キーボード上のRegion of Interest (ROI) に対して手形ジェスチャーを認識し、ジェスチャ機能により文字選択を行う非接触文字入力を提案した[4]。実験では平均は1文字あたり2.08 s、誤り率2.07%を報告している。一方で、文字選択の誤り要因として、ユーザが文字境界付近でジェスチャーを行うと誤認識が生じることを指摘している。このように、ROIや境界への依存が強い方式では、わずかな手ぶれや個人差が誤選択につながりやすく、HMD環境の長文入力において安定性や修正コストの面で不利になり得る。

2.2 本研究のアプローチ

そこで本研究では、これら既存研究の問題点を踏まえ、外部デバイスの追加や精密な接触操作に依存せず、HMD内蔵ハンドトラッキングで成立する両手動作型の日本語文字入力手法を提案する。提案手法は、(i) 両手で役割分担することで片手の連続的な移動と保持を分散し、1文字あたりの腕動作量の抑制を狙う。また、(ii) 特定の領域を精密に狙う操作ではなく、手の形と角度により日本語の子音行を選択し、別の手で母音列を選択する構造とすることで、境界付近で誤選択が生じるという課題を構造的に回避し、入力の安定性を確保する。(iii) 文字確定のタイミングを母

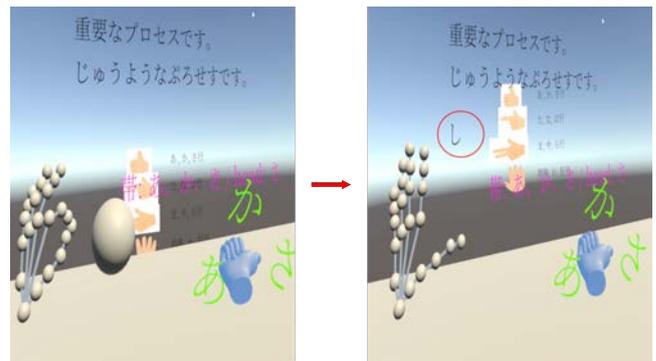


図1 「し」を入力している例

音選択側に集約し、子音行の選択は確定前に何度でも再選択できる設計とする。これにより、行選択の誤りが生じても、文字を一度確定して削除する手順に依存せず、手の形や角度の調整のみで修正できるため、誤り訂正行動の負担低減が期待できる。

加えて、操作は少数で分かりやすいジェスチャに限定し、学習コストを過度に増やさない範囲で設計する。一方で、入力速度や誤り訂正行動については、身体的負担低減を優先した上での副次的な設計目標として位置づけ、比較手法と比較して著しい悪化を生じさせないことを目指す。

3. 提案手法

本研究では、HMD環境におけるテキスト入力の操作負担を低減するため、外部デバイスを用いず、HMD内蔵のハンドトラッキングのみを用いて日本語(かな)を入力可能な両手ジェスチャ入力手法を提案する。提案手法は、(1) 右手で子音行(あ行、か行、さ行、...)を選択し、(2) 左手で母音(あ・い・う・え・お)を選択して1文字を確定する、という両手動作に基づく二段階入力である。これにより、各手の移動量を抑えつつ、1段階目動作による文字の訂正・再選択が可能な入力を実現する。

3.1 入力方法

ユーザは右手の指の伸展状態、手首のロール角度、および左手スワイプ動作を用いて入力を行う。濁音・半濁音・促音(小文字)などの特殊表記は変換用の行を左手で選択する。

3.1.1 基本入力(清音)

1文字の入力は、右手による子音行選択と左手による母音選択の2段階で構成される。まず右手で行グループを選択し、次に同グループ内の子音行を選択する。最後に左手で母音を選択し、確定操作によってテキストへ反映する。入力方法の詳細は以下のとおりである。

- (1) 右手の指の形(伸展本数)により行グループを選択する。
- (2) 手首のロール角度により、同グループ内の子音行を選

択し、子音行を決定する。

- (3) 左手の確定ジェスチャ（親指と人差し指をピンチし、対応する母音の方向にスワイプを行ったのち指を離す）により文字を確定し、入力文字列へ反映する。

図1は実際に「し」の入力を行っている様子である。右手で「し行」を選択し、左手で親指と人差し指をつまむ動作を行い、左へ動かしたのちに、指を離すことで入力される。

3.1.2 行グループ選択（右手の指形状）

右手の指の伸展状態を4種類に分類し、行グループ選択に用いる。表1は指形状と行グループの対応を示すものである。

| 右手の指形状（例） | 選択される行グループ |
|------------|------------|
| 親指のみ伸展 | あ行・か行・さ行 |
| 親指+人差し指 | た行・な行・は行 |
| 親指+人差し指+中指 | ま行・や行・ら行 |
| 全指を伸展 | 特殊・わ行・記号 |

