



WISS 2025

WISS2025 開催にあたって

塚田 浩二*

本論文集は、日本ソフトウェア科学会インタラクティブシステムとソフトウェア (ISS) 研究会が主催し、2025 年 12 月 03 日 (水) から 12 月 05 日 (金) にかけて、定山溪ビューホテル (北海道札幌市南区定山溪温泉東) で開催される「第 33 回インタラクティブシステムとソフトウェアに関するワークショップ」(Workshop on Interactive Systems and Software: WISS2025) での発表をまとめたものです。

WISS は、伝統的に 2 泊 3 日の泊まり込み形式で開催されており、例年 150 名以上の参加者が朝から深夜まで深い議論を行ってきました。Covid-19 のために、オンライン／ハイブリッド開催された年もありましたが、WISS2025 では約 250 名が参加して現地開催されます。

昨年の WISS2024 では、「開かれた WISS」を目指して、「登壇発表種別をショート枠相当に統一」「ロングティザー発表の廃止」等の様々な整理を行いました。こうした整理は、参加者フィードバックでも概ね好評でしたので、WISS2025 では大きな変更はせず、昨年の方針を踏襲しています。

WISS2025 では、「登壇発表」「国際学会招待発表」「デモ発表」「WISS Challenge」の四つのカテゴリの発表が行われます。

登壇発表には、各論文に対して 4 名の査読者 (内プライマリ査読者 1 名) を割り当てて査読を行い、採否判定会議での議論を経て、37 件の投稿に対して 24 件の論文を採択しました。できるだけ多くの論文を採択・議論するために、一部の論文は条件付き採録とし、査読者と著者による修正プロセスを経て、完成度を高めています。登壇発表の論文は、本予稿集に採録理由と併せて掲載しています。

国際学会招待発表には、11 件の投稿があり、いずれにも利害関係のない 4 名の審査員による審査を行いました。その結果、5 件の登壇発表／5 件のデモ発表が招待発表として実施されます。CHI/UIST 等のトップカンファレンス論文を中心に、研究内容はもちろん、採択までの道筋等の有用な知見も共有頂けると幸いです。

デモ発表については、近年締切より大幅に前に発表枠が埋まり対応に追われる状況が続いており、昨年途中から導入した方針を最初から示すことで、混

乱を未然に防ぐように配慮しました。具体的には、募集時の発表枠の制限を撤廃して、(1) デモ発表の申し込み件数に応じてデモ会場を増設する、(2) 増設で対応できない程多くの投稿があった場合のみ、追加審査を行い採否を決定する、という方針です。今年も (関係者の尽力により) 会場の増設のみで対応でき、124 件のデモ発表を採択しました。これと先の 5 件の招待発表に加えて、19 件の登壇デモ (登壇発表される論文のデモ発表) や、7 件のスポンサー展示を含めて、3 日間で 155 件の発表が行われます。

WISS Challenge は、学会での議論や交流等を活性化するシステムを募集するユニークなカテゴリであり、今年は 4 件の論文 (企画) が採択され、学会中に運用されます。

さらに、「WISS から広がる道」と題した特別講演を行います。過去に WISS で登壇発表され、HCI 関連分野で博士号を取りつつ、現在は多様な分野で活躍される 4 名の講演者 (小松 弘幸氏、小室 真紀氏、門村 亜珠沙氏、神原 啓介氏) をお招きして、過去の WISS での発表内容や現在の仕事、さらにインタラクティブ分野の研究経験がどのように「現在」に繋がっているか等をお話し頂きます。こうした発表・議論を通して、特に学生の皆さんに、WISS から広がる未来の可能性を考えるきっかけとなればと思います。

WISS2025 は、20 年以上振りの北海道開催となり、交通の便等を踏まえて昨年より小規模になると予想していたのですが、結果的には参加者数／デモ発表件数共に昨年を上回る規模となり、コミュニティの拡大を肌で感じています。私事となりますが、私が初めて北海道を訪れたのは、前回の北海道開催となる WISS2002@函館大沼であり、その後縁あって、はこだて未来大の教員となりました。実は教員になる以前はこの一度しか北海道に来たことがなく、WISS が繋いだ縁といえるかもしれません。この WISS2025 での体験／出会いが、多くの人の未来に価値あるものになればと思います。

最後に、登壇発表をはじめとした全ての発表にご投稿頂いた皆様、スポンサー企業の皆様、協賛学会関係者の皆様、WISS2025 のプログラム委員／運営委員の皆様に深く感謝いたします。

* 公立はこだて未来大学、WISS2025 プログラム委員長

WISS 2025 プログラム委員長

塚田 浩二（はこだて未来大）

WISS 2025 実行委員長

松村 耕平（立命館大）

WISS 2025 副実行委員長

佐藤 俊樹（北陸先端大）

WISS 2025 運営委員

会計	丸山 一貴（明星大） 橋本 直（明治大） 松村 耕平（立命館大）
出版・ WISS Design	石井 綾郁（明治大） 越後 宏紀（明治大） 高木 健（東大）
広報・ 学会リエゾン	辻田 眸（無所属）
WISS Challenge	門本 淳一郎（東大）
会場	佐藤 俊樹（北陸先端大）
投稿システム	三浦 元喜（千葉工大）
参加者管理・ 交流	西田 健志（神戸大）
動画生中継	高田 峻介（神戸高専） 藤田 和之（東北大）
デモ・ポスター	高橋 治輝（立命館大） 渡邊 拓貴（はこだて未来大） 雨坂 宇宙（阪大）
ネットワーク	奥本 隼（TwoGate） 米田 英正（Helpfeel）
Web・ サーバー	加藤 淳（産総研） 湯村 翼（北海道情報大）
投票	高橋 亮（東大）
懇親会	五十嵐 悠紀（お茶大） 吉田 博則（はこだて未来大） 山田 渉（NTT ドコモ）
表彰	三上 浩司（東京工科大） 松浦 昭洋（東京電機大）
企業 スポンサー	秋山 耀（チームラボ） 清木 昌（ほぼ日）
学生 ボランティア	宮藤 詩緒（東京科学大） 大西 鮎美（神戸大）

表紙デザイン

巻田 直樹（東大）

WISS 2025 プログラム委員

秋山 耀（チームラボ）
雨坂 宇宙（阪大）
五十嵐 悠紀（お茶大）
石井 綾郁（明治大）
越後 宏紀（明治大）
大西 鮎美（神戸大）
奥本 隼（TwoGate）
加藤 淳（産総研）
門本 淳一郎（東大）
後藤 真孝（産総研）
米田 英正（Helpfeel）
坂本 大介（北大）
佐藤 俊樹（北陸先端大）
志築 文太郎（筑波大）
清木 昌（ほぼ日）
高木 健（東大）
高田 峻介（神戸高専）
高橋 治輝（立命館大）
高橋 亮（東大）
塚田 浩二（はこだて未来大）
辻田 眸（無所属）
寺田 努（神戸大）
西田 健志（神戸大）
橋本 直（明治大）
藤田 和之（東北大）
松浦 昭洋（東京電機大）
松村 耕平（立命館大）
丸山 一貴（明星大）
三上 浩司（東京工科大）
宮藤 詩緒（東京科学大）
山田 渉（NTT ドコモ）
湯村 翼（北海道情報大）
吉田 博則（はこだて未来大）
渡邊 恵太（明治大）
渡邊 拓貴（はこだて未来大）

ISS 研究会連絡先

日本ソフトウェア科学会
「インタラクティブシステムとソフトウェア研究会」
主査: 塚田 浩二
公立はこだて未来大学
〒041-8655 北海道函館市亀田中野町 116 番地 2
E-mail: tsuka@mobiquitous.com

本研究会に関する情報は <https://www.wiss.org/>
からアクセス可能です。

もくじ

まえがき: WISS2025 開催にあたって

塚田 浩二 *i*

セッション 1: 創作

クリエイティブコーディングにおけるユーザの内面的利益を考慮したプログラミング環境

石井 行, 加藤 淳, 福里 司 1

試行錯誤や創造性が求められるクリエイティブコーディングにカスタマイズされた生成 AI チャットを組み込んだプログラミング環境を実装した研究である。システム設計の妥当性や新規性については評価された一方で、論文の記述や図の視認性について問題が多く、内面的利益を向上させるという研究目的に対する有用性を読み取りづらいなど改善の必要性が大きかったことから、条件付き採録と判定された。

VL-SensorIDE: VLM を用いた仮想センサ IDE によるインタラクション設計の対象拡張と容易化

戸田 竜星, 宮下 芳明 8

本研究は、VLM による動画像内容理解を型付きデータフローの出力を持つ「センサ」として一般化し、これを用いたプログラミングを支援する統合開発環境 VL-SensorIDE を提案するものである。既存研究 Gensor と比べ、値の型が多様化し、コードエディタも統合したプログラミング体験を実現している点、多様な実例を報告している点で異なる。実世界入出力で課題となるデバッグ支援が弱く技術的新規性が乏しいなど懸念はあるものの、WISS での活発な議論が期待されることから条件付き採録と判定された。

漫画制作者のためのネーム制作支援システムの開発

北野 和紀, 五十嵐 健夫 15

本論文は漫画制作工程の分析に基づき、シナリオ分割と吹き出し配置を支援する新規性あるシステムを提案している。対応した制作工程が限定されており、コマ割り工程への対応や実用に際しては質の向上が望まれる。有用性については課題が残るものの、LLM を有効活用した新規性のある手法である点を評価し、条件付き採録と判定された。

OpticDesigner: 多様な光学レンズ効果をカスタマイズ可能な 3D プリントレンズフィルターデザインシステム

吉本 健義, 大友 千宙, 宮下 芳明 22

本論文では、ユーザ自身が求める光学効果をデザインし、3D プリンタとスピンコーティングを用いて実用的な光学品質をもつレンズフィルタを製作できるシステム LightMorph が提案されている。個人レベルでレンズフィルタの光学効果を自在に設計・実現できるという発想は一定の新規性があり、また実際の写真・映像制作の事例を伴うワークショップを通じて有用性を示している点は評価できる。論文全体の構成や記述も明瞭であり、完成度は全体的に高い。一方で、一部論文中に客観性に欠ける表現や、デザインツールの操作方法や操作可能なパラメータ等を含む詳細の記述や設計意図の記述が不足している点などの指摘もあった。以上の理由から、条件付き採録と判断された。

セッション2: ディスプレイ

波紋表示による球面ディスプレイにおける視野外ポインティングの可視化

竹内 大空, Jana Hoffard, Hsueh Han Wu, Kelvin Cheng, 宮藤 詩緒, 小池 英樹..... 29

本研究では、球面ディスプレイ上での波紋表示によるポインティング位置推定を扱っている。先行研究から続いているものであり、今回の論文のみでの新規性は限定的だが、有用性と実装・評価について一定評価されている。定義や位置づけが不明瞭であることや、将来的な目的への議論が不足している点などの懸念はあるものの、WISSで議論することで研究の発展が期待される。以上の理由から、条件付き採録と判断された。

高速投影と時間方向の色分解による視線移動方向に応じた映像提示手法

宮崎 竜輔, 宮藤 詩緒, 小池 英樹..... 36

本研究は、高速プロジェクタを用いて、視線の移動方向によって異なる画像が知覚される映像提示手法を提案している。既存の視線インタラクションや多視点提示の研究と比べ、原理や効果に新規性が認められる。また、提案手法について実装とパラメータ調整の実験が行われており、その内容に正確性と有用性が認められる。応用例も大変興味深く、複数人利用や公共空間での利用の可能性も含め、将来の発展が期待される。一方で、記述内容の一貫性の問題や、原理や制約に関して追記が望まれる箇所、図や文章の誤りや不明瞭点などが指摘されている。以上の理由から、条件付き採録と判断された。

Dishplay: シリアルの磁気応答を利用した可食ディスプレイシステム

大津 陽翔, 山本 匠, 杉浦 裕太..... 44

鉄分添加されたシリアルおよび可動磁石マトリクスを用いて、液体に浮かぶシリアル粉末をコントロールできる、動的かつ可食であるユニークなディスプレイ技術の研究である。有用性では議論があるもののアイデアの面白さは査読者全員の認めるところであり、ディスプレイとしての性能等を追記することを条件とし、条件付き採録と判定した。

ヒトに偏光知覚能力を付与するウェアラブルシステム

大城 亨太, 大西 鮎美, 寺田 努, 塚本 昌彦..... 52

本論文は、偏光フィルタレンズを回転させて人間に偏光を知覚させる新しい人間拡張を提案している。着想の新規性は認められるが、偏光知覚の拡張ではなく既知現象の再確認にとどまる点が懸念される。また、偏光カメラを用いない光学的手法の必然性や実装に関する議論も不足している。有用性についても知見の新規性や議論の深さが不十分であると指摘された。ただし、着想自体はWISSで議論する意義がある。以上の理由から、条件付き採録と判断された。

セッション3: フィールド

ペンさく: 描いて探すことで深まるペンギン観察手法と水族館来訪者を対象とした実証実験

中川 由貴, 中村 聡史, 芦刈 治将, 板東 恵理子, 與倉 陵太, 渡邊 果南, 岩永 七海..... 59

ペンギンの腹部模様を描くことでペンギンの個体を検索する手法については既発表であるが、水族館の現場での課題・ニーズを丁寧に検討して設計しているシステムであり、アプリケーションの完成度も高い。また、実環境である水族館で実証実験を行っている点、システムの有用性・他施設への導入可能性を議論した点が評価できる。知見が多く、WISSで議論する価値がある研究であると判断した。以上の理由から、採録と判定された。

GPStroke: 複数の歩行軌跡に基づく GPS アート制作手法の提案

尾本 悦基, 塚田 浩二 67

日常生活の歩行ルートを GPS で記録・収集し、集めたルートの地図上の軌跡形状をストロークとして組み合わせたアート作品を制作できるシステムを提案した研究である。GPS アートを手軽に取り組めるようにする発想の新規性はポジティブに評価された一方で、歩行促進・アート制作の2つの狙いがうまく整理して記述されていないこと、どちらの狙いに対しても大きな効果が見られていない点がネガティブに評価された結果、条件付き採録と判定された。

Purikura-Fab: 3D プリンタ未経験者の造形プロセス理解と利用意欲を促進する体験型システムの開発

白井 義人, 高橋 治輝, 松村 耕平 73

3D プリンティング未経験者に対し魅力的な 3D プリント体験を与えることを目指した研究である。Web アプリを通して所望の写真にプリクラのような加工を行い、これを送信することで所望のリソフエンを造形できる。体験全体としては新規性があり、実装の正確性は評価できる。一方、既存の 3D プリント支援手法との比較不足や有用性が限定的であるという指摘、技術的項目の記述不足といった問題点が指摘された。よって、これらの修正を条件とした条件付き採録と判断された。

セッション 4: 探索

ユーザの能動的な探索を可能にする Deep Research システム

関口 祐豊, 樋口 啓太 80

本研究は Deep Research にタイル型 UI によって従来の一方向的な調査プロセスを緩和し、人間と AI エージェントが協動的に調査を進められる新しい枠組みを提示しており、コミュニティへの貢献は大きい。32 名のユーザー実験で本手法の網羅性や効率性の向上が確認された。以上の理由から、再録と判定された。

Paper Gallery: 論文要約スライドをコミュニティの資産に

牧野 倫太郎, 高橋 治輝, 松村 耕平 88

研究コミュニティに導入が容易で、論文要約スライドの視覚的価値を活かして、コミュニティ内の繋がりや議論を誘発できるという三つのデザイン指針を満たすシステム「Paper Gallery」を提案し、ゼミでの 10 か月の運用で得た実証的知見を報告している。既存 Wiki システムを拡張して実現され、差分が見えにくい点が懸念だが、論文要約スライド活用を多角的に試行した点が新しく、研究活動支援のありかたについて WISS で活発な議論が期待されるため、条件付き採録と判定された。

PrefTolerance: 歌声・伴奏・歌詞ごとの嗜好と許容度を考慮できる楽曲推薦インタフェース

中澤 一輝, 佃 洸樹, 中野 倫靖, 後藤 真孝 95

本研究は、楽曲推薦システムにおいて、歌声・伴奏・歌詞という三つの要素それぞれに対して個別に「許容度」を設定できる新たな仕組みを提案したものである。システムの実装は丁寧で完成度が高く、論文全体も明快にまとめられており、レビューは好印象であった。一方で、「許容度」と従来の「重要度」との概念的な違いが十分に説明されていない点や、システムの有効性を示す評価が限定的である点が指摘された。それを踏まえた委員会での議論の結果、WISS における議論を促す意義ある研究として評価され、採択に至った。

セッション5: XR

片眼のみに表示されるポインタを用いた xR 向け選択手法

鈴木 悠真, 折居 篤, 志築 文太郎..... 103

本論文は両眼立体視を行う環境でオブジェクトを選択する際に、ポインタを片目の視界のみに表示することでオブジェクトの選択時間とエラー率が改善されることを示している。新規性や有用性が認められ、WISS 登壇発表での議論に相応しい研究である。一方で、論文の記述については一部に不十分な箇所があるため、各査読者からのコメントを確認して反映されたい。以上の理由から、条件付き採録と判断された。

効率的な遠距離オブジェクトの回転操作を可能とするトラックボール型 VR コントローラ

岩井 望, 阿部 優樹, 坂本 大介..... 110

VR コントローラにトラックボールを組み込み、特に遠距離オブジェクトの回転操作における効率性と直感性を向上させる新しいインタラクション手法を提案する研究である。プロトタイプ実装と比較実験により定量・主観の両面から有効性を検証している。その結果、パフォーマンスと主観的評価に優れる可能性が示されている。近距離の回転操作では従来手法に劣る可能性も示されているものの、新規性と将来展望の点で高く評価できる。以上の理由から、条件付き採録と判定された。

VR 環境におけるレイキャストによるオブジェクトの選択成功率推定ツールの提案

奥野 達也, 清水 春翔, 笠原 暢仁, 本間 大一優, 山中 祥太, 宮下 芳明..... 118

査読者一同、VR 空間でのポインティングの成功率を推定する数値モデルは有用であり、将来性がある研究であることを認めている。一方で、今回モデル構築のために実施された実験から得られたデータは、本研究で提案している VR 環境での UI 構築時のオブジェクト成功率の推定に利用するには限定的なデータであり、実験計画を含めて再検討の余地がある。大規模なデータによるモデル構築を含めて、今後の発展が期待される。以上の理由から、条件付き採録と判定された。

座位での物理的な制約の緩和を目指した知覚されない閾値に基づいたリダイレクション手法の提案と評価

小林 広夢, 鈴木 湧登, 日下部 完, 坂本 大介..... 126

本研究は、座位環境での物理的な制約を緩和するために、頭部と胸部の回旋を同時に知覚閾値内で増幅するリダイレクション手法を提案し、その有効性を検証したものである。回旋ゲインを頭部と胸部へ同時適用することの新規性に加え、物理的制約の緩和と VR 体験の質の保持を両立するための設計指針を提供していることの実用性が高く評価された。一方で、一部主張のオーバークレイムや、一部評価指標の位置づけの不明瞭さが見られたことから、条件付き採録と判定された。

セッション6: 支援

PracToneVis: 管楽器奏者のための LLM を用いた基礎練習支援システム

池谷 優希, 伊藤 貴之..... 134

本研究では、管楽器のロングトーン練習を対象に、音量・音程・音色の3点を解析して可視化し、LLMで自然言語フィードバックするシステムを提案した。倍音構造に基づく音色クラスタリングを LLM と組み合わせる自然言語フィードバックするという点、アプリとして実装できている点、10名の奏者を対象にしたユーザ調査により評価されている点が評価された。一方で、先行研究との差分が明示的ではない点、評価方法や音色クラスタリングの妥当性に関する懸念が示された。以上の理由から、条件付き採録と判定された。

後方伸身宙返りの段階的学習をサポートする練習支援システム

高橋 洸汰, 岡部 誠 142

本論文は、器械体操の床種目における後方伸身宙返りを対象に、段階的学習支援と言語フィードバックを統合した練習支援システムを提案しており、類似度マップによる段階的学習支援機能に関する新規性、システムの実装の質も高いという点が評価された。一方で、システムの中核の機能で新規性があるという意見が多かった段階的学習支援の部分の記述不足と、提案システムの支援対象者が不明瞭である点が指摘された。以上の理由から、条件付き採録と判断された。

Graph Guide: 低視力者支援のためのセマンティック Focus+Context グラフ表示

SECHAYK Yotam, LI Yuan, RAVE Hennes, COLLEY Mark, SHAMIR Ariel, 五十嵐 健夫 . 150

本研究は、弱視者がグラフをよみとりやすくするために、グラフのカーソル付近を拡大表示しつつ、そのすぐ周囲に文脈的要素（例：軸、凡例）を常に表示する、いわば focus+context 表示手法を提案している。本研究の進め方は堅実であり、まず弱視者 5 名が参加した予備実験から要件を抽出した上で、この要件に基づき提案手法を設計している。また、弱視者 6 名が参加した 3 条件（None、Minimap、提案手法）を比較するユーザテストを通じて、提案手法がグラフの読み取りに役立つことや手法に関する方向性が報告されており、この点でも本研究の貢献は大きい。以上の理由から、本論文は採録と判断された。

セッション 7: 入力

KanaShark: SHARK2 に基づく日本語ジェスチャ入力手法の検討

豊田 真吾, 日下部 完, 坂本 大介 158

本論文は、スマートウォッチにおける日本語入力手法として、12 キー配列の子音を一筆書きでなぞる入力方式 KanaShark を提案し、その有効性を検証している。提案手法を含む 3 手法を 9 名を対象とした、10 日間の実験で比較しており、入力速度やエラー率において提案手法の優位性を示している。また NASA-RTLX や SUS スコアなども評価しており、実験結果やアイデア自体の有用性は査読者らによって高く評価されており、採録と判定した。

溝間隔の違いによる筆圧変化を活用したシート埋め込み型 ID 認識手法

瀬崎 夕陽, 関口 祐豊, 中村 聡史 166

本技術は溝パターンによる筆圧変化の読み取りに新規性があり、論文の質も高いと評価した。一方で、提案技術の有効性が既存操作との比較で不明瞭な点、および認識精度や個人差への対応など技術的な課題も残る。以上の理由から、条件付き採択と判断された。

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 173

本論文では、VR 環境でのオブジェクトの移動を目的として、頭部や視線による操作と発声による操作を組み合わせた手法を提案している。頭部・視線操作と発声操作の組み合わせは新規であり、詳細な比較実験の結果に基づいて、先行手法（発声のみ）よりも優れた選択手法であることを示した。一方で、この手法が具体的な VR アプリケーションで、どのように使用されるかについて、十分に議論されていないという指摘があった。委員による議論の結果、採録の条件ではないものの、具体的な使用例を示しながら手法の有用性を説明することが強く推奨された。以上の理由から、採録と判断された。

クリエイティブコーディングにおけるユーザの内的利益を考慮したプログラミング環境

石井 行* 加藤 淳† 福里 司*

概要. クリエイティブコーディングに生成 AI 技術を活用することで、ユーザの生産性や効率性の向上が見込める。しかし、既存手法の多くはプログラムを自動生成したり自動修正したりするため、ユーザの試行錯誤や内省の機会が失われてしまう可能性がある。そこで本研究では、ユーザの内的利益（内在的な能力、自己認識、自己内省、感情的なウェルビーイング）を維持し、その向上を支援することを目的として、生成 AI ベースの対話的インタラクションを活用する手法を提案する。具体的には、ユーザの制作途中のプログラムを踏まえつつ (1) ユーザの思考と内省を促す質問を行う対話式のチャット AI 及び、(2) 対話ログに基づく改善案を画面に提示するビジュアル機能をコードエディタに統合する。本システムにより、ユーザの認知的な負荷を無理なく増加させ、ユーザの思考を促すことで内的利益を向上させる効果が期待できる。

1 はじめに

人間の創造的活動を促進・補助する技術は、ヒューマン・コンピュータ・インタラクション分野における重要な研究テーマである。その中でも特に、プログラミングによって画像・アニメーションなどを生成する「クリエイティブコーディング」の活動を支援するために、生成 AI 技術を導入する試みが世界的に注目されている。しかし既存研究の大半（例：Spellburst [1]）は、生成 AI によって出力されたプログラムコードをそのまま使用する、またはユーザが書いたコードを自動修正するものである。このアプローチは、生産性や効率性を向上させることはできるものの、ユーザの自己内省の機会を奪ってしまい、生成 AI への過度な依存を引き起こす危険性が存在することが、Wang ら [18] によって指摘されている。Cox ら [16] は、創造性支援ツールの評価方法がアウトプットの生産性や品質に偏重している点に加え、創造的活動を通じて得られた個人の心理的・内的な成長こそが、人間が創造性を発揮する本質的な価値となることを論じている。また、制作における作業効率の向上は、必ずしもユーザの満足度や学習効果に繋がらない。これらの指摘は、学習プロセスにおける試行錯誤を促すための「意図的な摩擦」を重視する教育者の知見と一致するものである [14]。以上のことから、生産性や効率性だけでなく、ユーザの内面的影響（例：内在的な能力、自己効力感、自己内省、感情的なウェルビーイング）を評価することが重要であると考えられる。

そこで本稿では、クリエイティブコーディング活動におけるユーザの内在的な能力、自己効力感、自

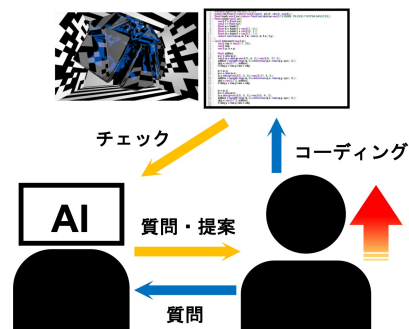


図 1. 本システムの概念図。AI は直接的な「答え」ではなく、「質問」を提示することで、ユーザの思考と内省を促す。

己内省、感情的なウェルビーイングを維持し、その向上を支援するために生成 AI を活用するインタラクション手法及び、これを実装したプログラミング環境を提案する（図 1）。具体的には、従来のようにプログラムコードを自動生成・修正するのではなく、(1) ユーザのコードを参照して思考や内省を促す質問を投げかけるチャット機能と、(2) 対話の文脈に基づいた改善案（修正案）を視覚的に確認できるビジュアル機能を、コードエディタに組み込んだシステムを実現した。本システムの有用性を示すために、クリエイティブコーディング経験者による本システムを用いた作品を紹介する。また、現状の課題や解決策を示し、将来への展望について述べる。

2 関連研究

2.1 クリエイティブコーディング環境

クリエイティブコーディングを支援するための技術は、これまで多数提案されている。例えば加藤ら [10] は、音楽に合わせタイミンよく歌詞が動く

Copyright is held by the author(s).

* 早稲田大学

† 産業技術総合研究所

インタラクティブな視覚表現を「リリックアプリ」と名付け、その開発支援フレームワークを提案した。また, Chugh ら [5] は, 画面上の出力結果 (例: 図形) を直接編集し, その結果をプログラムコードに反映させるシステムを考案した。生成 AI 技術を用いた支援システムの例としては, Spellburst が挙げられる [1]。このツールでは, 生成 AI 技術による出力結果をノードベースで管理・提示することでアーティストの生産性向上に成功したとされている。しかし, これらの例すべてにおいて, 各技術の採用がユーザーの内面に与える影響は議論されていない。ユーザー自身の内面的な利益を評価している創造性支援ツールは, Cox らによる評価手法の包括的レビュー [16] によれば, 全体の 15% に留まることが報告されている。更に, McNutt ら [15] は, 「クリエイティブコーディングにおける芸術的表現」と「学校教育におけるプログラミング教育」の二つの側面から, プログラミングツールが重視する要素について, 学生を対象とした調査を通じて議論している。この調査では, 生産性を高めるツールは, ユーザーの深い理解を促すという教育上の目標とは相反しうることが示唆されている。これらを踏まえ, 本研究では, 生成 AI に基づくプログラムコードの自動生成ではなく, ユーザーが生成 AI と対話しながらプログラミング作業を行うことで, 試行錯誤しながら芸術的表現を模索できるシステムを提案する。

2.2 直接的な答え提示による弊害について

生成 AI 技術を用いた「答え」の自動提供は, ユーザーの AI への過度な依存や, 批判的思考 (ある考えについて前提となる事実を明らかにしながら, 多角的・論理的に考えること) の能力自体を低下させてしまう危険性がある [19]。更に, 生成される結果 (答え) は, あくまでも AI が生成できるデザイン範囲に限定されてしまうため, デザイナーの想像力を制限してしまう可能性がある。そこで Danry ら [6] は, 批判的思考の能力を向上させるために, ユーザーに「答え」ではなく「質問」を投げかける手法を提案している。また, AI による認知負荷は「問題を解決する能力」の向上にも寄与することが指摘されている [17]。この知見を受けて本研究では, ユーザーの批判的思考力や問題解決能力といった内在的能力を向上させるために, 「質問」ベースの対話型 AI 機能を構築する。更に, 候補提示型 UI [9][12] を参考に, 本研究では, ユーザー自身のコードと対話ログの文脈情報に基づき, ビジュアル案を複数提示する機能も検討する。

2.3 創造的な活動における内省について

クリエイティブプロセスにおける「内省」は重要な要素である一方, その概念の曖昧さから創造性支援に関する研究ではあまり言及されてこなかった [3]。そこで Ford ら [7] は, 内省を「現在の制作プロセス



図 2. 提案ユーザインターフェースのスクリーンショット:
(a) コードエディタ, (b) 描画パネル, (c) チャットパネル, (d) ビジュアル候補提示パネル。

への内省」「自己への内省」「過去の経験への内省」「実験を通じた内省」の 4 種類に分類し, 評価尺度を提案している。この分類に沿うと, 芸術的表現と計算論的思考を要するクリエイティブコーディング活動では, 自己への内省 (芸術的表現) とプロセスへの内省 (コーディング) の両方が重要となる領域である。換言すれば, 「表現したいイメージ」と「それをどうコードに落とし込むか」を両立させるための支援が必要と考えられる。そこで本研究では, 従来研究で取り組まれてきたコーディング支援だけでなく, 芸術的表現の面にも焦点を当て, とくに「自己への内省」を促す対話方法を検討する。

3 ユーザインターフェース

本システムは, クロスプラットフォームアプリケーションフレームワークである Qt (バージョン 6.8.3) を用いて, シェーダ言語 OpenGL Shading Language (GLSL) 向けのクリエイティブコーディング環境として実装している。対話式のチャット AI には, Google の提供する Gemini 2.5 flash の API を用いた。開発したユーザインターフェースを図 2 に示す。

画面構成としては, 標準的なクリエイティブコーディングの環境 (コードエディタ, 描画結果をプレビューする画面) に加え, AI との対話を行うための

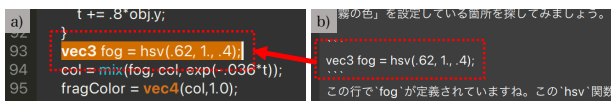


図 3. AI の返答によるプログラムコードのハイライト機能. (a) コードエディタ, (b) チャットパネル.

チャットパネルが常に表示されている. ビジュアル提案機能については, ユーザが必要な時に呼び出すことができる.

3.1 コードエディタ

従来のコードエディタと同様に, ユーザは GLSL コードを入力したり編集したりできる (図 2(a)). ユーザが入力したプログラムコードは, 文字単位でコンパイルされ, 描画パネルに結果が表示される. コードにエラーがある場合, エラーメッセージが表示される. このように, 明示的なコンパイル操作不要のライブプログラミング [2][11] を実現している (図 2(b)).

3.2 チャットパネル

ユーザは, 制作途中の作品について, 対話式のチャット AI と会話することができる (図 2(c)). AI への入力には, ユーザのチャット入力とエディタ上の GLSL コードに加え, 過去の対話履歴が含まれる. さらに, AI の振る舞いを規定するために, 以下のような役割とルールを定義したシステムプロンプトを付与している.

- **目的:** ユーザの自己内省や技術的試行, 概念の探求などを促す短い質問を返す.
- **役割:** クリエイティブコーディングにおける「思考/表現のパートナー」として振る舞う.
- **ルール:** 新たなコードスニペットを提示せず, あくまでユーザ自身の思考を促すことに徹する.

このプロンプト設計により, AI は単に答えを提示するのではなく, ユーザの内省を支援するパートナーとしての対話を実現する. また, AI 側がエディタ上のプログラムコードに言及した際, 該当する範囲をハイライト (橙色) する機能を有している (図 3). ハイライト機能を用いることで, チャットパネルとエディタパネルを確認する際の外的認知負荷を減らすことができる.

3.3 ビジュアル提案機能

チャットボット右部のボタンをクリックすると, 現在のコードを基にした改善案が三つ生成され, その出力結果のビジュアルがチャットパネルの下部に表示される (図 2(d)). これらの改善案を生成するために, AI に対して「現状のコードに対して, 対話ログから要約したユーザの意図を反映した完成形のコー

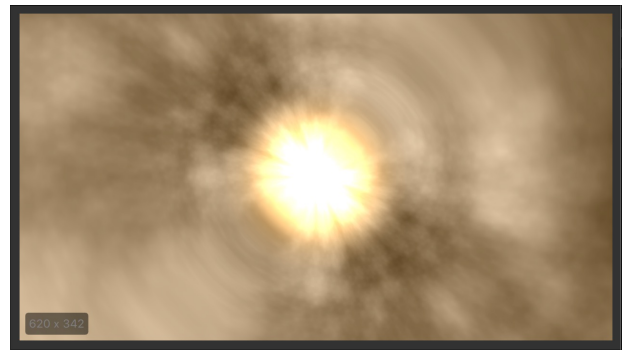


図 4. アイデア出しの段階から本システムを用いたアニメーション作品の例 (作品テーマ: 幼少期の心象風景). 制作時間: 60 分, AI への質問回数: 7 回, ビジュアル提案機能の使用回数: 1 回.

ドを, 3つのバリエーションで生成してください」というシステムプロンプトを与えている. このプロンプトによって, AI は大規模言語モデル (LLM) が有する応答の確率的な性質を基に多様な提案を行う. ユーザが表示されたビジュアル候補の中から一つの案を選択すると, コードの差分を解説する. 解説は, 「ユーザの現在のコードと選択された候補のコードとの技術的な差分を分析し, どのような変更が加えられたのかを自然言語で簡潔に説明してください」というシステムプロンプトにより生成される. 但し, エディタパネル上のプログラムコードを自動的に更新する機能はなく, ユーザ自身が更新する必要がある. これは, ユーザの主体性を担保しつつ, AI とユーザのビジュアルを介した共通理解を確立することを目的としている.

4 本システムを用いた制作事例

本章では, 本システムを用いた制作事例を 2 つ, その制作過程に着目して紹介する. なお, 本稿で示す各事例の制作者 (第一著者) は, 2 年程度の GLSL によるコーディング経験と, p5.js や Processing を用いたクリエイティブコーディング経験を有する. 更に C, C++, Python, JavaScript を用いたツール及びゲーム制作経験を有する.

4.1 事例 1: アイデア出しからの制作

図 4 は, アイデア出しの段階から本システムを用いて制作したアニメーション作品 (作品テーマ: 幼少期の心象風景)¹である. 本事例を通して, 制作の初期段階から作品を完成させるまでのプロセスを示す.

4.1.1 初期段階におけるアイデア出し

制作したい作品や目的が十分に定まっていない場合において, AI との対話機能を「アイデア出し」

¹ <https://www.shadertoy.com/view/WcBcz1>



図 5. AI による質問の一例（どんな感情や記憶、あるいは過去の体験が、今「何かを作りたい」という気持ちにつながっていますか？）。

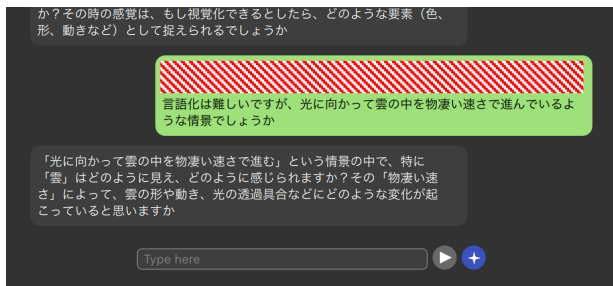


図 6. ユーザの過去や記憶を基にアイデア出しを行った場合（赤枠：プライバシーにかかわる内容）。

ツールとして利用することができる。但し、本システムは、ユーザの内省を促すことを目的としているため、AI は作品テーマを直接指示するのではなく、あくまでもユーザの内省を促すような「質問」を提示する。本事例では、AI は「どんな感情や記憶、あるいは過去の体験が、今『何かを作りたい』という気持ちにつながっていますか?」といった質問を提示した（図 5）。このようなユーザの過去の経験や記憶に基づく内省を促す際には、プライバシーや心的外傷（トラウマ）に関わる情報に触れる可能性がある点に留意する必要がある（図 6）。

また、ユーザが AI のテキストベースの返答だけで作品の具体像を描けない場合は、ビジュアル提案機能を用いることで新たな着想を得ることもできる。本事例では、図 7(a) のような選択肢が提示された。そして、選択肢をクリックすると、「hsv」関数と「mix」関数を組み合わせることで、fbm_val の値に応じて色相が 0.08（赤みがかったオレンジ）から 0.16（オレンジがかった黄色）の間で滑らかに変化する、暖色系のグラデーションが生成されるようになっています。」という解説が提示された（図 7(b)）。コードそのものが自動で書き換わることはないため、ユーザは解説をヒントに、表示された解説を参考にしながらコーディングを始めることができる。

4.1.2 ユーザと AI による試行錯誤

目的がある程度明確になったら、通常のコーディング作業と同様、エディタパネルにてコーディング作業を行う。ユーザが入力したプログラムコードは



図 7. ビジュアル提案機能の利用例。(a) ビジュアル候補, (b) ユーザのコードと選択肢のコードの差分の解説。

リアルタイムにコンパイルされ、描画パネルに結果が表示されるため、ユーザは試行錯誤しながら作品を制作できる。コーディングの途中で不明点が生じた場合、ユーザはチャット機能を通じて AI に質問を投げかけることができる。本事例では「theta が 2π から 0 に変わる境界に線が入ってしまいます。滑らかに繋げるにはどうすればいいですか?」というユーザの具体的な質問に対して、AI が「theta の値が 0 から 2π の間を巡るという『周期性』は、どのような図形に例えることができるでしょうか? その図形上の連続的な位置を、fbm の 2 次元入力として表現できるとしたら、どのようにすれば『離散的な切り替わり』を避けて、なめらかに繋がれると思いますか?」という返答を提示した（図 8）。この返答を基に、ユーザは不明点の解決を行うことができる。このように、AI が問題解決に至る思考を促す質問を返答することで、ユーザは不明点を解決するための思考を促され、自己効力感の向上が期待できる。

本作品のアイデア出しの段階から、ユーザが実際にコーディングを開始するまでに約 20 分、その間に AI からは 3 回の質問が提示された。本システムを用いることで、ユーザは従来のコーディング作業と比べ、自己への内省を反映した作品を制作することができ、満足感が得られた。最終的に、本作品の制作時間は 60 分程度、AI への質問は 7 回、ビジュアル提案機能の使用は 1 回であった。

4.2 事例 2：既存作品の改善

本節では、既存作品の改善において本システムを活用した事例を紹介する。図 9 は、ユーザが作成した途中段階のプログラムコード（作品テーマ：Struggling Structure）のパラメータ調整に本システムを用いたものである。途中段階の作品（図 9(a)）では、コードの各所にアニメーションを制御するパラメータが偏在する。つまり制作者は美的感覚に基づいてパラメータ調整をしなければならず、作品テーマを十分

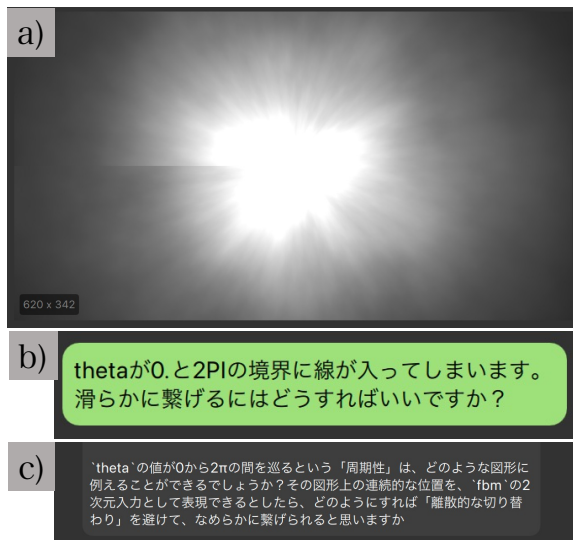


図 8. ユーザと AI による試行錯誤の様子. (a) 制作途中の描画パネル, (b) ユーザによる質問, (c) AI による返答.

に表現できていない状態にあった.

本システムを用いることで, 4.1.1 と同様にユーザーの内省に関する対話ログが作成される. 本事例でユーザーは「もがき苦しむような不安定な動き」をイメージしていると AI に伝えた. この対話ログを基に, AI は具体的なコード箇所 $p.xy *= \text{rot2}(\dots)$ を指摘し, 『不安定さ』や『もがき苦しむ』ような感覚をこの回転に加えるには, この $\cos$ 関数の引数や係数にどのような調整を加えることが考えられるでしょうか?』といった質問を提示した. これにより, ユーザーはアニメーションの動きを制御するパラメータの調整を意図的に行うことができた. このように, 本システムはユーザーの内省に基づいた試行錯誤を補助する.

本作品の改善 (図 9(b)) に 20 分程度を要した. AI への質問は 5 回, ビジュアル提案機能の使用は 2 回であった.

5 議論

本章では, システムを実用した制作事例を通して得られた知見をまとめ (5.1 節), クリエイティブコーディングにおいてユーザーの内面的利益を考慮したシステムを設計する上で重要と考えられるデザイン指針 (5.2 および 5.3 節) や, 今後必要とされる研究について議論する (5.4 節).

5.1 クリエイティブコーディングにおけるユーザーの内面的利益に関する考察

関連研究でも触れたように, クリエイティブコーディングにおいては「表現したいイメージ」と「そ

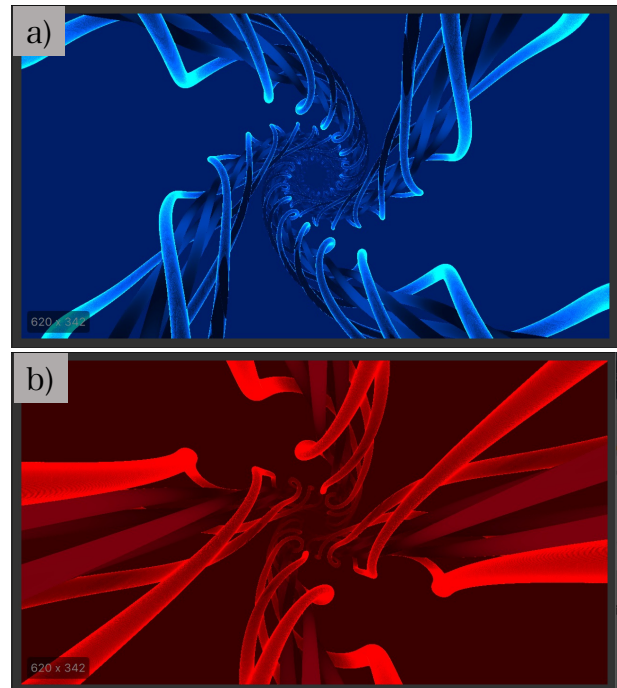


図 9. 本システムを用いたアニメーション作品の例 (作品テーマ: Struggling Structure). (a) ユーザによる制作結果 (途中段階), (b) 本システムを用いて改善した結果. 制作時間: 20 分, AI への質問回数: 5 回, ビジュアル提案機能の使用回数: 2 回.

れをどうコードに落とし込むか」の両立が課題である. 「表現したいイメージ」だけを深く考えてもコーディングが難しくなる一方, 「コーディング」だけ考えても表現の幅が狭まってしまう.

本システムを使った際 (特にアイデア出しの段階) には, チャット機能が, コードによる実現性とは無関係に内省を促す問いを提示する. そのため, 第一著者による制作事例においては, 通常のコーディング時よりも深く内省を行った実感があった (例: 幼少期の心象風景). その後, AI はイメージをどうコードに落とし込むかを考えるためのサポートとなる問いを提示した (例 1: 「もし視覚化できるとしたら, どのような要素 (色, 形, 動きなど) として捉えられるでしょうか?」という抽象的な問いや, 例 2: 「theta の値が 0 から 2π の間を巡る ...」という具体的な問い). 第一著者の普段行うクリエイティブコーディングでは, 自身のコーディング技術によって, どのような表現が可能かを探求するケースが多かった. その一方, 本システムで自身の通常の発想の外にある問いを新たに提示してもらったことで, 従来であれば選択肢に入らないような実装方法を発見できた. このように, 本システムにより, 先行する「表現したいイメージ」と「それをどうコードに

落とし込むか」というギャップを埋める過程に取り組みやすく、また、コーディングに関する学びが得られたようにも感じられた。

一方で、クリエイティブコーディングでは、アイデア出しが終わった後、単線的なコーディング作業だけではない点に注意が必要である。例えば、「表現したいイメージ」が明確な時でも、多くの場合、完成像を写真のように思い浮かべているわけではない。そのため、コーディング作業を通して予想外の新しい効果を発見し、表現の幅が広がるというセレンディピティを体験することがある。これも、クリエイティブコーディングの醍醐味である。しかし、内省した内容を直接プロンプトとして生成 AI を用いた場合、たとえイメージに近い結果が出力されたとしても、そこからユーザ自身が試行錯誤できる範囲が限られてしまう上に、作品への所有感が薄れてしまう危険性がある。本システムの場合は、自分で理解している感覚を持ちつつコーディングを行うことができたため、通常のコーディングと同じ水準で試行錯誤できたという実感が得られる。

このように、本システムは、生成 AI を単なる自動化の手段としてではなく、ユーザの内省を支え、ユーザ自身のコーディングに関する理解を助けるツールとして提供することで、ユーザの主体性を犠牲にせず創作過程の試行錯誤を支援できる可能性を示したと言える。

5.2 ユーザの内省支援に関する示唆

- **対話履歴の視覚化による内省支援機能**：AI との対話履歴をタイムラインやマップとして視覚化し、ユーザが自身の思考プロセスを客観的に振り返り、メタ認知を促す機能を実装する。
- **対話文脈への復帰を支援する UI 設計**：本システムを利用する際、ユーザが AI との対話をしている途中でコーディング作業を再開する様子が繰り返し観察された。そうした場合には、対話の文脈が飛躍してしまうケースがある。そこで、過去の対話コンテキストへ円滑に復帰できるように、未解決の対話スレッドを管理するボタンなど、AI との対話にスムーズに戻れる UI 機能を構築する。
- **外在的認知負荷を軽減する補助機能**：試行錯誤の過程で生まれる複数のプログラムコードのバージョンを管理・比較する機能をはじめ、コーディング作業以外の負荷（外在的認知負荷）を軽減するための補助機能を実装する。

5.3 AI との対話設計に関する示唆

- **自己効力感を維持するための解答提示機能**：本システムは、ユーザ自身の思考によって問題を解決すること（自己効力感の向上に繋げること）

を目的としているため、AI からは直接的な答えを提示せず、常に「質問」を投げかけるような形式となっている。しかし、いつまでもユーザが自力では解決できない状況において、答えを提示しないことが「学習意欲の低下」を招いてしまう危険性がある。そこで、一定時間思考しても解決に至らない場合に、ヒントや解答を段階的に提示する機能を設計する。

- **AI による能動的な介入機能**：ユーザのコーディング作業の停滞を検知し、ユーザからのテキスト入力なしに AI から質問や提案を行う機能を実装する。
- **ビジュアルアーティファクトへの対応拡張**：logomotion [13] のように、制作した画面結果を直接認識し、それに基づいた改善案やサポートを行う機能を実装する。

5.4 今後の展望

- **長期利用におけるユーザ行動の分析**：本システムは、ユーザの試行錯誤と内省を支援する機能を有しているが、実際にどのような観点で内面的利益がどの程度向上するのかについては、第一著者による制作事例を通して定性的に検証したのみである。今後、ユーザ実験を通して、各機能が自己効力感の向上や内省の増加へどのように寄与しうるか、より詳細に、Creativity Support Index [4] など併用しながら長期的に評価する必要がある。

なお、各機能はユーザの思考スキル向上を促すことを目的としているため、ユーザが本システムを将来的に使用しなくなることは、ユーザの思考スキルが十分に発達し、本システム自体への依存から脱却したことを意味し、必ずしもネガティブな結果とはいえない。そこで、長期的評価においては、ユーザのスキル、モチベーション、そして利用継続・離脱の動向を分析し、思考支援 AI とユーザの理想的な関係性を探求することが必要不可欠であろう。

- **内省を分析するツールとしての展開**：近年、学習や創造の過程における内省を評価するために質問票ベースの研究 [8] が多数発表されている。本システムは、ユーザと AI との対話を通じて内省の過程がログとして詳細に記録できるため、クリエイティブコーディングにおける学習者の内省プロセスを解明するための分析ツールとしてシステム展開が期待される。
- **多言語への対応**：現在は GLSL 言語に限定されているが、今後は Processing や、その JavaScript 版である p5.js にも対応する予定である。

謝辞

本研究の一部は、JSPS 科研費 JP24K20910 及び JST ACT-X JPMJAX22A3 の支援を受けたものである。

参考文献

- [1] T. Angert, M. Suzara, J. Han, C. Pondoc, and H. Subramonyam. Spellburst: A Node-based Interface for Exploratory Creative Coding with Natural Language Prompts. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST)*, pp. 100:1–100:22, New York, NY, USA, 2023. ACM.
- [2] S. Burckhardt, M. Fahndrich, P. de Halleux, S. McDirmid, M. Moskal, N. Tillmann, and J. Kato. It's Alive! Continuous Feedback in UI Programming. In *Proceedings of the 34th ACM SIGPLAN Conference on Programming Language Design and Implementation (PLDI)*, pp. 95–104, New York, NY, USA, 2013. ACM.
- [3] L. Candy. *The Creative Reflective Practitioner: Research Through Making and Practice*. Routledge, 2019.
- [4] E. Cherry and C. Latulipe. Quantifying the Creativity Support of Digital Tools through the Creativity Support Index. *ACM Trans. Comput.-Hum. Interact.*, 21(4), June 2014.
- [5] R. Chugh, B. Hempel, M. Spradlin, and J. Albers. Programmatic and Direct Manipulation, Together at Last. In *Proceedings of the 37th ACM SIGPLAN Conference on Programming Language Design and Implementation (PLDI)*, pp. 129–144, New York, NY, USA, 2016. ACM.
- [6] V. Danry, P. Pataranutaporn, Y. Mao, and P. Maes. Don't Just Tell Me, Ask Me: AI Systems that Intelligently Frame Explanations as Questions Improve Human Logical Discernment Accuracy over Causal AI explanations. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI)*, pp. 352:1–352:13, New York, NY, USA, 2023. ACM.
- [7] C. Ford and N. Bryan-Kinns. Towards a Reflection in Creative Experience Questionnaire. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI)*, pp. 277:1–277:16, New York, NY, USA, 2023. ACM.
- [8] C. Ford, A. Noel-Hirst, S. Cardinale, J. Loth, P. Sarmiento, E. Wilson, L. Wolstanholme, K. Worrall, and N. Bryan-Kinns. Reflection Across AI-based Music Composition. In *Proceedings of the 16th Conference on Creativity & Cognition (C&C)*, pp. 398–412, New York, NY, USA, 2024. ACM.
- [9] T. Igarashi and J. F. Hughes. A Suggestive Interface for 3D Drawing. In *Proceedings of the 14th Annual ACM Symposium on User Interface Software and Technology (UIST)*, pp. 173–181, New York, NY, USA, 2001. ACM.
- [10] J. Kato and M. Goto. Lyric App Framework: A Web-based Framework for Developing Interactive Lyric-driven Musical Applications. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI)*, pp. 124:1–124:18, New York, NY, USA, 2023. ACM.
- [11] J. Kato, T. Nakano, and M. Goto. TextAlive: Integrated Design Environment for Kinetic Typography. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI '15*, pp. 3403–3412, New York, NY, USA, 2015. ACM.
- [12] Y. Koyama and M. Goto. BO as assistant: Using Bayesian Optimization for Asynchronously Generating Design Suggestions. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology (UIST)*, pp. 77:1–77:14, New York, NY, USA, 2022. ACM.
- [13] V. Liu, R. H. Kazi, L.-Y. Wei, M. Fisher, T. Langlois, S. Walker, and L. Chilton. LogoMotion: Visually-Grounded Code Synthesis for Creating and Editing Animation. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI)*, pp. 157:1–157:16, New York, NY, USA, 2025. ACM.
- [14] A. M. McNutt, S. Cohen, and R. Chugh. Slowness, Politics, and Joy: Values That Guide Technology Choices in Creative Coding Classrooms. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI)*, pp. 54:1–54:16, New York, NY, USA, 2025. ACM.
- [15] A. M. McNutt, A. Outkine, and R. Chugh. A Study of Editor Features in a Creative Coding Classroom. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI)*, pp. 363:1–363:42, New York, NY, USA, 2023. ACM.
- [16] S. Rhys Cox, H. Bøjer Djernæs, and N. van Berkel. Beyond Productivity: Rethinking the Impact of Creativity Support Tools. In *Proceedings of the 2025 Conference on Creativity and Cognition (C&C)*, pp. 735–749, New York, NY, USA, 2025. ACM.
- [17] J. Sweller, J. J. G. van Merriënboer, and F. G. W. C. Paas. Cognitive Architecture and Instructional Design. *Educational Psychology Review*, 10(3):251–296, 1998.
- [18] A. Wang, Z. Yin, Y. Hu, Y. Mao, and P. Hui. Exploring the Potential of Large Language Models in Artistic Creation: Collaboration and Reflection on Creative Programming. *arXiv preprint arXiv:2402.09750*, 2024.
- [19] A. Woodruff, R. Shelby, P. G. Kelley, S. Rousso-Schindler, J. Smith-Loud, and L. Wilcox. How Knowledge Workers Think Generative AI Will (Not) Transform Their Industries. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI)*, pp. 641:1–641:26, New York, NY, USA, 2024. ACM.

VL-SensorIDE: VLM を用いた仮想センサ IDE による インタラクション設計の対象拡張と容易化

戸田 壱星* 宮下 芳明*

概要. 本研究は, VLM をユーザがプロンプトで再定義できる仮想センサとして活用し, そのままアプリコードへ配線できる IDE「VL-SensorIDE」を提案する. ユーザは, 測定値のフィードバックを見ながら「食事の進行度を 0-1 で返す」といった記述を試行錯誤しながら修正し, 新たなセンサを構築できる. これにより, 物理センサの設置が困難な対象や, 従来の認識モデルでは実装コストが高かった主観的基準・抽象的事象の測定が可能となる. 8 名の評価では 20 分で 53 件の多様なセンサが作成された. 本研究の成果は, VLM を仮想センサとしてコーディング環境に直結させることで, センシングの対象を拡張し, その設計と実装の自由度を高める点にある.

1 はじめに

インタラクション設計におけるセンシングは, 何をどの粒度で測るかという意味的な要件と, どのように測るかという実装手段を適切に結び付けることが要点である. しかし実際の開発では, 複数の障壁が存在する. まず, 多くのシステムは固定ラベルや固定閾値を前提としており, 導入後も含めて個人の主観や用途に合わせて測定基準を柔軟にカスタマイズするのが難しい. また, 思い出のぬいぐるみのような侵襲が許されない対象や, センサの取り付けが負担となる生物, 液体・影など取り付け自体が不可能な対象では, 物理センサを能動的に取り付けることができない. さらに, 空間の雰囲気, ユーザの作業負荷のような構成要素が不明瞭な抽象概念を測ろうとすると, 要素の洗い出し, 個別センサやコンピュータビジョン (CV) モデルの調達, データ統合という長い開発ループを踏む必要があり, 試行錯誤を阻んできた.

近年の Vision-Language Model (VLM) は, 複雑な視覚入力に対して自然言語で応答できる. この性質を利用し, 本研究では VLM をプロンプトの書き換えによって測定基準を再定義できる仮想センサとして活用する. 以後, VLM を用いた仮想センサを Vision-Language-based Sensor (VLS) と呼び, 視覚入力から数値・真偽・文字列の値を返すものとして定義する. VLS は従来は要素分解や個別実装が前提だった主観的・複合的概念も同一の枠組みで扱える. 加えて, 非接触であるため, 直接取り付けが難しい対象でも測定対象にできる.

この着想に基づき, VLS の設計から実プログラムでの活用までの手順を簡略化し, 試行錯誤のサイクルを高速化するために, IDE として VL-SensorIDE

を設計・実装した. VL-SensorIDE は, (1) VLM に渡す映像ソースの選択, (2) テーブルで VLS の設計, (3) 任意の VLS をドラッグ&ドロップでコードへ変数として配線, (4) プログラム実行時, 配線された VLS の測定値を変数へ代入, の手順を IDE 内で完結させる.

本稿の貢献は, 測定基準を自然言語で定義できること, 出力を数値・真偽・文字列として同じ枠で扱えること, 作成した VLS を IDE のコードに直接配線できる点にある. これにより, インタラクション設計で扱えるセンシング範囲を広げ, その実装を容易にする. さらに, 本アプローチの適用可能性を実証するためユーザスタディを実施し, 20 分で 53 件の多様な仮想センサが構築されることを確認した. この結果から, VL-SensorIDE が, センシングの適用範囲を拡張できる有効なアプローチであることが示唆された.

2 関連研究

VLM 登場以前のセンシング支援は, 主にアルゴリズムの実装と分析の効率化に主眼が置かれていた. Eyepatch [11] や Exemplar [4] は GUI での例示により分類器を対話的に構築したが, 出力は主に分類ラベルや検出領域であり, 基準変更には再学習を要した. プログラマ向けの IDE も探求され, VisionSketch は GUI によるデータフロー構築を [5], Gestalt は機械学習の実装と分析の統合を [12], DejaVu はカメラベースのプログラムの記録・再生機能を提供した [6]. これらの IDE はアルゴリズムの実装やデバッグを支援したが, 処理ロジックの定義はコード記述などに依存していた. 特に実世界由来の非再現的なデータはデバッグが困難であり, Exemplar や DejaVu は時系列データの可視化や記録再生機能でこの課題に取り組んだ. エンドユーザ向

Copyright is held by the author(s).

* 明治大学

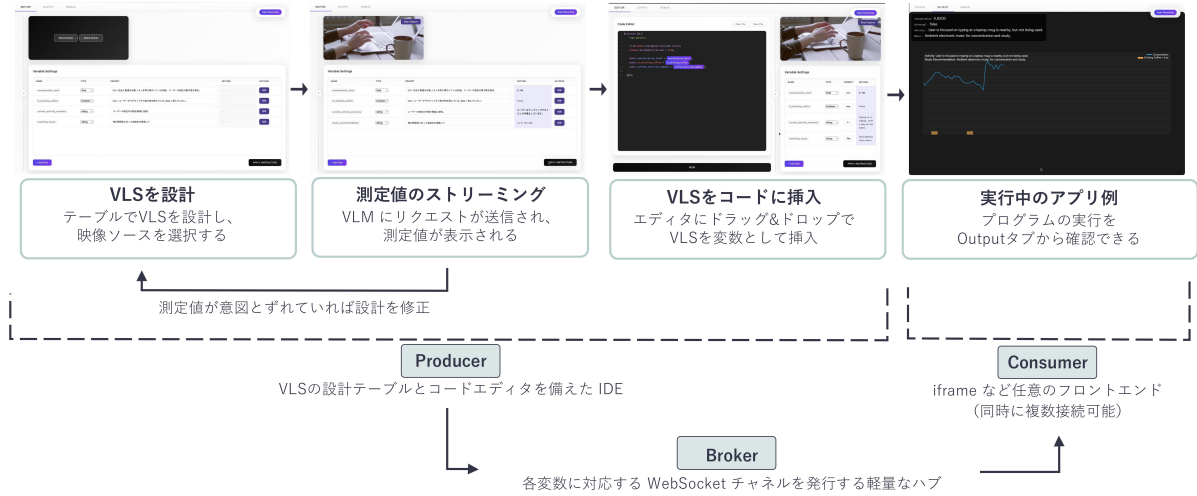


図 1. VL-SensorIDE のシステム構成と処理フロー。

けには Teachable Machine が GUI での学習から JavaScript をエクスポートするワークフローを普及させたが、明示的なデータ収集と学習が前提であり、単一プロンプトでの定義には対応しておらず、また出力も主に分類ラベルのため連続値などを扱う設計ではなかった [2].

VLM の登場は、実装中心のアプローチから、より宣言的なセンシング定義を可能にした。この潮流に先立ち、Synthetic Sensors は複数の環境センサ信号から高位の状態、例えば電子レンジの完了、を後付けの仮想センサとして抽出可能にした [8]. これにより、従来のスマートカメラの硬直性に対して、後からセンサの定義を追加できる柔軟さを示した。Zensors は自然言語の質問に対しクラウドワーカーが人力で即時応答するハイブリッドな手法を提示したが、常時運用には人件費が課題として残った [7]. この人力による解釈という役割を VLM は部分的に置き換える可能性を持つ。Gensors は、「子どもが家具によじ登ったら」のような任意プロンプトを真偽値で判定したが、外部アプリ連携は手作業であった [10]. GPT-4V をスマートホーム自動化へ流用する事例でも、文脈理解の利点とともに実装や運用上の摩擦が報告されている [14]. 一方で、VLM を連続値の計測に使う試みも進んでいる。CrowdCLIP は教師無しで群衆人数を数値として推定し、NumCLIP は数値の順序の理解を強化して画像から数値を得る利用範囲を広げた [9, 3]. ただし、いずれもタスク定義は固定で、運用中に測定対象や基準を自由に差し替える前提ではなかった。

以上を踏まえ、本研究は VLM を、プロンプトの書き換えで再定義できる仮想センサとして扱う。そしてこの仮想センサを IDE に直結し、開発フローの工数を抑える運用設計を提案する。本稿の貢献は、測定基準を自然言語で定義できること、出力を数値・

真偽・文字列として同一の枠で扱えること、作成した VLS を IDE のコードに直接配線できる点にある。これにより、短周期の更新で基準の記述・出力確認・修正のループを容易に回せる。導入後も、基準変更や出力型の切り替え・組み合わせをコードと一貫した環境で試せる。

3 提案システム

本研究は、VLM を用いた VLS の定義からアプリケーションへの接続までを同一の IDE 内で完結できる IDE 「VL-SensorIDE」を提案する。図 1 に示すように、本システムは (1) VLS の定義, (2) 測定値の確認と修正, (3) コードへの挿入, (4) アプリの実行という一連の流れを同一の IDE 内で完結できる。本章では、まずシステムの UI と基本的な操作フローを説明し、次いでそれを支えるアーキテクチャと実行サイクルの詳細を述べる。

3.1 UI と操作フロー

本システムの UI と操作フローを図 2 に示す。ユーザはまず、図 2(d) で入力映像ソースを選択する。次に、図 2(e) のテーブルで名前、型、プロンプトを宣言し、「Apply Instruction」ボタンを押すことで VLS の定義をシステムに反映させる。定義が反映されると、システムは 1.0s 周期でその VLS の測定を開始し、結果を同テーブルの Return 列に表示する。プロンプトを編集して再度ボタンを押すと、次の更新周期から新しい定義が反映されるため、ユーザは Return 列の値の変化を見ながら対話的に測定基準を試行錯誤できる。また、図 2(f) を押下することで図 3 のようにエディタを折りたたみ、VLS の定義に特化したビューで作業することも可能である。

定義した VLS は、図 2 に示すように、変数テー

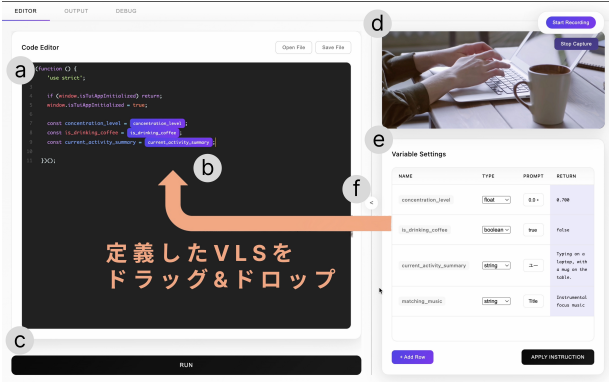


図 2. VL-SensorIDE の画面構成. a: エディタ, b: 変数ブロック, c: Run, d: 映像ソース, e: 変数テーブル, f: エディタ切替.

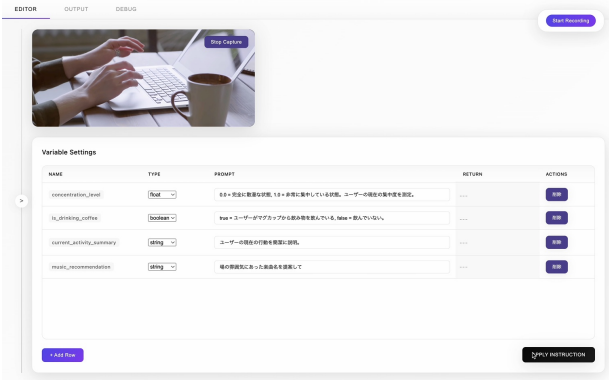


図 3. エディタ折りたたみ時の変数定義ビュー.

ブルからエディタへドラッグ&ドロップすることでコードに挿入できる. 挿入された VLS は、エディタ内で図 2(b) のようなブロックとして表示される. 最後に、図 2(c) の「Run」ボタンを押すと、これらの VLS が配線されたアプリケーションが Output タブ内で起動する.

3.2 アーキテクチャ

本システムは、役割を分けた Producer-Broker-Consumer の三層構成を採用している. Producer は、ユーザが操作する変数宣言 UI とコードエディタを備えた IDE 本体であり、入力映像の取得と VLM への推論要求の送出を担う. Broker は、Python で実装された WebSocket サーバとして機能し、Producer から受け取った測定値を送信する. そして Consumer は、その値を受け取り描画や反応を行う Web アプリケーションであり、複数クライアントの同時接続をサポートする. この三層構成によって、システムの拡張や再利用が行いやすくなっている.

3.3 実行サイクルと通信

システムのストリーミング処理は、映像キャプチャ、推論、送信、受け取りの順で構成される. このサイクルは 1.0s 周期で実行される. 前の処理が周期内に完了しない場合、そのサイクルは棄却され、常に最新のフレームが次周期の評価対象となることで、システムの応答性を維持している.

入力フレームの推論には、Gemini-2.0-Flash-Live を採用した. このモデルは永続的な WebSocket セッションをサポートしており、一度の接続で続けて推定を行えるため、リクエストごとの HTTP/TLS ハンドシェイクを省ける. また、セッション継続により前回値との整合を保ちやすい. アプリケーション実行では、初回ロード時にコード内のプレースホルダ変数をその時点の値で展開し、以降の更新は新しい測定値のみを差し替える軽量更新で行う. この構成により、編集時の把握しやすさと実行時の更新速度を両立する.

4 評価

4.1 手順

本システムの評価のため、プログラミング経験を持つ学部生 8 名（男性 7・女性 1, 平均 20.5 歳）を対象にユーザスタディを実施した.

評価実験は三つのフェーズで構成した. 実験環境として、参加者は全員が同一のハードウェアとブラウザを使用した. まず導入（約 3 分）で VLS をコードへ挿入する操作も含めた IDE の基本操作を実演・習熟させた後、探索・発想フェーズ（20 分）に移った. このフェーズで参加者は、任意の映像ソースを対象に自由に仮想センサを定義・修正し、発想した変数と応用アイデアを記録した. なお、本評価は仮想センサの作成と、それをどのようなプログラムで用いるかの回答までを対象とし、コーディングは評価範囲に含めない. これは、実装の成否や所要時間が個々の技能に依存し、本研究の検証対象を VLS の定義と活用構想に置いているためである. この記録から発想成果を評価した後に、事後フェーズ（約 15 分）で半構造インタビューを実施し質的知見を得るとともに、日本語版 SUS [1, 13] を回収してユーザビリティを測定した.

4.2 定量的結果

表 1 に示すように、20 分のタスクで合計 53 件（平均 6.6）の仮想センサが生成された. 型は float が 66% と最多で、boolean と string が続いた. SUS は平均 61.4 と、初期版 IDE として許容域に達していることを示す.

表 1. 定量的結果

項目	結果
生成センサ総数	53 (平均 6.6/人, SD 1.8)
型内訳	float 35 (66%)/boolean 8 (15%)/string 10 (19%)
SUS (n = 8)	平均 61.4, SD 8.8

表 2. ユーザが作成した仮想センサの分類

クラス	例	件数
A. 主観・情動系	かわいさ, 盛り上がり, 論理一貫性, 幸福度	16
B. 環境/生活モニタ	明るさ, 机の汚さ, 風力, 混雑度	10
C. TUI/物体操作	コップ水量, 水筒保持, 目線座標, お絵描き支援	10
D. 自己管理・ヘルスケア	集中度, 体調, 食事健康度	7
E. eSports/開発支援	FPS スキル, カメラワーク, CPU 負荷	6
F. 生成・創作系	ストーリー生成, レシピ提案	4

4.3 仮想センサの作例

参加者が作成したセンサは、表 2 に示すように多様なカテゴリに及んだ。主観・情動系が全体の 30% を占めた一方で、62% は環境センシングや自己状態の定量化など、既存ハードセンサの代替・補完を狙う実利指向であった。さらに 8% は「ストーリー」や「レシピ」など、文字列型の出力を試みる例で構成された。

例えば、主観・情動系のカテゴリでは、「キモ可愛さ」やイベントの「盛り上がり度合い」といった、個人の主観に基づき測定基準を柔軟にカスタマイズする必要がある対象が多数試された。ある参加者 (P8) は、Return 列のライブ値を見ながらプロンプトを対話的に修正することで自身の感覚に近い出力を得られたと述べた。こうした主観的指標が容易に得られることから、「作品の格付けチェック」や「共同作業の活性度を可視化する」といった応用アイデアが構想された。

TUI/物体操作のカテゴリでは、「コップ内の水量」や「ティッシュが使用されたか」を測定する VLS が作成された。これらは、物理センサの取り付けが困難な液体や、軽量かつ使い捨てであるため従来のセンサでの検知が難しい対象に対し、非接触でセンシングを行える本システムの特徴を示している。参加者

からは、これらの VLS を Tangible User Interface (TUI) における入力手法として活用するという応用が提案された。

生成・創作系のカテゴリでは、「卓上の物体を登場人物とした物語」や「映像に映っている食材からのレシピ提案」といった文字列型の VLS が試された。前者の例では、物体が動いたり新しい物体が登場したりすると、その状況変化を反映した物語が生成されていき、子供向けのインタラクティブなストーリーテリングへの応用が考案された。後者は、料理動画と冷蔵庫の中身を組み合わせた献立提案アプリへの応用が想定された。これらは、数値データを返す従来のセンサとは異なり、視覚的文脈の解釈そのものをセンサ値として直接テキストで出力する応用例である。これらの事例は、VL-SensorIDE が、物理量の測定から主観的な事柄の定量化、さらには文脈的なテキスト生成まで、広範な対象を同一の枠組みで扱えることがうかがえた。

4.4 質的分析

インタビューの質的分析から、主に以下のテーマが浮かび上がった。参加者からは、プロンプトによって測定基準を柔軟に定義できるため、かわいさや論理一貫性といった従来は定量化が困難であった主観的概念も計測対象にできた点が指摘された。また、ハードウェアを導入せずに仮説検証が行えることから、本格的な実装前の試行段階で有用なプロトタイプングツールとしての価値も確認された。一方で、測定基準のテンプレートや変数名補助といったプロンプト設計支援機能の必要性が示唆された。さらに、VLM にプロンプトを提案してほしいといった、より対話的な設計支援への要望も複数聞かれた。加えて、医療や安全に関わる高リスクなドメインにおいては、VLM による単独判定に依存することへの懸念も示された。

作成された仮想センサの有効性については、測定対象の性質によって傾向が見られた。解釈の幅を許容できる主観的な概念、例えばキモ可愛さやストーリー生成では、自身の感性と合っていたなど肯定的な評価が得られた。また、人の数やティッシュの使用有無のような要素が単純な判定も概ね安定して機能した。これらのケースでは、出力が短時間で更新されるため、参加者は自身の意図と結果を迅速に照合できる点を評価していた。これに対し、fps の上手さや面白さのように、複数の要素を統合して判断する高次な概念をプロンプトで定義した場合は、意図と出力の間に乖離が顕在化した。しかし複数の参加者は、これはシステムの限界ではなく、プロンプトに具体的な観点を示すことで改善できると考えていた。

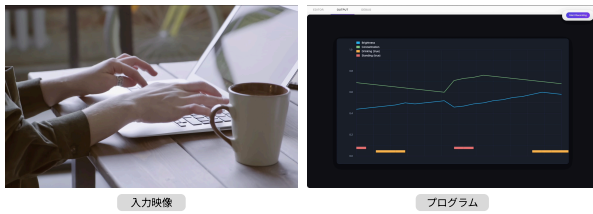


図 4. デスクワークモニタリングの実行画面。

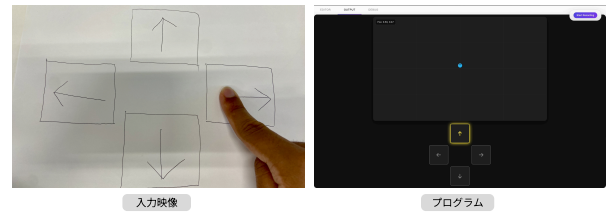


図 6. 紙を用いた即席 TUI の実行画面。

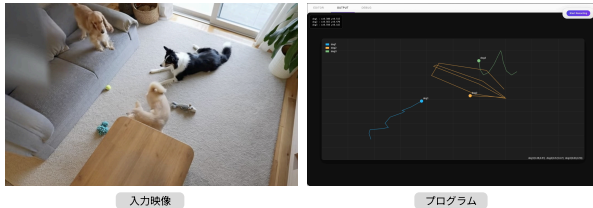


図 5. 複数オブジェクト追跡の実行画面。

5 応用事例

本章では、VL-SensorIDE の具体的な活用例として、性質の異なる 3 つのユースケースを詳述する。各事例では、その目的と実装プロセスを述べ、得られる考察を示す。

一つ目に、スマートオフィスなどにおいて、従来は複数の物理デバイスを要した環境・ユーザ状態の把握を、単一の映像ソースで低コストに実現する応用例を構築した (図 4)。ウェブカメラ映像のみを入力とし、「brightness (float)」、「concentration_level (float)」、「is_drinking (boolean)」といった、性質の異なる複数の仮想センサを同時に定義する。これにより、追加ハードウェアなしで作業環境とユーザの状態を統合的にモニタリングできる。本事例は、単一の非接触センサから、物理量、ユーザの内的な状態、断続的な行動といった、複数の異なるレイヤーの情報を同時に抽出できることを示している。

二つ目に、センサ装着が困難または負担となる対象に対して、非接触で行動追跡を行う応用例を構築した (図 5)。本アプローチの特徴は、CV 手法のように事前に定義された検出対象や基準に縛られない点にある。CV モデルは学習済みの特定カテゴリしか認識できないが、本システムではプロンプトを「青い首輪の犬を追跡して」のように修正するだけで、マーカーや特定の種に依存することなく、その場で柔軟に対象を再定義できる。

最後に、専用ハードウェアを用いることなく、TUI を即席に構築する応用例を示す (図 6)。紙に描いた十字キーへの指の接触を検出するため、「is_up_arrow_touched (boolean)」や「is_down_arrow_touched (boolean)」といった仮想センサ群を定義した。これらの出力をキー入力にマッピングする

ことで、ウェブカメラが紙に描かれた十字キーを機能的なコントローラーとして認識し、操作入力へ変換する。特別な機材を必要とせず、身近な素材でインタラクティブな装置をその場で具現化できる手軽さは、物理デバイスの形態を多様にする可能性がある。

6 議論

6.1 設計判断とトレードオフ

本研究は、専用学習器や追加ハードウェアではなく、VLM へのプロンプト定義だけで視覚入力から数値・真偽・文字列の値を得る設計を採った。この方針により、導入後でもプロンプトの再定義だけで測定基準を切り直せる。ユーザスタディでは、参加者がプロンプトを修正しつつ自らの要求に合う測定基準へ収束していく様子が観察され、この再定義の容易さが試行錯誤のサイクルを加速させていることがうかがえた。また、VLS をコードへ直接配線できるため、抽象概念に対しても長い開発ループを経ずに試行を開始できる。

一方で、非接触で測定が可能であるため、センサの装着自体が困難な対象でも、まずは測定の当たりを付けることができる。ただし、本方式は視覚手掛かりに限定され、視野外・遮蔽・外観が似た対象に弱く、VLM の出力揺らぎへの対処も不可欠である。さらに、曖昧な基準では値が不安定になりやすく、高リスク領域で単独判定に依存することには慎重さが求められる。また、本システムの VLS 測定は使用した VLM の推論処理時間の制約に基づいて 1.0 s 周期で実行している。そのため、視線移動や高速ゲームプレイのような短時間の事象は取りこぼしうる明確な制約がある。

また、VLM API を直接利用する場合は連続実行や複数プロンプト管理、コード反映といった負担が大きい。本手法は IDE がこれらを隠蔽することで、開発者がプロンプトの試行錯誤やプログラムの作成に集中できる点に利点がある。その反面、本 IDE の設計は、API リクエストのタイミングやエラーハンドリングといった低レイヤーの自由度を制約する側面もある。

以上より、本設計は初期の実装負担を最小化し、基準探索と適用開始までの時間を大幅に短縮する利

点をもたらす一方で、視覚情報への依存、出力の安定性、IDE 設計に起因する適用範囲・信頼性といった制約がある。

6.2 VLS の妥当性

4.4 節の所見から、妥当性は測定対象の性質に強く左右される。人数や有無、簡単な程度のように要素が単純な事象は多くの場面で安定して扱えた。一方で、厳密な精度が求められる場面では、VLM の性能や視覚条件に起因する出力の不確実性が課題となる。対照的に、感情や雰囲気といった主観的・抽象的な事象は、プロンプトを整えていくことで出力を意図に近づけやすい。プロンプトについては、測定するにあたっての観点や条件などを与えるといった一定の労力を必要とする。しかし、全てのユーザが即座に的確なプロンプトを記述できるわけではない。加えて、毎回プロンプトを手で詰めるのは負担が大きい。ユーザが効果的なプロンプトを効率的に見つけ出すためのデバッグ機能や補助ツールが重要である。この課題は、6.4 節で述べる VLS の設計補助機能の重要性へと直結する。また、VLS が実務でどの程度の水準を満たすかを示すための定量的な性能確認は今後の課題とする。

6.3 文字列出力の位置づけ

本稿では、数値型のセンサを単一軸を単一変数で測る一次元の測定と捉える。これに対し文字列型のセンサは、明るさ・色味・にぎやかさ・落ち着きのような複数の意味軸を同時に含む多次元の測定として位置づける。この前提のもとで、数値と文字列の関係を圧縮と抽出という変換で整理できる可能性を示す。圧縮は、複数の数値を一つの文字列にまとめることを指す。これは、個別の数値を並べて表示するだけでは捉えにくい、全体的な状況や文脈を一つの意味的なまとまりとして表現する試みである。多数の数値を個別に解釈せずとも全体像を直感的に把握でき、文字列は単なるキーワード検索を超え、意味的な近さに基づく検索の手がかりとなりうる。抽出は、文字列から必要な数値だけを取り出すことを指す。例えば「西日の差す、穏やかな午後」という一つの測定値から暖色度、明るさ、落ち着きといった複数の指標を推定する。これにより、当初測定対象として想定していなかった観点、例えば「集中しやすさ」といった指標を、後から同じ文字列データを用いて推測できる可能性がある。なお、可逆性と手続きの一般化は今後の検証課題とする。

6.4 VLS 設計補助

現行プロトタイプは、変数定義 UI と値プレビュー、ログ記録を備える一方で、型推定、変数名補完、測定基準テンプレート、および対話的にプロンプトを作成できる仕組みは未実装である。ユーザス

タディではこれらへの要望が複数挙がり、基準づくりの初動を支援することが試行速度と精度安定の双方に寄与する示唆が得られた。加えて、評価参加者からは「プロンプトを改善すれば、より望んだ値が得られそう」という声が複数聞かれた。これは、本アプローチの中心的な作業が、プロンプトを対話的に洗練させるプロセスであることを示唆している。しかし現状の値のプレビューだけでは、予期せぬ値が出力された際に、どの瞬間の入力映像が原因だったのかを特定したり、過去のプロンプト修正との比較を行ったりすることが難しい。このプロンプトの洗練プロセスをさらに深化させるため、関連研究で示された記録・再生のようなアプローチに基づき、入力映像、プロンプトの変更履歴、そして出力値の時系列を統合的に可視化・再生できるタイムラインベースのデバッグ機能が有効だと考える。

6.5 限界と展望

本評価は 8 名・短時間の探索的条件にとどまり、長期運用や効果量を伴う統計的検証は今後の課題である。また Gemini-2.0-Flash-Live 以外のモデルでの検証也未実施である。さらに本システムが扱えるのは視覚的手掛かりに限られ、他モダリティを必要とする事象は測定できない。今後は、設計補助、デバッグ支援、時間分解能の向上、定量的検証、および文字列出力と数値出力との往還に関する可逆性と手続きの一般性の検証に取り組む。これにより、本システムの適用可能な領域と使用上の前提を明確化し、再現性と運用上の信頼性を段階的に高める。

7 結論

本研究は、プロンプト定義という統一されたインタフェースを通じて、従来は個別の専門技術を要した多様なセンシングタスクを、エンドユーザが専門知識なしに定義・実装できる IDE「VL-SensorIDE」を提案した。8 名による初期評価では、20 分という短時間で 53 件の仮想センサが生成され、本アプローチの即時性と実用性が示された。本研究の貢献は、測定基準を自然言語で定義できること、出力を数値・真偽・文字列として同じ枠で扱えること、作成した VLS を IDE のコードに直接配線できる点にある。これにより、従来の物理量や状態の推定に留まらず、文脈的な解釈を文字列として出力するという、新しいセンシング形態をも扱う、新たな設計思想を提示した。

参考文献

- [1] J. Brooke. SUS: A Quick and Dirty Usability Scale. In *Usability Evaluation in Industry*, pp. 189–194. Taylor & Francis, London, UK, 1996. Originally published by Digital Equipment Corporation in 1986.

- [2] M. Carney, B. Webster, I. Alvarado, K. Phillips, N. Howell, J. Griffith, J. Jongejan, A. Pitaru, and A. Chen. Teachable Machine: Approachable Web-Based Tool for Exploring Machine Learning Classification. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI EA '20, pp. 1–8, New York, NY, USA, 2020. Association for Computing Machinery.
- [3] Y. Du, Q. Zhai, W. Dai, and X. Li. Teach CLIP to Develop a Number Sense for Ordinal Regression. In *Computer Vision – ECCV 2024: 18th European Conference, Milan, Italy, Sept. 29–Oct. 4, 2024, Proceedings, Part LXXXV*, Vol. 15143 of *Lecture Notes in Computer Science*, pp. 1–17. Springer, Cham, 2024.
- [4] B. Hartmann, L. Abdulla, M. Mittal, and S. R. Klemmer. Authoring Sensor-based Interactions by Demonstration with Direct Manipulation and Pattern Recognition. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, pp. 145–154, New York, NY, USA, 2007. ACM.
- [5] J. Kato and T. Igarashi. VisionSketch: Integrated Support for Example-centric Programming of Image Processing Applications. In *Proceedings of Graphics Interface 2014*, GI '14, p. 115–122, Toronto, Ontario, Canada, 2014. Canadian Information Processing Society.
- [6] J. Kato, S. McDirmid, and X. Cao. DejaVu: Integrated Support for Developing Interactive Camera-Based Programs. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology*, UIST '12, pp. 189–196, New York, NY, USA, 2012. ACM.
- [7] G. Laput, W. S. Lasecki, J. Wiese, R. Xiao, J. P. Bigham, and C. Harrison. Sensors: Adaptive, Rapidly Deployable, Human-Intelligent Sensor Feeds. In *Proceedings of the 33rd Annual CHI Conference on Human Factors in Computing Systems*, CHI '15, pp. 1935–1944, New York, NY, USA, 2015. Association for Computing Machinery.
- [8] G. Laput, Y. Zhang, and C. Harrison. Synthetic Sensors: Towards General-Purpose Sensing. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pp. 3986–3999, New York, NY, USA, 2017. Association for Computing Machinery.
- [9] D. Liang, J. Xie, Z. Zou, X. Ye, W. Xu, and X. Bai. CrowdCLIP: Unsupervised Crowd Counting via Vision–Language Model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, CVPR '23, pp. 2893–2903, Piscataway, NJ, USA, 2023. Institute of Electrical and Electronics Engineers.
- [10] M. X. Liu, S. Petridis, V. Tsai, A. J. Fianaca, A. Olwal, M. Terry, and C. J. Cai. Sensors: Authoring Personalized Visual Sensors with Multimodal Foundation Models and Reasoning. In *Proceedings of the 30th International Conference on Intelligent User Interfaces*, IUI '25, pp. 755–770, New York, NY, USA, 2025. Association for Computing Machinery. arXiv:2501.15727.
- [11] D. Maynes-Aminzade, T. Winograd, and T. Igarashi. Eyepatch: Prototyping Camera-based Interaction through Examples. In *Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology*, UIST '07, pp. 33–42, New York, NY, USA, 2007. ACM.
- [12] K. Patel, N. Bancroft, S. M. Drucker, J. Fogarty, A. J. Ko, and J. A. Landay. Gestalt: Integrated Support for Implementation and Analysis in Machine Learning. In *Proceedings of the 23rd Annual ACM Symposium on User Interface Software and Technology*, UIST '10, pp. 37–46, New York, NY, USA, 2010. ACM.
- [13] K. Sato, N. Mitomi, K. Kon, and H. Haruna. Reliability of the System Usability Scale in the Field of Prosthetics and Orthotics. *Journal of the Japanese Academy of Prosthetists and Orthotists*, 30(1):32–37, 2022.
- [14] S. Yun and Y. kyung Lim. What If Smart Homes Could See Our Homes?: Exploring DIY Smart Home Building Experiences with VLM-Based Camera Sensors. In *CHI Conference on Human Factors in Computing Systems*, CHI '25, pp. 1–22, New York, NY, USA, 2025. Association for Computing Machinery.

漫画制作者のためのネーム制作支援システムの開発

北野 和紀* 五十嵐 健夫*

概要. 漫画生成を行う研究はこれまでに数多く行われてきたが、制作プロセスに注目したものは少ない。実際の現場では、漫画家の知識や経験を最大限に活かしつつ、漫画制作を効率化できるシステムが求められる。本研究では、漫画家を使用することを想定したネーム制作の支援システム「NameGen」を提案する。NameGen は、漫画のシナリオを入力として受け取り、各コマのラフスケッチと吹き出しを生成する。この生成は、漫画シナリオをコマごとの小さなシナリオに分割した後、各コマごとにラフスケッチを生成し、実際の漫画のコマを参照として吹き出しを配置することで行う。実際の漫画シナリオを入力とした実験では、NameGen が漫画家が作成したものと同じようなラフスケッチと吹き出しを生成できることを示した。さらに、我々がこれまでに取り組んできた他の生成手法と比較することで、NameGen の有効性を確かめた。

1 はじめに

漫画は日本の文化として長年制作され続けてきた。研究の分野でも、映画やゲームなどの映像コンテンツを漫画に変換する研究 [16, 17, 22, 24] や、入力画像を漫画風のレイアウトにして配置する研究 [8, 23] などが行われてきた。さらに、LLM の登場以降は、シナリオから漫画画像を生成する研究 [11, 25] も行われるようになった。しかし、これらの研究は実際の漫画の制作プロセスに注目していない。以下に漫画会社へのヒアリングに基づく実際の漫画制作プロセスを記載する。

1. 漫画全体のシナリオを書く
2. シナリオをコマに分ける
3. コマ割りをする
4. ラフスケッチと吹き出しを描く
5. コマ割りを調整する
6. 仕上げを行って実際の漫画を描く

なお、5 の過程では、目線を引くために重要なコマを大きくしたり、ページ数の制約を満たすために複数のコマを一つにまとめたりする。

入力として人物やシーンの画像を与える必要がある既存研究は、シナリオから描き始める上記の漫画制作過程に当てはめることができない。LLM を用いてシナリオから漫画生成を行う研究は、シナリオとは別にレイアウトを入力する必要がある研究 [25] と吹き出しの生成ができない研究 [11] のみである。本研究では、上記のうちに 2～4 の過程を対象として、漫画家の制作支援のためのネーム生成システム「NameGen」を提案する。NameGen では、漫画全体のシナリオのみを入力とし、コマごとのラフスケ

ッチと吹き出しを生成する。具体的には、LLM を用いて全体のシナリオをコマごとのシナリオに分割した後、ファインチューニングされた画像生成モデルによってコマごとのラフスケッチを生成する。その後、各ラフスケッチに対して、類似する実際の漫画コマを検索し、それをもとに吹き出しを配置する。ただし、NameGen のコマ割りは 4 コマ漫画のように正方形を縦に並べるものとした。これは、実際の漫画制作現場でも、まずは正方形のコマを縦に並べたのち、ページ数の条件やコマの中身の重要度に合わせて大きさを調整している場合があるからだ。

実際の利用においては、漫画家は自身が書いたシナリオを NameGen に入力することで各コマごとのラフスケッチと吹き出しの候補を得る。その中から気に入ったものを使用し、コマ割りをページ数に合わせて調整することで漫画の下書きであるネームを素早く作成する。

以下の章では、漫画制作に関する先行研究の紹介とそれに対する本提案の位置付けを説明する。次いで、本提案の詳細な実装と我々が行った別の生成手法について紹介する。実際の漫画をもとにしたいくつかの生成例を紹介した後、現状の課題や改善の方向性を述べて、結言とする。

2 関連研究

2.1 画像を入力とする漫画生成

画像を入力とする漫画生成は、コマ割りと吹き出しの配置を通じて行われる。なお、この節では、映像を入力とする漫画生成も画像を入力とする漫画生成として扱っている。この種の研究は、映像から漫画に使用する画像フレームを抽出しているため、抽出以降の過程は画像を入力とした漫画生成として捉えることができる。

Copyright is held by the author(s).

* 東京大学

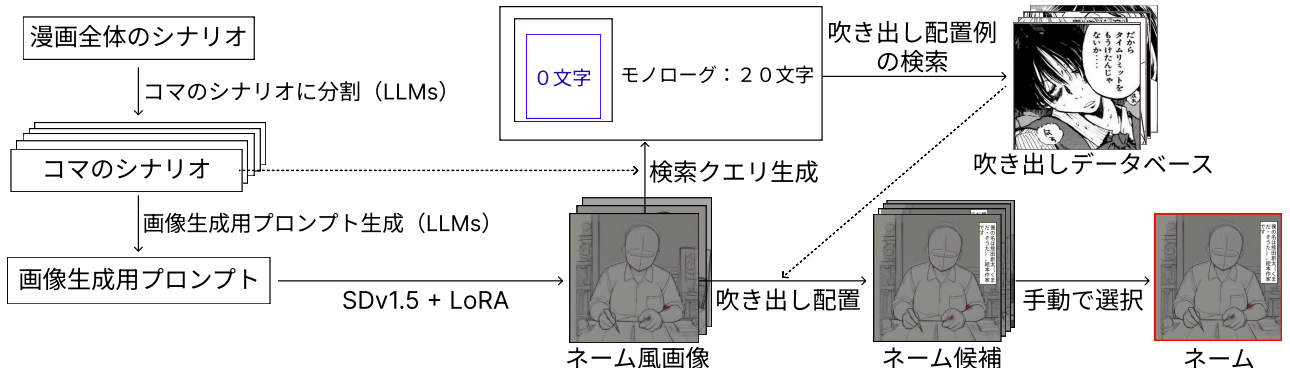


図 1. NameGen の全体像.