3.1.3 手首のロール角度による行グループ内子音選択

行グループ内の3候補選択（例：あ行・か行・さ行）は、親指を用いて計算される手首のロールに基づき決定される。表2は左手首ロールによって選択される子音行の例である。

| 親指方向（離散） | 選択される子音行 |
|----------|----------|
| 左方向 | あ行 |
| 上方向 | か行 |
| 右方向 | さ行 |

3.1.4 左手による母音選択

左手は母音（あ・い・う・え・お）の選択と文字確定を担う。母音は、左手親指と人差し指をつまむ動作を行った後、中心または、上下左右の4方向のいずれかへ動かしたのちに、指を離すことで入力される。表3は左手動作の対応を示すものである。

| 親指方向（離散） | 母音 |
|------------|----|
| 中心（ニュートラル） | あ |
| 左方向 | い |
| 上方向 | う |
| 右方向 | え |
| 下方向 | お |

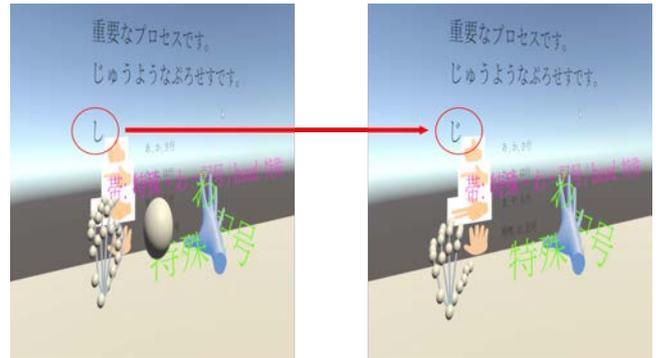


図2 「し」を「じ」に変換している例

3.1.5 特殊表記（濁音・半濁音・促音など）

濁音・半濁音・促音・文字削除などは、基本文字入力後に変換ジェスチャで付与する。本研究では右手で「特殊行」を選択した状態をトリガとして、左手スワイプ方向により変換種別を切り替える方式を採用する。

- **文字削除**：直前に文字があるとき、右手で「特殊行」を選択後、左手を動かさず、中心を選択することにより文字削除。
- **濁音**：直前文字が対象（か行・さ行・た行・は行）であるとき、右手で「特殊行」を選択後、左手を左方向に動かすことにより濁点付与（例：か → が）。図2は直前に入力された文字「し」を、右手で「特殊行」を選択し、左手を左方向に動かすことで「じ」に変換している例である。
- **半濁音**：直前文字が「は行」であるとき、「右手で「特殊行」を選択後、左手を右方向に動かすことにより濁半濁点付与（例：は → ぱ）。
- **促音・小文字**：直前文字が対象（例：ア行・た行・や行等）であるとき、右手で「特殊行」を選択後、左手を上方向に動かすことにより小文字化（例：つ → っ、や → ゃ）。

3.2 実装

3.2.1 動作環境

提案手法はUnity 2021.3.45f1上で実装し、Meta All-in-One SDKを用いてハンドトラッキング情報を取得する。HMDにはMeta Quest 3を用い、Quest LinkによりPCへ接続した状態で実装したソフトウェアを動作させている。

3.2.2 入力判定と状態管理

入力は「右手の手の形による行グループ選択」「右手のロール角度による子音行選択」「左手による母音選択」「確定」の流れに沿って状態遷移として管理する。各状態では、手の指形状（伸展本数）および手首のロール角度（親指向きベクトル）を参照して候補を決定し、確定操作により文字列へ反映する。



図 3 画面中央に表示される手の形に関する画像

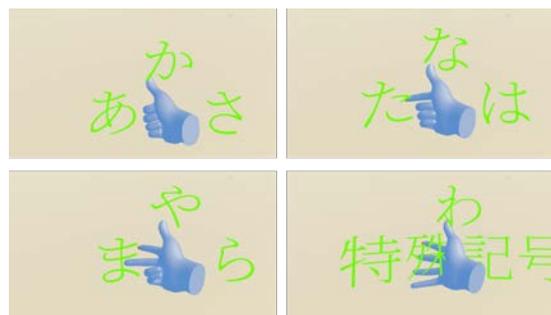


図 4 手首周辺に表示される各子音の配置

3.2.3 手首のロール角度の算出と離散化

親指の向きベクトルは、親指付け根付近の関節点から親指先端点への方向として定義する。これを手のローカル座標系に変換したのち、水平面内角度（ロール）を、子音行選択（3 領域）に用いる。

3.2.4 誤入力抑制と安定化

ハンドトラッキングの揺らぎに起因する誤入力を抑制するために母音選択時、一定時間は次入力の受付を抑制し、多重確定を防ぐ。

3.2.5 視覚フィードバック

ユーザが文字を入力する際に、手の形を忘れることや混乱を防ぐために、画面中央に行グループを選択するための手の形の対応を画像を用いて表示した。図 3 は画面中央に表示される手の形に関する画像を示したものである。また、ロール角度と候補の対応関係を提示する視覚フィードバックとして、手首周辺に各子音をロール方向に応じて配置した。図 4 は各指の形に対応して付近に表示される子音の表示を示したものである。

4. 評価実験

提案手法の有効性を検証するために実施した評価実験について述べる。

4.1 実験方法

本実験では、HMD 環境における日本語テキスト入力を対象とし、入力速度、バックスペース率、および NASA-TLX による主観的ワークロードを指標として、提案手法とフリック入力手法を比較評価する。

4.1.1 実験手順

手法の提示順序による影響を抑えるため、2 グループに分け、手法順序は参加者間でカウンタバランスをとった。各手法における手順は以下の通りである。参加者は当該手法に関する説明を受けた後、練習を 5 分間行い、続いて 5 分間の入力テストを 2 回実施した。すべての入力テスト前には 5 分間の休憩を設け、疲労の蓄積を抑制した。また、各手法の入力テスト 2 回目終了後に、NASA-TLX（6 項目）への回答を求めた。

4.1.2 参加者

参加者は立命館大学に在籍する 20 歳から 22 歳の 4 名の大学生である。全員が日本語を母語とし、視力は裸眼または矯正により日常生活に支障がない者である。また、全ての参加者はスマートフォンにおいてフリック入力を日常的に利用している。

4.1.3 入力文

入力文は、1 文あたりのひらがな表記の文字数が 10~20 文字となるように、生成 AI を用いて複数文を事前に生成した。文中には清音に加えて濁音、半濁音、促音、および句読点が含まれ得るが、これらの出現は各文で必須ではない。生成した文章群から各テストで提示する文章セットを作成した。文章セットはテストごとに異なるように割り当て、セット間で同一文が重複しないようにした。

4.1.4 記録データ

評価指標の算出のため、入力中の文字列変化履歴（挿入・削除）、各イベントの時刻（ミリ秒精度）、確定文字、削除操作、および特殊操作（濁音・半濁音・小文字）をログとして記録した。加えて、課題文の入力に必要な手の動きを評価するため、左右手の位置および姿勢を記録した。

4.1.5 評価指標

客観指標として入力速度、誤り訂正の負担、および入力に必要な手の運動量、主観指標として NASA-TLX を用いた。入力速度は Characters Per Minute (CPM) として算出し、誤り訂正の負担はバックスペース率として算出した。入力に必要な手の動きは、入力中に計測した左右手の移動距離の総和を正解文字数で正規化した 1 正解文字あたり運動量 $L_{\text{per_correct}}$ [m/char] として算出した。主観評価では NASA-TLX の 6 項目（精神的要求、身体的要求、時間的圧迫感、作業達成度、努力、フラストレーション）について各項目スコアおよび平均スコアを算出した。

4.2 実験結果

4.2.1 入力速度 (CPM)

入力速度の結果は練習区間を除外したテスト区間のみを対象として集計した。図 5 に示すように、CPM の平均値は比較手法が 28.59 (SD = 2.75)、提案手法が 12.12 (SD = 3.43) であった。提案手法は比較手法と比較して CPM が低い傾向を示し、平均値ベースでは約 57.6% 低下

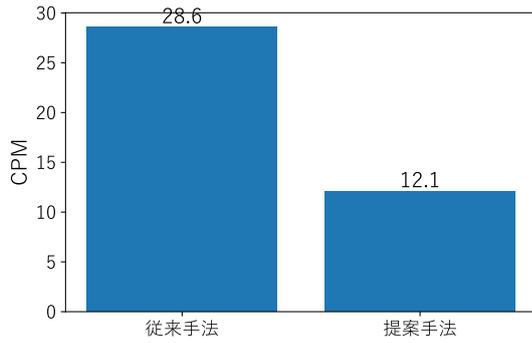


図 5 手法間の平均 CPM (練習除く)