2.1.1 コマ割り

複数の画像から漫画風のコマ割りを生成する先行研究 [6–8, 12, 14, 16–18, 22–24] は主に、ルールやテンプレートベースのもの [6, 12, 14, 16, 18, 21–24] と、機械学習を用いるもの [8, 17] に分けられる。ルールやテンプレートベースの手法では生成されるコマ割りの種類に限りがあったため、Cao ら [8] は初めて機械学習によるコマ割りを行った。本研究は正方形を並べる簡易的なコマ割りとしたが、機能の拡張としてこれらを用いることが想定される。

2.1.2 吹き出しの配置

吹き出しの配置に関する研究はわずかしが存在しない。Comic Chat [18] はコマの上側に貪欲的に吹き出しを配置したが、キャラクターと吹き出しの位置が離れてしまった。Chun ら [13] は、吹き出し同士の距離、及び、話者と吹き出しの距離を考慮して吹き出しの配置を行った。しかし、映像の漫画化を行うこの研究は、コマに含まれるセリフの量が多く、モノログとダイアログの区別もしていないため、コマに含まれるセリフの量や種類が多様な実際の漫画の吹き出し配置とは異なる。Look Over Here [9] は、機械学習によって読者の視線移動に基づく吹き出し配置を行った。しかし、視線移動はキャラクターの位置とコマ割りから予測されるため、それらが全て揃っている状況でしか吹き出しの配置ができない。これらの研究とは対照的に、NameGen は、コマごとに実際のコマに近い多様な吹き出し配置を行うことができる。NameGen では、実際の漫画のコマで構成されるデータベースに対して、生成されたネーム風画像と似た構造を持つコマを検索し、参照することで吹き出しの配置を行う。

2.2 テキストを入力とする漫画生成

LLM の登場以降、漫画をテキストから生成する研究が行われるようになった [11, 25]。Chen ら [11] は、プレーンテキストを入力として漫画を生成する研究を初めて行った。コマ内とコマ間のトランスフォー

マーブロックの設計により、ストーリーのテキストに基づいて 1 枚の漫画ページを生成した。しかし、吹き出しの配置や生成はできなかった。DiffSensei [25] は、コマ割り、登場人物と吹き出しの配置、各登場人物の画像、各コマの描写を入力として、描写に沿った登場人物のバリエーションの画像を生成した。DiffSensei は入力として登場人物や吹き出しの配置が与えられることを前提としているため、登場人物と吹き出しの配置を行う本提案とは目的が異なる。

3 方法

今回取り組むタスクは、漫画全体のシナリオを入力として各コマごとにラフスケッチと吹き出しを生成するものである。我々は三つの手法でこのタスクに取り組んだ。そのうち最も良い結果が得られた手法を NameGen と呼ぶことにし、3.1 で詳細に説明する。3.2 では、残りの二つの手法について概要のみを簡単に説明する。

3.1 NameGen

図 1 に示すように、NameGen は、漫画全体のシナリオを入力とし、コマごとのラフスケッチと吹き出しを生成する。具体的には、まず、LLM を用いて全体のシナリオをコマごとのシナリオに分割 (3.1.1) した後、コマごとの画像生成用プロンプトを生成する (3.1.2)。次に、Stable Diffusion [3] とラフスケッチ画像でファインチューニングされた LoRA [2] を用いて、前述の画像生成用プロンプトからラフスケッチを x 枚生成する。その後、事前に構築した吹き出しデータベース (3.1.3) に対して、各ラフスケッチごとに作成された検索クエリ (3.1.4) を用いて、類似する構造の漫画コマを k 件検索する (3.1.5)。検索の結果得られた漫画コマを参考として、ラフスケッチに吹き出しを配置する (3.1.6)。最終的に得られる $x \times k$ 枚の候補から人間が最も良いと感じたものを選択して最終結果とする。

3.1.1 シナリオの分割

漫画は主に複数のコマによって構成される。しかし、実際の漫画制作現場と同様、NameGen ではコマごとに分かれていない全体のシナリオのみが与えられる。したがって、最初に漫画全体のシナリオをコマごとのシナリオに分割する必要がある。

実際の漫画のシナリオ分割の観察をもとに、LLM に下記に述べる条件を与えてシナリオを分割する。

- 場面が変わる描写やセリフは同じコマに含めない
- 一つのコマに含まれる行動は一つのみとする
- 一つのコマに含まれる会話は一往復までとする
- 一つのコマに含まれる同じ人物のセリフは一つまでとする

さらに、セリフを表す文については、話者抽出、及び、モノログかダイアログかの判定を行った。

なお、ここで分割された各コマのシナリオは必ず元の文章から構成される。LLM による文章の生成や要約は行われていない。

3.1.2 画像生成用プロンプトの生成

本項では、画像生成モデルに渡すラフスケッチ生成用のプロンプト生成について述べる。プロンプト生成では、LLM を3つのステップで使用する。まず最初のステップでは、コマのシナリオに含まれる登場人物のローマ字読みを取得する。これによって、固有名詞をそのまま与えることによる別の名詞との混同を避ける: 固有名詞の「蝶子」-昆虫の「蝶」、固有名詞の「カエデ」-植物の「カエデ」。次のステップでは、最初のステップで得られた固有名詞とローマ字読みの関係を与えた上で、複数の文章（セリフの文章と描写の文章）を一つの文章にまとめる作業を行う。最後のステップでは、以前のステップで一つにまとめた文章を画像生成用の詳細な文章に変換する。ここで生成したプロンプトは画像生成モデル (Stable Diffusion v1.5 + LoRA) に与えられ、ラフスケッチの生成に活用される。

3.1.3 吹き出しのデータベース構築

NameGen では事前に構築したデータベースに対して、ラフスケッチと似た構造を持つ漫画コマを検索して参照することで吹き出しの配置を行う。本項では、高速な検索を可能にする吹き出しのデータベースの構築方法について述べる。

吹き出しのデータベース構築には Manga109 [5, 15, 20], Manga109Dialog [19] を用いる。Manga109 は 109 冊の漫画について、1 ページごとに画像とアノテーションデータを記したデータセットである。Manga109Dialog は、Manga109 についての追加のアノテーションデータであり、各吹き出しに対して話者の対応づけをしている。

まず、Manga109 のアノテーションデータをもとにアノテーションと画像を各コマごとのものに分割する。この時、複数のコマに跨っている吹き出しは除外する。その後、Manga109Dialog をもとにキャラクターの分類と吹き出しの分類を行う。キャラクターについては、吹き出しと対応づけられている場合は話者、そうでない場合は非話者として分類する。吹き出しについては、対応する話者がコマ内に存在する場合はダイアログ、そうでない場合はモノログと分類する。この分類方法では、キャラクターの心情を表す吹き出しとコマにいないキャラクターによる吹き出しを区別できない。本来モノログは前者のみを表すが、吹き出し配置の参考として使用する際には、コマに話者がいない吹き出しという点で同一視できる。

最後に、高速な検索を可能にするためにインデックス付けを行う。具体的には、コマに含まれる非話者数、話者数の組み合わせごとにデータベースをいくつかのグループに分割する。

3.1.4 吹き出し配置のためのクエリ生成

本項では、吹き出しのデータベースからラフスケッチと類似するコマを検索する際に使用する検索クエリの生成方法について述べる。

検索クエリはモノログ文字数、レイアウトという2つの要素によって構成される。モノログ文字数は話者をコマ中に持たないセリフの総文字数を表す。NameGen では、コマごとのシナリオにおけるモノログのセリフの文字数を合計することで取得する。レイアウトは、コマ内の人物の位置とその人物のセリフの文字数を記載したものである。レイアウトを得るために、まず、ラフスケッチに Openpose [10] を適用して人物の位置を得る。その後、右に位置するものから順番にコマごとのシナリオにおけるダイアログのセリフを割り当てる。「得られた人物の数 > そのコマのダイアログのセリフ数」の場合は、セリフを割り当てられなかった人物は非話者として位置だけを記載する。なお、「得られた人物の数 < ダイアログのセリフ数」の場合は、以降の処理は行わずに中断する。

3.1.5 吹き出し配置の検索

本項では、検索クエリから吹き出し配置を検索する方法について述べる。

まずは、条件をもとに簡単な絞り込みを行う。吹き出しのデータベース内のデータは、3.1.3 で述べたように話者数と非話者数の組み合わせでグループ分けされている。まずは、検索クエリの話者数と非話者数の組み合わせをもとに候補を絞り込む。その後、検索クエリのモノログ文字数と候補レイアウトのモノログ文字数を比較して、一定以上離れているものは除外する。

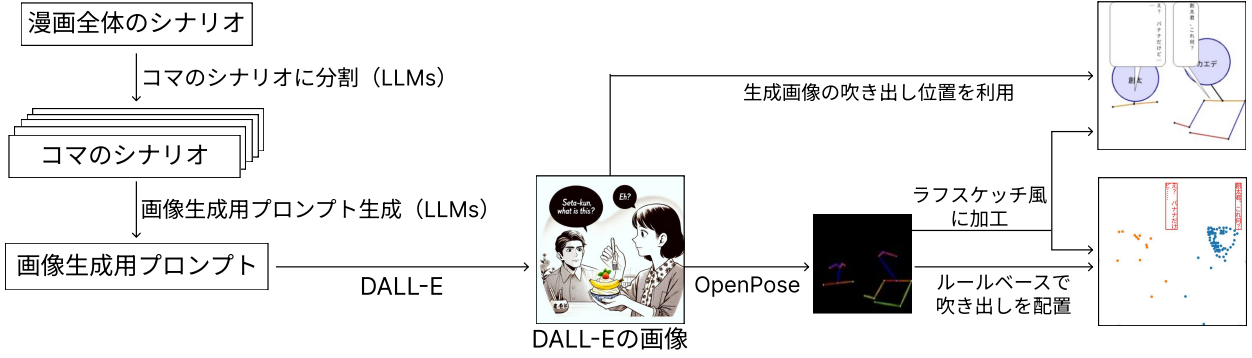


図 2. ネーム生成の他の手法 (生成画像の吹き出しの位置を利用する手法とルールベースで吹き出しの位置を決定する手法.)

次に, 上記絞り込みによって残った全ての候補に対して類似度を計算する. 類似度の計算は, 検索クエリのレイアウトと候補のレイアウト間の重み付き二部グラフの最適マッチングによって行う.

レイアウト中の話者と非話者をレイアウト要素として定義し, e または $\hat{e}$ を用いて定義する. 検索クエリのレイアウト $l = \{e_1, e_2, \dots, e_m\}$ と候補レイアウト $l' = \{\hat{e}_1, \hat{e}_2, \dots, \hat{e}_n\}$ を考える.

類似度は以下の最適化問題を解くことで得られる.

$$\begin{aligned} \max \quad & \sum_{i=1}^m \sum_{j=1}^n w(e_i, \hat{e}_j) \cdot \gamma_{ij} \\ \text{s.t.} \quad & \gamma_{ij} \in \{0, 1\}, \forall i \in \{1, 2, \dots, m\}, \forall j \in \{1, 2, \dots, n\}, \\ & \sum_{j=1}^n \gamma_{ij} \leq 1, \forall i \in \{1, 2, \dots, m\}, \\ & \sum_{i=1}^m \gamma_{ij} \leq 1, \forall j \in \{1, 2, \dots, n\}. \end{aligned} \quad (1)$$

ただし, $\gamma_{ij} \in \{0, 1\}$ は, 要素のペア $(e_i, \hat{e}_j)$ がマッチングされる時に 1 となり, それ以外の場合は 0 となる.

$e_i, \hat{e}_j$ 間の重み $w(e_i, \hat{e}_j)$ は次のように定義する.

$$w(e_i, \hat{e}_j) = \begin{cases} k \cdot \text{IoU}(e_i, \hat{e}_j) + (1 - k) \cdot G(t_i, \hat{t}_j), & \text{if } T(e_i) = T(\hat{e}_j) = \text{Speaker}, \\ \text{IoU}(e_i, \hat{e}_j), & \text{if } T(e_i) = T(\hat{e}_j) = \text{NonSpeaker}, \\ 0, & \text{otherwise} \end{cases}$$

ただし, $T(e_i)$ と $T(\hat{e}_j)$ はそれぞれ, $e_i, \hat{e}_j$ の属性を表し, 要素が話者の場合は Speaker, 非話者の場合は NonSpeaker となる. 要素の属性が Speaker のとき, t_i と $\hat{t}_j$ はそれぞれ, $e_i, \hat{e}_j$ が持つセリフの長さを表す. また, G は以下の式で表されるガウス関数

である. 2つの Speaker 要素が持つセリフの長さが近いほど高い値を取る.

$$G(t_i, \hat{t}_j) = \exp\left(-\frac{(t_i - \hat{t}_j)^2}{2\sigma^2}\right)$$

ただし, $k \in [0, 1]$ は, IoU と G の線型結合の割合を調整するためのパラメータである.

この最適化問題をハンガリアン法を用いて解くことで類似度を計算する.

3.1.6 吹き出しの配置

ラフスケッチの吹き出しの配置は, 前項で検索した実際の漫画コマの吹き出し配置を参照して行う. 検索の結果得られた漫画コマは複数の吹き出しを含むことがある. ラフスケッチの吹き出しを配置する際に参照するものを参照吹き出しとして, 以下では参照吹き出しを選択する方法を述べる.

参照吹き出しの選択方法は, セリフがモノログの場合とダイアログの場合で異なる. セリフがモノログの場合は, 検索結果の漫画コマのモノログの吹き出しのうち最も右側に位置するものを選択する. セリフがダイアログの場合は, 次に示すアルゴリズムによって選択する.

参照吹き出しを選択した後, 参照吹き出しの右上を開始点として縦書きで文字を配置する. この時, 縦幅は参照吹き出しと同じにし, 横幅を文字数に合わせて調整する. その後, モノログ要素の場合は長方形, ダイアログ要素の場合は楕円で文字全体を囲む.

3.2 ネーム生成の他手法

ここでは, NameGen に至るまでに行った二つの別手法について説明する. 図 2 に概要を示す. どちらの手法も NameGen と同じように画像生成用プロンプトを生成し, それに基づいて画像生成を行うが, 画像生成では Stable Diffusion ではなく OpenAI の DALL-E を使用した. その後, OpenPose によって

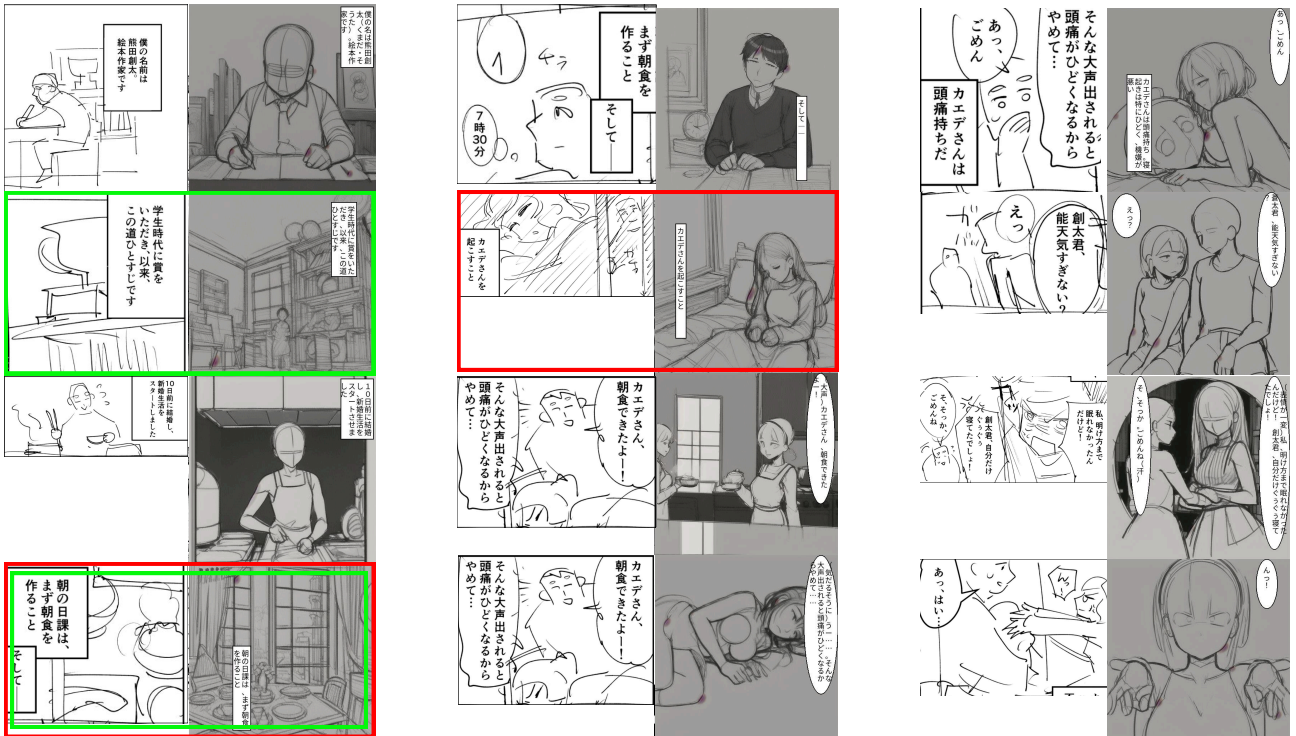


図 3. 『モラハラ DV 妻から命からがら逃げました』の実際の実原稿をもとに生成した結果の例。各コマごとに、左の画像が漫画家が書いたネーム、右の画像が生成されたラフスケッチと吹き出し。赤枠は吹き出しを左側に配置できた例。緑枠は人物がいないコマを生成できた例

Algorithm 1 ダイアログ要素の参照吹き出しを決定するアルゴリズム

```

0: 入力: 検索結果の漫画コマ  $l'$ , コマの原稿  $panel$ 
0:  $l'$  の要素  $\hat{e}_1, \hat{e}_2, \dots$  を並びかえる. 話者要素は
   非話者要素よりも前に来る. 話者要素同士は関
   連づけられている吹き出しが右にあるものほど
   前に来る.
0:  $pointer$  変数を 0 として初期化.
0: for  $\hat{e} \in l'$  do
0:   while  $panel[pointer]$  がダイアログではない
     do
0:      $pointer \leftarrow pointer + 1$ 
0:   end while
0:   if  $pointer \geq len(panel)$  then
0:     処理を終了
0:   end if
0:    $bbox \leftarrow \hat{e}$  が持つ最右の吹き出し
0:    $bbox$  を  $panel[pointer]$  の参照吹き出しに選択
0: end for

```

人物のポーズを抽出し、一つ目の手法では生成画像に含まれる吹き出しの位置、二つ目の手法ではルールベースで決定した位置に基づいて吹き出しの配置を行う。

一つ目の手法では、画像生成モデルが生成する画

像にすでに含まれている吹き出しを利用する。ファインチューニングを Manga109 で行った物体検出モデル [4] を用いて吹き出しを検出する。物体検出の信頼度が高い吹き出しのうち、最も右側に位置するものから順番に吹き出しの位置の参照として利用する。生成画像はラフスケッチではないので、OpenPose によって抽出した人間の姿勢情報をもとにラフスケッチに近くなるように加工する。

二つ目の手法では、吹き出しを右上から左側に向かって順番に配置するように試みる。その過程で吹き出しが人物と重なった場合は、人物を縮小して再度配置を試みる。

これらの方法で生成した結果と NameGen の結果の比較は 4.3 に記載した。

4 実験と結果

本システムを使用して『モラハラ DV 妻から命からがら逃げました』¹の実際の実原稿をもとにラフスケッチと吹き出しを生成した。また、NameGen と 3.2 で述べた他の手法との比較も行った。

¹ © 創太・ミカイ/パルソラ



図 4. ネームの生成がうまくいかない例

4.1 実装と手順

本システムの実装で使用する LLM は全て OpenAI の gpt-4o を用いた。

NameGen では、画像生成に Stable Diffusion v1.5 を使用した。チェックポイントは T-anime-v4 [1], LoRA は sketch anime pose [2] を採用した。吹き出し配置の検索で使用するガウス関数 G のパラメータ σ は 10, IoU と G の線型結合の割合を調整するパラメータ k は 0.4 とした。コマのシナリオへの分割、画像生成用プロンプトの生成はそれぞれ 1 回のみ行い、Stable Diffusion と LoRA を用いた画像生成は 3 回行った。生成された各画像に対して吹き出し配置の検索により類似度の高い順に 4 つの参考コマを取得した。そして、それぞれの参考コマごとに参考コマをもとにした吹き出し生成を行った。最後に、生成された 12 個（3 つのラフスケッチ × 4 つの参考コマ）の結果の中から我々が最も良いと考えるものを 1 つ選び、生成結果とした。

4.2 実際の実稿からの生成

漫画家が生じたネームと生成された結果の比較を図 3 に示す。各列の左側の画像が漫画家が生じたネーム、右側の画像が生成されたラフスケッチと吹き出しである。なお、漫画家が生じたネームは、生成結果の吹き出しのセリフをもとに対応すると考えられるものを選択した。シナリオの分割では、図 3 の吹き出しの中身に注目すると、二列目の下二コマ以外は実際の漫画と同じように分けることができた。ラフスケッチと吹き出しの生成では、赤枠で囲ったコマのように吹き出しが左に位置するコマの生成も行うことができた。また、緑枠で囲ったコマのように人物がいないコマに対してもラフなスケッチを提供することができた。一方で、コマを通して人物の一貫性が保たれなかった。また、図 4 に示すように、ハグをしてもらっているべき場面で単一の人物のみが生成されたり、長いモノローグが人物に重なってしまったりするなどの問題も確認された。

4.3 NameGen と他手法の比較

図 5 に手法ごとの結果の比較を載せる。それぞれで 5 枚ずつ生成したもののうち最も良いものを抜粋した。生成画像の吹き出し位置を利用する手法では、吹き出しが顔に被ったり、不自然な配置になったりする不備が多く見られた。ルールベースで吹き出し

を配置する手法と NameGen はどちらも似たような吹き出しの配置になった。しかし、ルールベースの手法は、どのコマでも似たような吹き出しの配置になってしまう問題があった。NameGen ではルールベースの手法で可能な標準的な吹き出しの配置だけでなく、図 3 で示したように多様な種類の吹き出しの配置ができた。

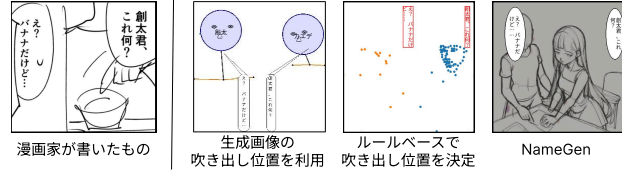


図 5. それぞれの手法ごとの結果の比較

5 課題と改善の方向性

NameGen は漫画シナリオのみから多様なパターンのラフスケッチや吹き出しを生成できる一方で課題も存在する。ここでは特に重要と考えられる三つを述べる。一つ目は、コマ割りで正方形のコマを縦に並べるのみとしている点である。これは、Automatic Stylistic Layout [8] などの先行研究を用いることで克服できるだろう。二つ目は、生成結果の中から最も良いものを人の手によって選んでいるという点である。これは、人物と吹き出しが被っているものを除外するなどのルールベースによるフィルタや、LLM に判断基準を与えて生成画像を評価するなどのランク付けによって改善できると考える。三つ目は、セリフの話者と人物の特徴が一致しない点である。例えば、女性のセリフにも関わらず話者が男性である場合などがこれに該当する。生成結果の画像に対して物体検出を行い、どの登場人物か判定した上でセリフを割り当てるなどの工夫が考えられる。

6 まとめ

本研究では、漫画製作者の支援を目的としてネームの制作支援システム NameGen を提案した。このシステムでは、漫画全体のシナリオをコマごとに分割し、画像生成モデルや実際の漫画コマに基づく吹き出し配置によって、様々なバリエーションを持つラフスケッチと吹き出しの生成を可能にした。将来的には、コマ割りや人間が選ぶ手間の削減を行う予定である。

謝辞

図中の漫画画像は Manga109 [15] の画像を一部使用した (© 竜正, 浅月 舞, 島田 ひろかず, 花田 朔生, 記伊 孝)。

参考文献

- [1] Civitai T-Anime. <https://civitai.com/models/69552/t-anime-v4-pruned>.
- [2] Sketch Anime Pose. <https://civitai.com/models/106609/sketch-anime-pose>.
- [3] Stable Diffusion v1.5. <https://huggingface.co/stable-diffusion-v1-5/stable-diffusion-v1-5>.
- [4] YOLO. <https://github.com/ultralytics/ultralytics>.
- [5] K. Aizawa, A. Fujimoto, A. Otsubo, T. Ogawa, Y. Matsui, K. Tsubota, and H. Ikuta. Building a Manga Dataset “Manga109” with Annotations for Multimedia Applications. *IEEE MultiMedia*, 27(2):8–18, 2020.
- [6] J. Boreczky, A. Girgensohn, G. Golovchinsky, and S. Uchihashi. An interactive comic book presentation for exploring video. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’00, p. 185–192, New York, NY, USA, 2000. Association for Computing Machinery.
- [7] J. Calic, D. P. Gibson, and N. W. Campbell. Efficient Layout of Comic-Like Video Summaries. *IEEE Transactions on Circuits and Systems for Video Technology*, 17(7):931–936, 2007.
- [8] Y. Cao, A. B. Chan, and R. W. H. Lau. Automatic stylistic manga layout. *ACM Trans. Graph.*, 31(6), Nov. 2012.
- [9] Y. Cao, R. W. H. Lau, and A. B. Chan. Look over here: attention-directing composition of manga elements. *ACM Trans. Graph.*, 33(4), July 2014.
- [10] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh. OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields, 2019.
- [11] S. Chen, D. Li, Z. Bao, Y. Zhou, L. Tan, Y. Zhong, and Z. Zhao. Manga Generation via Layout-controllable Diffusion, 2024.
- [12] W.-T. Chu, C.-H. Yu, and H.-H. Wang. Optimized Comics-Based Storytelling for Temporal Image Sequences. *IEEE Transactions on Multimedia*, 17(2):201–215, 2015.
- [13] B.-K. Chun, D.-S. Ryu, W.-I. Hwang, and H.-G. Cho. An Automated Procedure for Word Balloon Placement in Cinema Comics. In G. Bebis, R. Boyle, B. Parvin, D. Koracin, P. Remagnino, A. Nefian, G. Meenakshisundaram, V. Pascucci, J. Zara, J. Molineros, H. Theisel, and T. Malzbender eds., *Advances in Visual Computing*, pp. 576–585, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.
- [14] A. Durrant, D. Rowland, D. S. Kirk, S. Benford, J. E. Fischer, and D. McAuley. Automics: souvenir generating photoware for theme parks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’11, p. 1767–1776, New York, NY, USA, 2011. Association for Computing Machinery.
- [15] A. Fujimoto, T. Ogawa, K. Yamamoto, Y. Matsui, T. Yamasaki, and K. Aizawa. Manga109 dataset and creation of metadata. In *Proceedings of the 1st International Workshop on CoMics ANalysis, Processing and Understanding*, MANPU ’16, New York, NY, USA, 2016. Association for Computing Machinery.
- [16] L. Herranz, J. Calic, J. M. Martinez, and M. Mrak. Scalable Comic-Like Video Summaries and Layout Disturbance. *IEEE Transactions on Multimedia*, 14(4):1290–1297, 2012.
- [17] G. Jing, Y. Hu, Y. Guo, Y. Yu, and W. Wang. Content-Aware Video2Comics With Manga-Style Layout. *IEEE Transactions on Multimedia*, 17(12):2122–2133, 2015.
- [18] D. Kurlander, T. Skelly, and D. Salesin. Comic Chat. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH ’96, p. 225–236, New York, NY, USA, 1996. Association for Computing Machinery.
- [19] Y. Li, K. Aizawa, and Y. Matsui. Manga109Dialog: A Large-scale Dialogue Dataset for Comics Speaker Detection, 2024.
- [20] Y. Matsui, K. Ito, Y. Aramaki, A. Fujimoto, T. Ogawa, T. Yamasaki, and K. Aizawa. Sketch-based Manga Retrieval using Manga109 Dataset. *Multimedia Tools and Applications*, 76(20):21811–21838, 2017.
- [21] D.-S. Ryu, S.-H. Park, J.-w. Lee, D.-H. Lee, and H.-G. Cho. CINETOON: A Semi-automated System for Rendering Black/White Comic Books from Video Streams. In *2008 IEEE 8th International Conference on Computer and Information Technology Workshops*, pp. 336–341, 2008.
- [22] A. Shamir, M. Rubinstein, and T. Levinboim. Generating comics from 3D interactive computer graphics. *IEEE Computer Graphics and Applications*, 26(3):53–61, 2006.
- [23] H. Tobita. Comic Engine: interactive system for creating and browsing comic books with attention cuing. In *Proceedings of the International Conference on Advanced Visual Interfaces*, AVI ’10, p. 281–288, New York, NY, USA, 2010. Association for Computing Machinery.
- [24] M. Wang, R. Hong, X.-T. Yuan, S. Yan, and T.-S. Chua. Movie2Comics: Towards a Lively Video Content Presentation. *IEEE Transactions on Multimedia*, 14(3):858–870, 2012.
- [25] J. Wu, C. Tang, J. Wang, Y. Zeng, X. Li, and Y. Tong. DiffSensei: Bridging Multi-Modal LLMs and Diffusion Models for Customized Manga Generation, 2025.

OpticDesigner：多様な光学レンズ効果をカスタマイズ可能な3D プリントレンズフィルターデザインシステム

吉本 健義* 大友 千宙* 宮下 芳明*

概要. 写真家や映像制作者をはじめとする多くの撮影者は、それぞれ理想の表現を持っている。こうした表現を実現する手段は、カメラ内効果による撮影前処理とソフトウェア編集による撮影後処理に大別される。しかし、後処理では完全な再現が困難な効果もあり、多くの場合は前処理に依存する。だが、カメラ内の光学レンズ効果を構成するレンズやフィルターは個人ユーザが独自に調整することが難しい。本稿では、ユーザが自身の求める効果に応じて形状をカスタマイズできるレンズフィルターデザインシステム OpticDesigner を提案し、複数の光学レンズ効果の統合や強度、適用範囲の制御を可能にした。作製には光造形方式 3D プリンタを使用し、スピンコーティング法による後処理を加えることで、撮影用レンズフィルターとして利用可能な水準の透明度を実現した。さらに、写真・映像制作経験者を対象としたワークショップを実施し、ユーザの意図に応じてデザインされた光学レンズ効果の有効性とデザインの多様性を示した。

1 はじめに

撮影者にはそれぞれが求める理想の表現があり、写真や映像を構成する構図、露出等には全て撮影者の意図が存在する。写真・映像制作において、撮影者が求める表現を実現する手段は、カメラ内効果による撮影前処理とソフトウェア編集による撮影後処理に大別される。後処理でのみ可能な表現として、時間軸の編集による動きの変化や、撮影時には存在しない対象の追加などがある。しかし、収差やボケをはじめとする光学的なアーティファクトに由来する効果は完全な再現が困難である。これらは、レンズやフィルターなどを用いた撮影前処理だからこそ表現できる要素である。そのため、個別のニーズを満たすカメラ内効果を得るためには、撮影時に光学レンズ部品を調整する必要がある。近年の映像制作では、撮影者が求める最適な表現を生み出すため、その場面に適したレンズの設計・作製が行われている [6]。映画「Oppenheimer」では、理想の表現や光学性能を得るため、複数回の試作・撮影・改良を経てカスタムレンズを開発している。しかし、光学レンズの作製には精密な形状加工が求められ、それに要する時間や設備を伴うため製造コストは高い。そのため、個人がカスタムレンズ制作において試行錯誤を重ねることは難しい。多くの場合は工業的に生産された量産品を使用せざるを得ず、ユーザが扱える表現の幅や制御可能な領域が制限されてきた。

そこで、本稿では個人ユーザが光学レンズ効果の設計・作製を可能とすることを目的に、3D プリントを用いたレンズフィルターデザインシステム OpticDesigner を提案する。OpticDesigner は、レ

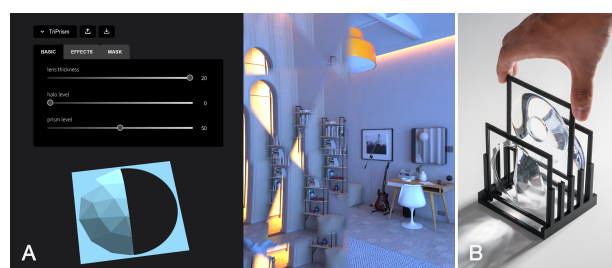


図 1. (A) OpticDesigner のインターフェース画面。
(B) 設計・作製されたレンズフィルター。

ンズフィルターの形状デザインを通じて複数の光学レンズ効果を統合し、その強度や適用範囲を制御できる。また、ユーザはシミュレーションによって作製したレンズフィルターの視覚効果を確認し、その結果に基づいた調整によって、自身の求める光学レンズ効果を実現できる。これにより、表現意図に応じた光学レンズ効果を柔軟に制御することが可能となり、シーン単位で変化する多様な視覚表現を低コストで実現する。作製したレンズフィルターは既存の光学系に外部から装着可能であり、カメラ本体の光学性能を保持したまま効果を付加できる (図 1B)。レンズ部分の作製には光造形方式 3D プリンタを用い、造形物にスピンコーティング法による後処理を加えた。これによって、写真や映像表現のためのレンズフィルターとして利用可能な水準の透明度を実現した。本稿では、OpticDesigner の機能を紹介し、設計・作製におけるワークフローを示す。さらに、著者らによる作例および写真・映像制作経験者を対象としたワークショップによるデザイン事例を通じて、本システムによって作製された光学レンズ効果の有効性と表現の多様性を検証する。

Copyright is held by the author(s).

* 明治大学

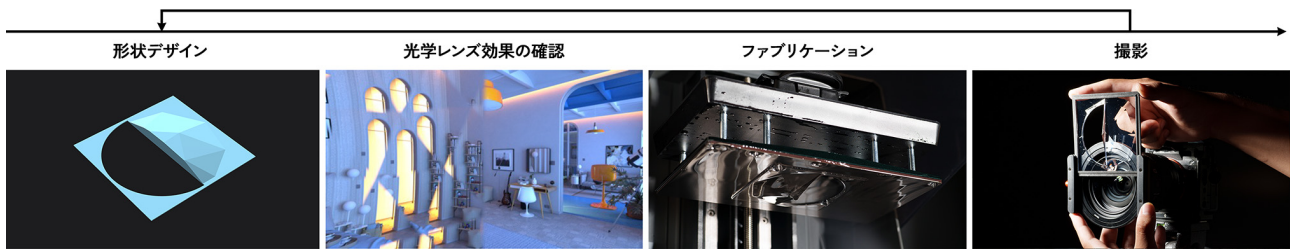


図 2. OpticDesigner によるワークフロー。

2 関連研究

2.1 光学レンズ効果の編集

写真・映像に対する後処理においては、Adobe Photoshop や DaVinci Resolve といったソフトウェアが用いられる。こうしたソフトウェアでは、レンズに直接光が差し込んだ際に発生するレンズフレアなどの光学的アーティファクトを意図的に付与し、その効果を調整することができる。また、本研究以外にも光学的アーティファクトを写真表現の重要な要素として扱っているものがある。例えば、Hullin らは物理的な正確さと表現の自由度を併せ持つレンズフレアのレンダリング手法を提案した [5]。また、Spencer らは光の散乱現象であるグレア効果を既存の画像に追加するアルゴリズムを提案している [8]。

だが、多くの研究では単一の光学的アーティファクトを対象としており、ユーザが求める複数の光学レンズ効果を写真や映像に付与することはできない。さらに、プリズムによる多重像やレンズ内の複雑な反射によるゴーストなどの効果は、視覚的な擬似再現は可能であるが、違和感のない完全な再現を既存の編集ソフトウェアで実現することは困難である。

2.2 3D プリント光学素子

3D プリント技術の発展により、積層造形を応用したフレネルレンズ [1] やアルバレスレンズ [4] をはじめとする多様な機能性をもつ光学素子の作製が注目されている [10]。特に、低価格帯の光造形方式 3D プリンタの普及とスピニング法による薄膜形成を用いた後処理 [9] により、光学用途にも応用可能なレンズを短時間かつ低コストで造形することが可能となった。Berglund らは、スピニング法により 3D プリントした平凸レンズの表面を平滑化し、射出成形プラスチックレンズと同程度の表面粗さを実現した [7]。撮像性能はガラスレンズには及ばないものの、高解像を必要としない用途では十分な性能を示した。さらに、Christopher らによる研究では同様の手法を用いることで、顕微鏡用のレンズ群を設計・作製し、明視野観察や蛍光観察用の対物レンズとして応用できることを示した [3]。一方、Chen らは、投影型マイクロ光造形技術 (Projection Micro-stereolithography) にグ

レースケール光重合とメニスカス平衡二次硬化を組み合わせて、低コストかつ高速に光学素子を造形する手法を提案した [2]。この手法により作製した非球面レンズをスマートフォンのカメラに装着することで、高解像度撮影を実証するとともに市販のカメラ装置への実装可能性を示した。

これらの研究では、3D プリント技術によって形状自由度の高い光学素子を低コストで作製している。しかし、多くの先行研究では光学素子作製の精度向上と性能評価が目的であり、光学レンズ効果のデザインによる表現支援に取り組む事例はほぼ存在しない。本研究は、ユーザ自ら光学レンズ効果をデザインできるシステムを提供する。さらに、先行研究で用いられたスピニング法によって、撮影用レンズフィルターとして利用可能な水準の透明度を実現した。これにより、ユーザの意図や目的に応じた光学レンズ効果のデザインを可能とし、ユーザの表現領域の拡張を促すことができる。

3 OpticDesigner

OpticDesigner におけるワークフローは図 2 に示すように、形状デザイン・光学レンズ効果の確認・ファブリケーション・撮影の 4 フェーズから構成される。ユーザはこれらのフェーズ内での試行錯誤によって、自身の求める表現をもたらすレンズフィルターをデザインする。OpticDesigner の対象ユーザは写真・映像を作品として仕上げた経験を有するプロ、もしくはハイアマチュアのカメラマンである。光学に関する専門的な知識は要求しないが、求める表現や意図を持ち、それらを反映させるための能力を有することを前提としている。本研究における表現や意図の範囲は、被写体及び構図、さらには構図内の特定領域に付与する効果の種類や強度とする。これらの要素を踏まえ、撮影者はそれぞれを組み合わせながら自身が求める画を構築していく。

3.1 デザインソフトウェア

OpticDesigner では図 1A に示すように画面左側でフィルター形状をデザインし、画面右側で光学レンズ効果を確認する。フィルター形状に関するアルゴリズムは Rhinoceros/Grasshopper で、光学レンズ効

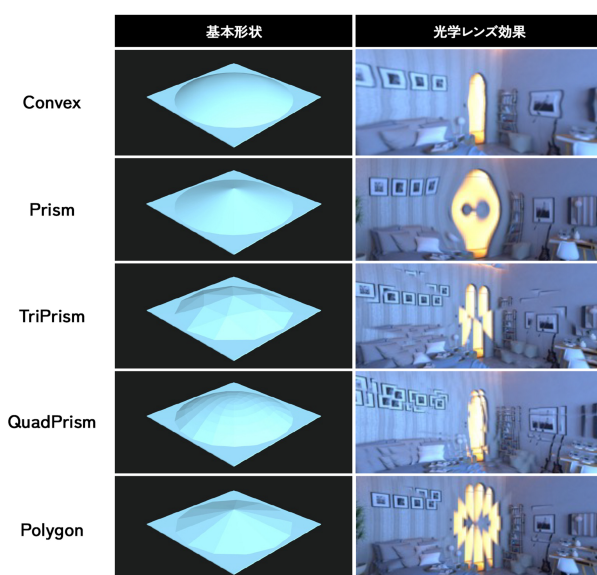


図 3. OpticDesigner による 5 種類の基本形状. ユーザはまずこの基本形状を選択し, 各種パラメータを制御することで自身が求める表現をもたらすフィルターをデザインしてゆく.

果のレイトレーシングシミュレーションは V-ray for Rhino によって実装した. また, Rhinoceros/Grasshopper の使用経験がないユーザでも OpticDesigner を扱えるように, UI を JavaScript によって実装し, Web アプリケーションとしている. ユーザはスライダーやプルダウンメニューといった一般的な UI 要素の操作によって, フィルター形状をデザインする. 一般に, カメラマンは形状編集を前提とする CAD ソフトウェアの操作経験を有していないことが多いが, Lightroom や Capture One 等の写真編集ソフトウェアの操作には慣れていると考えられる. そこで本研究では, これらのソフトウェアで多く用いられるスライダー形式の UI で実装した.

OpticDesigner では図 3 に示す 5 種類の基本形状があり, これらの基本形状を各種パラメータによって制御することができる. 本研究で実装した基本形状は, レンズフィルターとして市販されているものの中から, 光造形方式の 3D プリントにおいて技術的に再現可能なものを選んだ. 形状パラメータは, 効果の強度を制御するものと効果の範囲を制御するものの 2 種類に分類される. 前者には, レンズの厚みや多重像の大きさ・形状など, 効果の強度や特性に直接影響を与えるものが含まれる. 一方, 後者には, マスク範囲の指定, 穴の形状や大きさの制御, 形状表面に作成する平面の大きさを調整するものなどが含まれる. さらに, フィルター裏面に溝を形成することでクロスフィルター効果を制御する機能や, 任意のテキスト形状に沿って穴を作成する機能も実装しており, 光造形方式 3D プリンタの形状自由度を最大

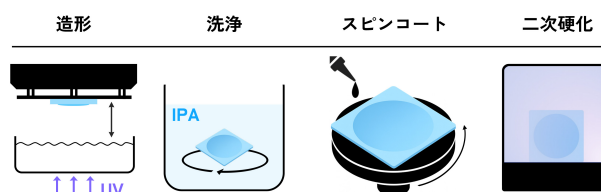


図 4. レンズフィルターのファブリケーション手順. スピンコーティング法によって 3D プリント由来の積層痕を緩和し, フィルターとして求められる光学性能を向上させている.

限に活用したデザインができる. OpticDesigner でデザイン可能な形状は全 10 万通り以上であり, ユーザは自身が求める表現のための 1 枚をデザインすることができる. また, OpticDesigner は STL 形式での保存や CSV 形式での形状パラメータの入出力に対応している. そのため, ユーザは気に入った形状を保存して次の形状をデザインすることや実際の撮影後にさらなる微調整を行うことが可能である.

OpticDesigner では画面左下にフィルター形状のプレビューを, 画面右にその形状による光学レンズ効果のプレビューを表示している. 形状のプレビュー部ではマウス操作によって様々な角度からフィルター形状を確認することができる. ユーザが 1 つでもパラメータを変更すると形状の再計算が行われ, そのプレビューが画面左下に表示される. 一方, 光学レンズ効果のプレビューはその処理時間の長さを考慮して, ユーザが形状を出力しないで行わない仕様とした. そのため, ユーザは自身の求める表現をもたらすであろうフィルター形状をある程度デザインしてから, その光学レンズ効果を確認することとなる.

3.2 ファブリケーション

本研究におけるフィルターのファブリケーション手順を図 4 に示す. デザインソフトウェアが STL 形式で出力したフィルター形状は, 光造形方式 3D プリンタ (Saturn 4 Ultra 16K, ELEGOO) で造形する. 著者らによる試行錯誤ののち, マテリアルにはフィルターとしての使用に適した高い透明度を持つレジン (新高透明度レジン, SK 本舗) を採用した. また, 造形パラメータの試行錯誤も行い, OpticDesigner でデザインした形状が確実に造形されるようレイヤー高さ 0.02 mm・露光時間 1.75 秒に設定している. 造形後は, IPA による 10 分間の洗浄の後, 透明度を向上させるためにスピンコートを行う. 洗浄後のフィルターに片面ずつ UV コーティングレジン (Kirari, SK 本舗) を滴下し, 段階的に回転速度を上昇しながら約 1 分間回転させる. スピンコートによって UV コーティングレジンを均一に塗布した後, Form Cure を用いて 2 分間の二次硬化を行う. また, 二次硬化前に予め FDM 方式 3D

プリンタによって造形しておいた 100 mm 幅の枠をフィルターに接着する。市販のフィルターも 100 mm 幅で製造されていることが多いため、既存のフィルター用ホルダーにそのまま装着することができる。

4 作例

OpticDesigner による光学レンズ効果のデザイン及び制御の有効性を示すため、著者らによって撮影された 2 つの作例を紹介する。これらの作例によって、本システムで設計・作製されたレンズフィルターの実際の写真・映像制作における応用可能性を示す。

図 5A は、部分的に開口したプリズム構造を持つレンズフィルターで撮影された写真である。プリズム効果とは、レンズフィルター内で光が屈折する際に分散や反射が生じ、多重像やスペクトルの分離といった視覚効果を生み出す現象である。市販のプリズムレンズフィルターは種類や効果のバリエーションが多く、ユーザ自ら自身のニーズを満たすものを選択し、その製品が持つ表現力の範囲で制作せざるを得なかった。特に、多重像を生み出す効果は幻想的な表現を可能にする反面、効果が画面全体に及ぶことで意図しない冗長な表現となり、フレーム内の主題の明瞭さを損なう恐れがある。OpticDesigner では、光学レンズ効果が及ぶ範囲や強度をユーザがデザインするため、背景の光源のみを強調しつつ、被写体となる人物を明瞭に描写できる。図 5A では、画面中央から左上にかけて背景のレインボーブリッジのハイライト部分が多重像として現れている一方で、残りの領域を開口部とすることでシャープで自然な表現を実現している。さらに、本作例では開放絞りで撮影することでフィルターの開口部を周囲の風景と自然になじませ、一枚の写真の中で幻想的な雰囲気と写実的な描写を両立している。

また、図 5B は、複合された効果を持つレンズフィルターによって撮影された写真である。従来のレンズフィルターでは、1 枚につき 1 種類の光学レンズ効果が基本であり、複数の効果を得るためには複数のレンズフィルターを重ねて使用する必要があった。しかし、フィルターの厚みが増すと周辺減光や光学収差が生じやすくなるという課題がある。OpticDesigner では、複数の光学要素を 1 枚のフィルターに統合し、意図した効果を効率的に得ることができる。本作例では、凸レンズとクロスフィルターの形状を組み合わせ、中央に開口部を設けたレンズフィルターを作製した。図 5B に示すように、中央の開口部ではレンズ本来のシャープな描写が保たれ、周辺部ではクロスフィルター構造によって点光源が拡散されている。また、凸レンズ構造による歪曲収差が加わることで、光芒が放射状に広がる印象的な表現が生まれ、複合的な視覚効果が実現している。



図 5. 作製したレンズフィルターを用いて撮影した写真。
(A) 部分的に開口させることで、効果の適用範囲を制御している。(B) 複数の光学要素を組み合わせ、複雑な光の拡散を表現している。

5 ワークショップ

OpticDesigner によって可能となる表現の事例を収集し、そのデザインの多様性を示すことを目的として、ワークショップを実施した。参加者は 22-28 歳の男性 4 名であり、写真・映像制作経験を有するプロ 1 名とアマチュア 3 名で構成された。いずれも自らの機材を用いた 1 年以上の制作経験を有しており、本研究の対象ユーザとして適切であると判断した。

5.1 プロセス

ワークショップでは OpticDesigner によって 2 枚以上のレンズフィルターをデザインするタスク（以下、デザインタスク）と、そのフィルターを用いて写真・映像を制作するタスク（以下、撮影タスク）の 2 つを行った。デザインタスクはワークショップの初日に実施され、全体の所要時間は約 1 時間であった。撮影タスクは参加者がフィルターを入手してから約 5 日間の間で、各参加者が自由なタイミングで実施した。各タスク終了後に、参加者はアンケートと半構造化インタビューへ回答するように求められた。

デザインタスクでは基本形状が異なるフィルターを 2 枚以上デザインすることを制約とし、時間は 1 枚につき 30 分を目安とした。デザインタスク実施前に本研究の概要説明を行い、撮影データの取り扱いに関する同意書への署名を得た後、OpticDesigner の使用方法を説明した。また、全ての基本形状と 4 節での作例を参加者に提示し、本システムによる表現の可能性を示した。参加者は必要に応じてそれらを手に取り、カメラを通して実際の見え方を確認しながらデザインタスクに取り組むことができた。デ

デザインタスク終了後にこれまでの制作経験に関するアンケートと、各フィルターをデザインした意図について尋ねる半構造化インタビューを実施した。

撮影タスクではフィルターの入手後約5日間、自由なタイミング・題材で写真・映像を制作してもらった。撮影タスク終了後、参加者には自身がそのフィルターによって表現したかったことが表れている作例をフィルターごとに3つ程度選出させた。選出した作例について、その選出意図や撮影時に感じたことについてを半構造化インタビューによって回答した。

5.2 デザイン事例

図6に、参加者がデザインしたレンズフィルター形状及び作例写真を示す。各参加者は2種類のレンズフィルターを作製した。ほとんどの参加者はConvexやPrismなどの画面全体に歪曲収差を生じさせる形状とTriprismやPolygonなどの多重像を生み出す形状をそれぞれ一枚ずつ選択していた。また、全ての参加者が効果範囲を制御するために、レンズフィルターの一部に開口部を設けていた。さらに、中央部への穴あけや断面を付与する機能、テキストの形状に穴を空ける機能を併用し、既製品ではないような複雑な形状をデザインする傾向が見られた。

5.2.1 ユーザ意図と予期せぬ効果がもたらす表現の多様性

参加者がデザインしたレンズフィルターは、デザインタスク時に意図した表現を実際の撮影においても発揮することができていた。例えば、参加者Cはデザインタスク時に、日常の風景において非現実感を現実と融合させるような異世界感を表現したいという意図を持ってフィルター作製を行っていた。図6C-2に示すように、形状としては、異世界感を表現するために中央部には大きなプリズムの形状を、周辺部には風景をそのまま映すために開口部を作成した。実際の作例では、画面中央において被写体の多重像が現れ、参加者Cが狙った「異世界感」を表現できたと自ら述べていた。また、参加者Dはデザインタスクにおいて、画面全体に歪曲収差を付与しつつ中央部では被写体を明瞭に描写したいと述べていた。その意図に基づき、薄い凸レンズの中央部を開口したような形状を作成した。作例(図6D-2)では、画面中央で蛇口を鮮明に描写しながら、周辺領域では水しぶきや逆光によるゴーストを歪ませることができていた。さらに、参加者Bは、対称的な多重像の配置による幾何学的な光の広がりを用意し、三角形のプリズム構造と開口部を組み合わせた。図6B-1から、生じた多重像を被写体の手前に配置することで、従来の写真表現における前ボケに近い効果が得られ、印象的な重畳表現を生み出すとともに、空間に奥行きや広がりを与える表現が確認できる。このように、撮影タスクでは意図通りの

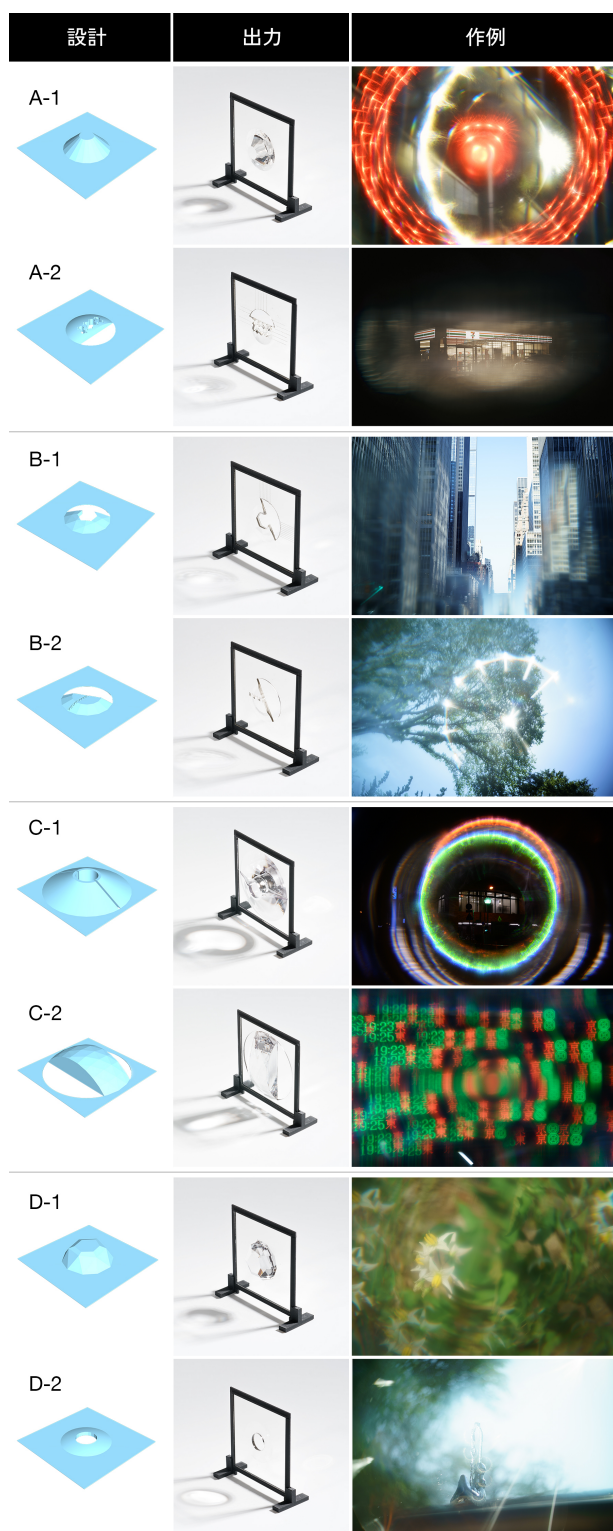


図6. ワークショップ参加者によって作製したレンズフィルターのデザイン事例。左から順に、参加者によってデザインされた形状、3Dプリントされた出力結果、およびそのレンズフィルターを用いて撮影された作例を示す。各参加者(A-D)は2種類のレンズフィルターをデザインし、いずれも異なる基本形状を選択している。

効果を得る過程で、撮影者が効果を応用的に用いることにより、当初のデザインタスク時の想定を上回る新たな表現が実現される場合もあった。一方、撮影タスク後のインタビューでは多くの参加者が「デザイン時に意図していなかった効果が得られた」と述べた。図 6A-1・A-2・B-2・C-1 に示す作例では、点光源を画角内に配置した際に、デザイン時には想定されていなかった光輪や、開口部・断面部の形状を通過した光源の拡散といった効果が確認された。これらの結果は、偶発的に生じる効果が写真表現の多様性を広げる可能性を持つことを示している。

5.2.2 デザインにおけるプレビューの有効性

参加者の多くは、事前に撮影するシーンを想定した上でレンズフィルターのデザインを行っていた。具体的には、まずシミュレーション機能で全体の見え方を検討し、おおよその構成を決定した後に、形状プレビューを用いて細部を調整するという利用方法が観察された。一方で、参加者の中には効果のプレビューに触発され、そこから表現を構築した者もいた。彼らはあるパラメータに極端な値を入力し、そこから得られる効果を実際の撮影にどう活かすかを検討しながらデザインを進めていた。例えば、参加者 A はレンズ部の厚みを最大値にすることで、強力な多重像の効果が得られることを発見した。その上で、主題の明瞭さを損なわない範囲で中央の解像感と周辺部の拡散のバランスをとる方針で図 6A-1 の形状を作り上げていた。インタビューにおいては、レンダリングのタイミングの違いから、パラメータを変更する度に出力が変化する形状プレビューが試行錯誤において有効であるとの意見も得られた。全ての参加者が OpticDesigner のプレビュー機能、シミュレーションの必要性を指摘していた。本研究の対象ユーザは、レンズ形状から得られる効果を直感的に予測できるほどの光学知識を持たないため、プレビュー機能の有効性が示されたといえる。

5.2.3 既存の撮影体験との違い

これまでの撮影では、汎用的なレンズやフィルターを個別のシーンに適用する形が一般的であり、量産品の中からニーズを満たすものを選択して撮影に臨んでいた。これに対し、OpticDesigner では、撮影者自身が特定のシーンや表現に合わせて設計・作製したレンズフィルターを撮影に用いる。この違いにより、参加者からは「設計・試作・撮影の試行錯誤を繰り返すことで表現を探索できる可能性を感じた」、「特定のシーン専用により作られたレンズフィルターを設計できれば、本システムの有効性を最大限に発揮できる」といった意見が得られた。これを裏付けるように、「OpticDesigner は一発勝負の現場には適さない可能性がある」という指摘も得られた。このような指摘は、従来のレンズやフィルターを用いた撮

影にはない考え方である。これは、量産品による汎用的な機材での撮影とは対照的に、OpticDesigner での撮影シーンや照明条件に応じた光学レンズ効果の調整によって、ユーザ自身のニーズに即した表現を生み出せる可能性を感じたことを示唆している。

6 制約

本稿では、従来の光学システムや 3D プリント光学素子に対して行われるような光学的な評価を実施していない。4 節と 5 節で示したように、OpticDesigner によって作製されるレンズフィルターの形状は多岐にわたる。そのため、特定の形状で得られた光学的評価と同様の結果が他の形状で得られるかは定かではないことから、光学的評価を実施しなかった。

また、OpticDesigner におけるソフトウェア上の制約として、効果適用範囲や形状設計の自由度が低いことがある。現状では、開口部は常にフィルター中央の一つ配置され、基本形状は事前に用意された 5 種類のみである。ワークショップ参加者からもこれらの制約に関する指摘が得られたことから、今後は複数の開口部を自由に配置できるインタフェースの開発と、複数形状の融合や基本形状以外の多様な形状を生成可能なアルゴリズムの実装を予定している。

さらに、光造形方式 3D プリンタによるレンズフィルター造形においては、スピンコーティングを開口部を持つ形状に適用するとコーティングむらが生じるという制約がある。これは、レジン滴下時に開口部を通じて非コーティング面へレジンを流出することに起因する。この課題に対しては、後処理時のみ開口部を一時的に塞ぐ逆形状のパーツを造形し、レジンの流出を防止する方法が有効である可能性がある。

7 おわりに

本稿では、ユーザ自身が求める光学レンズ効果を設計・作製可能にするレンズフィルターデザインシステム「OpticDesigner」を提案した。光造形方式 3D プリンタとスピンコーティング法を組み合わせることで、実用的な光学品質を実現しつつ、複数の効果の統合や適用範囲の制御を可能とした。さらに、写真・映像制作経験者を対象としたワークショップを通じて、ユーザが意図した効果を具現化できること、そして予期せぬ効果を含めた多様な表現が生まれることを示した。これにより、従来の汎用的な量産レンズでは実現できなかった、シーン固有のニーズに応じた撮影体験を提供できることが明らかとなった。

今後の研究では、複数の開口部を自由に配置できるシステムの実装とコーティング手法の改良によって、レンズフィルターの設計自由度と光学的品質を向上させる。また、スマートフォンやレンズ固定式カメラなどの撮影機器に対しても OpticDesigner を適用し、さらなるアプリケーションの拡大を目指す。

参考文献

- [1] M. Ali, F. Alam, N. Vahdati, and H. Butt. 3D-Printed Holographic Fresnel Lenses. *Advanced Engineering Materials*, 24(9):2101641, 2022.
- [2] X. Chen, W. Liu, B. Dong, J. Lee, H. O. T. Ware, H. F. Zhang, and C. Sun. High-Speed 3D Printing of Millimeter-Size Customized Aspheric Imaging Lenses with Sub 7 nm Surface Roughness. *Advanced Materials*, 30(18):1705683, 2018.
- [3] J. Christopher, L. M. Rooney, M. Donachie, D. Uttamchandani, G. McConnell, and R. Bauer. Low-cost 3D printed lenses for bright-field and fluorescence microscopy. *Biomed. Opt. Express*, 15(4):2224–2237, 2024.
- [4] D. Gonzalez-Utrera, B. Villalobos-Mendoza, R. Diaz-Urbe, and D. Aguirre-Aguirre. Modeling, fabrication, and metrology of 3D printed Alvarez lenses prototypes. *Opt. Express*, 32(3):3512–3527, 2024.
- [5] M. Hullin, E. Eisemann, H.-P. Seidel, and S. Lee. Physically-based real-time lens flare rendering. *ACM Trans. Graph.*, 30(4), 2011.
- [6] I. Marcks. Creating Custom Optics for *Oppenheimer*. <https://theasc.com/articles/optics-oppenheimer-sasaki>, 2023. (2025/10/23 参照).
- [7] L. M. Rooney, J. Christopher, B. Watson, Y. S. Kumar, L. Copeland, L. D. Walker, S. Foylan, W. B. Amos, R. Bauer, and G. McConnell. Printing, Characterizing, and Assessing Transparent 3D Printed Lenses for Optical Imaging. *Advanced Materials Technologies*, 9(15):2400043, 2024.
- [8] G. Spencer, P. Shirley, K. Zimmerman, and D. P. Greenberg. Physically-based glare effects for digital images. In *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '95, p. 325–334, 1995.
- [9] M. D. Tyona. A theoretical study on spin coating technique. *Advances in materials research*, 2:195–208, 2013.
- [10] Y. Zhu, T. Tang, S. Zhao, D. Joralmon, Z. Poit, B. Ahire, S. Keshav, A. R. Rajee, J. Blair, Z. Zhang, and X. Li. Recent advancements and applications in 3D printing of functional optics. *Additive Manufacturing*, 52:102682, 2022.

波紋表示による球面ディスプレイにおける視野外ポインティングの可視化

竹内 大空* Jana Hoffard* Hsueh Han Wu† Kelvin Cheng‡ 宮藤 詩緒* 小池 英樹*

概要. 本論文では、球面ディスプレイを用いたインタラクティブな遠隔コミュニケーションを実現するため、新しいポインティング手法として波紋型ポインティングを提案する。提案手法は、球面上に表示された360度映像内の指示位置を起点に同心波紋を等方的に伝播させ、観察者が可視領域に現れる波形から視野外の指示位置を推測できるようにするものである。本手法のために、全天周映像の球面マッピングと波紋生成・同期表示から成るUIを実装した。評価として、既存の局所マーカー提示法との比較による参加者実験を行い、視野外ポインティングにおける位置推定正答率・主観評価を指標に有効性を検証した。さらに、遠隔協調や案内支援などへの応用可能性についても言及する。

1 はじめに

遠隔コミュニケーションにおいて、参加者が同じ対象を同時に参照し共有するためには、広い視野の獲得と、指差しやハイライト等の指示を確実に伝えるポインティングが不可欠である。視野が狭い場合、対象の同定や位置関係の解釈に齟齬が生じやすく、またポインティングが不十分だと、言語化の負荷増大やタスク効率の低下につながる。

これらの課題に対し、入力(input)側では全天周カメラ(360度カメラ)を用いて環境全体を撮像し、受信側に広視野映像を提供することで、状況理解や相互参照を支援する研究が蓄積されている[12, 3, 15, 9, 7]。一方、出力(output)側の提示手法としては、全天周画像を球面ディスプレイに表示する取り組みがある[6, 14, 17]。この手法では、HMD(ヘッドマウントディスプレイ)で得にくい、複数人が同一の実体を囲みながら共有する、協調作業や、HMD等の装着負荷なく正確な相対位置関係を把握できる点が注目されている。球面という実体を介した提示は、身体配置(立ち位置や身体の向き)と視覚情報が直結するため、空間的合意参照に適している。しかし、球面上でのポインティングには固有の課題が残る。点光やマーカーを単純に表示しただけでは、球体の自遮蔽により裏面側の観察者からポインティング箇所が見えず、対象の位置を共有できない、という問題がある。

そこで本研究では、球面ディスプレイ上に「波紋表示(波紋型ポインティング)」を行い、ユーザーが球面上のどこを見てもポインティング箇所を推定できる手法を提案する。具体的には、指示点を起点として球面上を等方的に伝播する同心状のパターンを生成し、観察者は可視領域に現れる波紋の位相・

間隔・伝播方向から、視野外(裏側)にある指示点の方位と概略距離を推測することが可能となる。これにより、裏面に位置する参加者であっても、球面全体で一貫した指示位置を認識することが可能となる。本論文では、全天周映像を用いた遠隔コミュニケーションにおける球面ディスプレイ上の波紋型ポインティング手法を提案し、波紋の設計と球面上表示の実装方法を示す。さらに、参加者実験でポインティング精度を評価し、遠隔協調への応用例を紹介する。

2 関連研究

遠隔コミュニケーションにおいて、ポインティングは重要な役割を果たす。このため、さまざまなポインティング手法がこれまでに検討されてきた。その中でも、直感的かつ即応性に優れる手法として、触覚刺激を利用する方法が挙げられる[10]。しかし、多人数でのコミュニケーションにおいて触覚手法を用いる場合、すべてのリモートユーザがデバイスを装着する必要がある。これに対し、環境へのある程度のセットアップは必要とするものの、参加者に追加の装着を求めない代替手法として、ポインティングを視覚的に提示する方法が提案されている。たとえば、ビデオコミュニケーションにおいては、卓上に指さしの手を表示する方法[2, 16]や、デバイス画面上に重ねて表示する方法[4]が開発されている。

ポインティングの大きな課題の一つは、ユーザの視野外にある対象をどのように通知するかという点である。これに対処するため、2次元インタフェース上で視野外ターゲットを可視化する手法がいくつか提案されている。その一例として、ターゲットの周囲に円弧を表示し、ユーザがその中心を推定できるようにする方法がある[1]。その他にも、ARにおける放物線状の軌跡を用いる方法[13]、VRにおけるレーダー様のビジュアルモニタを用いる方法[8]、あるいは群れのような視覚的手がかりを提示し、ユー

Copyright is held by the author(s).

* 東京科学大学

† 楽天モバイル

‡ 楽天技術研究所

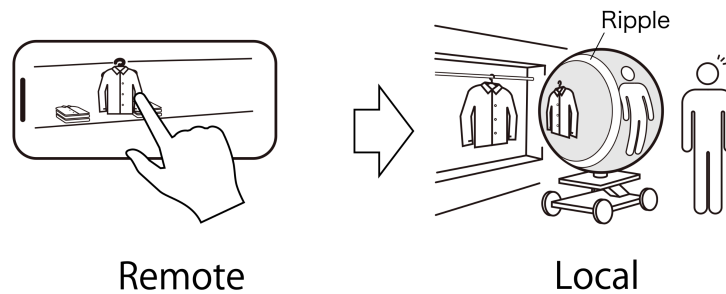


図 1. 波紋型ポインティングのイメージ図

ザの視線を誘導する方法 [5] が提案されている。

本研究では、球面ディスプレイと遠隔のモバイルデバイスとを組み合わせた非対称的なコミュニケーション場面を想定する。したがって、上述のようにユーザにポインティング位置を推定させることが可能な可視化ベースの手法が適していると考ええる。

3 提案手法

3.1 システムの概要

本研究は、球面ディスプレイと 360 度カメラを用いた、既存のマルチリモートコミュニケーションシステムが存在する ([11])。システムは、球面ディスプレイや 360 度カメラを備えたローカルサイドと、遠隔地からスマートフォンなどのデバイスを用いて参加する複数のリモートサイドから構成される。

リモートユーザは各々のデバイスから専用のアプリケーション（以下アプリ）にアクセスし、画面に表示されるローカルサイドのリアルタイム 360 度映像を視聴する。そして、アプリ内のジョイスティックを用いて、視点を操作し、ローカルサイドの球面ディスプレイを操縦する。また、アプリにはポインティング機能があり、画面をタップすることで、他のリモートユーザやローカルサイドへ注目点を伝達できる。

ローカルサイドでは、球面ディスプレイの上部に 360 度カメラを設置し 360 度映像を鏡像的に表示する。これらを遠隔操縦可能なオムニウィールに乗せる。リモートユーザのポインティングは球面ディスプレイ上の 360 度映像に重ねて表示し、その上部に顔の映像を配置する。オフラインのコミュニケーションでは、話者の顔や指をさす方向から注目点を推測できるが、本システムではポインティングに顔の映像を連動して表示させることで、これを再現している。

3.2 既存手法と波紋型ポインティングの提案

既存手法のポインティングは、リモートユーザのアプリ画面、球面ディスプレイ双方において、ユーザごとに色のついた正方形のマーカーを表示している。しかし、第 1 章でも述べた通り、球面ディスプ

レイは常に裏面を視認できず、既存の静的なポインティングではローカルユーザの視野外にあるそれを視認することができないという問題点があった。

そこで、リモートユーザがタップした箇所から白い円が波紋のように球面上を伝播していく波紋型ポインティングを提案する。ローカルユーザは、360 度映像上のあらゆる場所のポインティングに気がつけるだけでなく、視野外のポインティングの位置を推測することができる (図 1)。

4 実装

4.1 システム全体の実装

まず、システム全体の実装について概要を述べる。ローカルサイドのハードウェアを図 2(a) に示す。構成要素とその説明は以下の通りである。

- 360 度カメラ (RICOH THETA V)：ローカルサイドの映像を撮影する。
- 球面ディスプレイ：360 度カメラの映像やその他リモートサイドから送られた情報をリアルタイムで表示する。
- オムニウィール：リモートサイドからシステム全体を操縦できる。
- ポータブルバッテリー：球面ディスプレイやオムニウィールの電源。
- Arduino：オムニウィールの進行方向を制御する。
- ノートブック PC：360 度カメラで得られた映像を球面ディスプレイに表示する。またリモートサイドとの通信を行うとともに進行方向を Arduino に伝達する。

続いて、リモートサイドのアプリ画面を図 2(b) に示す。球面ディスプレイの映像や UI、リモートアプリは Unity で作成し、360 度映像や両サイドの情報を処理するサーバは Python で処理している。Unity プロジェクトと Python の通信は ZeroMQ による TCP/IP 通信を採用。音声通信とリモートユーザの顔映像は Zoom 経由としている。

波紋表示による球面ディスプレイにおける視野外ポインティングの可視化

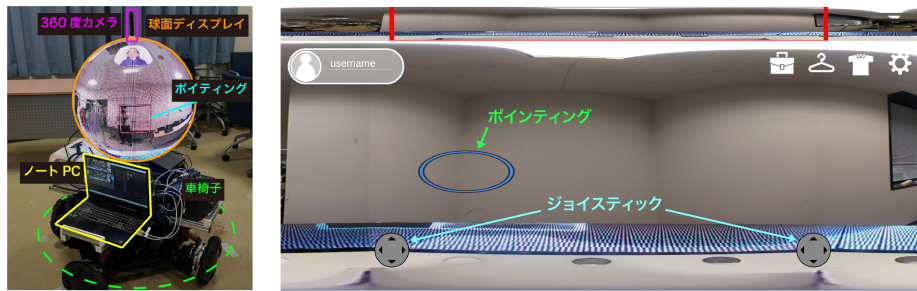


図 2. (a) ローカルサイドのハードウェア. (b) リモートユーザのアプリ画面.

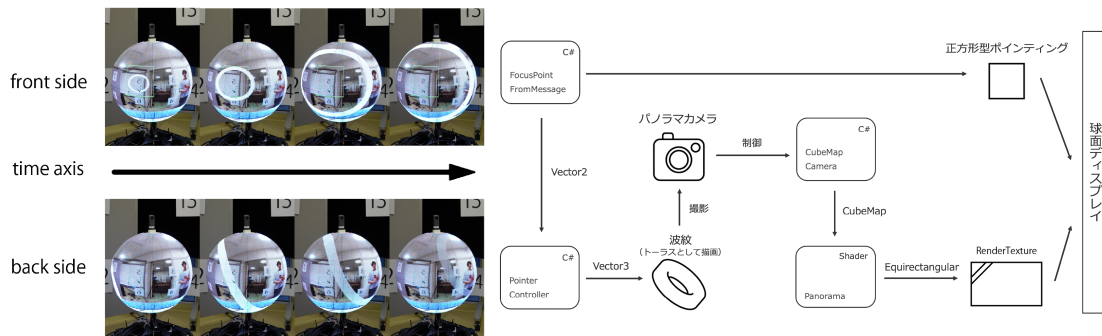


図 3. (a) 波紋型ポインティング. 伝播方向から指示点を推測できる. (b) 波紋型ポインティング生成のワークフロー.

リモートアプリでは、ローカルサイドの360度映像を画面上部に表示し、その内120度の視野をディスプレイ全体にリアルタイムで映し出す。右スティックで視点を操作、左スティックでオムニウィールを遠隔操縦し、上にあるシークバースクリーンで360度映像のうちリモートアプリ画面に表示されている範囲を赤枠で示している。リモートユーザは映像の任意の場所をタップすることで、ローカルユーザや他のリモートユーザの画面にポインティングを表示させることができる。

4.2 波紋型ポインティング生成の流れ

ローカルサイドのUnityプロジェクトにおいて、canvas内にはPlaneが配置されており、360度カメラから受け取ったequirectangular形式のパノラマ映像をRender Textureを用いてそこに表示している。そして、リモートユーザがタップした点のxy座標をサーバに送り、リサイズした後planeに被せるようにポインティングのプレハブを生成している。

続いて、波紋型ポインティング生成のワークフローを図3(b)に示す。Unity3D空間上の原点にカメラオブジェクトを配置し、得られたシーン映像をシェーダを用いてCubeMapからEquirectangular形式に変換、その映像を上記のPlaneに重畳して表示する。これにより、ローカルサイドの環境をスケール変換したメタバース空間としてUnity空間を扱うことができる。リモートサイドから受け取ったポインティング座標を三次元ベクトルに変換し、Unity空間上

にTorusプレハブを生成、原点を中心に球面の軌道を描くよう移動させる。torusは、始発点から2秒かけて収縮点に到達するスピードで球面の軸を等速直線運動し、1.2秒かけて徐々に透明度を下げる(図3(a))。これは、ポインティングの始発点と収縮点の差別化を図る為である。また、球面ディスプレイでの波紋表示と既存のマーカーによるポインティングを組み合わせを行っている。静的なポインティングと合わせて表示することで、ローカルユーザは波紋が伝播した後でも厳密な指示点を認識できる。

5 波紋型ポインティングの認識精度評価

5.1 実験概要

提案手法は、球面ディスプレイにおいてリモートユーザがどの方向からポインティングしているかを認識しやすくするために実装された。そこで、ポインティングの認識精度を評価するべく、四角形のポインティングのみの既存手法と、それに波紋型ポインティングを加えた提案手法の比較実験を行った。

実験を行った環境を図4に示す。壁で仕切られた幅3mほどの空間の中心にローカルサイドのハードウェアを配置し、その空間に被験者を立たせ、ポインティングを観察してもらった。壁には8方位上下2箇所の計16箇所に番号付きパネルを貼り付け、実験者がリモートユーザのアプリから特定の番号付きパネルをタップして球面ディスプレイ上にポインティングを表示させた。被験者は表示されたポイン

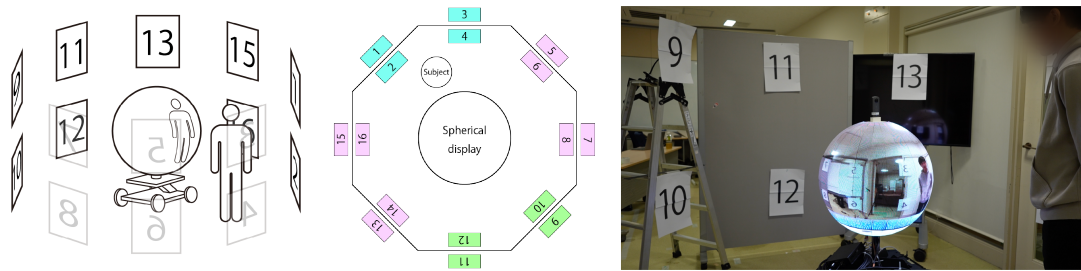


図 4. (a) 実験環境の全体図. (b) 実験環境を上から見た図. (c) 実験の様子.

Task1, Task2, Task3で2～5を繰り返す

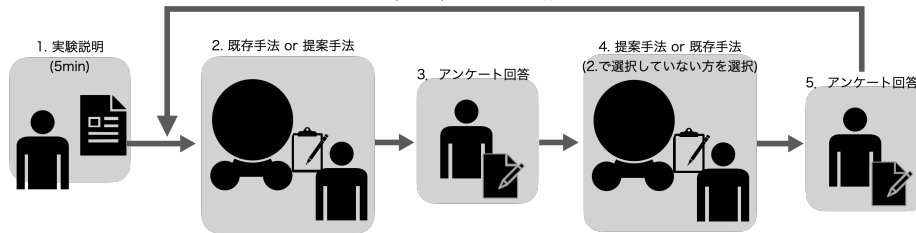


図 5. 実験のワークフロー

ティングを見て番号を記録してもらった。実験は条件を変えて3種類行った。実験のフローチャートを図5に示す。実験1では被験者を定位置（目印の1,2と3,4の中間にあたる場所）に固定し、足を動かさない範囲でポインティングを探してもらった。ポインティングは各番号を一度ずつ計16回、ランダムな順番で行われ、確認のために1つの番号につき二度ずつ行った。実験2では、被験者を実験1と同じ位置に固定した上で、記録するポインティングの数を2つに増やした。16つの番号を被験者を基準に正面（1,2,3,4）、側面（5,6,7,8,13,14,15,16）、背面（9,10,11,12）の3グループに分割し、グループごとの総当たり6回を試行した（図4(b)）。1秒間隔で2つの番号をタップした後、確認のため3秒間隔をあけてもう一度ポインティングを行った。実験3では、被験者を定位置に固定する制約を無くし、自由に動き回れる条件下で実験1を行った。しかし、スタートする位置は前と同じ1,2と3,4の中間にし、記録が終わるたびにそこに戻るよう指示した。被験者には自身のスマートフォンを持参してもらい、専用Googleformに記録してもらった。回答の選択肢には1から16の番号の他に「わからない」という欄を設け、当てずっぽうの回答は禁止にしたが、波紋の波形から見えない位置のポインティングを推測することは可とした。Googleformではその他に、各実験後には二つの手法における認識の容易さの違いについて、以下の質問に答えてもらった。

- 波紋の有無によってポインティングの認識のしやすさに変化はありましたか。具体的にお答えください。
- ポインティングが複数の場合、波紋の有無に

よる変化はありましたか。具体的にお答えください。

- 動き回れる状況下では、波紋の有無は認識に影響しましたか。具体的にお答えください。

また、実験後には感想欄とSUS（システムユーザビリティスケール）の設問を用意した。被験者にはそれぞれの手法に関する設問に答えてもらった。また、実験の様子を定点カメラで録画しながら実験を行った。

本実験は研究室内の学生と職員10人（男性9名、女性1名）を対象に行い、年齢は22歳から30歳（標準偏差=2.81）であった。実験順序を公平にするため被験者を二つのグループに分け、一方は既存手法を先に、もう一方は提案手法を先に行い、実験ごとに手法の順序を変えた。実験を実施するにあたって、東京科学大学の「人を対象とする研究倫理審査委員会」の倫理審査を受けた。

5.2 仮説

波紋型ポインティングは裏面の情報がわからないという球面ディスプレイの制約を克服するために考案したものである。したがって、以下の仮説を立てた上で本実験を行う。

仮説 H_1 ：提案手法によりポインティングの認識精度が改善。

既存手法である正方形のみのポインティングでは、ローカルユーザがポインティングに対し反対側にいる場合認識することができなかった。一方提案手法では、球面のあらゆる場所から波紋を視認し、正方形のポインティングからその場所を正確に認識する

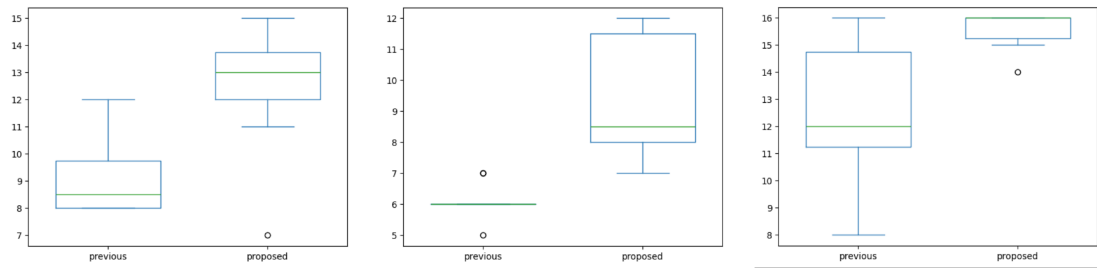


図 6. (a) 実験 1 (定位置単数探索) における正解数. (b) 実験 2 (定位置複数探索) における正解数. (c) 実験 3 (自由位置単数探索) における正解数.

ことができる。したがって、既存手法では視認できないポインティングの認識精度が向上すると予想する。

仮説 H_2 ：提案手法によりポインティングの場所が推測可能。

波紋型ポインティングの特徴として、ローカルユーザは波紋の形状を見て指示点の予想を立てることが可能になる。実験では視野外のポインティングの認識精度が向上し、被験者から推測の手助けになったという意見が得られることを期待する。

これら二つの仮説に対し、定量評価と定性評価の二つの観点から検証を行う。

5.3 実験結果

5.3.1 定量評価

各実験における既存手法と提案手法の正答率の箱ひげ図を図 6(a),(b),(c) に示す。

実験 1 では既存手法の平均値は 9.1, 中央値 8.5 であったのに対し、提案手法では平均値 12.4, 中央値 13.0 となった。実験 2 では既存手法の平均値は 6.1, 中央値 6.0 であったのに対し、提案手法では平均値 9.3, 中央値 8.5 となった。実験 3 では既存手法の平均値は 12.4, 中央値 12.0 であったのに対し、提案手法では平均値 15.6, 中央値 16.0 となった。それぞれの実験における、有意水準 $\alpha=0.05$ の Wilcoxon test の片側検定の結果を表 1 に示す。

表 1. Wilcoxon test の結果

実験	p value
実験 1	0.005859375
実験 2	0.001953125
実験 3	0.007526315

Wilcoxon test の結果から、全ての実験において p value が有意水準 $\alpha=0.05$ を下回った。

また、SUS スコアについては、提案手法は既存手法の平均値は 68.75, 中央値 73.75 であったのに対し、提案手法では平均値 83.75, 中央値 82.5 となった。Wilcoxon test により、両者間に有意差が認められた ($p = 0.03 < 0.05$)。

5.3.2 定性評価

本節では、実験における二つの手法の定性的評価を、フォームの設問や自由記述欄で得られた意見、動画分析により行う。

得られた意見 実験 1 に関しては、ポインティングの推測について言及する人が多かった。既存手法では認識できないポインティングでも、提案手法では波紋の形状から推測することができたという意見が多く得られた。しかし、上下の差異に関しては認識に苦戦したと述べた人も見受けられた (P1, P2, P10)。

実験 2 に関しては、ポインティングの推測に加え、認識のしやすさについて述べている人を見受けられた。タスクが増えた分、ポインティングを認識すること自体が難しくなったが、提案手法の場合はポインティングのタイミングを認識しやすかったという意見が得られた。中にはそれを活用し、1 回目のポインティングで大体の位置を認識し、2 回目でその周辺を注視することで詳細な位置を把握する、といった方法を確立している人もいた (P4)。

実験 3 に関しては、実験 2 と同様その認識のしやすさから、動き始めるタイミングや方向を掴みやすかったという意見が多く見受けられた。提案手法では、1 回目のポインティングで大体の位置が推測できるため、迷いなく探することができた様子だった。その他、提案手法で大体の方向を認識し、既存手法の正方形を視認することで確実性が増したという意見が見られた (P8, P10)。

動画分析 実験の様子を録画した定点カメラの映像をもとに、特徴的だった被験者の行動を紹介する。

- P1 は動けない条件の実験 1, 2 のにおいて、視野外のポインティングを目視しようと積極的に身を乗り出していた。提案手法においても、壁に貼られた目印ではなく、ディスプレイ上に表示された番号を視認しようとしていた。
- P5 は各実験の既存手法において、ポインティングが消えるタイミングがわからなかったため、ポインティングが終了した際に合図を求

めた。その後はその合図をもとに選択肢を選び、回答のペースが良くなった。

- P7は実験3の既存手法において、視野外のポインティングを球面に対し常に右回りで探索していた。一方提案手法では最短ルートになるよう右回り左回りを使い分けて着目点に向かっている様子だった。
- P10は実験3の既存手法において、動けるといふ条件下であっても途中からあまり動かず、視野外のポインティングはわからないを選択している様子だった。提案手法に関しては積極的に着目点を探索していた。

5.4 考察

Wilcoxon testの結果から、三種の実験全てにおいて、既存手法のポインティングの認識精度が向上していることがわかり、仮説 H_1 が示された。また、SUSスコアも既存手法と提案手法で有意差が認められた。SUSスコアの評価グレードは既存手法がC、提案手法がAであり、波紋型ポインティングがユーザビリティの観点でも優れていることを示している。さらに、アンケートや動画分析から、既存手法では視認できなかった背面のポインティングも、提案手法では押されているタイミングを認識しやすかったという意見が得られた。定性的な観点からも、提案手法は被験者のポインティング認識を補助していたと言える。

実験1と2に関して、被験者はその場を移動できないという条件下で行なっているため、視野外のポインティングに関しては推測する必要があった。二つの実験においてWilcoxon testにより両手法に有意差が認められたことは、波紋の形状によりポインティングの推測ができたことが一因と考えられる。定性評価においても、波紋の形状から位置を推定することができたとの意見が得られた。したがって、仮説 H_2 に関して、おおむね推測可能であると言える。しかし、アンケートでは、上下の判別に関しては推測に苦戦したという意見もあった。今回実験の環境による制限から、左右に比べて上下の数字間の間隔が近くなってしまったことが原因である。したがって、厳密なポインティング位置に関しては、本手法では判別が少し難しいということがわかった。

6 議論

本研究の提案手法は球面ディスプレイ上の全方位に対応可能なポインティング認識補助システムである。既存の静的なポインティングでは、球面ディスプレイの利用者は裏面の情報を認識することが困難であったが、提案手法により、視野外のポインティングを視認し指示点の位置を推測することが可能となった。本手法は、球面ディスプレイを用いたコミュ

ニケーション支援、特にマルチユーザの遠隔コミュニケーションにおいて、利点を活かすことができると考える。

一つ目の例として、球面ディスプレイを用いたオンラインショッピングシステムへの応用が考えられる。客はローカルサイドである実店舗の様子をリモートから視聴し、店員とコミュニケーションをとる。波紋型ポインティングにより、店員は一度に複数の客の着目している商品を把握することができる。一度に複数の客と対応することができ、マルチリモートショッピングのクオリティをより向上できると期待する。

同じような一対多のコミュニケーションのシーンとして、観光案内への応用を提案する。ローカルサイドのハードウェアを観光地に配置し、リモートから参加するユーザとポインティングを使って新たな気づきを共有することができる。ツアーガイドが本システムを利用することで、観光客が逸れる等の心配がなくなるほか、提案手法により一方的な説明ではなく相互的なコミュニケーションが実現すると考える。

しかし、今回実装した波紋型ポインティングは、視野外の正確な着目点が予測できるわけではない。実験1は実験3と比べ認識精度が落ちていたことから、着目点を視認できない条件下では認識精度が低下してしまうことが問題点として挙げられる。また、提案手法において、波紋が経過する際に他のUI表示を邪魔してしまう可能性がある。ユーザ同士のポインティングの衝突や、他のUIとの衝突により、どちらかのUIが認識しづらくなるおそれがある。

今後の展望としては、他手法のポインティングおよび多岐にわたる分野への応用を模索したい。例えば、虫眼鏡のようにポインティングした箇所を拡大するような手法では、詳細な位置を伝えられるのではないかと考える。また、他分野の応用に関しては、今プロジェクトの特徴である一対多のコミュニケーション構造を当てはめられるサービスを考えていきたい。

7 謝辞

本研究は、楽天モバイルとの共同プロジェクト研究により実施されたものである。

参考文献

- [1] P. Baudisch and R. Rosenholtz. Halo: a technique for visualizing off-screen objects. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, p. 481–488, New York, NY, USA, 2003. Association for Computing Machinery.
- [2] D. Iwai, R. Matsukage, S. Aoyama, T. Kikukawa, and K. Sato. Geometrically

- Consistent Projection-Based Tabletop Sharing for Remote Collaboration. *IEEE Access*, 6:6293–6302, 2018.
- [3] S. Johnson, I. Rae, B. Mutlu, and L. Takayama. Can You See Me Now? How Field of View Affects Collaboration in Robotic Telepresence. New York, NY, USA, 2015. Association for Computing Machinery.
- [4] B. Jones, A. Witcraft, S. Bateman, C. Neustaedter, and A. Tang. Mechanics of Camera Work in Mobile Video Collaboration. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, p. 957–966, New York, NY, USA, 2015. Association for Computing Machinery.
- [5] S. Kergaßner, N. Doerr, M. Wieland, M. Fuchs, and M. Sedlmair. HiveFive360: Extending the VR Gaze Guidance Technique HiveFive to Highlight Out-Of-FOV Targets. In *Proceedings of Mensch Und Computer 2024*, MuC '24, p. 11–20, New York, NY, USA, 2024. Association for Computing Machinery.
- [6] Z. Li, S. Miyafuji, E. Wu, H. Kuzuoka, N. Yamashita, and H. Koike. OmniGlobe: An Interactive I/O System For Symmetric 360-Degree Video Communication. In *Proceedings of the 2019 on Designing Interactive Systems Conference*, DIS '19, p. 1427–1438, New York, NY, USA, 2019. Association for Computing Machinery.
- [7] Y.-T. Lin, Y.-C. Liao, S.-Y. Teng, Y.-J. Chung, L. Chan, and B.-Y. Chen. Outside-In: Visualizing Out-of-Sight Regions-of-Interest in a 360° Video Using Spatial Picture-in-Picture Previews. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, UIST '17, p. 255–265, New York, NY, USA, 2017. Association for Computing Machinery.
- [8] S. J. Liu, R. Henrikson, T. Grossman, M. Glueck, and M. Parent. RadarVR: Exploring Spatiotemporal Visual Guidance in Cinematic VR. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, UIST '23, New York, NY, USA, 2023. Association for Computing Machinery.
- [9] M. Manabe, D. Uriu, T. Funatsu, A. Izumihara, T. Yazaki, I.-H. Chen, Y.-Y. Liao, K.-Y. Liu, J.-C. Ko, Z. Kashino, A. Hiyama, and M. Inami. Exploring in the City with Your Personal Guide: Design and User Study of T-Leap, a Telepresence System. In *Proceedings of the 19th International Conference on Mobile and Ubiquitous Multimedia*, MUM '20, p. 96–106, New York, NY, USA, 2020. Association for Computing Machinery.
- [10] A. Matsuda, K. Nozawa, K. Takata, A. Izumihara, and J. Rekimoto. HapticPointer: A Neck-worn Device that Presents Direction by Vibrotactile Feedback for Remote Collaboration Tasks. In *Proceedings of the Augmented Humans International Conference*, AHs '20, New York, NY, USA, 2020. Association for Computing Machinery.
- [11] S. Miyafuji, J. Hoffard, S. Takeuchi, Y. Makimoto, H. H. Wu, K. Cheng, and H. Koike. Remote Online Shopping System with 360-Degree Streaming and Spherical Display. Workshop in VR-HSA. IEEE, 2025.
- [12] J. Norris, H. Schnädelbach, and G. Qiu. CamBlend: an object focused collaboration tool. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, p. 627–636, New York, NY, USA, 2012. Association for Computing Machinery.
- [13] H. Oshimi, M. Perusquía-Hernández, N. Isoyama, H. Uchiyama, and K. Kiyokawa. LocatAR: An AR Object Search Assistance System for a Shared Space. In *Proceedings of the Augmented Humans International Conference 2023*, AHs '23, p. 66–76, New York, NY, USA, 2023. Association for Computing Machinery.
- [14] L. Takagi, S. Miyafuji, J. Pardomuan, and H. Koike. LUNACHair: Remote Wheelchair System Linking Users to Nearby People and Assistants. In *Proceedings of the Augmented Humans International Conference 2023*, AHs '23, p. 122–134, New York, NY, USA, 2023. Association for Computing Machinery.
- [15] A. Tang, O. Fakourfar, C. Neustaedter, and S. Bateman. Collaboration with 360° Videochat: Challenges and Opportunities. In *Proceedings of the 2017 Conference on Designing Interactive Systems*, DIS '17, p. 1327–1339, New York, NY, USA, 2017. Association for Computing Machinery.
- [16] N. Yamashita, K. Kaji, H. Kuzuoka, and K. Hirata. Improving visibility of remote gestures in distributed tabletop collaboration. In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work*, CSCW '11, p. 95–104, New York, NY, USA, 2011. Association for Computing Machinery.
- [17] T. Yonezawa, N. Hayashida, K. Urano, J. Przybilla, Y. Kyono, and N. Kawaguchi. MetaPo: A Robotic Meta Portal for Interspace Communication. In *ACM SIGGRAPH 2022 Posters*, SIGGRAPH '22, New York, NY, USA, 2022. Association for Computing Machinery.

高速投影と時間方向の色分解による視線移動方向に応じた映像提示手法

宮崎 竜輔* 宮藤 詩緒* 小池 英樹*

概要. 本論文では、ユーザの視線の移動方向に応じて知覚される映像が変化する映像表示手法を提案する。本手法によって知覚される映像は、視線が静止している状態では視線静止時の目的画像のみであるが、視線を特定の方向に移動させながら映像を閲覧することにより、埋め込まれた視線移動時の目的画像を視認することができる。また、上下左右のうち異なる2方向について、それぞれの方向への視線移動時に異なる画像を提示することも可能である。本手法は、人間の臨界融合周波数を超える高速での映像表示が可能な高速プロジェクタ等のデバイスと、特定の方法での画像の分割手法により実現される。本論文では、観察者に目的の色を知覚させる手法や色、解像度、表示速度に関する制約を表示実験を通して議論する。また、各ユーザの視線状態に応じて異なる映像を提示するアプリケーション例を紹介する。

1 はじめに

ディスプレイやプロジェクタでの情報提示は社会において様々な場面で使用されている。特に公共の場など不特定多数の閲覧者が存在する空間においてこれらのデバイスを用いて情報を提示することには、単一のデバイスで多くの閲覧者に同時に情報を提示できるという利点がある。なかでも、プロジェクタは曲面など複雑な面への投影が可能である他 [3, 21], ディスプレイも構造物の壁面や床といった部分に埋め込むことで、構造物の表面への映像投影が可能となる。カメラやセンサを用いて閲覧者をトラッキングすれば、インタラクティブな情報提示を行うことも可能となる。

しかし、一般的なディスプレイでは、異なる閲覧者に対してそれぞれに対応した異なる映像を提示することは困難である。ヘッドマウントディスプレイでは、各閲覧者ごとにデバイスを装着させることにより各個人への異なる情報提示が可能だが、不特定多数の閲覧者を対象とする場合は、閲覧者の増加に伴い必要となるデバイス数が増加する問題が生じる。また、カメラやセンサなどによるトラッキング例についても、特に不特定多数の閲覧者がいる公共空間においては、精度やプライバシーに関する懸念がある。これらの課題を解決する方法として、トラッキングなどを用い映像そのものを閲覧者に合わせて変化させるのではなく、映像自体は変化しないが閲覧者の状態によって知覚される内容が変化する表示手法が考えられる。本論文では、特に閲覧者の視線の動きに着目する。

本論文では、各観察者に個別の情報表示を目的として、表示映像が観察者の視線が特定の方向に移動したときにのみ特定の画像を提示する新たな映像表

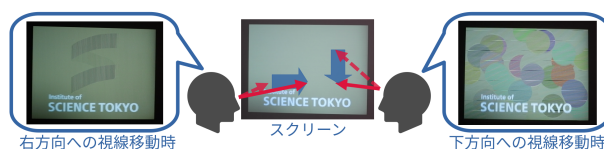


図 1. 本提案の概要。最大2方向の視線移動に対して、それぞれの移動方向に応じた異なる画像を提示することが可能となる。

示手法を提案する(図1) 本手法では、観察者の視線が静止している状態時は「静止時目的画像」のみが知覚される。しかし、視線を特定の方向に移動させながら映像を閲覧すると、「視線移動時目的画像」を視認することができる。さらに、上下左右のうち異なる2方向について、それぞれの方向への視線移動時にそれぞれ異なる視線移動時目的画像を提示することも可能である。本手法は、視線移動方向に応じて映像が変化する手法 [13, 14] を発展させ、従来手法では不可能であったカラー表示への対応を行う。具体的には、従来同色領域に対して適用していた手法を改良し、特定のピクセル領域ごとに「視線移動時目的画像」の色と、その補正色を用いたフレームに分解し、高速プロジェクタなどの高速表示が可能なデバイスを用いて映像の表示を行う。これにより、カメラやセンサ等のトラッキングを用いずに複数の閲覧者にそれぞれ異なる情報を提示することが可能となる。

2 関連研究

2.1 視線を用いたインタラクション手法

プロジェクタによる投影映像やディスプレイによる表示映像とのインタラクション手法としては、ハンドジェスチャをトラッキングするものや [2], 他の物体やデバイスを入力として利用するもの [20, 23], プロジェクタユニットそのものの動きを入力とする

Copyright is held by the author(s).

* 東京科学大学

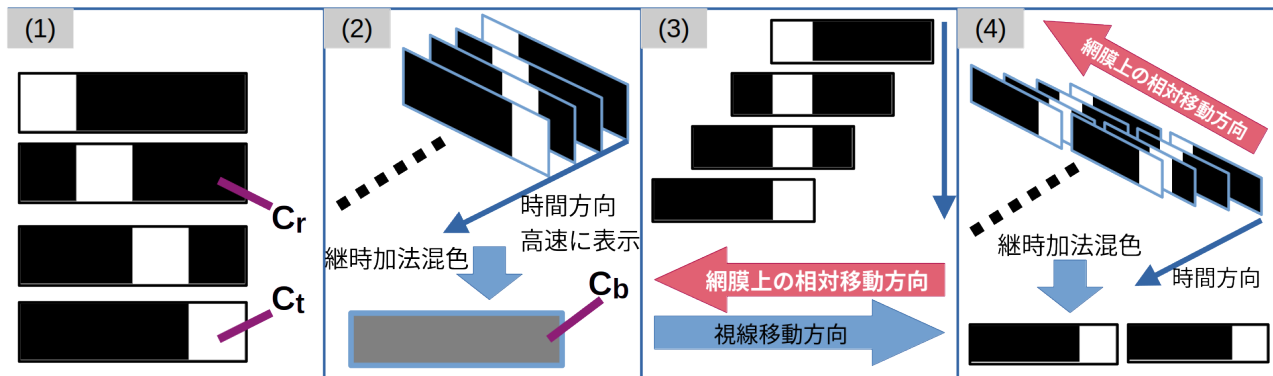


図 2. 色分解手法の説明.

もの [16, 9] など多数の方法が研究されてきた. これらの例では, ユーザやデバイスの動作を検出する方法として, 主にカメラなどを用いたトラッキングが使用される.

インタラクション手法として, 視線を利用する方法が存在する [25, 6]. この方法は, 身体の大きな動きを必要としない. そのため, 両手が塞がっている状態での入力手法や, VR 環境における入力手法として身体のジェスチャを用いた入力と併用するなどして利用される [7, 11]. また, 視線の動きはユーザの意識を顕著に反映するため, ユーザの注意や興味を分析し利用することができる [4]. これらの手法は基本的に専用のアイトラッカを利用し視線を検出することで実現されている. キャリブレーション不要な手法も研究されているが, 多くは専用のデバイスとキャリブレーションが必要であり, 同時に体験することができる人数に制限がある. 本提案手法も, 視線のなめらかな移動を利用したインタラクションを可能としているが, トラッキングを行わず, 視線の動きに応じて知覚される画像が直接変化する. また, 視線移動を直接インタラクション手法として用いているのではなく, 視線移動を原理の一部として利用しているという点で従来とは異なる立ち位置である.

2.2 トラッキングを使用しない手法

上記の手法では, 映像とのインタラクションには, トラッキングなどを併用する必要であった. Sakaue ら [19] は, トラッキングなどを用いずに, プロジェクタと投影先のスクリーンの距離の変化に応じて映像が変化する投影手法を提案している. これは, 本提案と同様移動方向などに応じて映像を変化させることができる一方, 移動方向はスクリーンの法線方向に限定されている.

Ikeda ら [8] は, 投影映像を左右に特定の速度で移動させることにより知覚される投影映像を変化させる手法を提案している. これも本提案と似たコンセプトであるが, 静止時には意味を持たない模様が投影されてしまい, 静止時の情報提示を妨げるという問題があった.

他には, レンチキュラーレンズなどを用いスクリーンに特殊な加工を施すことや [22], ディスプレイの前に無数の穴をあけたバリアを設置すること [15, 10] などにより, 見る角度により異なる映像を提示するものがある.

2.3 高速表示を利用したディスプレイ

500Hz 以上といった高速で映像を提示することが可能なプロジェクタやディスプレイが存在する. 高速投影の利点として, 人間の視覚特性を利用したアプリケーションが可能となる点がある. 人間が感知できる光の点滅の周波数の限界は臨界融合周波数 (CFF) と呼ばれており, 様々な条件にもよるがおよそ 60Hz である [5]. この速度を十分に超えた投影を行うことにより様々な映像表示やインタラクションが可能となる [17]. Phyxel は, 高速投影と回転する質感を持ったスクリーンを用いて投影に質感を提示している [28]. Kohtani らは, 通常の映像投影に加えて, 映像の一部を特定のパターンに投影色を分解し投影を行うことで, 映像の一部に視線誘導を行っている [12]. また, Tosa らは同様に知覚限界に近い速度での色の点滅を用いて視線誘導を行っている [24]. また, 高速な視線移動であるサッケード運動を利用し, サッケード運動により 1 次元表示デバイスを用いて 2 次元画像を知覚させる例がある [1, 26].

3 提案

本論文では, 観察者の視線移動方向に応じて異なる画像を提示する映像投影手法を提案する. 本提案の概要を図 1 に示す.

3.1 継時加法混色と視線移動時の知覚

人間の視覚は一般に 60Hz 程度である CFF を超える高速で複数の色刺激を提示された場合, それぞれの色を峻別することはできず, 一定期間に提示された色刺激を混色し知覚する [5]. そのため, 図 2(1) のような n 枚の画像フレームを, 1/60 秒間に等間隔で順番に高速に表示した場合, 図 2(2) のように各画素は時間方向に平均化され知覚される. この高

速表示を行う際に、視線を左から右になめらかに移動させる場合を考える。この際、人間の網膜上の座標系では、表示フレームが相対的な右から左に移動する。図2(3)のようにこの相対的な移動速度と、表示映像における白色部分の仮想的な移動速度が一致した場合、図2(4)のように時間方向に混色が行われるため、白黒の模様が出現したように知覚される。この手法を用いることにより、視線の静止時と特定方向への視線移動時で異なる模様を提示することが可能となる。

3.2 色分解手法の拡張

上記手法を、任意の色を提示するために拡張する。以下、色の表現には赤緑青それぞれ0～255の8ビット階調を使用する。視線移動時に顕在化させることを目的とするターゲット色を $C_t = (R_t, G_t, B_t)$ とし、視線静止時に提示することを目的とする背景色を $C_b(R_b, G_b, B_b)$ とする。 C_t を打ち消す補正部分の色を $C_r = (R_r, G_r, B_r)$ とする。使用するフレームの枚数を n ($n \geq 3$) とした場合に、これらの間には以下の式が成立する。

$$\frac{C_t + (n-1)C_r}{n} = C_b \quad (1)$$

ここで、表示するフレームに使用される色である C_t, C_r は、 $0 \leq R_t, G_t, B_t < 256$ 及び $0 \leq R_r, G_r, B_r < 256$ である。 C_t について、 R_t, G_t, B_t が下限である0のとき、 R_r, G_r, B_r の範囲より、

$$0 \leq R_b < 256 \times \frac{n-1}{n} \quad (2)$$

となる。 R_t, G_t, B_t が上限である255のとき、

$$\frac{255}{n} \leq R_b < \frac{256(n-1) + 255}{n} \quad (3)$$

となる。これらの共通範囲より、この範囲において C_b は以下の範囲に制限される。

$$\frac{255}{n} \leq R_b, G_b, B_b < 256 \times \frac{n-1}{n} \quad (4)$$

この制約内において、視線移動方向に n ピクセル分の領域を、視線静止時は C_b 、視線移動時は1ピクセルの C_t と $n-1$ ピクセルの C_r の領域に分解することが可能となる。

上記の分解手法を用い、目的とする画像の各画素値に応じた色に分解することで、視線移動により知覚される目的画像を映像中に埋め込むことができる。ただし、1ピクセルの C_t と $n-1$ ピクセルの C_r を提示した際に、人間の視覚やカメラのセンサがその領域を何色であると知覚するかについては、表示例の章にて説明するが、 C_t の反対色であるように知覚することが示唆されている。そのため、目的画像の各画素の値に対して、その反対色を C_t の値とし

て採用することにより、目的画像を埋め込むことが可能となる。詳細な手順については実装の章にて説明する。

3.3 2方向の視線移動への対応

上記手法により、ある1方向の視線移動が行われている際にのみ目的画像を提示することが可能となった。次に、2方向の視線移動に対してそれぞれ異なる画像を提示する方法を説明する。まず、各視線移動方向と対応する方向に色 C_t の領域が移動するように分解した2種類の画像セットを用意する。そして、これらをそれぞれ順番に交互に出力する。これにより、各視線移動方向に対応する目的画像を表示することが可能となる。

3.4 なめらかな視線移動の誘発

本提案手法はなめらかな視線移動を前提としている。しかし、一般に人間はある位置から別の位置へ視線を移動しようとした際に、サッケード運動が発生する[30]。これはなめらかな動きではなく、瞬時に注視点が移動し、また移動中の視覚の情報はサッケード抑制により知覚されない。そのため、他の方法で視線移動を誘発する必要がある。この方法には、意図的な視線移動と意図しない視線移動の2種が考えられる。意図的な視線移動の誘発方法としては、表示面上に移動する点を重畳表示しその点を注視する、指で表示面をなぞりその指を注視するといった方法が考えられる。これらの例のように移動する物体を追跡する場合、注視点は物体の0.7～0.9倍程度の速度で移動し、位置の差は定期的にサッケード運動で補完される[18]。意図しない視線移動の誘発方法としては、表示デバイスや映像そのものが急に動く場合[29]や、スマートフォンなどの手に持った別の物体に注視した状態で歩くなど、他の物体に注視しながら観察者や表示デバイスが移動するなどの例が考えられる。これらを利用したアプリケーション例について、アプリケーション例の章で議論する。

4 実装

4.1 ソフトウェア

本提案手法をもとに、ある静止時目的画像と視線移動時目的画像に対して、それぞれを視線の静止時と移動時に知覚させるような処理方法を説明する。処理の手順は図3に示してある。まず、分割数 n 及び対応する視線移動方向を決定する。次に、静止時目的画像が制約条件(4)を満たすように、各画素値 c を $0 \leq c < 256$ から $255/n \leq c' < 256 \times (n-1)/n$ にスケールする。視線移動時目的画像は各画素を反対色に変換する。反対色とする理由は考察の章にて述べる。次に、静止時目的画像と視線移動時目的画像の両方を、視線移動方向が長辺となるように $1 \times n$



図 3. 目的画像からの画像分解手順.

ピクセルの長方形領域に分割し、その領域の代表色を領域内の色の平均で定める。そして各領域に対して、 C_b を静止時目的画像におけるその領域の代表色、 C_t を色反転後の視線移動時目的画像におけるその領域の代表色として定め、 C_r を計算し、提案手法により n 枚のフレームに分解する。

4.2 ハードウェア

本提案では、CFF を超える非常に速い速度での映像の出力が可能であるデバイスが必要となる。本実装では、投影速度の柔軟性や実装の容易さなどの理由から高速プロジェクタ DynaFlash [27] を使用した。これは、 1024×768 ピクセル 24bit フルカラー画像を最大およそ 900Hz の高速で投影が可能なデバイスである。

5 表示例

5.1 画像分割数及び埋め込み枚数と視認性の関係

本提案手法による表示の実現性及び各パラメータと視認性の関係を確認するために、画像の分割数 n 及び視線移動の方向数 x を変え投影を行った。投影周波数 F は、分割したフレームの 1 周期分を $1/60$ 秒以内に投影することを目安に、 $F = 60 \times n \times x$ と定めた。投影した映像をカメラを動かしながら撮影した。

5.2 撮影方法

投影及び撮影の様子は図 4 に示した。撮影には X-A7(FujiFilm) を使用し、シャッタースピードを人間の視覚に近づけるため $1/60$ 秒とし、ISO 感度 2500、露出補正 0 とした。カメラを対応する方向に振ることで視線移動を再現し、その様子を動画で撮影した。

6 考察

投影結果を図 5 に示した。この結果より、以下の示唆が得られた。



図 4. 投影環境のセットアップ.

6.1 分割数 n 、視線移動方向数と視認性の関係

分割数 n を大きくするほど、視線移動時目的画像の視認性が向上した。しかしその一方で、目的画像の処理時に n に比例する大きさの領域に画像を分割するため、 n を大きくするほど画質が荒くなり、文字などの識別可能性は低下する。また、 n に比例して CFF を考慮した投影周波数は増大するため、投影デバイスの最大出力可能周波数というハードウェア面の制約も考慮する必要がある。視線移動方向の数については、2 方向埋め込み時は 1 方向埋め込み時と比較して視線移動時画像の視認性が低下しており、また投影周波数も倍となるため、よりハードウェア面の制約を受ける。

6.2 C_t と視認される色の関係

図 5 左の移動時元画像を $1 \times n$ ピクセルの領域に分割し各領域内で平均をとったものが C_t であるが、投影の結果知覚される色はおよそその反対色となった。そのため、画像の前準備段階で提示したい目的画像の反対色画像を作成しそれをもとに C_t を計算する必要があることが示唆された。このように反対色が知覚された理由としては、画像分解手法の関係上色が C_r となる領域が存在することが関係していると考えられる。 C_r は視線移動時の目的色 C_t を打ち消して視線静止時の目的色 C_b となるように計算されているため、背景画像次第ではあるが基本的に C_t の反対色の明度を下げた値に近くなる。投影の

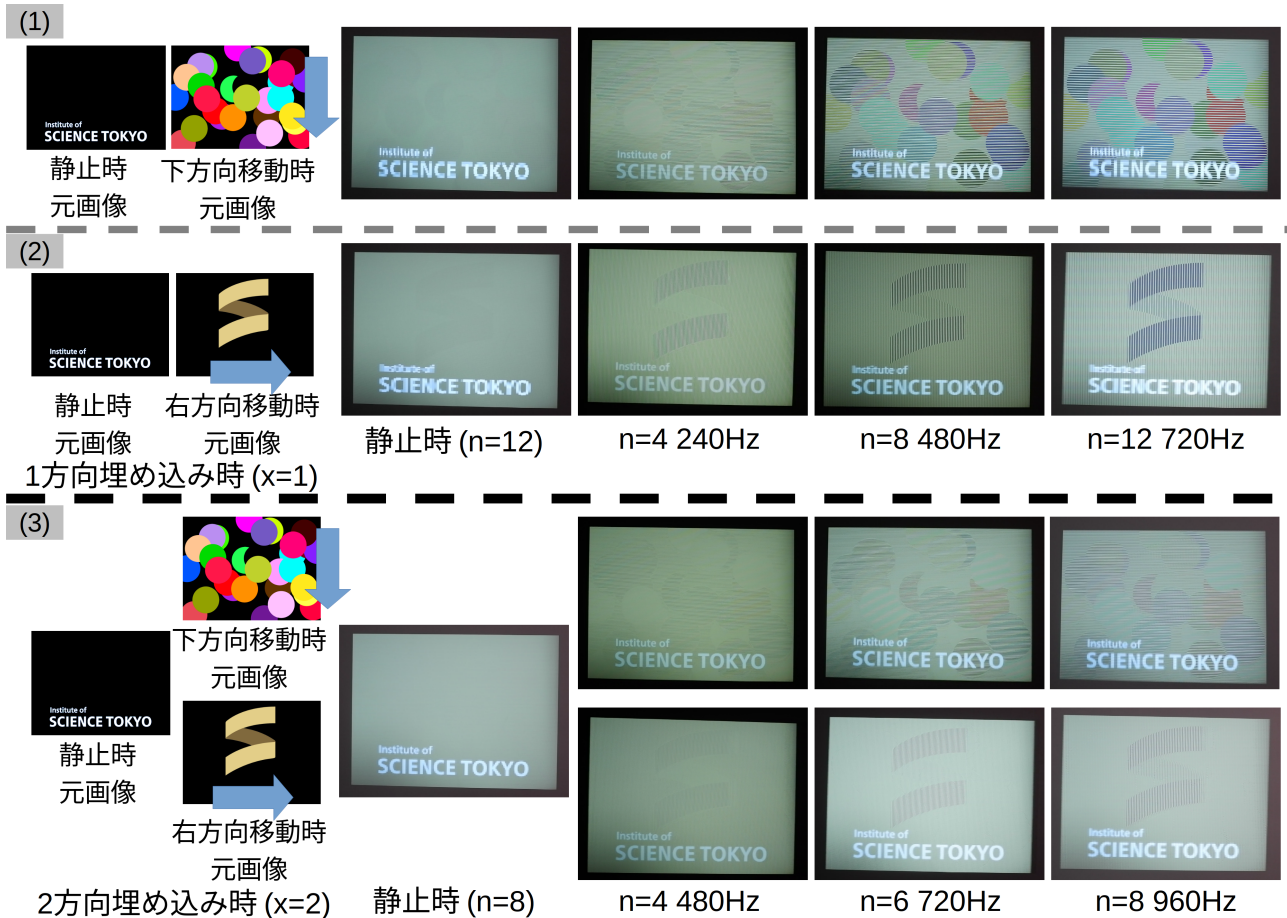


図 5. 投影結果。ただし、視線移動時元画像は色反転後の状態を表示している。(1) 下方の視線移動に対して知覚される画像のみを埋め込んだ場合の結果。(2) 右方向の視線移動に対して知覚される画像のみを埋め込んだ場合の結果。(3) 下及び右方向に対応する画像を同時に埋め込んだ場合の結果。

明るさや視線静止時の画像などの条件次第で C_r が C_t と比較してより強く知覚されたため反対色のように見えたと考えられる。これらがどの程度関係しているかの解明は今後の課題である。

7 制限

本提案で提示できる画像について、視線移動時、視線静止時の提示画像両方について、 $1 \times n$ ピクセルの領域に分割するため解像度が下がる他、視線静止時の画像については式 (4) の色の制約が存在する。また、視線移動時に知覚される色が元画像の色をどれだけ再現できているかどうかは検証の必要がある。

本提案では高リフレッシュレートでの映像表示が可能なデバイスとして高速プロジェクタを選択した。そのため、本提案がゲーミングディスプレイといった高速なディスプレイ装置でも可能かどうかは今後の課題である。特に、表示の明るさなどの影響によって C_r と C_t のどちらがより強く知覚されるかにより目的の色を提示する際の最適な分解手法は変化する可能性がある。

また、本提案では上下左右から 2 方向について、

それぞれの方向への視線移動時に知覚できる画像の埋め込みを実験したが、3 方向以上については未検証である。また、斜め方向の視線移動に対応する場合は、基本的に立方体である画素に対して、どのように斜め方向の領域に分解するかなどの課題が残る。

8 アプリケーション案

本提案手法は、視線移動により生じる網膜上の相対的な映像の移動によって知覚される画像が変化する。この相対的な移動を意図的に誘発する方法及び意図せず誘発する方法を用いたアプリケーション案をそれぞれ述べる。

8.1 意図的な視線移動

8.1.1 指を用いた映像とのインタラクション

このアプリケーション案は、通常時は文字が見えないが、指でスクリーンをなぞりながらその指を見つめることで画像を知覚することができるというものである [13]。このアプリケーションの概要を図 6 左上に示す。このアプリケーション案では、あるユーザが指でスクリーンをなぞり、それを複数名でみた

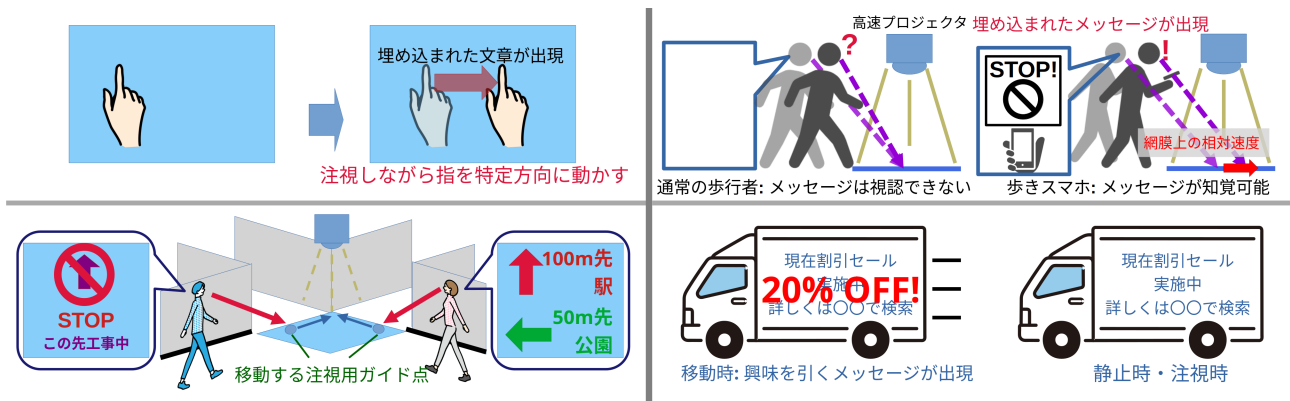


図 6. アプリケーション案の概要.

場合でも、指に視線を合わせているかどうかで画像を知覚できるかどうかが変わる。これにより、スクリーン上を指でなぞることで画像情報を取り出すというインタラクションが生まれるほか、視線の合わせ方で異なる人に異なる映像を同時に同位置に提示することが可能である。このアプリケーションにおいて、ユーザは指を動かすことで、能動的に視線移動を誘発し、文字を知覚するかどうかを選択することが可能となる。

8.1.2 交差点などでの情報提示

また、異なる複数方向への視線移動で異なる情報を提示できることを利用し、交差点など、不特定多数の人が異なる方向へ移動する場所の床などに情報を投影することが考えられる。概要を図6左下に示す。これにより、異なる方向へ移動するユーザにそれぞれに対応した情報を提示することができる。交差点の床などに本提案手法を配備し、各移動方向ごとに視線移動を誘発する移動点を複数重畳表示する。ユーザは、どの移動点を注視するかによって知覚する映像を選択することができる。本提案は、投影している映像は変化させることなく、ユーザの状態により知覚される情報が変化する。そのため、同一位置にある情報でも、ユーザの動きや視線によって異なる知覚を与えることができる。そのため、同一位置に、対象となるユーザによってそれぞれ異なる情報を提示することが可能となる。

8.2 意図しない視線移動

8.2.1 歩きスマホ者向けディスプレイ

また、本提案手法を床や壁などに配備することにより、歩きスマホを行っている歩行者にのみ情報を提示することが可能となる [14]。この例の概要を図6右上に示す。歩きスマホを行っている歩行者は、手元のスマートフォンに視線を固定している。その状態で歩行することにより、背景は相対的になめらかに後ろへ移動しているようにみなすことができる。一方で、通常の歩行者が配備箇所を見つめた場合でも、視線が固定されなめらかな移動は発生しない。

そのため、歩きスマホを行っている歩行者のみが映像を知覚できる。この例は、意図しない視線移動を利用し知覚される映像を変化させる例である。ユースケースとして、歩きスマホへの警告や、移動方向などの誘導などを行うサインを表示する例が考えられる。

8.2.2 車両の側面などでの広告表示

車両の側面など、移動する物体に本提案手法を適用することにより、移動しているときと停止しているときで周囲の閲覧者に異なる情報を提示することができる。概要を図6右下に示す。例として、車両が走行中は大きな文字など移動中でも読みやすい情報を投影して、停止時はその文字が消え細かい説明が見えると言ったものが考えられる。ユーザが興味を持ち車両に視線を合わせた場合は、ユーザから見た相対速度は0となるため、車両の側面に描かれていた広告などの細かい文字を集中して読むことができる。車両が走行中の場合は、ユーザが車両を視認してから車両を追従できるまでの時間に、埋め込まれた映像を意図せず知覚できる。そのため、大きな字で目を引くような情報を埋め込み、広告表示に重畳するといった用途が考えられる。

9 結論と展望

本論文では、視線移動方向に応じて異なる画像を知覚させることができる映像表示手法を提案した。今後の展望として、表示実験を通して示唆を得た表示パラメータと視認性の関係などについて、表示手法の評価とユーザ実験を行う予定である。

謝辞

本研究は JSPS 科研費 JP25K03164 の助成を受けた。

参考文献

- [1] H. Ando, J. Watanabe, T. Amemiya, and T. Maeda. Full-scale saccade-based display:

- public / private image presentation based on gaze-contingent visual illusion. In *ACM SIGGRAPH 2007 Emerging Technologies*, SIGGRAPH '07, p. 5-es, New York, NY, USA, 2007. Association for Computing Machinery.
- [2] H. Benko and A. D. Wilson. Multi-point interactions with immersive omnidirectional visualizations in a dome. In *ACM International Conference on Interactive Tabletops and Surfaces*, ITS '10, p. 19–28, New York, NY, USA, 2010. Association for Computing Machinery.
- [3] C. Boletsis. The Gaia System: A Tabletop Projection Mapping System for Raising Environmental Awareness in Islands and Coastal Areas. In *Proceedings of the 15th International Conference on Pervasive Technologies Related to Assistive Environments*, PETRA '22, p. 50–54, New York, NY, USA, 2022. Association for Computing Machinery.
- [4] G. Buscher, A. Dengel, and L. van Elst. Query expansion using gaze-based feedback on the sub-document level. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '08, p. 387–394, New York, NY, USA, 2008. Association for Computing Machinery.
- [5] H. de Lange Dzn. Research into the Dynamic Nature of the Human Fovea→Cortex Systems with Intermittent and Modulated Light. I. Attenuation Characteristics with White and Colored Light. *J. Opt. Soc. Am.*, 48(11):777–784, Nov 1958.
- [6] H. Drewes and A. Schmidt. Interacting with the computer using gaze gestures. In *Proceedings of the 11th IFIP TC 13 International Conference on Human-Computer Interaction - Volume Part II*, INTERACT'07, p. 475–488, Berlin, Heidelberg, 2007. Springer-Verlag.
- [7] A. Esteves, E. Velloso, A. Bulling, and H. Gellersen. Orbits: Gaze Interaction for Smart Watches using Smooth Pursuit Eye Movements. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, UIST '15, p. 457–466, New York, NY, USA, 2015. Association for Computing Machinery.
- [8] H. Ikeda, T. Hayakawa, and M. Ishikawa. Bilateral Motion Display: Strategy to Provide Multiple Visual Perception Using Afterimage Effects for Specific Motion. In *Proceedings of the 25th ACM Symposium on Virtual Reality Software and Technology*, VRST '19, New York, NY, USA, 2019. Association for Computing Machinery.
- [9] N. Joshi, A. A. R. Irudayaraj, J. Hartmann, and D. Vogel. An Instrumented Office Chair with a Steerable Projector for Personal Spatial Augmented Reality. In *Proceedings of the 2022 ACM Symposium on Spatial User Interaction*, SUI '22, New York, NY, USA, 2022. Association for Computing Machinery.
- [10] A. Karnik, W. Mayol-Cuevas, and S. Subramanian. MUSTARD: a multi user see through AR display. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, p. 2541–2550, New York, NY, USA, 2012. Association for Computing Machinery.
- [11] J. Kim, S. Park, Q. Zhou, M. Gonzalez-Franco, J. Lee, and K. Pfeuffer. PinchCatcher: Enabling Multi-selection for Gaze+Pinch. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, CHI '25, New York, NY, USA, 2025. Association for Computing Machinery.
- [12] A. Kohtani, S. Miyafuji, K. Urakaki, H. Kat-suyama, and H. Koike. MOSion: Gaze Guidance with Motion-triggered Visual Cues by Mosaic Patterns. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI '24, New York, NY, USA, 2024. Association for Computing Machinery.
- [13] R. Miyazaki, S. Miyafuji, K. Kim, and H. Koike. A Display Method Visible Only During Gaze Fixation on Moving Objects using High-Speed Projection and Striped Pattern. In *Adjunct Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology*, UIST Adjunct '25, New York, NY, USA, 2025. Association for Computing Machinery.
- [14] R. Miyazaki, S. Miyafuji, and H. Koike. Be Smart-Phone Zombie!: Guidance Display for Texting While Walking Using Striped Pattern and High-Speed Projection. In *Adjunct Proceedings of the 27th International Conference on Mobile Human-Computer Interaction*, Mobile-HCI '25 Adjunct, New York, NY, USA, 2025. Association for Computing Machinery.
- [15] A. Nashel and H. Fuchs. Random Hole Display: A non-uniform barrier autostereoscopic display. In *2009 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video*, pp. 1–4, 2009.
- [16] B. Patnaik, H. Peng, and N. Elmqvist. Vis-Torch: Interacting with Situated Visualizations using Handheld Projectors. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI '24, New York, NY, USA, 2024. Association for Computing Machinery.
- [17] J. Rekimoto. Temporal Phase Shift: 時間位相差を持つプロジェクションシステムによる錯視現象とそのモデル化. *WISS2009*, p. 0028, 12 2009.
- [18] D. A. Robinson, J. L. Gordon, and S. E. Gordon. A model of the smooth pursuit eye movement system. *Biological Cybernetics*, 55(1):43–57, oct 1986.
- [19] F. Sakaue and J. Sato. Active 3D Motion Visualization Based on Spatiotemporal Light-Ray Integration. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1977–1985, 2020.
- [20] T. Sato, D.-H. Hwang, and H. Koike. Mli-oLight: Projector-camera Based Multi-layered Image Overlay System for Multiple Flashlights Interaction. In *Proceedings of the 2018 ACM International Conference on Interactive Surfaces*

- and Spaces*, ISS '18, p. 263–271, New York, NY, USA, 2018. Association for Computing Machinery.
- [21] T. Sato, J. Pardomuan, Y. Matoba, and H. Koike. ClaytricSurface: An Interactive Deformable Display with Dynamic Stiffness Control. *IEEE Computer Graphics and Applications*, 34(3):59–67, 2014.
 - [22] A. Shirai, H. Suzuki, Y. Yamaguchi, and K. Yanaka. Glassless augmented display for public Signage. In *Proceedings of the 2016 Virtual Reality International Conference*, VRIC '16, New York, NY, USA, 2016. Association for Computing Machinery.
 - [23] M. Spindler, S. Stellmach, and R. Dachsel. PaperLens: advanced magic lens interaction above the tabletop. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, ITS '09, p. 69–76, New York, NY, USA, 2009. Association for Computing Machinery.
 - [24] R. Tosa, S. Hattori, Y. Hiroi, Y. Itoh, and T. Hiraki. ChromaGazer: Unobtrusive Visual Modulation using Imperceptible Color Vibration for Visual Guidance, 2024.
 - [25] M. Vidal, K. Pfeuffer, A. Bulling, and H. W. Gellersen. Pursuits: eye-based interaction with moving targets. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '13, p. 3147–3150, New York, NY, USA, 2013. Association for Computing Machinery.
 - [26] J. Watanabe, T. Maeda, and H. Ando. Gaze-contingent visual presentation technique with electro-ocular-graph-based saccade detection. *ACM Trans. Appl. Percept.*, 9(2), June 2012.
 - [27] Y. Watanabe, G. Narita, S. Tatsuno, T. Yuasa, K. Sumino, and M. Ishikawa. High-speed 8-bit Image Projector at 1,000 fps with 3 ms Delay. 12 2015.
 - [28] T. Yoshida, Y. Watanabe, and M. Ishikawa. Phyxel: Realistic Display of Shape and Appearance using Physical Objects with High-speed Pixelated Lighting. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, UIST '16, p. 453–460, New York, NY, USA, 2016. Association for Computing Machinery.
 - [29] 日本視覚学会. 視覚情報処理ハンドブック. 朝倉書店, 東京, Japan, 2017.4 2017.
 - [30] 日本視覚学会. 図説視覚の事典. 朝倉書店, 東京, Japan, 2022.11 2022.

Dishplay：シリアル磁気応答を利用した可食ディスプレイシステム

大津 陽翔* 山本 匠* 杉浦 裕太*

概要. 本稿では、鉄分強化シリアルを用いた可食ディスプレイシステム「Dishplay」を提案する。鉄分強化シリアルは磁気応答性を有しており、磁石に引き付けられる。この性質を利用した情報ディスプレイとして、永久磁石を土台に配置することでパターン表示する静的なディスプレイ、リニアアクチュエータアレイを用いて磁石を上下させることによる動的ディスプレイを実装した。提案手法に関して、シリアル量や液体の種類・量、時間経過などが表示精度に与える影響やパターン遷移に要する時間を調査した。システムの応用例として、祝祭的なメッセージの提示や実用シナリオについて議論した。

1 はじめに

HCI 研究では、食体験をより豊かにすることを目的として、食品をメディアとして活用する研究が行われている [7]。これらの研究は、視覚や聴覚、さらには触覚などの多様なモダリティを活用することで、食体験をインタラクティブな体験へと拡張してきた [6]。中でも、視覚提示は、食品の見た目を変化させることで味覚に影響を与える [22]、データを可視化することでエクササイズなどのモチベーションを向上させる [17] などの例が提案されている。

より多様な視覚情報を食卓に重畳する方法として、プロジェクション技術が注目されている。プロジェクション技術は、食品や食器の表面に映像を付与できるという利点を有し、例えば、子どもの食習慣の改善などに利用されている [5, 31]。しかし、照明条件に制約されるほか、食品自体の動的な視覚的变化を直接的に生じさせるものではない。

食品そのものを情報提示媒体として視覚的变化させる研究も行われている。例えば、レンチキュラ効果を持つ寒天 [29] や、光の当て方によって模様が変化するクッキー [28] などが提案されている。しかし、依然として表現力や制御性に制限がある。

本研究では、鉄分強化シリアルを用いたディスプレイを提案する。市販の鉄分強化シリアルに含まれる金属鉄の磁気応答性に着目し、液体中のシリアル片を磁場により動かすことで、食品自体を動的な視覚ディスプレイとして機能させる (図 1)。本研究ではリニアアクチュエータアレイを用いた動的ディスプレイと永久磁石を手動で配置する静的ディスプレイの 2 つを実装した。加えて、液体の種類などの要素が表示精度に与える影響と、パターン遷移に必要な時間について調査した。

本稿の貢献は、(1) 市販食品に含まれる鉄を活用した動的な可食ディスプレイ手法の提案、(2) 表示

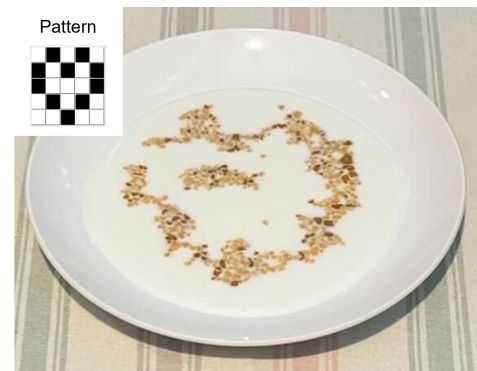


図 1. ターゲットパターン (左) と、Dishplay によって生成された実際のパターン (右)。

品質やパターン遷移に必要な時間に影響を与える要因の知見の報告、(3) 本手法の制御可能性と表現力の観点から整理した今後の課題の提示、となる。

2 関連研究

2.1 Human-Food Interaction (HFI)

近年、食に関するインタラクションを探索する HFI 研究が拡大している。従来、食事は単に消費される対象として扱われてきたが、HFI の文脈では、食を用いたインタラクションの設計が行われている。例えば、3D プリンティングによる食品の構造制御 [12, 19, 20, 23]、動く食品の開発 [25, 4]、音響出力 [3, 26] や振動 [2]、光 [1]、AR・VR [27, 21] を用いた体験拡張、食べられるセンサの提案 [16] など、多様なアプローチで研究が進められている。

2.2 食品への視覚提示

HFI におけるアプローチのひとつに、視覚提示を用いた手法がある。これは大きく「プロジェクションを用いた手法」と「食品自体を利用する手法」に分けられる。

Copyright is held by the author(s).

* 慶應義塾大学

プロジェクタを用いて食事環境や食品表面に映像を重ねる研究は多数存在する [32]. 特に、動きを伴う視覚的な演出によって食体験の拡張を図る研究が進められている. 例えば、食事のプロセスをゲーム化することで子どもの食習慣を改善する研究がある. 例えば、FoodWorks は、野菜が食べられるとお皿の上に報酬のバッジを投影し [5], FunEat は、アニメーションを通じて偏食の改善を促す. [31]. また、味覚の知覚に影響を与えるため、動的な視覚提示を行う研究も行われている. Suzuki らは、カレーに「ぐつぐつと煮える」映像を投影し、食欲や塩味の知覚を増強させる効果を報告した [24]. Huisman らは、ヨーグルトの周囲に投影する映像が、酸味の評価に影響を与えることを見出している [9].

これに対して、食品自体の物理的特性を活かして視覚的变化を実現する研究も存在する. Yoshimoto らは、透明なゼリーをレンチキュラーレンズ状に加工し、視点に応じて始点に応じて異なる情報を提示する手法を提案した [29, 30]. Yamamoto らは、クッキー生地特定の角度で微細な穴をあけ、焼成後の凹凸が作り出す影によって視点依存の像を形成する手法を提案した [28]. さらに、液体そのものを表示媒体とする研究も行われている. Ishii と Shiio は、電気分解を用いて発生する泡の集合体を制御し、ディスプレイとして活用する手法を提案した [15]. 同氏らはこの技術を発展させ、泡による高解像度なグレースケール表示を実現している [13]. 加えて、電気分解で生じるイオンと指示薬の化学反応による発色を利用し、食品表面などに直接カラーパターンを描画する手法も提案している [14].

本稿の提案は、市販化された食品素材そのものを活用した可食ディスプレイとなり、動的にパターンの変更が可能となる.

2.3 食品への鉄分添加に関する動向

食品への鉄分添加（鉄強化）は、世界的な鉄欠乏症対策として重要な役割を担っている. 一般的な食品に含まれる鉄は、動物性食品由来のヘム鉄と、植物性食品や多くの強化食品に用いられる非ヘム鉄に大別される [8, 11]. これらは人の体内で吸収されるが、強い磁気応答性を示さないため、外部磁場による操作には不向きである. 一方、シリアルのような乾燥・長期保存が求められる食品では、金属鉄である還元鉄が利用されている. 還元鉄は、味や色への影響が少なく安定性が高い上に、製造コストが比較的低い利点から、強化成分として使用される [10, 18].

3 Display

3.1 原理

関連研究で述べた還元鉄は、他の鉄化合物と異なり、特有の物理的特性を持つ. 還元鉄は酸化や錯形

成をしていない金属鉄粒子であり、強磁性や常磁性を保持しているため、外部磁場に対して可視的に応答することができる.

本研究では、1 g あたり 0.46 mg の還元鉄を含有する General Mills Multi Grain Cheerios¹ を材料として選択した. 予備実験の結果、シリアルを細かく粉砕することで、金属鉄粒子の表面積が増加し、磁場への応答性が大きく向上することが確認された. さらに、液中での沈降を防ぐため、粉砕後のシリアル粉末をふるいにかけ、極端に微細な粒子を除去して使用した. 粉砕後の粒子の大きさは一辺 1.0 mm から 7.6 mm 程度である. 磁石は、円柱型のネオジウム磁石（直径 12 mm、高さ 3 mm、磁束密度約 89.74 mT）を用いて、静的ディスプレイと動的ディスプレイの2つを実装した. また PET 樹脂素材の皿² を使用し、80 ml, 120 ml, 160 ml の水を入れた時の皿が置かれた面からの高さは 10.6 mm, 12.3 mm, 13.8 mm である.

3.2 静的ディスプレイ

静的ディスプレイシステムは、あらかじめ磁石を固定配置し、鉄分強化シリアルの粉砕片を磁力で吸着させることで、パターンを表示する構造である. 静的システムは小型で複製や準備が容易かつ低コストであるため、簡便な食材ディスプレイとして有用である. 5 × 5 のマトリクスに対応した薄型土台を 3D プリントにより製作し、ネオジウム磁石を任意のセル（一辺 21 mm）に配置可能とした. これにより、磁石の位置にシリアル片が吸着し、任意の静的パターンを形成できる.

3.3 動的ディスプレイ

動的ディスプレイは、25 個のリニアアクチュエータによって各磁石の高さを個別に制御し、液体表面に浮遊するシリアル片を動的に操作する. 本システムは、パターン入力用の GUI（図 2（A））と、ハードウェア機構（図 2（B））から構成されている. ユーザは GUI 上の 5 × 5 のグリッド形式で表示パターンを入力でき、表示状態を切り替えることが可能である. さらに複数のパターンをラベル管理し、選択・編集が可能である. 作成したパターンは「Send to Arduino」ボタンを押すことでシリアル通信でハードウェアへ送信される. これにより、液体の中で鉄分強化シリアルを用いたパターン表示が可能となる（図 2（C））.

制御系は Arduino Uno を用いて実装した. Arduino はシリアル通信でグリッドデータを受信し、I2C プロトコルを介して各モータへ制御信号を送信する. 25 台のサーボアクチュエータは PCA9685 PWM ドライバによって同時に駆動されており、それ

¹ <https://amzn.asia/d/6VZF5A>

² <https://www.nitori-net.jp/ec/product/8961864>

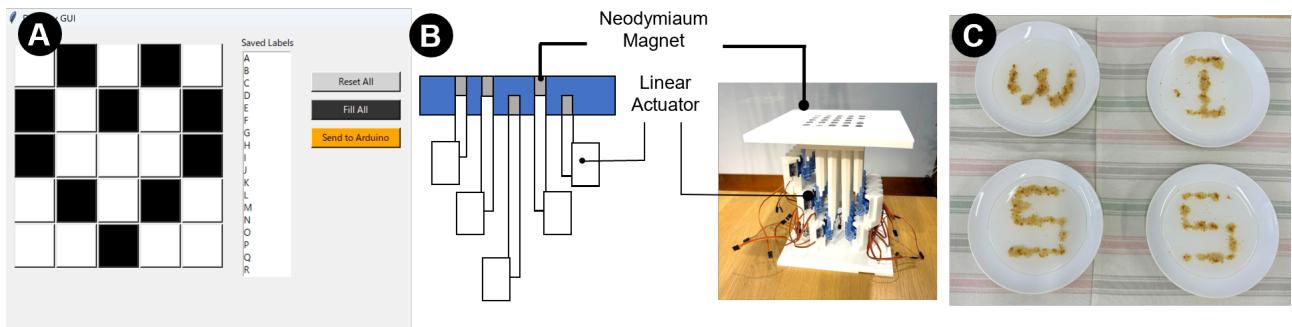


図 2. (A) パターンデザインの GUI, (B) ハードウェア構造, (C) パターン「WISS」

それぞれがグリッドの1セルに対応している。PCA9685は安定化された外部電源（MCH-K3010D）から給電され、Arduino本体はホストPCのUSB給電により動作している。

モーターに関しては、Kitronik Linear Actuator Kitsを利用した。このキットはサーボモータとラック・ピニオン機構が含まれており、サーボモータはラック・ピニオン機構を介してサーボモータの回転運動を直線運動に変換することができる。その先端にはネオジム磁石5つ（高さ $3 \times 5 = 15$ mm）を装着している。25個のKitronik Linear Actuator Kitsを 5×5 の格子状に配置している。この磁石が上下動することで、液体表面に浮かぶ粉碎した鉄分強化シリアル片を磁力で局所的に操作し、動的なパターン形成を実現している。セルの寸法は一辺21 mm、装置全体の高さは約一辺300 mmの卓上設計で実現している。各アクチュエータは3Dプリントで作られた凹凸構造を活かした組み合わせ設計であるフレームで垂直方向に固定されている。

4 表示精度の調査

4.1 手順

水やシリアル量の量などの表現に影響を与える可能性のあるパラメータを変更し、撮影した画像を用いてディスプレイとしての精度を評価した。これにより、各パラメータが表示精度に与える影響を定量的に把握することを目的とした。評価対象として、 5×5 の白いマス目からなるハート形のパターン（図3（C））を作成し、このパターンをDishplayの静的ディスプレイシステムによって再現した。

比較するパラメータには以下の7つを採用し、ベースライン条件から1因子ずつ変更して実験を行い、条件ごとに5回行った。ベースライン条件は以下の因子のうち太字の通り設定した。

1. シリアル散布後の撮影までの待機時間（**1分**, 3分, 5分）
2. 液体添加量（80 ml, **120 ml**, 160 ml）

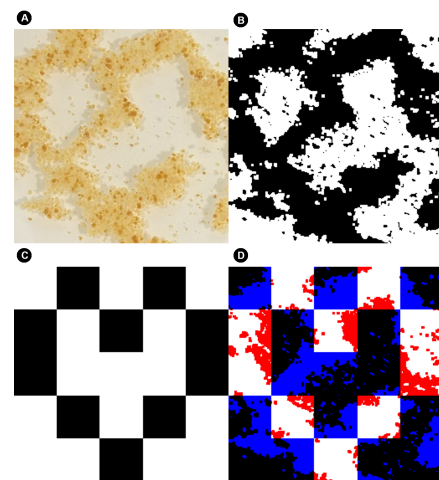


図 3. (A) リサイズされた撮影画像 (B) 前処理後の撮影画像 (C) 比較するターゲット画像 (D) 類似度の視覚化（白：両方白、黒：両方黒、赤：ターゲット黒・撮影白、青：ターゲット白・撮影黒）

3. シリアル投入量（0.5 g, **1.0 g**, 1.5 g）
4. シリアル散布方法（**均一散布**, イラストに沿った散布）
5. 外部振動の有無（**なし**, 指による机への振動操作）
6. シリアルの浸漬時間（**0分**, 15分）
7. 液体種類（**水**, 牛乳）

ディスプレイ再現後、撮影条件（照明、距離、角度）は全実験で統一し、上方からiPhone14で撮影した。

撮影した画像は比較しやすいように以下の前処理を行った。まず、画像サイズを 500×500 ピクセルに統一した（図3（A））。次に、局所的なコントラスト強調（CLAHE: Contrast Limited Adaptive Histogram Equalization）を適用した。局所的な明暗差により黒・白の境界が不明瞭になる部分に対し、

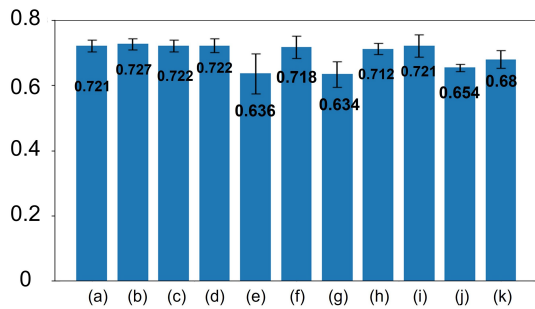


図 4. 各パラメータにおける類似度の平均と標準偏差.
(a) ベースライン, (b) 待機時間 3 分, (c) 待機時間 5 分, (d) 液体量 80ml, (e) 液体量 160ml, (f) シリアル量 0.5g, (g) シリアル量 1.5g, (h) イラストに沿った散布, (i) 指による振動 (j) 浸漬時間 15 分, (k) 牛乳.

暗い領域や明るい領域のディテールが強調され、ターゲットパターンの黒塗り部分と白背景の判別が容易になる。さらに、コントラスト強調後のグレースケール画像に対して、Otsu 法による自動二値化を適用した。Otsu 法は、画像のヒストグラムを基に自動で閾値を算出し、ピクセルを黒 (0) または白 (255) に分類する手法である。最後に、二値化により生じる小さなノイズや点状の誤判定を除去するため、モルフォロジー処理の「開運動態 (opening)」を適用した (図 3 (B))。

撮影画像の再現精度を定量化するために、類似度 (黒・白両方を含む一致率) を指標として用いた (図 3 (D))。類似度は、ターゲットパターンと再現画像の各ピクセルについて一致した場合を 1、不一致の場合を 0 として集計し、全ピクセル数で割った値として定義される。さらに、画像を 5×5 のセルに分割し、セル単位で類似度を算出した。このセル単位類似度をヒートマップとして可視化することで、パターンのどの領域が高精度で再現されているかを直感的に把握することが可能となる。

4.2 結果・考察

各パラメータの全体類似度の平均と標準偏差を図 4 に、セルごとの類似度の平均のヒートマップを図 5 に示す。

4.2.1 ベースライン

類似度の平均は 0.721、標準偏差は 0.018 となった。他の条件と比較しても高くかつ安定した精度が得られており、その結果、本実験におけるパラメータ設定の工夫が有意に反映されていない可能性が示唆される。この結果は、サンプル数の少なさによりベースラインの性能が過大評価されたか、あるいは本手法においては類似度が概ね 0.72 付近で頭打ち

となる傾向があるのか、さらなる検討を要する。

4.2.2 待機時間

シリアルを散布してから撮影までの待機時間を 3 分および 5 分に延長した結果、ベースラインと比較して類似度および標準偏差に顕著な変化は見られなかった。当初は、磁石からの引力を長時間受け続けることでシリアルがよりパターンに沿うと想定していたが、結果は異なった。この要因として、ベースラインの精度が既に高いこと、あるいは本実験における評価手法が適切でなかった可能性が考えられる。なお、類似度指標では差異が確認されなかったものの、実験者による目視観察では、撮影までの待機時間を延長することで再現されるパターンがより精緻化されていく傾向がわずかに感じられた。

4.2.3 液体量

液体量を 80 ml に減少させた場合、ベースラインと比較して類似度および標準偏差に顕著な変化は見られなかった。当初は、液体量が減少することでシリアルと磁石との距離が短縮され、クーロンの法則に基づき磁力が強まると想定され、精度の向上が期待されたが、結果は異なった。この要因については、待機時間の検討で考察した内容と同様の可能性が示唆される。さらに、ヒートマップにおいても類似した傾向が見られたことから、同様の要因が作用したと考えられる。

液体量を 160 ml に増加させた場合は、ベースラインと比較して類似度が低下し、標準偏差が増加した。これは、クーロンの法則に基づきシリアルが磁石から受ける磁力が弱まったためであると推察される。また、標準偏差が大きくなったことから、再現性の低下も確認された。なお、類似度は 0.636 と必ずしも低い値ではなかったが、実験者による目視観察では、多くの場合においてハート形状のパターンは判別困難であった。

4.2.4 シリアル量

シリアル量を 0.5 g に減少させた場合、ベースラインと比較して類似度に顕著な変化は見られなかったが、標準偏差は増加した。ヒートマップからは、磁石が配置されているセルにおける類似度が低く、それ以外のセルでは高い傾向が確認された。このことから、シリアル量が不足しているため必要なセルに十分に集まらず、一方で不要なセルにも集まりにくい状況が生じたと推察される。また、標準偏差が大きくなったことから、シリアルの分布に関して再現性が低下していることが示唆される。さらに、類似度はベースラインと同程度であったものの、セルの欠落が散見されたため、実験者による目視観察では、数値上の類似度よりも再現性が低く判断された。

一方、シリアル量を 1.5 g に増加させた場合、ベー

Dishplay: シリアル磁気応答を利用した可食ディスプレイシステム

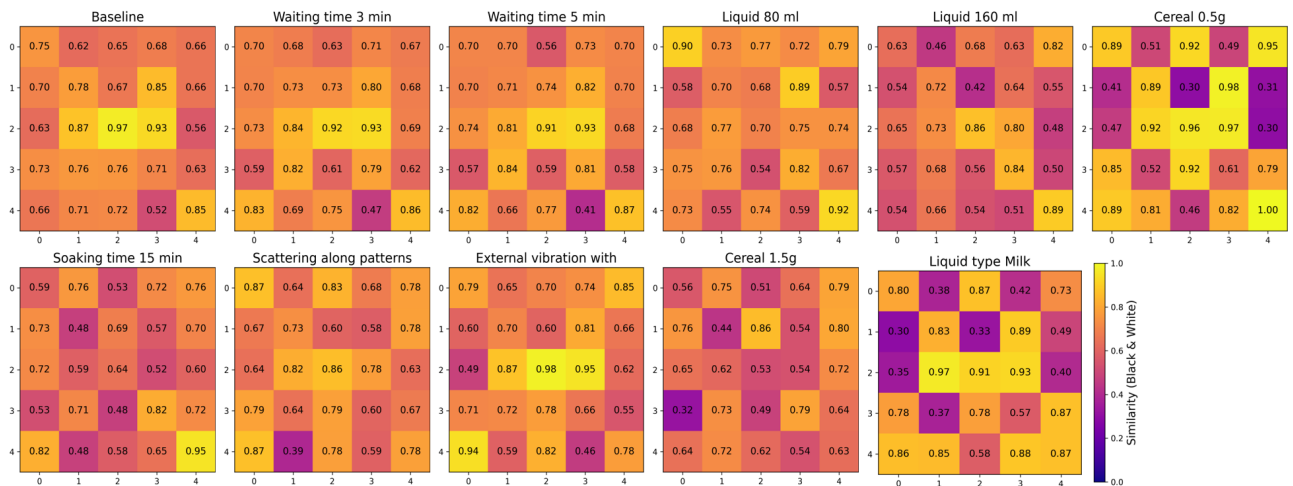


図 5. 各パラメータにおけるセルごとの類似度の平均ヒートマップ。(黄色に近いほど高精度、紫に近いほど低精度)

スラインと比較して類似度は低下し、標準偏差は増加した。ヒートマップの解析では、磁石が配置されているセルにおいて類似度が高く、それ以外のセルにおいて低い傾向が見られた。この現象は、シリアル量が過剰であるために、0.5 g の場合とは逆の挙動を示したと考えられる。また、標準偏差が大きいことから再現性が低く、160 ml の条件と同様に、実験者による目視観察では、数値上の類似度よりも明らかに低い再現度であると認識された。

4.2.5 散布方法

パターンに沿ってシリアルを散布した場合、ベースラインと比較して類似度および標準偏差に顕著な変化は見られなかった。均一散布の場合には、磁石が配置されているセルとシリアルとの距離が遠く磁力が小さいことに加え、シリアルが特定のセルに集中あるいは不足することで、類似度が安定しない傾向が生じる。そのため、イラストに沿って散布することでこれらの問題が軽減され、安定して高い類似度が得られると当初は予想していたが、結果は異なった。これについては、待機時間に関する検討で考察した内容と同様の可能性が示唆される。なお、磁石を配置せずに散布のみでパターンを再現した場合、シリアルは定着せず、意図した形状は形成されなかったことが確認された。

4.2.6 外部振動

実験者が指によって机を叩き外部振動を与えた場合、ベースラインと比較して類似度に顕著な変化は見られなかったものの、標準偏差は増加した。当初は、振動によって動いているシリアルは静止しているシリアルに比べて磁力の影響を受けやすく、精度の向上が期待された。しかし、結果は異なった。この要因としては、待機時間に関する検討で考察した内

容と同様の可能性に加え、一部のサンプルにおいて類似度が低下し、その影響で平均類似度が低下したことが、標準偏差の増加として現れたと考えられる。

4.2.7 浸透時間

ディスプレイ化する前にシリアルを 15 分間水に浸透させた場合、ベースラインと比較して類似度は低下したが、標準偏差に顕著な変化は見られなかった。この要因として、長時間の浸透により液架橋力が作用し、シリアル片同士が凝集した可能性が考えられる。当初は、浸透によってシリアルから鉄成分が分離し、磁力の影響が弱まると予想していた。しかし、実験者による目視観察では、ベースラインと同様にシリアルが磁石に引き付けられる様子が確認された。このことから、外部振動などを加えて凝集したシリアルを分離することで、ベースラインと同等の類似度が得られる可能性が示唆される。

4.2.8 液体種類

液体を牛乳に変更した場合、ベースラインと比較して類似度は低下し、標準偏差は増加した。この要因の一つとして、液体の粘度が影響した可能性が考えられる。一般に水の粘度は約 1 mPas であるのに対し、牛乳は約 2 mPas と水より高く、シリアルの移動が制限されるため、類似度が低下し標準偏差が増加したと推察される。また、牛乳が白濁していることにより、ある程度沈下したシリアルが見えなくなることも、標準偏差増加の一因と考えられる。

5 パターン遷移性能の調査

動的システムのパターン遷移性能を評価するため、開始パターン (図 6 (A)) からの目標パターン (図 6 (B)) への遷移を対象とし、外部振動の有無が与える影響を検証した。各条件下で 5 回の試行を行

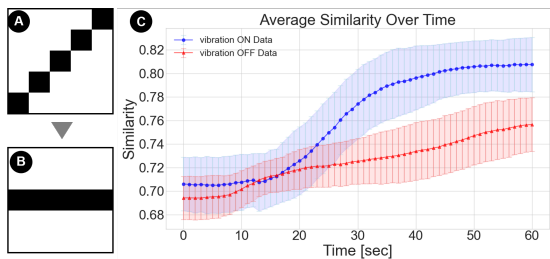


図 6. 遷移実験で用いた (A) 開始パターンと (B) 目標パターン, (C) パターン遷移における類似度の時間変化

い、遷移開始から 60 秒間の類似度の時間変化を 1 秒ごとに記録した。なお、本実験では液体量を 80 ml, シリアル量を 0.5 g に設定した。

各条件における類似度の時間変化の平均と標準偏差を (図 6 (C)) に示す。類似度は緩やかに上昇し、60 秒時点での平均値は 0.757 であった。同時点でも上昇傾向は継続しており、遷移が完了するにはさらに時間を要することが示唆される。対照的に、振動ありの条件では、遷移開始後 20 秒から 30 秒の区間で類似度が急峻に上昇し、50 秒付近で定常状態となった。60 秒時点での平均値は 0.808 であった。これにより、高速なパターン遷移が確認された。また、両条件下で遷移開始直後の数秒間は類似度に大きな変化が見られないが、これはアクチュエータの物理的な移動時間に起因すると考えられる。

これらの結果は、静的表示の場合とは異なり、動的なパターン遷移において外部振動が遷移の速度と精度の両方を向上させる上で極めて有効であることを示している。

6 応用例

本システムの具体的な活用法として、静的・動的それぞれの特性を活かした応用例と、より実践的な食事シーンへの展開を示す。

静的システムは、その準備の容易さを活かして複数連携させることが可能である。本研究では学会名「WISS」を 4 つの皿で表示したように (図 2(C)), 誕生日や記念日といったイベントにおいて、テーブル全体で祝祭的なメッセージを提示できる。

動的システムのリアルタイム制御という特徴は、食事に時間的な演出を加えることを可能にする。例えば、GUI からの指示で液面上に「3, 2, 1」といったカウントダウンをシリアルで表示させれば、参加者の期待感を高めるインタラクティブな体験を創出できる (図 7 (A))。

より実践的な応用を目指し、実際の料理であるスープを液体として用いる検証も行った (図 7 (B))。粘度が高いという課題に対し、精度実験で得られた

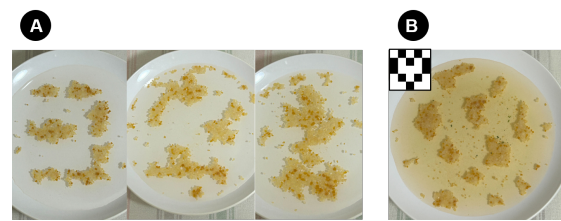


図 7. (A) 動的システムのカウントダウン (B) たまねぎスープを用いた実践的シナリオ

知見を応用することで表示精度の向上を試みた。この試みは、例えば結婚式で提供されるスープに新郎新婦のイニシャルを描き出すといった、実用シナリオに向けた初期検討である。

7 議論・今後の展望

今後の展望として、第一にシステムのハードウェア改良を通じ、表現力と再現性の両面から性能向上を目指す。より小型なアクチュエータで装置の薄型化を図り、ピクセルを高密度化することで複雑なパターン表現を可能にする。

第二に、類似度評価に加え、体験価値を評価するためのユーザスタディを実施する。アンケートやインタビューを通じて主観的な評価を取り入れ、人間の知覚に即した評価指標の確立を目指す。最終的には、レストランやイベントといった実環境での実証実験を通じて、多様な条件下での実用性を検証し、本システムのさらなる改良と応用可能性を拡大する。

本研究では、情報表示のためにシリアルを砕いた状態で利用している。このプロセスによってシリアル本来の食感が変化してしまう可能性がある。実験者による定性的な評価では、粉碎後のシリアルは、元のシリアルを液体に 5 分程度浸して柔らかくした状態の食感に近いことが確認された。一方で、味覚的な変化は特に感じられず、喫食する上での不快感もなかった。

8 結論

本稿では、鉄分強化シリアルが持つ磁気応答性を活用した可食ディスプレイシステム「Dishplay」を提案した。静的パターンを表示する小型システムと、リニアアクチュエータアレイにより動的表示を可能にするシステムの 2 種類を設計・実装した。さらに、表示品質を定量的に評価するための精度実験を行い、液体量、シリアル量、散布方法といった複数のパラメータが表示精度に与える影響を明らかにした。今後はより小型なアクチュエータを用いて解像度の向上や、主観的な評価を行うことを目指す。

謝辞

本研究の一部は JSPS 科研費 JP24KJ1957 の助成を受けたものです。

参考文献

- [1] W. Chen, Q. Gao, Z. Hu, K. Minamizawa, and Y. S. Pai. Living Bento: Heartbeat-Driven Noodles for Enriched Dining Dynamics. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, CHI '25, New York, NY, USA, 2025. Association for Computing Machinery.
- [2] W. Chen, Y. Yang, K.-H. Liu, Y. S. Pai, J. Yamaoka, and K. Minamizawa. Cymatics Cup: Shape-Changing Drinks by Leveraging Cymatics. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI '24, New York, NY, USA, 2024. Association for Computing Machinery.
- [3] J. Deng, Y. Li, H. Wang, Z. Fang, Florian 'Floyd' Mueller. Sonic Delights: Exploring the Design of Food as An Auditory-Gustatory Interface. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, CHI '25, New York, NY, USA, 2025. Association for Computing Machinery.
- [4] J. Deng, H. Yang, A. Saini, U. D. Gaudenz, L. Yao, P. Olivier, Florian 'Floyd' Mueller. Dancing Delicacies: Designing Computational Food for Dynamic Dining Trajectories. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*, DIS '23, p. 244–262, New York, NY, USA, 2023. Association for Computing Machinery.
- [5] S. Ganesh, P. Marshall, Y. Rogers, and K. O'Hara. FoodWorks: tackling fussy eating by digitally augmenting children's meals. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational*, NordiCHI '14, p. 147–156, New York, NY, USA, 2014. Association for Computing Machinery.
- [6] T. Gayler, C. Sas, and V. Kalnikaitis. Exploring the Design Space for Human-Food-Technology Interaction: An Approach from the Lens of Eating Experiences. *ACM Trans. Comput.-Hum. Interact.*, 29(2), Jan. 2022.
- [7] A. Grimes and R. Harper. Celebratory technology: new directions for food research in HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, p. 467–476, New York, NY, USA, 2008. Association for Computing Machinery.
- [8] Harvard T.H. Chan School of Public Health. Iron—The Nutrition Source. The Nutrition Source, Harvard T.H. Chan School of Public Health. Accessed: 2025-08-12.
- [9] G. Huisman, M. Bruijnes, and D. K. J. Heylen. A Moving Feast: Effects of Color, Shape and Animation on Taste Associations and Taste Perceptions. In *Proceedings of the 13th International Conference on Advances in Computer Entertainment Technology*, ACE '16, New York, NY, USA, 2016. Association for Computing Machinery.
- [10] R. Hurrell. Iron Fortification Practices and Implications for Iron Addition to Salt. *The Journal of Nutrition*, 151:3S–14S, 02 2021.
- [11] R. Hurrell and I. Egli. Iron bioavailability and dietary reference values. *The American Journal of Clinical Nutrition*, 91(5):1461S–1467S, 2010.
- [12] A. Ishii. Food Printing with Electrolysis Bubbles for Texture Control. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, CHI EA '24, New York, NY, USA, 2024. Association for Computing Machinery.
- [13] A. Ishii, M. Fukushima, N. Tanaka, Y. Matoba, K. Ikematsu, and I. Siio. Electrolysis Bubble Display based Art Installations. In *Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction*, TEI '21, New York, NY, USA, 2021. Association for Computing Machinery.
- [14] A. Ishii, K. Ikematsu, and I. Siio. Designing Electrolysis Ion Display on Everyday Open Wet Surfaces. In *Proceedings of the 2022 ACM Designing Interactive Systems Conference*, DIS '22, p. 1020–1034, New York, NY, USA, 2022. Association for Computing Machinery.
- [15] A. Ishii and I. Siio. BubBowl: Display Vessel Using Electrolysis Bubbles in Drinkable Beverages. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, UIST '19, p. 619–623, New York, NY, USA, 2019. Association for Computing Machinery.
- [16] K. Kato, K. Ikematsu, H. Nakamura, H. Suzaki, and Y. Igarashi. FoodSkin: Fabricating Edible Gold Leaf Circuits on Food Surfaces. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI '24, New York, NY, USA, 2024. Association for Computing Machinery.
- [17] R. A. Khot, D. Aggarwal, R. Pennings, L. Hjorth, and F. F. Mueller. EdiPulse: Investigating a Playful Approach to Self-monitoring through 3D Printed Chocolate Treats. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, p. 6593–6607, New York, NY, USA, 2017. Association for Computing Machinery.
- [18] K&J Magnetics, Inc. Magnetic Cereal: Magnets Reveal Hidden Cereal Ingredient! K&J Magnetics Blog, Jan. 2025. Accessed: 2025-08-12.
- [19] Y.-J. Lin, P. Punpongsanon, X. Wen, D. Iwai, K. Sato, M. Obrist, and S. Mueller. FoodFab: Creating Food Perception Illusions using Food 3D Printing. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing*

- Systems*, CHI '20, p. 1–13, New York, NY, USA, 2020. Association for Computing Machinery.
- [20] M. Miyatake, K. Narumi, Y. Sekiya, and Y. Kawahara. Flower Jelly Printer: Slit Injection Printing for Parametrically Designed Flower Jelly. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA, 2021. Association for Computing Machinery.
 - [21] T. Narumi, T. Kajinami, T. Tanikawa, and M. Hirose. Meta cookie. In *ACM SIGGRAPH 2010 Emerging Technologies*, SIGGRAPH '10, New York, NY, USA, 2010. Association for Computing Machinery.
 - [22] M. Nishizawa, W. Jiang, and K. Okajima. Projective-AR system for customizing the appearance and taste of food. In *Proceedings of the 2016 Workshop on Multimodal Virtual and Augmented Reality*, MVAR '16, New York, NY, USA, 2016. Association for Computing Machinery.
 - [23] J. Sun, Z. Peng, W. Zhou, J. Y. H. Fuh, G. S. Hong, and A. Chiu. A Review on 3D Printing for Customized Food Fabrication. *Procedia Manufacturing*, undefined.
 - [24] Y. Suzuki, T. Narumi, T. Tanikawa, and M. Hirose. Taste in Motion: The Effect of Projection Mapping of a Boiling Effect on Food Expectation, Food Perception, and Purchasing Behavior. *Frontiers in Computer Science*, Volume 3 - 2021, 2021.
 - [25] W. Wang, L. Yao, T. Zhang, C.-Y. Cheng, D. Levine, and H. Ishii. Transformative Appetite: Shape-Changing Food Transforms from 2D to 3D by Water Interaction through Cooking. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, p. 6123–6132, New York, NY, USA, 2017. Association for Computing Machinery.
 - [26] Y. Wang, H. O. Obie, Z. Li, F. D. Salim, J. Grundy, Florian 'Floyd' Mueller. GustosonicSense: Towards understanding the design of playful gustosonic eating experiences. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI '24, New York, NY, USA, 2024. Association for Computing Machinery.
 - [27] P. Weber, K. Krings, J. Nießner, S. Brodesser, and T. Ludwig. FoodChattAR: Exploring the Design Space of Edible Virtual Agents for Human-Food Interaction. In *Proceedings of the 2021 ACM Designing Interactive Systems Conference*, DIS '21, p. 638–650, New York, NY, USA, 2021. Association for Computing Machinery.
 - [28] T. Yamamoto, T. Amesaka, A. Withana, and Y. Sugiura. ShadoCookies: Creating user viewpoint-dependent information displays on edible cookies. *Computers & Graphics*, 127:104158, 2025.
 - [29] T. Yoshimoto, N. Kasahara, and H. Miyashita. Fabrication of Edible lenticular lens. In *ACM SIGGRAPH 2023 Posters*, SIGGRAPH '23, New York, NY, USA, 2023. Association for Computing Machinery.
 - [30] T. Yoshimoto, Y. Minato, and H. Miyashita. Edible Lens Array: Dishes with lens-shaped jellies that change their appearance depending on the viewpoint. In *Adjunct Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*, UIST Adjunct '24, New York, NY, USA, 2024. Association for Computing Machinery.
 - [31] Y. Zhao, C. Yu, J. Nie, M. Dong, Y. Sang, F. Ying, and G. Wang. FunEat: An Interactive Tableware for Improving Eating Habits in Children. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI EA '21, New York, NY, USA, 2021. Association for Computing Machinery.
 - [32] 鳴海 拓志. プロジェクションマッピングによる食体験の拡張. *日本画像学会誌*, 61(1):22–32, 2022.

ヒトに偏光知覚能力を付与するウェアラブルシステム

大城 亨太* 大西 鮎美* 寺田 努* 塚本 昌彦*

概要. 規則的に振動する光を偏光といい、これに対してあらゆる方向に振動する光を無偏光という。人間は偏光と無偏光が同じように見えるため偏光を知覚できない一方で、偏光を知覚し生活に活用する生物が存在する。このことから、人間も偏光を知覚することで得た情報を活用できる可能性がある。そこで本研究では、人間が偏光をどのように活用できるかを明らかにするため、人間に偏光を知覚させるウェアラブルデバイスを提案する。提案デバイスでは、偏光子を眼前で回転させ、偏光のみが明滅して見えるようにすることで偏光知覚を実現した。また、日常生活において偏光がどのように利用できるかを探索的なアプローチにより調査した結果、偏光は太陽の位置の推定や水たまりの検出等に役立てられることがわかった。

1 はじめに

人間は目や耳などの感覚器官を用いてさまざまな情報を知覚しており、情報技術による知覚拡張は日常生活の様々な面で有用である。例として、音の周波数を操作するウェアラブルデバイスにより、人間の聴力では従来聴こえなかった音を可聴化する研究がある [1]。このデバイスは、ガスの漏洩時に発生する高周波音を認識して異常機器を検出したり、電気自動車の走行時に発生するスイッチングノイズを認識して衝突事故を防いだりできる。また、ゴーグルに取り付けたシャッターを一定時間間隔で開くことで、ゴーグル着用者の視覚のフレームレートを下げる研究がある [2]。このゴーグルでは、高速に移動する物体の姿が鮮明に捉えられたり、回転するコインの表裏が見えたりと、通常では認識できない運動体の情報を認識できるようになる。これらのように、人間が知覚拡張によって通常では得られない情報を得られるようになると、それらの情報を生活に役立てられる可能性がある。

人間が周囲の環境などを認知する際に視覚は重要な役割を果たすが、人間が感知できる光の情報は一部である。本研究では、人間が通常では得られない情報として偏光を採り上げる。光は電磁波であり、光の進行方向に対して電場と磁場が互いに直交するように振動している。太陽光や蛍光灯の光などの一般的な光はあらゆる方向に振動しており、これを無偏光という。これに対して反射などの特定場面では、振動する方向が規則的な光が生じることがあり、これを偏光という [3]。本論文では、自身が見ている光が偏光か無偏光か判別できる状態を、偏光を知覚できる状態とする。人間は偏光に対する感覚器官をもっておらず、偏光と無偏光が同じ見え方をするため、一般には偏光を知覚できない [4]。一方で、いく

つかの生物は偏光を生活に活用している。ミツバチは、偏光の情報を利用して太陽の位置を検知することで方角を推定しており、偏光を巣に戻るためのナビゲーションとして活用している [5]。またハリガネムシは、宿主であるカマキリが偏光を知覚していることを利用し、カマキリが水面での反射の際に生じる偏光を目印にして水に飛び込むように操る [6]。これらの偏光を利用する生物のように、人間も偏光を知覚して生活に役立てられる可能性がある。

本研究の最終的な目的は、人間が適切な方法で偏光を知覚し、得た情報を自らの判断や行動に活かせるようにするデバイスの実装である。人間が偏光をどう活用できるのかがわかれば、偏光を知覚させる適切な方法が検討できる。しかし人間は自身の目では目に入った光が偏光であるかわからず、人間が偏光をどう活用できるのかは調査されていない。そこで本研究では、最終目的に向けた第一段階として、偏光を知覚できる状態を人間が体験できるようにするウェアラブルデバイスを提案する。また、デバイスを着用した状態で生活することで、生活の中で偏光が発生している場面を探索し、偏光を知覚して何に活用できるかを調査する。偏光知覚は、偏光と無偏光の見え方が異なるようにしたり、偏光であることを音で知らせたりすることで実現できる。今回は偏光子を眼前で回転させることで、偏光と無偏光の見え方を変えて偏光知覚を実現する。回転する偏光子を通して偏光を見たとき、偏光子が透過する光の振動方向は時間的に変化し、目に入る偏光の光量は周期的に変化する。そのため明滅して見えるものは偏光、見え方が変わらないものは無偏光と判別できる。

本研究では探索的なアプローチをとり、日常の中で偏光がどのように利用できるかを調査した。参加者は日常の中で偏光知覚デバイスを装着し、偏光が発生する場面を観察した。この実験により、提案デバイスが日常生活における偏光知覚を可能にすることが示された。さらに、観察された偏光に基づき、

Copyright is held by the author(s).

* 神戸大学大学院工学研究科 Graduate School of Engineering, Kobe University

それが人間にとってどのように役立つかを議論し、偏光を生活に活用できるようにするうえで適切な偏光知覚の方法を検討した。

2 関連研究

2.1 人間拡張に関する研究

情報技術を用いて人間の身体能力や知覚などの増強や拡張を行う技術を人間拡張という。近年、人間拡張についての研究は数多く行われている。人間拡張は大きく分けて身体拡張、認知拡張、知覚拡張の3つに分類できる [7]。本研究は知覚拡張に着目する。

Jain らは、ヘッドマウントディスプレイを用いて音を可視化する聴覚障害者向けのシステムを提案した [8]。このシステムでは、矢印などのアイコンによって音源の方向が示され、アイコンの大きさで音の大きさが示される。これにより聴覚障害者は、音を空間的に捉えることで視覚から音の情報を得られる。Watanabe らは、音の周波数を操作するウェアラブルデバイスにより聴力を自在化する手法を提案した [1]。この手法では、非可聴域の周波数を可聴域の周波数に変換することで、人間の聴力では聞こえない音を可聴化できる。このデバイスは、ガスの漏洩時に発生する高周波音を認識して異常機器を検出したり、電気自動車の走行時に発生するスイッチングノイズを認識して衝突事故を防いだりと、日常生活のさまざまな場面で汎用的に活用できる。Koizumi らは、レンズに取り付けたシャッターを一定時間間隔で開く Stop Motion Goggle を提案した [2]。このシステムは、ゴーグル着用者の視覚のフレームレートを下げることで、高速に移動する物体の姿が鮮明に捉えられたり、回転するコインの表裏が見えたりと、通常では認識できない運動体の情報を認識できるようになる。このゴーグルはスポーツ観戦に応用でき、ボールの動きが速すぎて目で追えない場合などに有用である。

これらの例から、人間が知覚拡張によって通常では得られない情報を得られるようになると、それらの情報を生活に役立てられる可能性がある。本研究では、偏光知覚を実現することで、人間が偏光を生活に役立てられることを目指す。

2.2 偏光を活用する生物に関する研究

偏光を生活に活用する生物が存在する。これらの生物による偏光の活用例を紹介する。ミツバチは、巣や餌場の位置を記憶するために方角を把握しながら飛行する。この習性について Rossel らは、青空の偏光パターンを提示した状態で、餌が置かれた場所に対してミツバチがどのように動くかを分析した [5]。その結果、ミツバチは青空の偏光を光の明るさや色の変化により知覚することで方角を認識することが解明された。カマキリを宿主とする寄生虫のハリガ

ネムシは、繁殖のためにカマキリが水に飛び込むように操る。この習性について Obayashi らは、光の条件が異なる2つのプールを用意し、ハリガネムシに寄生されたカマキリがどちらに飛び込むかを調査した [6]。その結果、ハリガネムシに寄生されたカマキリは、光強度に関わらず水面での反射により生じた水平方向の偏光を目印にして水に飛び込むことが明らかになった。また海底に生息するシャコは、円偏光を識別して偏光を獲物や捕食者の発見に活用している [9]。これらの活用例から、人間も偏光を知覚できれば生活に役立てられる可能性がある。

次に、生物が偏光を知覚する方法を紹介する。コオロギは青空の偏光を認識することで太陽の位置を把握する。Sakura らは、コオロギがどのように偏光の情報を得ているかを解明するため、偏光の方向を変化させたときにコオロギがもつニューロンの応答の変化を調査した [10]。その結果、コオロギは特定の向きの偏光に対して強い興奮性の応答を示す POL1 ニューロンを3種類もち、各ニューロンでの応答するニューロン数の比率により偏光の情報を得ている可能性が示唆された。しかし、偏光を感知する感覚器官をもつ生物とは異なり、人間にはそのような生物学的機構が存在しない。よって人間が偏光を知覚するには人工的な手段に頼らざるを得ない。

2.3 特定の目的のために偏光を用いた研究

偏光を活用した研究は多く存在し、偏光を用いた様々な技術が提案されている。Umeyama らは、物体の表面での反射光である鏡面反射成分と、物体内部の色素粒子での散乱光である散乱反射成分とを分離する手法を示した [11]。この手法では、異なる方向の偏光子を通して撮影した画像を分析することで、偏光を含む鏡面反射光と無偏光である拡散反射光に分離できる。Wolff らは、偏光の情報を解析することで、物体表面の材質を推定する手法を提案した [12]。この手法では反射光の偏光特性の違いが用いられており、特に非導電体と高導電性の金属が判別できることが示された。さらに Cula らは、複数光源のもとで撮影された画像を、それぞれの光源1つで照らされた画像に分解する手法を提案した [13]。この手法では、偏光の情報をを用いて物体の表面での反射光がどの光源によるものかを分離できる。

このように、偏光を特定の目的に利用する研究は数多く存在する。しかし人間が自ら偏光を知覚することで得られる利点を明らかにした研究は存在せず、これらの他にも偏光の活用法が新たに見つかる可能性がある。そこで本研究では、人間が偏光を知覚できるデバイスを実装することで、偏光を知覚して生活に活用できる可能性を探索する。

3 提案デバイス

3.1 偏光知覚の定義

本論文では、視界中の各部分の光が偏光か無偏光か判別できる状態を、偏光を知覚できる状態とする。人間は偏光に対する感覚器官をもっておらず、偏光と無偏光が同じ見え方をするため偏光を知覚できない。よって人間が偏光を知覚するには人工的な手段を採る必要がある。具体的には、偏光と無偏光の見え方が異なるようにする方法や、偏光であることを音で知らせる手法が考えられる。偏光カメラを用いれば、偏光を検出してディスプレイ上で偏光を可視化することは可能であるが、ディスプレイ映像では光に対する知覚印象が実際の空間で感じるものと異なる[14]ことが懸念される。よって本研究では、人間が光そのものの変化を視覚的に感じられることを重視し、デジタル変換を介さずに偏光を直接知覚させる光学的手法を採用する。

3.2 偏光知覚を実現する手法

本論文では視覚から偏光を知覚させる。その方法はいくつか考えられるが、偏光を知覚する生物のように[10]、人間も異なる振動方向の偏光を比較して見ると偏光を知覚できる可能性がある。異なる振動方向の偏光を比較して見る方法としては、昆虫の複眼のように、視覚内で映像を複数並べるような比較方法も考えられるが、この方法では人間の普段の世界の見え方と大きく異なる。そこで本研究では、偏光子を眼前で回転させて時間的な変化により偏光を比較する手法を提案する。偏光子とは、入射光の振動方向によって透過光の強度を変化させる光学素子である。透過光の強度 I [W/m^2] はマリュスの法則により、入射光の強度 I_0 [W/m^2]、入射光の振動方向と偏光子の透過軸がなす角 θ を用いて、式(1)のように表される[15]。

$$I = I_0 \cos^2 \theta \quad (1)$$

よって偏光子を眼前で回転させると、特定の振動方向をもつ偏光は目に届く明るさが周期的に変化し、あらゆる振動方向をもつ無偏光は目に届く明るさが一定となる。本論文では、偏光子の回転により偏光が明滅するように見せることで偏光知覚を実現する。本手法は、回転する偏光子を通した光学的变化をそのまま提示することで、カメラや映像処理を介さずに光の偏光特性を直接知覚できることに特徴がある。すなわち、本手法は映像として再構成された偏光情報を見るのではなく、観察者が自らの視覚で偏光現象を体験的に捉えることを可能にする。

3.3 提案デバイスの構成

提案デバイスを装着したときの様子を図1に示す。提案デバイスは眼鏡フレーム、マイコン、サーボモータ、偏光子、プーリーとベルトからなる回転機構で

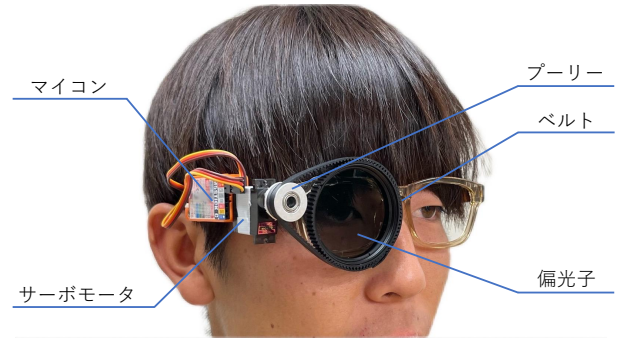


図 1. 提案デバイスを装着したときの様子

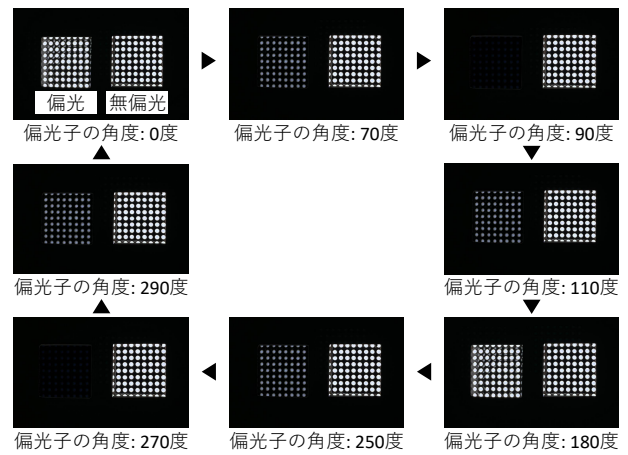


図 2. 提案デバイスを通した偏光と無偏光の見え方

構成される。サーボモータに取り付けたプーリーからベルトを通して偏光子に力を伝え、偏光子を回転させる。回転機構の動力源はサーボモータで、マイコンがモータの回転速度を制御する。また、マイコンの電源としてモバイルバッテリーを使用している。マイコンは M5Stack Technology 社製の M5StickC である。サーボモータは梅本社製の MG90D で、同一方向に連続回転できる。偏光子はマルミ工機社製で、直径は 49 [mm] である。

偏光子の回転が遅いと明滅がわかりにくくなり、偏光が知覚しにくくなる可能性がある。逆に、偏光子の回転が速いと偏光が激しく明滅し、デバイス使用者にストレスを与える可能性がある。第一著者がいくつかの回転速度で試した結果、1.4 秒に 1 回程度の明滅が妥当であったため、提案デバイスでは偏光子が約 2.8 秒で一回転するように設計した。また装着時の負荷を考慮して軽量に設計し、デバイスの総重量は 75.0 [g] に抑えた。このように、日常的な使用を想定し、偏光子の回転速度や装置重量を調整して知覚の明瞭さと装着快適性の両立を図った。

3.4 提案デバイスを通したときの見え方

提案デバイスを通した偏光と無偏光の見え方を図2に示す。各画像の左側の光は LED に偏光子を固

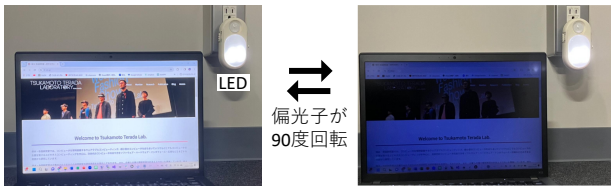


図 3. 液晶ディスプレイでの偏光



図 4. 青空での偏光



図 5. ガラス面での偏光

定して被せているため偏光で、右側の光は何も被せていない LED であるため無偏光である。偏光の振動方向を 0 度とすると、これらの画像は左上の画像から時計回りに、提案デバイスの偏光子の角度がそれぞれ 0 度、70 度、90 度、110 度、180 度、250 度、270 度、290 度の画像である。提案デバイスを通すと無偏光は常に明るく見え、見え方に変化はない。一方で偏光は明滅して見えている。よって、提案デバイスによって偏光と無偏光が判別できる。

4 実験

本章では、提案デバイスを用いて偏光が日常のどのような場面で発生しているかを探索する実験について述べる。その後、実験により見つかった偏光とその活用例について述べる。なお本実験は、神戸大学大学院工学研究科人を直接の対象とする研究倫理審査委員会の承認(承認番号 05-55)を得て行ったものである。

4.1 実験の概要

本実験の目的は、偏光が日常のどのような場面で発生しているかを探索することである。被験者は、提案デバイスの装着により偏光が知覚できる状態で同伴者と自由に歩行し、ものが明滅して見えたときに同伴者に逐次報告した。同伴者は被験者からの報告を受け、偏光が発生している場面や、どのような見え方であったかを被験者から聞き取り記録した。被験者は 20 代から 50 代の男女 7 名で、実験時間は 1 人あたり約 1 時間である。実験場所は神戸大学キャンパス周辺や住宅地であり、屋内外両方で探索を行った。また実験は晴天時と雨天時に行った。

4.2 見つかった偏光とその活用例

実験で見つかった偏光の例を図 3 から図 12 に示す。これらはすべて提案デバイスを通して撮影した。左右の画像では偏光子の角度が 90 度異なり、各場面における明滅の最も明るい状態と暗い状態の画像である。

液晶ディスプレイの見え方

図 3 は、液晶ディスプレイで発生していた偏光の見え方である。提案デバイスを通すと液晶ディスプレイが明滅して見え、液晶ディスプレイから出ている光が偏光であるとわかる。対して右上の LED ライ

トの明るさは変化せず無偏光だとわかる。液晶ディスプレイは 2 枚の偏光子を利用しており、電圧により偏光の振動方向を制御することで液晶ディスプレイの明るさが調整できる [16]。よって液晶ディスプレイの光は偏光であり、提案デバイスを通すと偏光子の角度によって液晶ディスプレイの明るさが違って見え、特定の角度では消灯しているように見えた。よって偏光が知覚できると液晶ディスプレイの映像と他の映像を区別できる。

青空の見え方

図 4 は、青空で発生していた偏光の見え方である。この画像は太陽と反対方向の青空を撮影しており、強く明滅して見えた。一方で、太陽付近の青空はほとんど見え方が変わらなかった。光の強度に占める偏光の強度の割合を偏光度といい、青空では太陽からの距離が離れるにしたがって偏光度が大きくなり、一定の距離を超えるとまた小さくなる [17]。また青空において偏光の振動方向は太陽を中心とした同心円を形成している [18]。図 4 は太陽と反対方向の青空であり偏光度が大きかったため、青空が強く明滅していたと考えられる。

よって偏光度や偏光の振動方向がわかると、太陽の位置が推定できるようになる可能性がある。人間は太陽の位置で方角を推定できるが、街中などで視界が遮られ、太陽の位置がわからない場合でも太陽の位置を偏光から推定できる可能性がある。太陽の位置が常に知覚できると、時間がわかりやすくなることや、方向感覚が付くことで地図が読みやすくなることなどが期待できる。また地図アプリを使用中に GPS の不調などにより方向がわからなくなることがあるが、青空での偏光から方角を知ることによってナビゲーションに偏光を活用できる可能性がある。

ガラスへの映り込みの見え方

図 5 は、ガラス面で発生していた偏光の見え方である。ガラスへの映り込みが明滅して見えたが、ガ



図 6. 石垣と路面での偏光

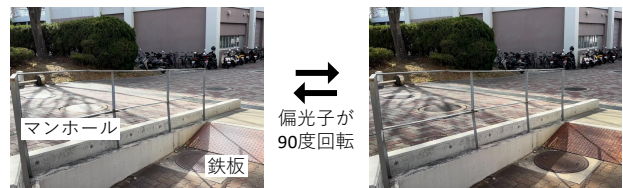


図 8. 晴れた日の路面での偏光



図 7. 自動車の車体での偏光

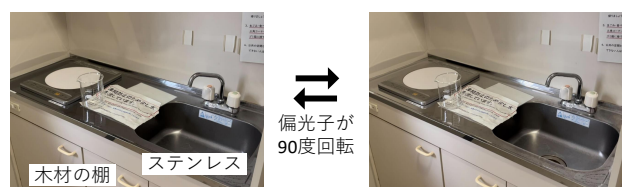


図 9. シンクでの偏光

ラス背後の実物体は明滅していなかった。つまり、ガラスでの反射光は偏光，ガラス背後からの透過光は無偏光だとわかった。

よって偏光はガラスへの映り込みとガラス背後の実物体を分離して知覚させられる可能性がある。通常はガラスへの映り込みとガラス背後の実物体が重なって見えるためこれらを判別できず，ガラスの扉での出会い頭の衝突などが発生することがある。しかし，偏光知覚によりこれらを分離して認識できると，ガラスの手前側と奥側の双方の状況を同時に把握できるようになる。一方で，ガラスを正面から見た場合には偏光が発生しておらず，ガラスを見る角度によって明滅の強さが異なっていた。これは反射角によってガラスでの反射光の偏光度が異なるためである。よってガラスへの映り込みとガラス背後の実物体を分離して認識するには，ガラスを斜めから見た視点が必要であることを考慮しなければならない。

面方向が異なる物体の見え方

図 6 は，石垣と路面で発生していた偏光の見え方である。石垣と路面での反射光はともに明滅して見え，偏光が発生しているとわかった。石垣と路面では明滅するタイミングが異なっており，石垣が明るいときに路面が暗く，石垣が暗いときに路面が明るく見えた。図 7 は，自動車の車体で発生していた偏光の見え方である。自動車のボディ面での反射光が明滅して見えたが，面によって明滅するタイミングが異なっていた。これらのことから，提案デバイスでは，物体の面方向の違いによって生じる偏光の振動方向の差は，明滅するタイミングの違いによって知覚できることがわかった。

よって明滅のタイミングの違いによって偏光の振動方向がわかると，物体の面方向の推定に活用できる可能性がある。物体の面方向が推定できると，物体の形状を把握し面の歪みなどを認識することに役立てられる。特に自動車においては，ボディ面の面方向を推定して傷やへこみを認識できる可能性がある。

材質が異なる物体の見え方

図 8 は，晴れた日の路面で発生していた偏光の見え方である。日向の路面が全体的に明滅して見えたが，特にマンホールやスロープの鉄板で強く明滅していた。図 9 は，シンクで発生していた偏光の見え方である。ステンレスの部分が明滅して見えたが，木材で作られた棚の部分は見え方がほとんど変わらなかった。図 10 は，植木で発生していた偏光の見え方である。やわらかい葉をもつ右手前の植木 A は弱く明滅して見え，硬くて艶のある葉をもつ左奥の植木 B は強く明滅して見えた。これらにおいて，強く明滅していたのはマンホールやスロープの鉄板，ステンレス，および硬くて艶のある葉をもつ植木 B であった。これは，光沢のある物体は偏光である反射光を生じやすいためだと考えられる。

よって偏光度がわかると物体の材質の推定に役立てられる可能性がある。人間の肌は皮脂によるてかりが生じたり，乾燥して肌の光沢が失われたりすることがあり，偏光を知覚できると肌の状態を推定できるようになる可能性がある。

濡れた路面や水たまりの見え方

図 11 は，路面が湿った箇所が発生していた偏光の見え方である。路面は全体的に明滅して見えたが，湿った箇所では周囲と比較して強く明滅して見えた。図 12 は，雨天時の路面で発生していた偏光の見え方である。雨により路面が湿った箇所が強く明滅して見えた。また，水たまりとなった箇所でも偏光が発生していたが，周囲の路面が濡れた箇所と比較すると明滅が弱く感じられた。よって，乾いた路面や水たまりと比較して，湿った路面では偏光が発生しやすいことが確認できた。これは，湿った路面では偏光である反射光が発生しやすいためだと考えられる。

このことから，偏光を知覚できると路面の濡れや水たまりを検出できるようになる可能性がある。雨天時には，どこに水たまりがあるか把握して避けな



図 10. 植木での偏光



図 12. 雨天時の路面での偏光

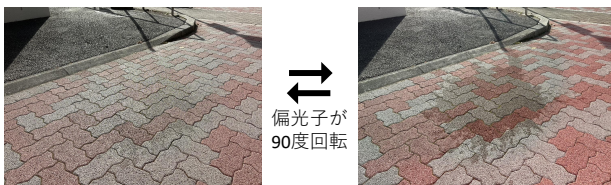


図 11. 湿った路面での偏光

がら歩行するが、誤って水たまりを踏んでしまうことがある。そこで、偏光知覚により水たまりを検出できると水たまりを避けやすくなる可能性がある。また濡れた路面は滑りやすく、特に自動車の運転時には注意する必要があるが、偏光知覚により路面状況を把握しやすくなり安全性が高まる可能性がある。

5 議論

5.1 明滅による偏光提示手法の妥当性と限界

提案デバイスを用いた実験では、偏光子の回転によって偏光が明滅して見えることで偏光を視覚的に知覚できた。明滅の強さや光の強度から偏光度をおおよそ推定できることが確認された一方で、偏光の振動方向については明滅のタイミングの違いから相対的な差異を把握できるのみであり、偏光がどの方向に振動しているかは直接的には知覚できなかった。被検者からは装着時に大きな違和感や不快感は報告されず、おおむね自然な装着感で偏光を観察できた。また、無偏光の部分は見え方が変わらないため、基本的には普段通りの生活ができたとの評価を得た。一方で、長時間装着した際の疲労感や集中力への影響は今後の評価が必要である。よって明滅による偏光の提示手法は、偏光を知覚できる状態を体験する初期段階としては妥当であると評価できる。一方で、偏光の特性をより深く理解して応用へとつなげるためには、偏光度や偏光の振動方向といった複数の要素を同時に提示することで、偏光をより総合的に知覚できるデザインへと発展させる必要がある。

5.2 汎用的活用に向けた偏光提示手法の検討

実験では、偏光の活用例ごとに必要な情報が異なり、提案デバイスでは偏光を活用するうえで十分な情報が得られない場面があった。そこで本節では、実験で得られた知見をもとに、偏光という新たな視覚情報を人間が自然に理解し、日常生活の中で偏光を汎用的に活用できるようにするための提示手法に

ついて検討する。

探索的実験の結果、提案デバイスには主に次の二つの課題があることが明らかになった。第一に、偏光の振動方向がわからないという課題である。提案デバイスでは明滅の強さから偏光度を推定できたが、偏光の振動方向については明滅のタイミングの違いから相対的な差異を把握できるのみであった。これに対しては、偏光子の透過軸を視野内に表示し、偏光の振動方向を直接的に把握できるようにすることが有効であると考えられる。このような機能の導入により、太陽の位置や物体の面方向の推定といった場面でも、より正確に偏光の情報を活用できるようになることが期待される。第二に、偏光が遮断される時間が短いという課題である。特に、ガラスへの映り込みとガラス背後の実物体を分離する場面においては、ガラス面での反射光が遮断される時間が短く、ガラス背後の実物体を十分に視認できないという問題があった。これに対しては、ユーザが状況に応じて偏光子の回転速度を調整できる制御機能を追加し、明滅の速さを自ら操作できるようにすることが有効であると考えられる。

今後はこれらの機能を導入することで、人間が偏光を自然に理解し、日常生活の中で活用できるようになる汎用的な偏光知覚デバイスの実現を目指す。

6 まとめ

本論文では、人間が偏光を知覚して生活に活用できるようにするための第一段階として、偏光を知覚できる状態を人間に体験させるウェアラブルデバイスを提案した。提案デバイスは偏光子の回転により偏光を明滅させることで偏光知覚を実現しており、明滅するものを偏光、明るさが一定であるものを無偏光と判別できる。実験では、提案デバイスを着用した状態で自由に生活させ、どのような場面で偏光が発生しており、偏光を知覚して何に活用できそうか探索した。実験の結果、偏光は太陽の位置の推定や水たまりの検出等に役立てられる可能性が示唆された。またこれらの活用例に基づき、偏光を活用するうえで必要な機能を整理し、適切な偏光提示手法を検討した。今後はこれらの機能を導入することで、人間が適切な方法で偏光を知覚し、得た情報を生活に活用できるようにするデバイスを実装する。

謝辞

本研究の一部は、科学研究費補助金基盤研究 (A) (25H01157) の支援によるものである。ここに記して謝意を表す。

参考文献

- [1] H. Watanabe and T. Terada: Manipulatable Auditory Perception in Wearable Computing, *Proc. of the Augmented Humans International Conference 2020 (AHS '20)*, pp. 9:1–9:12, (Mar. 2020).
- [2] N. Koizumi, M. Sugimoto, N. Nagaya, M. Inami, and M. Furukawa: Stop Motion Goggle: Augmented Visual Perception by Subtraction Method Using High Speed Liquid Crystal, *Proc. of the 3rd Augmented Human International Conference (AH '12)*, pp. 1–7 (Mar. 2012).
- [3] J. J. Foster, S. E. Temple, M. J. How, I. M. Daly, C. R. Sharkey, D. Wilby, and N. W. Roberts: Polarization Vision—Overcoming Challenges of Working with a Property of Light We Barely See, *The Science of Nature*, Vol. 105, pp. 1–26 (Mar. 2018).
- [4] S. E. Temple, J. E. McGregor, C. Miles, L. Graham, J. Miller, J. Buck, N. E. Scott-Samuel, and N. W. Roberts: Perceiving Polarization with the Naked Eye: Characterization of Human Polarization Sensitivity, *Proc. of the Royal Society B: Biological Sciences*, Vol. 282, No. 1811 (July 2015).
- [5] S. Rossel and R. Wehner: How Bees Analyse the Polarization Patterns in the Sky, *Journal of Comparative Physiology A*, Vol. 154, pp. 607–615 (Sep. 1984).
- [6] N. Obayashi, Y. Iwatani, M. Sakura, S. Tamotsu, M. C. Chiu, and T. Sato: Enhanced Polarotaxis Can Explain Water-Entry Behaviour of Mantids Infected with Nematomorph Parasites, *Current Biology Magazine*, Vol. 31, No. 12, pp. 777–778 (June 2021).
- [7] S. Villa, J. Niess, T. Nakao, J. Lazar, A. Schmidt, and T. Machulla: Understanding Perception of Human Augmentation: A Mixed-Method Study, *Proc. of the 2023 CHI Conference on Human Factors in Computing Systems*, No. 65, pp. 1–16 (Apr. 2023).
- [8] D. Jain, L. Findlater, J. Gilkeson, B. Holland, R. Duraiswami, D. Zotkin, C. Vogler, and J. E. Froehlich: Head-mounted Display Visualizations to Support Sound Awareness for the Deaf and Hard of Hearing, *Proc. of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*, pp. 241–250 (Apr. 2015).
- [9] N. W. Roberts, T.-H. Chiou, N. J. Marshall, and T. W. Cronin: A Biological Quarter-Wave Retarder with Excellent Achromaticity in the Visible Wavelength Region, *Nature Photonics*, Vol. 3, No. 11, pp. 641–644 (Nov. 2009).
- [10] M. Sakura, D. Lambrinos, and T. Labhart: Polarized Skylight Navigation in Insects: Model and Electrophysiology of e-Vector Coding by Neurons in the Central Complex, *Journal of Neurophysiology*, Vol. 99, No. 2, pp. 667–682 (Feb. 2008).
- [11] S. Umeyama and G. Godin: Separation of Diffuse and Specular Components of Surface Reflection by Use of Polarization and Statistical Analysis of Images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 26, No. 5, pp. 639–647 (May 2004).
- [12] L. B. Wolff and T. E. Boulton: Constraining Object Features Using a Polarization Reflectance Model, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 13, No. 7, pp. 635–657 (1991).
- [13] O. G. Cula, K. J. Dana, D. K. Pai, and D. Wang: Polarization Multiplexing for Bidirectional Imaging, *Proc. of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, Vol. 2, pp. 1116–1123 (June 2005).
- [14] A. Çevik, Z. T. Kazanasmaz, G. Tambellini, G. Salvadori, and F. Leccese: Perception of Light in Museum Environments: Comparison between Real-Life and Virtual Visual Experiences, *Sustainability 2022*, Vol. 14, No. 14288 (Nov. 2022).
- [15] M. Monteiro, C. Stari, C. Cabeza, and A. C. Martí: The Polarization of Light and Malus' Law Using Smartphones, *The Physics Teacher*, Vol. 55, No. 5, pp. 264–266 (May 2017).
- [16] J. Xiong and S. T. Wu: Planar Liquid Crystal Polarization Optics for Augmented Reality and Virtual Reality: from Fundamentals to Applications, *ELight*, Vol. 1, No. 3, pp. 1–20 (June 2021).
- [17] M. L. Brines and J. L. Gould: Skylight Polarization Patterns and Animal Orientation, *Journal of Experimental Biology*, Vol. 96, No. 1, pp. 69–91 (Feb. 1982).
- [18] T. Ma, X. Hu, L. Zhang, J. Lian, X. He, Y. Wang, and Z. Xian: An Evaluation of Skylight Polarization Patterns for Navigation, *Sensors*, Vol. 15, No. 3, pp. 5895–5913 (Mar. 2015).