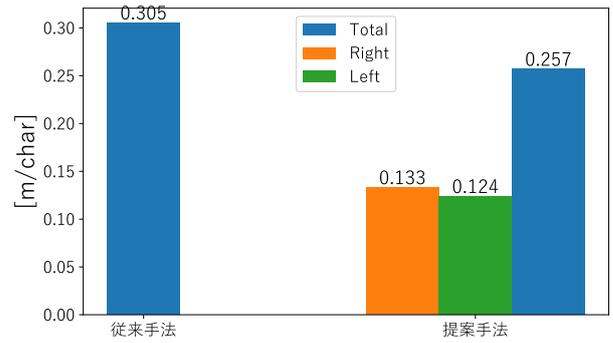


図 7 1 正解文字あたり運動量

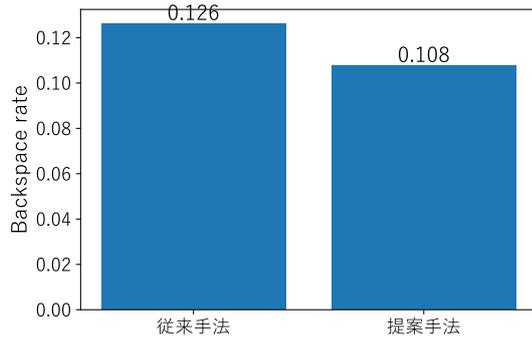


図 6 手法間の平均バックスペース率 (練習除く)

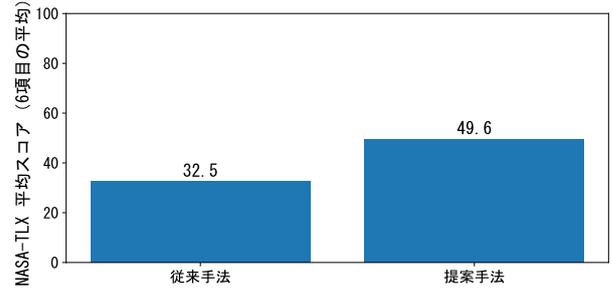


図 8 手法ごとの NASA-TLX スコア (6 項目平均)

した. 対応のある Wilcoxon 符号付順位検定を行った結果, 比較手法と提案手法の CPM に有意差は認められなかった ($p = 0.125$).

4.2.2 バックスペース率

誤り訂正の傾向を評価するため, 入力中のバックスペース率を算出した. 図 6 に示すように, バックスペース率の平均値は比較手法が 0.126 (SD = 0.046), 提案手法が 0.108 (SD = 0.023) であった. 提案手法は比較手法と比較してバックスペース率が低い傾向を示し, 平均値ベースでは約 14.3%低下した. 対応のある Wilcoxon 符号付順位検定を行った結果, 比較手法と提案手法のバックスペース率に有意差は認められなかった ($p = 0.625$)

4.2.3 運動量

なお, 本研究における運動量は, HMD 内蔵ハンドトラッキングにより取得した親指先端の 3 次元位置を用い, テスト区間中の連続サンプル間のユークリッド距離を積算した総移動距離 L_{total} を指す. 右手および左手の移動距離をそれぞれ

$$L_{right} = \sum_{t=2}^T \mathbf{p}_t^{\text{right}} - \mathbf{p}_{t-1}^{\text{right}}, \quad L_{left} = \sum_{t=2}^T \mathbf{p}_t^{\text{left}} - \mathbf{p}_{t-1}^{\text{left}}$$

とし, $L_{total} = L_{right} + L_{left}$ と定義した. ここで $\mathbf{p}_t^{\text{right}}, \mathbf{p}_t^{\text{left}}$ はそれぞれ時刻 t における右手・左手の親指先端位置である. さらに, 1 正解文字あたり運動量は

$$L_{per_correct} = \frac{L_{total}}{N_{correct}}$$

として算出した. ただし, $N_{correct}$ はテスト区間における正解文字数である.

得られた各手法ごとの 1 正解文字あたり運動量 $L_{per_correct}$ [m/char] を図 7 に示す. $L_{per_correct}$ の平均値は比較手法が 0.305 (SD = 0.075) [m/char], 提案手法が 0.257 (SD = 0.058) [m/char] であり, 提案手法は比較手法に比べて 1 正解文字あたり運動量が約 15.9%低い傾向を示した. 一方で, 対応のある Wilcoxon 符号付順位検定では有意差は確認されなかった ($p = 0.625$).

4.2.4 主観評価 (NASA-TLX)

図 8 に示すように, NASA-TLX スコア (6 項目平均) は比較手法が 32.5, 提案手法が 49.6 であり, 提案手法の方が主観的ワークロードが高い傾向を示した.