ペンさく：描いて探すことで深まるペンギン観察手法と水族館来訪者を対象とした実証実験

中川 由貴* 中村 聡史* 芦刈 治将† 板東 恵理子†
 與倉 陵太† 渡邊 果南† 岩永 七海†

概要. 動物園や水族館では複数の個体が同時に展示されることが多く、来訪者の観察は全体を眺めるにとどまりやすい。特にペンギンは数十羽単位で飼育されることが一般的であり、一部の施設では個体名を付けて一覧掲示を行うなど、個体に着目した観察を促す工夫がなされている。個体識別することがその個体への共感につながることは示されているものの、実際に観察中の個体と掲示情報を結びつけるのは容易ではない。そこで我々はこれまで、腹部模様を描画して個体を検索できるシステムを提案し、その有用性を検証してきた。本研究では、実環境におけるシステムの効果を検証するため、都市型水族館で一般来訪者を対象とした実証実験を実施した。また実験にあたり、新たに個体情報を一覧できる図鑑機能と観察履歴を保存できるアルバム機能を追加した。実験の結果、個体名を用いた会話や模様の違いへの気づき、展示前での滞在時間の増加が確認され、主体的な観察や動物への関心を高める有効性が示唆された。

1 はじめに

水族館や動物園は、多くの人々にとってレジャーの場として楽しめる一方、教育・研究・保全といった社会的役割を担う重要な施設である [15]。来訪者はそこでの体験を通じて動物への理解や保全意識、自然とのつながりを深めることが期待される [11, 3]。

しかし、水族館や動物園では複数の個体を同時に展示することが一般的であるため、来訪者の観察は展示全体を俯瞰して眺めるだけにとどまりやすい。この課題に対応するため、多くの施設では個体に名前を付け、性格や特徴を紹介する掲示を設置するなど、来訪者が個体に着目できるような工夫を行っている。こうした個体に対する工夫は、ライオンやカバ、アシカやタコといったように少数で展示されている動物に対しては有効に働くが、数十頭、数十羽規模で飼育・展示するような動物の場合は効果が薄れてしまう。

本研究で扱うペンギンは、数十羽規模で飼育・展示されることが多く、来訪者の観察は俯瞰的なものになりやすい。こうした問題に対し、多くの施設ではフリッパーバンドや名前による個体識別をして、個体一覧や相関図を掲示するなどの展示の工夫を行っている。この個体に着目した情報の提示は、リピーターを増やすことにも繋がっており、推しペンギンなどの言葉もうまれている。また、イルカにおいては個体を識別することがその個体に対する共感や保全意識の高まりにつながるということが明らかになっている [30] ため、個体識別を容易化することは重要であ

ると考えられる。一方で、既存のフリッパーバンドによる識別には、視認性の低さや個体への負担の課題があることに加え、来訪者が観察中の目の前の個体とその情報を直接結びつけることは容易ではない。

我々はこれまで、ペンギンの腹部斑点模様に注目し、来訪者がその模様を描画することで個体を検索可能とする手法を提案・実装し、有用性を示してきた [26]。しかし、これまでの研究は大学生を対象とした小規模な実験であり、実際の水族館の一般来訪者における有用性は十分に検証されていなかった。また、検索対象となる個体数にも制限があった。

そこで本研究では、都市型水族館であるサンシャイン水族館とともに研究を進め、一般来訪者を対象に実証実験を行い、実環境におけるシステムの有用性および他の施設への導入可能性を検証する。実験にあたり、水族館で飼育されている全個体をシステム上で検索対象とした。また、子どもを含む幅広い年代の来訪者の利用を想定し、ユーザインタフェースをわかりやすいものに再設計した。さらに、システムの利用モチベーション向上のために、個体情報を一覧できる図鑑機能と、過去に描画・検索した個体を保存し閲覧できるアルバム機能を追加した。これらの改良を施したシステム「ペンさく」を用いて、サンシャイン水族館にて6日間の実証実験を実施し、来訪者の観察行動に与える影響を考察する。

2 関連研究

2.1 動物園・水族館の役割について

動物園や水族館は単なる娯楽施設ではなく、さまざまな役割を担う施設へと発展してきた [23, 31, 32]。現在では、娯楽・教育・保全・研究の4つの目的の

Copyright is held by the author(s).

* 明治大学

† サンシャイン水族館

もと運営されており [15], 5つ目の柱として福祉やウェルビーイングを位置づける必要性も議論されている [14]. また, 娯楽についても消費的な楽しみだけでなく, 動物への共感を促すエンゲージメントとして再定義すべきとの議論もある [29].

こうした施設への訪問が来訪者に与える影響は, 多くの研究で報告されている. 例えば, 動物園や水族館への訪問が保全に関する知識や自己効力感, 行動意欲の向上につながることや [6, 21], 大人の来園者の保全に対する態度と理解にも影響を与えることが明らかになっている [11]. さらに, Jensen ら [16] は訪問によって得られた生物多様性に関する学習成果が訪問後も長期にわたって持続する可能性があることを示しており, Bruni ら [3] は動物園体験が自然とのつながりを潜在的に促進することを述べている.

一方で, これらの教育的効果や態度変容が時間とともに薄れることや実際の行動変容にはつながらないという指摘 [1] もある. Malamud ら [19] は態度変容や教育, 保全への関心促進への説得力のある証拠が存在しないと主張しており, 来訪者の体験の効果を高めるためには, 展示手法や教育プログラムに対する工夫が求められている.

本研究では, ペンギンをただ眺めるだけの鑑賞にとどまらず, 個体に着目して観察する体験を提供することで来訪者の学びや意識の変容を促すことを目指している. ペンギンの腹部模様を描いて個体を識別するという手法を通じて, 来訪者により深い観察を促し, 動物への興味や共感, 学びにつながる新たなアプローチを提案する.

2.2 動物園の来園者に関する研究

動物園や水族館の来訪者は, 娯楽や癒しといった目的に加え [24, 5], 教育的な体験を求めて訪れる場合も多い [33, 2]. 実際, Davey [9] は来園者の主要な訪問動機として教育と珍しい動物の観察を挙げており, 来園者は一般市民や学生に比べて動物園や飼育動物に対して肯定的な認識を持つことが示されている. また, Packer [28] は, 来訪者が学ぶことを意図していなくても, 学習体験に引き込まれたり, 無意識的に学びを求めたりすることを示している.

特に, 参加型展示や自然に近い環境での展示は, 来訪者の滞在時間や関心の向上につながることを示されている [12, 22]. さらに, 展示に教育的な介入を組み合わせることで, 知識の習得や理解に寄与することが明らかになっている [8, 34]. また, 来訪者が施設の動物や展示に対して抱く肯定的な認識がもっと学びたいという意欲につながることや, 動物とのインタラクティブな体験を通じて学習効果が高まることも指摘されている [13, 7]. さらに動物への共感 は保全行動の変容にも結びつく可能性を持つ [5, 35]. このように, 動物園や水族館での学習や保全意識を高めるためには, 教育的な介入や展示環境の工夫を

して, 動物への関心や共感を高めることが重要であるといえる.

Luebke ら [18] は, 動物を見る際の情緒的な反応が来訪者の感じる楽しさと強く関連していることを示している. Lucardie [17] は, 楽しさや喜びが記憶や学習の効果を高めることを明らかにしており, 来訪者の楽しみを引き出し, 関心や学習効果を高める仕組みが求められる.

本研究では, ペンギンの模様を自ら描いて検索するという能動的かつインタラクティブな観察手法を用いている. 個体ごとの特徴に着目し, その違いを確かめたり, 個体情報を検索したりすることで来訪者の気づきや理解が促進され, 動物への関心や共感につながると考えられる.

2.3 生き物の個体識別に関する研究

種の適切な管理や野生動物管理のため, 動物の個体識別が行われている. Burghardt ら [4] は, 野生のアフリカペンギンの集団の管理や保全などを目的として, 腹部模様の映像からリアルタイムに個体識別するプロトタイプシステムを開発している. また, RGB 画像による外見ベースのペンギン検出と, ハイパースペクトル情報によるスペクトルベースの個体識別を組み合わせた手法により, ペンギンの群れ画像から個体を識別するシステムも提案されている [27].

他にも, 個体数把握や行動研究などを目的として, ヒレなどの傷跡からアシカの個体を識別する手法 [20] や, 深度画像を用いたニワトリの個体識別手法 [36] が提案されている. Duyck ら [10] は生態系の監視や保全のために, 動物の個体識別をする画像検索システムを開発している. このように, 個体固有の模様などの特徴を用いて画像から識別した研究は多く行われている.

一方で, 本研究はユーザが能動的に観察し, 描いた斑点描画の特徴から個体を識別することを目指しており, ユーザの描画には曖昧性やばらつきがあるため単純な画像処理での認識ではなく, 人の描画特性を考慮した類似度判定が必要となる.

3 システムデザイン

3.1 システム設計

本研究では, 実際の水族館での一般利用を考慮し, 以下の必要要件を考慮してシステムを設計した.

- **観察中の個体の特定ニーズ**: サンシャイン水族館には, 個体の名前を確認する手段がなく, 来訪者は観察している個体とその情報を結びつけにくい状況にあった. そこで, 腹部模様を描画して検索する機能を主とし, 観察対象の個体を特定できる仕組みを実現する.
- **幅広い来訪者への対応**: 小さな子どもでも使

えるよう、描画インタフェースやボタン表記はひらがなを用いる必要がある。

- **検索網羅性の確保**：過去の研究では、システム上で対象とする個体数が限られていた。本研究では水族館からすべての個体情報を提供してもらい、全 47 羽を検索対象とする。
- **個体情報へのアクセス性**：描画が難しい場合にも個体情報にアクセスできるようにするため、サンシャイン水族館のすべてのペンギンを一覧で把握可能とする。サンシャイン水族館では草原・天空・バックヤードの 3 エリアにペンギンが分かれているが、図鑑機能では来訪者が直接観察できないバックヤード個体も含めて一覧化する。
- **来館時以外の利用ニーズ**：館内で観察するだけでなく、自宅や他の場面でもペンギンの情報を参照したい来訪者のために、アルバム機能を導入する。これにより、利用者は過去に描画・検索したペンギンの情報を保存し、過去の観察の振り返りや、思い出としての閲覧が可能となる。

なお、基本的な描画インタフェースと検索アルゴリズムは、我々のこれまでの研究 [26, 37] に基づいている。

3.2 検索用データセット構築

本研究では、サンシャイン水族館で飼育されているケープペンギン全 47 羽を対象に検索用データセットを構築した。各個体の正面写真を提示し、20~24 歳の大学生・大学院生 30 名（男性 21 名、女性 9 名）に腹部模様を描画してもらい、その描画データからテンプレートデータを作成した。これは、ひとによる描画のばらつきを抑えるとともに、検索時にユーザ描画との類似度を算出し、個体を推定する際の基準として用いるものである。得られた描画は先行研究 [26] と同様に腹部を 3×3 に分割してベクトル化し、個体ごとに平均をとることで検索に用いるテンプレートデータとした。

3.3 システム概要

3.1 節の要件をもとに、システム「ペンさく」を実装した。ペンさくは以下の画面群によって構成されている。

- **ホーム画面**：ユーザはログインまたはゲストとして利用を選択し、「おえかき」ボタンから描画を開始することができる。ログインして利用することで、アルバム機能が利用可能となる（図 1 左上）。
- **描画面**：展示中のペンギンを観察しながら、腹部模様を描画する。なお、障害物やペンギ



図 1. ペンさくの各画面の UI：上から順に、ホーム・描画、検索結果・個体詳細、図鑑・アルバム画面

ンの動きなどにより腹部の一部が観察できない場合は「みえないところ」として領域を指定することができる。ここで指定された領域は検索における類似度計算から除外される。描画が終了し「かんせい」ボタンを押すと、検索結果画面に遷移する（図1右上）。

- **検索結果画面**：描画が完了すると、入力された模様とテンプレートデータとの類似度に基づき、候補となるペンギンがランキング形式で提示される（図1左中央）。
- **個体詳細画面**：検索結果から特定の個体を選択すると、その個体の画像・名前・バンド色・プロフィール情報などを閲覧できる。観察していた個体と一致していた場合、「この子だった!」ボタンを押すと描画データがデータベースに送信され、ユーザーのアルバムにも記録される（図1右中央）。
- **図鑑画面**：サンシャイン水族館で飼育されているすべてのペンギンを一覧形式で閲覧することができる。各個体の顔画像をクリックすると、その個体の画像・名前・バンド色・プロフィール情報を表示するモーダルウィンドウが開く（図1左下）。
- **アルバム画面**：ログイン状態のユーザーは、過去に描画・検索したペンギンの描画データ、該当個体の画像、検索日時を記録・閲覧できる。アルバムはホーム画面からアクセス可能であり、再訪時の振り返りや思い出として活用できる（図1右下）。

また、ユーザーはホーム画面・描画面・図鑑画面をヘッダーのナビゲーションメニューから自由に遷移することができる。

本システム「ペンさく¹」はフロントエンドをJavaScriptのフレームワークであるVue.js、バックエンドをPHPで実装し、データベースとしてMySQLを使用して実装した。

4 実証実験

本研究では、システム「ペンさく」を用いてペンギンの腹部模様を描画しながら個体を検索することが、来訪者の観察行動にどのような影響を与えるかを調査するため、サンシャイン水族館にて実証実験を実施した。

実験期間は2024年10月15日（火）～21日（月）のうち、19日（土）を除く計6日間であった。実験はペンギン展示場「草原エリア」および「天空エリア」にて行い、来訪者が自然にシステムにアクセスできるよう、それぞれの展示前にペンさくの説明と



図2. 掲示物の例（左から草原エリア，天空エリア）



図3. 実験の様子

QRコードを記載した掲示物（図2）を設置した。実験の様子を図3に示す。

実験期間中、掲示物は常時掲示し、QRコードを読み込んだ来訪者が自身のスマートフォンからシステムにアクセスできるようにした。また、必要に応じて掲示物を読んでいる来訪者に対して実験者が声をかけ、システムの概要や使い方について口頭で説明を行う場面もあった。

実験中は3名の実験者が毎日10:00～16:00に展示場周辺に待機し、来訪者の行動を観察した。主に、掲示物を読む様子、QRコードを読み込む様子、システムを操作している様子などを対象とし、観察された行動や発話内容は逐次メモとして記録した。

5 結果と考察

5.1 ペンさく利用者の概要

実験期間中、実験者が展示場にて観察・記録した来訪者の行動メモをもとに、筆頭著者が観察および発話の特徴を集計した。行動メモの総文字数は13,222字であった。行動メモからペンさく利用者の数を曜日別・使用形態別にまとめたものを表1に、年代別・性別にまとめたものを表2に示す。

実験では多くの来訪者が展示前の掲示物を興味深

¹ <https://pensaku.com/>

く読み、QRコードを通じてシステムにアクセスする様子が確認された。6日間の実験期間中、記録できた範囲で167組271名がペンさくを利用していた。ペンさくの使用形態については、複数人で1つの端末を用いて使用していた来訪者が79組と最も多かった。特に子連れの親子が、親の端末を使って一緒に描画している様子が数多く確認できた。また、水族館の来訪者が多くなる週末から月曜にかけて利用者が増加する傾向がみられ、最終日には1日で55組が利用していた。年代別では、小学生を中心とする10歳以下が最も多く78名、次いで20代が80名、30代が44名と、全体として子どもとその保護者を中心とした幅広い年齢層の来訪者が利用していた。

ペンさくのホーム画面に設置した任意回答形式のアンケートには11名の回答が得られた。回答者の多くはリピーター層であり、ペンギンへの好感度が高い来訪者であった。訪問目的は「子どものため」が最多で、1人や家族連れが中心であった。また、ぼてとや武蔵など具体的な個体名を覚えたことが報告された。

5.2 コミュニケーションの促進

ペンさくの利用者からは、ペンギンの個体名を用いた会話が頻繁にみられた。サンシャイン水族館ではこれまで個体名を来訪者が直接確認する手段がなかったが、ペンさくを通じて名前を検索できることで「この子はぽてとだった」「オレライムは点が多いな」など、個体名を介した会話が発生していた。

特に、親がスマートフォンを操作しながら子どもに描画を促す協調的な利用が多く観察された。兄弟間で端末を交代しながら利用する様子や、同行者と検索結果を互いに見せ合いながら話し合う姿も記録され、個体を中心とした自然なコミュニケーションが促進されていた。

5.3 観察の変化

ペンさくの利用により、個体ごとの違いに着目した観察が促されていた。描画を通じて腹部模様の特徴に気づき、「この子は点が多い」「模様が違うの知らなかった」といった発話が数多く確認された。また、「似たような柄の子がいる」「ペアのペンギンも見つけた!」といった反応もあり、個体間の関係性を考察する姿勢がみられた。

さらに、ペンさくを利用した来訪者は他の来訪者と比較して展示前での滞在時間が増加していた。実際に、「これ見てると時間経っちゃう」といった発話に加えて、子どもが夢中で描画に取り組む様子や、その結果親が何度もそろそろ移動しようと促す場面も観察され、観察体験の継続性が高まっていた。滞在時間は動物への関心度を測る指標とされており[25]、この結果はペンギンへの興味関心の高まりを示すものといえる。

表 1. 使用形態別の利用組数（組）

使用形態	火～金	日・月	合計
1人	16	17	33
複数人・1端末	28	51	79
複数人・複数端末	25	30	55
合計	69	98	167

表 2. 年代・性別別の利用者数（人）

年代区分	男	女	合計
～10歳	36	42	78
～20歳	2	4	6
～30歳	35	45	80
～40歳	15	29	44
～50歳	15	22	37
～60歳	9	11	20
60歳～	1	4	5
合計	113	157	271

5.4 興味・親しみの形成

ペンさくを利用した来訪者のペンギンへの関心や親しみが強まる様子も観察された。アンケートには「今まで名前が分からなかったから嬉しい」「ペンギンファンとしても楽しい経験ができた」といった好意的な意見が多数寄せられた。また、検索結果をスクリーンショットで保存したり、腹部以外の余白に自分なりの観察結果を残すコメントを書いたり、システムを拡張した自主的な観察行動をとっている来訪者も確認できた（図4）。また、図鑑機能には草原や天空に加えてバックヤードで飼育されている個体も含めて一覧できるようにしたことで、来訪者のニーズに応えると同時に、展示では見られない個体への理解や愛着を広げる役割も果たしていた。

さらに、個体情報から「誕生日が近い」といったように自己との関係性を見出し親近感を示す来訪者もいた。一度検索した個体を覚えて、後に再び探そうとする子どもの姿も確認され、ペンギンへの親しみや愛着につながっていた。これらは、ペンギンを単なる展示対象ではなく関係性のある存在として認識するきっかけとなっていたと考えられる。



図 4. 応用的な描画の例

5.5 課題

実験では、掲示を読んでもシステムにアクセスしない来訪者が数多くおり、高齢者や外国人来訪者ではスマートフォン操作や言語が障壁となった可能性がある。また、「模様が见えない」「動いてしまって描きづらい」といった声があり、観察環境による制約があった。現状では見えない部分を除外する機能を備えているが、さらなる工夫が求められる。

検索精度に関しては、個体を特定できない事例や「当たらない」という指摘があったため、検索アルゴリズムやテンプレートデータの改善により検索精度を向上させることが必要である。本研究のデータセットは静止画を用いた描画タスクによって構築されており、実際の観察環境で見られるようなペンギンの動きや部分的に見えない状況は十分に考慮されていない。そのため、動画を用いた描画タスクや、多様な観察条件を含むデータ収集を行い、実環境に適したテンプレート構築をする必要がある。

さらに、一部の子どもからはペンギンを観察せず画面を塗りつぶすなど自由な落書きとして利用する姿もみられた(図5)。このような利用は、観察支援という本来の目的から離れてしまう可能性があるため、観察を自然に促す仕組みが必要と考えられる。

5.6 他施設への導入可能性

本システムは、サンシャイン水族館における実証実験を通じて有用性を確認した。今後、他の水族館や動物園に導入するためには、まず施設ごとの個体情報(名前、バンド色、性別、説明など)の整備が不可欠である。特に誕生日については来訪者の反応が大きく、親近感や関係性を感じるきっかけとなっていたため、他施設での活用においても積極的に取り入れるべきである。

また、本研究ではバックヤードを含む全個体を閲覧できる図鑑機能を実装した。図鑑に対する肯定的な反応が多く得られ、他施設で導入する際にも効果的であると考えられる。

導入にあたり必要な作業は、これら基本情報の整理に加え、検索のためのデータセットの構築が必要であり、個体の写真撮影や模様の特徴をもとにしたテンプレートデータの作成を行う必要がある。本研究では大学生を対象とした描画タスクによってテンプレートを作成したが、今後は効率的な収集方法を検討する必要がある。

これらの準備を整えることで、本システムは他施設にも導入可能であり、来訪者の観察行動や意識変容を支援する仕組みとして活用できると考えられる。

5.7 今後の展望

今後は、検索精度の向上に向けてアルゴリズムやテンプレートデータを改良するとともに、外国語対応やユーザインタフェースの改善を進めていく。ま

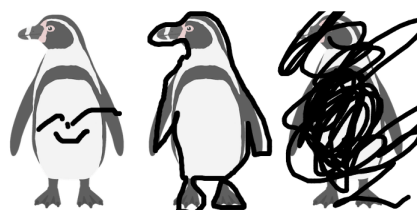


図 5. 模様に基づかない自由描画の例

た、システム上で描画・検索・閲覧のログを記録・分析し、来訪者行動をより詳細に把握する予定である。

さらに、本実験は短期間の実施であったため、今後は長期的な運用を通じて継続的な利用や利用者の行動変容に与える影響を検証する。また、他の施設での実証に向けた計画も進めており、他の水族館への展開を通じて異なる来訪者層や施設環境における有用性を評価する予定である。将来的には教育や福祉との連携を視野に入れた実装を進めていく。

6 おわりに

本研究は、水族館の来訪者にペンギンの個体に着目した観察を促すため、ペンギンの腹部模様の描画から個体を検索するシステム「ペンさく」を開発し、水族館における実証実験からシステムが人の観察行動に与える影響を調査した。その結果6日間の実験期間中、167組271名がシステムを利用し、同行者とのコミュニケーションの促進や、個体に焦点を当てた観察が促される様子が確認された。展示前での滞在時間の増加や、個体名を通じた興味・親しみの高まりも見られ、観察体験を豊かにする効果が示された。

また、サンシャイン水族館と連携し、実際の飼育個体すべてを対象とした検索や図鑑機能を来訪者が自由にアクセスできる利用形態を整え、より広く水族館・動物園で資する仕組みづくりの第一段階を示した。今後は、検索精度の改善をしたうえで、システムの長期的な運用や、複数施設での実証実験を行い、異なる環境におけるシステムの有用性や観察の継続性について詳細な調査をしていく。

さらに本研究は、動物観察における主体性や関心を高める新たな支援のあり方を提示し、水族館における教育・福祉的価値の向上につながる仕組みへと発展させていくことを目指す。来訪者が自ら個体を識別し、その違いや特徴に気づく体験は、単なる観賞を超えて動物との関係性を築くきっかけとなる。また、来訪者による個体の観察結果を施設と共有する仕組みが実現すれば、飼育者だけでは把握しきれない異常や変化を早期に察知し、動物の健康や福祉の向上に役立つ可能性がある。こうした経験は来訪者にとっても学びや保全に対する参加意識を育み、動物園・水族館の役割に貢献すると考えられる。

参考文献

- [1] L. M. Adelman, J. H. Falk, and S. James. Impact of National Aquarium in Baltimore on visitors' conservation attitudes, behavior, and knowledge. *Curator: The Museum Journal*, 43(1):33–61, 2000.
- [2] R. Ballantyne, J. Packer, K. Hughes, and L. Dierking. Conservation learning in wildlife tourism settings: Lessons from research in zoos and aquariums. *Environmental Education Research*, 13(3):367–383, 2007.
- [3] C. M. Bruni, J. Fraser, and P. W. Schultz. The value of zoo experiences for connecting people with nature. *Visitor Studies*, 11(2):139–150, 2008.
- [4] T. Burghardt, B. Thomas, P. J. Barham, and J. Calic. Automated visual recognition of individual African penguins. In *Fifth International Penguin Conference*, 2004.
- [5] S. Clayton, J. Fraser, and C. D. Saunders. Zoo experiences: Conversations, connections, and concern for animals. *Zoo Biology: Published in affiliation with the American Zoo and Aquarium Association*, 28(5):377–397, 2009.
- [6] S. Clayton, A.-C. Prévot, L. Germain, and M. Saint-Jalme. Public Support for Biodiversity After a Zoo Visit: Environmental Concern, Conservation Knowledge, and Self-Efficacy. *Curator: The Museum Journal*, 60(1):87–100, 2017.
- [7] C. K. Collins, S. McKeown, and R. O'Riordan. Does an animal–visitor interactive experience drive conservation action? *Journal of Zoological and Botanical Gardens*, 2(3):473–486, 2021.
- [8] C. Collins, I. Corkery, S. McKeown, L. McSweeney, K. Flannery, D. Kennedy, and R. O'Riordan. Quantifying the long-term impact of zoological education: a study of learning in a zoo and an aquarium. *Environmental Education Research*, 26(7):1008–1026, 2020.
- [9] G. Davey. Public perceptions in urban China toward zoos and their animal welfare. *Human Dimensions of Wildlife*, 12(5):367–374, 2007.
- [10] J. Duyck, C. Finn, A. Hutcheon, P. Vera, J. Salas, and S. Ravela. Sloop: A pattern retrieval engine for individual animal identification. *Pattern Recognition*, 48(4):1059–1073, 2015.
- [11] J. H. Falk, E. M. Reinhard, C. L. Vernon, K. Bronnenkant, J. E. Heimlich, and N. L. Deans. Why zoos and aquariums matter: Assessing the impact of a visit to a zoo or aquarium. 2007.
- [12] E. J. Fernandez, M. A. Tamborski, S. R. Pickens, and W. Timberlake. Animal–visitor interactions in the modern zoo: Conflicts and interventions. *Applied Animal Behaviour Science*, 120(1-2):1–8, 2009.
- [13] A. M. Godinez and E. J. Fernandez. What is the zoo experience? How zoos impact a visitor's behaviors, perceptions, and conservation efforts. *Frontiers in Psychology*, 10:1746, 2019.
- [14] P. J. Greenwell, L. M. Riley, R. Lemos de Figueiredo, J. E. Brereton, A. Mooney, and P. E. Rose. The Societal Value of the Modern Zoo: A Commentary on How Zoos Can Positively Impact on Human Populations Locally and Globally. *Journal of Zoological and Botanical Gardens*, 4(1):53–69, 2023.
- [15] JAZA. (公社) 日本動物園水族館協会の 4 つの役割.
- [16] E. A. Jensen, A. Moss, and M. Gusset. Quantifying long-term impact of zoo and aquarium visits on biodiversity-related learning outcomes. *Zoo biology*, 36(4):294–297, 2017.
- [17] D. Lucardie. The impact of fun and enjoyment on adult's learning. *Procedia-Social and behavioral sciences*, 142:439–446, 2014.
- [18] J. F. Luebke and J. Matiassek. An exploratory study of zoo visitors' exhibit experiences and reactions. *Zoo biology*, 32(4):407–416, 2013.
- [19] R. Malamud, R. Broglio, L. Marino, S. O. Lilienfeld, and N. Nobis. Do zoos and aquariums promote attitude change in visitors? A critical evaluation of the American zoo and aquarium study. *Society & Animals*, 18(2):126–138, 2010.
- [20] S. D. McConkey. Photographic identification of the New Zealand sea lion: a new technique. 1999.
- [21] X. McNally, T. L. Webb, C. Smith, A. Moss, and J. Gibson-Miller. A meta-analysis of the effect of visiting zoos and aquariums on visitors' conservation knowledge, beliefs, and behavior. *Conservation Biology*, 39(1):e14237, 2025.
- [22] L. J. Miller. Visitor reaction to pacing behavior: Influence on the perception of animal care and interest in supporting zoological institutions. *Zoo Biology*, 31(2):242–248, 2012.
- [23] R. Miranda, N. Escribano, M. Casas, A. Pino-del Carpio, and A. Villarroya. The Role of Zoos and Aquariums in a Changing World. *Annual Review of Animal Biosciences*, 11:287–306, 2023.
- [24] J. M. Morgan and M. Hodgkinson. The motivation and social orientation of visitors attending a contemporary zoological park. *Environment and behavior*, 31(2):227–239, 1999.
- [25] A. Moss and M. Esson. Visitor interest in zoo animals and the implications for collection planning and zoo education programmes. *Zoo biology*, 29(6):715–731, 2010.
- [26] Y. Nakagawa and S. Nakamura. Drawing-type Search Method Focusing on Penguin's Abdominal Patterns for Enriching Observation Experiences in an Aquarium. In *Proceedings of the 2024 International Conference on Advanced Visual Interfaces, AVI '24*, pp. 1–3, 2024.
- [27] Y. Noboru, Y. Ozasa, and M. Tanaka. Appearance-and-Spectral-Based Identification

- System for Penguin Individuals. *ITE Transactions on Media Technology and Applications*, 13(2):211–220, 2025.
- [28] J. Packer. Learning for fun: The unique contribution of educational leisure experiences. *Curator: The Museum Journal*, 49(3):329–344, 2006.
- [29] P. E. Rose and L. M. Riley. Expanding the role of the future zoo: Wellbeing should become the fifth aim for modern zoos. *Frontiers in Psychology*, 13:1018722, 2022.
- [30] P. Smith, J. Mann, and A. Marsh. Empathy for wildlife: The importance of the individual. *Ambio*, 53(9):1269–1280, 2024.
- [31] S. L. Spooner, S. L. Walker, S. Dowell, and A. Moss. The value of zoos for species and society: The need for a new model. *Biological Conservation*, 279:109925, 2023.
- [32] S. Thomas. Social Change for Conservation – The World Zoo and Aquarium Conservation Education Strategy.
- [33] A. Villarroja, R. Miranda, A. Pino-del Carpio, and M. Casas. Social Perception of Zoos and Aquariums: What We Know and How We Know It. *Animals*, 14(24):3671, 2024.
- [34] B. Wagoner and E. Jensen. Science learning at the zoo: Evaluating children’s developing understanding of animals and their habitats. *Psychology & Society*, 3(1):65–76, 2010.
- [35] A. Young, K. A. Khalil, and J. Wharton. Empathy for animals: A review of the existing literature. *Curator: The Museum Journal*, 61(2):327–343, 2018.
- [36] B. Zhang, Y. Qiu, X. Wang, H. Lu, and F. Wang. Research on the method of individual identification of chickens based on depth image. In *Journal of Physics: Conference Series*, Vol. 1631, p. 012018, 2020.
- [37] 中川由貴, 中村聡史. ペンギンの腹部模様に着目した斑点描画型検索・観察手法とその検証. 情報処理学会論文誌, 65(12):1842–1853, 2024.

未来ビジョン

本研究は、来訪者の観察体験を技術によって拡張するだけでなく、動物園・水族館の重要な役割である教育や動物福祉意識の向上に寄与する点で意義を持つ。来訪者が自らペンギンを「個体」として識別する体験を通じて観察行動や理解がどのように変化するかを明らかにし、教育・福祉・エンタテインメントを横断する新たな動物との関わり方を提案する。

今後はまず、来訪者の行動変容や理解の深まりを定量的に捉える評価指標を確立し、展示設計や福祉評価の基盤的知見を提供することを目指す。例えば、来訪者による個体の観察記録が、飼育担当者にとって異常行動や健康状態の兆候を早期に把握する手がかりとなるなど、展示環境や飼育管理の改善に結びつく可能性を探る。さらに、来訪者による個体観察の深さや持続性を動物福祉を考慮した展示評価の新

たな指標として位置づけることにより、展示設計や施設運営に新しい視点を提供できると考える。

また本手法は、個体ごとに視覚的特徴を有する他種にも展開可能であり、より良い観察行動への変容を促す一般的アプローチとしての応用可能性も見込まれる。

本研究の取り組みは、レジャーとしての目的だけでなく動物園・水族館における展示・教育・研究のあり方を再定義する一助となることが期待される。将来的に多くの施設で導入されることで、来訪者の観察行動や学びを広く比較・分析できる基盤が整い、楽しみながら学べる仕組みとして社会に定着していくことを目指す。

GPStroke: 複数の歩行軌跡に基づく GPS アート制作手法の提案

尾本 悦基* 塚田 浩二*

概要. 歩行活動は健康維持に効果があることが知られているが、多忙な現代人にとって定期的な運動時間の確保は困難である。そこで本研究では、日常生活の中で手軽に歩行活動を取り入れるために、日常の断片的な歩行軌跡を「ストローク」として収集し、後から組み合わせて GPS アート作品を制作するシステム「GPStroke」を提案する。本システムではユーザは通勤・買い物などの日常移動時に軌跡を記録し、後でこれらの歩行軌跡をストロークとして利用して、作品を制作できる。歩行活動と作品制作を時間的に分離することで、忙しい人々でも日常の移動を活用し、楽しみながら作品を制作できる仕組みを目指す。本稿では、GPStroke のシステム設計と実装について紹介し、さらに基礎的な評価実験を通して提案システムの効果や課題について検討する。

1 はじめに

歩行活動を含む身体活動は、心血管疾患、糖尿病、がんなどの非感染症疾患のリスクを低減し、メンタルヘルスの向上にも効果があることが報告されている [9]。WHO は成人に週 150 分以上の中強度運動を推奨しているが、2024 年の調査 [8] では、世界の成人の約 31 %がこの推奨レベルを満たしておらず、身体活動不足の割合は 2000 年の 23.4 %から増加傾向にある。身体活動の障壁に関する研究 [1][4] では、身体活動不足の大きな要因として「時間不足」が一貫して指摘されており、仕事や家事といった日常生活の責務と運動の両立が困難であることが明らかにされている。このような背景から、限られた時間の中でも実践可能な歩行促進手法が求められている。

従来の歩行促進アプリケーションとして、Fitbit¹ や Apple Health²などが広く利用されている。これらは歩数や移動距離といった数値的フィードバックに基づいており、身体活動の把握に有用である一方で、歩行そのものを楽しむ仕組みや継続的な動機づけには限界がある。また、PokémonGO[5] のようにゲーミフィケーションを活用した位置情報ゲームは、

歩行意欲の向上に一定の効果を示しているが、プレイのためにまとまった時間を必要とするため、多忙な人々には導入が難しい。

本研究では、GPS アートを活用して、楽しみながら歩行活動を促進する手法に着目した。GPS アートは歩行経路を地図上の線画として可視化することで、移動に創造的な価値を付与する活動である。この手法は、単なる移動を作品制作という目的のある活動に変換し、歩行への動機づけを高める可能性を持つ。しかし従来の GPS アートは、事前に描く作品のテーマを決め、移動経路を厳密に計画して一筆書きで作品を描く手法が一般的であった。これは創造的で魅力的な活動である一方で、まとまった時間と計画性が必要であり、日常生活に気軽に取り入れることは困難であった。

そこで本研究では、日常生活の中で得られる断片的な歩行軌跡をストロークとして収集し、それらを後から組み合わせて GPS アート作品を制作するシステム GPStroke を提案する (図 1)。この仕組みにより、通勤や買い物といった日常の移動を創作の素材として活用し、短時間の歩行を積み重ねながら創造的な体験を楽しむことが可能となると考える。本論文では、まず歩行軌跡を組み合わせた新たな GPS アートの制作手法を実装し、小規模な運用を通してその基礎的な効果を検証する。さらに、将来的な歩行促進手法としての可能性を議論する。なお、本研究は

Copyright is held by the author(s).

* 公立はこだて未来大学

¹ <https://www.fitbit.com/>

² <https://www.apple.com/health/>

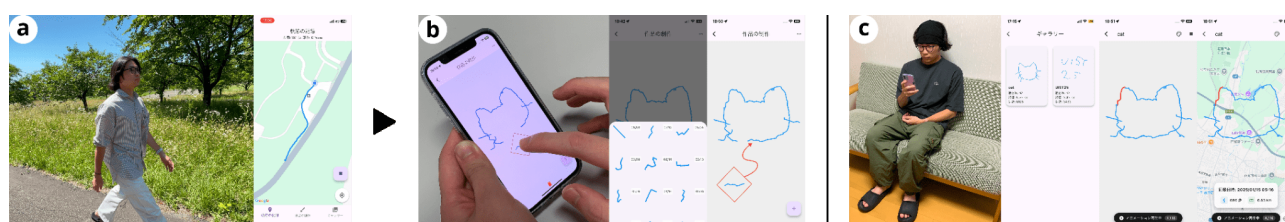


図 1. GPStroke の概要. (a) 日常生活での軌跡記録, (b) ストロークの組み合わせによる作品制作, (c) 作品の閲覧

WISS2024 のデモ発表 [12] で発表し、SIGGRAPH Asia の Posters で発表予定の内容に加筆したものである。

2 関連研究

本研究の関連研究を「日常での歩行促進手法」「GPS アートを活用した取り組みや支援手法」の観点から紹介する。

2.1 日常生活での歩行促進手法

PokémonGO[5] は、スマートフォンの GPS を利用し、現実世界を舞台にプレイヤーが移動しながらポケモンを捕まえたりアイテムを獲得したりする位置情報ゲームである。このゲームは、歩行距離に応じた報酬システムや、地域ごとのイベントを活用することで、プレイヤーの継続的な移動を促進する仕組みを備えている。

SPOBY[2] は本来乗り物で移動すべき距離を歩行や自転車で移動した場合に、抑制された CO2 排出量を脱炭素量として計測し、スコア化するアプリケーションである。脱炭素量や活動量に応じてポイントが蓄積され、アプリ内に掲載される景品と交換できる報酬システムを通じて歩行を促進している。

古田ら [11] は、歩行データを収集する作業のモチベーションを向上させるために、ゲーミフィケーションを活用した位置情報連動型陣取りゲームを提案している。このシステムでは、ユーザが歩行データを収集することで、現実世界をフィールドに見立てた陣地を獲得でき、陣地獲得数に応じてランキングが表示される。2 週間にわたる評価実験では、システムが歩行データの投稿数や歩行距離の増加に寄与し、ユーザの歩行活動を促進したことが示された。

テクテクライフ [10] は、ユーザが歩いた経路を地図上に色塗りして可視化することで、行動範囲の拡大や歩行意欲を高めることを目的としたアプリケーションである。ユーザが歩行を通して地図を塗りつぶしていくことで、自身の移動範囲が視覚的に把握できるようになり、達成感を得られる仕組みになっている。また、同じ経路を繰り返し歩くと色が濃くなるため、継続的な利用を促している。

2.2 GPS アートを活用した取り組みや支援手法

GPS アートは、移動経路を活用して地図上に描画を行う創造的な活動であり、地域活性化や観光促進、健康促進の手段としても注目されている。

地域振興を目的とした事例として、「北陸 100 文彩プロジェクト」[3] では、GPS アートを活用して北陸地方の地形や、名所を題材とした作品を制作する活動が行われている。このプロジェクトでは、参加者が地域を探索しながらアートを完成させることで、地域の歴史や文化に触れる機会を提供している。

Strava Art[6] は、ランニングやウォーキングの経路を地図上に描画することで、ユーザが自身の運動記録を創造的なアート作品として楽しむプラットフォームである。ユーザは自身の移動経路を SNS 上で共有することができ、他のユーザとの交流を通じて運動意欲を高める効果が報告されている。

Waschk ら [7] はスケッチや形状を入力として、マップ上でウォーキングルートを自動生成するシステムを提案し、従来手動で行われていたルート計画の労力軽減を図っている。

2.3 本研究の独自性

本研究の独自性は、通勤や買い物といった日常移動の歩行軌跡を記録し、それらを後から組み合わせて作品を制作するアプローチにある。複数の断片的な軌跡を組み合わせることで、運動のためのまとまった時間を確保せずに短時間の歩行を繰り返しながら作品を完成させることが可能となる。このように、時間制約の課題を解決しつつ GPS アートの制作を支援して、将来的に歩行促進につなげることを目指している。

3 提案: GPStroke

3.1 システム概要

本研究では複数の歩行軌跡に基づく GPS アート制作支援手法 GPStroke を提案し、モバイルアプリケーションとして実装する。本システムの利用の流れは図 1 の通りである。まず、ユーザはスマートフォンの GPS ログ機能を用いて日常移動時の歩行軌跡を記録する。記録された軌跡はアプリ内にストロークとして保存される。次に、ユーザは蓄積されたストロークの中から任意のものを選択し、キャンバス上で配置・移動・回転を行うことで作品を制作する。完成した作品はアプリ内に保存され、後から閲覧や振り返りが可能である。本システムは、以下の 3 つの主要な機能から構成されている。各機能の詳細については 4 章の実装にて詳しく説明する。

1. ユーザの歩行軌跡を記録する機能
2. 複数の軌跡を組み合わせて作品を制作する機能
3. 制作した作品を閲覧する機能

3.2 システム構成

GPStroke のシステム構成図を図 2 に示す。スマートフォンに内蔵された GPS ログ機能を用いてユーザの歩行軌跡を記録し、同時に歩数や移動距離のデータを取得する。技術的な実装として、GPS トラッキングではユーザの移動速度に応じて位置情報の取得間隔を動的に調整し、バッテリー消費を抑えながら正確な軌跡記録を実現している。作品制作機能では、タッチ操作による軌跡の配置・移動・回転を可

能にし、複数の軌跡を自由に組み合わせることが出来る。データ管理には、ローカルストレージとクラウドサーバー（Firebase Firestore）を併用し、軌跡データや作品情報をサーバー側で一元管理することで、デバイス間でのデータ共有を可能にしている。本アプリでは、iOS および Android のスマートフォンを対象としている。開発言語にはクロスプラットフォーム対応が可能な Flutter を使用し、これにより iOS と Android の両方で動作する。制作された作品はアプリから一覧して振り返ることができる。振り返り機能では、各ストロークがどの場所で記録されているのか等を合わせて楽しむことができる。

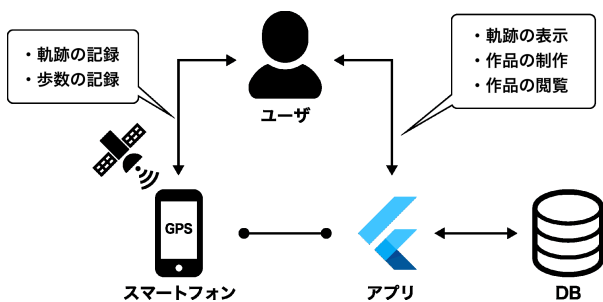


図 2. システム構成図

4 実装

本章では、システムの実装を「軌跡を記録する機能」「作品を制作する機能」「作品を振り返る機能」という機能に分けて説明する。

4.1 軌跡を記録する機能

ユーザは通勤や買い物などの日常の移動時に、記録ボタンをタップするだけでスマートフォンの GPS を用いて歩行軌跡をストロークとして記録できる。記録されたストロークは日付や移動距離といった情報とともに保存され、のちの作品制作のために蓄積される。記録中の軌跡は Google Maps API を用いて地図上に Polyline としてリアルタイムに描画される。これにより、ユーザは必要に応じて画面を確認すれば現在の移動経路を逐次確認でき、記録が正常に行われていることを視覚的に把握できる（図 3）。また、本アプリはバックグラウンド動作に対応しており、記録ボタンをタップした後はアプリを閉じて位置情報の取得が継続される。そのため、歩行中に画面を注視する必要はなく、通常の移動行動を妨げることなく利用できる。記録の停止は再度ボタンをタップすることで行われ、取得されたデータはローカルデータベースおよび Firestore に保存される。

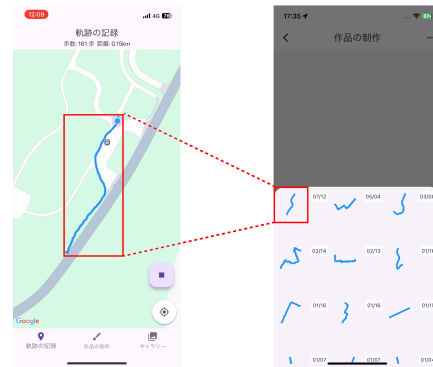


図 3. 軌跡記録機能の外観

4.2 作品を制作する機能

ユーザは、蓄積されたストロークのリストから素材として使いたいものを選択する。ユーザは、記録された道の形状からインスピレーションを得て、例えば「カーブした道」を「動物の耳」に見立てるなど、パズルのようにストロークを組み合わせることで 1 つの作品を創り上げる。記録済みの軌跡は一覧表示され、ユーザは任意のストロークを選んでキャンバスに配置できる（図 4 (a)）。未使用のストロークはリストに残り、使用済みのものは除外されるため、毎回新しい歩行を取り込みながら制作できる設計となっている。配置されたストロークはドラッグ操作で自由に移動でき（図 4 (b)）、さらにツイストジェスチャを用いて回転が可能である（図 4 (c)）。また、キャンバス上で不要になったストロークは画面下部のゴミ箱アイコンをタップすることで取り除くことができる。一方で、本システムではストロークの拡大・縮小機能はあえて実装していない。これは、元々の歩行軌跡の形状や大きさをそのまま活かすことを重視し、記録された移動の痕跡を改変せず作品へ反映することを目的としているためである。

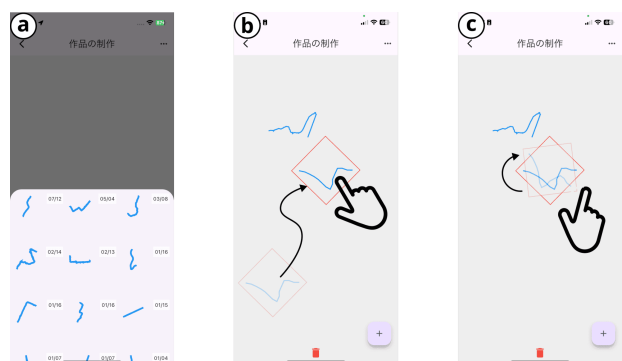


図 4. 記録した軌跡の配置/移動/回転

4.3 作品を振り返る機能

完成した作品はタイトルを付けて保存でき、これまで作った作品を一覧して振り返ることができる。一覧画面では、各作品のプレビュー画像が表示され、画像をクリックすると作品詳細画面に移行する(図5)。作品詳細画面では、色分け機能やアニメーション機能等を用意している。色分け機能では、画面右上のボタンを押すことでストローク毎に異なる色を描画し、作品がどのような移動軌跡から作られているかを分かりやすく表示できる(図6)。アニメーション機能では、完成作品を全体表示した上で、構成するストロークを記録順に1つずつ強調表示する。各強調表示に合わせて、地図上の位置情報等を合わせて表示することもでき、制作過程を段階的に振り返ることが可能となる(図7)。

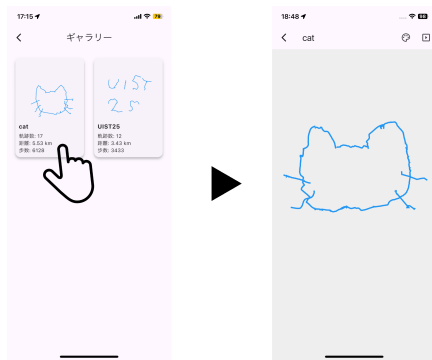


図 5. 作品の一覧画面

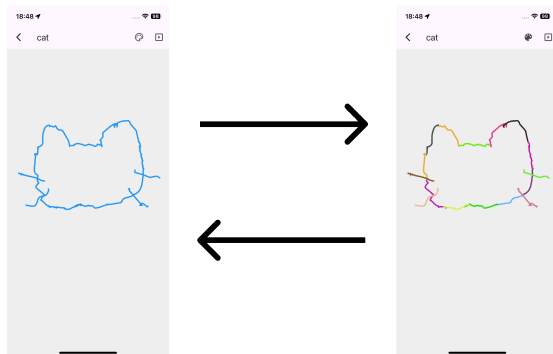


図 6. 作品閲覧時の色分け機能

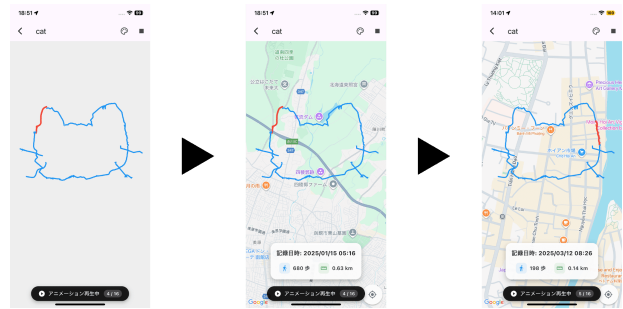


図 7. 作品閲覧時のアニメーション機能

5 運用評価

5.1 目的

小規模な運用評価として、ユーザに GPStroke を日常生活の中で利用してもらい、GPS アート制作手法としての基礎的な性能および効果を検証する。さらに、将来的に歩行促進手法として活用するための課題を調査する。これらの結果をもとに今後の改良に向けた指針を得ることを目的とする。

5.2 手法

被験者にアプリの概要を説明した後、iPhone にインストールして5日間利用してもらった。被験者には日常生活の中で歩行を記録し、記録した軌跡を用いて2つ以上の作品を制作するよう指示した。利用後にはアンケート(5段階評価及び自由記述)を実施し、楽しさ、歩行促進効果、継続使用意欲について回答を得た。被験者は21~22歳の大学生4名(男性3名、女性1名)であった。

5.3 結果と考察

5段階評価の結果を図8に示す。また被験者が制作した作品例を図9に示す。「軌跡記録機能の使いやすさ」は平均4.5と高評価で、「ワンアクションで開始・停止できる」「記録精度が高い」といった意見が得られた。一方で「開始ボタンを押し忘れることがあるため、自動記録機能がほしい」という要望もあった。「自身の軌跡で作品制作を楽しめたか」は平均3.75でやや高めだったが、標準偏差1.63とばらつきが大きく、個人差が見られた。自由記述では「他者の作品を見たい」「自分の作品を共有したい」といった要望も寄せられた。「歩行活動が増えたと感じたか」は平均2.5と低く、標準偏差1.65とばらつきが大きかった。この結果から、本アプリによる歩行促進効果は現時点では限定的であったと考えられる。自由記述では、「日常の歩行が固定化されており、特に新たな歩行活動を生む仕組みが必要」「歩行を促す通知機能やルート案内のような仕組みがあればよい」との意見が得られた。また、本評価時点では作品完成後のアニメーション機能などが未実装であっ

たため、歩行軌跡と完成作品の連携による振り返りがしづらく、歩行活動の意識的増加につながりにくかったことも要因と考えられる。一方、「今後もアプリを利用して歩行を続けたいと思うか」は平均3.75と比較的高く、歩行促進効果はまだ限定的であるものの、アプリの仕組みや可能性に対する期待を感じる利用者が多い結果となった。自由記述では、「自分の軌跡が残ることに価値を感じた」との意見があり、一方で「記録開始ボタンを押す手間やバッテリー消費への懸念がモチベーションの維持に影響する」との指摘もあった。今後は、記録の自動化や低消費電力モードの実装などにより、利用負担を軽減していきたい。

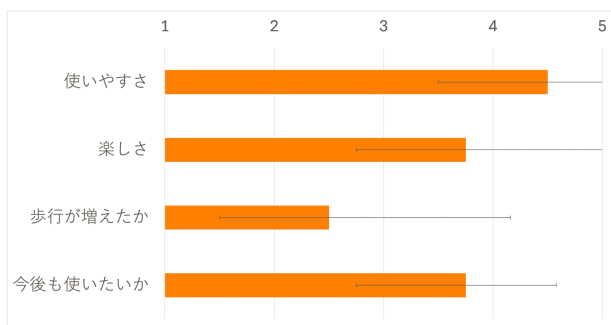


図 8. 運用評価結果

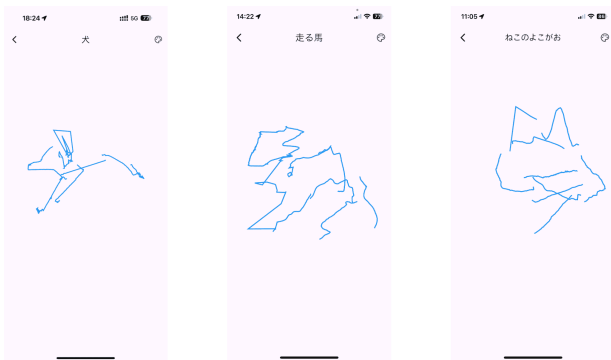


図 9. 被験者が制作した作品例（左：犬，中央：走る馬，右：猫の横顔）

6 議論

運用評価の結果、GPS アートの制作手法としては一定の効果が得られ、ユーザが自身の歩行軌跡を素材として創作を楽しむことが確認された。一方で、歩行促進の観点では新たな行動を喚起する仕組みが十分ではなく、歩行変容を支援する機能の拡充が必要である。これを踏まえ、次節以降では歩行促進を強化するための「ルート提案機能」と、歩行促進と創作体験向上の双方に寄与する「ソーシャル機能」の2つの方向から改善を検討する。

6.1 ルート提案機能

まず、「ルート提案機能」について議論する。運用評価では「日常の歩行が固定化されており、特に新たな歩行活動を生む仕組みが必要」「歩行を促す通知機能やルート案内があればよい」というニーズが確認された。現状のシステムでは、歩行軌跡は日常的な移動に依存しており、ユーザが能動的に「描きたい形」を実現することが難しいという課題がある。この課題を解決するために、ユーザが意図する軌跡形状に基づいてルートを自動生成・提示する機能が有効であると考えられる。具体的には、ユーザは地図上で目的地を指定し、キャンパス上で描きたい軌跡の形状を指でなぞることで、なぞった形状に近い実際に歩行可能なルートが提示される仕組みである（図10）。例えば、通勤途中に星の形を描きたいと思ったユーザが、キャンパス上で星型をなぞると、駅から職場までの複数のルート候補が提示され、その中から最も星型に近い経路を選択できる。この機能により、歩行軌跡を組み合わせる程度作品を制作した段階から、ユーザがデザイン的に必要なストロークを描いて、対応する歩行ルートをシステムから提案できる。また、ルートの距離や所要時間も表示することで、ユーザが日常生活の中で無理なく歩けるかを判断できるように配慮する。これにより、歩行が作品制作のための創造的な活動となり、能動的に外出する動機づけが強化されると期待している。

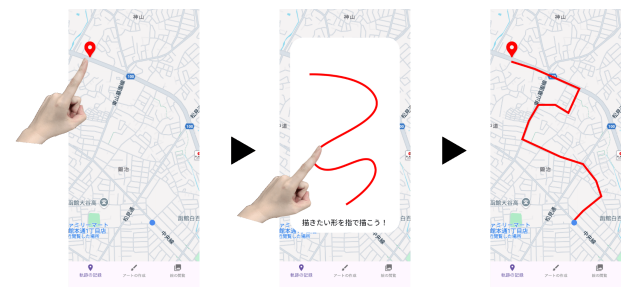


図 10. ルート提案機能のイメージ図

6.2 ソーシャル機能

次に、「ソーシャル機能」について議論する。運用評価では「他者の作品を見たい」「自分の作品を共有したい」というニーズが確認された。これは、作品制作が単なる個人の活動にとどまらず、他者との交流や競争、協働を通じてより魅力的な体験に発展しうること示している。この課題に対して、作品のアニメーション出力機能と SNS 共有機能を追加することが有効と考える。ユーザは作品を簡易なレビューとして短時間のアニメーションとして書き出し、X や Instagram などのプラットフォームで発信できるようにする。これにより、アプリ未使用者にも歩行と創作の魅力が伝わり、新規ユーザの獲得

も期待される。さらに発展的な機能として、ユーザー同士が軌跡データを交換し合える仕組みも検討する。例えば、あるユーザーが制作した美しい軌跡パターンを他のユーザーがダウンロードし、自分の地域で類似のルートを歩くことで、地域を超えた共創活動が期待される。このように、ソーシャル機能を通して、作品を通じて他者からのフィードバックを受け取れる機会が生まれ、ユーザーの創作活動や歩行継続への動機づけが促進されることを期待している。

7 まとめと今後の展望

本研究では、日常の断片的な歩行軌跡を組み合わせて GPS アートを制作するシステム GPStroke を提案した。本システムは、通勤や買い物といった日常の移動を創作の素材として活用し、短時間の歩行を積み重ねながら創造的な体験を楽しむことができる。

今後の展望として、議論で述べたような新機能を実装すると共に、異なるユーザー層や利用状況における有効性を検証し、システムの一般化を目指す。将来的には、本システムを用いた歩行促進機能についても併せて検証・改善し、人々の健康的な生活習慣形成に貢献していきたい。

謝辞

本研究の一部は、一般社団法人新雪の 2025 年度北海道 IT クリエータ発掘・育成事業（新雪プログラム）の支援を受けて実施されたものである。

参考文献

- [1] M. F. Cavallini, et al. Lack of time is the consistent barrier to physical activity and exercise in 18 to 64 year-old males and females from both South Carolina and Southern Ontario. *Journal of Physical Activity Research*, 5(2):100–106, 2020.
- [2] S. S. INC. SPOBY - 健康と脱炭素を叶えるエコライフアプリ. <https://spoby.jp/>, 2023.

(2025/11/17 確認).

- [3] Motoi Ishikawa (ほくりく GPS アート部). 北陸 100 文彩プロジェクト. <https://hokuriku-gpsart.com/100bunsai.html>, 2023. (2025/11/17 確認).
- [4] Y. S. Koh, P. Asharani, F. Devi, K. Roystonn, P. Wang, J. A. Vaingankar, E. Abdin, C. F. Sum, E. S. Lee, F. Müller-Riemenschneider, et al. A cross-sectional study on the perceived barriers to physical activity and their associations with domain-specific physical activity and sedentary behaviour. *BMC public health*, 22(1):1051, 2022.
- [5] Niantic, Inc. Pokémon GO. <https://pokemongo.com>, 2016. (2025/11/17 確認).
- [6] Strava. Strava: Track Running, Cycling & Swimming, 2009. (2025/11/17 確認).
- [7] A. Waschke and J. Krüger. Automatic route planning for GPS art generation. *Computational Visual Media*, 5(3):303–310, 2019.
- [8] WHO. Nearly 1.8 billion adults at risk of disease from not doing enough physical activity. <https://www.who.int/news/item/26-06-2024-nearly-1.8-billion-adults-at-risk-of-disease-from-not-doing-enough-physical-activity>, 2024. (2025/11/17 確認).
- [9] WHO. Physical activity. <https://www.who.int/news-room/fact-sheets/detail/physical-activity>, 2024. (2025/11/17 確認).
- [10] テクテクライフ公式サイト. テクテクライフ公式サイト. <https://www.tekutekulife.com/>, 2020. (2025/11/17 確認).
- [11] 古田瑛啓, 奥川和希, 村山優作, 宮田章裕. 陣取りゲーム要素を導入した歩行データ収集システムの検証. *情報処理学会論文誌*, 64(1):33–42, 2023.
- [12] 尾本悦基, 塚田浩二. 複数の歩行軌跡に基づく GPS アートを用いた歩行促進アプリの提案. 日本ソフトウェア科学会 WISS 2024 予稿集, デモ発表 3-B12, 2024.

Purikura-Fab：3D プリンタ未経験者の造形プロセス理解と利用意欲を促進する体験型システムの開発

臼井 義人* 高橋 治輝* 松村 耕平*

概要. 3D プリンタが安価で扱いやすくなったことで、個人によるものづくりが広まりつつある。しかし、3D プリンタの使用には、依然として 3D データの入手やスライサパラメータの設定など様々な処理が必要であり、3D プリンタで何かを作りたいという動機がなければ利用には至らないだろう。そこで本研究は、3D プリンタ未経験者に動機と手軽な利用条件、簡潔なワークフローを提供することで、理解と利用意欲の向上を目指す。提案システムは、写真をもとにリソフェン形状の 3D モデルを作成し、造形に必要な一連の処理を自動化・簡略化する。体験者は、スマートフォン上の Web アプリから造形したい写真を送信した後に、3D プリンタ上のボタンで造形対象を選ぶだけで良い。さらに、画像という性質を活かして、プリクラのようにテキストやスタンプを重ねる画像編集要素を取り入れた。本稿では、提案システムが満たすべき要件とその実装について述べ、実際に造形された制作例を紹介する。

1 はじめに

2025 年現在、Blender¹や Tinkercad²といったモデリングツールの簡易化や、Bambu Lab³に代表される安価かつ高性能な 3D プリンタの登場により、パーソナルファブリケーション [3, 7] と呼ばれる個人によるものづくりが広まりつつある。教育現場 [2] では、学生が 3D プリンタを通じて技術的理解や創造性を高めるためのワークショップが開かれている [8]。日常生活 [5, 6, 10] においても、家庭での 3D プリンタの将来的な活用法が模索されている [9]。3D プリンタを活用できるユーザにとっては、さまざまな立体形状を高精度に作り、活用できるような技術が確立されていると言えるだろう。

しかし、未経験者にとっては「3D プリンタを使いたい」と思うきっかけが不足している。日常生活や趣味のための小物や備品などが 3D プリンタで作れることを知らなかったり、具体的な利用動機を持ちにくかったりするため、ものづくりのための手段とならないのが現状である。さらに、いざ利用しようとしても、3D モデルの作成や入手、3D プリンタ用の制御コード (Gcode) に変換するスライサソフトの操作、材料であるフィラメントの選定といった作業が必要となる。失敗しやすい箇所も多く、一度失敗するとモチベーションの維持が難しい [4]。

依然として 3D プリンタを使ったことがない者が多い状況を踏まえ、本研究では、「未経験者には気軽に魅力的な 3D プリント体験の機会を与えるべきで



図 1. Purikura-Fab の造形例

あり、その体験により基礎理解や利用意欲の向上につながる」という仮説を立てた。この考えに基づき、本稿では未経験者を対象とした体験型 3D プリントシステムを提案する。本システムでは、写真を 3D モデルへ変換し、造形に必要な処理を自動化・簡略化することで、未経験者でも容易に 3D プリントを体験できる環境を提供する。体験者は、システムから造形してみたい写真を送信した後に、自動で処理された造形対象を 3D プリンタ上のボタンで選ぶだけで良い。さらに、「プリクラ」のようにテキストやスタンプを重ねる画像編集要素を取り入れた。これは、従来のワークフローにおける 3D モデリングの代替として機能するだけでなく、自分だけの作品を制作している感覚を提供することにも寄与する。この体験を通じ、図 1 のような作品を作る中で、体験者に 3D プリンタに関する基本的な知識を習得させ、将来的な利用意欲の向上を目指す。

2 予備調査

様々な 3D プリント手法の中で、3D プリンタ未経験者にとって理解や利用意欲を高めるための手法

Copyright is held by the author(s).

* 立命館大学

¹ <https://www.blender.org>

² <https://www.tinkercad.com>

³ <https://bambulab.com/ja-jp>

として有効なものを明らかにし、その有効性を調査した結果についてまとめる。

2.1 3D プリント体験を促すための事例と手法

3D プリントに必要な 3D モデルの用意を省略したり、あるいは簡潔に 3D プリントを行う手法は数多く存在している。

例えば、Bambu Lab などが手掛ける 3D プリンタでは、事前にいくつかのサンプルデータが用意されているため、そのデータを選択するだけで造形することができる。また MakerWorld⁴では、web 上に掲載されているデータを選択するだけで、3D プリントすることができる。これらの手法は手軽に 3D プリントができるという観点においては優れているが、ユーザが直接選択したり制御することが少ないため、作成物に対して愛着を感じにくい傾向があると考えられる。

p5.fab[11] や G-coordinator⁵は、ユーザがコードを書くことで 3D プリンタを制御することができる。これらのツールは、プログラミング経験が豊富なユーザにとっては、自身のスキルを活用しやすく、3D プリンタに対する理解を深めるツールとしても、利用意欲を高めるツールとしても有効であると考えられる。一方で、プログラミング経験があまりないユーザにとっては、取っ掛かりにくく、3D プリントの理解を深め、利用意欲を高めるツールとしては有効だと考えにくい。

生成 AI を活用したツールとして、PrintMon Maker⁶ や Meshy AI⁷などが存在する。これらのツールを活用した体験は、昨今の生成 AI の普及により、多くのユーザにとって魅力的であると考えられるが、3D プリントを行う際に制御するパラメータが多く、手軽にできる体験とは考えにくい。

さらに、写真や画像から 3D プリンタのためのデータを作成する手法 [1] が存在する。この手法は、画像の輝度情報に応じて 3D モデルの厚みを変えることで、パターンやテキスト、エンボス加工などに応用している。本研究ではこの手法を活用したもので、画像の輝度をもとに厚みの異なる板に変換する「リソフェン」を採用する。リソフェンは、厚みの違いを利用して光の透過率を調整し画像のデザインを表現することができる技法である (図 1)。この技法は、基本的に単色の材料で造形物を作るため、材料の切り替えや写真に応じた調整が不要である。また、構造がシンプルなため造形の失敗も少なく、誰でも手軽に造形できると考えられる。さらに、背後から光を当てるとは入力された画像がはっきりとわかる



図 2. 学内で実施したリソフェンを使用した 3D プリント体験会の様子

ないため、出力された形状とそれを実際に手に取ったときの違いに驚きを感じることがあり、魅力的な体験につながる可能性がある。リソフェンを使用したツールとして、すでに Image to Lithophane⁸ や Make My Lithophane⁹などが存在する。しかし、これらのツールはいわゆる 3D モデリングツールの拡張版のような形で提供されており、未経験者が扱うには、パラメータの設定やスライス処理、データの受け渡しなどに課題がある。

2.2 リソフェンを使用した 3D プリント体験

自分の持っている写真でリソフェンを制作する体験は、未経験者にとって魅力的か。そして、利用可能なツールを用いて自分ひとりで制作することができるのか。こうした前提を確認するため、予備調査として 3D プリント体験会 (図 2) を実施した。

体験会は、立命館大学大阪いばらきキャンパス H 棟にある開けたスペースで、5 月 13 日の 12-15 時に実施された (なお、事前に大学内で張り紙などで宣伝を行った)。参加者は、3D プリンタと本体験会についての説明を受け、造形してみたいと思った写真をスタッフに提出した。その後、事前に用意した PC を用いて、提出した写真をスタッフのサポートのもと、Image to Lithophane を用いて 3D モデルに変換した。3D プリンタは、FlashForge の Adventurer 5M Pro¹⁰、スライサは同社の FlashPrint 5 を使用した。造形には 1 時間程度要し、制作物を受け渡し時に、イベントに参加した動機、3D プリンタの利用経験や難しいと感じたこと、ひとりでも利用できそうかなどについてインタビューを行った。

体験会には、10 名の参加者 (すべて同大学の学生) があり、この内 7 名は 3D プリンタ未経験者であったものの、全員が写真を用いた 3D プリントを体験することができた。体験会の参加の動機としては、「通りかかって面白そうだった」という回答

⁴ <https://makerworld.com/ja>

⁵ <https://github.com/tomohiron907/G-coordinator>

⁶ <https://share.bambulab.com/PrintMonMakerLaunch>

⁷ <https://www.meshy.ai>

⁸ <https://3dp.rocks/lithophane/>

⁹ <https://makerworld.com/ja/makerlab/makeMyLithophane?from=makerlab>

¹⁰ <https://flashforge.jp>

に加え、「自分の写真を使って作品がもらえるのが魅力的だった」といった回答が3件ほどあった。このため、写真を題材とすることは未経験者にとって魅力的である可能性が示された。また、体験会以前から3D プリンタに興味はあったが、実際に使う機会がなかったと述べる参加者もあった。

体験会が参加者に与えた影響として、ある体験者（一度だけ3D プリンタ経験あり）は、「（3D プリンタに対して）難しいというイメージはあるが、表現の幅がいろいろあって、もっと取り組んでみたいと思った」と回答した。その他の参加者にも、一度体験してみることで、「使うハードルが下がった」、「思ったより簡単にできた」、「写真がきれいに立体化されて驚いた」といった前向きな意見が得られた。さらに、「研究や趣味で使ってみたい」、「自作グッズを作りたい」といった利用意欲が示され、本体験が3D プリンタ利用の動機付けとなる可能性が示唆された。

一方で、体験会で難しかったこととして、10名全員が「スライサの設定」と「用語の意味」を回答しており、これらが未経験者にとって障壁となることが確認された。また、3D プリンタの操作を補助するスタッフがおらずとも造形までできると答えた参加者は1名のみであり、現行のツールを提供するだけでは初心者が自力でワークフローを完遂するのは困難であることも分かった。

以上より、写真を用いた3D プリントは、未経験者にとって「自分の写真を立体化する」というわかりやすく魅力的な体験となっており、結果として3D プリンタの理解や利用意欲を高める効果を持つと考えられる。一方で、提案システムでは、未経験者が躓いていた処理や設定を適切に支援する必要がある。

3 提案システム：Purikura-Fab

ここでは、本研究の提案システムの要件とそれらをどのように満たしたのかについて紹介する。

・手元にある写真で手軽に3D プリント体験

提案システムで第一に満たすべき要件は、誰もが手軽に3D プリントを体験できる仕組みである。3D モデリングや3D モデルを探す準備段階の手間を省き、造形時間が長くなりすぎない工夫も必要である。

提案システムは、写真からリソフェン形状を作るため3D データの準備が不要となる。体験者は、自身のスマートフォンにある写真をシステムにアップロードするだけで良い。また、作りたい写真が手元にはない場合は、その場で画像を検索したり新たに撮影しても良い。そして、造形完了時間を1時間前後に調整することで、品質と大きさを担保する。

・魅力的な3D プリント体験

手軽さだけでなく体験者にとって魅力的な体験であることが重要である。つまり、安易に体験を省略するだけでは、制作物のオリジナリティが損なわれる可能性がある。実際に、著者らの先行研究[12]では、

テトリスを応用した造形体験を実現したものの、造形物に自分で作った感がないという課題があった。

写真からリソフェンを制作する手法は、自分で選んだ写真が3D プリントされるため、それだけで体験者にとって魅力的な制作物になる可能性がある。提案システムでは、この特徴をさらに有効活用すべく、入力された写真にテキストやスタンプを追加する画像編集機能を実装する。この画像編集は単純であるものの、編集結果がリソフェンの形状にまで反映されるため、従来の3D モデリング作業を代替し得る。

前述した写真を用いた手軽な体験とこの画像編集機能は「プリクラ」を参考にしている。商業施設やゲームセンターで気軽に体験できるプリクラは、その場で写真をデコレーションするという手軽さでステッカーを作ることができる。これは提案手法の重要な特徴であるため、我々はシステムを「Purikura-Fab」と命名した。

・造形に至るまでの処理の自動化

予備調査の体験会で明らかになったように、未経験者は設定や何気ない用語の理解で躓く可能性がある。また、PCと3D プリンタとのデータのやりとりなど、3D プリントに直接は関係しない手間も生じる。こうした未経験者にとって障壁になり得る要素は、可能な限り取り除かれなければならない。

提案システムでは3D データの作成、スライス処理、3D プリンタへの転送を自動化することとした。体験者が行う必要のある作業は、①写真や画像のアップロード、②各処理が完了するまで待機、③3D プリンタに転送された造形データの選択のみである。

・パラメータ調整の余地

予備調査では、スタッフの手を借りながらも3D プリント体験を経ることで、「思ったより簡単にできる」という理解を示すことがわかった。つまり、すべての処理を自動化したり隠蔽したりするだけでは、体験によって得られる学びの機会をも損失してしまう可能性がある。

造形プロセスの理解と利用意欲の促進を目指すため、提案システムでは積層ピッチと造形速度のパラメータを体験者が調整可能な形で残すこととした。これらのパラメータは、スライサの基本的な設定項目であり、仕上がりの品質やプリント時間に影響を与える。体験者自身が調整する機会があれば、これらの造形結果への影響を通じて速度と品質のトレードオフを学べることが期待される。また、この調整が体験者にとって体験の障壁とならないように、なにも変更しなくても造形できるように設計する。あくまでも、体験者自身が体験の過程で興味があれば変更できるように設計を行う。

・誰もが利用できる運用

予備調査の参加動機で、「通りかかって面白そうだった」という意見があったように、3D プリンタは手

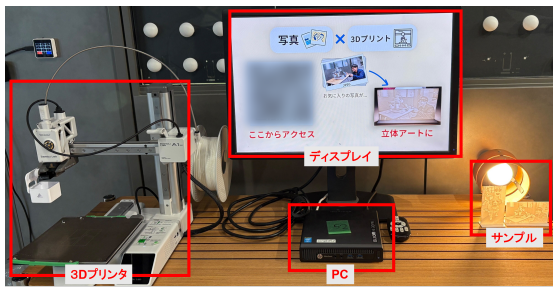


図 3. 提案システムの設置図

軽に体験できる場所にあり、そのシステムがどのような体験をもたらすのかひと目で分かるようになっていなければならない。

そこで、提案システムは誰もがアクセスできるような場所に 3D プリンタを配置する想定で実装する。システム説明や 3D プリンタの状態を示すためのディスプレイを併設することで、誰でも体験できるという情報も提供できるようにする。

4 実装

4.1 システム構成

提案システムは、主に 2 つの要素から構成されている。1 つ目の要素は、Gcode を作成する Web アプリケーションである。この Web アプリケーションは、ユーザが画像編集を施した写真と選択した造形パラメータ、事前に用意したプリセットをもとに Gcode の作成を行う。生成した Gcode はクラウドストレージ (Firebase) に送信されて管理される。

2 つ目の要素は、3D プリンタと通信を行うサーバである。このサーバは、クラウドストレージに保存されている Gcode をダウンロードし、3D プリンタへデータを送信する役割と、3D プリンタの状態や造形完了時間をユーザにディスプレイから提示する役割を備えている。

提案システムの設置図を、図 3 に示す。3D プリンタは、価格の安さと造形速度の速さを考慮して Bambu Lab A1 mini¹¹を採用した。フィラメントは、Bambu Lab PLA Basic(白)を採用した。そして、造形物のサンプルを展示するため、3D プリンタとディスプレイの近くにライトを設置している。

4.2 Web アプリケーション

システムの画面および 3D モデルへの変換処理は、Image to Lithophane¹²を参考に実装した。追加で実装したインターフェースやボタンの入力処理には、HTML や CSS, JavaScript を用いて実装した。画像編集の機能は、キャンバス API を用いており、絵

文字スタンプとテキストの配置と大きさの変更が可能である。また、ユーザはスライサパラメータをラジオボタン形式で選択できる。積層ピッチは、0.2 mm から 0.36 mm まで 0.04 mm 間隔で選択できるように設計しており、造形速度は 230 mm/s から 270 mm/s まで 10 mm/s 間隔で選択できるように設計した。またそれぞれのパラメータの役割がユーザにわかるようにイラストと説明を含めた提示も行う。編集した画像をもとに 3D モデルを生成後、ユーザが選択したパラメータの数値、事前に用意したプロファイルをもとに、スライサエンジンである CuraEngine を使用して Gcode を作成する。生成された Gcode を Firebase へアップロードする。

4.3 3D プリンタと通信を行うサーバ

開発言語として、Python を用いて実装した。ユーザに、3D プリンタの造形状況やファイルの転送状況をディスプレイに提示する画面は Pygame で実装した。Bambu Lab A1 mini の稼働状況や造形ファイルの情報の取得は MQTT, Gcode の送信は FTP で実装した。FTP 転送やプリンタ監視、Firebase 監視などの並行処理を行うために Threading モジュールを用いて並列処理を行っている。システムは、常に Firebase の監視を行っており、造形データが Web アプリケーションからアップロードされるたびに随時ダウンロードする。ダウンロードが完了したのちに、3D プリンタへの転送を行う。なお、システム運用のため、3D プリンタの状況を Google スプレッドシートへ記録するようにした。

4.4 制作するリソフェン

制作するリソフェンの外形を図 4 に示す。入力された画像はまずグレースケール画像に変換され、輝度値に応じて各部分の厚みが決定される。具体的には、画像の最も明るい部分が最も薄く、最も暗い部分が最も厚くなるように厚みが線形的に割り当てられる。このとき、輝度の 1 階調 (0 から 255 までの 256 段階) ごとに高さを少しずつ変化させる仕組みを用いている。輝度の値は、最小厚さ (0.8 mm) と最大の厚さ (3 mm) を基準に、1 階調ずつ厚さが増えるように換算される。最小の厚さと最大の厚さは著者が試作する中で、造形の安定性と品質の兼ね合いから導いた。また、画像を 3D モデルに変換する際には、1mm あたり数ピクセルの密度で高さデータが割り当てられ、これが「ピクセル比率」として造形の解像度を決める。

生成されるリソフェンは高さ方向に積層され、全体を支える台座部分が約 30 mm 突き出している。入力画像の縦横比を保持したまま自動で拡大・縮小され、長辺が最大 100 mm となるように調整される。造形時間はおおよそ 1~1.5 時間である。

¹¹ <https://bambulab.com/ja-jp/a1-mini>

¹² <https://github.com/MarkDurbin104/3dp.rocks/>

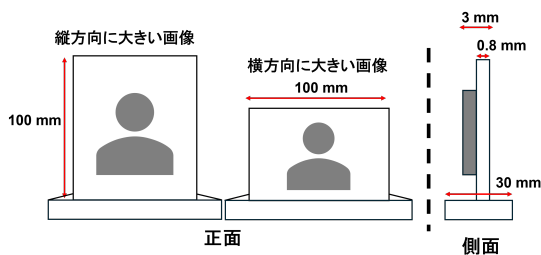


図 4. 制作するリソフェンの図

5 システム利用の流れ

体験者は自身のスマートフォンから、ディスプレイに表示されている QR コードを読み取ることで提案システムにアクセスする(図 3)。その後、出力したい写真を選択する。写真を確認し、写真に画像編集を施すか、そのまま写真を確定させて造形パラメータの選択を行うかを選択する(図 5 左)。

「写真を編集する」を押す(図 5 左)ことで、画像編集画面(図 5 中央)に遷移し、ユーザはスタンプやテキストの入力を行う。最終的な画像を確認した後(図 5 右)、次の画面へ移行する(図 6 ①)。



図 5. 画像編集を行う画面

次に 3D モデルの確認と、スライス処理に必要なパラメータ設定を行う。プレビューに生成された 3D モデルが表示される(図 6 ②)。この時、生成される 3D モデルに対して高さや厚みを変更できる。変更がなければ、画面をスクロールし、スライス処理のパラメータ設定を行う。

スライス処理でユーザが変更できるパラメータは、積層ピッチと造形速度である。ユーザは説明文を確認(図 6 ③)し、自由にパラメータを選択する。パラメータ選択後、アップロードを行う(図 6 ④)。アップロードが完了するまで、ローディング画面が表示される(図 6 ⑤)ので、ユーザは画面を閉じずに待機する。アップロード完了後、ディスプレイ上には、3D プリンタへ転送中のファイルのファイル名と転送の進捗が表示される(図 7 ②)。

Gcode を 3D プリンタへと送信完了すると、ディスプレイにはファイル選択画面(図 7 ③)が表示さ



図 6. スライス処理を行う画面

れるので、ユーザが 3D プリンタのディスプレイを操作し、造形ファイルを選択する。

造形が開始されるとディスプレイ上に造形の終了予定時刻が表示される(図 7 ④)。ユーザは表示された時刻を目安にして、造形物を回収するために再訪すれば良い。3D プリンタが造形中は常にこの画面が表示されている。以上の手順を踏むことで図 1 に示すようなリソフェン作品が完成する。

6 システム運用

提案システムを立命館大学大阪いばらきキャンパス H 棟 5F の廊下に設置し(図 3)、実際の利用状況やシステムの動作を確認した。8 月 7 日からシステムを設置し、8 月 31 日現在において、その間に制作された作品は計 12 件であった。

システムによって造形された例を図 8 に示す。制作されたものの多くはペットやお気に入りの写真を題材としていた。なお、プライバシーの観点から、体験者から同意を得られていない制作物を論文に載せることは避けた。

システムは体験者によって自由に使用されており、著者らによるサポートは必要なかった。運用期間中に制作された作品の大半は問題なく造形されたが、いくつか失敗事例も確認された。たとえば、ベットが正しく設置されていなかったため、造形途中に 3D プリンタのベッドから造形物が剥離するケースがあった。さらに、高輝度領域が連続する写真を入力としたリソフェンでも造形失敗が見られた。図 9a はその一例であり、元の画像を確認すると、画像の上部が白く、薄い層が続くため失敗したと考えられる。このため、青空や白背景が広い画像では薄い層が続く

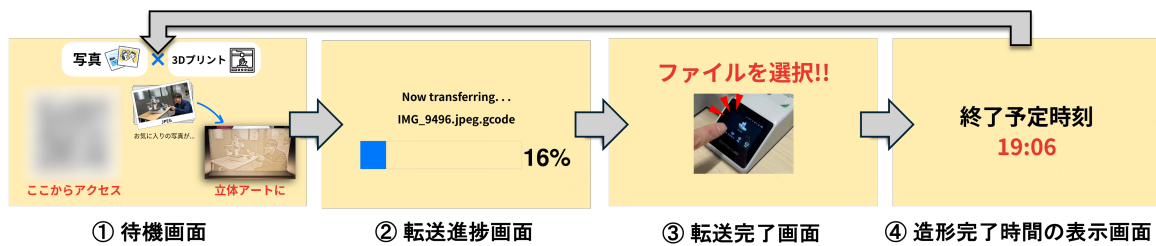


図 7. ディスプレイの表示画面



図 8. 造形例

ため、最後まで造形できずに失敗することもある。また、画像編集で追加されたテキストやスタンプが綺麗に造形されないケースも一部見られた。図 9b に示すように、文字と背景写真の輝度値が近いと、リソフェン変換後に十分なコントラストが得られず、文字輪郭が判別しにくくなるがあった。その結果、完成物では文字が背景に埋もれ、編集した内容がわかりづらいケース (図 9 c) が確認された。

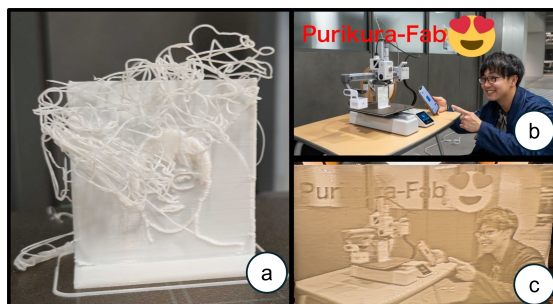


図 9. 失敗例。(a) リソフェンに薄い部分が多いと積層に失敗することがある。(b, c) 画像編集で追加された文字が写真に埋もれてしまう場合がある。

利用者からの意見としては、ローディング画面 (図 6⑤) の表示時間が長く、システムのエラーと誤解される点や、選択可能なパラメータの基準が不明瞭である点が指摘された。また、造形完了予定時刻が造形開始後にしか提示されない点についても、不満の声が寄せられた。これらの問題に対して、以下の対応策を考えている。まず、スライス処理中のロー

ディング画面については、単に静止画を表示を提示するだけでなく、処理進捗やアニメーションを加えることで映像に変化を持たせ、利用者が処理の継続を直感的に理解できるようにする。次に、パラメータ選択に関しては、積層ピッチや造形速度を変更した場合の仕上がりや造形時間の差異をサンプル画像や比較図として提示し、利用者が結果を想像しやすい形でガイドを与える。さらに、アップロード処理の段階で、想定される造形時間をあらかじめ提示することで、利用者が出力までの待機時間を把握できるようにし、不満軽減につなげることを目指す。

7 今後の展望

運用中の意見をもとにシステムを修正していき、より多くの人が利用できる場所での評価実験を通じて、提案システムが、3D プリンタ未経験者に対してどのような影響を及ぼすかを調査していく。

特に、本研究ではユーザが提案システムを体験する中で、どのような過程を経て 3D プリンタに対する理解や利用意欲を深めていくのかに注目する。現在の提案システムでは、造形完了時間表示画面 (図 7④) など、映像に変化が見られない箇所が存在する。そのような箇所にフィラメントや Gcode、モデリングなどの情報を提示することで、システム体験者に対し、より理解や興味を促すことができるかもしれない。リソフェン作品の土台に、選択したパラメータの値や造形時間を印字することも有効だと考える。自分の制作物を通じて、どのパラメータを変更させるとどのような結果がでるかを理解することで、3D プリント体験そのものに対する理解が深まるかもしれない。

8 おわりに

本研究では、3D プリンタ未経験者を想定して、3D プリントに対する基礎理解や利用意欲の向上を目指した。ユーザが自身のスマートフォンの写真をリソフェン形状の 3D モデルに変換して造形できるシステムを開発した。今後は評価実験を通じて、システムが 3D プリンタ未経験者に与える影響を調査する。

参考文献

- [1] M. Alexa and W. Matusik. Reliefs as images. In *ACM SIGGRAPH 2010 Papers*, SIGGRAPH '10, New York, NY, USA, 2010. Association for Computing Machinery.
- [2] J. Günther, L. Brehm, H. Günzel, and A. Humpe. Teaching 3D Printing Technology Hands-on. In *2020 IEEE Global Engineering Education Conference (EDUCON)*, pp. 953–957, 2020.
- [3] N. Gershenfeld, 田中 浩也, 糸川 洋. Fab ーパーソナルコンピュータからパーソナルファブリケーションへ. オライリー・ジャパン, 2012.
- [4] N. Hudson, C. Alcock, and P. K. Chilana. Understanding Newcomers to 3D Printing: Motivations, Workflows, and Barriers of Casual Makers. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, p. 384–396, New York, NY, USA, 2016. Association for Computing Machinery.
- [5] J. Kim, A. Guo, T. Yeh, S. E. Hudson, and J. Mankoff. Understanding Uncertainty in Measurement and Accommodating its Impact in 3D Modeling and Printing. In *Proceedings of the 2017 Conference on Designing Interactive Systems*, DIS '17, p. 1067–1078, New York, NY, USA, 2017. Association for Computing Machinery.
- [6] B. Lee, G. Han, J. Park, and D. Saakes. Consumer to Creator: How Households Buy Furniture to Inform Design and Fabrication Interfaces. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, p. 484–496, New York, NY, USA, 2017. Association for Computing Machinery.
- [7] C. Mota. The rise of personal fabrication. In *Proceedings of the 8th ACM Conference on Creativity and Cognition*, C&C '11, p. 279–288, New York, NY, USA, 2011. Association for Computing Machinery.
- [8] B. Norouzi, M. Kinnula, and N. Iivari. Making Sense of 3D Modelling and 3D Printing Activities of Young People: A Nexus Analytic Inquiry. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA, 2021. Association for Computing Machinery.
- [9] R. Shewbridge, A. Hurst, and S. K. Kane. Everyday making: identifying future uses for 3D printing in the home. In *Proceedings of the 2014 Conference on Designing Interactive Systems*, DIS '14, p. 815–824, New York, NY, USA, 2014. Association for Computing Machinery.
- [10] E. Stemasov, A. Botner, E. Rukzio, and J. Gugenheimer. Ephemeral Fabrication: Exploring a Ubiquitous Fabrication Scenario of Low-Effort, In-Situ Creation of Short-Lived Physical Artifacts. In *Proceedings of the Sixteenth International Conference on Tangible, Embedded, and Embodied Interaction*, TEI '22, New York, NY, USA, 2022. Association for Computing Machinery.
- [11] B. Subbaraman and N. Peek. p5.fab: Direct Control of Digital Fabrication Machines from a Creative Coding Environment. In *Proceedings of the 2022 ACM Designing Interactive Systems Conference*, DIS '22, p. 1148–1161, New York, NY, USA, 2022. Association for Computing Machinery.
- [12] 高橋 治輝, 白井 義人, 松村 耕平. ボタンひとつで3Dプリントが体験できるシステムの提案と運用. 日本ソフトウェア科学会 WISS2024, 2024.

ユーザの能動的な探索を可能にする Deep Research システム

関口 祐豊^{*†} 樋口 啓太^{*}

概要. 大規模言語モデルを活用した AI エージェントアプリケーションである Deep Research は、ユーザに代わって多段階的なウェブ検索を実行し、複数セクションから構成されるレポートを生成するシステムである。従来のシステムでは、ユーザは初期クエリの入力とエージェントが生成したリサーチプランの承認またはコメントのみが可能であり、リサーチ開始後はユーザが結果に至るまで介入できないため、しばしばユーザの関心と乖離したレポートが生成されるという課題があった。本論文では、ユーザが能動的に介入可能な Deep Research システムを提案する。提案システムでは、まずユーザのクエリに基づいて複数のリサーチ候補を提示し、ユーザはそこから関心のある候補を複数選択できる。さらに、リサーチ結果に基づいて関連するリサーチ候補の追加や、動的な要約、構造化されたレポート生成機能を提供することで、ユーザは理解を深めながらさらなる探索を進めることが可能となる。ユーザ実験の結果から、提案システムは従来の Deep Research と比較して、網羅性の高いレポートを効率的に低負担で生成できることが示唆された。

1 はじめに

大規模言語モデルの登場により、ChatGPT などの AI 対話システムが急速に普及している。初期の AI 対話システムでは、モデルの挙動を定義するシステムプロンプトと、ユーザの入力クエリ（ユーザプロンプト）を入力することで、言語モデルがユーザの要望に応じた出力を生成している。近年では、ReAct [19] を代表とする、大規模言語モデルによる推論とツール実行を組み合わせた、エージェント構築の枠組みが登場したことで、ユーザの要望を答えるために様々な機能を対話システムから呼び出し可能にしている。具体的には、ユーザのクエリに対して必要なツールを推定し、情報検索に関するクエリではウェブ検索を呼び出したり、数理的な質問には動的なプログラムの生成と実行を組み合わせている。

代表的な AI エージェントアプリケーションに「Deep Research」と呼ばれる機能がある。これは、ユーザのクエリに基づいてリサーチ計画を立案し、多段階的なウェブ検索を実行することで、複数のセクションから構成されるレポートを生成する仕組みである。Deep Research は 2025 年 2 月に ChatGPT においてリリースされ、2025 年 8 月現在では、Google Gemini や Microsoft Copilot など主要 AI 対話サービスにて同様の機能が利用可能となっている。Deep Research システムでは、ユーザが最初にリサーチトピックを入力すると、トピックに基づいてエージェントが複数の段落から構成されるリサーチ計画を生成する。ユーザは、生成された計画に対して承認や修正のためのコメントをすることができ

る。一度、ユーザが承認をすると、それぞれの段落のリサーチを行うエージェントを呼び出し、すべてのリサーチが完了すると序章やまとめを記述して、レポートとして出力する。Deep Research の利用において、ユーザは満足するリサーチ計画が生成されるまで介入することができるが、リサーチ中は一切の介入ができないため、出力されたレポートがユーザの満足するものとなる保証はない。そのため、しばしばユーザは出力された結果に基づいてクエリを修正し、改めて Deep Research の実行を承認し、満足する結果が得られるまで長時間待機が必要なセッションを何度も繰り返すことが必要となる。

本論文では、Deep Research に能動的探索の枠組みを導入することで、ユーザが満足するまで 1 つの対話的なセッションを続けることを可能とし、効率的なリサーチを実現するシステムを提案する（図 1）。提案システムはユーザがリサーチクエリ（例えば大規模言語モデルに関する最新研究についてなど）を入力すると、クエリに基づいてセクションの候補を複数個生成しタイル状の UI に表示する。ユーザが興味に基づいてリサーチタイトルが表示されたタイルを選択すると、リサーチエージェントが起動し選択されたセクションの検索を開始する。ユーザは複数のタイルを同時に選択することができ、リサーチは同時並行で実行される。リサーチが完了するとシステムはユーザの能動的探索を支援するために、1) リサーチが完了したセクションのサマリ生成と、2) ユーザが選択したトピックに関連するセクションを新たに生成してタイルに表示を実行する。これを繰り返すことで、ユーザは自らの関心に基づいた能動的探索を満足するまで続けることができる。さらに、リサーチ結果表示画面では、結果の編集や再調査などを行うことができる。本研究では提案システムを

Copyright is held by the author(s).

^{*} SB Intuitions

[†] 明治大学

All authors contributed equally to this work



図 1. 提案システム画面. リサーチ画面 (a) では、ユーザがクエリを入力すると 10 個のタイルが生成される (a-1). タイルをクリックすると調査が開始され、完了後にはセクションのサマリ (a-2) と関連するトピックの新規タイル (a-3) が生成される. また、画面下部のチャット欄から新規トピックのタイルが生成できる (a-4). リサーチ結果表示画面 (b) では、調査結果の閲覧やセクションの再調査 (b-1)、不要なセクションの非表示 (b-2) を行うことができる.

評価するために、既存の Deep Research と比較するユーザ実験を実施した. 実験の結果、ユーザは提案システムにより、与えられたトピックに関するレポート作成をより効率的に行えることを明らかにした.

2 関連研究

2.1 エージェント AI システムと Deep Research

エージェント AI とは、言語モデルが考えること (推論) と外部の道具を使うこと (検索やコード実行など) を行き来しながら課題を解く仕組みである. 代表例の ReAct は、この往復を素直に文章として出力しつつ必要なときに行動を挿入する [19]. MRKL は外部知識や計算手段をモジュールとしてつなぐ設計を示し [8], Toolformer は「どの道具をいつ使うか」を自己学習させる [16]. さらに、AutoGen は複数エージェントが会話で協力する形を整え [18], HuggingGPT は言語モデルをコントローラとして外部モデル群を呼び分ける [17]. Web 上の調査という文脈では、ブラウザ操作と出典提示を前提にした WebGPT が長文の調査応答の基盤をつくった [12].

Deep Research は、ユーザの問いから計画を立て、段階的なウェブ探索と要約を経てレポートを出力するタイプのアプリケーションの総称である [13, 5, 11]. 従来は実行前の計画確認はできても、実行中の探索に介入しにくく、途中の気づきをすぐ反映しづらいという課題があった. 本研究はこの枠組みに、ユーザが途中で選択や方向転換を行える能動的探索を組み込み、調査の流れをユーザが握り続けられる Deep Research として位置付ける.

2.2 能動的探索 UI

能動的探索とは、ユーザが情報を少しずつ集めながら理解を深め、必要に応じて方針を更新していく進

め方である. Bates のベリーピッキングモデルは、この段階的な集め方を端的に説明し [2], Marchionini は検索を「発見・学習・理解」の過程として位置付けた [9]. また、Hearst は検索結果を意味のある切り口で見せることで、曖昧な意図に足場を与える UI を示した [7]. 本研究は、これらの考え方を踏まえつつ、単なる検索ボックス中心の UI ではなく、エージェントが調査を担い、ユーザが興味に沿って選択・深掘り・方針転換を繰り返せる UI を目指す.

本研究の UI では、最初にクエリから複数のセクション候補をタイルで提示し、ユーザが選ぶたびにエージェントが並列に調査を進め、短い要約を返す. 要約に基づき関連候補を自動追加し、ユーザは結果を見ながら即時に編集や再調査を指示できる. この循環は、会話型情報検索における明確化質問や文脈維持の考え方 [15, 20, 1, 3] を踏襲している. 本研究ではユーザの能動的探索行動が、ツール実行を伴うエージェントの行動計画の決定につながる、新しい Human-AI Agent Interaction の方式を提案する.

3 能動的な探索が可能な Deep Research システム

3.1 従来の Deep Research のワークフロー

本研究では、基盤となる Deep Research 手法として、LangChain のオープンソースソフトウェアである Open Deep Research を使用する. このソフトウェアは、一般的な Deep Research のアルゴリズムとユーザインタラクションを持ち、ソースコードが公開されているため、基盤として採用した. OpenAI や Google などが提供する既存の Deep Research システムは、内部アルゴリズムやエージェント構成が非公開であり、同一条件での制御やログの取得、パラメー

ユーザの能動的な探索を可能にする Deep Research システム

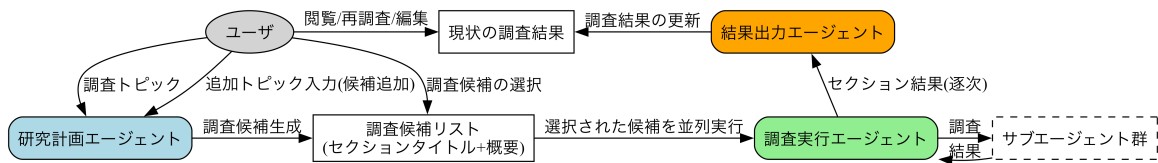


図 2. 提案手法のワークフロー

タの調整が困難である。そのため、本研究では、再現性および制御可能性を重視し、エージェント構成の統一や実行過程の記録が可能な Open Deep Research をベースラインとして採用した。Deep Research は、ユーザの検索クエリに対して研究計画エージェントと調査実行エージェント、結果出力エージェントが順次的に動作するワークフローをしている。研究計画エージェントは、ユーザからのクエリを受け取ると複数のセクションから構成される研究プランを生成する。システムはユーザにプランを提示し、承認するかコメントによる修正をするかを問う。ユーザは満足するプランが生成されるまで修正を繰り返すことができる。ユーザの承認が得られると、調査実行エージェントにプランが渡され、それぞれのセクションについての調査をサブエージェントに実行させる。この時のサブエージェントのアルゴリズムや調査対象の情報ソースは各サービスやその設定に依存する。すべてのサブエージェントの実行が完了すると、調査結果が結果出力エージェントに渡され、序論やまとめ文章の生成や文章整形が実行され、ユーザに結果が表示される。このプロセスの中でユーザはクエリの入力と調査計画の修正指示までしか介入できず、その後は結果が出力されるまで待機が必要となる。本研究では、このワークフロー中のコンポーネントを提案システム開発に活用するとともに、従来手法としてユーザ実験のベースラインにも採用する。

3.2 提案手法のワークフロー

提案ワークフローを図 2 に示す。提案手法は研究計画と調査実行、結果出力それぞれのエージェントをシステムの構成要素として活用しながら、能動的探索を実現している。ユーザが最初に調査トピックを入力すると、研究計画エージェントが複数個の調査候補をユーザに提示する。ユーザは調査候補セクションのタイトルと概要から、調査を開始する調査候補を決定する。調査は並列で実行されるため、ユーザは複数の候補を選択することができる。システムは、ユーザが調査候補を選択すると、調査実行エージェントを呼び出して調査を開始して、結果を新たに生成する。また、調査を実施したい候補がない場合は、ユーザは追加でトピックを入力することで、さらに調査候補が追加される。上記を繰り返すことで、ユーザは自らの興味に基づいた調査を進めることができる。システムはこの際、調査が完了し

たセクションを元に内容の要約などを随時更新する。ユーザはいつでも現状の調査結果を確認でき、必要に応じて編集や再調査を行うことができる。

3.3 プロトタイプ UI

本研究では、前節で述べた能動的 Deep Research ワークフローを基盤として、タイル型リサーチシステムを構築する。システムの UI は探索の起点となるタイル型リサーチ画面 (図 1a)、調査が完了したセクションを編集可能なりサーチ結果表示画面 (図 1b) の 2 つの画面から構成される。

タイル型リサーチ画面 タイル型リサーチ画面は、調査候補の提示と選択を行うインターフェースである。能動的 Deep Research ワークフローに基づき生成された調査候補がグリッド状のタイルとして表示され、ユーザがクリックすることで調査が開始する。調査中のタイルにはプログレスバーが表示され、進捗状況をリアルタイムで把握できる。調査が完了すると、その成果に基づき関連トピックが新たに 10 件追加表示される。また、各タイルは階層的に構成されており、上位トピックと下位トピックの関係を視覚的に把握できるようになっている。下位トピックのタイルを選択して調査を行った場合、レポート内では上位トピックのセクション内の節として自動的に整理される。もしユーザが求める調査候補タイルがない場合、図 1(a-4) のチャット欄から追加の候補生成を要求でき、入力内容に基づいて新たな調査候補タイルを生成できる。

リサーチ結果表示画面 リサーチ結果表示画面は、調査結果の出力と編集を行うインターフェースである。調査結果はレポート形式で整理され、ユーザはそれを閲覧、編集、ダウンロードできる。また、各セクションに対して再調査を指示でき、必要な箇所のみ差分的に更新できる。図 1(b-2) では、各セクションに付与されたチェックボックスを外すことで不要なセクションを非表示・削除できる。さらに、セクションはドラッグ&ドロップによって順序を変更でき、ユーザはレポート構成を柔軟に調整可能である。

3.4 実装

本システムは、フロントエンドに React、バックエンドに FastAPI を用いて実装している。LLM のモ

デルは OpenAI 社の gpt-4.1-nano, gpt-4o-search-preview を主に使用し、前者は応答生成や要約などの軽処理に、後者は研究計画生成や Web リサーチなどの高負荷処理に用いている。リサーチ処理には、LangChain が提供するオープンソースである Open Deep Research を活用し、セクション候補の生成やレポート作成段階で利用している。さらに WebSocket によりリサーチの進捗状況をリアルタイムでフロントエンドへ送信する。

これらの実装のもとで動作時間を計測した。Deep Research システムにおいて、調査開始から研究計画プラン生成までの平均所要時間は約 20.08 秒であり、研究計画プラン承認からレポート完成までの平均所要時間は約 48.02 秒である。提案システムでは、1 トピックを対象とした場合のタイトル選択から章生成までの平均約 32.66 秒である。また、提案システムに初期表示される 10 個のタイトルのうち 5 つと、調査完了時に生成される関連トピック 10 件のうち 2 件は LLM から直接即時生成され、ユーザの待ち時間を 10 秒以内に抑えている。残りの 5 つのタイトルや 8 件の関連トピックは研究計画エージェントによる高品質な候補として提示される（初期表示タイトル生成完了までの平均所要時間：約 39.01 秒、関連トピック生成完了までの平均所要時間：約 60.40 秒）。

4 ユーザ実験

提案システムがユーザと AI エージェントが協調する調査タスクにおいて有効であるかを明らかにするために、ユーザ実験を実施した。本実験では提案システムと従来の Deep Research システムの比較を行った。実験参加者としては所属企業のデータアナレータ 32 名（平均年齢 41.3 歳（SD: 8.5）、内 19 名女性）が実施した。31 名が対話 AI についての経験を持っており、また 21 名が一度は Deep Research 機能を使用したことがある。

4.1 実験設定

本実験ではユーザが Deep Research システム（Baseline）と、提案システム（proposed）を比較した。実験参加者は、事前に割り当てられた調査トピックをそれぞれのシステムにて調査し、その結果提出と体験アンケートの記入を行った。提出された結果に対して、LLM を使用した調査結果の評価を実施した。

ベースラインシステムの実装 ベースラインとなる Deep Research システムの実装は、Open Deep Research に Web UI を開発することで実現した。ユーザの体験は 3.1 節と同様の流れである。生成された調査報告は、不必要な章を除外してマークダウン形式でダウンロードすることができる。公平性のために提案システムが持つ、章ごとの再調査機能をベー

スラインでも使用可能としている。これは提案システムの操作性を再現し、両システム間で操作機会とエージェント条件を統一するために導入している。また、研究計画エージェントと調査実行エージェント、結果出力エージェントは同じロジックを利用しているため、生成 AI モデルやエージェントの条件を揃えた比較実験を行っている。

調査トピック 複数の分野の調査において有用であるかを調査するために、3 種類のカテゴリを用意した。各カテゴリには 2 つのリサーチトピックがある。各実験参加者には 1 つのカテゴリが均等に割り当てられ、2 つのシステムでそれぞれ 1 つのトピックを調査した。ただし、2 つのシステムの操作順はカウンターバランスが取れるようアサインしている。

A 環境問題

1. ヒートアイランド現象の原因と影響
2. 気候変動が農作物生産量に与える影響

B 教育

1. 学校でのアクティブラーニング導入効果
2. 教員研修制度と授業改善の関係性

C 健康

1. 運動習慣とメンタルヘルスの関係
2. ワークライフバランスの改善事例

実験の流れと調査タスク 実験参加者はまず本実験についての説明を受け参加に同意した。次に、実験参加者は割り当てられた調査カテゴリとシステム順序を確認した。ユーザはベースラインと提案システムのそれぞれで練習タスクを行った。練習タスクは「日本の祝日について」というテーマを 10-15 分程度で調査するというものである。その後、ユーザはデモグラフィック情報をフォームに入力し実験タスクを開始する。実験参加者は割り当てられたシステム順序と調査トピックに従って実験を進める。最初に実験説明書を確認してから調査タスクを開始する。実験タスクでは、操作するシステムを使用してトピックに関する調査を進め、最終的に 12 章のセクションからなる調査を提出する。両方のシステムにおいて、提出にあたり再調査機能を使用できるが、各セクションを直接編集することはできないとした。12 章のレポートを提出後、体験アンケートを記載した。

調査タスク実施にあたり、実験参加者は以下のような方針の説明を受けた。

この実験では、システムを使って指定されたトピックについて調査を行い、包括的かつ重要な事実を抑えたレポートを作成していただきます。システムが生成した 12 章を提出していただきます。直接調査内容を編集する必要はありません。時間はあまりかけすぎず実施してください。限られた時間の中で、最も有益な情報を整理しまとめてください。

Deep Research システムにおいては、ユーザは複数回の修正依頼や調査を繰り返すことで 12 章からなる調査レポートを提出する。Deep Research システムは最初の調査トピックを入力後、調査計画エージェントが 4-6 章からなる調査計画を生成する。ユーザが 12 章の報告を生成するために、2 通りの方法を選択できる。1 つ目の方法は、生成された調査計画に対して「〇〇という章を生成してください」と要求することで、調査する章を追加し 12 章からなる調査計画を生成し、調査を承認する方法である。この方法では、1 つの調査レポートをダウンロードし提出する。2 つ目の方法は、複数回の調査を実施することで、合計 12 章以上の章を生成する方法である。この両方の方法において、修正要求や 2 回目以降の調査トピックの内容は調査計画や結果を基に、実験参加者が考え入力している。12 章以上が生成された場合、実験参加者は必要な章のみを選択しダウンロードした。2 つ目の方法では複数の調査レポートをダウンロードし、合計 12 章に調整して提出した。また、調査レポートにはイントロダクションとまとめ章が生成されるが、それらのセクションはすべて除外してダウンロードするように指導した。

提案システムにおいては、一度のトピック入力により 12 章の調査レポートを生成し提出する。ユーザは好みの章候補が生成されない場合は、トピック生成要求をすることができる。また、深掘り候補として生成され、調査した節も 1 つの章としてカウントする。12 以上の調査を行った場合には、実験参加者が調査結果を吟味して 12 章を選択した。提案システムにおいては、各実験参加者は、1 つのマークダウンファイルのみを提出した。

評価指標 評価はアンケート評価及び LLM による調査結果の判定を実施した。アンケートでは 5 つの有用性と成果物の自己評価に関する質問を行った。アンケートは全て 7 段階のリッカートスケールで実施した。具体的な質問項目は「Q1: このシステムは、あなたの調査作業にとって有用だった」「Q2: 調査結果は十分に網羅的である」「Q3: 調査結果は十分に深掘りされている」「Q4: 今回作成した成果物の品質は十分高い」「Q5: 成果物の信頼性（根拠の確かさ）は十分に高い」である。実験参加者は、ユーザ体験測定を目的として UEQ-S のアンケート 8 項目を回答した。UEQ-S ではアンケート結果から実用性指標と、タスクに直結しない“快さ”の質であるヘドニック指標を計算して分析した。実験参加者はメンタルワークロード測定のために NASA-TLX(重み付けなし)も回答した。UEQ-S は実用性指標とヘドニック指標の両方を観測した。また、調査開始時刻と調査完了時刻も報告した。

調査結果の評価として大規模言語モデルを使用する LLM as a judge [6] を採用した。評価指標として

は、誤り率と調査結果の網羅性、そして全体的なレポートの質である。誤り率は提出された 12 章の中に何%のハルシネーションが存在するかを、langchain ライブラリの事実性確認機能により、確認を行った。この際、誤っていると判定された章はレポートから削除して、以後の網羅性及び質の評価を実施した。

網羅性は、RAG (Retrieval-Augmented Generation) の評価指標として TREC 2024 RAG Track で採用されている自動 nugget 評価法を使用した [14]。この方法は記載された複数の文書から重要な事実を抽出し、1 つの文書にどの程度抽出された事実が含まれているかを評価するものである。提出したレポートに適用するために以下のようなプロセスを行った。まず、ある 1 つの調査トピックについて、Deep Research と提案手法両方の全ての調査レポート (今回の設定の場合 1 つのトピックに関して提出される調査レポートはベースラインと提案システムそれぞれから 5,6 本である) を自動 nugget 抽出に適用する。その際に抽出された nuggets が 1 つの調査レポートに含まれている割合を評価する。この際には見逃してはいけない重要な事実の網羅率及び、重要だが必須ではない事実も含めた網羅率の両方を計算する。最後に、レポートの質を SummEval[4] 及び USR[10] にて提案された評価プロンプトを使用して 9 項目各 5 点の 45 点満点で評価した。

4.2 実験結果

被験者内要因で 2 つの手法 (Deep Research, 提案システム) を比較した。すべての指標に対して Wilcoxon 符号付順位検定 (両側) を適用し、Benjamini-Hochberg 法による p 値補正を行った。効果量は rank-biserial 相関 r_{rb} を用いた。タスク完了時間に関しては、報告が完全でないものや、システムエラーや API 呼び出しの制約により実験が中断されたものに関しては除外し、被験者内で両方が観測できた 25 ペアでの比較となった。

表 1 に全指標の結果を示す。すべての指標において、提案手法がベースラインを上回った。また、Q5 のレポートの信頼性への自己評価と、LLM による誤り率と品質の評価以外の指標において有意差が観測された。一方で、誤り率と品質の評価に関しては、提案手法とベースラインの差は小さく、有意差は観測されなかった。上記の結果から、提案システムは従来の Deep Research と比較して、効率的で網羅性の高いレポート生成が、ユーザにとって低負担で行えることが示唆された。実験参加者からは、「調査内容がタイル形式で出力されるため視覚的にわかりやすい」など、タイル型 UI を肯定的に評価する意見が多く得られた。一方で、調査の進行に伴いタイルの数が過剰に増える点を指摘する意見もあった。

表 1. ユーザ実験の結果

指標	N	Deep 平均 (SD)	提案平均 (SD)	中央値差	r_{rb}	p
アンケート結果（リッカートスケール：7段階）						
Q1（有用性）↑	32	3.531 (1.722)	4.969 (1.675)	-2.00	-0.713	0.001750**
Q2（網羅性）↑	32	3.594 (1.500)	5.000 (1.626)	-2.00	-0.772	0.000307**
Q3（深掘り）↑	32	3.500 (1.459)	5.125 (1.519)	-2.00	-0.926	0.000008**
Q4（成果物品質）↑	32	3.281 (1.420)	4.594 (1.456)	-2.00	-0.849	0.000094**
Q5（信頼性）↑	32	3.969 (1.356)	4.375 (1.362)	-1.00	-0.442	0.095512
UEQ-S（実用性・ヘドニック：-3.0～3.0）						
実用性 ↑	32	-0.781 (1.425)	0.953 (1.408)	-1.75	-0.800	0.000028**
ヘドニック ↑	32	-0.023 (1.224)	1.289 (1.227)	-1.00	-0.896	0.000001**
完了時間・メンタルワークロード						
タスク時間（分） ↓	25	68.200 (50.158)	45.040 (25.800)	14.00	0.492	0.031808*
NASA-TLX（5-100） ↓	32	67.656 (18.184)	51.224 (16.540)	16.67	0.827	0.000012**
レポートの LLM 評価（誤り率・品質・網羅性）						
誤り率 ↓	32	0.044 (0.085)	0.026 (0.045)	0.083	0.199	0.495079
品質（45 点満点） ↑	32	32.719 (2.630)	33.062 (2.169)	-1.00	-0.129	0.564731
網羅率（重要のみ） ↑	32	0.625 (0.241)	0.825 (0.103)	-0.18	-0.720	0.000316**
網羅率（全て） ↑	32	0.548 (0.187)	0.744 (0.122)	-0.19	-0.817	0.000070**

注）平均 (SD) は各指標で 2 つの手法のペアが揃った観測のみを用いて算出。差分は Deep- 提案（正は Deep が大）。 p 値は Benjamini-Hochberg 法による FDR 補正後の値。 p 列は有意であれば **太字+記号**（*： $p < .05$, **： $p < .01$ ）。

5 議論

実験の結果から調査タスクにおいて提案システムが優れている面が明らかになった。提案システムでは、システムが提案する調査項目を承認していくことで、新たな項目の調査や関連内容の深掘りなどを行うことができるため、ユーザが追加のトピックを自ら入力せずに調査を進めることができる。そのため、ユーザが最初に多くの項目の調査を実行した後、レポートを閲覧しながら情報の取捨選択が容易であった。一方で、Deep Research は、入力トピックに関連が深い調査項目のみが提案されるため、網羅性を高めるためには、ユーザがこれまでの調査結果から適切なトピックを改めて入力する必要があった。

信頼性への自己評価に関しては両システム間で同程度であったが、誤り率に関しても同程度であったため、調査の信頼性に関しては調査実行エージェントの性能に依存すると考えられる。提案手法は既存の Deep Research のフレームワークにおけるエージェントを活用しているため、各エージェントがより高性能になれば提案手法の有用性もさらに高まると考えられる。LLM による自動評価ではレポート品質に有意差は見られなかったが、アンケートによる主観評価では提案システムのレポート品質が有意に高く評価された。これは、ユーザが提示候補を選択して調査を進める過程で選択支持バイアスが働いた

可能性がある。また、長文の文章を LLM で適切に評価する方法が未確立であることも要因の一つと考えられる。しかし、品質が同程度だとしても、ユーザの負担が低く網羅性の高いレポートを生成できることは、提案システムの優位性だといえる。

実験参加者からは、提案システムが提示する調査候補タイルの数が調査を進めていくと増えすぎるといった意見もあった。そのため、今後は調査候補のランキングによる提示量抑制や類似トピックのクラスタリングなど、ユーザビリティ面の改善や、調査の網羅性をさらに高めるための検討を行いたい。

6 むすび

本研究ではユーザが能動的に探索可能な、ユーザと AI エージェントが協調的に調査タスクを行うための UI を開発し、その効果を検証した。提案システムは従来手法よりもインタラクティブな調査項目の選択が可能であるため、ユーザは幅広いトピックを容易に調査することができる。選択した項目に応じて深掘り調査項目の候補が自動追加されるため、ユーザがこれまでの結果に基づき新たな調査トピックを考えて入力する手間を低減させることができる。ユーザ実験では、提案システムは従来のシステムと比較して、ユーザが少ない負担でより網羅的なレポートを生成できることを明らかにした。

謝辞

ユーザ実験実施にあたり多くの協力をいただいた山崎天さん、吉田奈央さんに感謝いたします。

参考文献

- [1] M. Aliannejadi, H. Zamani, F. Crestani, and W. B. Croft. Asking Clarifying Questions in Open-Domain Information-Seeking Conversations. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*, pp. 475–484. ACM, 2019.
- [2] M. J. Bates. The Design of Browsing and Berrypicking Techniques for the Online Search Interface. *Online Review*, 13(5):407–424, 1989.
- [3] J. Dalton, C. Xiong, and J. Callan. TREC CAsT 2019: The Conversational Assistance Track Overview. In *Proceedings of the Text REtrieval Conference (TREC 2019)*, 2020.
- [4] A. R. Fabbri, W. Kryściński, B. McCann, C. Xiong, R. Socher, and D. Radev. SumEval: Re-evaluating Summarization Evaluation. *Transactions of the Association for Computational Linguistics*, 9:391–409, 2021.
- [5] Google. Gemini Deep Research — your personal research assistant. <https://gemini.google/overview/deep-research/>, 2025. Accessed 2025-08-01.
- [6] X. Gu, Jiawei anzd Jiang, Z. Shi, H. Tan, X. Zhai, C. Xu, W. Li, Y. Shen, S. Ma, H. Liu, et al. A survey on llm-as-a-judge. *arXiv preprint arXiv:2411.15594*, 2024.
- [7] M. A. Hearst. Clustering versus Faceted Categories for Information Exploration. *Communications of the ACM*, 49(4):59–61, 2006.
- [8] E. Karpas, O. Abend, Y. Belinkov, B. Lenz, O. Lieber, N. Ratner, Y. Shoham, Y. Levine, K. Leyton-Brown, D. Muhlgay, N. Rozen, E. Schwartz, G. Shachaf, S. Shalev-Shwartz, A. Shashua, and M. Tenenholz. MRKL Systems: A Modular, Neuro-Symbolic Architecture that Combines Large Language Models, External Knowledge Sources and Discrete Reasoning. *arXiv preprint arXiv:2205.00445*, 2022.
- [9] G. Marchionini. Exploratory Search: From Finding to Understanding. *Communications of the ACM*, 49(4):41–46, 2006.
- [10] S. Mehri and M. Eskenazi. USR: An Unsupervised and Reference Free Evaluation Metric for Dialog Generation. In D. Jurafsky, J. Chai, N. Schluter, and J. Tetreault eds., *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 681–707, Online, July 2020. Association for Computational Linguistics.
- [11] Microsoft. Researcher agent in Microsoft 365 Copilot. [https://techcommunity.microsoft.com/blog/microsoft365copilotblog/resea](https://techcommunity.microsoft.com/blog/microsoft365copilotblog/researcher-agent-in-microsoft-365-copilot/4397186)
- rch-er-agent-in-microsoft-365-copilot/4397186, 2025. Accessed 2025-08-01.
- [12] R. Nakano, J. Hilton, S. Balaji, J. Wu, L. Ouyang, C. Kim, C. Hesse, S. Jain, V. Kosaraju, W. Saunders, X. Jiang, K. Cobbe, T. Eloundou, G. Krueger, K. Button, M. Knight, B. Chess, and J. Schulman. WebGPT: Browser-assisted Question-Answering with Human Feedback. *arXiv preprint arXiv:2112.09332*, 2021.
- [13] OpenAI. Introducing deep research. <https://openai.com/index/introducing-deep-research/>, 2025. Accessed 2025-08-01.
- [14] R. Pradeep, N. Thakur, S. Upadhyay, D. Campos, N. Craswell, and J. Lin. Initial Nugget Evaluation Results for the TREC 2024 RAG Track with the AutoNuggetizer Framework. *arXiv:2411.09607*, 2024.
- [15] F. Radlinski and N. Craswell. A Theoretical Framework for Conversational Search. In *Proceedings of the 2017 ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR)*, pp. 117–126. ACM, 2017.
- [16] T. Schick, J. Dwivedi-Yu, R. Dessi, R. Raileanu, M. Lomeli, L. Zettlemoyer, N. Cancedda, and T. Scialom. Toolformer: Language Models Can Teach Themselves to Use Tools. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.
- [17] Y. Shen, K. Song, X. Tan, D. Li, W. Lu, and Y. Zhuang. HuggingGPT: Solving AI Tasks with ChatGPT and its Friends in Hugging Face. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.
- [18] Q. Wu, G. Bansal, J. Zhang, Y. Wu, B. Li, E. Zhu, L. Jiang, X. Zhang, S. Zhang, J. Liu, A. H. Awadallah, R. W. White, D. Burger, and C. Wang. AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation. *arXiv preprint arXiv:2308.08155*, 2023.
- [19] S. Yao, J. Zhao, D. Yu, N. Du, I. Shafran, K. Narasimhan, and Y. Cao. ReAct: Synergizing Reasoning and Acting in Language Models. In *Proceedings of the 11th International Conference on Learning Representations (ICLR) Workshop/ArXiv*, 2022. arXiv:2210.03629.
- [20] H. Zamani, J. R. Trippas, J. Dalton, and F. Radlinski. *Conversational Information Seeking*. Foundations and Trends in Information Retrieval. Now Publishers, 2022.

未来ビジョン

本研究で提案した能動的 Deep Research システムは、現行の大規模言語モデルおよび検索アルゴリズムの処理性能を前提として設計されている。将来的には、モデルの推論効率や探索アルゴリズムの改善により、情報探索の速度が飛躍的に向上することが予想される。そのような環境においても、本システムの設計理念は有効であると考ええる。本論文で強調したいのは、処理速度の向上そのものではなく、ユーザが探索の流れを能動的に制御し、自らの理解を形成していく過程を支える設計である。本手法の価値は、そのような人間中心の探索

体験を成立させる点にある。

また、探索の深度と処理速度の間には、トレードオフの関係が存在する。情報源を増やし、多段階的な推論を行うほど、処理時間は不可避免的に増加する。この関係は技術の進歩によって緩和されるとしても、完全に消えることはない。こうした制約の中で、人間と AI がどのように関わり、共に知を形成していくのが良いのだろうか。その答えは一つではないと思うが、まさにその問いを考え続けることこそが、AI 時代における HCI テーマの一つであり、本研究もその一端を描くものである。

Paper Gallery: 論文要約スライドをコミュニティの資産に

牧野 倫太郎* 高橋 治輝* 松村 耕平*

概要. ゼミや勉強会で作成される論文要約スライドは、テキスト中心の論文情報と比較して、研究内容を視覚的に理解できるという点で価値ある知的資産である。しかし、これらのスライドは発表後に散逸しがちであり、その活用は各研究者の管理や研究室ゼミの運営方針などの研究コミュニティに委ねられているのが現状である。本稿ではこの課題に対し、論文要約スライドを活用した、一連の研究活動を支援する Web システム「Paper Gallery」を提案する。本システムは、(1) 既存の情報共有プラットフォームとの連携によるユーザ負担のない自動データベース構築、(2) スライドの視認性を高めるギャラリー表示と検索機能による知識発見の促進、(3) 論文要約スライドを介した研究者間のコミュニケーション活性化を特徴とする。

1 はじめに

研究者にとって、日々多くの論文を読み、その内容を効率的に整理することは、研究活動を進める上で極めて重要である。この活動を支援するため、PDF リーダーにさまざまな機能を追加することで個人の論文読解を効率化するツール [6] が開発されるだけでなく、集団では研究室のゼミや勉強会といったコミュニティ単位での論文紹介も活発に行われている。

このコミュニティでの論文紹介において、研究内容をスライド 1 枚に要約する手法は、その視覚的な情報把握のしやすさから、研究室内外での情報共有や議論の促進に広く用いられている [8, 14]。例えば、CHI 勉強会はこの要約手法を推奨しており、資料作成要項で「論文の核となる図や写真をスライドに載せるなど、概要がすぐ見てわかるスライドを作成してください」と指示している¹。また、著者らの研究室では Cosense²を研究室の情報共有プラットフォームとして活用しており、CHI 勉強会と同様のフォーマットで作成した論文要約スライドを共有し、1 分間で論文内容を紹介するゼミを実施している。適切に作成された要約スライドは、論文内容を効果的に伝達するだけでなく、大量の論文から興味のあるものを発見するための優れた資源となりうる。

しかし、これらの論文要約スライドは有益な知的資産であるにもかかわらず、その多くは発表後に PDF や PowerPoint 等のファイル形式のまま保管され、内容の横断的な検索や一覧が困難である。また、その活用は各研究者の管理や研究室ゼミの運営方針などの研究コミュニティに委ねられているのが現状である。例えば、新たに研究室に配属された学生が過去の論文を概観したり、自身の研究に関連する知

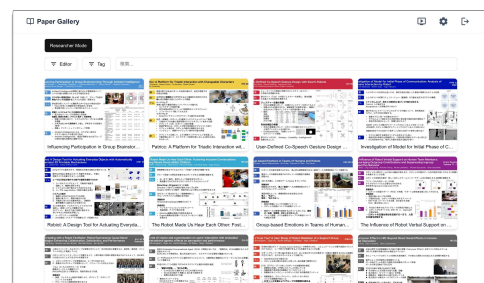


図 1. 提案システム。勉強会での発表後も論文要約スライドをデータベースとして活用できる。

見を持つメンバーを見つけたりといった活用が考えられるが、このようなコミュニティの集合知としてのさらなる活用には、支援の余地がある。この課題に対し、論文要約スライドを管理するために Wiki のような既存ツールを利用することは、テキスト主体であるためスライドの視覚的価値を損ない、論文以外の情報に知見が埋没する懸念がある。一方で、スライド共有専用のデータベースを新たに構築・運用するアプローチは、情報の手動入力やフォーマット統一といった作業負担がユーザの日常的な活動を阻害し、導入・定着の障壁となる。

そこで本研究では、コミュニティで日々作成される論文要約スライドを、特別な負担なく体系的に蓄積・活用する Web システム「Paper Gallery」を提案する (図 1)。本システムの目的は、発表後に散逸し、単なる記録にとどまりがちなスライドを、誰もが横断的に検索・再利用でき、新たなコミュニケーションを誘発する「コミュニティの資産」へと転換させることにある。

この目的を、Paper Gallery は主に三つの特徴によって実現する。第一に、コミュニティで既に運用されている情報共有プラットフォーム (本稿では Cosense) からデータを自動取得することで、ユー

Copyright is held by the author(s).

* 立命館大学

¹ CHI 勉強会 2025: 発表資料作成要綱, <https://sigchi.jp/seminar/chi2025/slide.html>

² Helpfeel: Cosense, <https://cosen.se/product>

ずに導入・運用の負担を強いることなくデータベースを自動構築する。第二に、このデータベースをスライドの視認性を最大限に活かしたギャラリー UI で提示し、効率的な情報発見と視覚的な内容の想起を支援する。第三に、論文要約スライドを起点とした議論やアイデア共有を促す機能を提供することで、単なる情報の閲覧に留まらない、コミュニティ内でのコミュニケーションを活性化させる。

2 関連研究

研究者にとって論文を読むことは重要であるが、日々増え続ける膨大な量の論文を読み解く負担は大きい。そのため、論文読解を支援する研究は数多い。支援手法の一つとして、PDF リーダーの機能を拡張し、論文読解体験の向上を目指す研究がなされている。例えば、専門用語や数式の定義をその場で提示する ScholarPhi [3]、自然言語処理を用いて論文内容を要約する PaperPlain [1]、そして個人の閲覧履歴に基づいて関連性の高い引用を提示する CiteSee [2] などがある。The Semantic Reader Project [6] は、これらを含む 10 の研究を統合し、読者の理解を支援する新たな論文 PDF の読書インターフェースを開発する取り組みである。また、論文を介した議論の促進という観点では、Huang らが提案した SURF [4] がある。このシステムは、論文とソーシャルメディア上の議論を結びつけ、個人の読解体験を豊かにすることを目指している。個人の読解支援に対し、コミュニティ単位での論文読解を支援する取り組みもある。具体的には、CHI 勉強会 [14] や、その運営を支援するシステムである Paper Guilds ³ がある。このシステムは、勉強会運営の負担を軽減し、コミュニティでの論文共有を支援する。

これらの研究を踏まえ、個人の読解支援に注目する The Semantic Reader Project [6] などの先行研究に対し、本研究はコミュニティを対象とする。また、Paper Guilds が勉強会の運営支援を目的とする一方、本研究は、勉強会で作成される論文要約スライドという視覚的情報をコミュニティの知的資産として捉え、その発表後における継続的な活用に着目する。これは、論文全体よりも要点が構造化・視覚化されたスライドの方が、コミュニティ内での知識の再利用や議論の活性化を促進しやすいという考えに基づいている。

本研究は論文要約スライドという視覚的な情報に注目している。同様に、ユーザが作成した視覚的な情報を用いて議論を促す研究も存在する。Viegas らの ManyEyes [9] は、ユーザがデータをアップロードし、インタラクティブな可視化を作成・議論できる Web サイトであり、ソーシャルなデータ分析を促進する。さらに、論文情報の視覚的な理解という点

では、Zhu らの研究 [11] がある。この研究は、抽象的な学術情報を現代アートのような表現を用いて直感的に理解可能にするデザインを探索している。本研究とは表現のアプローチこそ異なるものの、論文情報の視覚的な理解の促進という点で共通している。

3 システムデザイン

本システムは、運用しながら改善を加える反復的なプロセスで開発した。当初は、研究室内で蓄積される論文要約スライドの再活用を目的にシステムを実装した。その後、プロトタイプ [16] を用いたゼミでの運用や機能追加を繰り返す中で、論文要約スライドを「コミュニティの資産」へと転換させるという最終的な目的と、それを実現するために必要な以下の 3 つの設計方針が得られた。

3.1 (D1) 導入・運用障壁の最小化

コミュニティに新しいシステムを導入する際の課題は、メンバーに新たな負担を強いることである。著者らの研究室では、論文要約スライドにまとめて Cosense に提出し発表する、という既存の論文読み活動があった。これを阻害しないことが、システムを継続的に利用するうえで重要となる。もし厳格なフォーマットを強要すれば、それはメンバーの負担となるだけでなく、個々の工夫や多様性が失われ、論文読み活動のモチベーションを削いでしまう。我々が重視するのは、メンバーそれぞれが作成する多様な視点や特徴が反映されたスライドである。昨今では大規模言語モデルを用いて論文要約スライドを自動生成する技術も存在する⁴ [5] が、自身の研究関心に基づいて、論文のどの部分を重要と捉えるかという解釈、さらには図表の選び方や情報の配置といったデザインの工夫に、作成者ならではの個性が反映される。この多様性を維持するためにも、システムが利用者の活動を制限してはならない。例えば、共有データベースのために論文情報を定型フォーマットで登録するよう求められれば、利用者は自身のメモを追記したり、独自の観点で整理したりすることが難しくなる。利用者が既存の論文読みのやり方を変えることなく、できるだけ自由なフォーマットで記述できることが、活動を継続する上で重要となる。そこで本研究では、利用者に課す負担を、論文のタイトルづけや画像の配置といった最低限の規約に留め、**普段の活動を妨げることなくシステムを導入・運用できること**、を一つの設計方針 (D1) とした。

3.2 (D2) 知識の探索と活用の促進

従来、研究室メンバーが作成したスライドは誰でも閲覧できるものの、各々の個人ページにまとめら

³ Paper Guilds, <https://pg1.jp/>

⁴ GitHub: Proceedings-summarizer, <https://github.com/ShotaImamura/Proceedings-summarizer>

れ、一覧性に乏しく、タグ付けをしても関連するスライドを探し出すことが困難であった。その結果、スライドが蓄積されてもコミュニティの知見として有効活用されず、個々の記録に埋もれてしまうという課題があった。この論文要約スライドには高い視覚的価値がある。テキストよりも画像の方が記憶に残りやすいという画像優位性効果が知られており、実際にテキストと画像が含まれる Web サイトの閲覧タスクにおいて、利用者はテキスト領域の約 2 倍の時間をサムネイル画像の注視に費やしたと報告されている [10]。多くの論文データベースがテキスト情報に依存するのに対し、図やキーワードを視覚的に配置したスライドは、内容の直感的な理解と記憶の定着を助ける可能性がある。この視覚的価値は、研究室に配属されたばかりの学生など、その分野の専門知識がまだ十分でないメンバーにとって、論文読解の負担を軽減しながら分野全体を概観する、有効な手助けとなる。さらに、優れたスライドを多数一覧・比較することは、要約すべき情報の見極め方やデザインスキルを学ぶ、実践的な学習環境を提供する。また、蓄積されたスライドを俯瞰することは、新たな研究アイデアの創出も期待できる。このように埋もれた価値を最大化し、知識の探索と活用を促す必要性から、論文要約スライドの視覚的価値を活かし、利用者が膨大な知識の中から新たな発見を容易に行えることを二つ目の設計方針 (D2) とした。

3.3 (D3) コミュニケーションの誘発

スライドには単なる知識の記録に留まらず、人と知識、そして人と人を結びつける「触媒」となり得る可能性がある。このコミュニケーションの誘発は、大きく二つの場面に分けられる。一つは蓄積されたスライドを通じた非同期的な繋がりであり、もう一つは論文紹介ゼミのような場における同期的な議論である。この非同期的な繋がりにおいて、スライドの「誰が作成者か」という情報は、これまで十分に活用されてこなかった貴重な情報と言える。例えば、同じ研究室の先輩が作成したスライドは、自身の研究テーマに近い論文を見つけるための信頼できる指針となる。また、勉強会であれば、他大学の研究者など、同じ論文に興味を持つ専門家と繋がるきっかけにもなり得る。しかし、従来の個人ページによる管理では、こうしたスライドを偶然的に発見することは難しく、それをきっかけとしたコミュニケーションが発生しづらかった。一方、同期的な議論の場では、発表が一方的になりがちで活発な議論が生まれにくいという課題があった。そこで、スライドという共通の知識基盤を活用し、より多くの参加者が活発な議論を交わせるようにするためのインタラクティブな仕組みの必要性が高まった。このように、スライドの価値を最大限に引き出すには、過去の知識へのアクセスを容易にするだけでなく、対話の場

を豊かにする必要がある。この知見に基づき、スライドという知識を起点とし、コミュニティ内の新たな繋がりや議論を誘発することを三つ目の設計方針 (D3) とした。

4 提案システム: Paper Gallery

本稿で提案する Web システム「Paper Gallery」は、3 章で述べた反復的な運用と実装のプロセスを経て開発された。本章では、各設計方針に対応する主要機能と、その実装について述べる。

4.1 (D1) 導入・運用障壁を最小化する機能

既存 Wiki との連携による自動データベース構築

本稿ではデータソースとして、研究室や学会で広く利用実績があり [7]、著者らの研究室においても既に運用されている Cosense を活用する。ただし、これは実装の一例であり、API を通じてページ内容と画像を取得できる他の Wiki のようなツール (例: Notion⁵) でも代替できる可能性がある。また、過去の勉強会などで作成されたスライドも、後から Cosense にアップロードするだけで本システムを活用することができる。

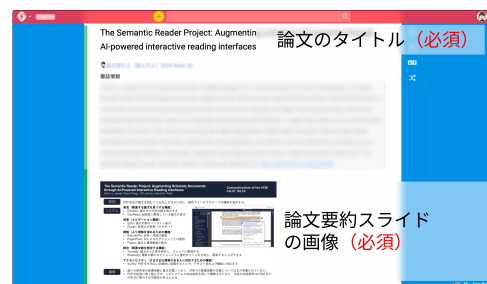


図 2. Cosense の論文まとめページの例

利用者は、図 2 のように論文タイトルをページ名とし、本文の最初に要約スライドの画像を配置するというシンプルな規約に従うだけでよい。システムが Cosense の API を介してこれらの情報を自動で取得・整形するため、利用者は特別な登録作業を意識せず、普段の活動の延長線上で知識資産を構築できる。さらに、Cosense のページは PDF などの従来の保存形式に比べ柔軟に編集でき、あとからメモや参照などの情報を追加できるという利点もある。

4.2 (D2): 知識の発見と再利用を促進する機能

俯瞰的な知識発見を促すギャラリー表示

論文要約スライドは、図 1 に示すギャラリー形式で表示される。従来のテキストベースの表示とは異なり、論文の核となる図やキーワードを視覚的に確

⁵ Notion: Notion, <https://www.notion.com/ja>

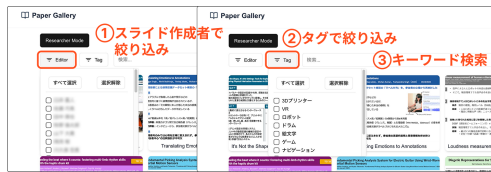


図 3. 検索・絞り込み機能

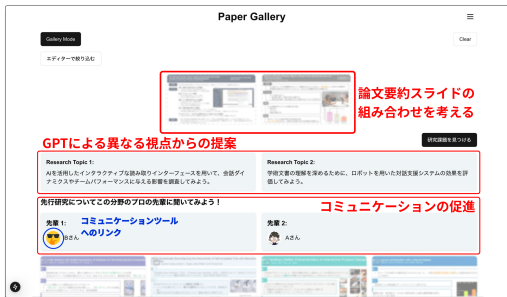


図 4. 新たな研究テーマを考える

認できるため、利用者は内容を掴みやすく、興味のある論文を直感的に探すことができる。研究分野に不慣れな学生にとっては、論文を探す心理的なハードルが下がり、分野全体を把握するきっかけとなる。

多角的な検索・絞り込み機能

目的の論文を効率的に探し出すため、図3に示す3種類の検索・絞り込み機能を提供する。多くの利用者は、自身のメモとしてCosense上に論文の著者名、学会名、関連キーワード、あるいは個人的な考察などを自由に記述している。これらを検索対象とすることで、登録作業を不要としつつ多角的な検索が可能なデータベースを構築した。スライド作成者での絞り込みでは、特定のメンバーがまとめた論文を一覧できる。タグでの絞り込みでは、システムがCosense本文から自動でキーワードを抽出し、タグを付与する。追加機能として利用者がCosense上に「#」記法で付与した任意のタグが利用可能である。キーワード検索では、著者名や学会名など任意のテキストで検索できる。これらの機能により、過去の論文要約スライドを容易に参照し、自身の研究内容の再確認や関連研究の効率的な参照が可能となる。

新たな研究テーマ創出支援機能

知識の再利用を促進するため、蓄積されたスライドを基に新たな研究テーマの着想を支援する機能を提供する。アイデアは既存の要素を組み合わせることで生まれるとされる[12]。本機能では、利用者が2件の論文要約スライドを選択すると、OpenAI APIを利用してGPTが両者の内容を組み合わせた新しい研究テーマを2つ提案する。例えば、論文精読とロボットに興味を持つユーザが、図4に示す画面で、

論文精読に関する先行研究[6]のスライドとロボットに関するスライドを組み合わせると、図4中のResearch Topicに示すように「AIを活用したインタラクティブな読み取りインタフェースを用いて、会話ダイナミクスやパフォーマンスに与える影響を調査してみよう。」「学術文書の理解を深めるために、ロボットを用いた対話支援システムの効果を評価してみよう。」という、2つの異なる視点からの研究テーマが提案される。このように、個人だけでは着想しにくい多様な視点を得ることが可能である。

4.3 (D3) コミュニケーションを誘発する機能 インタラクティブな議論を促すライブ機能

スライド発表の場における参加者間のコミュニケーションを活性化させ、活発な議論を誘発するため、図5に示すプレゼンテーションとライブチャットを統合した機能を提供する。



図 5. ライブチャット画面

このチャットとプレゼンテーションを統合した形式は、WISSでの発表においても活用されている[15]。また、発表者向けにスライドとタイマーのみが表示される全画面モードも備える。参加者は発表を聞きながらリアルタイムでコメントや質問を投稿できる。このチャット履歴はスライドと紐づけてCosenseに保存でき、その場で生まれた議論や気づきも知識として蓄積される。さらに、チャット機能のプロトタイプング時に利用者から得られた「感想のみではコメントしづらい」という意見から、意思表示のハードルを下げるため、絵文字によるリアクション機能も実装した。投稿された絵文字は全参加者の画面上にリアルタイムで表示され、数秒かけてフェードアウトすることで、活発な反応を視覚的に表現する。

研究者との連携支援

スライドを起点としたコミュニケーションを促進するため、スライドの作成者情報を活用し、コミュニティで既に利用されているコミュニケーションツールと連携する機能を提供する。利用者は、図6のようにスライドを拡大表示でき、作成者のアイコンをクリックすると、そのコミュニティで普段利用しているコミュニケーションツールへ遷移できる。現状

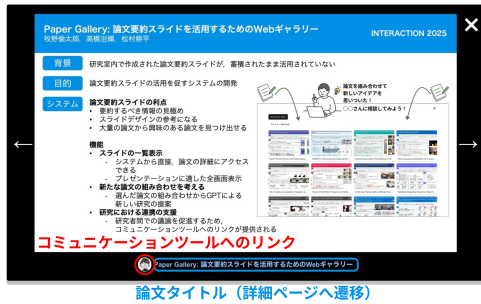


図 6. 各論文のページ

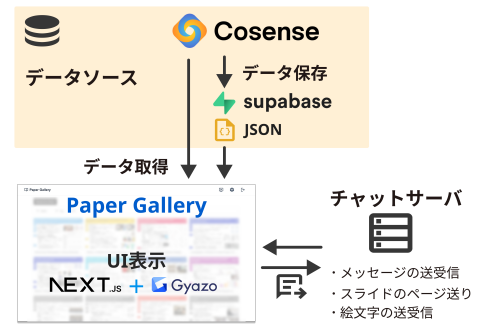


図 7. システム構成

では、著者らの研究室で日常的に Slack⁶が利用されているため、個人ごとのパブリックチャンネルへ遷移する実装となっているが、連携先は各コミュニティの運用に合わせて設定可能である。この連携は、新たな研究テーマを創出する際にも有効である。4.2 章の図 4 に示す画面では、GPT が提案した研究テーマと共に、「先行研究について聞いてみよう!」として関連スライドの作成者情報が表示される。これにより、例えばロボット研究に関心を持った学生が、その分野の専門家である先輩（図 4 中の A さん）に直接連絡を取り、自身の研究方針や読むべき論文について助言を求める、といった連携が生まれる。このように、スライドの作成者情報を活用することで、同じ興味を持つ人同士が交流するきっかけを創出する。

4.4 実装

本システムは、Next.js を用いた Web システムとして構築した。システム構成を図 7 に示す。専用のデータベースを持たず、データソースとして Cosense および画像共有サービス Gyazo⁷ の API を利用する。検索に必要な情報は Cosense 上に併記されているため、画像からテキストを読み取ることは不要である。これにより、画像データを中心にしながらも、必要な詳細情報をテキストで補完し、参照しやすい構造を実現している。システムのパフォーマンスと外部 API への負荷を考慮し、Cosense から取得した情報は外部のデータベース（本稿では supabase）や JSON 形式のファイルとして保存する。システムは主にこのデータを参照することで、スライド数が増えても高速なレスポンスと安定した動作を維持する。また、ライブチャット機能のメッセージ・絵文字の送受信とスライドのページ送りは、専用のチャットサーバを介して行われる。

5 活用事例

提案システムは、著者らの研究室において、10ヶ月間（2024 年 11 月～2025 年 8 月）にわたり継続

的に運用されている。本章では、この期間中に（大学の休暇期間などを除き）実施された週次の論文紹介ゼミでの活用事例も踏まえて、各設計方針を検証する。このゼミは、参加学生が論文要約スライドを用いて 1 分程度で論文内容を紹介する形式で、期間中に 22 回開催され、平均 21 名の学生が参加した。

5.1 (D1) 導入・運用障壁に関する事例

データベースには合計 928 件の論文要約スライドが登録された。この内訳は、期間中のゼミで新たに作成された 429 件に加え、システム導入以前に作成されていた 499 件の既存スライドである。これには卒業生が作成したものも含まれ、多様なメンバーによるスライドが蓄積されている。例えば、卒業生が残したスライドという知的資産が、研究室に配属されたばかりの学生にとって、自身の研究テーマに近い論文を探す際の一助となっている様子も観察された。また、ゼミを開催する際には、運営者が個別に発表資料を集約する手間を省けるため、ゼミ開催の負担軽減にもつながった。このように、システムが特別な負担なくゼミの既存のワークフローに組み込まれ、継続的に利用されたことを示している。

5.2 (D2) 知識の探索と活用促進に関する事例

運用期間中の観察において、本システムの検索・絞り込み機能が、知識の探索とコミュニティ内の新たな繋がりを促進する様子が確認された。

まず、検索・絞り込み機能は、世代を超えた知識の継承に有効であった。例えば、「トレーニング」というタグで検索すると、既に卒業した学生と新たに配属された学生のスライドが同時に表示される。これにより、教員は新入生へ同じ関心を持つ先輩の論文を具体的に紹介し、指導に役立てることができた。また、この機能は分野を超えた交流が生まれる可能性を可視化した。同じタグ検索から、ゲーム研究と音楽研究という異なる分野の学生が、共に「トレーニング」という共通のタグに含まれるという、これまで認識されていなかった事実が明らかになった。これは、本システムが専門領域を越えた交流のき

⁶ Slack: Slack, <https://slack.com/intl/ja-jp/>

⁷ Helpfeel: Gyazo, <https://gyazo.com/ja>

かけとなり得ることを示唆している。

加えて、ギャラリー表示はデザインスキルの学習環境としても機能した。多くの学生が自身の過去のスライドデザインをテンプレート的に利用する傾向や、その変遷を一覧することができた。ある学生は、より質の高い要約を目指して期間の途中でデザインを大きく変更したが、こうした個人の工夫や改善の過程を可視化し、コミュニティに共有できる点は、ギャラリー表示の大きな利点である。ただし、これらの利用がどの程度知識の活用につながったかという影響については本研究では評価していない。今後、利用者へのアンケートやインタビューを通して、この設計方針に関する評価を行う予定である。

5.3 (D3) コミュニケーション誘発効果の事例

システム運用中に機能追加したライブチャットの利用状況を評価する。図8のグラフは、機能導入後の週ごとのコメント投稿数の推移を示したものである。導入から時間が経過しても利用が落ち込むことはなく、1回の開催で最低でも80件以上（平均約100件）のコメントが投稿された。この結果は、機能が一時的な利用に留まらず、ゼミにおける議論を継続的に誘発する機能として定着したことが示された。

コメント内容を分析したところ、図8の分類に示すように、最も多かったのは「感想・意見」である。これには、論文内容への印象・評価や、研究内容への提案、発表に対する相槌といった多様な対話的なやり取りを含むものであり、全体の61%（646件）を占めている。このことから、本機能が場のインタラクティブ性の向上に貢献したことが示された。次いで多かったのが「質問」で、33%（355件）に上る。これには、「○○はどういう意味だろう?」という内容の理解を深めるためのものから、「○○の場合は?」、「××したらどうなるか?」といった具体的な代替案や新たな研究テーマの着想に繋がるような発展的な問いも含まれていた。その他、挨拶・リアクション（約4%、44件）や著者や学会といった「関連情報」（2%、19件）についての言及も見られた。

さらに、意思表示手段の一つとして、絵文字によるリアクション機能も利用されている。例えば、発表時間が長すぎる発表者には時計の絵文字が、優れた発表の後には拍手の絵文字が送信されるといった活用が見られた。この絵文字は匿名で送信することができ、文字のコメントとして直接的には指摘しづらいことを伝えるためにも活用されたと考えられる。

ただし、発言者の傾向として、全体のおよそ8割は教員によるコメントであった。これは、学生は自身の発表担当があることや、ゼミの教育的側面として、教員からのコメントには質問や発表資料への改善案が含まれることが要因として考えられる。この利用状況から、システムがコミュニケーションを誘発する土台として機能していることは確認できたが、

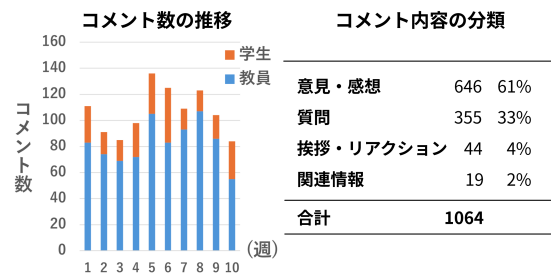


図 8. 週ごとの教員と学生のコメント数とその分類

より多様な参加を促す上での課題も示唆された。

6 展望

10ヶ月間の活用事例を通して、本システムがコミュニティの知識資産構築と知的コミュニケーション促進に有効であることが示された。その一方で、いくつかの改善点も明らかになった。これらの知見に基づき、今後の展望を述べる。

第一に、データベースの質的向上と作成者支援である。現状のチャットでの質問が多い一因として、スライドに記載されるべき情報の不足が考えられる。そこで、登録されたスライドを大規模言語モデルを用いて分析し、論文の核となる図やキーワードなど、含めるべき情報が欠けていないかを事前に確認する機能を検討している。

第二に、コミュニティ全体の議論活性化である。活用事例では、学生からの自発的な発言が少ないという課題が浮き彫りになった。この課題に対し、学生のコメント投稿を促すために、議論のきっかけとなるような問いをシステムが自動で提示する機能などを検討している。また、短い発表時間の中でチャットのタイミングがずれるという意見も寄せられたため、複数の時間軸を持つチャットシステムの研究 [13] を参考にインターフェースを改善し、前後のスライドにもコメントができるようにしたい。

第三に、個人の利用動機強化である。現状の検索・閲覧機能に加え、個々の利用者にとってのメリットをより強く感じられるよう、システムの閲覧ログなどを基に、各個人の研究テーマに合致する論文を自動で推薦する機能などを検討している。

7 おわりに

本研究では、コミュニティ内での論文読み活動で作成される論文要約スライドを、発表後も価値ある知識資産として活用するための Web システム「Paper Gallery」を提案した。既存の Wiki と連携することでユーザ負担のないデータベースを自動構築し、視認性の高いギャラリー表示やインタラクティブなライブ機能によって、コミュニティにおける知識の発見とコミュニケーションを促進することを目指した。

参考文献

- [1] T. August, L. L. Wang, J. Bragg, M. A. Hearst, A. Head, and K. Lo. Paper Plain: Making Medical Research Papers Approachable to Healthcare Consumers with Natural Language Processing. *ACM Trans. Comput.-Hum. Interact.*, 30(5), Sept. 2023.
- [2] J. C. Chang, A. X. Zhang, J. Bragg, A. Head, K. Lo, D. Downey, and D. S. Weld. CiteSee: Augmenting Citations in Scientific Papers with Persistent and Personalized Historical Context. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, New York, NY, USA, 2023. Association for Computing Machinery.
- [3] A. Head, K. Lo, D. Kang, R. Fok, S. Skjonsberg, D. S. Weld, and M. A. Hearst. Augmenting Scientific Papers with Just-in-Time, Position-Sensitive Definitions of Terms and Symbols. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA, 2021. Association for Computing Machinery.
- [4] R. Huang, A. K. Zhao, Z. Saghi, S. Sabouri, and S. Chattopadhyay. Beyond the Page: Enriching Academic Paper Reading with Social Media Discussions. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology*, UIST '25, New York, NY, USA, 2025. Association for Computing Machinery.
- [5] S. Imamura, H. Hiraki, and J. Rekimoto. Serendipity Wall: A Discussion Support System Using Real-time Speech Recognition and Large Language Model. In *Proceedings of the Augmented Humans International Conference 2024*, AHs '24, p. 237–247, New York, NY, USA, 2024. Association for Computing Machinery.
- [6] K. Lo, J. C. Chang, A. Head, J. Bragg, A. X. Zhang, C. Trier, C. Anastasiades, T. August, R. Authur, D. Bragg, E. Bransom, I. Cachola, S. Candra, Y. Chandrasekhar, Y.-S. Chen, E. Y.-Y. Cheng, Y. Chou, D. Downey, R. Evans, R. Fok, F. Hu, R. Huff, D. Kang, T. S. Kim, R. Kinney, A. Kittur, H. B. Kang, E. Klevak, B. Kuehl, M. J. Langan, M. Latzke, J. Lochner, K. MacMillan, E. Marsh, T. Murray, A. Naik, N.-U. Nguyen, S. Palani, S. Park, C. Paulic, N. Rachatasumrit, S. Rao, P. Sayre, Z. Shen, P. Siangliulue, L. Soldaini, H. Tran, M. van Zuylen, L. L. Wang, C. Wilhelm, C. Wu, J. Yang, A. Zamarron, M. A. Hearst, and D. S. Weld. The Semantic Reader Project. *Commun. ACM*, 67(10):50–61, Sept. 2024.
- [7] note・Satoshi Nakamura. [研究室ゼミ] 5分プレゼン・5分質疑の論文紹介. <https://note.com/nkmr/n/n101d391725fe>, (参照 2025-08-31).
- [8] slideshare・Yoichi Ochiai. 先端技術とメディア表現1 #FTMA15. <https://www.slideshare.net/slideshow/1-ftma15/47697911>, (参照 2025-08-31).
- [9] F. B. Viegas, M. Wattenberg, F. van Ham, J. Kriss, and M. McKeon. ManyEyes: a Site for Visualization at Internet Scale. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1121–1128, 2007.
- [10] C. Yangandul, S. Paryani, M. Le, and E. Jain. How many words is a picture worth? attention allocation on thumbnails versus title text regions. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*, ETRA '18, New York, NY, USA, 2018. Association for Computing Machinery.
- [11] S. Zhu and J. Chan. Sense and Sensability: Exploring Future Immersive Environments for Scholarly Sensemaking. In *Proceedings of the 2025 Conference on Creativity and Cognition*, p. 886–901, New York, NY, USA, 2025. Association for Computing Machinery.
- [12] ジェームス・W・ヤング. アイデアのつくり方. CCC メディアハウス, 1988.
- [13] 小倉 加奈代, 松本 遥子, 山内 賢幸, 西本 一志. Kairos Chat: 主観的時間の概念を導入したチャットシステム. インタラクシオン 2010 論文集, 2010(4):259–266, 2010.
- [14] 松村 耕平ほか. CHI 勉強会 2017: ネットワーク連携した勉強会とその支援システム. 研究報告ヒューマンコンピュータインタラクシオン (HCI), 2017.
- [15] 綾塚 祐二, 河口 信夫. 参加者が作る会議支援システム〜WISS Challenge〜. コンピュータ ソフトウェア, 23(4):76–81, 2006.
- [16] 牧野 倫太郎, 高橋 治輝, 松村 耕平. Paper Gallery: 論文要約スライドを活用するための Web ギャラリー. インタラクシオン 2025 論文集, pp. 709–712, 2025.

PrefTolerance:

歌声・伴奏・歌詞ごとの嗜好と許容度を考慮できる楽曲推薦インタフェース

中澤 一輝^{*†} 佃 洸[†] 中野 倫靖[†] 後藤 真孝[†]

概要. 楽曲に対する嗜好には、歌声・伴奏・歌詞といった音楽要素に対する嗜好の度合いがあると考えられる。しかし、従来の楽曲推薦システムでは、ユーザは推薦された各楽曲が自分の嗜好に合うかどうかを楽曲単位で評価することが一般的であるため、音楽要素ごとのユーザの嗜好を十分に反映した推薦が難しい問題がある。また仮に、ある楽曲のある音楽要素に対するユーザの嗜好の度合いが高いことがわかったとしても、その音楽要素とどの程度まで類似した楽曲であればそのユーザに推薦すべきかわからない問題もある。本論文では、これらの問題を解決するための楽曲推薦インタフェース PrefTolerance を提案する。具体的には、歌声・伴奏・歌詞の3つの音楽要素を対象とし、各音楽要素に対してユーザが嗜好の度合いを評価可能にする。さらに、各音楽要素に対して、ユーザがどの程度の類似度を嗜好に合う範囲として許容できるかの度合いを「許容度」として指定可能にする。PrefTolerance は音楽要素ごとの嗜好と許容度に応じて、嗜好に合う楽曲の範囲がどのように変化するかを可視化することで、許容度の低さ（こだわりの強さ）を重視した楽曲推薦ができる。

1 はじめに

「歌詞は好みだけど歌声は好みでない」あるいは「歌声は好みだけど伴奏が好みでない」のように、音楽の好みは楽曲を構成する音楽要素 [13, 21] によって変わる。さらに、「歌詞はこだわらないので好みと違ってても許容できるけど、歌声は好みの曲しか聴きたくない」のように、音楽要素ごとにどの程度まで好みと違ってても許容できるかという度合い（許容度）は異なる。しかし、これまでの楽曲推薦システム [1, 20, 25] では、楽曲単位での嗜好に基づいており、許容度のような観点も考慮されていなかった。

そこで本論文では、歌声・伴奏・歌詞の3つの音楽要素ごとの嗜好（preference）と、各音楽要素に対する許容度（tolerance）をユーザが指定可能な楽曲推薦インタフェース PrefTolerance を提案する。ここでは「許容度」を、音楽要素ごとにどの程度の類似度までを嗜好に合う範囲として許容できるかの度合いと定義する。PrefTolerance は、ユーザが様々な楽曲の3つの音楽要素に対する嗜好を5段階で評価できる機能と、それぞれの要素に対する許容度をいつでも変更できる機能を備えており、これらの機能に基づいて楽曲推薦を行う。ユーザが許容度を直感的に指定できるようにするため、音楽要素ごとに別々に楽曲間の類似度を俯瞰して把握できる可視化を行う。さらに、ユーザが一度の操作で満足する許容度を決定することは困難であるため、スライダを用いてユーザが許容度を調整できるようにし、その

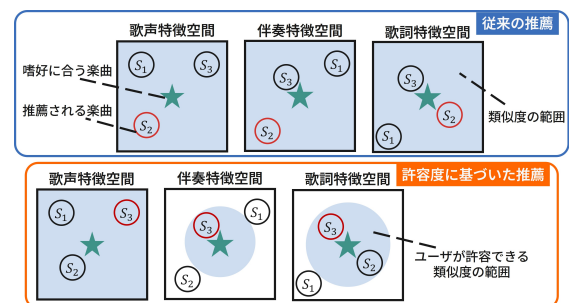


図 1. 各音楽要素の嗜好に基づいた推薦で、ユーザが許容できる類似度の範囲の指定が無い場合と、指定した場合。

変化に応じて嗜好に合う楽曲の範囲がどのように変わるかをインタラクティブに可視化する。

PrefToleranceにより、単なる嗜好を超えた、ユーザのこだわりを反映した楽曲との出会いを実現できる。ユーザがある楽曲の歌詞を嗜好に合うと評価した場合、それと歌詞がどの程度類似した楽曲であれば、そのユーザの嗜好に合うかは自明ではない。楽曲を聴く際に歌詞に対するこだわりが強い（許容度が低い）ユーザであれば、類似度の高い歌詞を持つ楽曲だけを嗜好に合うと判断する一方で、歌詞に対するこだわりが弱いユーザであれば、類似度が低い歌詞を持つ楽曲であっても嗜好に合うと判断するであろう。すなわち、音楽要素ごとの嗜好が得られただけでは、ユーザがその音楽要素に関してどの程度の類似度までを嗜好に合う範囲として許容できるかが分からないため、推薦すべき楽曲を適切に絞り込めないという問題がある。例えば、図1の上段の例

Copyright is held by the author(s).

^{*} 筑波大学

[†] 産業技術総合研究所

では、ユーザが各音楽要素に対して許容できる類似度の範囲の指定が無い場合、楽曲 S_1 , S_2 , S_3 のいずれもが推薦候補となり、その中で嗜好に合う楽曲と最も類似度が高い楽曲 S_2 が推薦されてしまう。この問題に対して、PrefTolerance では、図 1 の下段の例のように、ユーザが各音楽要素に対して許容できる類似度の範囲を指定可能にする。これにより、ユーザは自身のこだわりの強い音楽要素の類似度を重視した楽曲推薦を体験することができる。

2 関連研究

本章では関連研究に対する本研究の位置づけを述べる。

2.1 推薦手法

楽曲推薦インタフェースで主に用いられる方式として、協調フィルタリングと内容に基づくフィルタリングの 2 つが挙げられる [26]。協調フィルタリングは、他のユーザの聴取履歴などから得られる嗜好傾向を基に楽曲推薦を行う手法である [8, 18]。この手法は、楽曲の内容に関する特徴の解析を必要としない利点があるが、新規ユーザや未評価の楽曲に対応しづらい問題 [3] などがある。それに対して、内容に基づくフィルタリングは、楽曲の音色などの音響特徴 [9] を統合したベクトル間の類似度に基づいて、ユーザが好きな楽曲に類似した楽曲を見つけることで、嗜好に合った楽曲を推薦できる手法 [6, 14, 23] である。これは再生数に依存しないことから、前述した問題は起きにくく、PrefTolerance もこちらに位置づけられる。しかし、従来の内容に基づくフィルタリングでは楽曲の特徴全体に対するユーザの嗜好しか反映できなかった。PrefTolerance は、2 次元空間上で楽曲間の類似度を可視化でき、かつ音楽に詳しくないユーザであっても比較的判別しやすい歌声・伴奏・歌詞についての音楽要素ごとに嗜好を反映できる利点を持つ。

2.2 楽曲間の類似度とユーザの嗜好の可視化

楽曲間の類似度を俯瞰して把握できる可視化として、Hamasaki ら [5] は、同曲異唱楽曲を 2 次元空間上で可視化し探索できる機能を持つ Songrium を提案した。Vad ら [22] は、音響特徴の類似度を基に楽曲を 2 次元空間上の点として配置し、楽曲のムードやテンポなどの情報を点の色を用いて可視化するインタフェースを提案した。Sasaki ら [19] は、歌詞のトピックおよび類似度をレーダーチャートと 2 次元空間上のマッピングによって可視化し探索できる LyricsRadar を提案した。これらの研究は、特定の音楽要素に対する可視化を実現しているが、PrefTolerance では、音楽要素ごとの類似度をそれぞれ 2 次元空間で可視化できる点に特徴がある。

一方、楽曲に対するユーザの嗜好を可視化する研究として、Saito ら [17] は、楽曲のテンポや音量といった 5 種類の音響特徴の中から、ユーザが任意に 2 つの特徴を選択し、それらを軸とした 2 次元空間に配置し、ユーザが「好き」または「好まない」とした楽曲を異なる色で可視化する MusiCube を提案した。この研究では、ユーザの好みを空間上の点の色と位置で把握できるが、より細かな好みの分布は明らかではない。これに対して PrefTolerance では、楽曲の評価と許容度に応じて推定される嗜好をグラデーションのように可視化することで、ユーザは好みの強弱や広がりを感じ取る。直感的に把握できる。

2.3 許容度の指定

許容度と似た概念を扱う研究として、Bostandjiev ら [2] は、アーティストとウェブ情報に対して、ユーザがどの程度重視したいかをスライダで入力し、アーティストを推薦する TasteWeights を提案した。Millecamp ら [12] は、楽曲推薦において重要視したい属性をユーザがスライダもしくはレーダーチャートで指定できるインタフェースを提案した。これらの研究は、ユーザが楽曲の好みを判断する際に「歌詞の重要度が高い」といったように、音楽要素自体をどれくらい重視するかの度合いを指定できるが、歌詞を重視しつつも「この歌詞に非常に似た歌詞だけを聴きたい」や「この歌詞とある程度似ている歌詞ならばよい」のような類似度の範囲指定には対応できない。それに対して PrefTolerance では、「音楽要素ごとにどの程度の類似度まで許容できるかの度合い」を表す許容度を用いるため、そうした場合にも対応できる。このように、音楽要素に対する重要度と、類似度に対する許容度は異なる概念である。

3 PrefTolerance

PrefTolerance のインタフェースを図 2 と図 3 に示す。図 2 は、使用開始直後にユーザが 1 曲目に聴く楽曲を選択する画面である。その後、ユーザが楽曲の聴取・評価・許容度の指定を行う図 3 の画面に変わる。これは、(1) 音楽要素ごとの空間、(2) 許容度スライダ、(3) 動画再生プレーヤ、(4) 評価エリア、(5) 探索・推薦タブの 5 つの機能で構成されている。

PrefTolerance は音源を入手可能な歌詞を伴う任意の楽曲が対象だが、今回の実装では、ニコニコ動画¹に投稿された歌声合成楽曲の再生数上位 1 万曲のうち、歌詞の解析が可能であった 9103 曲とした。歌声合成楽曲を利用した理由は、様々なジャンルの楽曲が投稿されていることに加え、Songle Widget [4] を用いた、楽曲のサビ区間 (図 3 の動画再生プレーヤにおける下部のオレンジ色の四角) の試聴再生が容易にできるためである。

¹ <https://www.nicovideo.jp/>



図 2. ユーザが 1 曲目に聴く楽曲を選択する画面 (セレクトボックスを押すと、選択候補の楽曲がプルダウンで表示され、その中から選ぶか、最上部の検索ボックスから楽曲名を入力して選択することができる)。

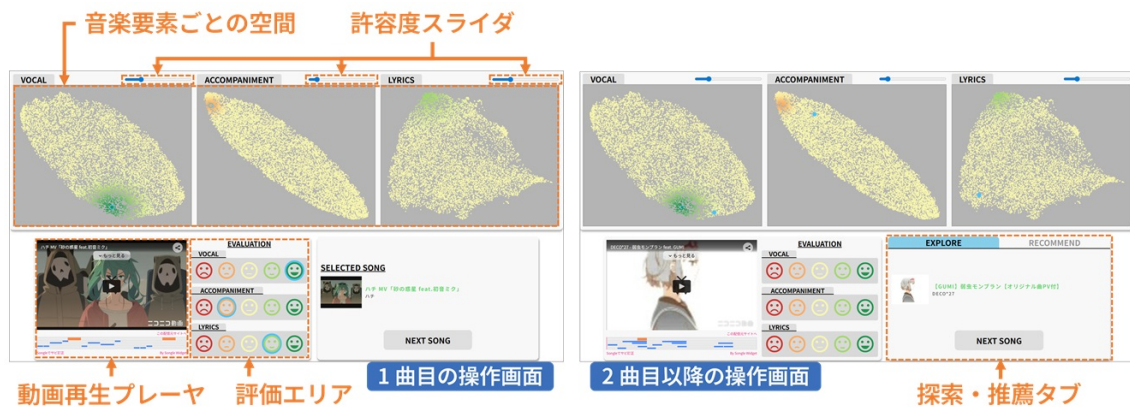


図 3. 選択した楽曲の聴取・評価・許容度の指定を行う画面 (左図が 1 曲目の操作画面、右図が 2 曲目以降の操作画面)。

3.1 操作手順

PrefTolerance の操作手順について述べる。1 曲目に関する操作は以下のとおりである。

- (1) ユーザは 1 曲目に聴く楽曲を選択する (図 2)。
- (2) 「START」ボタンを押すと図 2 から図 4 の 1 曲目の操作画面に切り替わる。
- (3) 選択した楽曲を動画再生プレーヤで聴取し、各音楽要素がどの程度嗜好に合うかを 5 段階で評価する。ただし、ユーザは全要素を評価する必要はなく、評価されていない要素は、中間の値 (黄色) が自動的に割り当てられる。
- (4) 音楽要素ごとの許容度を、許容度スライダからそれぞれ指定する。
- (5) 図 4 の 1 曲目の操作画面の右下にある「NEXT SONG」ボタンを押すことで、図 4 の 2 曲目以降の操作画面に切り替わる。

2 曲目以降の操作は以下のとおりである。

- (6) 以下の 3 種類の楽曲の中から、次に聴取する楽曲をユーザが選択する。
 - (a) 探索タブを押すと、ユーザが次に評価する楽曲がタブ内に 1 曲提示され、動画再生プ

レーヤで再生される。

- (b) 推薦タブを押すと、ユーザの嗜好に合う楽曲が 5 曲タブ内に提示され、その中の 1 曲が動画再生プレーヤで再生される。
- (c) 音楽要素ごとの空間上で気になった点を押すことで、その点に該当する楽曲が動画再生プレーヤで再生される。
- (7) (3) と (4) の操作を行う。
- (8) 図 4 の 2 曲目以降の操作画面の探索タブ内にある「NEXT SONG」ボタンを押すことで、探索・推薦タブで提示される楽曲が更新される。ただし、探索タブで提示される楽曲を評価しなければ、探索・推薦タブで提示される楽曲は更新されない。

以上の (6)～(8) をユーザは繰り返し操作しながら、楽曲の評価と許容度の入力を行う。(8)において「NEXT SONG」ボタンを利用できる条件として、探索タブで提示される楽曲を評価したときに限定した理由は、推薦タブで提示される楽曲が狭い範囲のみに偏った状態でユーザが繰り返し聴取・評価することを防ぐためである。適度に探索タブも利用することで、より適切に嗜好を反映した多様な楽曲

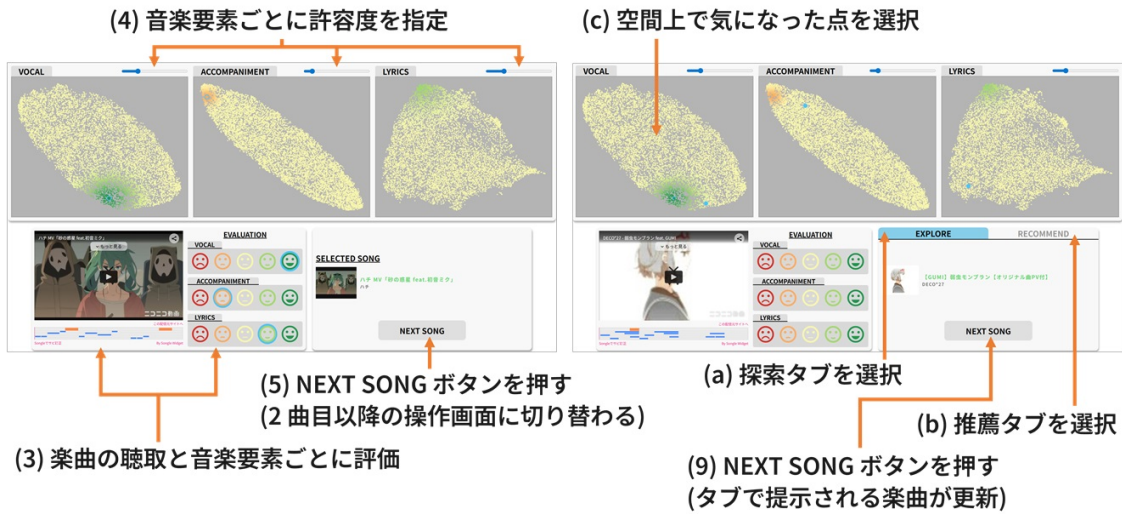


図 4. 1 曲目の操作画面（左）と、2 曲目以降の操作画面（右）における操作手順。

を推薦することが可能となる。

3.2 音楽要素ごとの空間

図 3 の画面上部における、歌声 (VOCAL)・伴奏 (ACCOMPANIMENT)・歌詞 (LYRICS) の 3 つの音楽要素の空間は以下の方法で作成した。歌声と伴奏に関しては、既存の音源分離技術 HT Demucs [16] を用いて各楽曲を歌声と伴奏に分離した。その後、自己教師あり学習で事前学習済みの音楽特徴抽出モデル MERT [7] を用いて、それぞれ 1024 次元のベクトルを取得し、歌声と伴奏のベクトル空間を作成した。歌詞に関しては、事前学習済みの文埋め込みモデル Sentence-BERT [15] を用いて、各楽曲の歌詞から 384 次元のベクトルを取得し、歌詞のベクトル空間を作成した。そして、各ベクトル空間の標準化を行い、既存の非線形次元圧縮技術 UMAP [11] を用いて 2 次元ベクトル空間に変換した後、横軸、縦軸ともに 0 から 1 の範囲を取るように正規化を行うことで各空間上の楽曲の座標値を決定した。UMAP は、t-SNE [24] と比べて、空間の局所的な構造だけでなく、大域的な構造も考慮できるという特長がある。例えば、歌詞の許容度が高いユーザーの場合、歌詞空間上でより広範囲の楽曲まで考慮する必要があるため、大域的な構造も考慮できる特徴を持つ UMAP が有効であると考えて採用した。この可視化により、ユーザは音楽要素ごとに、楽曲同士の類似度を俯瞰して把握できる。

3.3 評価エリア

PrefTolerance の図 3 における評価エリアでは、ユーザは聴取している楽曲の歌声・伴奏・歌詞の各音楽要素がどの程度嗜好に合うかを入力する。評価に関しては「嫌い」、「少し嫌い」、「どちらでもない」、「少し好き」、「好き」に対応した 5 段階のリッカー

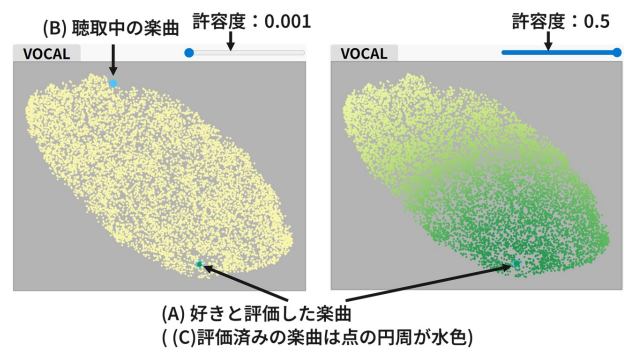


図 5. 歌声空間における評価値の例 (左図が許容度が 0.001 の場合、右図が許容度が 0.5 の場合)。

ト尺度を用いた。3 段階では粗すぎ、7 段階では多数の楽曲を評価する際に負担を感じるので 5 段階とした。ユーザが評価を直感的に入力できるように、フェイス・スケール [10] を用いた。

ユーザがある楽曲のある音楽要素に対して評価すると即座に、図 5 のように、その音楽要素の空間内で、その楽曲に対応する点が評価に応じて着色される。例えば、ある楽曲の歌声に関して「好き」(緑色のとても喜んでいる顔)と入力した際は、歌声の空間上でその楽曲に該当する点の色が緑色(図 5 (A))に変化する。これにより、ユーザは評価した楽曲が各空間上のどの位置にあり、どのような評価となったかを視覚的に確認できる。また、評価前で聴取中の楽曲の点を水色 (B)、評価済みの楽曲は点の円周を水色 (C) で表現する。

3.4 許容度スライダを用いた音楽要素ごとの空間とのインタラクション

ユーザは、図 3 の許容度スライダを左右に動かして各空間に対する許容度を指定できる。各空間にお

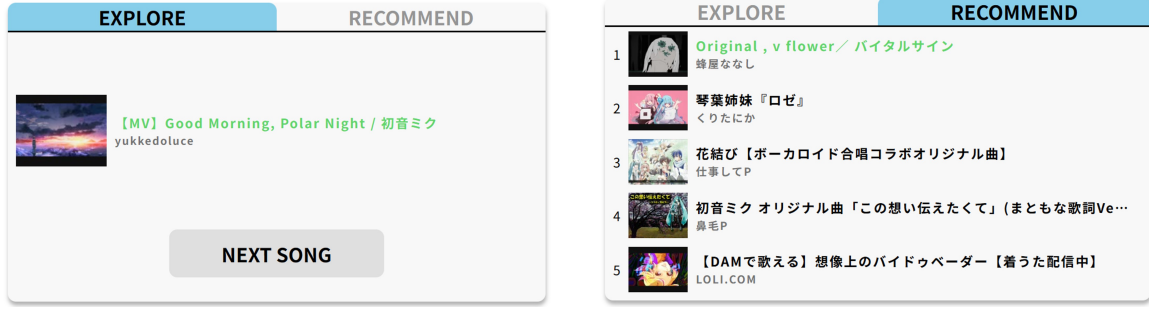


図 6. 探索結果を提示する画面 (左, 探索タブ) と推薦結果を提示する画面 (右, 推薦タブ). タブ内には楽曲のサムネイル画像と楽曲名, クリエイター名が表示され, 選択中の楽曲は楽曲名が緑色で表示される.

ける許容度および評価済み楽曲の位置と評価値に基づき, 未評価楽曲の点の色がグラデーションのように変化する. このグラデーションは, 未評価楽曲の各音楽要素に対する評価の推定値から決められる. 例えばユーザが歌詞を「好き」と評価した楽曲があったときに, 歌詞の空間内ではその楽曲との距離が遠くなるにつれて, 楽曲の歌詞に対する好みの度合いは徐々に下がると考えられる. そのため, 評価済み楽曲の位置と評価値を基に, 各空間における未評価楽曲の評価の推定値は 2 次元ガウス分布に従うと仮定する. ここで, 未評価の楽曲 i , 評価済み楽曲 j , 歌声を V , 伴奏を A , 歌詞を L とし, 音楽要素の集合を $E = \{V, A, L\}$ としたとき, 音楽要素 $e \in E$ における i の評価の推定値 Z_{e_i} を次式とする.

$$Z_{e_i} = \frac{w_{e_j}}{2\pi\sigma_{e_j}^2} \exp\left(-\frac{(x_{e_i}-X_{e_j})^2+(y_{e_i}-Y_{e_j})^2}{2\sigma_{e_j}^2}\right) \quad (1)$$

ここで, w は評価値に対応する重み (「嫌い」は -1 , 「少し嫌い」は -0.5 , 「どちらでもない」は 0 , 「少し好き」は 0.5 , 「好き」は 1), (x, y) は未評価の楽曲の座標 (横軸が x , 縦軸が y), 平均 (X, Y) は評価済み楽曲の座標, 標準偏差 σ は, ユーザから入力された許容度である. 2 次元ガウス分布は共分散行列の非対角成分が 0 で, 対角成分が同一の値 σ^2 とした. 許容度は, $0.001 \sim 0.5$ の範囲を 0.001 刻みで指定可能であり, 初期値は 0.05 である. 許容度 σ が大きい場合, 評価済み楽曲との類似度が低い (距離が遠い) 楽曲に対しての Z_{e_i} が高くなる.

評価済み楽曲が複数ある場合は, 評価済み楽曲からそれぞれ求められた Z_{e_i} の値を加算した合計値を用いた. ただし, 合計値が 1 より大きいときは 1 , -1 より小さいときは -1 の色とした. 例えば, ユーザがある楽曲の歌声に対して「好き」と評価して, 許容度が 0.001 の場合を図 5 の左, 許容度が 0.5 の場合を図 5 の右に示す. これらを比較すると, 実際に評価した楽曲以外に対しての, 評価の推定値が異なる (許容度が大きい右図の方が, より遠くまで影響を与える). この評価値と許容度に応じた Z_{e_i} の変化は, リアルタイムに更新され, ユーザは即座に

その効果を見ることができる. ユーザは, このような空間上の点の色や位置などを参考にしながら, 気になった点 (楽曲) をクリックして楽曲動画を再生したり, その楽曲の音楽要素に対する嗜好を評価したりできる.

3.5 探索・推薦タブ

ユーザがまだ評価していない領域から繰り返し点を選んで評価するのは煩雑なため, もっと効果的に探索できるように「探索」タブ機能を設けた. 一般的に, 楽曲の探索というとユーザが自身の嗜好に合う楽曲を探索することを指すが, 本論文では「インタフェースがユーザに評価してほしい楽曲を探索する」(次に評価すると効果的な楽曲をインタフェースがユーザに提示する)ことを意味する. ユーザが図 3 の「探索」タブを押すと, 自身とその周辺が未評価の楽曲が 1 曲提示される (図 6: 左). そのような楽曲の選定のために, 未評価の楽曲の中から, 評価の推定値 Z_{e_i} の絶対値が 0 に近い (評価済み楽曲から遠い) 楽曲を候補とする. 具体的には, $-\tau \leq Z_{e_i} \leq \tau$ となる楽曲数が最も多い音楽要素を対象とし (本論文では $\tau = 0.3$ とした), その範囲に含まれる楽曲の中からランダムに 1 曲を選択する. このような楽曲に対して評価が得られれば, Z_{e_i} の多くが評価済み楽曲に基づく形に更新され, より的確な推薦が可能となる.

一方, 「推薦」タブを押すと, 音楽要素ごとの嗜好の度合いと許容度に基づいた楽曲が提示される (図 6: 右). 具体的には, 未評価の楽曲 i に対して, Z_{e_i} の平均値 R_i を次式から求める.

$$R_i = \frac{Z_{V_i} + Z_{A_i} + Z_{L_i}}{3} \quad (2)$$

推薦タブ上で R_i が高い順に, 1 位から 5 位の上位 5 曲を推薦楽曲として提示する. 例えば, ユーザが歌詞の許容度が低く, ある楽曲を「好き」と評価した場合, 評価した楽曲との類似度が高い楽曲の Z_{L_i} が高くなり, 推薦で重視されやすい. そのため, R_i を用いることで, 嗜好の度合いが高い音楽要素のう

ち、許容度の低さ（こだわりの強さ）を重視した楽曲推薦を受けることができる。また、提示される楽曲は、1位の楽曲から順に動画再生プレーヤで再生されるが、ユーザが気に入った楽曲をクリックすることでその楽曲から再生することも可能である。

4 ケーススタディ

本章では、PrefToleranceの体験を深めるために、男性の学生2名（参加者A、Bとする）と1対1の対面形式によるケーススタディを行った。参加者はまず、インタフェースの概要と使用方法の説明を受け、操作練習を行った後にインタフェースを使用した。インタフェースの使用にあたっては、以下の2つの条件を設けた。1つ目は、インタフェースの探索・推薦タブで提示される楽曲が更新された際、参加者は推薦された楽曲を1曲以上評価することである。これは、最終的な推薦結果に対する満足度をインタフェースの使用終了時に回答してもらうためである。2つ目は、参加者が1曲目に選択する楽曲は、既に聴取したことがあり、音楽要素のいずれかについて「好き」または「少し好き」と評価できる楽曲とすることである。これは、参加者の嗜好を反映して使用を開始してもらうためである。また、楽曲の再生位置については指定せず、参加者が楽曲の先頭やサビなど任意の位置から再生して聴取できるようにした。インタフェースの使用時間は最大30分とし、参加者が満足する推薦結果を得られた時点で使用を終了できるようにした。インタフェースの使用後、参加者はアンケートに回答した。アンケートは、音楽要素ごとの評価に対する満足度および最終的な推薦結果に対する満足度のそれぞれに対して7段階（1～7で大きいほど満足度が高い）で回答する項目と、インタフェースを使用した感想を述べる自由記述欄の構成とした。

ケーススタディを行った結果、インタフェースの使用時間は参加者Aは26分、参加者Bは30分だった。評価した楽曲数は、参加者Aは18曲、参加者Bは20曲だった。音楽要素ごとの評価に対する満足度は、参加者Aは6、参加者Bは5と高かった。最終的な推薦結果に対する満足度も参加者Aは6、参加者Bは4と高かった。自由記述欄では参加者Aから「最終的な推薦結果の上位3曲は自分の好みに合っていた」という回答が得られた。

参加者Aがインタフェースの使用を終えたときのそれぞれ空間を図7に示す。図7について、それぞれの空間上の点の色で「好き」に該当する緑色の点の分布が多く見られたが、歌声空間と歌詞空間では「少し嫌い」に該当する点の分布が見られた。参加者Aのインタフェースの使用を終えるまでの許容度の推移を図8に示す。図8について、3曲目までにはいずれかの音楽要素で許容度の変化が起きていた。例えば、歌詞の許容度は0.050から0.213と変化して

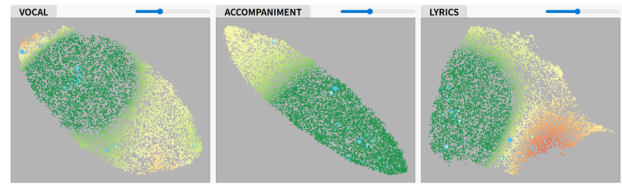


図 7. 参加者 A が PrefTolerance の使用を終えたときのそれぞれの空間。

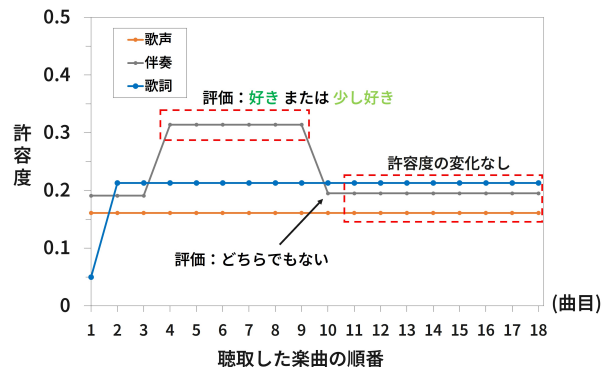


図 8. 参加者 A における許容度の推移。

いた。10曲目の伴奏の許容度は、0.314から0.195に低下しており、このときの伴奏の評価は4曲目から9曲目まで「好き」または「少し好き」の評価で、10曲目は「どちらでもない」であった。11曲目以降の許容度は、各音楽要素で変化していなかった。

5 おわりに

本論文の主な貢献を以下に述べる。

- ユーザが歌声・伴奏・歌詞の音楽要素ごとの嗜好と、各音楽要素でどの程度の類似度までを嗜好に合う範囲として許容できるかの度合いの「許容度」を指定できる楽曲推薦インタフェース PrefTolerance を提案した。
- ユーザが満足する許容度を指定可能にするため、音楽要素ごとの空間の可視化と、各空間における許容度および評価済み楽曲の位置と評価値に応じて、好みに合う楽曲の範囲がどのように変化するかを点の色のグラデーションで表現するインタラクションを実現した。
- ユーザの音楽要素ごとのこだわりを重視することができる推薦機能と、効率よくユーザの嗜好を反映することができる探索機能を実現した。

今後は、ユーザの許容度を自動的に推定する機能を設けることで、許容度の指定が得意でないユーザでも扱えるインタフェースの実現を目指していきたい。また、許容度を用いない手法との比較実験を行い、PrefToleranceの有効性を調査していきたい。

謝辞

本研究の一部は JST CREST JPMJCR20D4 と JSPS 科研費 JP25H01174 の支援を受けた。

参考文献

- [1] B. Amiri, N. Shahverdi, A. Haddadi, and Y. Ghahremani. Beyond the Trends: Evolution and Future Directions in Music Recommender Systems Research. *IEEE Access*, 12:51500–51522, 2024.
- [2] S. Bostandjiev, J. O'Donovan, and T. Höllerer. TasteWeights: a visual interactive hybrid recommender system. In *Proceedings of the 6th ACM Conference on Recommender Systems (ACM RecSys 2012)*, pp. 35–42, 2012.
- [3] Y. Deldjoo, M. Schedl, and P. Knees. Content-driven music recommendation: Evolution, state of the art, and challenges. *Computer Science Review*, 51:100618, 2024.
- [4] M. Goto, K. Yoshii, and T. Nakano. Songle Widget: Making animation and physical devices synchronized with music videos on the web. In *Proceedings of the 2015 IEEE International Symposium on Multimedia (IEEE ISM 2015)*, pp. 85–88, 2015.
- [5] M. Hamasaki, M. Goto, and T. Nakano. Songrium: Browsing and listening environment for music content creation community. In *Proceedings of the 12th Sound and Music Computing Conference (SMC 2015)*, pp. 23–30, 2015.
- [6] D. Kostrzewa, J. Chrobak, and R. Brzeski. Attributes Relevance in Content-Based Music Recommendation System. *Applied Sciences*, 14(2), 2024.
- [7] Y. Li, R. Yuan, G. Zhang, Y. Ma, X. Chen, H. Yin, C. Xiao, C. Lin, A. Ragni, E. Benetos, N. Gyenge, R. Dannenberg, R. Liu, W. Chen, G. Xia, Y. Shi, W. Huang, Z. Wang, Y. Guo, and J. Fu. MERT: Acoustic Music Understanding Model with Large-Scale Self-supervised Training. In *The Twelfth International Conference on Learning Representations (ICLR 2024)*, 2024.
- [8] X. Liu, Z. Yang, and J. Cheng. Music recommendation algorithms based on knowledge graph and multi-task feature learning. *Scientific Reports*, 14(1):2055, 2024.
- [9] B. Logan. Music Recommendation from Song Sets. In *Proceedings of the 5th International Society for Music Information Retrieval Conference (ISMIR 2004)*, pp. 425–428, 2004.
- [10] C. D. Lorish and R. Maisiak. The face scale: A brief, nonverbal method for assessing patient mood. *Arthritis & Rheumatism*, 29(7):906–909, 1986.
- [11] L. McInnes, J. Healy, and J. Melville. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv preprint arXiv:1802.03426*, 2018.
- [12] M. Millecamp, N. N. Htun, Y. Jin, and K. Verbert. Controlling Spotify Recommendations: Effects of Personal Characteristics on Music Recommender User Interfaces. In *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization (ACM UMAP 2018)*, pp. 101–109, 2018.
- [13] T. Nakano, K. Yoshii, and M. Goto. Musical Similarity and Commonness Estimation Based on Probabilistic Generative Models of Musical Elements. *International Journal of Semantic Computing (IJSC)*, 10(01):27–52, 2016.
- [14] M. Pulis and J. Bajada. Siamese Neural Networks for Content-based Cold-Start Music Recommendation. In *Proceedings of the 15th ACM Conference on Recommender Systems (ACM RecSys 2021)*, pp. 719–723, 2021.
- [15] N. Reimers and I. Gurevych. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP 2019)*, pp. 3982–3992, 2019.
- [16] S. Rouard, F. Massa, and A. Défossez. Hybrid Transformers for Music Source Separation. In *Proceedings of the 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (IEEE ICASSP 2023)*, 2023.
- [17] Y. Saito and T. Itoh. MusiCube: A Visual Music Recommendation System featuring Interactive Evolutionary Computing. In *Proceedings of the 2011 Visual Information Communication-International Symposium (ACM VINCI)*, pp. 1–6, 2011.
- [18] D. Sánchez-Moreno, A. B. G. González, M. D. M. Vicente, V. F. L. Batista, and M. N. M. García. A collaborative filtering method for music recommendation using playing coefficients for artists and users. *Expert Systems with Applications*, 66:234–244, 2016.
- [19] S. Sasaki, K. Yoshii, T. Nakano, M. Goto, and S. Morishima. LyricsRadar: A Lyrics Retrieval System Based on Latent Topics of Lyrics. In *Proceedings of the 15th International Society for Music Information Retrieval Conference (ISMIR 2014)*, pp. 585–590, 2014.
- [20] M. Schedl, P. Knees, B. McFee, and D. Boganov. Music Recommendation Systems: Techniques, Use Cases, and Challenges. In *Recommender Systems Handbook*, pp. 927–971. Springer, 2022.
- [21] K. Tsukuda, T. Nakano, M. Hamasaki, and M. Goto. Unveiling the Impact of Musical Factors in Judging a Song on First Listen: Insights From a User Survey. In *Proceedings of the 24th International Society for Music Information Retrieval Conference (ISMIR 2023)*, pp. 561–570, 2023.
- [22] B. Vad, D. Boland, J. Williamson, R. Murray-Smith, and P. B. Steffensen. Design and Evaluation of a Probabilistic Music Projection In-

- terface. In *Proceedings of the 16th International Society for Music Information Retrieval Conference (ISMIR 2015)*, pp. 134–140, 2015.
- [23] A. van den Oord, S. Dieleman, and B. Schrauwen. Deep content-based music recommendation. In *Proceedings of the 26th International Conference on Neural Information Processing Systems (NIPS 2013)*, pp. 2643–2651, 2013.
- [24] L. Van der Maaten and G. Hinton. Visualizing Data using t-SNE. *Journal of machine learning research (JMLR)*, 9(11):2579–2605, 2008.
- [25] 奥 健太. 楽曲推薦システム—プレイリスト, コンテキスト, インタクション—. 人工知能学会誌, 34(3):300–308, 2019.
- [26] 吉井 和佳, 後藤 真孝. 音楽情報処理技術の最前線: 7. 音楽推薦システム. 情報処理学会, 50(8):751–755, 2009.

未来ビジョン

本研究が目指す未来は大きく2つある。

まず、音楽推薦において「許容度」という新たな概念を人々に浸透させていくことで、単に「好き」や「嫌い」といった評価を伝え合うだけでなく、音楽要素ごとの嗜好と許容度を可視化した2次元空間を見せ合う形で、互いの好みや許容度を直感的に共有するようなコミュニケーションが可能になることを期待している。この空間を互いに見せ合うことにより、「あなたは歌声に対してこだわりが強い」、「あなたと私で伴奏の好みが全然違う」、「あなたと比べて私は歌詞に対して全然こだわっていない」といった違いを感覚的に理解しやすくなる。さらに、互いの嗜好や許容度を比較・探索するようなインタラクションも可能になるので、例えば自分と近い聴き方をする人を見つけてその人のイチ推しの楽曲を聴かせてもらうような

音楽との出会いも実現できる。

次に、許容度の概念は音楽に限らず様々な可能性を持っており、他のメディアコンテンツや日常生活にも応用されていくことを期待している。例えばメディアコンテンツでは、アニメやドラマなどの映像作品の推薦システムを実現する際にも、物語や映像表現、演出のトーンなどの特徴に対して、ユーザが自分の嗜好と許容度を設定できるような機能が考えられる。日常生活でも、「食の辛さや甘さなどに対する嗜好と許容度」や「服の色や素材、デザインなどに対する嗜好と許容度」など、多様な場面で好みと許容度を探索・可視化できるような発展が考えられる。

このように本研究をきっかけに「許容度」の概念が広まることで、音楽のみならず日常生活の様々な体験において、嗜好や許容度が可視化されて活用されやすくなり、相互理解がより円滑になる未来が切り拓ける。

片眼のみに表示されるポインタを用いた xR 向け選択手法

鈴木 悠真* 折居 篤* 志築 文太郎†

概要. xR におけるオブジェクト選択手法に関する課題として、両眼視差による問題がある。両眼視差とは、左右の眼の位置あるいは網膜像の違いを表す。これにより、奥行きが異なる複数の物体を見たときに片方の物体の像が二重に見える現象が引き起こされ、オブジェクトの選択が困難になる。この問題を生じない手法として、本研究は、片眼のみに表示されるポインタを用いた xR 向け選択手法を実装した。本手法の性能を評価するために、両眼に表示されるポインタを用いた従来手法に対する、Fitts の法則に基づく比較実験を行った。その結果、本手法は従来手法と比較して、選択のための移動時間、エラー率、スループット、作業負荷、およびユーザビリティについてより優れた性能をもつことが示された。また、ポインタを表示する眼の決定には、利き眼だけでなくユーザの好みを考慮する必要があることが示唆された。

1 はじめに

xR において、ユーザから離れた位置に存在する仮想オブジェクトを選択する操作は、不可欠なもののひとつである [5]。これについて、ポインタをオブジェクトにかざすことにより選択を行う手法 [9, 11] が研究されている。この手法は、xR において一般的に用いられるレイによる手法と比較してユーザの手の震えの影響を受けにくいと、特に奥行きが深く、小さいオブジェクトを選択する際に、正確な選択を行えるという特徴がある [11]。

しかしこの手法は、両眼視差による問題を生じうる。両眼視差とは、左右の眼の位置の違いにより生じる左右の網膜像の違いを表す。これにより人間は物体の奥行きを知覚できる。しかし、奥行きが異なる複数の物体を見たときには、片方の物体のみに焦点が合い、もう片方の物体の像が二重に見える現象が生じる。特に奥行きが深く、小さいオブジェクトに対してユーザが選択を行う際には、ポインタあるいはオブジェクトのいずれかが二重に見えることにより、選択が困難になる [8]。両眼視差による問題は人間の左右の眼の焦点が特定の 1 つの位置以外に合わないことにより引き起こされる。そのためこの問題の解決のために、片眼のみを用いてオブジェクトを見る工夫が考えられる。しかしこの手法では、xR の特色である立体視を行えない。

本研究の目的は、立体視を妨げないままに、両眼視差による問題を生じない xR 向け選択手法を実現することである。このために本研究は、視野闘争と呼ばれる人間の視覚に関する現象を利用する。視野闘争とは、左右の眼において異なる映像を見た時に、それぞれの眼に対応する映像が交わって認識される現象である。特に、左右の眼に映る映像のうち異なる

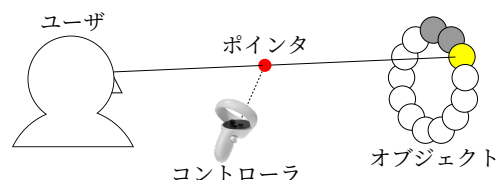


図 1. 本手法における選択の様子。ユーザはポインタをオブジェクトにかざすことにより選択を行う。ポインタはコントローラに付随しており、片眼のみに表示される。

部分が小さい、コントラストが高い、あるいは動く物体である場合、その部分は反対の眼に映る映像に抑制されることなく見え続ける傾向がある。また、立体視および視野闘争は異なる経路によって制御されるため、視野闘争は立体視による奥行き知覚を妨げない [13]。この現象に基づき、本研究は、立体視を妨げないままに両眼視差による問題を生じない、片眼のみに表示されるポインタを用いた xR 向け選択手法を示す (図 1)。また、本手法を実装した後、その性能を評価するための実験を行う。

2 関連研究

本節では、本研究に関連する研究として xR におけるオブジェクト選択手法および左右の眼に異なる映像を提示する xR 手法に関する研究を示す。また、それらに対する本研究の位置づけを述べる。

2.1 xR におけるオブジェクト選択手法

ポインタをかざすことによる選択手法および両眼視差による問題を解決するための選択手法に関する研究を示す。

Copyright is held by the author(s).