項目別平均を図 9 に示す. 提案手法は比較手法と比較して, 精神的要求が高く (78.8 vs 20.0), 作業達成度が高く (55.0 vs 25.0), 努力が高かった (67.5 vs 21.2). 一方で, 身体的要求は比較手法の方が高く (80.0 vs 46.2), 比較手法では身体的負担が大きいことが示唆された. 時間的圧迫感や両手法で同程度 (いずれも 17.5) であり, フラストレーションは提案手法がわずかに高かった (32.5 vs 31.2). 以上より, 提案手法は身体的負担の低減が期待できる一方で, 入力時の認知的負荷や達成度に関する負の評価, および努力量の増大が課題として示された.

4.3 考察

1 正解文字あたり運動量は提案手法で低い傾向を示しており, NASA-TLX の身体的要求が提案手法で低下してい

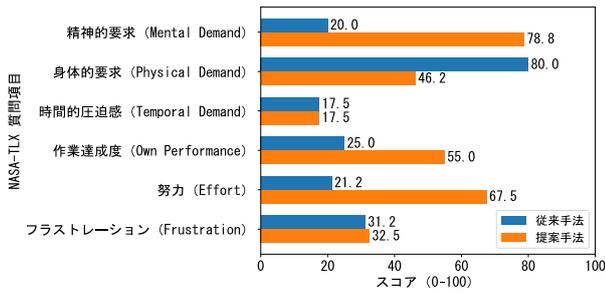


図 9 手法ごとの NASA-TLX 項目別平均 (0-100).

る結果とも整合していることから、主観評価においても身体的な「きつき」が軽減されたことが示唆される。したがって、提案手法は腕の移動や保持に由来する負担を軽減する方向に作用した可能性が高い。

一方で、提案手法は CPM が大きく低下し、NASA-TLX の精神的要求、作業達成度、努力が上昇している。これらは相互に関連した結果であると考えられる。すなわち、提案手法では入力操作の手順や判断が増えることで、どの操作を行うべきかを継続的に考える必要が生じ、その結果として入力が遅くなり、期待したパフォーマンスを達成できない感覚が強まり、より大きな努力を要した可能性がある。この解釈は、バックスペース率が大きく増加していない点とも整合する。もし速度低下が誤入力の頻発に起因するならば訂正操作が増えバックスペース率が上昇しやすいが、本結果では提案手法のバックスペース率はむしろ低い傾向である。したがって、速度低下の主因は誤り訂正の増加ではなく、入力単位あたりの選択・確定に要する時間や、判断・探索に伴う認知処理にある可能性が高い。

5. おわりに

本研究では、HMD 環境における日本語テキスト入力において、文字入力利用時の腕・肩・手指に生じる身体的負担が課題となる点に着目し、両手動作を用いた日本語文字入力手法を提案した。提案手法は、手の形と角度による行選択と、反対の手による列選択を組み合わせることで、精密な位置合わせを要求せずに文字選択を行う。評価実験の結果、1 正解文字あたり運動量は提案手法で低い傾向を示し、NASA-TLX の身体的要求も低下した一方で、入力速度 (CPM) は比較手法より低く、NASA-TLX の精神的要求、作業達成度、努力が高い傾向を示した。

これらの結果から、提案手法は身体的負担の低減に寄与し得る一方で、短期的には認知的負荷の増大により入力速度と主観的達成感が低下する可能性が示唆された。今後の課題として、熟達によって入力速度や主観評価がどの程度改善するかを明らかにするため、より長い練習期間を含む縦断的評価や、被験者数を増やした検証を行う必要がある。

参考文献

- [1] 山口泰生, 西川宜利, 志築文太郎: HMD 前面を用いた VR 用フリック文字入力手法の予備評価, 情報処理学会研究報告, Vol. 2023-HCI-205, No. 12 (2023).
- [2] Ogitani, T., Arahori, Y., Shinyama, Y. and Gondow, K.: Space saving text input method for head mounted display with virtual 12-key keyboard, in *2018 IEEE 32nd International Conference on Advanced Information Networking and Applications (AINA)*, pp. 342-349 IEEE (2018).
- [3] Chu, Q., Chen, C. P., Hu, H., Wu, X. and Han, B.: iHand: Hand Recognition-Based Text Input Method for Wearable Devices, *Computers*, Vol. 13, No. 3 (2024).
- [4] Rahim, M. A., Shin, J. and Islam, M. R.: Hand gesture recognition-based non-touch character writing system on a virtual keyboard, *Multimedia Tools and Applications*, Vol. 79, No. 17, pp. 11813-11836 (2020).