* 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

2.1.1 ポインタをかざすことによる選択手法

Diako ら [9] は VR において「EyeSeeThrough」という手法を示した。この手法においてユーザは、オブジェクトに対して行う操作が並んで表示された半透明メニューをオブジェクトにかざした後、視線を用いてオブジェクトを選択する。この手法はオブジェクトおよび行う操作を同時に選択することで、高速かつ直感的なオブジェクト操作を実現した。

Uta ら [12] は 3D 環境において「Gaze-Hand Alignment」という手法を用いた。この手法においてユーザは、ターゲットに対して視線を合わせた後に手を用いてポインティングを行い、この2つを一致させることによりオブジェクトを選択した。この手法は両眼視差による問題から、ターゲットの奥行きが増すほどポインティング性能が低下することが示された。

Yue ら [10] は AR においてターゲットをスマートフォンによって一人称視点的に遮蔽することによって選択する選択手法を示した。実装されたプロトタイプシステムでは、フロントカメラからユーザの虹彩の位置を識別し、同時にリアカメラからスマートフォンによって遮られた矩形領域を推定した。

本研究は、これらの研究と同様に、ユーザがポインタを選択の対象とするオブジェクトにかざすことによりオブジェクトを選択する手法を示す。

2.1.2 両眼視差による問題を解決するための選択手法

両眼視差による問題に対して、これまで様々なアプローチが行われてきた。

Ferran[1] らは立体ディスプレイ環境において、カメラのファインダのような効果を用いた選択手法を示した。ユーザが選択を行う方向に対して、その奥にあるオブジェクトを平面的に描画する仮想スクリーンを表示し、それを通して選択を行った。これにより、選択における両眼視差による曖昧性を回避した。

Joon[8] らは VR においてスマートフォンを用いた選択手法を示した。ユーザは透明なウィンドウとして描画されたスマートフォンをオブジェクトにかざし、その画面上にてタップすることによりオブジェクトを選択する。この際、選択操作中にのみオブジェクトを平面的に描画することにより、両眼視差による問題を生じない選択手法を実現した。

これらの手法は、両眼視差による問題を生じないために、特定の領域においてオブジェクトを平面的に描画するという方法をとった。しかし、この方法は立体視を妨げるとともに、選択の度に平面的なレンダリングへ切り替えるための操作およびそれを行うための時間を必要とする。

Adrian[11] らは VR においてユーザの頭部に固定された平面内を移動するカーソルを用いた選択手法を示した。ユーザは外部入力デバイスを用いてカーソルを操作し、2D デスクトップのように選択を行う

ことができた。このとき、オブジェクトの大きさや位置をオブジェクトの奥行きに合わせて動的に変化させることにより、両眼視差による問題に対処した。

本研究は、選択に用いるポインタを片眼のみに表示することにより、両眼視差による問題を生じない手法を示す。

2.2 左右の眼に異なる映像を提示する xR 手法

Colin[13] らは 3D ステレオグラフィック環境において、片眼のみに提示された 2D カーソルを用いた選択手法を示した。実験において参加者は、2D 選択においてはカーソルをターゲットの上に置くように、3D 選択においてはカーソルをターゲットの内部に合わせるように指示された。その結果、2D 選択は 3D 選択より有意に高速であり、またエラー率も有意に低かった。

郭ら [16] は xR において視野闘争を用いたユーザインタフェース提示手法を示した。画面の重畳による情報の欠落を防ぐため、ユーザの片眼のみに不透明なユーザインタフェース画面を提示した。その結果、半透明ユーザインタフェースと比べ視覚的印象が低下した一方、ビジュアルシーンの変化への認識精度と認識の確信度が向上することが示された。

山本ら [17] は VR において左右それぞれの眼に異なる第一人称視点を提示する、複数環境の同時体験インタラクション手法を示した。この手法においてユーザは、VR 酔いを誘発することなく、異なる2つの視点を同時に知覚できた。

現在、左右の眼に異なる映像を提示する手法を xR 向け選択手法に用いた研究は行われていない。

3 片眼のみに表示されるポインタを用いた xR 向け選択手法

本手法においてユーザは、利き手用コントローラに付随して表示されるポインタ（赤色の球体、コントローラの中心から前方の 0.2m の位置、大きさは奥行きに関わらず常に 0.5°、片眼のみに表示される）をオブジェクトにかざした状態において、コントローラの人差し指トリガボタンを押すことにより選択を行う。このとき、ユーザの眼のうちポインタが表示されている眼の位置を基準点として、そこからポインタの中心を結んだ直線上にあるオブジェクトのうち、ユーザに最も近いオブジェクトが選択される（図2左）。ポインタは片眼のみに表示されるため、ユーザはポインタに焦点を合わせる必要がない。したがって、両眼視差による問題が生じない。また、ポインタを除く他のオブジェクトは通常通り両眼に表示される。したがって、ユーザは選択の対象とするオブジェクトを両眼によって視認できるため、立体視が可能である。

なお本研究では、ポインタを表示する眼として利

片眼のみに表示されるポインタを用いた xR 向け選択手法

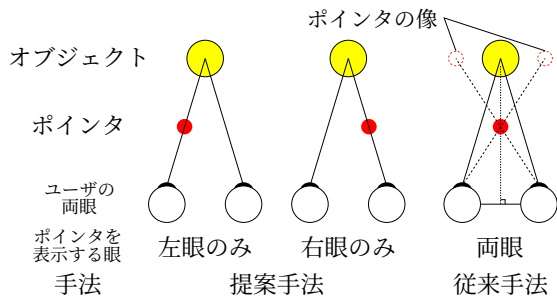


図 2. 各手法における選択の時のユーザーの両眼、ポインタ、およびオブジェクトの位置関係。特に従来手法においては、両眼視差によりオブジェクトに焦点を合わせた時ポインタの像が二重に見える。

き眼とする (Dominant) のが良いか、あるいは非利き眼とする (NonDominant) のが良いかについても併せて調査する。

4 参加者実験

本手法を評価するための参加者実験を行った。

4.1 参加者

参加者は研究室外の大学生および大学院生 12 名 (男性 11 名, 女性 1 名, 平均年齢 22.7 歳, $SD = 0.99$, 全員右利き) である。VR・AR 機器の使用頻度を、5 段階評価 (5: 日常的に使う, 4: 頻度は低いが定期的に使う, 3: 不定期で使うことがある, 2: 使ったことがある, 1: なし, の 5 段階) にて集計した。その結果, 2 名が 4, 3 名が 3, 6 名が 2, 1 名が 1 と回答し, 5 と回答した参加者はいなかった。また利き眼について Miles 法に基づく検査法 [19] による検査を行ったところ, 11 名が右眼, 1 名が左眼と判別された。

4.2 実験装置および環境

参加者は実験用アプリケーションがインストールされたヘッドマウントディスプレイ (以下, HMD) を頭部に装着した。HMD として, Meta Quest 2 を用いた。また, 利き手 (参加者が両利きである場合には右手) にコントローラを把持した。反対側の手は操作に使用しなかった。

実験のためのアプリケーションを実装した (図 3)。実装環境は Unity 2022.3.22f1 である。実験アプリケーションにおける配色は Adrian ら [11] による実験を, 各色のカラーコードはカラーユニバーサルデザイン推奨配色セット [15] をそれぞれ参考とした。アプリケーションが開始されると, 正面方向奥行き 1.5 m の位置に, 直径 0.1 m の青色の球体であるスタートターゲットが表示された。ユーザがこのターゲットを選択することにより, タスクが開始された。

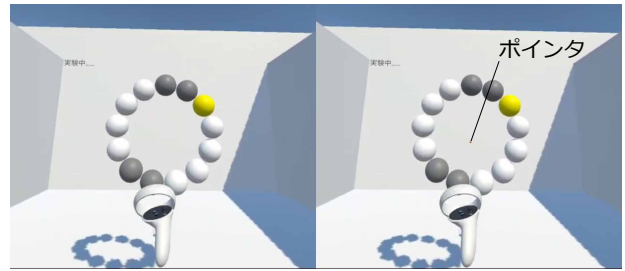


図 3. 実験用アプリケーションにおけるユーザーの視界。左眼に表示される映像 (左側) と右眼に表示される映像 (右側) が異なり, ポインタ (赤色の球体) は片眼のみに表示されている。(この図はイメージであり, 左側の画像には編集が加えられている)



図 4. 実験における参加者の様子。参加者は頭部に HMD を装着し, 利き手 (右手) に把持したコントローラを用いて選択を行った。

タスクが開始されるとスタートターゲットは消えた。タスクがひとつ終了すると再びスタートターゲットが表示された。これを選択することによりユーザは, 即座にセッション内の次のタスクを開始できた。

参加者は, 参加者の腕の動きを阻害しない程度に十分広い空間において, 椅子に座った状態にて実験に参加した (図 4)。実験中参加者は移動しないよう求められたが, 頭部の動きは制限されなかった。

4.3 タスク

実験に用いたタスクは, Fitts の法則に基づくタスク [6, 5, 12, 4] であった。

各条件において, タスクが開始されると参加者の正面に円形に配置された 13 個のターゲットが表示された。タスクの間, 参加者が現在選択すべきターゲットは黄色, まだ選択していないターゲットは白色, 既に選択したターゲットは灰色にて表示された。参加者が選択すべきターゲットにポインタをかざした状態にてボタンを押した場合, 選択に成功し, 次に選択すべきターゲットが新たに黄色に表示された。選択すべきターゲットの外側にポインタをかざした状態にてボタンを押した場合, 選択に失敗し, 現在選択すべきターゲットが赤色に点滅することにより強

調された。選択に成功するまで現在選択すべきターゲットは変化しなかった。

参加者は異なるターゲットの大きさおよびターゲット間の距離ごとに6条件×13ターゲット=78回の選択を行った。これを1タスクとして、ターゲットの奥行きごとに計3回のタスクを行った。これを1セッションとして、異なる手法ごとに計3回のセッションを行った。実験全体において参加者は、3タスク×3セッション×78回=702回の選択を行った。各タスクにおいて参加者は、あるターゲットの大きさについてターゲット間の距離の近い順に選択を行い、これをターゲットの大きさの大きい順に行った。このタスクを、各セッションにおいてターゲットの奥行きの小さい順に行った。セッションにおける手法の順番はラテン方格法により参加者間において均等になるよう設定された。

4.4 実験設計

本手法を評価するための比較対象として、従来手法である両眼に表示されるポインタを用いた選択手法 (Both) を用いた。この手法において選択されるオブジェクトを決めるための基準点としては、ユーザの両眼の中央の位置を用いた (図2右)。

本実験における独立変数を以下に示す。各値の基準はユーザがHMDを装着した時のユーザの頭部位置とした。

- 手法: Both, Dominant, NonDominant
- ターゲットの奥行き (D): 0.5 m, 1.0 m, 2.0 m
- ターゲットの大きさ (W): 6°, 4°, 2°
- ターゲット間の振幅 (A): 25°, 50°

また、従属変数には、移動時間 (以下, MT)、エラー率 (以下, ER)、および性能を評価するための指標として Throughput (以下, TP) を用いた。また、作業負荷およびユーザビリティを評価するための指標として、NASA-Task Load Index (以下, NASA-TLX) [3, 18] および System Usability Scale (以下, SUS) [2] を用いた。さらに、3つの手法のうちどの手法がより好みであったかについて、参加者に順位付けを求めた。併せて、記述式のアンケートにより各手法に対するコメントを求めた。

MT は、現在選択すべきターゲットが赤色に表示されてからそのターゲットが正しく選択されるまでの時間である。 ER はすべての選択に対する、参加者が選択すべきターゲットの外側にポインタをかざした状態にてボタンを押すことにより失敗とみなされた選択の割合である。 TP の算出には式1を用いた。

$$TP = \frac{ID}{MT}, \quad ID = \log_2\left(\frac{A}{W} + 1\right) \quad (1)$$

4.5 実験手順

参加者はまず本研究に関する説明を受け、実験前アンケートに回答した。次にHMDを装着し、楽な

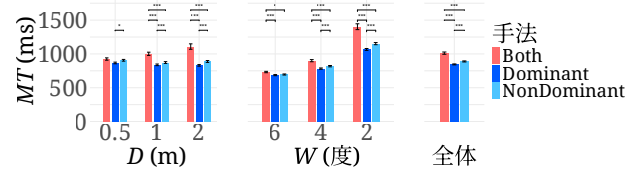


図 5. 各条件および手法における MT の平均値。エラーバーは標準偏差を示す。(*: $p < 0.05$, ***: $p < 0.001$)

姿勢をとって椅子に座った。その状態において、その時の頭部の位置がシステムにおける基準位置となるよう、実験用アプリケーションがリセットされた。次に参加者は実験用アプリケーションの操作について説明を受けながら、練習タスクを行った。練習タスクでは、 $D = 1.0$ m, $W = 4^\circ$, $A = 25^\circ$ の1条件について、13回の選択を行った。その後、参加者はなるべく速く、かつなるべく正確にタスクを行うよう指示された。各タスクは、スタートターゲットを選択することにより、参加者によって開始された。各タスクが終了したとき、参加者はHMDを装着したままその場で休憩をとることができた。本実験はゴリラアーム現象 [7] を引き起こす腕への負荷をもたらすため、参加者は実験前の説明にて、各タスクが終了したとき毎回1分程度の休憩をとるよう促された。参加者が休憩を十分取ったと判断したときに、参加者によってタスクは再開された。各セッションが終了したとき、参加者はその条件に関してNASA-TLX および SUS に基づくアンケートに回答した。すべてのセッションが終了したとき、参加者は記述式の実験後アンケートに回答した。またその内容についてインタビューを受けた。実験は最大で60分の時間を要した。

4.6 実験結果

MT , ER , TP のそれぞれについて、アンダーソン-ダーリング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、 D および W の3要因を独立変数として整列ランク変換 [14] を行った後、三元配置分散分析を行った。その後、下位検定としてフリードマン検定、およびウィルコクソンの符号順位検定による多重比較を行った。この際 p 値の補正にはホルムの方法を用いた。

4.6.1 MT

各条件および手法における MT を図5に示す。手法 ($F_{2,22} = 8.30$, $p < 0.001$)、 D ($F_{2,22} = 9.17$, $p < 0.01$)、 W ($F_{2,22} = 297.43$, $p < 0.001$) において主効果が認められた。また、手法および D ($F_{4,44} = 9.04$, $p < 0.001$)、手法および W ($F_{4,44} = 11.48$, $p < 0.001$)、 D および W ($F_{4,44} = 13.66$, $p < 0.001$) の各2要因間、および手法、 D および W の3要因間

片眼のみに表示されるポインタを用いた xR 向け選択手法

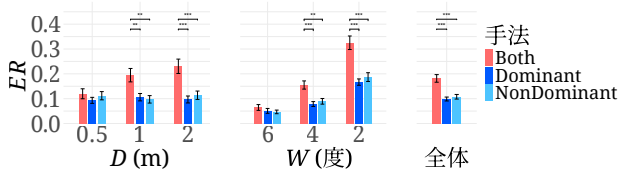


図 6. 各条件および手法における ER の平均値。エラーバーは標準誤差を示す。(**: $p < 0.01$, ***: $p < 0.001$)

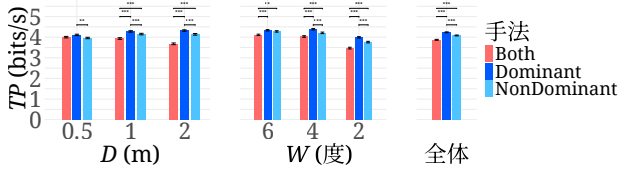


図 7. 各条件および手法における TP の平均値。エラーバーは標準誤差を示す。(**: $p < 0.01$, ***: $p < 0.001$)

($F_{8,88} = 3.50$, $p < 0.01$) で交互作用が認められた。各 D および W では、 $D = 1.0$ m, $D = 2.0$ m, $W = 4^\circ$, $W = 2^\circ$ においてすべての手法間で有意差が見られ、Both が最も遅く Dominant が最も高速であった。

4.6.2 ER

各条件および手法における ER を図 6 に示す。手法 ($F_{2,22} = 13.98$, $p < 0.001$), D ($F_{2,22} = 4.79$, $p < 0.05$), W ($F_{2,22} = 149.02$, $p < 0.001$) において主効果が認められた。また、手法および D ($F_{4,44} = 7.95$, $p < 0.001$), 手法および W ($F_{4,44} = 13.94$, $p < 0.001$), D および W ($F_{4,44} = 5.12$, $p < 0.01$) の各 2 要因間、および手法、 D および W の 3 要因間 ($F_{8,88} = 2.28$, $p < 0.05$) で交互作用が認められた。

各 D および W では、 $D = 1.0$ m, $D = 2.0$ m, $W = 4^\circ$, $W = 2^\circ$ において Both および Dominant 間と Both および NonDominant 間で有意差が見られ、Both が最もエラーが多く Dominant が最も正確であった。

4.6.3 TP

各条件および手法における TP を図 7 に示す。手法 ($F_{2,22} = 3.47$, $p < 0.05$), D ($F_{2,22} = 3.51$, $p < 0.05$), W ($F_{2,22} = 15.92$, $p < 0.001$) において主効果が認められた。また、手法および D ($F_{4,44} = 3.96$, $p < 0.01$), D および W ($F_{4,44} = 10.35$, $p < 0.001$) の各 2 要因間で交互作用が認められた。

各 D および W では、 $D = 1.0$ m, $D = 2.0$ m, $W = 4^\circ$, $W = 2^\circ$ においてすべての手法間で有意差が見られ、Both が最も性能が低く Dominant が最も性能が高かった。

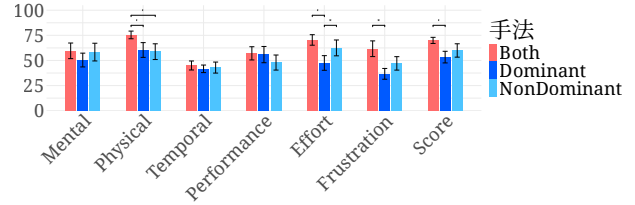


図 8. 各手法における NASA-TLX スコアの平均値。エラーバーは標準誤差を示す。(*: $p < 0.05$)

4.6.4 NASA-TLX

各手法のスコアの平均値はそれぞれ、Both は 69.97 ($SD = 10.59$), Dominant は 53.33 ($SD = 20.06$), NonDominant は 59.94 ($SD = 23.07$) であった (図 8)。このうち Both-Dominant 間には有意な差が見られた。各項目のスコアについては、Physical における Both-Dominant 間および Both-NonDominant 間、Effort における Both-Dominant 間および Both-NonDominant 間、Frustration における Both-Dominant 間で有意差が見られた。

4.6.5 SUS

各手法のスコアの平均値はそれぞれ、Both は 47.29 ($SD = 22.57$), Dominant は 82.50 ($SD = 13.10$), NonDominant は 74.79 ($SD = 17.07$) であった。このうち Both-Dominant 間 ($p < 0.01$) および Both-NonDominant 間 ($p < 0.05$) に有意な差が見られ、Dominant および NonDominant が高いユーザビリティを持つと評価された。

4.6.6 好み

各手法の好みに関する平均順位は、Both が 2.92 位 (1 位: 0 人, 2 位: 1 人, 3 位: 11 人), Dominant が 1.25 位 (1 位: 9 人, 2 位: 3 人, 3 位: 0 人), NonDominant が 1.92 位 (1 位: 3 人, 2 位: 8 人, 3 位: 1 人) であった。したがって、Dominant が最も好まれ、また Both が最も好まれなかった。

4.6.7 アンケート

Both について、参加者はポインタが「2 つに見える (P1, P2, P5, P6, P8, P10, P11, P12)」と感じた。ターゲットを選択するためにそれら 2 つの像の中央部分に合わせて狙う必要があり (P1, P2, P5, P6, P8, P11, P12), ターゲットの選択が難しいと感じた。あるいは、ポインタが「実際の位置と異なる (P3, P9, P11)」と感じたり、ポインタをターゲットに合わせても選択できない (P4, P5, P9) と感じたりした。

本手法について、参加者はポインタが「2 つに分かれない (P2, P3, P6, P7, P9, P12)」と感じた。このため「直感的 (P1, P10)」で「選択しやすい

(P2, P3, P5, P9, P12)」と感じ、また「眼への負荷が少ない (P5)」と感じた。左右の目に異なる映像が提示されることには、多くの参加者は違和感を覚えなかった (P1, P2, P4, P11, P12)。しかし同時に慣れが必要 (P4, P12) だと感じたりポイントが見えづらい (P3, P8) と感じた参加者もいた。

本手法の中でも NonDominant については、いくつかの参加者は違和感を覚えた。ポイントがずれていた (P4, P7, P10) と感じたり、「腕がふわっとするよう (P1)」に感じたり、「常に VR 酔いの少し手前のような感覚がした (P5)」参加者もいた。ポイントが「やや薄く (P9)」なったと感じた参加者や、「消えて見えることがあった (P6)」と回答した参加者もいた。しかし、Dominant と同じ (P2, P11, P12) であったと感じた参加者や、NonDominant のほうがポイントが「はっきり映っていて」操作しやすかった (P3, P8) と感じた参加者もいた。

4.7 考察

図 5, 図 6, 図 7 より、本手法は従来手法と比べて、奥行きが深く小さいターゲットを選択する際に有意に高速かつ正確に操作を行うことができ、また高いスループットを実現することが示された。これは、従来手法において参加者は両眼視差による問題によりターゲットに正確に狙いを定めることができずターゲットの選択に苦労したことに対して、本手法においてはその問題が発生せず容易に選択を行うことができたためと考えられる。

また、Dominant と NonDominant の間では、Dominant のほうが有意に高速であり、かつ高いスループットを実現することが示された。これは、NonDominant における問題が Dominant において発生しなかった参加者が多かったためと考えられる。

図 8 より、本手法は従来手法と比べて、Physical, Effort, および Frustration に関して作業負荷が低いことが示された。これは、Both においてポイントが「ぶれて 2 つに見え」ることによる選択時間の増加が腕への負荷に繋がったり、意図しない操作に繋がったりしたことが理由であると考えられる。また、Dominant は他の手法と比べて、有意に優れたユーザビリティをもつことが示された。

本手法について、アンケートにてより好みであると回答した手法と MT の平均がより速い手法が異なる参加者がいた (P1: Dominant がより好みと回答したが NonDominant の方が速い, P3: P1 と逆)。しかし、両参加者において MT の平均がより速い手法はいずれもより後に行った手法であり、この結果には学習による順序効果が影響した可能性がある。ここから、本手法には習熟による選択性能の向上の可能性があると考えられる。

実験後アンケートの結果から、NonDominant におけるポイントの見え方に関して個人差が見られた。

ここから、単に利き眼あるいは非利き眼と言っても、左右の眼それぞれへの依存度には個人差があることが示唆された。そのため、ポイントを左右どちらの眼のみに表示するかを決定する際には、機械的な判別ではなく、ユーザの好みや感じ方を反映できる方法を用いる必要があると考えられる。

5 今後の課題

5.1 サンプルの多様化

本実験の参加者は全員が 20 代かつ右利きであり、そのうち左眼が利き眼であったのは 1 人のみであった。さらに、参加者全体の人数も 12 人と少なかった。今後の実験においては、より多数の、かつより多様な属性をもつ参加者を対象とすることにより、より正確な実験結果が得られると考えられる。

5.2 他の多様な手法との比較

本研究においては、従来手法として、両眼に表示されるポイントを用いた選択手法を本手法の評価のための比較対象とした。しかし、実際の xR 環境においては、この手法以外にも多様な選択手法が存在する。これらに対する比較実験を行うことにより、より正確に本手法を評価できると考えられる。

5.3 実際の xR 環境を想定した参加者実験の実施

本研究における実験環境について、実験の単純化のために、背景には平坦で模様のない白色の壁を用いた。また、選択すべきターゲット以外に、紛らわしいオブジェクトは存在しなかった。しかし実際の xR 環境においては、よりオブジェクトとカーソルが視認しにくい環境での操作が求められる。このような実際の xR 環境を想定した状況における各手法の性能については、追加の調査が必要である。

6 まとめ

本研究は、xR におけるオブジェクト選択に関して、両眼視差による問題を生じない、片眼のみに表示されるポイントを用いた手法を示した。また、本手法を実装した後、その性能を評価するための Fitts の法則に基づく比較実験を行った。その結果、本手法は従来手法と比べて、より奥行きが深く小さいオブジェクトを選択する際に、操作時間、エラー率、スループット、作業負荷、およびユーザビリティに関してより高い性能をもつことが示された。また、ポイントを表示する眼の決定には、利き眼だけでなくユーザの好みを考慮する必要があることが示唆された。

参考文献

- [1] F. Argelaguet and C. Andujar. Visual Feedback Techniques for Virtual Pointing on Stereoscopic

- Displays. In *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology*, VRST '09, pp. 163–170, New York, NY, USA, 2009. Association for Computing Machinery.
- [2] J. Brooke. SUS: A quick and dirty usability scale. *Usability Evaluation in Industry*, pp. 189–194, 1996.
- [3] J. C. Byers, A. C. Bittner, and S. G. Hill. Traditional and raw task load index (TLX) correlations: Are paired comparisons necessary? In A. Vol. 1, pp. 481–485, 1989.
- [4] P. Fitts. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47(6):381–391, June 1954.
- [5] J. Gabel, S. Schmidt, O. Ariza, and F. Steinicke. Redirecting Rays: Evaluation of Assistive Raycasting Techniques in Virtual Reality. In *Proceedings of the 29th ACM Symposium on Virtual Reality Software and Technology*, VRST '23, pp. 1–11, New York, NY, USA, 2023. Association for Computing Machinery.
- [6] International Organization for Standardization. Ergonomic requirements for office work with visual display terminals (VDTs) - Part 9: Requirements for non-keyboard input devices, 2000.
- [7] S. Jang, W. Stuerzlinger, S. Ambike, and K. Ramani. Modeling Cumulative Arm Fatigue in Mid-Air Interaction based on Perceived Exertion and Kinetics of Arm Motion. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pp. 3328–3339, New York, NY, USA, 2017. Association for Computing Machinery.
- [8] J. H. Lee, T. Jin, S.-H. Lee, S.-J. Lee, and S.-H. Bae. Stereoscopic Viewing and Monoscopic Touching: Selecting Distant Objects in VR Through a Mobile Device. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, UIST '23, pp. 1–7, New York, NY, USA, 2023. Association for Computing Machinery.
- [9] D. Mardanbegi, B. Mayer, K. Pfeuffer, S. Jalaliniya, H. Gellersen, and A. Perzl. Eye-SeeThrough: Unifying Tool Selection and Application in Virtual Environments. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 474–483, 2019.
- [10] Y. Qin, C. Yu, W. Yao, J. Yao, C. Liang, Y. Weng, Y. Yan, and Y. Shi. Selecting Real World Objects via User-Perspective Phone Occlusion. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, pp. 1–13, New York, NY, USA, 2023. Association for Computing Machinery.
- [11] A. Ramcharitar and R. J. Teather. EZCursorVR: 2D Selection with Virtual Reality Head-Mounted Displays. In *Proceedings of the 44th Graphics Interface Conference*, GI '18, pp. 123–130, Waterloo, CAN, 2018. Canadian Human-Computer Communications Society.
- [12] U. Wagner, M. N. Lystbæk, P. Manakhov, J. E. S. Grønbaek, K. Pfeuffer, and H. Gellersen. A Fitts' Law Study of Gaze-Hand Alignment for Selection in 3D User Interfaces. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, pp. 1–15, New York, NY, USA, 2023. Association for Computing Machinery.
- [13] C. Ware and K. Lowther. Selection using a one-eyed cursor in a fish tank VR environment. *ACM Trans. Comput.-Hum. Interact.*, 4(4):309–322, Dec. 1997.
- [14] 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, New York, NY, USA, 2011. Association for Computing Machinery.
- [15] カラーユニバーサルデザイン機構. カラーユニバーサルデザイン推奨配色セットについて, 2018. https://cudo.jp/?page_id=1565 [アクセス日: 2025/08/26].
- [16] 郭凱, 下村祐樹, 細井十楽, 伴祐樹, 割澤伸一. ビジュアルシーンと重畳せずに認識可能な視野闘争を用いた視覚的な UI 提示手法の検討. インタラクション 2024 論文集, インタラクション 2024, pp. 1011–1016, 2024.
- [17] 山本航世, 渡邊恵太. ParaSights: 両眼視野闘争によって 2つの環境と並行してインタラクションできる空間提示手法. WISS 2024 予稿集, WISS 2024, 2024.
- [18] 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.
- [19] 北澤一樹, 勝山しおり, 古川智巳, 外里富佐江. 5種類の利き眼検査による一致性の検討. 長野保健医療大学紀要 Vol. 6, pp. 13–17, 2021.

効率的な遠距離オブジェクトの回転操作を可能とする トラックボール型 VR コントローラ

岩井 望* 阿部 優樹* 坂本 大介*

概要. 仮想現実 (Virtual Reality; VR) におけるコントローラを用いた従来のインタラクションは、腕や手首を繰り返し動かす必要があり、手の疲労に繋がりがやすい。特に、ユーザの手の届かない範囲にあるオブジェクトを効率的に回転させることが難しい。本稿では VR における効率的なオブジェクトの操作の実現のため、球形のインタフェースであるトラックボールを使用することを検討する。商用の VR コントローラを改良してプロトタイプを実装し、トラックボールを利用した回転手法を実装した。このトラックボール型のコントローラによる操作手法の性能を評価するため、2つの従来手法を含めた3つの手法間での比較実験を行った。その結果、トラックボール型コントローラによる操作手法は特に手の届かない遠距離のオブジェクトの回転操作に優れており、トラックボールを用いた回転手法の VR インタラクションにおける有用性が示された。

1 はじめに

仮想現実 (Virtual Reality; VR) 技術は、エンタテインメントやデザイン [8]、データ探索 [14] などの分野での活用が進んでいる。これらの VR アプリケーションでは、ユーザはコントローラを使用し、現実と同様に手でオブジェクトの操作を行うことができる。しかし、オブジェクトを回転させるには、腕や手首を回す操作を繰り返す必要があるため、疲労に繋がりがやすい [7]。また、ユーザの手が届かない遠距離のオブジェクト回転操作について、確立された手法が存在するとは言えない。

これらの課題に対し、腕や手首を回す操作に加えて、VR コントローラ上での操作を回転入力に利用する手法が提案されている [9] [10] [12] [13]。最も本研究に関連する重要な研究は Kim らの研究であり、彼らは VR コントローラのトラックパッドを活用する手法を提案した [9] [10]。この手法は、従来の腕や手首による回転操作はそのままに、親指のトラックパッド操作で追加の回転を行う。従来手法に比べて回転操作に伴う疲労が軽減され、特に遠距離のオブジェクトの回転を効率化できる可能性が示されている。しかし、トラックパッドのスワイプによる回転操作がユーザにとって直感的でない可能性も示唆された。

そこで我々は、Kim らの手法で用いられていたトラックパッドを、Blender [1] などの 3D グラフィックスソフトウェアで直感性が示されている [5][11] [20]、トラックボールに置き換える手法を提案する。トラックボールは球を転がして入力するため、スワイプ操作よりもオブジェクトの回転と入力のマッピングが

直感的になる可能性がある。

本研究の目的は、VR 上の 3D オブジェクトを効率的かつ直感的に回転させる手法を開発することである。その実現のため、まず既存のコントローラを改造し、トラックボールを埋め込んだプロトタイプを開発した。そしてトラックボールによる回転操作を実装し、その有用性と実用性を調査するため、近距離および遠距離のオブジェクトを操作するタスクを用いた実験を実施した。そのパフォーマンスとユーザビリティを、従来のコントローラによる手法、そして Kim らのトラックパッドを利用した手法の2つと比較評価する。

2 関連研究

2.1 VR における仮想 3D オブジェクトの操作手法

VR における仮想 3D オブジェクトの操作手法としては「ダイレクトハンド手法」が主流であり、ユーザは現実と同様に、オブジェクトを掴んで直感的に操作できる。しかし、長時間の使用は腕の疲労に繋がりがやすい [7] [17]。

この課題に対するアプローチの1つとして、腕や手首の操作だけでなく、VR コントローラ上のインタフェースへの入力を活用する手法が提案されている [9] [10] [12] [13]。Kim らは、従来のダイレクトハンド手法に加え、トラックパッドに Arcball [15] と呼ばれる回転操作を実装することで、ダイレクトハンド手法の効率性を維持しながら、大きな回転操作に伴う腕の疲労を軽減することを示した [9] [10]。

また、ダイレクトハンド手法は、遠距離のオブジェクトに対する操作が困難という課題も抱えている。遠距離操作を支援する手法としては、対象との距離

Copyright is held by the author(s).

* 北海道大学

に応じて手の動きをスケールする Direct HOMER (Hand-centered Object Manipulation Extending Ray-casting) [2] が挙げられるが、手の不安定さによる誤差も拡大されるために精度に課題が残る。Kim らはこの「HOMER 手法」とトラックパッドによる回転操作を組み合わせることで遠距離オブジェクトの操作を可能とした。HOMER 手法による 6 自由度 (Degrees of Freedom; DoF) の操作と、トラックパッドによる回転操作が分離されるため効率的であった一方、その 2DoF のスイープ入力を回転に変換する手法は直感的とは言えない可能性があった [10]。我々は、Kim らの手法をトラックボールによる 3DoF の回転操作に拡張することで、直感性も同時に兼ね備えた手法を開発することを目的とする。

2.2 3D オブジェクト操作のためのトラックボールユーザインタフェース

トラックボールは主に PC に対するポインティングデバイスとして用いられる。その一方、3D グラフィックスソフトウェアにおけるオブジェクトの回転制御への活用も提案されている [5][11] [20]。Froehlich らは、トラックボールを使用した 6DoF 入力デバイスを開発し、球の回転をオブジェクトの回転に直接対応させることで回転操作を可能とした [5][11]。商用の 6DoF 入力デバイスよりも、直感的で、速度やユーザビリティに優れる可能性が示されている。VR あるいは拡張現実感 (Augmented Reality) においてトラックボールを活用した研究は存在するが [19] [13] [18]、この球による回転操作を VR におけるオブジェクトの回転操作に応用した研究は我々の知る限り存在しない。

我々は、Kim らの手法を参考に、トラックボールによる回転操作を導入することで、直感的かつ効率的な、近距離および遠距離のオブジェクト操作が行えると考えた。本研究では、既存の VR コントローラのトラックパッドをトラックボールに換装し、球の回転によってオブジェクトを回転できる手法を開発する。

3 プロトタイプ

本研究では、既存の VR コントローラにトラックボールを搭載し、従来の 6DoF オブジェクト操作に加えて、トラックボールによる 3DoF の回転操作を可能にする手法を提案する。この手法を実現するため、プロトタイプを実装した。

3.1 ハードウェアとシステムの実装

プロトタイプの概観を図 1 に示す。VR コントローラには、Kim らの手法で用いられている HTC VIVE コントローラを採用した。まず、西村らがオープンソースとして公開している実装 [6] を参考に回転検知モジュールを作成した。球を設置する台座部



図 1. プロトタイプ実装の概観。34mm の球のプロトタイプ (a) 側面、(b) 内部構造。(c) 実装した 3 種類のプロトタイプ。球の大きさが左から 25mm, 34mm, 40mm。

は西村らの 3D モデルを使用しており、コントローラを傾けた際に球が落下しないよう、球の被覆部分を拡張して 3D プリンタで出力した。拡張は、球を真下に向けても球が落ちず、かつ親指で球を転がした際になるべく摩擦を感じない程度に行った。また、球は台座部から取り外し可能であった。この台座部に 2 つのイメージセンサを 90 度ずらして配置し、球の 3DoF の回転を検知した (図 1a)。2 つのイメージセンサには PixArt 社の PMW3360 を使用した。そして、これらのセンサはマイコンボード (Arduino Nano Every) に接続され、シリアル通信によってデータを Unity に送信することで、VR 上の 3D オブジェクトの回転に反映された (図 1b)。

次に、このモジュールを VIVE コントローラのトラックパッド部分に固定するための部品を、同じく 3D プリンタで出力した。上部アタッチメント (図 1b) については、公開されている 3D モデルを利用し [16]、それ以外の部品については Unity で使用される VIVE コントローラの 3D モデルを利用して作成した。これらの部品を用いて、台座部を、親指での操作にしやすいのために斜めに約 45 度傾けて取り付けた。イメージセンサについては接着剤で台座部に固定し、他の部品については取り外し可能な形状とした。また、3D プリンタのフィラメントには PETG フィラメントを使用した。

3.2 球のサイズの決定

次に、球のサイズを決定するため、この上述のプロトタイプを商用のトラックボールで一般的な直径 25mm, 34mm, 40mm の 3 種類の球で実装し (図 1c)、予備調査を実施した。それぞれの球にはエレコム社、サンワサプライ社、そして LZYDD 社の球を使用した。予備調査では、6 名の参加者に後述する実験タスクを実施し、使いやすさについての主観

的なフィードバックを収集した。ただし、評価に十分な時間を与えるため、試行回数や順序に制限は設けなかった。その結果、6名中3名が34mmのプロトタイプを最も好み、2名が25mmのプロトタイプを最も好んだ。また、25mmのプロトタイプについては「球を転がす際に摩擦を感じる」、「指に対して球が小さいために一部回転が行いにくい」という意見が得られたため、本研究では34mmのプロトタイプを採用し、実験で用いた。また、VIVE コントローラが203gであったのに対し、34mmのプロトタイプは237gであった。

4 実験

本実験では、VRにおけるトラックボールを用いた回転手法の有用性を調査する。Kimらの実験デザインを参考に、従来手法とKimらのトラックパッドを用いた手法を比較対象とし、近距離および遠距離のオブジェクトに対する操作タスクでそのパフォーマンスとユーザビリティを評価する。近距離タスクは手の届く場所でのタスクであるのに対し、遠距離タスクは直接手が届かないため、VRに特有の操作が必要となるタスクである。

4.1 実験参加者

実験参加者として12名(女性2名、平均年齢22.5歳、標準偏差4.17歳)が参加した。利き手は右利きが9名、左利きが3名で、全員が利き手で操作を行った。VRは8名に使用経験があり、その頻度は月1回程度が2人、過去数回程度の使用経験が6名だった。トラックパッドは全員が月1回以上の頻度で使用しており、10名が週1回以上の頻度で使用していた。トラックボールは5名に使用経験があり、1名が週1回以上使用、他4名は過去数回程度の使用経験であった。トラックパッドとトラックボールについてはPCに対するポインティングデバイスとしての使用経験が含まれた。実験の所要時間は約1時間半程度で、参加者には謝礼として1,000円分のQuoカードを渡した。また、本実験は、北海道大学大学院工学研究院及び大学院情報科学研究院 人を対象とする実験研究に関する倫理審査委員会の承認を得ている(承認番号 R7-07)。

4.2 実験環境

実験には、CPUにIntel(R) Core(TM) i9-11900K @ 3.50GHz、32GBのRAM、GPUにNVIDIA GeForce RTX 3060 Tiを搭載したPCを使用した。VRの頭部装着型ディスプレイ(Head Mounted Display; HMD)とコントローラにはHTC VIVE Proを用い、実験のアプリケーションはUnity ver.2020.3.30f1で開発した。

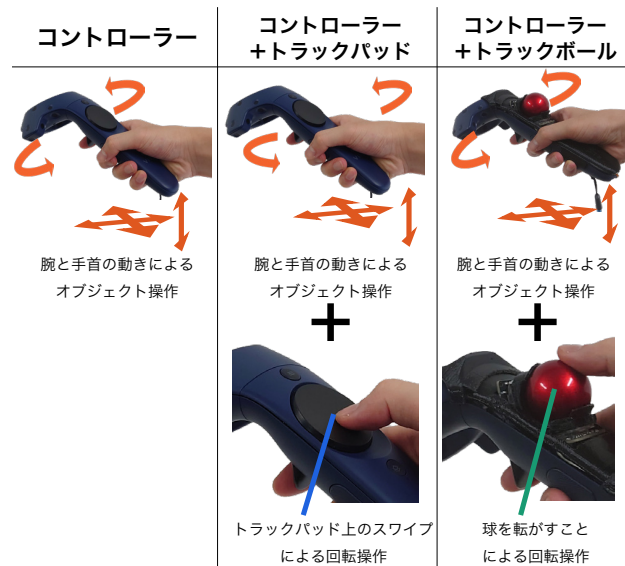


図 2. 3つの入力手法の概観。

4.3 比較手法

コントローラ手法(コントローラ)とコントローラ+トラックパッド手法(トラックパッド)、コントローラ+トラックボール手法(トラックボール)の比較を行った。図2にそれぞれの手法の概観を示し、それぞれの入力手法の詳細について以下に示す。

4.3.1 コントローラ手法(コントローラ)

コントローラ手法は、Kimらの先行研究と同様に本実験におけるベースラインであり、コントローラでオブジェクトを掴み、移動や回転を行う手法である。

近距離操作では、ダイレクトハンド手法により、コントローラがオブジェクトに触れた状態でトリガーを引くと対象を把持できた。その後、腕や手の動きがオブジェクトの位置と向きに反映された。遠距離操作では、HOMER手法により、コントローラから射出されるレイをオブジェクトに当て、トリガーを引くことで把持できた。その後、コントローラの移動量が掴んだ時点のユーザからオブジェクトまでの距離に応じ、スケールされてオブジェクトに反映された。回転は、コントローラの向きがオブジェクトの向きに1対1で対応されることで制御できた。

4.3.2 コントローラ+トラックパッド手法(トラックパッド)

コントローラ+トラックパッド手法は、Kimらの先行研究で提案された手法であり、コントローラ手法の操作にトラックパッドによる回転操作を追加した手法である[9]。ユーザはオブジェクトを把持した状態で、親指でトラックパッドをスワイプすることで、オブジェクトを回転させることができた。トラックパッド上のスワイプの指の動きが、HMDの座標に向かうベクトル(カメラ方向)の向きで仮想

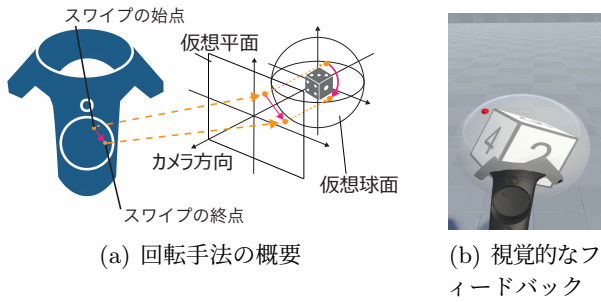


図 3. トラックパッド手法の回転手法の概要と実験参加者に与えられた視覚的なフィードバックの見た目。

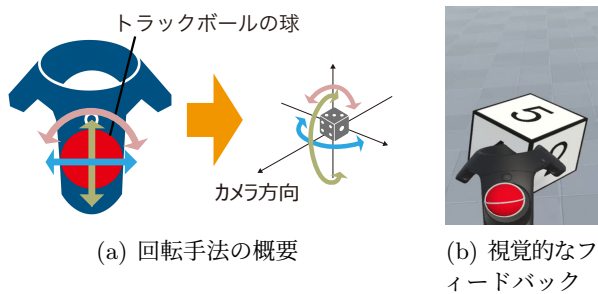


図 4. トラックボール手法の回転手法の概要と実験参加者に与えられた視覚的なフィードバックの見た目。

的な球面上（仮想球面）にマッピングされ、回転が計算される（図 3a）。図 3b に示すように、この仮想球面へのマッピングの状態は実験参加者に視覚的にフィードバックされた。

4.3.3 コントローラ+トラックボール手法（トラックボール）

コントローラ+トラックボール手法は、3 節で実装したプロトタイプを用いる本研究の提案手法である。トラックパッド手法と同様に、コントローラ手法に加えて、親指でトラックボールを操作することで追加の回転入力を行う。その際の球の回転が図 4a に示す対応関係においてオブジェクトに反映された。また、図 4b に示すように、手元の球の回転が実験参加者に視覚的にフィードバックされた。

4.4 実験タスク

タスクは Kim らの先行研究と同じく、操作対象のオブジェクトを、手本のオブジェクトと同様の位置と回転角度に配置するドッキングタスクである [10]。実験参加者は椅子に座った状態で HMD を被り、このタスクを遂行した。タスクでは、3 秒のカウントダウンの後に最初の試行を開始し、参加者の前方に白い不透明なサイコロのオブジェクトと、赤い半透明のサイコロのオブジェクトを 1 つずつ表示した。参加者は白いサイコロを各手法で操作し、赤いサイコロに位置と向きが一致するように移動および回転

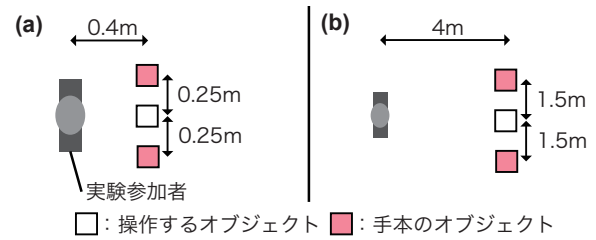


図 5. (a) 近距離タスクと (b) 遠距離タスクにおける、実験参加者とオブジェクトの位置関係。

させた。サイコロの向きは各面の数字によって特定でき、2つのサイコロの位置と向きがある閾値以下で合致すると赤いアウトラインを表示するようにし、その状態を 500ms 維持することで参加者は試行を完了できた。1つの試行が完了すると、1秒のカウントダウンの後に再びサイコロを表示し、次の試行を開始した。参加者はこの試行を繰り返し行った。

このタスクを近距離および遠距離のオブジェクトに対して実施した。近距離タスクでは各辺 10cm の白いサイコロを参加者の 0.4m 前方に、同じ大きさの赤いサイコロをその左右のどちらか 0.25m の位置に配置した（図 5a）。遠距離タスクでは各辺 40cm の白いサイコロを参加者の 4m 前方に、赤いサイコロをその左右のどちらか 1.5m の位置に配置した（図 5b）。誤差の閾値は、近距離タスクで位置 3cm、角度 10° 、遠距離タスクでは位置 10cm、角度 10° とした。各試行の開始時、白いサイコロは常に同じ初期状態で提示し、赤いサイコロはランダムな軸で時計回りに 135° 回転した向きで、左右どちらかの位置にランダムな順で提示した。ただし、左右の出現回数は参加者間で同一とした。

4.5 実験デザイン

本実験は参加者内実験計画を用いて行った。独立変数は手法（コントローラ、トラックパッド、トラックボール）であり、手法についてラテン方格法を考慮した順番でタスクを行った。参加者は各手法において、まず 20 試行で構成される近距離タスクを行い、次に同じく 20 試行で構成される遠距離タスクを行った。従って、各参加者は 3 手法 $\times$ 2 タスク $\times$ 20 試行 = 120 回のオブジェクト操作を行った。また実験参加者は 12 名であるため、合計 1,440 データが収集された。

従属変数として試行時間と総回転角度 [9] を測定した。総回転角度は 1 回の試行において実験参加者がオブジェクトを何度回転させたかを表す。回転操作を行えば行うほど総回転角度は大きくなるため、総回転角度は手法ごとの回転操作の容易さを測る指標として扱うことができる。また、主観的な評価として参加者は、各タスク完了後に NASA Raw Task Load Index (NASA-RTLX) [4] と System Usability Scale (SUS) [3] に基づく 2 つのアンケー

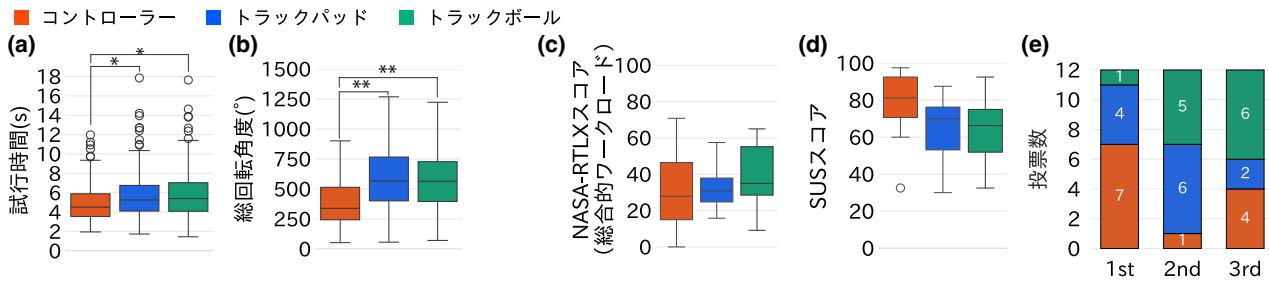


図 6. 近距離タスクにおける手法ごとの (a) 試行時間, (b) 総回転角度, (c) NASA-RTLX の総合的ワークロードの平均スコア, (d) 平均 SUS スコア, (e) 手法の好み (1st が最も好まれたことを意味する). 有意差は * $p < .05$, ** $p < .01$.

ト, そして全試行の終了後に手法の好みの順位とその理由についてのアンケートに回答した。

4.6 実験手順

実験参加者が実験室に到着後, 実験の概要を説明し, 実験参加の同意書への署名を依頼した。その後, 事前アンケートへの回答を求め, タスクと操作手法の詳細を説明した。説明が終了後, 参加者は HMD を装着し, 1 つ目の手法で近距離タスクの練習を 3 分以上, 実験参加者から操作に慣れたという回答が得られるまで行った。練習後に本番タスクを遂行し, 完了すると参加者は NASA-RTLX と SUS に基づく 2 つのアンケートに回答した。次に, 遠距離タスクの練習と本番タスクを同様にを行い, 完了後に再度アンケートへの回答を求めた。なお, 遠距離タスク前の練習には, 操作への慣れや疲労を考慮し, 最低時間は設けなかった。上記の手順を 3 つの手法全てで繰り返し, 各手法の間には最低 1 分間の休憩を設けた。全てのタスクが完了した後, 最後に手法の好みに関するアンケートへの回答を求め, そのアンケートの完了をもって実験を終了とした。

5 実験結果

それぞれの従属変数と, NASA-RTLX と SUS によって得られた値について, フリッドマン検定とウィルコクソンの符号順位検定を用い, Holm 法による補正を用いた。手法の好みの順位については加重平均を算出した。

5.1 近距離タスク

5.1.1 試行時間

試行時間に対して手法間に有意な差が確認された ($\chi^2_{2,N=12} = 10.17, p < .01$)。また, 事後検定により, コントローラはトラックパッド ($p < .05, r = .47$) とトラックボール ($p < .05, r = .57$) より有意に高速であった (図 6a)。

5.1.2 総回転角度

総回転角度に対して手法間に有意な差が確認された ($\chi^2_{2,N=12} = 15.17, p < .01$)。また, 事後検定に

より, コントローラの総回転角度はトラックパッド ($p < .01, r = .71$) とトラックボール ($p < .01, r = .67$) より有意に小さかった (図 6b)。

5.1.3 NASA-RTLX & SUS

NASA-RTLX の総合的ワークロードスコア (低いほど良い) および SUS の総合スコア (高いほど良い) についての平均スコアを図 6c, d に示す。NASA-RTLX の総合的ワークロードスコアに手法間で有意な差は認められなかった。また, 同様に SUS スコアについても, 総合スコアに有意な差は認められなかった。

5.1.4 手法の好み

実験参加者 12 名中 7 名がコントローラ, 4 名がトラックパッド, 1 名がトラックボールを最も好んだ (図 6e)。コントローラ, トラックパッド, トラックボールの手法の好みの順位の加重平均はそれぞれ 1.75, 1.83, 2.42 であった (低いほど良い)。

コントローラは「直感的だと感じた (P2, P3, P4, P5, P9, P10, P11)」という理由で好まれ, トラックボールは「意図した方向に転がしづらかった (P3, P6, P7, P8)」や「親指が疲れた (P4, P5, P12)」といった理由で好まれなかった。

5.2 遠距離タスク

5.2.1 試行時間

試行時間に対して手法間に有意な差が確認された ($\chi^2_{2,N=12} = 19.50, p < .01$)。また, 事後検定により, トラックボールはコントローラ ($p < .01, r = .71$) とトラックパッド ($p < .01, r = .53$) より有意に高速で, トラックパッドはコントローラより有意に高速であった ($p < .01, r = .71$) (図 7a)。

5.2.2 総回転角度

総回転角度に対して手法間に有意な差が確認された ($\chi^2_{2,N=12} = 18.00, p < .01$)。また, 事後検定により, コントローラの総回転角度はトラックパッド ($p < .01, r = .71$) とトラックボール ($p < .05, r = .45$) より有意に小さく, トラックボールの総回

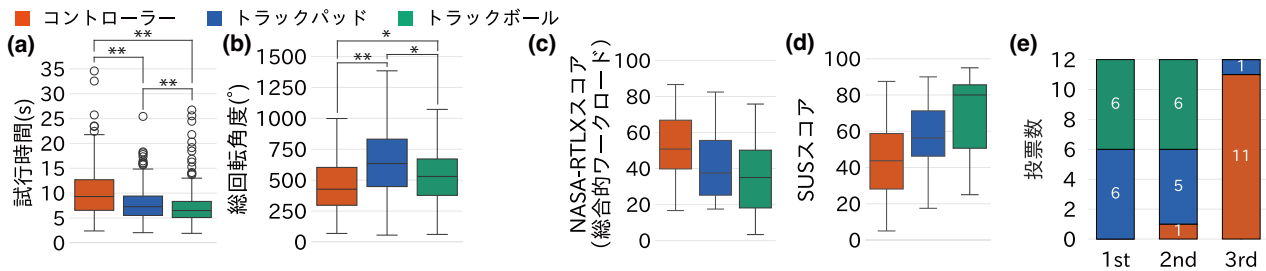


図 7. 遠距離タスクにおける手法ごとの (a) 試行時間, (b) 総回転角度, (c) NASA-RTLX の総合的ワークロードの平均スコア, (d) 平均 SUS スコア, (e) 手法の好み (1st が最も好まれたことを意味する). 有意差は * $p < .05$, ** $p < .01$.

転角度はトラックパッドより有意に小さかった ($p < .05$, $r = .51$) (図 7b).

5.2.3 NASA-RTLX & SUS

NASA-RTLX の総合的ワークロードスコア (低いほど良い) および SUS の総合スコア (高いほど良い) についての平均スコアを図 7c, d に示す. NASA-RTLX の総合的ワークロードスコアに手法間で有意な差は認められなかった. また, 同様に SUS スコアについても, 総合スコアに有意な差は認められなかった.

5.2.4 手法の好み

実験参加者 12 名中 6 名がトラックパッド, 6 名がトラックボールを最も好んだ (図 7e). コントローラ, トラックパッド, トラックボールの手法の好みの順位の加重平均はそれぞれ 2.92, 1.58, 1.50 であった (低いほど良い).

トラックボールは「トラックパッドよりも回転させやすかった (P1, P3, P8, P11)」という理由から好まれ, コントローラは「腕や手首の負担を感じた (P1, P2, P10)」や「直感的ではないと感じた (P5, P12)」という理由から好まれなかった.

6 実験のまとめと考察

本実験で得られた結果を以下に要約する.

- 近距離タスクでは, コントローラが最も高速で, 最も好まれる手法であった. 一方, トラックボールは最も好まれない手法であり, 球の操作に伴う親指への負担が指摘された.
- 遠距離タスクでは, トラックボールが最も高速で, かつ好まれる手法であった. また, トラックパッドの総回転角度は最も大きく, その回転操作が直感的でない可能性があった.
- 実験全体を通じて「トラックパッドよりトラックボールの方が直感的と感じた (P6, P7)」という意見が得られた.

以上の結果から, 近距離のオブジェクト操作においてトラックボールは操作効率の向上に寄与しない

ことが示唆された. NASA-RTLX スコアはコントローラ操作による一定の身体的疲労を示したものの, 現実と同様にオブジェクトを直接掴んで操作できる直感性が, その負担を上回ったと考えられる.

対照的に, 現実とは異なる操作が求められる遠距離タスクでは, コントローラによる操作がかえって直感的ではなかったことが示唆された. 本実験結果は, トラックボールがこの課題を解決する可能性を示している. これは, 物理的な球の回転をオブジェクトの回転に 1 対 1 で反映するメタファが, コントローラによる操作, さらには 2DoF のスワイプ入力を回転に変換するトラックパッド手法よりも, 参加者にとって直感的であったためだと考えられる.

本実験により, トラックボールの利用には利点と欠点が存在することが明らかになったが, これらの知見に基づき, 今後の VR コントローラへのトラックボールの導入を提案する. これにより, 従来のコントローラの近距離オブジェクトの操作や, 物理ボタン等の入力機構をそのままに, 遠距離のオブジェクトの回転操作の効率性の向上が期待できる. その実現のため, 今後はより実践的なタスクや長期利用での効果を検証したい. 特に, 本実験でトラックボール手法の試行時間の結果にばらつきが見られたことから, 長期使用における習熟過程を解析することが, さらなるパフォーマンスとユーザビリティの向上に繋がると考えられる.

7 まとめ

本研究では, VR における効率的な 3D オブジェクトの操作を目的とし, 既存のコントローラにトラックボールを組み込むプロトタイプを作成し, 球による回転操作を用いる手法を提案した. その有用性を検討するため, 近距離および遠距離の操作タスクにおいて, 2 つの従来手法との比較実験を行った. その結果, 特に遠距離タスクにおいて, 提案手法がパフォーマンスと主観的評価に優れる可能性が示された. 本研究で実装したトラックボール型コントローラはプロトタイプであり, 実用化には改善が必要だが, 確立された手法がまだない遠距離のオブジェクト操作に対し, 本手法が有効なアプローチとなり得ることを示せた点に意義があると考えられる.

謝辞

本研究は、JST 創発的研究支援事業、JPMJFR226S の支援を受けたものである。

参考文献

- [1] Blender. Blender - The Free and Open Source 3D Creation Software —blender.org. <https://www.blender.org/>, 2025. Accessed: August 31, 2025.
- [2] D. A. Bowman and L. F. Hodges. An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. In *Proceedings of the 1997 Symposium on Interactive 3D Graphics*, I3D '97, p. 35–ff., New York, NY, USA, 1997. Association for Computing Machinery.
- [3] J. Brooke. SUS: A quick and dirty usability scale. *Usability Eval. Ind.*, 189, 11 1995.
- [4] J. C. Byers, A. C. Bittner, and S. G. Hill. Traditional and raw task load index (TLX) correlations: Are paired comparisons necessary? In A. 1989.
- [5] B. Froehlich, J. Hochstrate, V. Skuk, and A. Huckauf. The GlobeFish and the Globe-Mouse: two new six degree of freedom input devices for graphics applications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, p. 191–199, New York, NY, USA, 2006. Association for Computing Machinery.
- [6] GitHub. ParRot · GitHub. <https://github.com/ParRot-3DMouse>, 2025. Accessed: August 31, 2025.
- [7] J. D. Hincapié-Ramos, X. Guo, P. Moghadasian, and P. Irani. Consumed endurance: a metric to quantify arm fatigue of mid-air interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, p. 1063–1072, New York, NY, USA, 2014. Association for Computing Machinery.
- [8] S. Kalantari and J. R. J. Neo. Virtual Environments for Design Research: Lessons Learned From Use of Fully Immersive Virtual Reality in Interior Design Research. *Journal of Interior Design*, 45, 05 2020.
- [9] S. Kim and G. Lee. Virtual Trackball on VR Controller: Evaluation of 3D Rotation Methods in Virtual Reality. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI EA '23, New York, NY, USA, 2023. Association for Computing Machinery.
- [10] S. Kim and G. Lee. Evaluating an In-Hand Ball-Shaped Controller for Object Manipulation in Virtual Reality. In *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pp. 10–19, 2024.
- [11] A. Kulik, J. Hochstrate, A. Kunert, and B. Froehlich. The influence of input device characteristics on spatial perception in desktop-based 3D applications. In *2009 IEEE Symposium on 3D User Interfaces*, pp. 59–66, 2009.
- [12] J. Lee, M. Sinclair, M. Gonzalez-Franco, E. Ofek, and C. Holz. TORC: A Virtual Reality Controller for In-Hand High-Dexterity Finger Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, p. 1–13, New York, NY, USA, 2019. Association for Computing Machinery.
- [13] C. Lim, J. Kim, and M. J. Kim. Thumble: One-Handed 3D Object Manipulation Using a Thimble-Shaped Wearable Device in Virtual Reality. In *Adjunct Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, UIST '22 Adjunct, New York, NY, USA, 2022. Association for Computing Machinery.
- [14] K. Marriott, F. Schreiber, T. Dwyer, K. Klein, N. H. Riche, T. Itoh, W. Stuerzlinger, and B. H. Thomas. *Immersive Analytics*. Vol. 11190. Springer, 2018.
- [15] K. Shoemake. ARCBALL: a user interface for specifying three-dimensional orientation using a mouse. In *Proceedings of the Conference on Graphics Interface '92*, p. 151–156, San Francisco, CA, USA, 1992. Morgan Kaufmann Publishers Inc.
- [16] Thingiverse. HTC Vive Joystick mod. by 3DaptiveClint - Thingiverse. <https://www.thingiverse.com/thing:2705398>, 2025. Accessed: August 31, 2025.
- [17] D. Yu, T. Dingler, E. Velloso, and J. Goncalves. Object Selection and Manipulation in VR Headsets: Research Challenges, Solutions, and Success Measurements. *ACM Comput. Surv.*, 57(4), Dec. 2024.
- [18] D. B. Zhao and W. Stuerzlinger. A Novel Bare-Handed Manipulation Technique for Distant Objects in Virtual Reality. In *2025 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 353–360, 2025.
- [19] 岩井 望, 崔 明根, 坂本 大介, 小野 哲雄. AR 入力デバイスとしての身体装着型トラックボールの検討. インタクション 2024 論文集, pp. 68–77, 2024.
- [20] 西村 陽向, 高橋 伊織, 成瀬 宏亮. トラックボール型 3D マウス ParRot の提案と評価. インタクション 2025 論文集, pp. 818–823, 2025.

未来ビジョン

既存の VR コントローラでは、入力インタフェースとしてジョイスティックやトラックパッドが主流であり、トラックボールの使用は検討されてこなかった。我々は本研究を基に、VR コントローラへのトラックボールの導入を推進したい。本研究は、トラックボール型の VR コントローラが、遠距離オブジェクトの操作効率を向上させる可能性を示した。これは、VR 空間内でのワールド制作といった大規模な創作活動を促進し、さらには、遠距離のオブジェクトを直感的な操作による新たなゲーム体験の創出にも寄与すると考える。

加えて、本コントローラは VR におけるアクセシビリティの向上にも繋がる可能性があると考えられる。トラックボールは少ない手の動きで操作が可能のため、上肢の運動機能に制約のあるユーザにとって扱いやすいデバイスとなり

得る。また、「ごろ寝マウス」や「空中マウス」と呼ばれる、臥位での使用を想定した手持ち型のトラックボールが市販されており（図 8）、本コントローラも寝ながらの VR 利用に適している可能性が示唆される。本研究では回転操作に着目したが、今後はアクセシビリティの観点からも調査を進め、多様なユーザが快適な VR 体験を享受できる未来の実現を目指す。



図 8. エレコム株式会社. ワイヤレスハンディトラックボール M-RT1DRBK

VR 環境におけるレイキャストによるオブジェクトの 選択成功率推定ツールの提案

奥野 達也* 清水 春翔* 笠原 暢仁* 本間 大一優* 山中 祥太† 宮下 芳明*

概要. XR デバイスの普及により三次元空間におけるインタラクションが一般化する中、UI 開発者はより優れたユーザ体験を実現するためにユーザビリティへの配慮が強く求められている。HCI 分野では、この課題解決のため古くからポインティング研究が行われ、三次元空間を対象とした研究も大きく進展している。しかし、研究が進む一方で、そこで得られた指針を実際の開発者が UI 改善のサイクルに直接活用するためには、実用的なツールを提供することが求められる。本研究は、VR 開発におけるこの「研究と開発のギャップ」を埋めることを目指し、開発ツール内でオブジェクトの選択成功率を推定するシステムを提案する。本稿ではその理論の検証を行い、ツールの機能について説明する。また、VR 開発者に本ツールを試用してもらい、そのフィードバックから有用性を検証する。

1 はじめに

仮想現実 (VR) 技術は、従来の 2D ディスプレイでは得られない高いリアリティとインタラクティブ性を提供する。VR 環境において、ユーザがユーザインタフェース (UI) 要素をコントローラーで選択する動作は、あらゆるインタラクションの基礎となる最も重要な操作の一つである。

VR におけるポインティングは、奥行き方向のターゲット配置の定義など、2D 環境にはない要因によってポインティング操作が複雑化することが指摘されている [1]。特に、ユーザから遠くに、あるいは小さく表示されるターゲットを選択する際の難易度は著しく上昇し、ユーザの操作ミスやフラストレーション、ひいては VR 体験全体の質の低下に繋がりがかねない。したがって、VR コンテンツの開発者やデザイナーが、UI 要素がどの程度選択しやすいかを定量的に評価できれば、VR 体験を向上させられる可能性がある。

本稿では、この課題に応えるため、VR 環境におけるポインティング操作の成功率を推定する数理モデルを構築し、そのモデルを用いた VR における UI 分析ツールの検討を行う。このモデルは、ポインティング操作時間の予測で広く知られるフィッツの法則 [6] を補完し、成功率に焦点を当てるものである。成功率推定には、VR 環境に特化した終点分布モデルである EDModel [23] の知見を用いる。最終的に、2D UI 分析ツール Tappy [25] のように実用的な形で提供することで、HCI 分野の理論的知見の VR コンテンツ開発の現場への応用を実現する。

2 関連研究

2.1 VR 空間でのポインティング

VR 空間でのポインティングには様々な選択手法が存在するが、特にコントローラーからレイを伸ばして対象を選択するレイキャストリングは、少ない身体的負荷で遠方のオブジェクトを操作できるため、広く採用されている。しかし、この手法はターゲットが小さい、または遠い場合に精度が低下する課題を抱えている [2]。この課題を解決するため、ポインティングを補助する様々なインタラクション技術 [4, 16] が提案されている。一方、ユーザのポインティング行動をモデル化し、パフォーマンスを理解・予測するアプローチも重要視されており [12]、多くの研究が Fitts の法則 [6] を用いて移動時間推定を行っている。

2.2 ポインティングの成功率

古くは Crossman や Welford が 1D タスクにおいて終点分布がガウス分布に従うことを示した [5, 17]。その後の研究では、終点分布が二変量ガウス分布に従うと仮定し、ポインティングの成功率を求め、難易度 (ID) に変換した報告 [7] や、有効ターゲット幅に基づいたマウスのポインティングにおけるエラー率モデルを構築し、そのモデルが 1 次元のバー形状のターゲットと 2 次元の円形ターゲットの両方に適用可能であることを示した報告がある [18, 19]。Bi らは指によるタップ座標の分布が、ターゲットの大きさに応じて変化する分布と、指の太さや身体的な操作精度に起因する絶対的な分布、という 2 つの独立したガウス分布の重ね合わせで説明できるという二重分布仮説 [3] に基づく Dual Gaussian Distribution Model を提案した [4]。このモデルを 3 次元や、移動ターゲットなど、様々なターゲット形状や条件に

Copyright is held by the author(s).

* 明治大学

† LINE ヤフー株式会社

応用したモデルが提案されている [8–10, 15, 22, 24].

その中でも EDModel [23] は, VR 環境におけるポインティング終点分布に特化したモデルである. このモデルは, ターゲットの幅 (W), 移動距離 (A), そして VR 特有の要素である奥行き (Z) が終点分布にどう影響するかを調査した. その結果, VR における終点分布もまた, 二変量ガウス分布で非常によく近似できることを示し, 各パラメータ (W , A) から分布の形状 (平均と共分散) を予測する回帰式を導出した. 分布の形状は回帰式を用いて式 (1) に示すような二変量ガウス分布に従う.

$$\mu = \begin{bmatrix} \mu_x \\ 0 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{bmatrix} \quad (1)$$

予測した分布の形状から, ターゲット D の選択成功率は式 (2) で求めることができる.

$$y = \iint_D \frac{1}{2\pi\sigma_x\sigma_y} \exp\left(-\left(\frac{(x-\mu_x)^2}{2\sigma_x^2} + \frac{y^2}{2\sigma_y^2}\right)\right) dx dy \quad (2)$$

しかし, 先行研究 [23] では, モデルを構成する複数のパラメータの組み合わせによって, 成功率の推定精度がどのように変化するかという点についての検証がなされていない. そのため, 本研究では異なるパラメータを持つモデルを用いて終点分布を予測し, 成功率の推定精度を比較検討する.

2.3 UI デザイン支援ツール

2D 平面上の UI デザインにおいては, 定量的な予測モデルに基づく実用的な支援ツールの開発が活発に行われている. Tappy [21, 25] は, Dual Gaussian Distribution Model を理論的基盤とし, スマートフォン用のウェブページやアプリケーションにおける UI 要素のタップ成功率を推定するツール群である. 本ツールは, テキストの周囲領域がタップイベントを受け付けているかなど, 外観からの判断が困難なタップ可能領域を正確に識別し, その成功率を表示する. Tappy を用いた評価によって, 開発者やデザイナーが想定よりもタップ成功率の低い UI 要素や実装上の誤りを発見するきっかけとなり, デザインの具体的な修正や, 定量的指標に基づいてチームでの意思決定を促進するなど, その有効性が示されている. また, Tappy をブラウザ上の拡張機能として実装した Tap Analyzer [11] も提供されている.

3 推定モデル構築のためのデータ収集実験

本研究の目的は, Tappy [25] を VR 上に拡張した成功率推定ツールの実現である. そこで, 既存の EDModel [23] を再検証するとともに, 成功率推定に有効な新たなパラメータを用いた代替モデルを検討するためのユーザ実験を行う.

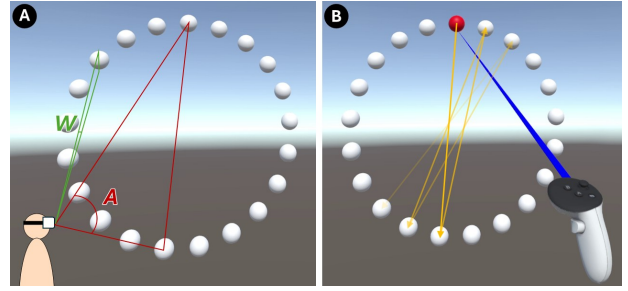


図 1. (A) ターゲットの幅 W , 移動距離 A の角度形式.
(B) 被験者は矢印で示された経路に従って, 順次ターゲットを選択した.

3.1 参加者・装置

18 歳から 24 歳 (平均: 21.0 歳) の大学生 18 名 (女性 2 名, 男性 16 名) が実験に参加した. 実験には Meta Quest 3 と標準の Touch Plus コントローラーを使用し, 実験プログラムは Unity で作成された. レイキャスティング手法を用い, トリガーにはコントローラーのボタンを使用した.

3.2 実験タスク

本実験では, ISO9241-411 の多方向タッピング課題 [14] のように, リング状に配置された 21 個の球状ターゲットを次々に選択するポインティングタスクを実施した (図 1A). 被験者は, 図 1B に示された順序に従ってターゲットを選択した. 選択が完了すると短い音が鳴り, 新たな目標ターゲットがハイライトされて次の試行が開始された. その後, 被験者はポインターを新たな目標ターゲットへ移動させる. 先行研究 [23] にならって, 選択の正誤を示すフィードバックは提供されなかった.

本実験では, ターゲット幅 W と移動距離 A を変化させた (図 1A 参照). これら 2 つの変数は, 角度形式で記述した [12, 13]. すなわち, W と A それぞれのなす角を各被験者の視点を基準に計算した.

また, 本研究では終点の誤差を物理的な距離ではなく角度を基準とした座標系で評価する. 本来, 3 次元空間上の一点は, 被験者からの距離 r と 2 つの角度 (θ, ϕ) を用いた球面座標系で表される. しかし, コントローラーから発せられるレイは無限の長さを持つと見なせるため, 距離 r の要素は無視できる. これにより, 終点の誤差の表現を, 2 つの角度成分のみを用いた 2 次元座標へと単純化できる.

そこで本実験では, 終点 p を 2 次元の角度座標 $p = (x, y)$ で定義した. ここで x は, ポインティングの移動方向 (開始ターゲットから目標ターゲットへの直線) に平行な角度の誤差, y はその移動方向に対して垂直な角度の誤差を表す. なお, 座標軸の原点は目標ターゲットの中心に設定しており, x と y は共に負の値を取り得る.

3.3 設計と手順

本研究では、2つの要因（ターゲットの幅 W と移動距離 A ）を用いた 8×3 の被験者内計画を採用した。ターゲットの幅 W は 1° から 4.5° まで 0.5° 刻みの8段階、移動距離 A は 30° 、 35° 、 40° の3段階とした。これらの値は先行研究 [23] および予備実験によって決定され、極端な条件下（ $W=4.5^\circ$ かつ $A=30^\circ$ ）でも配置されたリング上のターゲットが互いに重ならないことを確認した。被験者と各ターゲット間の距離は一定値（100m）とした。ターゲットの幅 W と移動距離 A の組み合わせの提示順序は、被験者ごとにランダムであった。先行研究 [23] にならない、各 W と A の組み合わせ（1つのリング）において、最初の試行は破棄され、20回の試行が分析対象として残った。最終的に、この実験から合計8640個のエンドポイント（ $8W \times 3A \times 18人 \times 20試行$ ）を収集した。

被験者は初めにVRデバイスと選択タスクについての説明を受けた。次に、VRヘッドセットを装着し、正面方向のキャリブレーションの後、練習試行を実施した。その後、本試行へ進んだ。実験の後に、被験者にアンケートへの記入を依頼し、年齢・性別・利き手の情報を収集した。

本実験では、各被験者が自身の快適なペースで各タスクを完了できるような選択行動を記録したいと考えたため、自然なスピードでターゲットを選択するように指示した。被験者には、試行間に小休憩を自由に取ることができると伝えた。

4 結果

4.1 データスクリーニングと正規性検定

実験において、8640試行のデータが得られた。先行研究 [23] に倣い、各被験者の各ターゲット条件において、平均移動時間および各試行の最初のクリック座標が各軸の中心から3標準偏差以上離れた外れ値を除去した結果、313試行が除外された。残ったデータを被験者ごと、条件ごとにグループ分けして分析した。

すべての終点座標は、コルモゴロフ-スミルノフ検定により、x軸およびy軸の両方において分布の正規性が確認された（ $\alpha = 0.05$ ）。次に、各集合に対して、x軸とy軸の両方で、最尤推定を用いてガウス分布の平均 μ と標準偏差 σ を推定した。2軸間の相関 ρ も計算した。したがって、5つの従属変数 μ_x 、 σ_x 、 μ_y 、 σ_y 、および ρ を用いて、平均ベクトル μ と共分散行列 Σ を計算した（式(1)参考）。

4.2 終点分布

5つの従属変数に対する反復測定分散分析（RM-ANOVA）を行った。自由度はGreenhouse-Geisser法により調整した。 W は μ_x （ $F_{2.81} = 9.95$ 、 $p <$

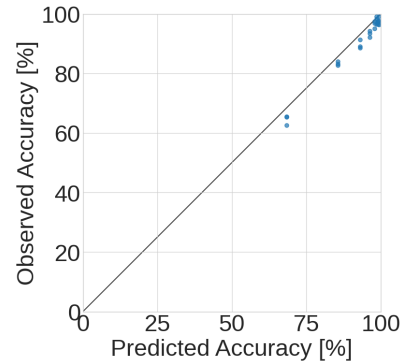


図 2. 球体ターゲットの観測成功率と、実験より導出した終点分布を用いて推定された成功率

0.001 、 $\eta_p^2 = 0.369$ ）と σ_x （ $F_{6.91} = 56.30$ 、 $p < 0.001$ 、 $\eta_p^2 = 0.768$ ）、 σ_y （ $F_{2.94} = 40.90$ 、 $p < 0.001$ 、 $\eta_p^2 = 0.706$ ）に有意な主効果を示した。その他の主効果や交互作用は見られなかった。

従属変数に主効果をもつ要因について線形回帰分析を実施した。回帰分析の結果、 W と μ_x （ $R^2 = 0.73$ ）、 W と σ_x （ $R^2 = 0.98$ ）、そして W と σ_y （ $R^2 = 0.89$ ）の間に線形関係が示された。

この結果に基づき以下の二変量ガウス分布を作成した。

$$\mu = \begin{bmatrix} eW + f \\ 0 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} (aW + b)^2 & 0 \\ 0 & (cW + d)^2 \end{bmatrix} \quad (3)$$

ここでそれぞれの定数は回帰の結果より $a = 0.1102$ 、 $b = 0.23130$ 、 $c = 0.0715$ 、 $d = 0.2311$ 、 $e = -0.0623$ 、 $f = -0.0846$ である。

4.3 成功率推定

次に、本実験より導出された二変量ガウス分布（式(3)）を使用して、式(2)に基づき各条件のターゲット選択の成功率を推定した。推定結果と観測値を図2に示す。観測された成功率と推定された成功率の最も大きい差は5.82%（ $W = 1.0^\circ$ 、 $A = 30^\circ$ ）で、平均絶対誤差（MAE）は2.35%、 R^2 は0.987であった。LOOCV（leave-one-out cross-validation）を行うと、 $MAE = 2.39\%$ 、 $R^2 = 0.985$ であった。

5 追加分析

本研究で提案するモデルの応用性と実用性を探るため、3つの主要なパラメータに関する追加分析を実施する。ここでの目的は、成功率推定ツールに適用した時の計算コストや使いやすさといった実用的な観点から、最適なモデル構成を明らかにすることにある。具体的には、以下の3つの観点からモデルの妥当性を検証する。

距離（A）の使用。 第一に、ターゲットまでの距離

表 1. 追加分析における各モデルの成功率推定精度の比較と LOOCV による汎化性能の評価

モデル	学習データ		LOOCV		AIC
	MAE (%)	R^2	MAE (%)	R^2	
ベースライン	2.354	0.987	2.394	0.985	82.50
距離パラメータ (A) あり	2.352	0.985	2.405	0.982	85.57
オフセット単純化	3.907	0.982	3.933	0.982	90.01
ワールド座標系	3.827	0.989	3.871	0.988	78.91

A を使用するモデルの妥当性を検証する。ハンドベースのレイキャスティングを用いた先行研究 [23] では、その終点分布予測モデルに距離 A が含まれていた。しかし、実験のタスクにおいては、RM-ANOVA の結果からその影響は限定的だった。また、距離を使用しないモデルの精度が十分であった場合、ターゲットのサイズのみを考慮し、成功率推定が行える。その結果、開発者はポインターとターゲット間の距離を考慮せずにツールを使用でき、使いやすさが向上する。そこで、距離を含むモデルと含まないモデルの性能を比較し、A を除外しても推定精度を損なうことなくモデルの簡潔性を高められるかを検討する。

オフセット (μ_x) の単純化。 第二に、中心からの誤差の平均 μ_x を 0 と仮定するモデルの妥当性を評価する。ポインティングの誤差に関するフィッツの法則に基づいた確率的モデル [14] などの多くの基本モデルは、ユーザは常にターゲットの中心を狙うと仮定し、中心からの誤差平均 (μ_x, μ_y) を 0 としている。本研究では、この仮定に基づくモデル (「 μ_x なし」モデル) と、ターゲット幅 W に応じて誤差の平均が動的に変動するモデル (「 μ_x あり」モデル) の性能を比較する。この比較を通じ、 μ_x を導入しない単純なモデルが実用上十分であることを検証する。

ワールド座標系の採用。 第三に、進行方向を基準とした座標系 (進行方向を x, その垂直方向を y) から、Unity 標準のワールド座標系へ置き換えた場合のモデルの有効性を検証する。進行方向基準の座標系は、先行研究 [23] でも使用されており、運動の解析に適していると考えられる。一方、ワールド座標系は、コントローラーのボタン押下時、力が加わって生じるポインターのズレなど、実世界における重力や慣性の影響を含めてモデル化できる可能性がある。加えて、ワールド座標系での成功率推定が可能であれば、ポインターの進行方向やターゲットの向きといった変動要素を考慮する必要がなくなり、ターゲットサイズのみで成功率を計算できるようになる。これは、応用ツールを開発する上で、計算コストを削減し、実装を簡素化できる利点をもたらす。

5.1 結果

上記 3 つの追加分析を実施し、本実験と同様の手順で分析を行った。各モデルにおける成功率の推

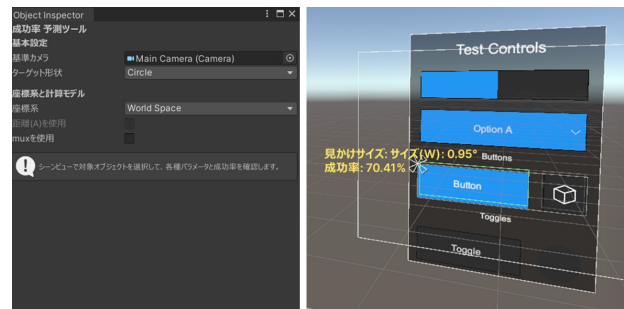


図 3. 左：成功率推定ツールのウィンドウ。ウィンドウ上で基準となるカメラ、ターゲット形状などのオプションを設定する。右：Unity のシーンビュー。あるオブジェクトをクリックすると、サイズ (単位：度)、選択成功率が表示される。

定結果を表 1 に示す。結果として、ベースラインとなるモデルと比較して推定精度は若干低下したものの、平均絶対誤差 (MAE) はいずれも 2-4% 程度に収まっていた。また、各モデルの AIC をベースラインと比較すると、その差は 10 未満であり、実用上十分に機能する推定が可能であることが確認された。よって、検討した 3 つのモデルはそれぞれ、精度への影響を許容範囲に留めつつ、使いやすさの向上、モデルの単純化、実装の簡素化や計算コストの削減といった利点をもたらす。そのため、応用場面に応じて各モデルを十分に利用可能であると考えられる。

6 成功率推定ツール

成功率推定モデルを VR コンテンツ開発者が容易に利用するため、VR 上の UI 分析ツールを開発した。これは、スマートフォンの UI 分析ツール「Tappy」[25] のコンセプトを VR に応用したものである。

このツールは Unity のエディタ拡張機能として実装されている。デザイナーや開発者は VR シーン的设计中に、任意の UI 要素 (ボタン、パネルなど) を選択することで、その選択成功率をリアルタイムに確認できる (図 3)。ツールは、選択された UI のサイズ、カメラからの距離 (奥行き)、想定されるポインティングの移動距離を自動で取得し、推定モデルに基づき成功率を計算・表示する。これにより、開

発者は「このボタンは遠すぎてクリックしにくい」「このメニューは小さすぎて誤操作を招く」といったユーザビリティ上の問題を、開発の初期段階で定量的に特定し、改善することが可能となる。

6.1 実装

先行研究 [25] および VR 環境の特性に基づき、VR 空間の UI の選択成功率を推定するツールの詳細な機能について述べる。

ターゲット形状の選択. 開発者は分析対象となるターゲットの形状を、球体または矩形から選択できる。本研究における実験は球体を対象としているため、球体はモデルの標準的な対応形状である。一方、矩形に関して本研究で直接的な実験は行っていない。しかし、実験において x 軸方向と y 軸方向の終点分布に相関が見られなかったため、各軸に対して独立に成功率を推定可能であると仮定し実装した。

また、矩形はポインターの進行方向によって有効な縦横サイズが変動する。この問題に対しては、追加分析によって得られた「ワールド座標系においても推定が可能である」という知見を用い、絶対座標系で矩形の寸法を計算する手法を導入した。

選択中心オフセットの考慮. 開発者は、ターゲットを指し示す際の座標（選択中心）を、ターゲットの幾何学的な中心と仮定するか否かを選択できる。実験では、ターゲットサイズが大きくなるにつれて、選択中心がターゲット中心からユーザの進行方向手前側へ偏移する傾向が確認されている。一方で、追加分析により、選択中心をターゲット中心と仮定した場合でもモデルの有効性が保たれることが示されているため、システムはこの仮定の有無を選択できるオプションを提供する。

ターゲットまでの距離の考慮. ポインターの初期位置からターゲットまでの距離を、成功率の算出に含めるか否かを選択するオプションを提供する。追加分析の結果、ターゲットまでの距離の有無によって推定精度に大きな差異は生じないことが明らかになっている。しかし、本研究では、距離が大きくなるほど選択成功率がわずかに低下する傾向も示されている。そのため、この影響を考慮したい開発者向けに本オプションを提供する。

6.2 使用方法

まず、開発者が基準となるカメラと、成功率を推定したいターゲットオブジェクトを選択する。するとエディタ上で指定されたターゲットの周辺にそのサイズと選択成功率が表示される（図 3 右）。これらの手順に加え、開発者は前節で述べたオプションを選択する（図 3 左）ことで、目的に合わせた推定モデルを使用することが可能である。

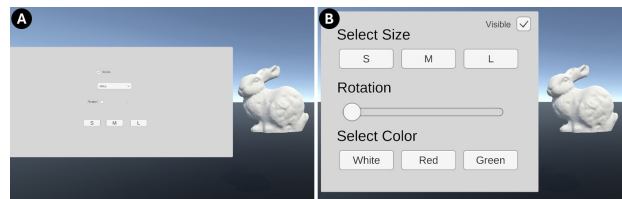


図 4. (A) 調整前の UI (B) 参加者が調整した後の UI. キャンバスは左手のコントローラーに紐づけられた。

7 開発経験者を対象としたユーザスタディ

本稿で提案する成功率推定ツールの有用性を検討するため、VR ゲーム開発経験がある 2 名を対象にユーザスタディを実施した。参加者の経歴を考慮し、それぞれ異なる形式でツールの試用とフィードバックの提供を依頼した。本ユーザスタディは 2 名の経験者を対象とした小規模なものであり、提案システムの有用性を網羅的に検証するものではなく、今後の改善に向けた質的な示唆を得ることを目的とした簡易的な調査である。

7.1 制御環境下でのユーザスタディ

参加者の 1 名は、過去 1 年間の VR 開発経験があるが、現在は VR 開発に従事していなかった。そのため、実環境でのツールの使用ではなく、制御下のタスクにおけるツールの有効性の調査を目的とした。

まず、参加者にツールの基本的な操作方法（UI 要素を選択し、成功率を確認する手順）を説明した。次にタスクとして、3D モデル（ウサギ）の色や大きさを調整するサンプル UI（図 4A）を提示し、「本ツールで表示される選択成功率を参考に、UI をより操作しやすく改善してください」と依頼した。参加者がタスクを完了した後の UI を、図 4B に示す。タスク完了後には、ツールが実際の UI 改善プロセスにおいてどのように役立ったか、また実用上の課題は何かについて意見を聴取した。

調査中、参加者はまず UI パネル上の各ボタンやスライダーを視認し、それぞれの機能的な重要度を判断した。例えば、色を変更するボタンは頻繁に使うが、3D モデルの表示・非表示を切り替えるボタンはあまり使わないといった判断が見られた。その後、本ツールを使い、個々の UI 要素の選択成功率を数値で確認した。使用頻度が高いにもかかわらず成功率が低いと表示されたボタンを発見すると、ツールで成功率を表示させながら、適切な大きさまで要素を拡大させる操作をした。重要度が高いと判断したボタン群については、参加者が「ツールで成功率を確認した後、ボタンサイズを 95% 以上の成功率が表示されるまで拡大する」という具体的な選択成功率の目標を設定して作業を進める様子が観察された。

参加者からは、実際に HMD を装着して何度も実

行・確認する手間をかけずに、選択成功率を分析し、UI を修正できる点が有用であるとの意見が得られた。また、開発者の主観や操作への慣れで見過ごされがちな問題を客観的な数値で判断できるため、UI 開発の一つの指標になるとされた。具体的な活用場面として、品質保証 (QA) フェーズの前段階で、プランナーやエンジニアが UI の操作性を議論する際に成功率などの数字が確認できると、一つの設計指針になるという意見があった。

7.2 実開発環境でのユーザスタディ

もう 1 名の参加者は、VR 開発経験 5 年の VR-Chat ワールド制作者である。参加者には 1 週間ツールを試用してもらい、アンケート形式でフィードバックを得た。アンケートはツールの使用感や今後の導入可能性、さらに普段の UI 開発における課題などについて、主に自由記述形式で回答を求めた。

アンケートで参加者は普段、UI のサイズや配置を過去の制作物での経験や体感といった主観的な尺度に頼って調整しており、本ツールがこれまで直感に頼りがちだった UI 設計に客観的な指標を与えた点が有用であるとした。特に UI 設計の初期段階で UI の大きさを決定する際に有用であるとし、今後の開発ワークフローにも導入したいと回答した。本ツールが提供する指標は、開発者ごとの設計判断のばらつきを抑制し、見た目は問題ないが、実は操作しにくい UI の削減に貢献できると期待された。また、本ユーザスタディは VRChat の UI 開発という環境で行われた。そこでは、VRChatSDK の仕様により Unity エディタ上での実機デバッグが困難であるという技術的制約が存在する。本ツールは、その影響を軽減し、VR 環境を介さずに UI のユーザビリティを検証する有効な手段となり得ることが示された。さらに、VRChat ではユーザが使用するアバターの身長が多岐にわたるため、ユーザによって視点が異なり、あらゆる身長のユーザにとってアクセスしやすい UI を設計することが困難という課題も存在する。本ツールを用いることで、複数の身長モデルで実機デバッグを繰り返すことなく、身長差が UI のアクセシビリティに与える影響を定量的に把握でき、この課題の解決にも寄与する可能性が確認された。

8 議論

本研究では、いくつかの前提条件のもとでモデルを構築しており、その一般化可能性については今後の課題が残されている。これは、ツールを多様な UI へ適用する上で不可欠な要素である。第一に、モデルの基盤となる実験設定の限界がある。本実験はユーザから等距離の円周上に配置された球体ターゲットを対象とした。しかし、実用的な UI はユーザからの距離や視野角が常に変動する平面上に配置されるこ

とが多い。加えて、視野の中心部と周辺部ではポインティング精度が異なる可能性もあり、本モデルを適用する際の予測精度については慎重な検討が求められる。また、ツールでは実験結果から各軸を独立に推定できると仮定し、矩形にも対応している。この仮定の妥当性を含め、本実験と異なる条件でのモデルの適用可能性は今後の重要な検証課題である。

第二に、被験者属性の偏りが挙げられる。本実験の参加者は平均年齢 21.0 歳の大学生に限定されていた。加齢による運動能力の変化や、VR 経験の有無がポインティング精度に与える影響は考慮されていない。より多様なユーザ層に対してモデルの一般化可能性を確保するためには、異なる年齢層や VR 習熟度のユーザを含めた追加検証が必要である。

第三に、入力デバイスと選択手法が限定的である点が挙げられる。本研究はコントローラーを用いたレイキャスティング方式を前提としている。ボタン押下時にポインターがずれるハイゼンベルク効果 [20] のように、入力デバイスの物理的な形状が選択精度に与える影響は無視できない。本モデルが異なるデバイスや選択方法で同様の精度を持つかは保証されない。さらに、バーチャルハンドやアイトラッキングといった他のポインティング手法に本研究の知見を適用できるかは、別途検証が必要である。

本研究では、開発したツールの有用性を検証するためにユーザスタディを実施した。その結果、本ツールが UI 設計プロセスに客観的な指標を提供し、改善のきっかけとなり得ることが示唆された。具体的には「これまで直感で設定していた UI のサイズに客観的な目安ができた」といった肯定的な意見が得られた。このことから、本ツールは HCI の知見と開発現場のギャップを埋める可能性があり、特に実機検証に労力を要し、身長や姿勢といった個人差が UI の操作性に影響する AR/VR などの XR 開発における有用性が示唆された。しかし、本ユーザスタディはツールの改善に向けた意見収集を目的としており、参加者数が少ない点や、参加者が経験者に限定されるという制約がある。そのため、定量的かつ網羅的な評価のためには追加の大規模な実験が必要であると考えられる。

今後の展望として、エディタ上での使用のみならず、HMD を装着した状態でツールを使用し、現在のポインターの位置から各 UI への選択成功率を表示する機能拡張などが考えられる。また、本システムを用いて設計した UI と利用せずに設計した UI を比較し、設計時間の短縮や品質向上の観点からシステムの有用性を定量的に評価することも考えられる。さらに、UI 設計支援にとどまらず、アクションゲームにおいて、ブロックのサイズや配置を調整することで、難易度を定量的に設計・分析するツールとしても応用が可能であると考えている。

参考文献

- [1] M. Amini, W. Stuerzlinger, R. J. Teather, and A. U. Batmaz. A Systematic Review of Fitts' Law in 3D Extended Reality. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, CHI '25, New York, NY, USA, 2025. Association for Computing Machinery.
- [2] F. Argelaguet and C. Andujar. A survey of 3D object selection techniques for virtual environments. *Computers & Graphics*, 37(3):121–136, 2013.
- [3] X. Bi, Y. Li, and S. Zhai. FFitts law: modeling finger touch with fitts' law. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, p. 1363–1372, New York, NY, USA, 2013. Association for Computing Machinery.
- [4] X. Bi and S. Zhai. Predicting Finger-Touch Accuracy Based on the Dual Gaussian Distribution Model. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, UIST '16, p. 313–319, New York, NY, USA, 2016. Association for Computing Machinery.
- [5] E. R. Crossman. *The measurement of perceptual load in manual operations*. PhD thesis, University of Birmingham, 1956.
- [6] P. M. Fitts. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology*, 47(6):381, 1954.
- [7] T. Grossman and R. Balakrishnan. A probabilistic approach to modeling two-dimensional pointing. *ACM Trans. Comput.-Hum. Interact.*, 12(3):435–459, Sept. 2005.
- [8] J. Huang, F. Tian, X. Fan, H. Tu, H. Zhang, X. Peng, and H. Wang. Modeling the Endpoint Uncertainty in Crossing-based Moving Target Selection. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, p. 1–12, New York, NY, USA, 2020. Association for Computing Machinery.
- [9] J. Huang, F. Tian, X. Fan, X. L. Zhang, and S. Zhai. Understanding the Uncertainty in 1D Unidirectional Moving Target Selection. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, p. 1–12, New York, NY, USA, 2018. Association for Computing Machinery.
- [10] J. Huang, F. Tian, N. Li, and X. Fan. Modeling the Uncertainty in 2D Moving Target Selection. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, UIST '19, p. 1031–1043, New York, NY, USA, 2019. Association for Computing Machinery.
- [11] K. Kobayashi. Tap Analyzer, 2024.
- [12] R. Kopper, D. A. Bowman, M. G. Silva, and R. P. McMahan. A human motor behavior model for distal pointing tasks. *International Journal of Human-Computer Studies*, 68(10):603–615, 2010.
- [13] J. Petford, M. A. Nacenta, and C. Gutwin. Pointing all around you: selection performance of mouse and ray-cast pointing in full-coverage displays. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2018.
- [14] R. W. Soukoreff and I. S. MacKenzie. Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research in HCI. *International journal of human-computer studies*, 61(6):751–789, 2004.
- [15] H. Usuba, S. Yamanaka, J. Sato, and H. Miyashita. Predicting Touch Accuracy for Rectangular Targets by Using One-Dimensional Task Results. *Proc. ACM Hum.-Comput. Interact.*, 6(ISS), Nov. 2022.
- [16] L. Vanacken, T. Grossman, and K. Coninx. Exploring the Effects of Environment Density and Target Visibility on Object Selection in 3D Virtual Environments. In *2007 IEEE Symposium on 3D User Interfaces*, 2007.
- [17] A. T. Welford. Fundamentals of skill. 1968.
- [18] J. O. Wobbrock, E. Cutrell, S. Harada, and I. S. MacKenzie. An error model for pointing based on Fitts' law. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, p. 1613–1622, New York, NY, USA, 2008. Association for Computing Machinery.
- [19] J. O. Wobbrock, A. Jansen, and K. Shinohara. Modeling and predicting pointing errors in two dimensions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, p. 1653–1656, New York, NY, USA, 2011. Association for Computing Machinery.
- [20] D. Wolf, J. Gugenheimer, M. Combosch, and E. Rukzio. Understanding the Heisenberg Effect of Spatial Interaction: A Selection Induced Error for Spatially Tracked Input Devices. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, p. 1–10, New York, NY, USA, 2020. Association for Computing Machinery.
- [21] S. Yamanaka, H. Usuba, J. Sato, N. Sasaya, F. Yamashita, and S. Yamaguchi. Tappy Plugin for Figma: Predicting Tap Success Rates of User-Interface Elements under Development for Smartphones, 2024.
- [22] S. Yamanaka, H. Usuba, H. Takahashi, and H. Miyashita. Servo-gaussian model to predict success rates in manual tracking: Path steering and pursuit of 1d moving target. In *Proceedings of the 33rd annual ACM symposium on user interface software and technology*, pp. 844–857, 2020.
- [23] D. Yu, H.-N. Liang, X. Lu, K. Fan, and B. Ens. Modeling endpoint distribution of pointing selection tasks in virtual reality environments. *ACM Trans. Graph.*, 38(6), Nov. 2019.
- [24] Y. Zheng, J. Huang, H. Zhang, Y. Bian, J. Liu, C. Yang, F. Tian, and X. Meng. 3D Ternary-Gaussian model: Modeling pointing uncertainty

of 3D moving target selection in virtual reality. *International Journal of Human-Computer Studies*, 198:103454, 2025.

[25] 山中 祥太, 薄羽 大樹, 山下 郁矢, 笹谷 奈翁美, 佐

藤 潤一. スマートフォン用ウェブページとアプリにおけるタップ成功率推定ツール Tappy の実用化. 情報処理学会シンポジウムインタラクション, No. 014 in INT25, pp. 119–128, 2025.

座位での物理的な制約の緩和を目指した 知覚されない閾値に基づいたリダイレクション手法の提案と評価

小林 広夢* 鈴木 湧登* 日下部 完* 坂本 大介*

概要. VR空間内では、上半身の回旋を伴うオブジェクトの把持・移動が頻繁に行われるが、座位では背もたれなどの障害物で回旋動作が困難となるなど、物理的な制約が大きい。そこで本研究では、頭部と胸部（上肢）の回旋量をユーザが知覚できない範囲で同時に増幅するリダイレクション手法を提案する。本手法は認知負荷とサイバー酔いを抑制しつつ物理的な回旋量を削減することで、これらの物理的な制約の緩和を目指す。まず提案手法の設計のため、回旋量の増幅がユーザに知覚されない閾値を調査する実験を行い、頭部と胸部の回旋を同時に約1.43倍まで増幅しても、ユーザに知覚されないことが示唆された。この知見に基づき提案手法を実装し、主観的な作業負荷・サイバー酔いの程度と物理的な回旋量削減への影響を評価した。その結果、提案手法は、主観的な作業負荷・酔いの程度に統計的な差は確認されなかったが、これまで困難だった「視認性向上」と「座位における物理的な制約の緩和」を両立できる可能性が示唆された。

1 はじめに

Virtual Reality (VR) はゲームなどのエンターテインメント分野で広く普及し始めている。多くのVRアプリケーションでは、上肢や胸部、頭部を動かし、ユーザを囲むように配置された仮想オブジェクトを把持・移動するといった直接作用するインタラクションが求められる [1, 2]。

しかし障害物（机など）の存在など、利用する環境により仮想オブジェクトとのインタラクションが困難であることがある。特に座位は下半身の動きが妨げられ、加えて背もたれが障害物となり、動きが制限されるため、立位と比較して並進移動や回旋動作が困難となる [3, 4]。並進移動に対しては、上半身の傾きに合わせてVR空間を移動する手法などが提案されてきた [5, 6]。回旋動作に対しては、回旋角度を2倍にして、VR映像として提示させる手法などが提案されている [7, 8, 9]。しかし回旋動作に関する手法は、平衡感覚と視覚情報間で齟齬が生じ、ユーザの認知負荷を高め、サイバー酔いを引き起こす可能性が指摘されている [10, 11]。

そこで、回旋動作における認知負荷とサイバー酔いを抑制しつつ、本来の課題である座位での動きの制限を緩和するため、リダイレクション手法の利用が提案されている [12, 13]。リダイレクションとは、ユーザの物理的な運動方向・量を知覚されない範囲で変容させ、VR映像として提示することで、無自覚に行動変容を促す手法である。これを利用し、頭部の物理的な回旋量と比較して、VR空間内での頭部回旋量を知覚できない程度に増幅することで、VR空間内での周囲の視認性向上を目指した手法である。VR空間での視覚操作がユーザに知覚されず、認知負荷やサイバー酔いを抑制につながる。しかし、このリダイレクションを用いた手法は頭部の回旋のみを増幅対象としており、胸部や上肢は連動しない。そのため、頭部回旋のみの増幅だけでは上肢の物理

的な動作量を削減できずに、上肢と障害物との干渉は依然として大きいままである。

そこで本研究では「上半身回旋時に、背もたれと上肢が干渉すること」と「座面により、下半身の動きが制限され、回旋が困難であること」という2点の物理的な制約の緩和を目指し、頭部と胸部（上肢）の回旋動作に同時にリダイレクションを適用する手法（図1）を新たに検討する。本手法は、ユーザがリダイレクションを知覚できない程度に頭部と胸部（上肢）の仮想的な回旋量を増幅する。これにより、認知負荷とサイバー酔いを抑制しつつ、物理的な回旋量を削減し、背もたれなどによる物理的な制約の緩和を目指す。本研究では、ユーザがリダイレクションを知覚できない程度に頭部と胸部（上肢）の回旋量を増幅する手法を設計し、かつ手法が実際に有用かを確認する。これを設計するためにまず頭部と胸部（上肢）に同時にリダイレクションを適用する際、ユーザに知覚されない回旋増幅の閾値（知覚閾値）はどの程度かを調査する実験を行う。その後、提案手法が主観的な認知負荷とサイバー酔いの増大を抑止しつつ、物理的な回旋量を削減可能かを評価する実験を行う。2つの実験を通して提案手法が座位によるVR空間でのインタラクションにどのように貢献するのかを検討する。

2 関連研究

2.1 VR空間内での可動性向上に向けた研究

座位での活動は下半身が固定され、かつ背もたれが障害物となるため、動作制約が大きい。この制約を緩和し、VR空間内での可動性を向上させるため、様々な研究が行われてきた [5, 6, 7, 8, 9, 12, 13]。例えば、上半身の傾きに合わせてVR空間を移動する手法 [5] や、VR空間内の手を仮想的に伸ばし、遠方のVRオブジェクトとのインタラクションを可能にする手法 [6] が挙げられる。その中でも頭部の運動量や回転量を操作し、VR空間内の視認性を向上させる手法がある [7, 8, 9, 12, 13]。これらの研究では主に Dynamic Rotation Gains と Scrolling と呼ばれ

Copyright is held by the author(s).

* 北海道大学

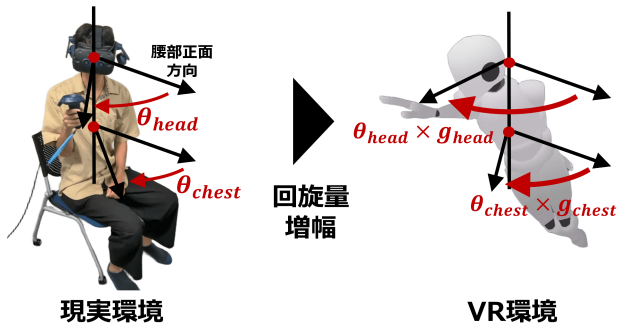


図 1. 上半身のヨー方向回旋量操作の概略図 (上) . 現実環境での腰部-胸部間角度 θ_{chest} と腰部-頭部間角度 θ_{head} (左) のそれぞれの角度に胸部回旋ゲイン g_{chest} と、頭部回旋ゲイン g_{head} を乗じて、VR 環境での各回旋角度を操作する (右) .

る 2 つの手法が提案されている [7, 8, 9]. 前者は頭部のヨー方向の回旋量に回旋倍率を乗じ、VR 環境の視界に反映させる手法である. これにより頸部の可動域では視認が困難な後方も視認可能となる. 後者の Scrolling 手法は、頭部の回旋角度が特定の閾値を超えると、同一回旋方向へ視界が自動で回転する手法である. これらの手法は、僅かな頭部の回旋動作でより広範な視野を確認可能にした.

しかし、これらの手法は平衡感覚と視覚情報間の不一致が増大し、ユーザの認知負荷やサイバー酔いが増大することが指摘されている [10, 11]. そこで、認知負荷やサイバー酔いを抑制する手法として、リダイレクション手法の利用が提案されている [12, 13, 14, 15].

2.2 リダイレクション手法

リダイレクションとは、ユーザの物理的な運動方向・量を知覚されない範囲で変容し、VR 映像として提示することで、疑似感覚を生起したり、無自覚な行動変容を促す手法である [12, 13, 14, 15, 16, 17, 18]. 代表的な研究では、リダイレクテッドウォーキングが挙げられる [16]. 現実環境で円弧状に歩行するユーザに対し、VR 環境では現実の動きと対応付けられた直進している映像を提示する. これにより直進歩行感覚を生起させる. 視覚と平衡感覚が不一致の場合に視覚が優位となる知覚心理学的特性 [10] を利用し、リダイレクション手法は VR 空間内の体験を現実として知覚させる. Steinicke らはこのリダイレクション手法を、「曲率操作」「並進移動量操作」「回旋量操作」の 3 つの基本操作に分類している [19].

本研究では上半身の仮想的な回旋量を増幅し、物理的な回旋量を削減することで、物理的な制約を緩和するため、基本操作の 1 つである回旋量操作を使用した. 回旋量操作とは、VR 空間内での回旋量を物理的な回旋量に対して増幅または減衰させる操作である. 回旋ゲインと呼ばれる係数を導入することで実現する. 回旋ゲインは物理空間での回旋量を VR 空間の回旋量に変換する際の倍率を意味し、物理的な回旋量に回旋ゲインを乗じ、VR 環境での回旋量として反映させることで回旋量を操作する. 回旋ゲインの値が 1 では現実と VR 環境で回旋量は一致し、

1 より大きい値では回旋量が増幅され、1 より小さい値では減少する. この回旋量操作を人体の各部位の動作に用いることで、例えばわずかに首を動かすことで VR 空間内の後方を視認可能になるなど、可動域などの制約を緩和できる. この回旋量操作は、VR 空間内での視認性向上を目指し後述する頭部の回旋に適用する研究が行われている [12, 13, 14].

2.3 頭部のリダイレクション

頭部のリダイレクションでは、物理的な頭部の回旋角度に対して、回旋ゲインを乗じて増幅した仮想的な回旋時の視界を、ヘッドマウントディスプレイ (HMD) を通じてユーザに提示する [12, 13, 14, 15]. これにより VR 空間内での視認性向上を目指す. 基本原理は Dynamic Rotation Gains と同じだが、リダイレクションでは操作がユーザに知覚されない範囲で行われるため、視覚・平衡感覚間の齟齬が小さく、認知負荷やサイバー酔いを低減できる.

しかし、既存のリダイレクション手法は頭部回旋のみに着目している. よって、頭部の回旋増幅に胸部や上肢が連動せず、上肢や手部の物理的な運動量の削減は難しい. そのため、障害物との干渉を十分に緩和できないという課題があった. そこで本研究では、知覚できない範囲で頭部と胸部 (上肢) の回旋量を増幅することで、認知負荷とサイバー酔いを抑制しつつ、物理的な回旋量を削減し、障害物との干渉を低減する手法を検討する.

3 提案手法の検討

本研究では認知負荷とサイバー酔いを抑制しつつ、座位時の物理的な制約を緩和することを目指し、頭部と胸部の回旋量を知覚できない程度に増幅する手法を提案する (3.1 節). この際、回旋増幅がユーザに知覚されない点は認知負荷とサイバー酔いを抑制するための手法設計に重要な要件である. 提案手法の回旋倍率をユーザに知覚されない適切な値にするため、頭部と胸部の回旋量を増幅した際に、ユーザが回旋増幅を知覚しない上限、すなわち知覚閾値を調査する実験も、設計に際して併せて行う (3.2 節).

3.1 「頭部+胸部」回旋動作リダイレクション手法

前提として本研究では、回旋動作について、Nasr らの上半身の身体モデリングを参考に腰部を回旋基準とし、胸部と頭部を独立したヨー方向の回旋軸と定義した [20]. また胸部と上肢の回旋運動は連動するため [21], 上肢の回旋運動は胸部の回旋量と同期させた. 以降に、頭部・胸部・上肢の各回旋動作を増幅する方法を記述する.

胸部の回旋増幅については、腰部回旋角度を回旋角度 0 とし、腰部回旋角度と胸部正面方向がなす角度を胸部回旋角度 θ_{chest} とした. そして、回旋量操作における胸部回旋ゲイン g_{chest} を乗じた $\theta_{chest} \times g_{chest}$ を、VR 空間内の回旋角度として反映させることで、胸部の回旋増幅を実現した (図 1).

頭部の回旋も、腰部回旋角度と頭部正面方向とがなす角度を頭部回旋角度 θ_{head} とし、頭部回旋ゲイン g_{head} を乗じて、VR 空間内の仮想的な回旋角度として反映することで頭部回旋を増幅した (図 1).

上肢の回旋運動の増幅は、胸部と上肢の回旋運動が連動しているため [21]、胸部回転ゲインによる回転量操作が作用する。胸部回旋角度と手部方向とがなす角度を手部回旋角度 θ_{hand} とすると、現実環境における腰部回旋角度と手部回旋角度とがなす角度 $\theta_{hand} + \theta_{chest}$ は、VR 環境では $\theta_{hand} + \theta_{chest} \times g_{chest}$ として反映させることで、上肢に対するリダイレクションを実現した (図 2)。この際、胸部の回旋と同期しない手部の独立的な運動は増幅しなかった。

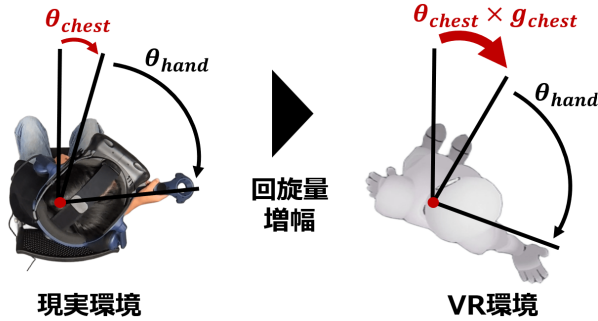


図 2. 上肢のヨー方向回転量操作の概略図。現実環境での腰部—手部間角度 $\theta_{hand} + \theta_{chest}$ (左) に対して、胸部回転ゲイン g_{chest} が作用することで、VR 環境での上肢のヨー方向回転を操作する。(右)

これらの回旋増幅操作は、認知負荷とサイバー酔いを抑制するため、ユーザが知覚しない範囲で行う必要がある。そのため、各回転ゲイン (g_{chest} , g_{head}) の知覚閾値を探索した (3.2 節)。

3.2 知覚閾値調査実験

認知負荷とサイバー酔いを抑制する手法を設計する際に必要となる、ユーザがリダイレクションによる回旋増幅を知覚しない閾値を調査する。

3.2.1 参加者と使用機器

本実験の参加者は 12 名だった (男性 9 名、女性 3 名、平均年齢 22.92 歳、標準偏差 1.75 歳)。実験では、HMD、コントローラーは HTC VIVE Pro2 を、トラッカーは VIVE トラッカー (3.0) を用いた。ソフトウェアは Unity アプリケーションを用いた。

3.2.2 実験デザイン

実験は参加者内計画として設計し、独立変数は以下の 2 つである。

1. 頭部回転ゲイン: (調査した値: 1/1.60, 1/1.40, 1/1.20, 1.00, 1.20, 1.40, 1.60)
2. 胸部回転ゲイン: (調査した値: 1/1.90, 1/1.60, 1/1.30, 1.00, 1.30, 1.60, 1.90)

頭部回転ゲイン、胸部回転ゲインとは、頭部、胸部でのヨー回転量を VR 環境に反映する際の回転量の倍率を意味する。各独立変数の調査する値は予備的に調査して決定した。各ゲインの 1 以下の値は先行研究に倣い [22]、1 以上の値の逆数とした。本実験の目的は VR 環境での上半身の回旋増幅がユーザに知覚されない閾値を調査することである。そのため頭部回転ゲインと胸部回転ゲインどちらか一方が角度増幅を引き起こし、他方が減衰を生じさせる組み合わせを調査する必要性が低いと判断し、頭部回

転ゲインと胸部回転ゲインともに 1 以上の全組み合わせと、頭部回転ゲインと胸部回転ゲインともに 1 以下の全組み合わせの計 31 種類の組み合わせを調査した (表 1)。組み合わせの提示順はランダムな順で提示した [12,13]。31 種類の組み合わせをランダムに並べ替えたタスク群を 1 セッションとし、参加者 1 人当たり 10 セッションの計 310 タスクを行った。

表 1. 独立変数の組み合わせ (丸が今回検討した独立変数組み合わせ)

		胸部回転ゲイン						
		1/1.90	1/1.60	1/1.30	1	1.30	1.60	1.90
頭部回転ゲイン	1/1.60	○	○	○	○			
	1/1.40	○	○	○	○			
	1/1.20	○	○	○	○			
	1.00	○	○	○	○	○	○	○
	1.20				○	○	○	○
	1.40				○	○	○	○
	1.60				○	○	○	○

従属変数は、角度増幅と減衰のどちらか一方が提示された際に「ユーザが増幅もしくは減衰を正しく判断できた正答率」である。先行研究に倣い [12,13]、二肢強制選択法という物理的な刺激に対する人間の感度を定量的に求める手法に沿って従属変数を設定した。二肢強制選択法ではまず、ある物理的な刺激を参加者に提示する。その後、参加者は刺激の程度について事前に用意された 2 つの回答のうち 1 つを選ぶ。そして、刺激に対して正しい回答をした割合を算出する。本研究における提示した物理的な刺激は、回旋量操作により物理的な回旋時とは異なった VR 環境の視覚情報である。事前に用意した回答は、VR 空間内での回旋量が物理的な回旋時と比較して「増幅」していたか、「減衰」していたかの 2 択とした。ここで「増幅」とは、物理的な回旋量よりも VR 内での回旋量が大きい状態を指し、「減衰」とは小さい状態を指す。これらのことから従属変数は、増幅もしくは減衰を正しく回答した正答率とした。

3.2.3 実験タスク

参加者は VR 空間内に配置されたオブジェクトに対し、上半身の回旋を伴う把持及び移動を行った。その後、参加者は自身の物理的な回旋量と比較して VR 空間内での視覚的な回旋量が「増幅」していたか「減衰」していたかを回答した。参加者に HMD を通して提示される視界は、物理的な頭部の動きと同期しつつも、各タスクごとにランダム順で設定した倍率で回旋量が増幅または減衰された。椅子は動かないよう固定され、VR 環境は野外風景を採用した [13]。オブジェクトは参加者の目の高さで奥行 0.5m、正面から左右に 90~110 度の範囲にランダムに配置し、直径 14cm の赤色球とした [9,13]。

タスクの詳細について、まず参加者は座位で正面を向く。次に、VR 空間内の左右に配置した 2 つの赤色球に、それぞれ上半身を回旋させて把持し、正面に移動する。左右 2 つの赤色球を正面の再配置場所に適切に移動すると、直前の回旋動作が「増幅していた」か「減衰していた」かを問う画面が提示される。参加者はこの問いに対してコントローラーを用いて回答する。ここまでの流れが 1 タスクである。

3.2.4 実験手順

参加者は周囲に障害物がない実験室内で椅子に着座したのち、年齢、性別などを尋ねる事前質問票に回答した。次に実験タスク内容を説明し、その後、腰部と胸部にトラッカーを、頭部と手部にHMDとコントローラを装着した。実験機器やタスクへの理解のため、練習タスクを実施した。本番タスクと同じ条件でゲイン値のみを変更し、参加者がタスクを理解したと申告するまで行った。その後、本番タスクを10セッション行った。セッション間には任意で休憩を設けた。実験全体の所要時間は約75分だった。

3.2.5 分析方法

二肢強制選択法で得られた正答率データに対し、3次元のロジスティック回帰分析を行う。基底関数には、物理的な刺激に対する人間の感度はシグモイド性を持つことから[23,24]、シグモイド関数を使用した。基底関数の数はk-分割交差検証(評価指標：負の対数尤度)を行い4つとし、最尤推定を用いた[12,13,14]。その後、分析で得られた回帰曲面から知覚閾値を算出する。知覚閾値は正答率75%時の回転ゲイン値とした[12,13,16]。しかし本研究では、3次元データとなるため正答率75%の閾値は頭部回転ゲインと胸部回転ゲインの組み合わせで構成される曲線として算出される。よって、この曲線上から提案手法に用いる適切な組み合わせを定める必要がある。身体部位の位置や運動に関する自己受容感覚と視覚情報との齟齬は、酔いや認知負荷の増大につながることを示唆されていることから[25]、本研究では頭部と胸部のゲイン値の差に起因する感覚間の齟齬を避けるため、閾値曲線上で頭部回転ゲインと胸部回転ゲインが等しくなる組み合わせを提案手法に用いる閾値として採用した。

3.3 実験結果

図3に3次元の「増幅」選択割合回帰曲面グラフを示す。また図4に二次元グラフ上で増幅操作時の正答率75%曲線を示す。頭部ゲインと胸部ゲインの値が等しくなる閾値曲線上の点は(頭部ゲイン, 胸部ゲイン) = (1.43, 1.43)だった。閾値曲線上で頭部ゲインの値が1の場合、胸部ゲインの値は1.90より大きく、収集したデータ群の外挿をとった。胸部ゲインが1の場合は頭部ゲインは1.46であった。

結果より、認知負荷やサイバー酔いの増大が小さいと考えられる、頭部回転ゲインと胸部回転ゲインがともに1.43となる組み合わせを、提案手法における回転倍率として採用した。

4 評価実験

導出した知覚閾値をもとに設計した提案手法が認知負荷・サイバー酔いの増大を抑制しつつ、頭部・胸部・手部の物理的な動作量を削減し、障害物との干渉の低減に寄与できるか調査する。

4.1 実験参加者と使用機器

本実験の参加者は12名だった(男性9名、女性3名、平均年齢22.92歳、標準偏差1.75歳)。実験機器は3.2節の実験と同じものを使用した。

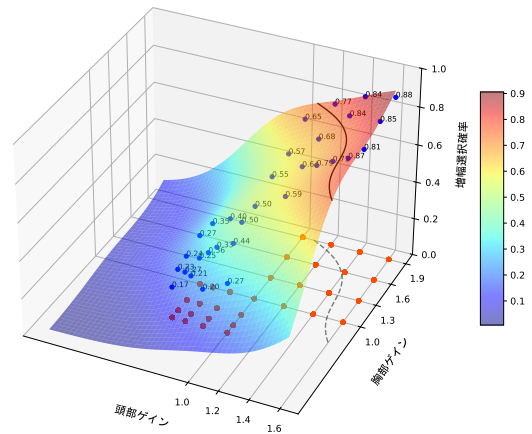


図3. 増幅回答割合の3次元ロジスティック回帰曲面(虹色曲面)。赤色点は独立変数組み合わせ、青色点は各独立条件組み合わせ時の増幅選択割合、黒色曲線は角度増幅操作時の正答率75%曲線である。

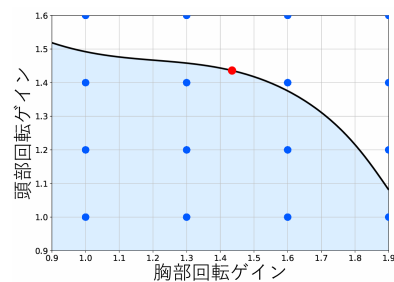


図4. 増幅操作時の正答率75%曲線(黒色曲線)。縦軸と横軸がそれぞれ頭部回転ゲインと胸部回転ゲイン、青色点が調査したゲイン組み合わせ、赤色点が後の評価実験で用いるゲイン組み合わせ、薄い青色の範囲は、正答率75%以下の領域。

4.2 実験デザイン

本実験は参加者内計画として設計し、独立変数は以下の4つを設定した。

1. **頭部+胸部増幅条件**: 頭部と胸部の回旋を同時に増幅する。ゲインは3.2節の実験で求めた知覚閾値から、(頭部回転ゲイン, 胸部回転ゲイン) = (1.43, 1.43)とした。
2. **頭部のみ増幅条件**: 比較対象となる既存手法として設定した。ゲインは3.2節の実験の結果に基づき、(頭部回転ゲイン, 胸部回転ゲイン) = (1.46, 1.00)とした。
3. **胸部のみ増幅条件**: 胸部のリダイレクション適用効果を分離・評価するために設定した。ゲインは3.2節の実験の結果より、(頭部回転ゲイン, 胸部回転ゲイン) = (1.00, 1.90)とした。
4. **ベースライン条件**: 回旋増幅を一切行わない条件である。

条件の提示順序は学習や疲労の効果を相殺するため、カウンターバランスを用いて参加者ごとに変更した。

評価指標は「動作遂行コスト」「作業負荷」「サイバー酔い」の3点を採用し、各指標に従属変数を設定した。動作遂行コストの従属変数は、頭部・胸部の累積回旋角度、頭部・胸部の回旋範囲、及び手部の累積移動量を設定した。加えて、物理的な動作量が削減できれば、動作の効率化によりタスク速度も

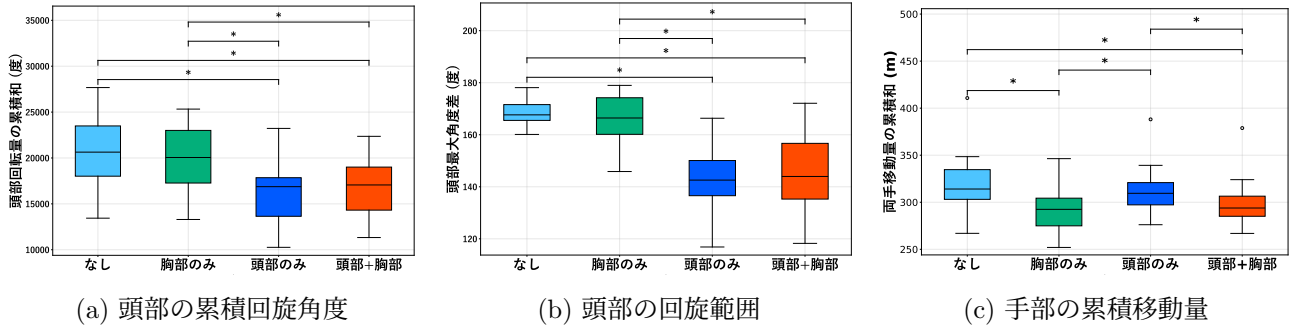


図 5. 頭部の累積回旋角度 (a), 頭部の回旋範囲 (b), 手部の累積移動量 (c) の条件間比較. 水色はベースライン条件, 緑色は胸部のみ増幅条件, 青色は頭部のみ増幅条件, 赤色は頭部+胸部増幅条件を示す.

向上するという仮説の元, タスク完了時間も設定した. 作業負荷は, 認知負荷や身体的負荷などの要素を包括的に含めた指標であり, 認知負荷を測定するために設定した. 従属変数には作業負荷を主観的に測定する NASA-RTLX のスコアを用いた [11]. サイバー酔いの指標には主観的な酔いの程度を測定する SSQ (Simulator Sickness Questionnaire) を用いた [13].

4.3 実験タスク

実験タスクは 3.2 節の実験のタスクと同様に, Langbehn らの先行研究を参考にし [7, 9, 13], 参加者の VR 空間内の周囲に配置した赤色球を, コントローラを用いて把持し, 正面に再配置するタスクを設定した. まず, VR 空間内の参加者の周囲に, 再配置対象となる赤色球を 1 つ配置する. 赤色球の位置は参加者の頭部座標を中心とする半径 0.5 メートルの円周上, かつ目の高さとした. 正面をヨー方向角度 0 度とし, -110 度, -80 度, -50 度, 0 度, 50 度, 80 度, 110 度の 7 つの離散的な方向からランダムに選択された. 参加者は提示された赤色球をコントローラのトリガー操作で把持し, 正面の半透明の球体で示した地点に移動することが求められた. 正しく再配置されると移動した赤色球は消滅し, 直ちに次の赤色球が前述のいずれかの角度に提示される. 参加者はこの一連の再配置操作を繰り返し行った. 各提示角度ごとに 8 回繰り返し, 合計 56 回 (7 角度 × 8 回) のタスクを 1 セッションとした. 各独立変数に対し 1 セッションずつ, 計 4 セッションを実施した.

4.4 実験手順

参加者は実験室に入室し着席した後, 参加者の年齢や性別などを回答する質問票と主観的な酔いの程度を測る SSQ に回答した. 次に, 参加者はタスク内容や使用機器, 注意事項に関する説明を受けた. この際, 各独立条件の内容は伏せられた. 説明後, 参加者は HMD (頭部), トラッカー (頭部および胸部), コントローラ (両手) を装着し, 着座した状態で実験に臨んだ. 実験は独立条件ごとにセッション単位で実施した. 各セッションの開始前には, 同一条件で 1 分の練習タスクを行った. 練習後, 各セッションを開始した. 各セッションの終了後, 参加者は NASA-RTLX, 及び SSQ に回答した. 回答後は休憩を挟み, 次の条件のセッションへ移行した. この一連の手続きを 4 つの独立条件全てにおいて繰り返した. 実験全体の所要時間は約 70 分だった.

4.5 分析方法

最初に, 全ての従属変数に対してシャピロ・ウィルク検定を行い, 正規性の仮定を満たすか確認した. ただし, NASA-RTLX と SSQ スコアは順序尺度であり, 非正規分布とみなした. 正規性が確認された変数は, パラメトリック検定として繰り返しのある一元配置分散分析を用いた. 球面性の仮定はモークリーの球面性検定で確認した. 分散分析で有意な主効果が見られた場合は, 下位検定にホルム法による補正を行った対応のある t 検定を実施した. 正規性が棄却された変数は, ノンパラメトリック検定に, フリードマン検定を用いた. フリードマン検定で有意差が示された際には, 下位検定としてホルム補正を行ったウィルコクソンの符号順位検定を実施した.

4.6 実験結果

4.6.1 評価指標: 動作遂行コスト

頭部の累積回旋角度, 頭部の回旋範囲, 胸部の累積回旋角度, タスク完了時間で正規性が認められた. その後, 後続の検定にかけた結果, タスク完了時間以外の従属変数で有意差が認められた. 各従属変数の検定結果を以下に記述する.

タスク完了時間を繰り返しのある一元配置分散分析にかけた結果, 有意差は確認されなかった ($F(3, 33) = 0.29, p=0.83$).

頭部の累積回旋角度を繰り返しのある一元配置分散分析にかけると有意差が認められた ($F(3, 33) = 19.26, p < .05$). 球面性の仮定も満たしたため ($p > .05$), ホルム補正を適用した対応のある t 検定にかけた結果, ベースライン条件-頭部のみ条件間 ($p < .05$), ベースライン条件-頭部+胸部条件間 ($p < .05$), 胸部のみ条件-頭部のみ条件間 ($p < .05$), 胸部のみ条件-頭部+胸部条件間 ($p < .05$) で有意差が認められた (図 5a).

頭部の回旋範囲は繰り返しのある一元配置分散分析の結果, 有意差が認められ ($F(3, 33) = 22.78, p < .05$), 球面性の仮定も満たした ($p > .05$). 対応のある t 検定では, 胸部のみ条件-頭部のみ条件間 ($p < .05$), ベースライン条件-頭部のみ条件間 ($p < .05$), 胸部のみ条件-頭部+胸部条件間 ($p < .05$), ベースライン条件-頭部+胸部条件間 ($p < .05$) で有意差がみられた (図 5b).

胸部の累積回旋角度は, 繰り返しのある一元配置分散分析の結果, 有意差が認められ ($F(3, 33) = 9.60, p < .05$), 球面性の仮定を満たした ($p > .05$).

対応のある t 検定では、ベースライン条件-頭部+胸部条件間 ($p < .05$)、ベースライン条件-頭部のみ条件間 ($p < .05$)、ベースライン条件-胸部のみ条件間 ($p < .05$) で有意差が認められた。

胸部の回旋範囲は、フリードマン検定の結果、有意差が認められた ($\chi^2 = 10.40, p < .05$)。後続のホルム補正を適用したウィルコクソンの符号順位検定を行った結果、ベースライン条件-頭部のみ条件 ($p < .05$)、ベースライン条件-頭部+胸部条件 ($p < .05$) で有意差が認められた。

手部の累積移動量はフリードマン検定の結果、有意差が認められた ($\chi^2 = 20.10, p < .05$)。その後のウィルコクソンの符号順位検定では、ベースライン条件-胸部のみ条件間 ($p < .05$)、胸部のみ条件-頭部のみ条件間 ($p < .05$)、ベースライン条件-頭部+胸部条件間 ($p < .05$)、頭部のみ条件-頭部+胸部条件間 ($p < .05$) で有意差が認められた (図 5c)。

4.6.2 評価指標：作業負荷・サイバー酔い

NASA-RTLX の各項目でフリードマン検定を行った結果、精神的要求度 ($\chi^2 = 3.57, p > .05$)、身体的要求度 ($\chi^2 = 2.39, p > .05$)、時間的要求度 ($\chi^2 = 0.92, p > .05$)、達成度 ($\chi^2 = 2.77, p > .05$)、努力 ($\chi^2 = 3.34, p > .05$)、フラストレーション ($\chi^2 = 1.65, p > .05$) の全項目で有意差は確認されなかった。

SSQ の各項目にもフリードマン検定を行い、吐き気 ($\chi^2 = 0.61, p > .05$)、眼球運動系 ($\chi^2 = 0.37, p > .05$)、見当識障害 ($\chi^2 = 1.87, p > .05$) の全項目で有意差は確認されなかった。

5 議論

5.1 結果のまとめ

NASA-RTLX スコアを分析した結果、認知負荷を含めた全ての項目で、全条件間に統計的な差は確認されなかった。SSQ スコアも同様に、全条件間で統計的な差は確認されなかった。

一方で動作遂行コストに関する従属変数の分析では多くの有意差が見られた。特に「頭部+胸部増幅」条件は、ベースライン条件と比較してタスク完了時間を除く全ての動作遂行コスト指標で改善が認められた。一方、比較手法である「頭部のみ増幅」条件および「胸部のみ増幅」条件もベースラインと比較して一部の指標を改善したが、手部の移動量または頭部の回旋動作で動作量の削減が見られず、頭部と手部の動作量を同時に削減した「頭部+胸部増幅」条件に比べて動作量の削減効果は限定的だった。

これらのことから、頭部回旋のみを増幅させた既存手法 [12, 13, 14] は、頭部回旋増幅に伴う視認性を向上することは可能であるが、上肢や手部の物理的な動作量を削減し、障害物と上肢間での干渉を低減するのは困難であることが示唆される。一方で、提案手法は、視認性向上と物理的な制約の緩和を同時に実現可能であることが示唆された。

タスク完了時間で改善が見られなかった原因は、タスク全体に占める物理的動作時間に対し、探索などの認知的処理時間の割合が支配的だったためと考

えられる。不確実性の高いタスクでは、認知的処理時間がタスク完了時間の主要因となる可能性が指摘されている [26]。本実験でも、ランダムで出現する仮想オブジェクトを探索する認知的処理がボトルネックとなり、提案手法による物理的動作時間の削減効果が、タスク完了時間の短縮に十分寄与しなかったと考えられる。よって、物理的動作のみならず、認知的処理を考慮した設計も今後調査する必要がある。

また本研究では、平衡感覚・自己受容感覚・視覚情報間での齟齬と、それに伴う認知負荷・サイバー酔いの増大を防ぐため、頭部と胸部の回転ゲインを等しく設定した。しかし、認知負荷・酔いを十分に軽減し、より大きな動作量を削減する非対称なゲイン組み合わせの存在は十分に考えられ、この最適な組み合わせを探索し特定する必要がある。

5.2 応用例

まず1つ目に、エンターテインメント領域での応用が考えられる。VR ゲームなどでは周囲に存在するアイテムを把持して利用するといった際に、頭部や上肢の回旋が求められる [1, 2]。しかし座位では、周辺の障害物により動作が阻害される。そこで本手法を用いることで障害物との干渉を低減し、上肢動作を伴った広い VR 体験の実現に寄与できると考えた。

そして2つ目の応用例として、上肢の機能障害を抱えた患者の機能回復への応用が考えられる。理学療法の分野では、タスクを達成できるという自信 (自己効力感) が訓練効果を高めることが知られている [27]。この知見から、手部の並進移動を増幅し、VR 映像として提示する「ハンドリダイレクション」という手法 [28] の利用が検討されている [18]。これは、上肢の機能障害により本来は困難なタスクを仮想的に実行可能にし、患者の自己効力感と訓練効果の向上を目指す。しかし、ハンドリダイレクション手法は、ほとんどが直線的な手部の運動のみに着目している [18, 28]。そこで、本研究の提案手法を利用することで、頭部・胸部・上肢の回旋動作を伴う、より現実に近い多様な訓練タスクが実現可能となり、訓練環境の多様化に寄与できると考えた。

6 結論

本研究では、認知負荷・サイバー酔いを抑制しつつ、背もたれなどによる物理的制約を緩和する手法として、座位姿勢時の頭部と胸部 (上肢) の回旋動作を、ユーザが知覚できない範囲で同時に増幅するリダイレクション手法を提案した。まず、認知負荷・サイバー酔いを抑制するために、手法の設計に不可欠であった頭部と胸部 (上肢) の回旋増幅時の知覚閾値を調査した。結果、両部位の回旋を同時に 1.43 倍まで増幅しても、ユーザに知覚されないことが示唆された。その後、この閾値を用いて手法を設計し、有効性評価を行った結果、頭部と胸部 (上肢) を同時に増幅する提案手法はベースライン条件と比較して主観的な作業負荷・サイバー酔いに統計的な差は確認されなかったものの、頭部のみ、または胸部のみへ適用する手法と比較して、頭部の視認性向上と回旋動作における物理的な制約の緩和を同時に実現できる可能性が示された。

謝辞

本研究は、JST 創発的研究支援事業、JPMJFR226S の支援を受けたものである。

参考文献

- [1] Owlchemy Labs Meta Platforms Inc. meta store, vacation simulator. <https://www.meta.com/ja-jp/experiences/vacation-simulator/2393300320759737/>, 2025. アクセス日: 2025 年 8 月 22 日.
- [2] Valve Corporation. steam ストア, half-life: Alyx. https://store.steampowered.com/app/546560/HalfLife_Alyx/?l=japanese, 2025. アクセス日: 2025 年 8 月 22 日.
- [3] Matthew P Reed, Matthew B Parkinson, and David W Wagner. Torso kinematics in seated reaches. *SAE transactions*, pp. 167–174, 2004.
- [4] Erika Zemková, Michal Jeleň, Ludmila Zapletalová, and Dusan Hamar. Muscle power during standing and seated trunk rotations with different weights. *Sport Mont*, Vol. 15, No. 3, 2017.
- [5] Sam Tregillus, Majed Al Zayer, and Eelke Folmer. Handsfree omnidirectional vr navigation using head tilt. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, p. 4063–4068, New York, NY, USA, 2017. Association for Computing Machinery.
- [6] Ivan Poupyrev, Mark Billinghurst, Suzanne Weghorst, and Tadao Ichikawa. The go-go interaction technique: non-linear mapping for direct manipulation in vr. In *Proceedings of the 9th annual ACM symposium on User interface software and technology*, pp. 79–80, 1996.
- [7] Eike Langbehn, Joel Wittig, Nikolaos Katzakis, and Frank Steinicke. Turn your head half round: Vr rotation techniques for situations with physically limited turning angle. In *Proceedings of Mensch Und Computer 2019*, MuC '19, p. 235–243, New York, NY, USA, 2019. Association for Computing Machinery.
- [8] Nahal Norouzi, Luke Bölling, Gerd Bruder, and Greg Welch. Augmented rotations in virtual reality for users with a reduced range of head movement. *Journal of rehabilitation and assistive technologies engineering*, Vol. 6, p. 2055668319841309, 2019.
- [9] Caroline Jay and Roger Hubbard. Amplifying head movements with head-mounted displays. *Presence*, Vol. 12, No. 3, pp. 268–276, 2003.
- [10] R Jürgens, T Boss, and W Becker. Podokinetic after-rotation does not depend on sensory conflict. *Experimental Brain Research*, Vol. 128, pp. 563–567, 1999.
- [11] Tianren Luo, Gaozhang Chen, Yijian Wen, Pengxiang Wang, yachun fan, Teng Han, and Feng Tian. Exploring the effects of sensory conflicts on cognitive fatigue in vr remappings. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*, UIST '24, New York, NY, USA, 2024. Association for Computing Machinery.
- [12] Benjamin Bolte, Gerd Bruder, Frank Steinicke, Klaus Hinrichs, and Markus Lappe. Augmentation techniques for efficient exploration in head-mounted display environments. In *Proceedings of the 17th ACM Symposium on Virtual Reality Software and Technology*, VRST '10, p. 11–18, New York, NY, USA, 2010. Association for Computing Machinery.
- [13] Chen Wang, Song-Hai Zhang, Yizhuo Zhang, Stefanie Zollmann, and Shi-Min Hu. On rotation gains within and beyond perceptual limitations for seated vr. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 29, No. 7, pp. 3380–3391, 2023.
- [14] Jason Jerald, Tabitha Peck, Frank Steinicke, and Mary Whitton. Sensitivity to scene motion for phases of head yaws. In *Proceedings of the 5th Symposium on Applied Perception in Graphics and Visualization*, APGV '08, p. 155–162, New York, NY, USA, 2008. Association for Computing Machinery.
- [15] 石川雄大, 崔高超, 井上康之, 田川和義, 唐山英明. Vr 環境におけるピッチ方向のリダイレクション操作に関する検討. 日本バーチャリアリティ学会論文誌, Vol. 29, No. 1, pp. 57–65, 2024.
- [16] Sharif Razzaque. *Redirected walking*. The University of North Carolina at Chapel Hill, 2005.
- [17] 山本達己, 松本啓吾, 鳴海拓志, 谷川智洋, 廣瀬通孝. ヨー方向とロール方向の複合的視点操作によるリダイレクション技術の検討. 日本バーチャリアリティ学会論文誌, Vol. 23, No. 3, pp. 159–168, 2018.
- [18] 森本浩輔, 渡邊恵太. バーチャルハンドのスケール変化を用いたハンドリダイレクション手法の提案. 日本ソフトウェア科学会インタラクティブシステムとソフトウェア研究会 予稿集, 2024. デモ発表.
- [19] Frank Steinicke, Gerd Bruder, Luv Kohli, Jason Jerald, and Klaus Hinrichs. Taxonomy and implementation of redirection techniques for ubiquitous passive haptic feedback. In *2008 International Conference on Cyberworlds*, pp. 217–223. IEEE, 2008.
- [20] Ali Nasr, Arash Hashemi, and John McPhee. Scalable musculoskeletal model for dynamic simulations of upper body movement. *Computer methods in biomechanics and biomedical engineering*, Vol. 27, No. 3, pp. 306–337, 2024.
- [21] W Ben Kibler and Aaron Sciascia. Current concepts: scapular dyskinesis. *British Journal of Sports Medicine*, Vol. 44, No. 5, pp. 300–305, 2010.
- [22] Keigo Matsumoto, Eike Langbehn, Takuji Narumi, and Frank Steinicke. Detection thresholds for vertical gains in vr and drone-based telepresence systems. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 101–107. IEEE, 2020.
- [23] Felix A Wichmann and N Jeremy Hill. The psychometric function: I. fitting, sampling, and goodness of fit. *Perception & psychophysics*, Vol. 63, No. 8, pp. 1293–1313, 2001.
- [24] Nicolaas Prins, et al. *Psychophysics: a practical introduction*. Academic Press, 2016.
- [25] Séamas Weech, Sophie Kenny, and Michael Barnett-Cowan. Presence and cybersickness in virtual reality are negatively related: A review. *Frontiers in Psychology*, Vol. Volume 10 - 2019, , 2019.

- [26] Steven C. Seow. Information theoretic models of hci: a comparison of the hick-hyman law and fitts' law. *Hum.-Comput. Interact.*, Vol. 20, No. 3, p. 315–352, September 2005.
- [27] Rachana Gangwani, Amelia Cain, Amy Collins, and Jessica M. Cassidy. Leveraging factors of self-efficacy and motivation to optimize stroke recovery. *Frontiers in Neurology*, Vol. Volume 13 - 2022, , 2022.
- [28] André Zenner and Antonio Krüger. Estimating detection thresholds for desktop-scale hand redirection in virtual reality. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 47–55, 2019.

未来ビジョン

リダイレクション手法は、頭部の回転や歩行動作に限らず、階段の昇降やドアの開閉など、多様な状況下における動作への応用が試みられてきた。

しかし既存研究の多くは、身体の一部位の動作や、身体全体を単一的に捉えた動作に焦点を当てている。しかし、実際の人体の動作は、複数の身体部位が協調することで成立しており、複雑かつ多様である。

そこで、本研究では、頭部・胸部・手部といった複数部位に対してリダイレクションを適用することで、より実際の身体動作に近い複雑な動作への拡張を試みた。これは、単一的な身体動作からさらに発展し、より複雑な身体動作における可動性の向上に寄与するものである。

将来的には、本研究をさらに発展させ、より多くの身体部位を考慮したリダイレクションを実現したい。そして、様々な身体動作が求められる VR ゲームなどのエンターテインメント領域から、身体機能に制限を有する者への身体動作の補助など、様々な状況下で柔軟に応用できるリダイレクションを単一の設計で実現できるようにしたい。

また、VR 技術と日常生活が溶け合った、さらに長期的な将来においては、多くの人々の日常生活に、リダイレクション手法が利用されることで、身体機能による制約を超えたより柔軟な身体活動を実現し、かつ身体機能に障害のある者と、健常者との身体機能の差が縮小することで、誰もが自由かつ平等に活動できる世界を実現したい。

PracToneVis：管楽器奏者のための基礎練習支援システム

池谷 優希* 伊藤 貴之*

概要. 管楽器奏者にとって基礎練習は演奏技術の土台となる重要な活動であるが、その成果を客観的に評価する機会は多くない。特に音色は定量的評価が困難であり、課題の把握が難しい。本研究では、ロングトーンを対象とした基礎練習支援システム「PracToneVis」を開発し、音響解析による可視化と大規模言語モデル(LLM)によるフィードバックを通じて、演奏傾向の理解を支援する。音響解析において、我々は音量・音程・音色の3点に着目している。音量と音程は基準値との差分に基づき評価し、音色は倍音構成から抽出した特徴ベクトルをクラスタリングして可視化する。さらにこれらの安定性を得点化し、LLMによる自然言語での解説とともに演奏者に提示する。実験では、10名のフルート奏者によるロングトーンを分析し、PracToneVisの出力に対する評価アンケートを実施した。その結果、音色に関する表示の有用性が示された。今後は、より可読性の高い可視化方法の検討と、問題点の自動検出を通じて、さらなる練習の効率化を目指す。

1 はじめに

近年、音楽教育において ICT や AI 技術を活用した学習支援が注目されている。特に、個人練習の質を向上させるために、演奏の可視化やフィードバックを自動化する研究が進んでいる。一方で、管楽器の基礎練習は、客観的な評価を得る機会は少ない。

ロングトーン練習は、息のコントロール、音程、音色の安定性などを養うための基本的な練習であり、多くの管楽器奏者が日常的に取り組んでいる。しかし、こうした練習においては、演奏者が自分自身の成長や課題を客観的に把握することが難しく、特に指導者不在の状況ではフィードバックの欠如が顕著である。これにより、非効率な練習や誤った癖の定着などが生じるリスクがある。

本研究では、こうした課題を解決するために、音響解析と大規模言語モデル(LLM)を活用した基礎練習支援システム「PracToneVis」を提案する。このシステムは、演奏音声の録音から音響的特徴（音量、音程、音色）を抽出・分析し、それらをもとに練習内容を数値化・可視化する。さらに、分析結果から LLM による自然言語でのアドバイスを生成し、専門的な知識がなくても演奏の改善点を理解しやすいよう支援する。

このシステムの特徴は、音色に焦点を当てた音響的な定量評価と、人間らしい言語による解説を組み合わせる点にある。これにより、練習者は自らの演奏の良し悪しを数値化および言語化された形で把握

し、改善に向けた次の一手を明確にすることが可能となる。また、分析結果は視覚的に提示されるため、練習の傾向や変化を時系列で追うことも可能であり、継続的な成長を促す要素として機能する。

本稿では、システム構成、音響分析手法、LLM によるフィードバック生成の仕組み、そして実際に 10 名のフルート奏者を対象に実施したユーザ評価実験の結果について報告する。

2 関連研究

2.1 管楽器の基礎練習

管楽器奏における基礎的な技能の習得には、ロングトーンやスケール練習といった反復的なトレーニングが不可欠である。

渡部らは、伝統的なロングトーン練習が、単なる基礎固めにとどまらず、音質形成、集中力の育成、姿勢保持など多面的な効果をもつことを指摘し、その文化的・身体的側面にも注目した[1]。また、ロングトーン練習が音楽表現力の向上に資する理想的な訓練形態であることが、練習法の検討や、脳科学の面から示されている[2, 3]。

さらに、近年では情報技術を応用した支援システムの開発が進んでいる。森下らは、音の安定性と持続性に着目し、ロングトーン練習の可視化によって初学者の定着を支援するシステムを提案した[4]。また毛利らは、スケール練習時の音程や音量をリアルタイムでフィードバックするシステムを構築し、正確性と音色の両面からの改善を図った[5]。しかしこれらの手法では、ロングトーンの反復に伴う音色の変化を読み取りにくいという課題が残っている。

Copyright is held by the author(s).

* お茶の水女子大学

これらの研究に加え、学習者自身の気づきを促すインタラクティブな支援環境も検討されている。Watanabe らは、Web 上で演奏を記録・分析・自己評価できる練習支援システム「Suiview」を提案した[6]。この研究は、音量・音程・音色のグラフ表示、それらの安定性得点化、ログ機能に加え、データ共有により指導者からフィードバックを送ることを可能にした。このアプリケーションは楽器の初学者を対象にしたものであり、練習効率の向上に大きく寄与することを証明した。実際に、中学生奏者を対象としてモデル演奏の視聴や自己評価、録音視聴の有効性を検証し、これらの組み合わせが演奏技術に正の影響を及ぼすことが示されている[7]。しかし、これらの取り組みは音のばらつきが大きい初学者を主な対象としている。一定以上の楽器経験を有する中級者にとって、音量や音程のみならず音色も実力向上のための重要な要因であるが、音色に関する情報や分析が十分に提供されているとは言い難い。また、毛利らの研究により実証されているように、管楽器の演奏経験の有無と音色の識別能力には強い相関がある[5]。そのため、音色を分類しただけでは、その特徴をユーザが正確に理解できるとは限らない。

一方で、機械学習や AI を活用した個別最適化の研究も進行している。具体的には、サクソフォン初学者の演奏傾向を分類し、タイプ別に異なるフィードバックを提示する手法[8]や、音響情報と姿勢データを同時に解析することで、演奏技能と身体的負荷のバランスを評価する手法[9]などが提案されており、これらは総合的な演奏支援の実現可能性を示唆している。

このように、管楽器の基礎練習に関する研究は、従来の実践知と科学的知見を融合しつつ、可視化技術や AI を取り入れた支援環境の構築へと発展している。

2.2 演奏の可視化

管楽器の基礎練習に限らず、演奏プロセスの機械学習と可視化は、演奏者の内省と技能向上を支援する手段として近年さまざまな形式で注目されている。とりわけ歌唱や楽器演奏において、音高変化や演奏傾向を可視的に把握する試みが進められている[10, 11, 12]。また Giraldo らは、バイオリン演奏における音の明るさ、豊かさ、安定性などを含む「音色の質」をリアルタイムで評価する機械学習モデルを開発した[13]。Pati らは学生による音楽演奏を自動的に評価するための DNN (深層ニューラルネットワーク) モデルを提案した[14]。Johnson らは、スマートフォンを用いて演奏中のピッチとタイミングをリアルタイムに評価・フィードバックするアプリケーション「MusicCoach」を提案した[15]。また、

スコアと関連付けることによって、演奏課題を可視化する手法も検討されている[16, 17]。Khulusi らは、これらの演奏の可視化に関する既存研究を俯瞰的に整理し、IMUTUS や Open Orchestra などの実装例を紹介しながら、学習者の自己評価や継続的練習の支援における可視化の有効性を論じている[18]。

このように、演奏の機械学習と可視化に関する研究は、パフォーマンス評価、構造理解、学習支援といった複数の目的を持ち、それぞれに応じた手法が開発されている。

2.3 音色の分析

音色は演奏表現の重要な要素であり、その分析と可視化は演奏理解や教育支援においても注目されている。中でも木管楽器の分析においては、特徴依存性に関する調査や、FFT 音響記述子と機械学習を用いた音色の分析など、さまざまな手法が採用されてきた[19, 20]。

本研究で実験の対象とするフルートについても、倍音に関する様々な検証がなされている。Yorita らは、音色の倍音構造を言語的イメージに対応づけた[21]。その後の研究により、奏者が倍音バランスを無意識に操作することで多様な表現を生み出していることが明らかにされた。さらに、演奏者による無意識的な音色調整の方法と、それを聴取者がどのように知覚するかが詳細に検討された[22, 23]。

また、Ferguson らは演奏中の音響特徴をリアルタイムに可視化することで、学習者が自身の音を直感的に理解できるようにするシステムを開発した[24]。球体オブジェクト群とパーティクル (粒子) を音に合わせて描画し、各パラメータを球や粒子の動きで表現している。Gkiokas らのクラリネットを対象とした音響分析においても、同様に球体で音の変化を表現している[25]。これらの研究は、音色という聴覚的・主観的な情報を視覚的に再構成し、分析・教育・創作の場面に応用する試みとして意義深い。

一方で、Yorita らによる倍音分析は音色評価構造の理解にとどまっている。本研究の新規性は、この倍音分析を基礎練習支援の文脈へ発展させた点にある。提案システム「PracToneVis」は、倍音構成を特徴量としたクラスタリングと色付けにより、音色の安定性や傾向を直感的に把握できる新しい可視化手法を実現した。さらに、大規模言語モデル (LLM) による自然言語フィードバックを統合し、音響データ解析と AI による出力を組み合わせたインタラクティブな評価を可能にした。これにより、データ駆

動型の客観指標と視覚的フィードバックを導入し、自律的な基礎練習を支援する枠組みを提示した。本研究では、Yorita らの音響学的研究成果を教育実践へと展開するモデルを示すとともに、LLM によるコメント生成を通じて、教師不在下でも学習者が「何をどのように改善すれば良いか」を理解できる自律学習支援環境を実現した点に意義がある。

3 提案手法

本研究では、単音主体の楽器練習（例：ロングトーン）における音響的な演奏傾向を定量的かつ視覚的に解析するための Web ベースの分析支援システム「PracToneVis」を構築した。本システムは録音音源を入力として、各音に対して音量・音程・音色の 3 つの項目を数値化・可視化するとともに、それらの安定性スコアを算出して総合評価として可視化する。また、音色クラスターの分析に基づいて、LLM（大規模言語モデル）による音質傾向の説明文も生成し、演奏者にとって可読性の高いフィードバックを提供することを特徴とする。なお、基礎練習において最も重要なのは自らの音を丁寧に聴くことであり、画面表示に過度に依存してしまうと演奏者の成長を妨げる可能性がある。そのため、本研究はリアルタイム解析ではなく録音音源を対象とした。

3.1 処理手順

3.1.1 音声の読み込みとセグメンテーション

音声ファイルは WAV 形式でアップロードされ、librosa ライブラリを用いて読み込まれる。まず基本周波数(f_0)とエネルギー(RMS)をフレーム単位で抽出し、RMS 値の閾値に基づいて有声音区間をマスク処理する。さらに、ユーザが指定するテンポおよび拍数情報を基に、楽譜上の音価に対応する長さ（例：全音符 8 拍）でセグメントを分割する。区間の検出には領域ラベリング法を用い、各セグメントの持続時間が所定の閾値を満たす場合にのみ分析対象として採用する。この処理により、各音を明確に分離して分析することが可能になり、音の傾向がつかみやすくなる。一方で、この可視化はタイミングの抽出には対応していない。

3.1.2 倍音構造に基づく音色特徴量の抽出

各セグメント内のスペクトル構造から音色に関する特徴量を抽出する。まず短時間フーリエ変換(STFT)によってピーク周波数を検出する。続いて最も低いピークを基音候補とし、他の周波数成分との相対比（基音の整数倍付近）に基づいて、最大 12 次までの倍音強度ベクトルを構築する。続いて 1 次成分を 1.0 とした相対値として各ベクトルを正規化して、「音色特徴量」として記録する。全てのセグ

メントから収集した特徴ベクトルに標準化を施し、次元削減とクラスタリングに利用する。基音とその倍音成分との相対的な強度比を算出することにより、音程の揺れによるクラスタリングへの影響を最小限に抑える。

3.1.3 音色クラスタリングと視覚化

音色特徴量に対して主成分分析(PCA)を適用して 2 次元空間に射影する。続いて K-means 法によるクラスタリングを適用する。ここで最適なクラス数（ただし最大 7 個とする）の決定にはシルエットスコアを用いる。各クラスに対して PCA 空間上の重心を求めた後、その順序に基づいて 7 色のカラーパレット（赤～紫の虹色スケール）を割り当てる。これにより、クラスターの音質傾向が視覚的に区別可能となる。

ここで、各セグメントに対して、以下の 3 つの要素を同一時間軸上に重ねて可視化する。

- ・ エネルギー（RMS）推移
- ・ チューニング誤差（単位：セント）
- ・ 音色クラスター（PCA 空間上の座標とクラスターの色）

この可視化により、演奏者は各音の安定性や個性を多角的に把握することが可能となる。

3.1.4 安定性スコアの算出

本研究では Watanabe らの「Suiview」[6]を参考にして以下の安定性スコアを設定し、セグメントごとに 0~100 点のスケールで算出する。

- ・ 音量安定性スコア：RMSの標準偏差に基づき、変動が少ないほど高得点とする。変動が大きい場合は「不安定」と判定される。
- ・ ピッチ安定性スコア：チューニング誤差（単位：セント）の絶対平均値に基づき、音高の揺らぎが小さいほど高得点とする。
- ・ 音色安定性スコア：セグメント内に含まれるクラスターラベルの分布から、最頻クラスターの占有率を算出する。単一クラスターに収束していれば高得点、分散していれば低得点とする。

一音ずつの得点に加え、各練習における安定性スコアの平均を「安定性総合スコア」として表示することで、練習全体における演奏傾向を俯瞰可能とする。なお、このスコアは音色の「良否」ではなく「安定性」を主眼に設計している。したがって、望ましくない音色であっても一貫して同一クラスターに属する場合、高得点となる可能性がある。この点は音質評価とは独立した設計上の制限であり、今後は良質音クラスターとの複合スコア化を検討する。

3.1.5 音色クラスターの特徴記述と言語化

本研究では、音響分析結果を専門的でない演奏者にも伝わりやすい形で提示するために、LLM (GPT-

PracToneVis

基礎練習を分析する

ファイルを選択 選択されていません

楽器を選択

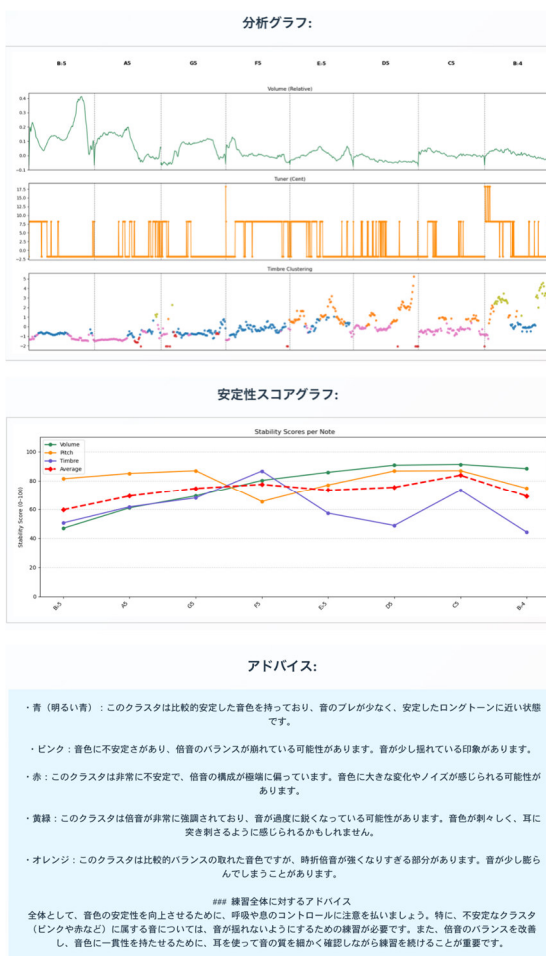
練習の種類を選択

速度 ↓ =

音程の基準: 442 Hz

解析実行

図 1. 音声アップロード・基本情報入力画面

図 2. PracToneVis による結果出力画面
(上から音量・音程・音色・安定性・AIの評価)

4a) を用いた評価文の自動生成を適用する。具体的には、音色の各クラスタについて以下の情報をプロンプトとして与える。

- ・ クラスタを構成する音の倍音強度ベクトルの平均
- ・ 音高ごとのクラスタ出現パターン (どの音高がどのクラスタに何回出現したか)
- ・ 割り当てられた色 (視覚的参照のため)

これらの情報をもとに、各クラスタの音色的特徴を「こもった」「明るい」「響きがある」「硬い」などの日本語の形容詞を用いて生成 AI に簡潔に回答させる。本システムでは生成 AI による回答に対して、専門語を避け、楽器演奏者に馴染みのある直感的な表現を促すようにチューニングをしている。さらに、全体的な演奏傾向に基づいた練習アドバイス (例: 「より芯のある音を目指しましょう」など) も自然言語で出力する。これらの出力は、AI に与えるプロンプトの内容によって傾向が大きく異なるため、ユーザ実験を通じて最良なプロンプトの内容、表現を検討する。

3.2 システム実装と出力形式

Python ベースの Flask によるバックエンドと React によるフロントエンドで構成される Web アプリケーションとして提案システムを実装した。ユーザはブラウザから音声ファイルをアップロードし、テンポ・拍数・基準ピッチ (例: 443Hz)・楽器名などのパラメータを図 1 の画面で入力する。このときシステムは即時に、図 2 のような分析結果を返す。分析グラフは上から順に音量・音程・音色を示す。音量は最頻値を基準として 0 に正規化し、音程は入力された基準ピッチを 0 として表示する。また、音色表示の色は AI によるアドバイス内の色と対応しており、各クラスタの特徴が言語的に説明される。安定性スコアグラフは、各音に対する得点をプロットすることで、音ごとに異なる課題の傾向を可視化する。

さらに、分析に伴い以下の 4 種類のデータが自動的に保存される。

- ・ 分析グラフ (PNG)
- ・ 安定性スコアグラフ (PNG)
- ・ AI によるアドバイス (PNG)
- ・ 倍音情報などの数値データ (JSON)

これらはすべて Web アプリ上でダウンロード可能となっており、演奏履歴の蓄積や教育目的での比較分析にも活用可能である。

表 1. 音色のクラスタリングに対する評価

	経験年数 (年)	回答
被験者 1	3	適切
被験者 2	10	適切
被験者 3	10	どちらかというと適切
被験者 4	20	どちらかというと適切
被験者 5	0.3	どちらかというと適切
被験者 6	10	どちらかというと適切
被験者 7	12	どちらかというと適切
被験者 8	11	どちらかというと適切
被験者 9	40	どちらかというと適切ではない
被験者 10	10	適切

表 2. 各プロンプトの特徴

	倍音と音色の関係性	特徴
A	×	フルートの倍音と音色の関係性の情報を全く与えない
B	○	「倍音という言葉を使わずに分かりやすく説明して。」という指示文を追加する
C	○	倍音と音色の関係は与えるが、その他詳細な指示文は削除する
D	○	「音の特徴を説明する際には、'響きのある'、'硬い'、'こもった'、'明るい'、'深みがある'といった言葉を適切に活用して。」という指示文を追加する

4 ユーザ実験

4.1 実験方法

実験では、フルート奏者のロングトーンに対する PracToneVis の解析結果と演奏者自身の自己認識との関係を明らかにすることを目的として、音響解析および主観的自己評価に基づく検証を実施した。

具体的には、10 名の異なる演奏歴を有するフルート奏者に対し、指定された 1 オクターブの音域 (例: C4~C5) に含まれる音を、ロングトーンで録音するよう依頼した。演奏は統一された録音条件下で、一定のテンポのもとで実施した。収録した各演奏について、PracToneVis での音響解析を実施した。

これらの音源に対する可視化結果を各演奏者本人に対して提示し、クラスタリングに対する評価、自然言語による出力に対する評価、さらに基礎練習に対する考えなどについてオンラインアンケート (Google Form) 形式で回答を求めた。学習者自身が自らの演奏傾向を主観的に理解し改善行動に繋げることも練習支援の目的であることから、本研究では主観的納得感を評価に含めることにした。ただし、初心者においては自己解釈が十分に形成されていない場合もあるため、今後は専門家による評価実験も実施する必要がある。

表 3. 得点化の結果

種類	A	B	C	D
票数	26	33	25	26

4.2 実験結果

4.2.1 音色のクラスタリングに対する評価

被験者 10 名に対して音色のクラスタリング (色分け) 結果は適切かどうかを「適切」、「どちらかといえば適切」、「どちらかといえば適切ではない」、「適切ではない」の 4 択で質問した結果、表 1 の回答を得た。クラスタリング結果は概ね適切と判断されたが、経験年数の長い被験者 9 のみ「どちらかといえば適切ではない」を選択した。

4.2.2 自然言語による出力に対する評価

次に、被験者 10 名に対して、4 種類のプロンプト (A~D) に基づく出力結果に対する評価を依頼した。各プロンプトの違いは表 2 に示す。ここで「倍音と音色の関係性」とは、Yorita らがフルートの音色を倍音構造で分類し、各分類に対する人間の印象を対応づけた知見を指す[21]。この情報をプロンプトに付与することで、LLM が各音の倍音構造に基づき特徴をより厳密に判断できると考えた。なお、この情報はプロンプト A を除くすべてに与えられている。さらに、その他の指示内容を差別化することで、プロンプトごとの出力傾向を検証した。特にプロンプト D では、「響きのある」「硬い」「こもった」など、倍音と音色の関係性を表す中核的な形容詞を明示的に指示している点が特徴である。アンケートでは 4 種類の結果を提示し、録音を聞いた後、自身の解釈に最も近いと感じた順に順位付けを依頼した。その順位に対して線形配点法 (1 位=4 点、2 位=3 点、3 位=2 点、4 位=1 点) による得点化を行った結果を表 3 に示す。最も高得点を得たのは B であり、A および D は同点でこれに続き、C はやや低い得点となった。さらに、奏者の経験年数と各出力の評価傾向を分析したところ、経験年数が短い奏者は B を高く評価する傾向があり、経験年数が長い奏者ほど D を支持する傾向が顕著であった。

4.2.3 自然言語による出力に対する評価

最後に、基礎練習の目的、普段感じている課題、PracToneVis の感想について自由記述で回答を求めた。表 4 は回答の抜粋である。目的に関しては、全員が「音の安定」と「音質の向上」を重視していることが確認された。一方で、演奏課題については奏者ごとに多様な内容が挙げられたが、全体としては音色の安定性や客観的評価の欠如に関する記述が最も多く見られた。さらに、PracToneVis の感想では、肯定的な意見が大半を占めたが、「良かった点も知りたい」という意見も寄せられた。

5 考察

ユーザ実験により、PracToneVis が提示する音色クラスタリングの結果は多くの被験者にとって妥当と評価された。特に初級から中級レベルの奏者に対しては、音色のばらつきや特徴を直感的に把握できる点で有用性が高いと考えられる。一方で、経験年数が長く音色が安定した奏者にとっては、本システムの出力が必ずしも自身の感覚と一致しない場合があり、分析モデルの改善余地が示唆された。現在のシステムでは、音が十分に安定している場合であってもクラスタ数を多めに設定してしまい、結果としてクラスタ間の音色の差異が小さいにもかかわらず過剰にクラスタを分割する傾向が見られた。このことから、クラスタ数の設定方法については再検討が必要である。

また、プロンプト B が最も高く評価された点は注目に値する。プロンプト A 以外の条件では、いずれも同じ倍音構造や音色に関する知識を与えているが、指示文によって出力のわかりやすさが大きく異なった。特にプロンプト B は、専門用語を避けて平易に説明するよう指示しており、直感的に理解しやすい表現につながったと考えられる。一方で、ある程度音が安定した熟練者にとっては、より精緻な言葉づかいや具体的な形容詞を含む説明が求められ、その結果、詳細な指示を与えたプロンプト D が相対的に高く評価されたと推察される。さらに、プロンプト A が比較的高い評価を得たことも興味深い。この傾向の背景として、初学者の演奏は音のばらつきが大きいいため、AI がもともと有している一般的な知識に基づく説明でも十分な特徴把握が可能であることが考えられる。すなわち、演奏者の熟練度によって求められるフィードバックの粒度が異なることが明らかとなったといえる。

自由記述の回答からは、演奏者の多くが「音の安定」と「音質の向上」を基礎練習の目的とすることが明らかになった。これらは「安定性のグラフ」や「音色の分類およびクラスタの説明」といった各機能にそれぞれ直結した回答であることから、本システムの有用性が示されたと言える。また、「改善点だけでなく良好な点も表示して欲しい」という要望も見られた。このことは、本研究の目的の一つである「基礎練習に対するモチベーションの維持」と強く関連している。従来の自動評価システムは誤りや弱点の指摘に重点を置きがちであるが、基礎練習においては「成長が実感できるポジティブなフィードバック」も演奏継続の大きな要因となる。したがって PracToneVis においても、良好な点を強調する仕組みを導入することで、練習者の前向きな感情を高め、ひいては練習意欲を高める効果が期待できる。

表 4. 自由記述の抜粋

質問	回答
基礎練習の目的	- 音質の精度向上 - 音程感のコントロール - 音を安定して鳴らす練習 - 指の連結、豊かな響きを出す
普段感じている課題	- 音の最初と最後が安定しない - 苦手な音がある 1人だと課題が分からない - 音によって音色が違う
PracToneVis の感想	- 音の安定性がわかりやすい - 音の傾向が分かって面白い - 良い点も教えて欲しい - 第三者目線の評価が有益

6 まとめ

本研究では、管楽器奏者の基礎練習を支援するシステム PracToneVis を提案し、フルート奏者を対象とした実験を通じてその有効性を検証した。PracToneVis は、録音音源を入力として音量・音程・音色を解析し、その安定性を数値化・可視化することにより、主観的評価に依存していた基礎練習を客観的に理解できる環境を提供する。また、音色のクラスタリング結果に基づいて LLM による自然言語フィードバックを生成し、専門知識を持たない学習者でも直感的に改善点を把握できることを特徴とする。これにより、演奏者は数値化・可視化結果の確認にとどまらず、音質傾向を言語的に理解することで、より効果的な練習サイクルを形成できる。

ユーザ実験の結果からは、音色クラスタリングの妥当性が概ね肯定され、自然言語による説明も有用であることが示された。特に「音の安定」と「音質向上」という基礎練習の主要な目的に対して、本システムが効果的に機能していることが確認された。一方で、経験豊富な奏者に対してはクラスタリングの精度や説明内容が必ずしも十分ではないことが示され、自由記述からはポジティブなフィードバックへの要望が示された。これにより、PracToneVis にはさらなる改善の余地があることが確認された。

また、本システムは現在フルートを対象として検証を行った段階にとどまっているが、他の管楽器にも応用可能である。今後は楽器ごとの音響的特性や倍音構造の違いを考慮し、より汎用的に利用できるよう機能の調整を進める必要があると考えられる。それに加えて、クラスタリング機能の精度向上と演奏者のモチベーションを高める仕組みの導入を実施し、基礎練習の質と継続意欲の双方を強化する支援環境の構築を目指す。

参考文献

- [1] 渡部謙一. 管楽器基礎奏法におけるテクニカルアスペクトの解析と再考: 伝統的基礎練習「ロングトーン」の本質. 北海道教育大学紀要 教育科学編, 71(2), 177-186, 2021.
- [2] 河内勇. クラリネット演奏における基礎練習に関する一考察 (1): 小・中学生の初心者のためのロングトーン及び跳躍練習について. 兵庫教育大学研究紀要: 学校教育・幼年教育・教育臨床・障害児教育・言語系教育・社会系教育・自然系教育・芸術系教育・生活・健康系教育・総合学習系教育, 39, 183-190, 2011.
- [3] Braase, R. Your Brain on Practice: Evidence-Based Strategies for Musical Training. Arizona State University, 2024.
- [4] 森下孟, 小口宙暉, 國宗永佳, 桐原礼, 本間喜子. ロングトーン練習に着目した初学者向け学校吹奏楽練習支援システムの開発. 信州大学附属図書館, 2018.
- [5] 毛利仁紀, 土屋雅人. 管楽器吹奏技術向上のためのスケール練習システムの研究. 日本デザイン学会研究発表大会第 68 回春季研究発表大会, 234, 2021.
- [6] Watanabe, M., Onoue, Y., Uemura, A., Kitahara, T. Suiview: A Web-Based Application that Enables Music in the AI Era: 15th International Symposium, CMMR 2021, Revised Selected Paper, 13770, 69-75. 2023.
- [7] Hewitt, M. P. The effects of modeling, self-evaluation, and self-listening on junior high. Journal of Research in Music Education, 49(1), 306-318, 2001.
- [8] 平野沙月. 機械学習を用いたサクソフォン初学者のための練習支援システムの提案. 卒業論文, 愛知県立大学 情報科学部, 2021.
- [9] Bagga, G., Wang, A., Qi, B., Perumal, S. Artificial intelligence in music—Evaluator. The Journal of Purdue Undergraduate Research, 14(1), 38, 2024.
- [10] Itoh, T., Nakano, T., Fukayama, S., Hamasaki, M., Goto, M. SingDistVis: Interactive overview+detail visualization for F0 trajectories of numerous singers singing the same song. Multimedia Tools and Applications, 84(2), 1057-1077, 2025.
- [11] 原彩月, 伊藤貴之. 同一楽曲におけるベースの演奏傾向の可視化. 第 86 回全国大会講演論文集, 2024(1), 817-818, 2024.
- [12] Knight, T., Boulliot, N., Cooperstock, J. R. Visualization feedback for musical ensemble practice: A case study on phrase articulation and dynamics. Visualization and Data Analysis 2012, SPIE 8294, 88-96, 2012.
- [13] Giraldo, S., Ramirez, R., Waddell, G., Williamon, A. A real-time feedback learning tool to visualize sound quality in violin performances. In Proc. MML 2017, 19-24, 2017.
- [14] Pati, K. A., Gururani, S., Lerch, A. Assessment of student music performances using deep neural networks. Applied Sciences, 8(4), 507, 2018.
- [15] Johnson, D., Han, D. Music Coach: Real-time evaluation of music performance using Nokia N900. Department of Computer Sciences, University of California, Santa Barbara, 2009.
- [16] Heyen, M. F., Sedlmair, M. Supporting music education through visualizations of MIDI recordings. IEEE VIS, Posers, 14, 2020.
- [17] Hiraga, R., Watanabe, F., Fujishiro, I. Music learning through visualization. In Proc. Second International Conference on Web Delivering of Music (WEDELMUSIC), 101-108, 2002.
- [18] Khulusi, R., Kusnick, J., Meinecke, C., Gillmann, C., Focht, J., Jänicke, S. A survey on visualizations for musical data. Computer Graphics Forum, 39(6), 82-110, 2020.
- [19] Gonzalez, Y., Prati, R. C. Similarity of musical timbres using FFT-acoustic descriptor analysis and machine learning. Eng, 4(1), 555-568, 2023.
- [20] Brown, J. C., Houix, O., McAdams, S. Feature dependence in the automatic identification of musical woodwind instruments. The Journal of the Acoustical Society of America, 109(3), 1064-1072, 2001.
- [21] Yorita, R., Clements, J. Using spectral analysis to evaluate flute tone quality. Proceedings of Meetings on Acoustics, 23(1), 035001, 2015.
- [22] 平岩樺生. フルートの演奏音における高周波成分と奏者による音色操作の関係. 国立音楽大学, 2024.
- [23] Hiraiwa, K., Miura, M. Unconscious overtone manipulation and transmission in flute performance: Insights into musical expression and perception. Frontiers in Psychology, 16(1393689), 2025. doi: 10.3389/fpsyg.2025.1393689
- [24] Ferguson, S., Vande Moere, A., Cabrera, D. Seeing sound: Real-time sound visualisation in visual feedback loops used for training musicians. In Proc. Ninth International Conference on Information Visualisation, 97-102, 2005.
- [25] Gkiokas, A., Perifanos, K., Nikolaidis, S. Realtime detection and visualization of clarinet bad sounds. In Proc. 11th International Conference on Digital Audio Effects (DAFx), 59-62, 2008.

未来ビジョン

本研究で提案した PracToneVis は、基礎練習を効率化する支援システムとして構築されたが、将来的には「誰もが平等に音楽を学び楽しめる社会」の実現に寄与できると考える。楽器演奏の習得の中には指導者や専門教育環境が不可欠な楽器もあり、都市部や経済的に恵まれた人々にしか十分な機会を見つけれない場合もあった。この問題に対して AI を活用することで、学習者は場所や環境に依存せずに自らの演奏を解析し、わかりやすい言葉で改善点や強みを把握できるようになる。その結果、独学者や教育資源が限られた地域の人々でも、音楽を継続的に学び、楽しむことが可能となる。こうした仕組みは、音楽教育における格差の解消と、文化としての音楽の民主化に大きく貢献する。

さらに、本システムは練習効率の向上にとどまらず、利用者のモチベーションを高める仕組みへと発展できる。第一に、演奏データをオンライン上で公開し、他のユーザがリアクションを付与できる機能を導入することで、学習者は他者からのフィードバックや共感を得られる。第二に、演奏データをベースとした学習型のゲーミフィケーションを組み込み、達成度に応じて新しい課題やステージが開放される仕組みを提供できる。これにより、学習者は基礎練習の習熟度を段階的に実感でき、練習そのものを「挑戦と達成のサイクル」として楽しむことが可能になる。これらの機能は、単調になりがちな基礎練習を継続的かつ主体的に取り組む学習体験へと変えることが期待される。

後方伸身宙返りの段階的学習をサポートする練習支援システム

高橋 洸汰* 岡部 誠*

概要. 器械体操の床種目における後方伸身宙返りを対象に、段階的学習支援と言語フィードバックを統合した練習支援システムを提案する。提案システムは、スマートフォンやタブレットで撮影した演技動画を姿勢推定し、得られた姿勢情報を採点規則に基づく減点リスクとしてアニメーション形式で可視化し、角度変化を演技評価に直結して理解できる機能を提供する。さらに、初級者から上級者までの演技を解析し、演技間の類似度マップを作成し表示する。学習者はマップの中から次の目標となる演技を選択し練習の参考とすることで段階的な学習を行うことができる。加えて大規模言語モデル(LLM)を活用し、目標の演技と自身の現在の演技の違いを解析させ、自然言語での改善指導を行う。初級者・中級者のスキルレベルからなる体操経験者を対象に評価実験を実施し、改善点の明確化や学習意欲向上に効果が確認された。

1 はじめに

器械体操は空中での回転運動や複雑な身体操作を含む競技であり、演技中に自らの姿勢や動きをリアルタイムで視覚的に確認することは困難である。そのため、技術の習得や改善には動画撮影や指導者からのフィードバックといった外部支援が必要となる。しかし、学校の部活動など多くの練習現場では指導者数が限られており、全ての選手に十分な指導を行うことが難しい。加えて部活動の練習時間が制限されており、短時間で効率的に学習できる仕組みが求められている。こうした状況下で特に初級者は、自身の演技動画を確認しても具体的な改善点を判断しにくい。模倣学習は有効であるが[7]、理想演技との技術的なギャップが大きい場合、改善の道筋を掴みにくく、効率的な上達を妨げる要因となっている。

従来の映像分析ツール[11, 12]は、演技動画の再生や関節角度の折れ線グラフ表示など、視覚的・定量的な情報の提示には有用である[17, 20]。しかし、提示された映像や角度を学習者自身が解釈して改善点を判断する必要がある。初級者にとって十分な学習支援とはいえない。さらに、次に何を学習すべきかといった学習経路の提示や指導者のような言語フィードバックを提供する機能も不十分である。

本研究では、図1(右)のように空中で腰や膝を曲げずに宙返りを行う、後方伸身宙返りに焦点を当てた練習支援システムを提案する。本システムでは従来の視覚的・定量的分析機能に加えて、採点規則に基づく減点リスクを円グラフのアニメーションとして可視化し、姿勢の乱れが採点にどのような影響を与えるかを直感的に理解できる機能を提供する。さ

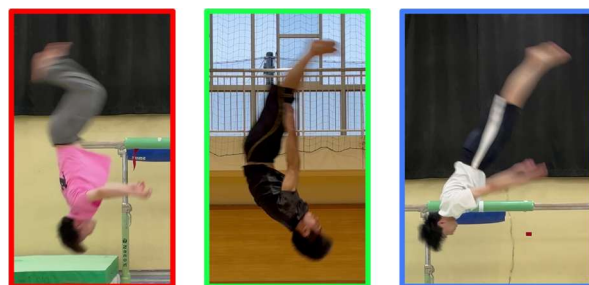


図 1. 初級者(左), 中級者(中), 上級者(右)の演技例。

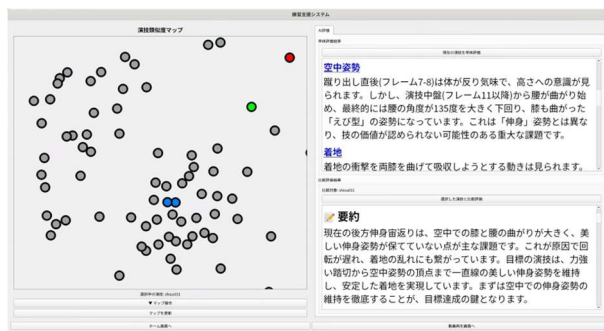


図 2. 段階的学習支援(左)と言語フィードバック(右)のUI画面。類似度マップ上の赤点・緑点・青点は図1の各枠色に対応した演技である。

らに、初級者から上級者までの演技データ(図1)を解析して類似度マップ(図2(左))を構築する。マップにより、学習者は現在の演技(赤点)を確認しながら次の目標となる演技(緑点)を参考にし、理想演技(青点)へと近づく段階的学習ができる。また、図2(右)の大規模言語モデル(LLM)[1]による言語フィードバック機能は、「空中で腰が曲がっています。腹筋と背筋に力を入れましょう。」といった具体的な改善点を自然言語で提示する。これにより、学習者は自身の課題を理解し、自律的に練習を進めることができる。

Copyright is held by the author(s).

* 静岡大学

2 関連研究

スポーツの上達支援には、補助器具や映像確認、戦術分析[6, 7, 10, 17, 20]など様々な手段がある。近年はICTの活用により[25, 26, 27]、動作の数値化や可視化、AI自動採点[8, 9]が進んでいる。器械体操分野でも3DカメラやLSTM[16]を用いた演技分析システム[11, 12]がある。これらは演技を客観的に可視化する点で有用だが、評価や改善点の解釈は選手に委ねられており、特に初級者の学習支援としては不十分である。また、上級者の演技を観察することで技術習得を促進することが報告されている[3]が、初級者は上級者との技術的ギャップが大きく、理想演技の模倣だけでは改善の方略が不明瞭になりやすい。そのため、技能を分解して段階的に学習効果を高める方法が有効とされている[4, 5]が、学習者のレベルに応じた中間目標を可視化し、学習経路の提示を支援する枠組みは十分に検討されていない。

3 後方伸身宙返り

後方伸身宙返りは、器械体操の床種目において伸身姿勢を維持した状態で後方に宙返りする技であり、空中姿勢の体幹コントロールやひねり技などの高度な技術の基盤となる基礎技である。伸身姿勢とは国際体操連盟(FIG)の採点規則[22, 23]によって「肩から腰を通り、足首までが一直線の姿勢(図1(右))であり、後方伸身宙返りのような動的な技ではわずかに胸を含む、あるいは反った姿勢も許容される」と定義されている。また、腰や膝が曲がった場合(図1(左・中))は、姿勢不十分と判断され減点となるだけでなく「屈伸宙返り」や「抱え込み宙返り」といった別の技として認定されてしまい、演技の得点に大きく影響する。この伸身姿勢は宙返り全体で維持される必要はなく、踏切(地面から足が離れた瞬間)から身体が逆さになった位置(頭が真下・足が真上の姿勢)の30度後までの区間において維持されることが求められる[24]。図1(右)は逆さ位置から身体が30度傾いた位置を示している。この区間以降は着地動作に備えて腰を曲げて減点対象とはならない。

4 ユーザインタフェース

インタフェースはデータ選択を行うホーム画面、視覚的・定量的分析機能を提供する画面、段階的学習支援とLLMによるフィードバックを提供する画面の三つで構成されている(詳細は付録の動画参照)。

4.1 動画撮影と入力

本研究では、部活動などの一般的な練習環境での実用性を重視し、高価な特殊機材を用いずに演技動画を撮影した。撮影には市販のスマートフォン

(iPhone15)やタブレット(iPad Air 第5世代)を用い、解像度4K、フレームレート30fpsとした。カメラは演技者の移動方向に対して真横の位置に三脚で固定し、演技者の全身が常にフレーム内に収まるように撮影した。本システムは以上の条件で撮影された演技動画と、その動画に対して姿勢推定(詳細は5.1節)を行って得られる姿勢情報の二つを入力とする。

4.2 視覚的・定量的分析機能

器械体操の自己評価は、まず自身の演技動画を確認することから始まる。本システムでは、基本的な動画再生(再生・一時停止・コマ送り・スロー再生)に加え、演技動画と関節を可視化した動画を切り替えて再生できる。ここでの関節可視化とは、演技者の主要な17の関節(鼻・左右目・左右耳・左右肩・左右肘・左右手首・左右股関節・左右膝・左右足首)を点として表示し、それらを線で結んだものである(図3)。本システムでは、左半身を緑色、右半身を青色、鼻を赤色で示し、さらに両肩と両股関節を結ぶ線をそれぞれ赤色で描画している。この関節可視化動画は、後述の角度グラフや減点リスクアニメーション、LLMによるフィードバックが提示する内容を直感的に理解する助けとなる。

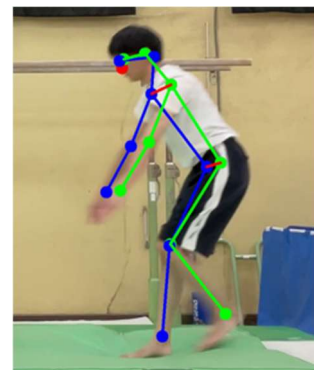


図3. 本システムの関節可視化図。

これらの基本的な動画再生機能に加えて、学習者自身が演技動画をより良く理解し、改善点を判断するために角度グラフと減点リスクアニメーションを提案する。以下に詳細を説明する。

4.2.1 角度グラフ

本機能は、「腰の角度」、「膝の角度」、「腕の挙上角度」の中から最大二項目を表示できる。図4(a)は上段に腰の角度、下段に膝の角度を示している。背景は踏切・空中・着地の三局面に応じて色分けされている(緑:踏切, 黄:空中, 赤:着地)ため、演技が現在どの局面にあるかを直感的に把握できる。図4(b)はこのグラフの演技動画から生成した連続写真であり、図4(a)と対応している。また、グラフ上には二本の基準線が表示される。赤線は動画の再生位置を示しており、動画の進行とともに横方向に移動する。

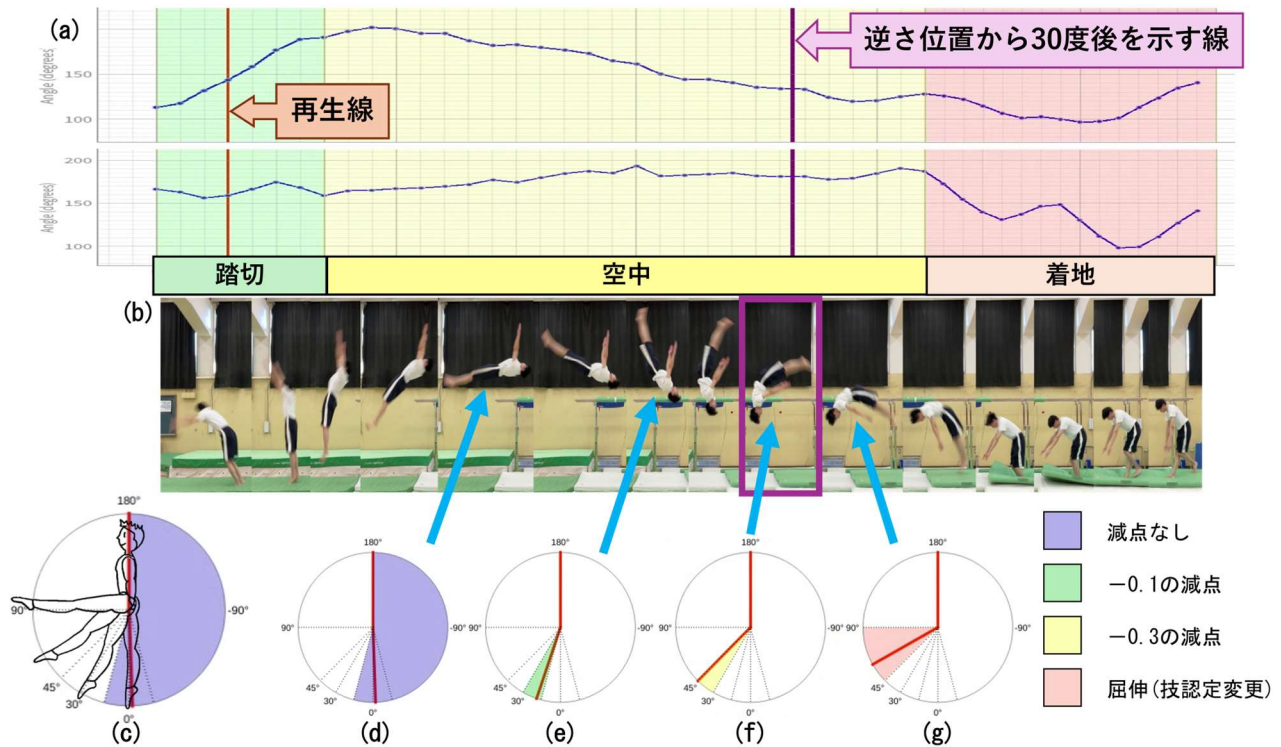


図 4. 本システムの可視化機能. (a) 角度グラフ. (b) 連続写真. (c) 減点リスクアニメーション.

紫の線は3章で述べた「伸身姿勢を維持すべき区間」の終端を示しており、宙返りにおける逆さ位置から、身体が30度以上傾いた瞬間のフレームを基準としている。学習者は空中局面の始まりから紫の線までの腰の角度を確認することで、演技が正しい伸身姿勢を維持できているかを視覚的に判断できる。

従来もこのようなグラフ表示はあったが[12]，三局面に応じた色分けや逆さ位置から30度後を示す線は今回我々が新しく提案するものである。

4.2.2 減点リスクアニメーション

本機能は、腰の角度に基づく減点リスクを円グラフのアニメーションとして可視化する。表示する角度は、理想的な伸身姿勢(180度)からの逸脱具合を表しており、値が大きいほど腰が曲がっていることを意味する。円グラフの赤線は図4(c)のように身体の軸を表し、上半身が固定され、下半身が動画の進行に伴い回転する。着色領域はFIGの採点規則[22, 23]に基づき定めた減点区分に対応している(表1)。本システムでは後方伸身宙返りが動的な技であることを考慮して腰の角度の逸脱が-15~15度の範囲であれば減点なしとした。

表 1. 角度ルール.

逸脱角度範囲(度)	技の判定	減点	色
-15~15	伸身姿勢	なし	青
15~30	伸身姿勢の逸脱	-0.1	緑
30~45	伸身姿勢の逸脱	-0.3	黄
45~90	屈伸姿勢	技認定変更	赤

図4(d)~(g)は連続写真に対応する四つの空中場面の減点リスクアニメーションの結果である。図4(f)が指す紫枠の場面は伸身姿勢を維持すべき区間の終端フレームである。図4(d)は青色を示しているため減点はない。この姿勢を維持できれば理想的だが、図4(e)では緑色を示しており減点対象(-0.1点)となる。さらに図4(f)では黄色を示している。伸身姿勢を維持すべき区間に表示された色の中で最も悪い評価が適用されるため、最終的にこの演技には-0.3点の減点が課される。図4(g)は赤色を示しているが、伸身姿勢維持の区間を過ぎているため、減点対象にはならない。このように、本機能は姿勢の変化と採点規則をリンクして可視化することで、学習者が自身の演技にどの程度の減点リスクが生じるのかを即座に理解できる。

4.3 段階的学習支援

本機能は、演技間の類似度マップを用いて、学習者が中間目標を参考に段階的に理想演技へ近づく学習アプローチを支援するものである。類似度マップとは、初級者から上級者までの様々な後方伸身宙返りの演技データを二次元上に点として配置したものである。マップ上では、点間距離が近いほど対応する演技同士が似ていることを意味する。これにより学習者は自らの演技が全体の中でどの位置にあるのかを視覚的に把握できる。

図5は、類似度マップを用いた学習の進め方を示

している。図 5(左)は、現在の演技(赤点)から理想演技(青点)へ直接到達しようとする例であり、両者の距離が大きく一度に技術的なギャップを埋めることは困難である。これに対し、図 5(右)は、現在の演技(赤点)に近い演技を順に中間目標として設定し、段階的に学習を進める例である。学習者はまず、マップ上で現在の演技(赤点)と理想演技(青点)の位置関係を確認する。次に、青点へ向かう方向に位置する近傍の演技を中間目標として選択する。距離が近い演技ほど動作構造が類似しており、無理のない改善が期待できるためである。中間目標の選び方としては、①自身と類似した身体的特徴を持つ演技者の演技を選ぶ、②理想演技(青点)の演技者が過去にどのように上達してきたかを辿る、③現在の演技(赤点)に近い他の演技者がその後どのように成長していったかを参考にする、といった多様な方法が考えられる。

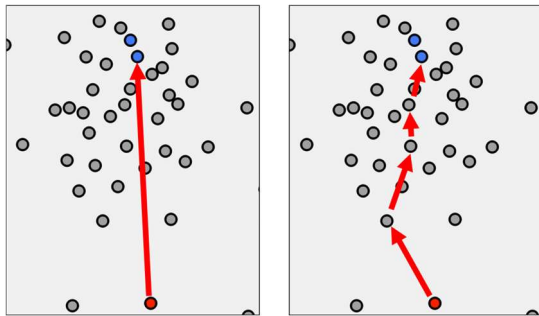


図 5. 類似度マップを用いた学習。直接理想演技を目指す場合(左)と中間目標を参考にする場合(右)。

中間目標を選択すると操作パネルが表示され、「動画を見る」「AI 評価用に選択」「手本演技として登録」といった分析アクションを実行できる。例えば、「動画を見る」を選択すれば、視覚的・定量的分析機能の UI 画面に遷移し、中間目標がどのような演技であるか確認できる。「AI 評価用に選択」では LLM によるフィードバック機能の比較対象としてこの中間目標を設定できる(詳細は 4.4.2 項)。これらを活用し、分析と練習を繰り返すことで、学習者はマップ上で自身の成長を確認しながら次のステップへ進むことができる。この「現状把握→中間目標の設定→比較・分析→練習・再評価」という学習サイクルにより、段階的に理想演技へと近づくことが可能となる。さらに、優れた演技を「手本演技として登録」することで、理想演技(青点)を更新できる。

4.4 LLM による言語フィードバック

本機能は、LLM を用いて演技进行分析し、言語フィードバックを提示することで、初級者でも具体的な学習の方向性を把握しやすくなる。提示するフィードバックは、単一の演技を評価する「単体評価」と

二つの演技を比較する「比較評価」の二種類である。

4.4.1 単体評価

LLM からの出力は、まず演技全体に対する 5 段階評価を提示し、続いて踏切・空中・着地の三局面ごとに技術的評価と改善点を示す。最後に、次の練習で意識すべき最重要ポイントを一つだけ提示する。これは、複数の改善点が同時に提示されても、学習者が一度に全てを意識することは困難であり、優先度の高い課題に焦点を絞ることで、次の練習に明確な目的を持って取り組むようにするためである。

4.4.2 比較評価

本機能は、段階的学習支援において中心的な役割を担い、類似度マップ上で選択した演技と現在の演技との具体的な差異を提示する。出力は、比較結果の要約が提示され、続いて三局面ごとに具体的な姿勢の違いや改善点が示される。さらに、両者の姿勢の差が最も大きい瞬間とその前後の姿勢が比較画像として表示される。最後に、単体評価と同様に次の練習でのワンポイントアドバイスが提示される。

5 データ処理とフィードバック生成

5.1 入力データの事前処理

撮影した演技動画は事前処理として、目視で踏切(地面から足が離れた瞬間)と着地の瞬間を特定し、踏切の 7 フレーム前から着地の 15 フレーム後までを切り出した。この切り出し範囲を、前章で述べた踏切・空中・着地の三局面に区分した。

5.1.1 姿勢推定

姿勢情報の定量化には MMPose[14]の二次元推定を用い、速度と精度のバランスに優れる RTMPose[13]を基盤モデルとして採用した。これにより、COCO データセット[15]準拠の全身 17 のキーポイント座標を時系列データとして取得した。ただし、後方伸身宙返りは高速でオクルージョンが生じやすく、完全な推定は困難である。そのため本研究では、誤推定を前提としつつ、分析結果が誤って評価されないよう必要最低限の修正を行う設計とした。実際、図 6(左)のように誤推定をそのまま用いると、良い演技が悪いと判定されてしまう場合がある。

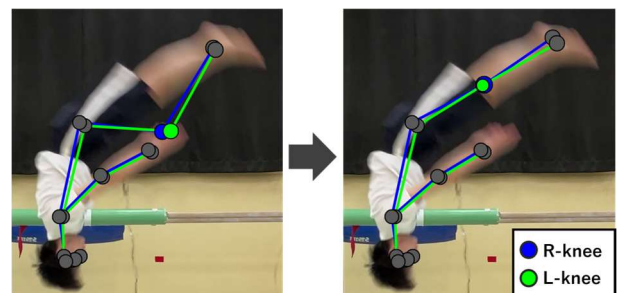


図 6. 両膝の誤推定修正。修正前(左)、修正後(右)。

またその逆も起こりうるため、図 6(右)のように修正を加える必要がある。これにより、膝の位置を精密に推定できたわけではないが、演技を適切に評価できる程度の姿勢推定結果を得ることができる。

5.1.2 誤推定箇所の修正

時刻 t におけるキーポイント座標を二次元ベクトル $\mathbf{P}_t = (x_t, y_t)$ とする。本研究では、①姿勢推定モデルが出力する信頼度スコア(推定結果の確からしさ)が閾値 σ_{conf} を下回った場合、または②加速度ベクトルの大きさが閾値 σ_{acc} を超えた場合に $\mathbf{P}_t$ を誤推定箇所として検出した。加速度の大きさは以下のように計算する。

$$\|\mathbf{a}_t\| = \|(\mathbf{P}_{t+1} - \mathbf{P}_t) - (\mathbf{P}_t - \mathbf{P}_{t-1})\| \quad (1)$$

検出された $\mathbf{P}_t$ は破棄し、前後の複数時刻における同一関節のキーポイント座標を制御点として三次スプライン補間により置換した。なお、誤検出の精度を向上させるため閾値 σ_{conf} と σ_{acc} は上半身(鼻・左右目・左右耳・左右肩)、腕部(左右肘・左右手首)、下半身(左右股関節・左右膝・左右足首)で個別の値を設定した。これは各部ごとに動きの激しさが異なるためである。以降の分析には、この修正済みのキーポイント座標を用いた。

5.2 分析アルゴリズムと LLM のプロンプト設計

本節では、事前処理されたキーポイント座標をもとに、正規化や角度計算、類似度算出を行い、本システムの各分析機能に用いるデータを生成するアルゴリズムと LLM のプロンプト設計について述べる。

5.2.1 キーポイント座標の正規化

キーポイント座標はピクセル座標であり、動画内での位置や演技者の体格に依存している。このままでは演技間の公平な比較が困難であるため、位置と大きさの二段階で正規化を行った。まず、全キーポイント座標を左右股関節の midpoint を原点とした相対座標に変換し位置の影響を除去した。次に大きさの正規化のための基準スケールを算出した。左右肩の midpoint と左右股関節の midpoint を結ぶユークリッド距離を体幹長と定義し、全フレームにおける体幹長の中央値を基準スケールとして採用した。最後に位置正規化後のキーポイント座標を基準スケールで除算し、体格に依存しない正規化キーポイント座標を得た。

5.2.2 関節角度の計算

正規化キーポイント座標から、腰の角度(肩－股関節－膝)、膝の角度(股関節－膝－足首)、腕の挙上角度(肘－肩－股関節)の三つの関節角度を左右それぞれ算出した。腰と膝の角度は FIG が定める伸身姿勢を評価するため、腕の挙上角度は、跳躍高と回転力を生み出す主要因とされる[18]ため採用した。

5.2.3 体幹角度の計算

宙返りの回転位相を求めるため、左右肩の midpoint と左右股関節の midpoint を結ぶベクトルと演技動画の水平軸とのなす角を体幹角度と定義した。この角度は、腰や膝の曲がりによる影響を受けにくく、演技の巧拙に依存しない一貫した位相を安定して表現できる。本研究では体幹角度を用いて、宙返り動作における逆さ位置から 30 度後の姿勢を特定した。

5.2.4 類似度マップの作成

同じ技でもタイミングや速度は異なるため演技長は一定ではなく、単純にフレーム番号を対応させると異なる姿勢同士を比較する恐れがある。そこで、時間的伸縮を許容する Dynamic Time Warping (DTW)[19]を用いて演技間の距離を算出した。

まず、宙返りの回転位相を表す体幹角度を指標とし、同一局面同士を対応付けるアライメントパスを求めた。この際、踏切・空中・着地の三局面ごとに独立して計算した後に連結することで、不適切な対応付けを防いだ。その後、このパスに沿って正規化キーポイント座標と関節角度を用いて演技間の距離を算出した。これは、同じ位相においてどの程度姿勢が異なるかを観察しながら演技の質を評価するためである。距離計算の際、空中局面は技の認定に直結するため最も重要であり、着地の完成度も減点に関わることから、局面ごとの重要度を反映させるため、踏切・空中・着地にそれぞれ 0.1 : 0.6 : 0.3 の重みを付与した。得られた距離は、パス長で除算することで平均化し、座標距離と角度距離の二つのスカラー値を算出した。両者はスケールが異なるため、全演技ペアの距離分布に基づいて 0~1 の範囲に最大最小正規化を施し、共通スケールに整えた上で加算して単一の演技間距離とした。最後に、距離行列の幾何構造を保ったまま演技間の関係性を視覚的に表現するため Multidimensional Scaling(MDS)[21]を適用して各演技を二次元マップ上に配置した。

5.2.5 LLM の入力プロンプト

本システムの言語フィードバック生成には Google Gemini[2]を利用し、プロンプトで「経験豊富な体操コーチ」としての役割を与えた。また、後述の入力データを分析させ、演技評価や改善点、アドバイスを提示するよう指示した。単体評価では、正規化キーポイント座標、関節角度、伸身姿勢を維持すべき区間のフレーム情報に加え、FIG の採点規則の要約を入力した。一方、比較評価では、現在の演技と中間目標の演技のキーポイント座標・関節角度に加え、DTW によるアライメントパス情報を入力した。これにより、LLM は単純なフレーム番号ではなく、技の同一局面同士を正確に対応付けて比較できる。比較評価では演技間の差異を分析することに重点を置くため、採点規則の要約は入力しない。

6 評価実験

練習支援システムとしての有効性と、各機能に対する受容性を検証するため、評価実験を実施した。今回の実験では、4名の器械体操経験者(男性2名、女性2名、年齢14~18歳、平均16.25歳、標準偏差2.06)が参加した。参加者のスキルレベルは、後方伸身宙返りを安定して実施できる中級者3名、そして同技を習得過程にある初級者1名で構成される。

6.1 練習支援システムとしての有効性評価

本実験では、①システムを使わず、演技動画を見て評価するベースライン条件と、②視覚的・定量的分析画面を用いて評価する条件、③段階的学習支援とLLMによるフィードバックを統合した画面(図2)を用いて評価する条件の三つで比較した。

参加者にはシステムの基本的な操作方法を説明した後、以下の手順で実験を実施した。まず、参加者の後方伸身宙返りの演技を撮影した。次に、研究者が事前処理を施し、本実験での分析対象として提案システムに入力した。その後、参加者は三つの条件をそれぞれ体験し、各条件の実施後に7段階のリッカート尺度(1=「全くそう思わない」、7=「非常にそう思う」)を用いたアンケートに回答した(表2)。三つの条件を体験する順序は参加者ごとにランダム化し、学習効果の順序による影響を排除した。

表 2. アンケート内容.

#	質問項目
Q1	演技を客観的に評価できましたか？
Q2	演技の改善点を明確に特定できましたか？
Q3	次の練習で何をすべきかがはっきりしましたか？
Q4	練習に対するモチベーションは向上しましたか？
Q5	演技を分析することへの興味・関心を高めましたか？
Q6	もっと自分の演技を改善したいと思いましたか？
Q7	操作は直感的でわかりやすかったですか？
Q8	表示された情報は一目で理解できるものでしたか？
Q9	演技を評価する作業はスムーズに行えましたか？

表2についてはQ1-Q3が自分の演技を客観的に評価できたかについて、Q4-Q6が練習のモチベーション向上に役立ったかについて、Q7-Q9が操作性についての質問項目となっている。図7に各条件実施後のアンケート結果を示す。

視覚的・定量的分析機能についてはQ1やQ7において動画再生のみと比較し、提案システムのスコアが高いことが分かる。この結果から、演技を客観的に評価する時に有用であり、システム操作に不明点がなかったことが予想される。LLMによる評価については、Q2やQ3において動画再生のみと比較し、提案システムのスコアが高いことが分かる。このことから、改善点を把握し、次の練習ですべきことが明確化されていることが分かる。提案システムを用いた二つの条件では、Q5, Q8, Q9のスコアがベ

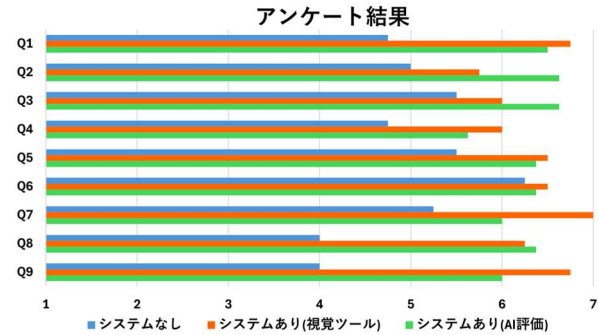


図 7. アンケート結果. 棒グラフは平均点を示す.

ースライン条件より高く、演技の分析意欲が高まり、表示情報も直感的に理解でき、評価作業も滞りなく行えることが確認された。Q6の結果から、参加者はシステムの有無にかかわらず改善意欲を示した。

6.2 各機能に対する受容性評価

全条件終了後に各機能に対する受容性評価アンケートを実施した。具体的には動画再生機能、角度グラフ、減点リスクアニメーション、連続写真、類似度マップ、LLMによる単体評価、LLMによる比較評価の7機能について「練習に役立ったか」を7段階のリッカート尺度で回答してもらい、併せて自由記述によるコメントも収集した。

図8にシステム内の各機能についてのアンケート結果を示す。いずれの機能においても高い評価を示した。また、参加者から、「減点リスクが動画に合わせて表示される機能がわかりやすい」、「類似度マップで現状を知れるのがよい」などの意見が得られた。

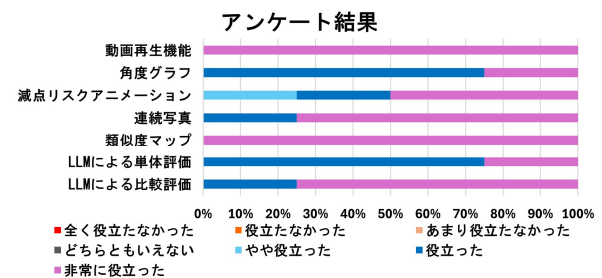


図 8. アンケート結果.

6.3 考察

評価実験(図8)の結果より、提案手法によって生成された類似度マップは高い評価を得た。これは、マップ上で同レベルの演技が近接して配置されており、ユーザにとって意味のある構造が形成されていることを示している。実験時にはマップ上に97個の演技データが存在したが、ユーザはいずれも目標演技を迷わず選択できていた。このことは操作性に対する評価(図7のQ7)からも裏付けられている。

今後はユーザスタディを通じて、本システムの有効性を客観的に検証していきたい。

謝辞

本システム開発にあたりデータの準備，システム活用に協力いただいた静岡大学体操競技部，および沼津ジュニア体操教室の皆様へ感謝します。

参考文献

- [1] K. Saab, T. Tu, W. H. Weng, R. Tanno, et al. Capabilities of gemini models in medicine. arXiv preprint arXiv:2404.18416, 2024.
- [2] Google Gemini. <https://gemini.google/about/>. (2025/08/30 確認).
- [3] R. Robertson, L. S. Germain, and D. M. Ste-Marie. The effects of self-observation when combined with a skilled model on the learning of gymnastics skills. *Journal of Motor Learning and Development*, 6(1):18-34, 2018.
- [4] J. S. Y. Chan, Y. Luo, J. H. Yan, L. Cai, and K. Peng. Children's age modulates the effect of part and whole practice in motor learning. *Human Movement Science*, 42, 261-272, 2015.
- [5] G. Irwin, and D. Kerwin. Coaching biomechanics interface: Competition and training. *ISBS-Conference Proceedings Archive*, 105-107, 2010.
- [6] J. Ochinn, G. Devineau, B. Stanciulescu and S Manitsaris. Game State and Spatio-Temporal Action Detection in Soccer Using Graph Neural Networks and 3D Convolutional Networks. In *Proc. ICPRAM*, pp. 636-646, 2025.
- [7] E. Boyer, R. G. Miltenberger, C. Batsche, and V. Fogel, V. Video modeling by experts with video feedback to enhance gymnastic skills. *Journal of Applied Behavior Analysis*, 42(4):855-860, 2009.
- [8] Y. Bai, D. Zhou, S. Zhang, J. Wang, E. Ding, Y. Guan, Y. Long, and J. Wang. Action quality assessment with temporal parsing transformer. In *ECCV*, pp. 422-438, 2022.
- [9] H. Fujiwara, and K. Ito. ICT-based judging support system for artistic gymnastics and intended new world created through 3D sensing technology. *Fujitsu scientific & technical journal*, 54(4), 66-72, 2018.
- [10] C. A. Aiken, J. T. Fairbrother, and P. G. Post. The effects of self-controlled video feedback on the learning of the basketball set shot. *Frontiers in psychology*, 3:338, 2012.
- [11] B. Reily, H. Zhang, and W. Hoff. Real-time gymnast detection and performance analysis with a portable 3D camera. *Computer Vision and Image Understanding*, 159, 154-163, 2017.
- [12] I. Masmoudi, J. Link, S. Möck, P. Nissinen, and A. Koelewijn. A SYSTEM FOR AUTOMATIC AND FAST GYMNASTICS POSE ESTIMATION AND KEY FRAME IDENTIFICATION. *ISBS Proceedings Archive*, 42(1), 636, 2024.
- [13] T. Jiang, P. Lu, L. Zhang, N. Ma, R. Han, C. Lyu, Y. Li, and K. Chen. RTMPose:Real-Time Multi-Person Pose Estimation based on MMPose. arXiv preprint arXiv:2303.07399, 2023.
- [14] MMPose Contributors. Openmmlab pose estimation toolbox and benchmark 2020.(2025/08/30 確認). <https://github.com/open-mmlab/mmpose>.
- [15] T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollar, and C. L. Zitnick. Microsoft COCO: Common objects in context. In *ECCV*, pp. 740-755, 2014.
- [16] S. Hochreiter and J. Schmidhuber. Long Short-Term Memory. In *Neural Computation*, 9(8):1735-1780, 1997.
- [17] M. Guadagnoli, W. Holcomb, and M. Davis. The efficacy of video feedback for learning the golf swing. *Journal of Sports Sciences*, 20(8):615-622, 2002.
- [18] B. Mkaouer, M. Jemni, S. Amara, H. Chaabène, and Z. Tabka. Kinematic and kinetic analysis of two gymnastics acrobatic series to performing the backward stretched somersault. *Journal of human kinetics*, 37(1):17-26, 2013.
- [19] S. Sempena, N. U. Maulidevi, and P. R. Aryan. Human action recognition using dynamic time warping. In *Proc. ICEEI*, pp.1-5, IEEE, 2011.
- [20] F. Potdevin, O. Vors, A. Huchez, M. Lamour, K. Davids, and C. Schnitzler. How can video feedback be used in physical education to support novice learning in gymnastics? Effects on motor learning, self-assessment and motivation. *Physical education and sport pedagogy*, 23(6):559-574, 2018.
- [21] J. B. Kruskal, and M. Wish. Multidimensional scaling, 11, Sage, 1978.
- [22] MAG Code of Points 2025-2028(2025/08/30 確認). https://www.gymnastics.sport/publicdir/rules/files/en_1.1%20-%20MAG%20Code%20of%20Points%202025-2028.pdf
- [23] WAG Code of Points 2025-2028(2025/08/30 確認). https://www.gymnastics.sport/publicdir/rules/files/en_1.1.%20WAG%20Code%20of%20Points%202025-2028.pdf
- [24] MAG Help Desk March 2024(2025/08/30 確認). <https://www.gymnastics.sport/publicdir/rules/files/>

en_WAG%20Help%20Desk%203%C2%BAEdition
%2015%20th%20cycle%20%E2%80%93%20March
%202024.pdf

- [25] 余 志皓, 廖 振傑, 小池英樹. VR リダイレクションを用いたゴルフスイングトレーニングシステム. WISS 第 31 回インタラクティブシステムとソフトウェアに関するワークショップ, pp. 10-16, 2023.
- [26] 松本 高, Erwin Wu, Chen-Chieh Liao, 小池 英樹. 拡張現実を用いた姿勢可視化によるスキートレーニングシステム. WISS 第 31 回インタラクティブシステムとソフトウェアに関するワークショップ, pp. 17-24, 2023.
- [27] 鈴木 湧登, 坂本 大介, 小野 哲雄. Gino .Aiki: 合気道の身体の使い方の習得を支援する MR ソフトウェア. WISS 第 30 回インタラクティブシステムとソフトウェアに関するワークショップ, pp. 46-55, 2022.

Graph Guide: 低視力者支援のためのセマンティック Focus+Context グラフ表示

SECHAYK Yotam* LI Yuan* RAVE Hennes† COLLEY Mark‡
SHAMIR Ariel§ 五十嵐 健夫*

概要. データ可視化グラフは情報伝達的手段として広く利用されているが、弱視者 (Low Vision Individuals: LVI) にとっては依然として困難を伴うものであり、注目している情報とその周囲にある文脈情報の両方にアクセスすることが難しい。LVI を支援する最も一般的なツールは画面拡大ツールであるが、一様な拡大によってグラフの一部が視野から外れ、パン操作やレイアウトの記憶が求められるため、高い認知的負荷を引き起こす。これらの課題を明らかにするため、5名の LVI を対象とした予備的調査を実施した。その結果、参加者は凡例といった文脈的要素を推測に頼りながら探ることが多く、それが余分な認知的負荷を生じさせていた。これらの知見に基づき、GRAPH GUIDE を開発した。GRAPH GUIDE は Focus+Context の発想に基づき、視野外にある文脈的要素 (例: 軸や凡例) を意味的に抽出し、視野内へ投影する手法である。さらに GRAPH GUIDE を検証するため、6名の LVI 参加者を対象に予備的評価を実施した。その結果、GRAPH GUIDE は既存の従来ツールによるワークフローと比較して、知覚的アクセスを改善し、労力を削減し、システムユーザビリティ尺度 (System Usability Scale: SUS) においてより高いスコアを達成した。本研究の知見は、新たな機能を画面拡大ツールと統合する際に生じる視覚的な情報過多の管理といった課題を明らかにしており、今後さらなる研究の必要性を示唆している。

1 はじめに

折れ線グラフや棒グラフといったデータ可視化グラフは、教育のおよび職業的な環境において情報伝達的手段として広く利用されている [19]。グラフを効果的に理解し読み取るためには、個々のデータ点といった現在注目している情報だけでなく、それらが位置づけられる周囲の文脈全体を視認できることが必要である [10]。しかし、弱視者 (Low Vision Individuals: LVI) は、完全には矯正できない視覚的制約を有しているため [38]、文脈の中で各データ点を読み取ることは困難な課題である [37][28]。

LVI は、低視力、視野狭窄、あるいはその両方といった多様な形で視覚情報へのアクセスに課題を抱える (図 1)。視覚情報を言語情報に変換する方法なども存在するが、LVI は残存視力を用いて自ら直接視覚情報にアクセスすることを好む傾向が報告されている [33]。

デジタルデバイスにおいて LVI の視覚的アクセスを支援するために最も広く利用されているツールは、画面拡大ツールである [22][35][36]。画面拡大ツール [2][1] は通常、画面上のコンテンツを一様に拡大し、全画面モードあるいはレンズモードで使

とが最も多い。

しかし、画面拡大ツールを用いてグラフを読む場合、拡大によって一部の領域が視野から外れたり、表示が遮られたりする [36][28][35]。特にグラフのような複雑な情報にアクセスする際には、LVI はパン操作、ズーム操作、レイアウトの記憶、さらには心的モデルの構築を行う必要があり、それが大きな認知的負荷につながる [22][28][33][23]。

LVI が多様な一般的グラフの読解課題にどのように取り組み、その文脈的ニーズを理解するために、5名の LVI を対象とした形成的評価を実施した。先行研究と同様に、本研究の結果からも、画面拡大ツールが依然として LVI を支援するほぼ唯一のツールであることが確認された [37][28][33][30]。さらに、グラフ上で関連する文脈を探る際、参加者はしばしば推測に頼ってパン操作を行い、それが誤りや時間・労力の浪費につながるということが明らかになった。

我々は、グラフ読解において重要となる文脈要素

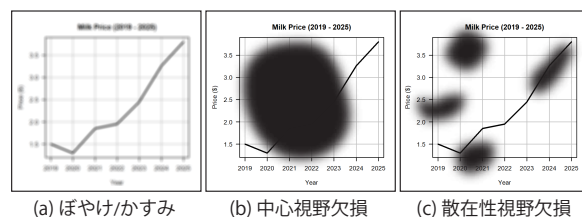


図 1: LVI として視覚問題の類型例。

Copyright is held by the author(s).

* 東京大学

† University of Münster

‡ University College London

§ Reichman University

として、軸、凡例、およびガイドラインを特定した。これらの知見に基づき、我々は LVI が既存のツールを用いて細部に焦点を当てながら、同時に文脈情報へアクセスできるように支援するインタラクション手法である GRAPH GUIDE を開発した。GRAPH GUIDE は画面拡大ツール [2][1] と連携して動作し、カーソル位置に基づいて視野外にある文脈情報を動的にユーザの視野内へ投影する。

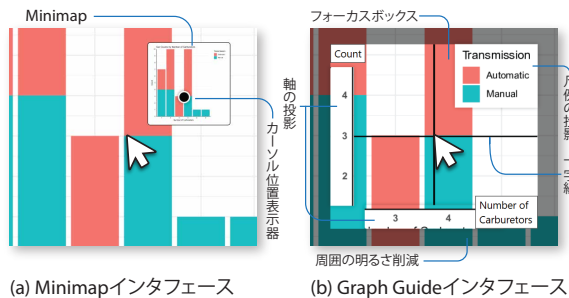


図 2: ユーザインタフェースの概要。

GRAPH GUIDE の予備的評価を、6 名の LVI を対象にグラフ読解シナリオで実施し、縮小表示のオーバービューペインといった従来手法や、参加者が日常的に利用している支援ツールとの比較を行った (図 2)。その結果、画面拡大ツールに加えて、LVI はしばしば複数の拡大モード (例: ピンチズーム、身体的な画面位置の調整) を組み合わせて利用しており、GRAPH GUIDE 使用時にはこれらへの適応が求められることが示唆された。また、GRAPH GUIDE が追加的な視覚情報を提示することにより、一部の利用者にとっては視覚的負担が増加する可能性があることも確認された。しかし一方で、GRAPH GUIDE は参加者の知覚的アクセスを有意に改善し、全体的な労力を軽減するとともに、グラフ読解における有用性を高めることが示された。特に、システムユーザビリティ尺度 (System Usability Scale: SUS) [7][4] において有意に高いスコアが得られた。

質的評価からは、参加者がグローバルな画面拡大よりも局所的な拡大を好み、事前に情報を提示するプロアクティブ型支援 (例: クリックして非表示) よりも、必要に応じて能動的に操作して情報を得るリアクティブ型支援 (例: クリックして表示) を好む傾向が明らかとなった。さらに、LVI の利用パターンが多様であることから、GRAPH GUIDE に他の手法 (例: 縮小表示のオーバービュー) の要素を取り入れることが、グラフへのアクセス性を一層向上させる可能性が示唆された。とはいえ、依然として重要な課題は、提示する視覚情報量と利用可能な表示領域とのバランスをいかに取るかである。

2 関連研究

棒グラフや折れ線グラフを含むデータ可視化グラフは、視覚的複雑さのために LVI にとって依然として困難である [28][37]。拡大機能を使用すると、データ点と注釈との関係が分断され、文脈を失う原因となる。近年の研究では、グラフを表形式に変換する手法 [25]、音によるインタラクション [9]、あるいは会話型エージェントを用いる手法 [31][21][15] などによって、これらの障壁に対処しようとする試みが行われている。また、一度に表示するデータ量を制限することで、拡大ツールに依存しないアクセスを実現するツールも提案されている [27]。しかし、これらのツールはグラフの再構築 [27] に依存していたり、会話型エージェントの利用 [31] など、LVI に新たなナビゲーション方法の変更を求めたり、あるいは表形式への変換 [25] のように視覚に基づくナビゲーションを省略してしまうことが多い。後者は LVI にとって必ずしも望ましい方法ではない [33]。

これに対し、GRAPH GUIDE は元のグラフを保持しつつ画面拡大ツールを支援し、視覚的探索へのアクセスを提供する。その結果、グラフ再構築に伴う誤りを回避し、既存の支援ツールにおける LVI の利用を補完し、視覚的探索時の負担を軽減する。

Focus+context および overview+detail のパラダイムは、情報可視化において細部と広い文脈の把握の両立を支援する [14][10][6]。Focus+context 手法は、フィッシュアイビュー [3] などに代表され、歪みや適応的スケーリングを通じて細部と文脈を単一のビュー内に統合する。一方、overview+detail は、それらをリンクされたウィンドウに分離することで実現される [34][40][8]。これらのパラダイムの拡張として、ズーム可能なインタフェースにおいて要素の詳細を表示または抑制する *semantic zooming*、および要素の視覚的外観を変えずに小さな対象や遠方の対象とのインタラクションを容易にする *semantic pointing* が提案されている [10]。しかし、これらの手法は主に晴眼者向けのインタラクション最適化を目的としており、特に重要な意味的要素 (例: 軸、凡例) が拡大視野の外にある場合に生じる LVI のアクセシビリティニーズには十分に対応していない。

本研究は Focus+context の概念を基盤とし、視野外にある意味的に重要な文脈情報を視野の中へ投影することで、LVI によるより効果的なナビゲーションと解釈を支援する。

3 予備的調査

多様なグラフにおける読解課題に対して LVI がどのように取り組み、またその文脈的ニーズを理解するために、5 名の LVI を対象とした形成的評価を実施した。本研究では、グラフの読解力を評価するために広く用いられている可視化リテラシー評価

表 1: 形成的調査の参加者 (N=5). 拡大機能はスクリーン拡大ソフトの使用を示す.

ID	年齢/性別	診断名	発症	法的失明	拡大機能
F1	28/女性	斜視, 神経学的視覚障害	先天性	はい	はい
F2	32/男性	左眼失明, 視野狭窄	先天性	はい	はい
F3	30/女性	網膜色素変性症 (RP), 視野狭窄	先天性	はい	いいえ
F4	41/女性	アルビニズム, 光過敏	先天性	はい	はい
F5	37/男性	アルビニズム, 乱視	先天性	はい	はい

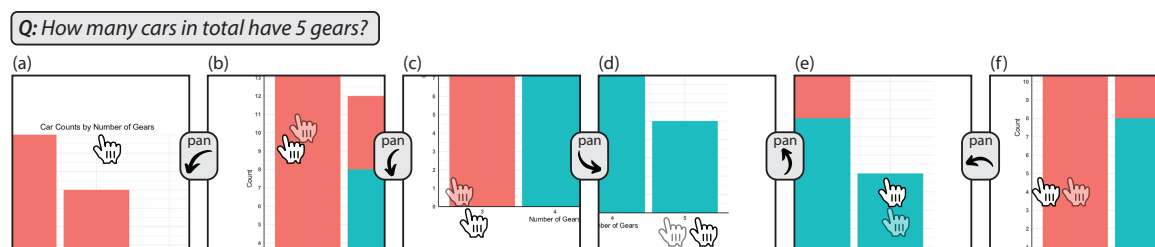


図 3: デフォルトツールを用いた積み上げ棒グラフでのインタラクションフロー.

テスト (Visualization Literacy Assessment Test: VLAT) [24][26] を利用した. VLAT は棒グラフ, 折れ線グラフ, 面グラフなど 12 種類の代表的な可視化手法に基づいて設計されており, グラフを読み取り理解する能力を測定できる. その簡略版である Mini-VLAT[26] は, 各グラフタイプにつき 1 問のみで構成されている. Mini-VLAT を用いることで, 実際のグラフ読解シナリオへのアクセスを可能にしつつ, 疲労を抑制することができる. 参加者はまずデジタル版を体験し, 続いて紙媒体版を用いた振り返りを行った.

3.1 調査手順

知人を通じて 5 名の参加者を募集した (表 1). 参加者は全員が弱視であり, 残存視力を用いてグラフを読むことを好んでいた. 5 名のうち 4 名は, 日常的に画面拡大ツールを使用していた. 参加者には, 1 時間の実験参加に対して 20 米ドルの謝礼を支払った. 本研究は所属機関の倫理審査委員会の承認を得て実施された.

各セッションは, 参加者の基本的な属性に関する質問およびグラフへのアクセス経験に関する質問から始めた. その後, 参加者は通常利用しているアクセシビリティ支援ツール (例: 画面拡大ツール) を用いながら, Mini-VLAT クイズ¹に回答した. デジタル版のクイズを体験した後, 提供された PDF 版²を用いた紙媒体版のクイズ体験を含む半構造化

インタビューを実施した. 各セッションは記録され, OpenAI Whisper [29] を用いて逐語録に書き起こし, 分析に供した.

3.2 調査結果

多くの参加者は教育的あるいは職業的な文脈においてデータ可視化グラフを利用した豊富な経験を有しており, 参加者 F5 は主にニュース報道の中でそれらに接していた. 先行研究と一致して, 参加者は低コントラストの配色や低解像度のラスター画像といった視覚的課題を報告した [28][37].

さらに, 複雑な視覚表現が負担となり, グラフの解釈を妨げる場合があると述べた. 例えば, 参加者 F1 は次のように述べている. 「一つの領域に非常に多くの情報が密集している場合, 例えばドットプロットや散布図のようなものでは, 軸上でどこにクラスタが存在するのか, あるいは実際にいくつの点があるのかを見分けることが難しい. ただ一つの大きな塊として認識してしまうことが多い.」

参加者はまた, グラフ内の異なる領域に視線を移動しなければならないことに伴うナビゲーション上の課題を指摘し, その過程で生じる不便さに対する不満を表明した (図 3). 彼らは, グラフ間でレイアウトが一貫していることが, こうした課題の軽減につながると強調した. 画面拡大ツールを使用する参加者 F2 は次のように述べている. 「凡例や軸を何度も参照しなければならない. 特にデータが動的に変化している場合, 全体像を把握するのは大きな課題である.」 視野が狭い参加者 F3 も次のように共有した. 「自分の視野では, 一度に全体を見渡すことができないので, 異なる領域に焦点を合わせ, 物理

¹ <https://tools.visualdata.wustl.edu/experiment/>; Accessed 30.08.2025

² https://washuvis.github.io/minivlat/Mini-VLAT_V2.pdf; Accessed 30.08.2025

的に移動しながらプロット全体を見回さなければならぬ。」参加者は、凡例の典型的な配置といった慣習的なデザインに依拠して効率的にグラフを解釈していた。しかし、グラフがこうした期待から逸脱した場合、読み取りやナビゲーションは一層困難になった。参加者 F1 は次のように述べている。「ヒストグラムは棒グラフと非常によく似ているので難しい。どの種類のグラフを見ているのか、そしてどのような情報を読み取るべきかを理解するのに時間がかかることがある。」

その結果、参加者はグラフのレイアウトをより柔軟に制御できるようにすることで、自身の期待に適合させ、ナビゲーションを容易にしたいという要望を表明した。参加者 F1 は次のように述べている。「レイアウトを少し調整できるオプションがあれば非常に助かる。」参加者 F2 もまた、次のように強調した。「全体像と詳細を同時に把握する必要がある。」Mini-VLAT クイズにおいて、参加者 F3 は次のようにコメントしている。「凡例の位置が一貫している方がずっと良い。『どこを見ればよいか』をそれほど探す必要がなくなる。」一部の参加者は、レイアウトがよりコンパクトなため紙媒体版の方が容易だと感じたが、総じてデジタル拡大ツールの快適さを評価し、電子版での実施を好んだ。

電子版と印刷版に対する参加者の嗜好は、画面拡大ツールなどの支援ツールへの習熟度によって多様であった。情報がより限られた空間に配置されているため視覚的なナビゲーションの負担が軽減されるとして、印刷版を好む参加者もいた。また、レイアウトがよりコンパクトであることから印刷版の方が使いやすいと感じる者もいた。しかし最終的には、拡大ツールが快適さを提供するため、すべての参加者がデジタル版を支持した。

4 Graph Guide の設計

予備的調査の知見に基づき (3 節), LVI のグラフアクセスを支援する手法として GRAPH GUIDE を開発した。GRAPH GUIDE を用いることで、LVI は画面拡大ツールやカラーフィルタ (例: 反転) などの支援ツールを継続して利用しながら細部を確認でき、さらにカーソル付近の位置で関連する文脈情報を動的に得ることが可能となる (図 2)。GRAPH GUIDE は、軸や凡例といったグラフの文脈情報をユーザのカーソル付近に投影することで機能する。これを実現するために、GRAPH GUIDE には凡例や軸のバウンディングボックス、さらに存在する場合はタイトルといった最小限の情報が必要となる。提供されるインタフェースは表示が一貫しており、元のグラフの視覚的アイデンティティを保持している。また、設定項目によって利用者のニーズに合わせた調整が可能である。

GRAPH GUIDE は React [12] を用いた Web コ

ンポーネントとして実装されている。現在の実装では、凡例や軸などの要素は、著者らが事前に手動で切り出している。将来的には、画像理解 AI などを利用して、自動化していく予定である [16][41][25]。次節では、LVI を支援するために GRAPH GUIDE に実装された各種機能について詳述する。

4.1 機能とパラメータ

フォーカスボックス カーソルの周囲に文脈情報を表示する領域を定義する。利用者は幅と高さを独立して変更でき、枠線の可視性、色、および太さを調整することが可能である。

周囲の明るさ削減 フォーカスボックスの周囲を暗くし、画面上の要素による視覚的な雑然さを低減する。この機能は利用者がオンまたはオフに切り替えることができる。

軸の投影 グラフに軸が存在する場合、その軸をフォーカスボックス内に投影する。X 軸は領域の下端に、Y 軸は左端に固定する。また、軸タイトルが存在する場合は、それらも表示される。利用者は投影される軸の割合 (0% から 100%) を調整できる。投影される軸のサイズは、元のグラフにおけるサイズと同一である。

凡例の投影 グラフに凡例が存在する場合、それをフォーカスボックス内の右上に投影する。投影される凡例のサイズは、元のグラフにおけるサイズと同一である。

十字線 ボックスの中心で交差する縦線と横線の 2 本を表示する。この十字線はガイドラインとして機能し、グラフにおける一般的な機能である。利用者は十字線の太さ、色、透明度、および拡張範囲 (ボックスの大きさまで) を調整できる。

動的な文脈表示 フォーカスボックス内の文脈情報は、グラフ上のカーソル位置に基づいて動的に変化する。例えば、X 値が「1985」、Y 値が「10」の棒にカーソルを合わせた場合、投影された軸には対応する値が表示される。

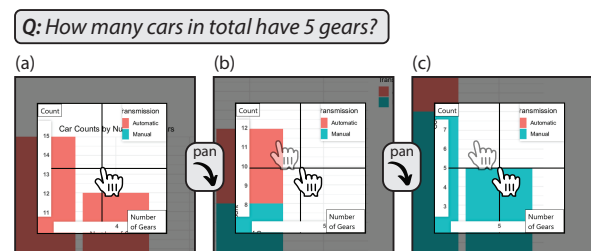


図 4: GRAPH GUIDE の操作フロー。

表 2: 参加者 (N=6) . 拡大機能はスクリーン拡大ソフトの使用を示す.

ID	年齢/性別	診断名	発症	法的失明	拡大機能
P1	49/ノンバイナリー	両側視神経症, 視野狭窄	後天性	はい	いいえ
P2	43/アジェンダー	未熟児網膜症, 視野狭窄	先天性	はい	いいえ
P3	66/男性	視神経症	先天性	はい	はい
P4	28/男性	無虹彩症, 光過敏	先天性	はい	はい
P5	35/男性	アルビニズム, 眼振, 光過敏	先天性	はい	はい
P6	39/女性	糖尿病網膜症, 視野狭窄	後天性	いいえ	はい

4.2 Graph Guide の体験

GRAPH GUIDE を利用する際, 利用者は自身の支援ツールに合わせて設定を調整する. 例えば光過敏と視界のぼやけを持つ弱視者 Tom は, 全画面拡大に収まるようフォーカスボックスを調整し, 雑然さを避けるため暗転と枠線を有効化, 十字線を高コントラストかつ太線で端まで表示する.

グラフを表示すると (図 4), Tom は拡大率 300% でタイトルへパンし, 軸情報も同時に確認できる. 質問に答えるため左右にパンして X 軸へ移動し, 下方へ移動して項目を読む. このとき十字線によりカーソルを動かさず視線をずらすだけで Y 値を確認でき, 操作は 2 回に抑えられ, 従来の 5 回から大幅に効率化された (図 3).

5 予備的評価

GRAPH GUIDE の予備的評価を, 6 名の LVI を対象にグラフ読解シナリオで実施し, ツールを使用しない条件および Minimap を用いた overview+detail 手法との比較を行った. 3 節と同様に Mini-VLAT[26]を使用した, 本評価では追加で 2 種類のバリエーションを作成した. 予備的調査参加者 5 名と予備的評価参加者の 6 名との間に重複はない.

バリエーションは, データにノイズを加えた並べ替え操作と, 質問の方向性や目的の変更 (例: 「最大」を「最小」に変える) という 2 つの操作によって作成された. これらのグラフと質問の改変は, データ可視化の専門家によって承認された. 順序効果を防ぐため, 条件とバリエーションの提示順をランダム化し, カウンターバランスを取った.

5.1 実験手順

弱視のある 6 名の参加者を, 団体を通じた募集およびスノーボールサンプリングによって集めた (表 2). セッションは 1.5 時間にわたり実施され, 参加者には 30 米ドルが支払われた. 本研究は所属機関の倫理審査委員会の承認を得て実施された.

各セッションは, 属性情報の収集から始まり, その後 3 つの条件が提示された: *None* (追加ツールなし), *Minimap* (カーソル付近に固定されたオーバー

ビューウィンドウ), および 4 節で述べた GRAPH GUIDE (順序はランダム化およびカウンターバランス化) である. すべての条件において, 参加者は自身の支援ツールを利用することができた. 課題は短い練習用クイズから開始され, その後, 参加者は設定を調整し, 標準時間の 150% に設定された制限時間内でグラフ読解クイズ (折れ線, 円, 棒) を完了した [26, 20]. 各条件の終了後, 参加者はワークロード評価 (NASA-TLX [17]), ユーザビリティ評価 (SUS [7]), および独自の体験評価質問紙に回答した. 研究の最後には半構造化インタビューを実施した. セッションは記録され, OpenAI Whisper [29] を用いて文字起こしが行われ, 分析に供された.

データ分析は Shapiro-Wilk 検定 [32] から開始された. 正規分布が確認されたデータは, 線形混合効果モデル [5], ANOVA[13], および Tukey の HSD[18] を用いて分析された. 一方, 正規性が確認されなかった場合には, ART ANOVA[39] および ART コントラスト [11] を用いた.

5.2 実験結果

全体として, 個別の NASA-TLX スコア [17] に基づきタスク負荷の低減傾向が見られたが, 統計的に有意ではなかった. 初期の結果は SUS スコア [7] に対する条件の有意な主効果を示した ($F(2,10) = 7.02$, $p = 0.012$, $\eta_p^2 = 0.58$). 事後のペアごとの比較では, GRAPH GUIDE 条件 ($\mu = 81.67$, $\sigma = 16.02$) および Minimap 条件 ($\mu = 76.25$, $\sigma = 10.09$) のいずれも, None 条件 ($\mu = 49.17$, $\sigma = 24.38$) より有意に高いスコアを示した (それぞれ $p = 0.015$, $p = 0.038$). ただし, GRAPH GUIDE 条件と Minimap 条件の間には有意差は認められなかった. 練習用クイズの結果には, 有意な差は認められなかった.

全体的な体験は両ツール条件において改善が見られた (図 5). 特に *information access* において最も強い効果が確認された ($F(2,8) = 6.35$, $p = 0.022$, $\eta_p^2 = 0.61$). この結果, GRAPH GUIDE 条件 ($\mu = 3.83$, $\sigma = 0.75$) は Minimap 条件 ($\mu = 2.67$, $\sigma = 1.03$, $p = 0.021$) より有意に優れていた. さらに

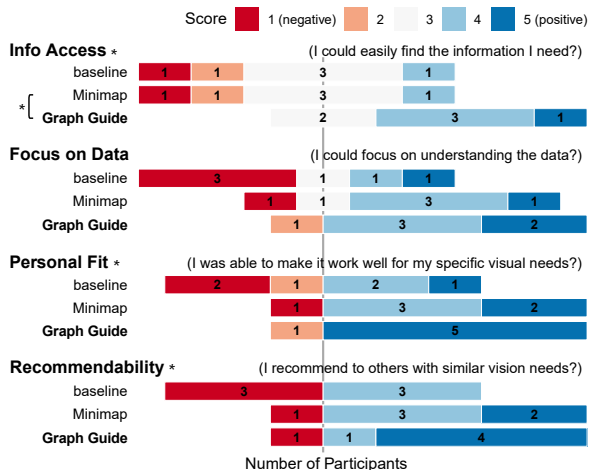


図 5: アンケートの得点結果. *は $p < 0.05$.

に、*Personal Fit*、*Recommendability*、*Focus on Data* においても GRAPH GUIDE を支持する改善が示されたが、事後比較では有意には至らなかった。これはサンプルサイズの小ささによる可能性がある。

Graph Guide に対する総合的評価 参加者は概ね、GRAPH GUIDE が視野外の文脈情報を自らの視野内に投影する点を評価した。特に十字線については影響が大きいと指摘され、推測に頼る必要を減らし、自信を高めたと報告された (P2, P3, P4, P6)。例えば、参加者 P3 は次のように述べている。「十字線によってより多くの制御感を得られ、その結果として自信が高まり、十分な時間があれば正答を導く可能性も高まった。」同様に、参加者 P5 は次のように述べている。「異なるグラフにおける凡例がボックス内に表示される点が良かった。」総じて、すべての参加者は従来の拡大のみのワークフローと比較して、効率性と安心感の両方を得られたと述べた。

拡大手法 すべての参加者において、システムの画面拡大ツール、ブラウザレベルのズーム、ピンチズーム操作など、複数の拡大モードを組み合わせる一貫した傾向が観察された。これらはしばしば、画面に身体を近づけるといった物理的な戦略と併用されていた。例えば、参加者 P1 はブラウザズームと画面拡大を組み合わせ、P5 はピンチズームとシステム拡大を切り替えて使用していた。これらの行動は、GRAPH GUIDE が単一のアクセスモードを前提とするのではなく、多様な拡大手法に適応する必要があることを示唆している。

局所的拡大機能への期待 参加者はしばしば、GRAPH GUIDE 自体が局所的な拡大を提供し、静的な文脈ボックスというよりも拡大レンズのように機能することを期待していた (P1, P2, P3, P5, P6)。参加者 P2 は次のように述べている。「テキストサイ

ズを大きくしたい。ボックスが拡大鏡のように機能し、そこに情報も表示されるようになってほしい。」これは既存の画面拡大ツールへの不満を示しており、ツール内に局所的な拡大機能を組み込むことで対応できることを示唆している。

視覚的負荷 利点がある一方で、一部の参加者 (P1, P2, P4, P5, P6) は、GRAPH GUIDE が煩雑さや圧迫感を引き起こす可能性を指摘した。例えば、参加者 P6 は次のように述べている。「時々少し圧倒される感じはあった。特に複視があるので、それでも実際に頼れる手段が得られるため価値はあった。」同様に、参加者 P2 はフォーカスボックスに関連する情報を覆い隠す可能性があるとして指摘した。多くの参加者はこのツールを有用だと感じていたが、これらのコメントは、追加文脈情報と限られた表示領域のバランスを慎重に取る必要があることを示している。

先行手法との組み合わせ 複数の参加者 (P1, P4, P5) は、GRAPH GUIDE に overview+detail 機能 (例: Minimap) を統合することで、局所的な文脈に加えて全体的な位置把握も支援できると提案した。参加者 P4 は次のように述べている。「Minimap と GRAPH GUIDE を組み合わせたかった。両方に利点があるからだ。Minimap はグラフ全体での位置把握を助け、GRAPH GUIDE はグラフ内部での位置把握を助けてくれた。」このことは、Focus+Context と overview+detail の支援を組み合わせたハイブリッド設計の可能性を示している。ただし、視覚的な煩雑さへの対応には十分な配慮が必要である。

6 おわりに

データ可視化グラフは情報伝達に有効だが視覚的に複雑であり、LVI にとってアクセスが困難である。画面拡大ツールは詳細と文脈の同時把握を妨げ、過度なパン操作を生む。我々は 5 名の LVI を対象に Mini-VLAT で予備的調査を行い、グラフ読解時の課題を調査した結果、参加者はしばしば推測に基づいてパンを行っており、一貫したレイアウトの有効性が示唆された。この知見を基に、我々は GRAPH GUIDE を提案した。GRAPH GUIDE はカーソル位置に応じて軸や凡例など視野外の意味的に重要な文脈情報を投影する focus+context 手法である。6 名の LVI による予備的評価では、GRAPH GUIDE がグラフアクセスを改善し、SUS スコア 81.67 を達成して None および Minimap 条件を上回った。一方、参加者は複数の拡大手法を併用しており、より柔軟な設計が求められることに加え、視覚的な雑然さが課題として指摘され、今後の検討すべき方向性であることも示された。我々のアプローチを画面拡大ツールと統合することで、より広範なアクセス支援へ発展できる可能性を示す。

謝辞

本研究にご協力いただいた参加者の皆様に深く感謝いたします。本研究の遂行にあたり、多大なご支援をいただいたソフトバンク株式会社および Beyond AI 研究推進機構に心より感謝申し上げます。

参考文献

- [1] Change Zoom settings for accessibility on Mac. Apple Support, 2025. Accessed August 30, 2025; instructions for the built-in Zoom feature in macOS.
- [2] Magnifier (Windows). Wikipedia, 2025. Accessed August 30, 2025; includes overview and history of the Windows screen magnifier app.
- [3] A. J. Afram, J. Briedis, D. Fujiwara, R. J. Jacob, C. G. Cao, and D. Kahle. Evaluation of Semantic Fisheye Zooming to Provide Focus+context. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 51(5):459–463, Oct. 2007.
- [4] A. Bangor, K. , Philip T., and J. T. and Miller. An Empirical Evaluation of the System Usability Scale. *International Journal of Human-computer Interaction*, 24(6):574–594, July 2008.
- [5] D. Bates, M. Mächler, B. Bolker, and S. Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48, 2015.
- [6] P. Baudisch, N. Good, V. Bellotti, and P. Schraedley. Keeping Things in Context: A Comparative Evaluation of Focus plus Context Screens, Overviews, and Zooming. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '02, pp. 259–266. Association for Computing Machinery, Apr. 2002.
- [7] J. Brooke. SUS: A quick and dirty usability scale. Technical Report TR 38, Digital Equipment Corporation, Reading, UK, 1996. Originally written in 1986, published in 1996 in P. W. Jordan, B. Thomas, B. A. Weerdmeester, & I. L. McClelland (Eds.), *Usability Evaluation in Industry*. London: Taylor and Francis.
- [8] S. Burigat, L. Chittaro, and E. Parlato. Map, Diagram, and Web Page Navigation on Mobile Devices: The Effectiveness of Zoomable User Interfaces with Overviews. In *Proceedings of the 10th International Conference on Human Computer Interaction with Mobile Devices and Services*, MobileHCI '08, pp. 147–156, New York, NY, USA, Sept. 2008. Association for Computing Machinery.
- [9] P. Chundury, Y. Reyazuddin, J. B. Jordan, J. Lazar, and N. Elmqvist. TactualPlot: Spatializing Data as Sound Using Sensory Substitution for Touchscreen Accessibility. *IEEE Transactions on Visualization and Computer Graphics*, 30(1):836–846, Jan. 2024.
- [10] A. Cockburn, A. Karlson, and B. B. Bederson. A Review of Overview+detail, Zooming, and Focus+context Interfaces. *ACM Comput. Surv.*, 41(1):2:1–2:31, Jan. 2009.
- [11] L. A. Elkin, M. Kay, J. J. Higgins, and J. O. Wobbrock. An Aligned Rank Transform Procedure for Multifactor Contrast Tests. In *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST '21)*, pp. 754–768. ACM, 2021.
- [12] I. Facebook. React - A JavaScript library for building user interfaces. <https://reactjs.org>, 2024. Accessed: 2024-09-12.
- [13] R. A. Fisher. *Statistical Methods for Research Workers*. Oliver & Boyd, Edinburgh, UK, 1925.
- [14] G. W. Furnas. Generalized fisheye views. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '86)*, pp. 16–23. ACM, 1986.
- [15] J. Gorniak, Y. Kim, D. Wei, and N. W. Kim. VizAbility: Enhancing Chart Accessibility with LLM-based Conversational Interaction. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*, UIST '24, pp. 1–19, New York, NY, USA, Oct. 2024. Association for Computing Machinery.
- [16] K. Goswami, P. Mathur, R. Rossi, and F. Deroncourt. ChartCitor: Multi-Agent Framework for Fine-Grained Chart Visual Attribution, Feb. 2025.
- [17] S. G. Hart and L. E. Staveland. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In P. A. Hancock and N. Meshkati eds., *Advances in Psychology*, Vol. 52 of *Human Mental Workload*, pp. 139–183. North-Holland, Jan. 1988.
- [18] T. Hothorn, F. Bretz, and P. Westfall. Simultaneous inference in general parametric models. *Biometrical Journal*, 50(3):346–363, 2008.
- [19] J. In and S. Lee. Statistical Data Presentation. 70(3):267–276.
- [20] T. Kellaghan and V. Greaney. Public Examinations Examined. *World Bank Publications*, pp. 1–353, 2019.
- [21] N. W. Kim, G. Ataguba, S. C. Joyner, C. Zhao, and H. Im. Beyond Alternative Text and Tables: Comparative Analysis of Visualization Tools and Accessibility Methods. *Computer Graphics Forum*, 42(3):323–335, 2023.
- [22] R. L. Kline and E. P. Glinert. Improving GUI Accessibility for People with Low Vision. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '95*, pp. 114–121, Denver, Colorado, United States, 1995. ACM Press.
- [23] H.-N. Lee, S. Uddin, and V. Ashok. Table-View: Enabling Efficient Access to Web Data Records for Screen-Magnifier Users. In *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '20, pp. 1–12, New York, NY, USA,

- Oct. 2020. Association for Computing Machinery.
- [24] S. Lee, S.-H. Kim, and B. C. Kwon. VLAT: Development of a Visualization Literacy Assessment Test. *IEEE Trans. Vis. Comput. Graph.*, 23(1):551–560, 2017.
 - [25] F. Liu, J. M. Eisenschlos, F. Piccinno, S. Krichene, C. Pang, K. Lee, M. Joshi, W. Chen, N. Collier, and Y. Altun. DePlot: One-Shot Visual Language Reasoning by Plot-to-Table Translation. In *Annual Meeting of the Association for Computational Linguistics*, Vol. abs/2212.10505, Dec. 2022.
 - [26] S. Pandey and A. Ottley. Mini-VLAT: A Short and Effective Measure of Visualization Literacy. 42(3):1–11.
 - [27] Y. Prakash, P. A. Khan, A. K. Nayak, S. Jayarathna, H. N. Lee, and V. Ashok. Towards Enhancing Low Vision Usability of Data Charts on Smartphones. Vol. 31, pp. 853–863, 2025.
 - [28] Y. Prakash, A. K. Nayak, S. Jayarathna, H. N. Lee, and V. Ashok. Understanding Low Vision Graphical Perception of Bar Charts. In D. R. Flatla, F. Hwang, T. J. V. Guerreiro, and R. Brewer eds., *The 26th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS 2024, St. John's, NL, Canada, October 27-30, 2024*, pp. 59:1–59:10, St. John's NL Canada, 2024. ACM.
 - [29] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, and I. Sutskever. Robust speech recognition via large-scale weak supervision. In *International conference on machine learning*, pp. 28492–28518. PMLR, 2023.
 - [30] Y. Sechayk, A. Shamir, and T. Igarashi. ShowMe: 対話的な強調表示と拡大表示によるプレゼンテーションビデオの視覚的アクセシビリティの改善. In *Proceedings of the 32nd Workshop on Interactive Systems and Software (WISS 2024)*, 2024. Authors: Yotam Sechayk (University of Tokyo), Ariel Shamir (Reichman University), Takeo Igarashi (University of Tokyo).
 - [31] J. Seo, S. S. Kamath, A. Zeidieh, S. Venkatesh, and S. McCurry. MAIDR Meets AI: Exploring Multimodal LLM-Based Data Visualization Interpretation by and with Blind and Low-Vision Users. In *Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS '24*, pp. 1–31, New York, NY, USA, Oct. 2024. Association for Computing Machinery.
 - [32] S. S. Shapiro and M. B. Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611, 1965.
 - [33] S. F. A. Szpiro, S. Hashash, Y. Zhao, and S. Azenkot. How People with Low Vision Access Computing Devices: Understanding Challenges and Opportunities. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS '16*, pp. 171–180. Association for Computing Machinery.
 - [34] D. S. Tan, B. Meyers, and M. Czerwinski. WinCuts: Manipulating Arbitrary Window Regions for More Effective Use of Screen Space. In *CHI '04 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '04, pp. 1525–1528, New York, NY, USA, Apr. 2004. Association for Computing Machinery.
 - [35] M. Tang, R. Manduchi, S. Chung, and R. Prado. Screen Magnification for Readers with Low Vision: A Study on Usability and Performance. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS '23*, pp. 1–15. Association for Computing Machinery.
 - [36] R. Wang, L. Zeng, X. Zhang, S. Mondal, and Y. Zhao. Understanding How Low Vision People Read Using Eye Tracking. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–17, Hamburg Germany, 4 2023. ACM.
 - [37] Y. Wang, Y. Zhao, and Y.-S. Kim. How Do Low-Vision Individuals Experience Information Visualization? In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, Chi '24. Association for Computing Machinery.
 - [38] Wikipedia contributors. “Congenital blindness.”. Wikipedia, The Free Encyclopedia, Aug. 2025. Accessed August 30, 2025.
 - [39] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins. The aligned rank transform for nonparametric factorial analyses using only ANOVA procedures. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*, pp. 143–146. ACM, 2011.
 - [40] Y. Yang, M. Cordeil, J. Beyer, T. Dwyer, K. Marriott, and H. Pfister. Embodied Navigation in Immersive Abstract Data Visualization: Is Overview+detail or Zooming Better for 3D Scatterplots? *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1214–1224, Feb. 2021.
 - [41] H. Zheng, S. Wang, C. Thomas, and L. Huang. Advancing Chart Question Answering with Robust Chart Component Recognition. pp. 5741–5750, Feb. 2025.

KanaShark: SHARK²に基づく日本語ジェスチャ入力手法の検討

豊田 真吾* 日下部 完* 坂本 大介*

概要. スマートウォッチのような小さな画面での日本語入力の困難さを解決するため、単語単位のジェスチャ入力手法 KanaShark を提案する。KanaShark は、SHARK² や Swype 入力の手法を日本語入力に用いられる 12 キー配列に応用し、単語の子音のみを一筆書きでなぞることで入力する。経路形状から単語を確率的に推定するため、操作の曖昧さを許容しつつ高速な入力を可能とする。本研究では KanaShark の有効性を検証するため、9 名の参加者による 10 日間の評価実験で、フリック入力手法、および子音タップ入力手法と比較した。その結果、KanaShark は学習効果が顕著であり、フリック入力手法より有意に高速かつ少ない誤入力での入力が可能であることが示された。一方で、初期学習コストの高さが課題として示された。これらの結果に基づき、本研究の最後では、スマートウォッチの日本語入力の効率性の観点から KanaShark の将来展望を議論する。

1 はじめに

スマートウォッチは、健康管理や通知の確認など、人々の日常生活において広く活用されている。しかし、スマートウォッチという限られた矩形上でのテキスト入力は操作が煩雑になりやすい。その結果入力ミスが発生しやすく、ユーザビリティの低下を招いている [3, 10, 11, 14]。そのため、代わりに音声入力や定型文の選択による文字入力がよく使用されているが、これらは利用環境やプライバシーの観点から適用範囲が限定されるため [6, 8]、ディスプレイ上のタッチ操作を前提とした文字入力手法の発展が依然として重要である。

ラテン文字圏でよく使われている QWERTY 配列は、アルファベットだけでも 26 個のキーが必要であり、スマートフォンのような小さな画面上にこれらを配置すると、各キーのサイズが小さくなりやすく操作性の課題が生じやすい。一方、日本語圏で広く利用されている 12 キー配列は 12 個のキーの配置からなるレイアウトであり、ラテン文字圏の一般的なキー配置よりも各キーの面積を大きくできる利点がある。2017 年の長澤らの調査では、大学生の 90% が 12 キー配列を基にしたフリック入力やトグル入力をスマートフォンで利用していることが報告された [16]。しかし、スマートウォッチのようなディスプレイが極めて小さいデバイスでは、それでもタップ操作の煩雑さが依然として課題となる。

小さなキーへの入力の問題を解決するため、ラテン文字圏では SHARK² [13] や Swype [12] に代表される QWERTY 配列上で一筆書きのジェスチャから単語を確率的に推定する手法が、スマートフォンやスマートウォッチの効率的な文字入力手法として活用

され、タップによる文字入力が抱える問題を克服してきた。一方で日本語入力においても一筆書きのジェスチャ入力手法の研究が存在する。BubbleGlide [21] は、特定の単語の子音情報（あ、か、さ、た、な、は、ま、や、ら、わ）のみを選択することで文字を入力するジェスチャ入力手法である。しかし、ラテン文字圏で行われたような一筆書きの経路全体の形状から単語全体を確率的に推定するアプローチは十分に検討されていない。

本研究では、日本語 12 キー配列において子音情報のみを一筆書きで入力し、経路全体に基づく確率推定で単語候補を提示する手法 **KanaShark** を提案する。本手法はスマートウォッチ上での高速な日本語入力を実現することを目的とし、評価実験を通じて提案手法の有効性を検証することで、入力体験の向上に寄与することを目指す。

2 関連研究

2.1 ラテン文字圏におけるジェスチャ入力技術の発展

タッチディスプレイにおけるジェスチャ入力手法は、1990 年代の文字単位の認識から始まり、当時の Personal Digital Assistant (PDA) 向けの入力手法として Unistrokes [9] や Graffiti [17] が提案され、一筆書き可能なように簡略化されたアルファベットを手書きで入力する手法について研究された。その後 SHARK² に代表される単語単位のジェスチャ入力が登場し、ユーザの曖昧な単語入力を確率的に解釈するパラダイムが確立され、Swype などの製品として広く普及した。近年では、スマートウォッチのような小型デバイスに最適化されたジェスチャ入力手法として、WatchWriter [11] や C-QWERTY [7]、SwipeRing [18] などが提案されている。本研究は、日本語 12 キー配列においてこのような単語単位の

Copyright is held by the author(s).

* 北海道大学

ジェスチャ入力を提案するものである。

2.2 スマートウォッチの効率的な日本語入力

スマートウォッチでの日本語文字入力においては、12 キー配列をはじめとするような子音のみが配置されたキーボードが主流であり、その上でフリックやジェスチャ操作などによる工夫が検討されてきた。PonDeFlick[2] は、画面端に配置された子音を選択し、その後画面中央でフリック操作を行う方式である。これにより、キーボードの画面占有を画面端に限定しつつ、画面中央の広い領域をフリック領域として有効活用できるようにした。SliT[19] は、画面外からのスライドインジェスチャを用いて文字を入力することで、キーボードが画面を大きく占有しない入力手法である。BubbleGlide[21] や BubbleOne[20] は、入力したい単語の子音のみに注目する入力手法として検討され、BubbleGlide と比較して文字入力速度が約 1.1 倍向上した。

BubbleGlide では、円周上に配置された子音を一筆書きすることで、1 文字ずつ逐次的に入力する手法が提案され、本研究とは子音を一筆書きする点が共通している。しかし、逐次的な認識手法の場合、一般的なグリッド状のキー配列は離れたキー間を結ぶ経路の途中で別のキーを通過し、誤認識を招くため利用できず、これを回避するためにキーを円周上に配置する必要があった。我々は、BubbleGlide のような子音一筆書きによる単語入力のアプローチを基に、さらに一筆書きの経路全体から単語を確率的に推定するアルゴリズムを導入することで、曖昧な入力にも対応可能な 12 キー配列の日本語文字入力手法を実現できるのではないかと考えた。本研究では、日本語 12 キー配列上で単語の子音のみをなぞった経路から、SHARK² に基づく確率的推定により入力単語を決定するシステムを検討する。

3 KanaShark の設計

本章では、KanaShark のジェスチャ入力インタフェースの設計と経路から単語を確率的に推定するアルゴリズムを説明する。なお、KanaShark のソースコードは以下の脚注の URL からアクセスできる。¹

3.1 インタフェース設計

KanaShark のインタフェースは、図 1 に示すようにジェスチャ入力画面と候補語選択画面の二つの画面から構成される。ジェスチャ入力画面では、12 キー配列上に「あ、か、さ、た、な、は、ま、や、ら、わ」の子音情報が表示され、ユーザは入力したい単語の子音情報を直線で結ぶように一筆でなぞる。この際、入力対応表 (表 1) に示すように、濁音・半濁音・小書き仮名への変換や連続する同一子音は省



図 1. KanaShark における「ありがとう」の入力例。

略する。ユーザが指を離すと候補語選択画面へ遷移し、左上を起点とし、左から右、上から下へ尤度順に候補語を表示する。ユーザが候補語をタップして選択すると入力語が確定され、再びジェスチャ入力画面に戻る。

表 1. KanaShark の入力対応表。

単語	入力する子音情報
りんご	らわか
みかん	まかわ
コンピュータ	かわはやわた
インタラクション	あわたらかさやわ
かたたたき	かたか

3.2 辞書の作成

KanaShark の候補語推定に用いる辞書は、日本語 Web コーパス 2010[1] を基に構築した。変換候補として自然な語を収録するため、自立語を先頭に持つ語または付属語のみからなる語を抽出した。例えば、「私は」のような語は日本語の文脈として自然であり変換候補として適切であるが、「は学校」のように付属語が先頭に現れる語は今回は変換辞書から除外した。そのうえで、頻度の高い語から順に 1-gram を 10 万件、2-gram を 10 万件、3-gram および 4-gram を各 2 万件、5-gram を 1 万件収集した (n-gram は形態素単位)。これにより、合計 25 万語を変換辞書として採用した。各語彙については、MeCab (Python ラップとして fugashi[15] を利用) を用いて読み仮名を自動生成し、コーパスにおける語の出現回数を頻度として統合し表 2 に示すような辞書データを作成した。

3.3 候補語推定アルゴリズム

KanaShark では、SHARK² の文字推定アルゴリズムを参考に前処理と入力時推定の 2 フェーズによって文字を推定する。前処理では図 2 に示すように、単語辞書の各語に対して、キーの中央を直線で

¹ <https://github.com/toyoshin5/KanaShark>

表 2. KanaShark で用いる単語辞書データの例.

語	読み仮名	頻度
北海道	ほっかいどう	9395168
北海道札幌	ほっかいどうさっぽろ	287662
北海道大学	ほっかいどうだいがく	109236
北海道で	ほっかいどうで	432722
北海道に	ほっかいどうに	409883
北海道の	ほっかいどうの	1722061

結ぶ理想的なジェスチャ経路を生成し、線形補間を用いて経路を等間隔の 20 点からなる座標データに変換する. その後, アスペクト比を維持したまま水平方向の値域が 0 から 1 の範囲に収まるように正規化する.

入力時には図 3 に示すように, ユーザの入力経路データを同様に整形し, 正規化された座標系列に変換する. 次に, 位置スコア, 形状スコア, 頻度の三つの観点 [13] から各単語との類似度を計算し, これらのスコアを統合してベイズ推定に基づく総合的な尤度を算出する. 最終的に, 尤度が一定のしきい値 (0.1%) を超えた単語のみを候補語として画面に表示する. 位置スコア, 形状スコア, 頻度を統合する際の重みや候補語を表示するしきい値は, 予備調査で探索的に決定した. 処理時間については, 25 万語の辞書に対して, Apple Watch Series 10 では前処理は約 5 秒, 入力時推定は約 0.1 秒で完了する.

4 評価実験

KanaShark, フリック入力手法, 子音タップ入力手法の 3 種類の日本語入力手法を比較するために, 計 10 日間にわたる継続的な評価を行った. 本実験は参加者内比較計画で実施し, 各参加者は 1 日 1 回のセッション内で 3 種類すべての入力手法を試した. 手法の提示順序に起因する影響を排除するため, 参加者 3 グループに対しラテン方格法に基づいて手法の提示順序を変更した. ただし, 日ごとの学習効果を正しく評価するため, 各参加者毎に割り当てられた順序を 10 日間を通して固定した.

4.1 実験条件

本実験では, 3 節で説明した KanaShark に加え, フリック入力手法および子音タップ入力手法を含めて 3 種類の日本語入力手法を比較した. 具体的なインタフェースは図 4 に示すとおりで, いずれの手法も watchOS の標準キーボードと同じキーの大きさ (7.47 mm × 3.85 mm) と配置を採用した.

フリック入力手法は, 日本国内で広く利用されているスマートフォン向け日本語入力手法である. 本研究では, watchOS の標準キーボードをそのまま用いるとキー入力のイベント自体を取得できないこ

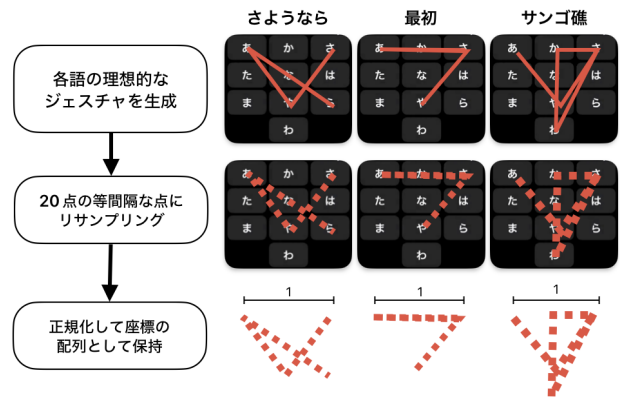
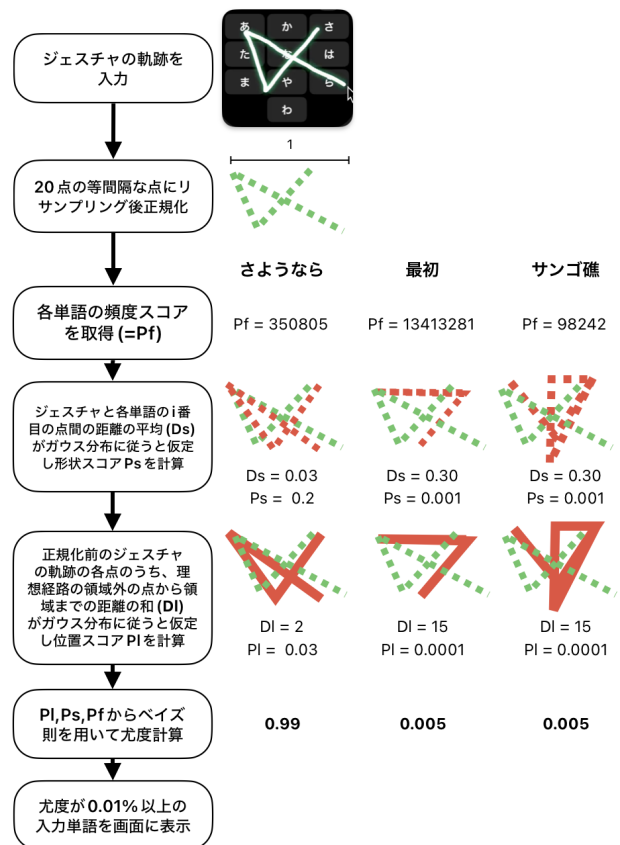


図 2. 候補語推定の前処理.



本図は処理の理解を助けるための模式図であり, 数値は仮の値である.

図 3. 候補語推定の入力時処理.

とや, 実験に影響を及ぼす可能性のある不要なインタフェース要素が含まれることから, 図 4 に示すように, 標準キーボードのレイアウトを再現した独自実装を用いた. 実験タスクに含めない句読点や英数字・記号入力は実装せず, 入力語の変換は「変換」キー押下によって図 1 の候補語選択画面と同形の画面が表示される仕様とした.

子音タップ入力手法は, フリック入力手法に対する



図 4. watchOS の標準キーボード（左）と本実験で比較する入力インタフェース（右）。

対照手法として設計した。入力したい単語の子音をタップし、決定ボタンを押下することで、図1の候補語選択画面と同様の画面が表示される。入力がジェスチャではない点を除き、その他の仕様は KanaShark と同一である。

4.2 評価項目

性能評価指標は文字入力速度 (Characters Per Minute, CPM) と誤入力率 (Errors Per Char, EPC) とした。CPM は、手本となる短文のかな漢字交じり表記の文字数を入力に要した時間で割った値であり、EPC は手本となる短文のかな漢字交じり表記の文字数に対する削除キー押下回数の割合である。また、主観評価として、10 日間の実験終了後には NASA Raw Task Load Index (NASA-RTLX) [5] と System Usability Scale (SUS) [4] に基づくアンケートを実施し、手法の好みとその理由についての記述式アンケートを行った。公平性を保つため、実験全体を通して KanaShark の手法名は「子音ジェスチャ入力手法」として提示した。

4.3 実験環境

実験には、文字入力用のスマートウォッチとして Apple Watch Series 10 46mm (watchOS 11.5) を用いた。実験用アプリケーションは SwiftUI を用いて開発し、Apple Watch 上で動作するように設計した。また、入力指示の提示のために iPhone 15 Pro Max (iOS 18.5) を机上のスマートフォンスタンドに設置して、実験参加者に iPhone 上の専用アプリケーションを通じて手本となる短文を提示した。

4.4 実験参加者

実験には 9 名（男性 7 名、女性 2 名、平均年齢 22.67 歳、標準偏差 0.67）の大学生・大学院生が参加した。利き手の分布は右利き 6 名、左利き 3 名であり、いずれの参加者も利き手ではない側の手首に Apple Watch を装着して実験を行った。フリック入力経験年数の分布は、10 年以上が 2 名、7～9 年間

が 5 名、3～4 年間で 1 名、経験なしが 1 名であった。

4.5 実験手順

実験参加者は、椅子に座り、机上に設置された入力指示用のスマートフォンの表示を確認しながら文字入力用のスマートウォッチを操作した。その際に、入力性能への影響を避けるため、参加者には入力中は肘や腕を含め身体を机に接触させないように指示した。各日のセッションは「練習フェーズ」と「本番フェーズ」で構成した。練習フェーズでは、各手法ごとに本番タスクの前に最大 3 分間の練習時間を設け、練習用の短文は本番タスクで用いるものとは異なるものとした。本番フェーズでは、iPhone の画面に提示される手本の短文を Apple Watch を用いて入力するよう指示し、各手法につき 10 個の短文（平均 8.6 文字）を入力させた。短文セットは、個人差や短文の内容への依存を排除するために全参加者・全日程で共通とし、「あ」から「ん」までのかなが 1 回以上出現するひらがなパングラムを満たしている。また、短文の提示順序は手法・日ごとにランダムとした。入力中に誤りがあった場合は、iPhone 上で該当文字が赤く表示され、必ず削除キーを使用して修正するよう指示した。10 日間の全セッション終了後には、主観評価のためのアンケートを実施した。

5 結果

5.1 文字入力性能

各手法の 10 日間の CPM の推移を図 5 に示す。KanaShark は 1 日目から 10 日目の間に入力速度が約 1.9 倍向上し、1 日目から 4 日目までは他の手法に比べて入力速度が低かったものの、5 日目以降は他の手法を上回る速度を達成した。本研究では、学習効果が安定し各手法の習得後の性能が発揮される区間として、実験期間の後半 1/3 に該当する最終 3 日間を評価区間とし、統計量を表 3 に示す。この区間に対して Friedman 検定を実施した結果、図 6 の (a) に示すように 3 つの入力手法間に統計

的に有意な差が認められた ($\chi^2_{2,N=9} = 13.00, p < .01$)。Holm 法で補正した Wilcoxon の符号順位検定による多重比較を行ったところ、提案手法である KanaShark はフリック入力手法と比較して有意に高速であり ($Z = -2.67, p < .05, r = .89$)、子音タップ入力手法もフリック入力手法より有意に高速であった ($Z = -2.67, p < .05, r = .89$)。一方で、KanaShark と子音タップ入力手法の間に有意な差は認められなかった。

EPC については、実験期間全体を通して大きな変化は見られなかった。同様に最終 3 日間に対して Friedman 検定を実施した結果、図 6 の (b) に示すように 3 つの入力手法間に統計的に有意な差が認められた ($\chi^2_{2,N=9} = 15.94, p < .01$)。Holm 法による補正を行った Wilcoxon の符号順位検定による多重比較を行ったところ、すべての手法間に有意差が認められ、KanaShark はフリック入力手法 ($Z = -2.67, p < .05, r = .89$)、子音タップ入力手法 ($Z = -2.49, p < .05, r = .83$) と比較して有意に低かった。

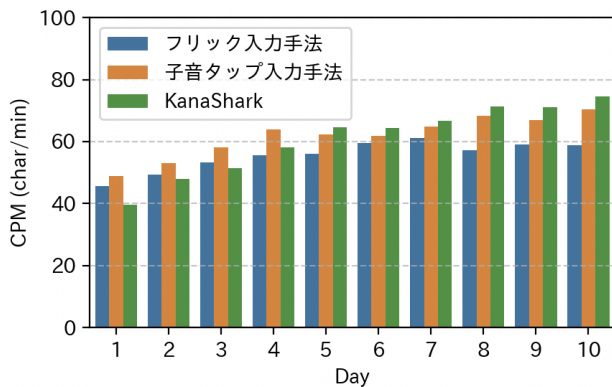


図 5. 各入力手法における CPM の推移。

5.2 メンタルワークロードとユーザビリティ

10 日間の全セッション終了後に実施した NASA-RTLX の Overall スコアを図 7 (a) と表 4 上段に示す。子音タップ入力手法はフリック入力手法と比較して有意に低いスコアを示し ($p < .05$)、メンタルワークロードが低いことを示した。しかし、KanaShark は標準偏差が大きいことから、中央値が最も低いにもかかわらず他の手法に対して有意差は認められなかった。また、NASA-RTLX の 6 つの個別評価項目ごとにも Friedman 検定を実施したが、いずれの項目も手法間で有意な差は認められなかった。

SUS のスコアを図 7 (b) と表 4 下段に示す。子音タップ入力手法と KanaShark は標準偏差が大きいことから、Friedman 検定では手法間で総合スコアに有意な差は認められなかった。しかし、10 の個別評価項目ごとには手法間で有意差が認められ、項

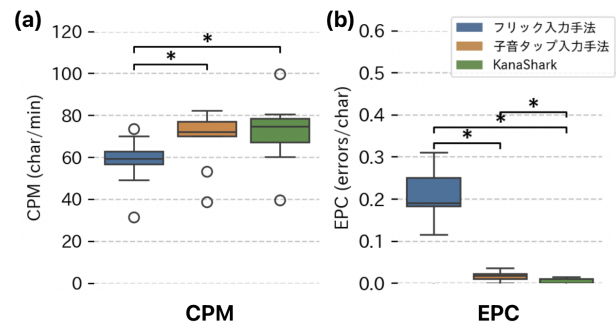


図 6. 最終 3 日間の (a) 平均 CPM. (b) 平均 EPC. 有意差は $*p < .05$.

表 3. 最終 3 日間の CPM と EPC に関する統計量。

CPM	Mean ± SD	Median
フリック入力手法	58.37 ± 12.34	59.53
子音タップ入力手法	68.54 ± 13.94	72.12
KanaShark	72.32 ± 16.29	74.86
EPC	Mean ± SD	Median
フリック入力手法	0.205 ± 0.062	0.191
子音タップ入力手法	0.018 ± 0.012	0.019
KanaShark	0.005 ± 0.006	0.000

目 4「この操作手法を利用するのに専門家のサポートが必要だと感じる」と項目 10「この操作手法を利用し始める前に知っておくべきことが多くあると思う」において、KanaShark はフリック入力手法よりもスコアが有意に高く、学習の負担が大きいと評価された (それぞれ $p < .05$)。

5.3 手法の好み

実験参加者 9 名中、2 名がフリック入力手法、5 名が KanaShark、2 名が子音タップ入力手法を最も好んだ。フリック入力手法、子音タップ入力手法、KanaShark の好みの順位の加重平均は、それぞれ 2.33, 2.11, 1.56 (数値が低い方が良い) であった。

KanaShark を好んだ理由としては、「ジェスチャの経路が多少ずれてもシステムが意図を汲み取って変換してくれるため、操作に気を使わなくても正確に入力できる (P2, P4, P7, P9)」、「指を画面に置いたまま操作できるので楽だった (P3, P7)」、「慣れれば最も速く入力できそうだと感じた (P2, P7)」といった点が挙げられた。一方で、「操作中に指が引っかかる感覚があった (P8)」、「自分の指で次のキーが見えなくなる (P8)」といった課題も指摘された。

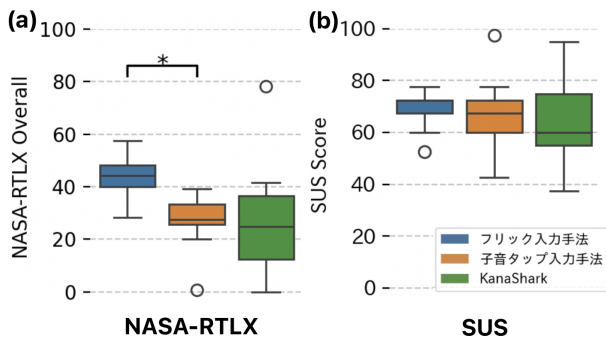


図 7. 実験終了後に実施した (a) NASA-RTLX の Overall. (b) SUS のスコア. 有意差は $*p < .05$.

表 4. NASA-RTLX と SUS スコアに関する統計量.

NASA-RTLX	Mean $\pm$ SD	Median
フリック入力手法	42.78 $\pm$ 9.67	44.17
子音タップ入力手法	26.67 $\pm$ 11.42	27.50
KanaShark	28.33 $\pm$ 23.03	25.00
SUS	Mean $\pm$ SD	Median
フリック入力手法	68.61 $\pm$ 7.92	72.50
子音タップ入力手法	66.94 $\pm$ 15.35	67.50
KanaShark	62.78 $\pm$ 18.13	60.00

6 考察と今後の課題

6.1 学習による入力速度向上

KanaShark は、学習に伴い入力速度が継続的に向上し、10 日間でフリック入力手法を上回る入力速度を達成した。CPM の推移では、実験初期はフリック入力手法に劣後したが、5 日目以降は他手法を上回り、最終 3 日間ではフリック入力手法に対する優位性が統計的に確認された。これは、同じ文字を入力するのに必要な動作が少ないためだと考えられる。KanaShark と子音タップ入力手法の間に有意差は認められなかったが、KanaShark の平均値は子音タップ入力手法以上であり、さらに学習曲線は 10 日間で飽和に達していないため (図 5)、さらなる習熟による向上が期待できる。

一方、EPC では、KanaShark が他の 2 手法より有意に低い値を示した。これは KanaShark が、スマートウォッチのような小さな画面での操作のずれを効果的に許容していることを裏付けるものである。これが結果的に修正操作の減少に繋がり、高速な入力を実現する上で重要な貢献を果たしていることが示唆される。

6.2 習熟過程における学習コストの課題

NASA-RTLX の Overall スコアによるメンタルワークロード評価では、KanaShark のスコアは参加者間でばらつきが大きく、他の手法と比較して統

計的な有意差は認められなかった。一方、SUS のスコアでは、KanaShark はフリック入力手法と比較して学習の負担が大きいと評価された。この学習負担の高さと評価のばらつきは、KanaShark の習熟過程に起因すると考えられる。

完全な習得のためには、子音の配置をある程度記憶する必要があり、参加者 (P8) が「自分の指で次のキーが見えなくなる」と指摘したように、初期段階ではオクルージョン問題が操作感を損なう要因となる。このような習熟上の課題が、一部の参加者にとっては特に高い認知的負荷として現れ、結果として参加者間での評価のばらつきを生んだことが示唆される。

6.3 今後の課題

評価タスクはスマートウォッチの利用文脈を踏まえ、辞書に存在する語による短文入力を中心に設計した。この前提は、日常的な返信が定型化されやすい実態を反映している一方で、辞書に存在しない固有名詞や専門語、文脈依存の語彙選択を含む状況に対する可能性は未検証である。この種の課題に対して Swype では、キーを一つずつタップして単語を入力する機能を併せ持つことで未知語の入力に対応している。これを参考に KanaShark においても、他の操作によって未知語入力を可能にする機能や入力された単語を学習する動的な辞書拡張機能を実装することが今後の課題となる。

また、システム側の実装要件として、語彙規模の拡大とリアルタイム推定の両立は非常に重要である。事前計算の再利用やデータ構造の改善によって、より大規模な辞書に対してもリアルタイムでの候補語推定を可能にすることが求められる。

さらに、環境に対するロバスト性の観点では、歩行中や公共交通機関内など身体への加速度が生じる状況における性能は未評価である。今後はそのような実利用環境での評価を進める。

7 おわりに

本研究では、スマートウォッチの 12 キー配列で、子音だけをなぞった経路から確率的に単語候補を推定し、文字を選択する入力手法 KanaShark を提案し、10 日間の継続評価でその有効性を確認した。実験の結果、10 日間の学習後はフリック入力手法より速く少ない誤りでの入力を実現した。一方で、初期の学習時の負荷についての課題も明らかとなった。本研究は、スマートウォッチにおける日本語入力に新たな選択肢を与えるものである。今後は、実利用環境における有効性の検証や、未知語の入力方法の検討を進めることで、より実用的なシステムの実現を目指す。

謝辞

本研究は、JST 創発的研究支援事業、JPMJFR 226S の支援を受けたものである。

参考文献

- [1] 日本語ウェブコーパス 2010. <http://www.s-yata.jp/corpus/nwc2010/ngrams/>, 2010. Accessed: 2025-08-02.
- [2] K. Akamine, R. Tsuchida, T. Kato, and A. Tamura. PonDeFlick: A Japanese Text Entry on Smartwatch Commonalizing Flick Operation with Smartphone Interface. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI '24, New York, NY, USA, 2024. Association for Computing Machinery.
- [3] P. Baudisch and G. Chu. Back-of-device interaction allows creating very small touch devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, p. 1923–1932, New York, NY, USA, 2009. Association for Computing Machinery.
- [4] J. Brooke. *SUS – a quick and dirty usability scale*, pp. 189–194. 01 1996.
- [5] J. C. Byers, A. C. Bittner, and S. G. Hill. Traditional and raw task load index (TLX) correlations: Are paired comparisons necessary? In A. 1989.
- [6] B. S. Chaparro, J. He, C. Turner, and K. Turner. Is Touch-Based Text Input Practical for a Smartwatch? In C. Stephanidis ed., *HCI International 2015 - Posters' Extended Abstracts*, pp. 3–8, Cham, 2015. Springer International Publishing.
- [7] G. Costagliola, M. De Rosa, R. D'Arco, S. De Gregorio, V. Fuccella, and D. Lupo. C-QWERTY: a text entry method for circular smartwatches (S). pp. 51–57, 07 2019.
- [8] C. Efthymiou and M. Halvey. Evaluating the Social Acceptability of Voice Based Smartwatch Search. In S. Ma, J.-R. Wen, Y. Liu, Z. Dou, M. Zhang, Y. Chang, and X. Zhao eds., *Information Retrieval Technology*, pp. 267–278, Cham, 2016. Springer International Publishing.
- [9] D. Goldberg and C. Richardson. Touch-typing with a stylus. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems*, CHI '93, p. 80–87, New York, NY, USA, 1993. Association for Computing Machinery.
- [10] J. Gong, Z. Xu, Q. Guo, T. Seyed, X. A. Chen, X. Bi, and X.-D. Yang. WrisText: One-handed Text Entry on Smartwatch using Wrist Gestures. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, p. 1–14, New York, NY, USA, 2018. Association for Computing Machinery.
- [11] M. Gordon, T. Ouyang, and S. Zhai. WatchWriter: Tap and Gesture Typing on a Smart-watch Miniature Keyboard with Statistical Decoding. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, p. 3817–3821, New York, NY, USA, 2016. Association for Computing Machinery.
- [12] S. Inc. Swype: Text Input for Screens. <https://web.archive.org/web/20080911004604/http://www.swypeinc.com/index.html>, 2008. Accessed: 2025-08-04.
- [13] P.-O. Kristensson and S. Zhai. SHARK2: a large vocabulary shorthand writing system for pen-based computers. In *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology*, UIST '04, p. 43–52, New York, NY, USA, 2004. Association for Computing Machinery.
- [14] X. Lin, Y. Chen, X.-W. Chang, X. Liu, and X. Wang. SHOW: Smart Handwriting on Watches. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 1(4), Jan. 2018.
- [15] P. O. McCann. fugashi: A Cython MeCab wrapper. <https://github.com/polm/fugashi>, 2019. Accessed: 2025-08-04.
- [16] N. Nagasawa. How Japanese university students type on smartphone and PC. 43:67–72, 12 2017.
- [17] I. Palm. Graffiti Arrives! – Press Release enclosed. <https://www.grot.com/zoomer/zoomer-list/zoomer-list-3Q1994/0048.html>, 1994. Accessed: 2025-08-24.
- [18] G. Rakhmetulla and A. S. Arif. SwipeRing: Gesture typing on smartwatches using a segmented QWERTY around the bezel. In *Proceedings of Graphics Interface 2021, GI 2021*, pp. 166–177. Canadian Information Processing Society, 2021.
- [19] 秋田 光平, 田中 敏光, 佐川 雄二. 画面占有率の低いスマートウォッチ向け文字入力手法 SiT. ヒューマンインタフェース学会論文誌, 21(1):131–140, 2019.
- [20] 戸羽 遼太郎, 加藤 恒夫, 田村 晃裕. BubbleOne: スマートウォッチ向け日本語曖昧入力インターフェースの試作と評価. 情報処理学会 インタラクシオン 2021, 2021.
- [21] 戸羽 遼太郎, 加藤 恒夫, 山本 誠一. BubbleGlide: N-gram を用いたスマートウォッチ向け日本語曖昧入力ジェスチャーキーボード. 情報処理学会 インタラクシオン 2020, 2020.

未来ビジョン

スマートウォッチの小さな画面は、テキスト入力の大きな障壁とされてきた。画面が小さいためにキーは必然的に小型化し、密集して配置される。この「キーが小さく密集した」環境は、正確な位置をタップする必要がある従来の手法にとっては大きな弱点となり、誤操作を誘発する原因となっていた。

しかし本研究を通して感じたのは、一筆書きでなぞるジェスチャ入力はこのような環境と相性が良いということである。「キーが密集している」ことは、すなわちキー間の距離が短いことを意味し、ジェスチャ入力にとってこの短い移動距離は指の操作量を減らし、素早い入力を可能にする。また、「キーが小さい」ことに対して、ジェスチャ入力は影響を受けにくいと感じた。経路全体の形状で単語を認識するため、

個々のキーを正確に通過する必要がないからである。実際に、キー1つの面積が小さいラテン文字圏のスマートフォンのQWERTY配列で、Swypeなどのジェスチャ入力が広く普及したことも、このような利点を反映していると考えられる。

今後は、本文で述べた今後の課題の対応が進めば、スマートウォッチはより実用的なテキスト入力環境を獲得すると考えている。さらに、大規模自然言語モデルなどの組み合わせにより文脈とジェスチャの形状から最適な語を推定できるような未来も考えられる。

このように、スマートウォッチの小さな画面への入力における障壁に対して、確率的に文字を推定するジェスチャ入力は相性が良く、日本語圏においても新たなインタラクションデザインとして受け入れられるのではないかと感じている。

溝間隔の違いによる筆圧変化を活用したシート埋め込み型 ID 認識手法

瀬崎 夕陽* 関口 祐豊* 中村 聡史*

概要. 本研究では、デジタルペンを用いた手書きを補助するものとして、筆圧変化を利用して筆記対象物を識別可能とする手法を実現し、それを用いた支援システムを実装した。具体的には、一定間隔で溝を設けた透明な筆記対象物をディスプレイ上に配置し、筆記時の筆圧変化をもとに筆記対象物の識別を行う、既存のディスプレイとデジタルペンのみで実現可能な新しい筆記対象物識別手法を用いる。評価実験では、溝の幅が異なる 4 種類のシートを作成し、筆圧のピークに基づき判定する手法の精度検証を行った。実験の結果、識別精度 0.78 で推定可能であることがわかった。さらに、本手法の有用性を検証するため、物理カラーパレットや学習支援教材などのアプリケーション例を実装し、実用性を検証することで、分野にとらわれない様々な場所での応用可能性を示した。

1 はじめに

教育のデジタル化によりタブレットが広く配布され、デジタルペンを用いた手書きが急速に普及している。ここでデジタルペンを用いた手書きでは、手書き入力と同時に画面上のメニュー操作を行う必要があるなど、ユーザにとって分かりにくい場合が多い。このような手書き入力の操作性を向上させる方法として Tangible User Interface (TUI) があり、例えば DataTiles[12] では、RFID タグを埋め込んだ透明タイルの認識と、タイル上の溝を活用した入力により、手書きとコンピュータ上での作業を効率化している。また、物理的なブロックを動かして画面内の図形を操作する手法 [5] や、ドーナツ状のオブジェクトを回転させて拡大する手法 [18] などがあり、物理的な制約を利用しつつ操作できるため、専門知識をもたないユーザや子どもでも悩むことなく楽しく使えるものとなっている。

ここで TUI の多くは、物体の存在や種類を認識するために外部カメラや赤外線センサ、あるいは物体自体への電子回路の組込みを前提としている。例えば、Lumino[2] では赤外線光源とカメラで物体を検出しており、SLAP Widget[15] では専用テーブルを用いる必要がある。また、導電パターンや電子部品を組み込み、タッチディスプレイ上で物体識別を行う手法も提案されており、TUIC[16] では高周波のタッチ間隔により、DataTiles[12] では物体に RFID タグをディスプレイにリーダを組み込み、種類や位置の認識を行っている。しかしこのような制約は、設置環境や製造コストの面で、普及や導入の障壁となる。

Acoustic Barcodes[7] はこうした問題を一部解決できるものであり、物体をなぞる際に生じる振動音

を利用して入力を可能としている。しかし、対象物の認識は可能であるものの、それと絡めた操作は困難である。ここで、筆記対象物の表面に凹凸がある場合、ペン入力時に筆圧の微細な変化が生じる。例えば、ザラザラした粗い表面では筆圧が小刻みに変化し、ツルツルした滑らかな表面ではその変化は少ない。こうした表面の凹凸は、TUI としての触覚的フィードバックを生むと同時に、筆圧変化という識別のための信号源にもなり得る。

そこで本研究では、普及しているペン入力デバイスをそのまま活用し、追加デバイスを必要としないシート型 TUI の実現を目的とし、透明なシート上に彫られた溝間隔の違いによる筆圧変化を活用した、シート埋め込み型 ID 認識手法を実現する。具体的には、溝間隔の異なる浅い溝を施した透明アクリル板のシートを筆記面に敷き、ペン入力デバイスで通常の手書きを可能としながら、ペン入力デバイスの筆圧値のピーク数やピーク間隔を用いてシートの種類を推定する。

本手法は識別にとどまらず、その筆記行為自体を入力操作として活用できる点にも特徴がある。また、本手法の実用性と応用可能性を示すため、デザインやエンタテインメント、セキュリティなど幅広い利用シナリオを想定したユースケースを実装し、その具体的な活用例を提示する。これにより、「描く」という日常的行為を基盤に、人間の創造・学習・遊びを支え、Human-Computer Interaction (HCI) 分野への展開可能性を示す。

2 提案手法

2.1 概要

本研究では、タブレットディスプレイ上に敷いた規則的な溝を有するシート上で筆記する際に生じる筆圧変化を利用し、シートの識別を行う手法を提案する (図 1)。具体的には、異なる溝間隔幅をもつ

Copyright is held by the author(s).

* 明治大学

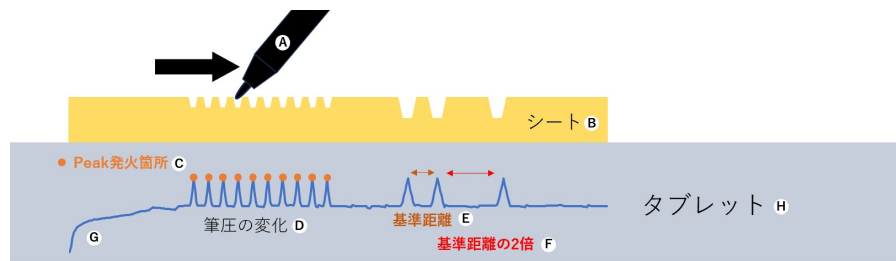


図 1. タブレット (H) 上にシート (B) が置かれており、ペン (A) が溝上を筆記すると、筆圧変化 (D) が生じる。筆記始めは筆圧が急激に上がる (G)。その後、ペンが溝を通過するたびに左右の隣接点よりも筆圧が高くなりピークとして検出される (C)。また、溝の間隔幅によってピークの発生間隔が異なる (E, F)。このように筆圧変化を利用することで、シートの識別を可能とする。

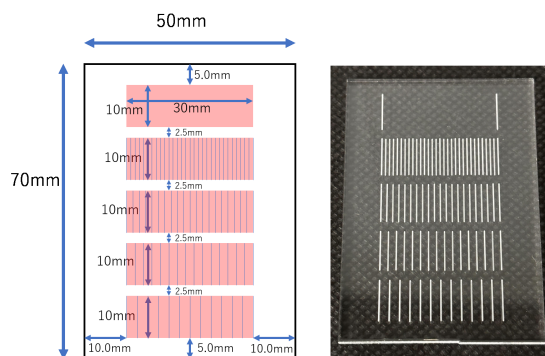


図 2. 異なる溝領域を持つ筆記対象物

アクリル板シート (図 1B) を用い、その上をペン先が通過する際に生じる筆圧の変化パターンに着目する (図 1D)。

ペン先が溝に接触することで生じる筆圧の増減は、溝の設置間隔や規則性に依って異なる特徴を示すことが予想される (図 1E, F)。例えば、1.0mm 間隔の溝領域では、1.0mm ごとに筆圧値にピーク (図 1C) が生じると考えられる。本手法は、この変化の回数や間隔を利用し溝領域を推定する。

本稿では、このアプローチの実現可能性を検証するため、事前に収集したストロークデータに対して特徴量抽出および識別手法の検証を行う。識別には、ランダムフォレストによる機械学習モデルによって、多様な筆記行動に対する識別性能の評価を行う。

2.2 シート作成

本研究ではシートに溝を施し、その上を筆記することにより生じる筆圧変化を用いてシートの識別を行う。シートに施す溝を特徴的なものにするによって、それに依った筆圧変化を起こさせることを意図している。今回はシートの素材には厚さ 3mm の透明アクリル板を用い、溝は光学レーザカッター (Trotec) により施す。レーザカッターの出力は 30% で速度は 2%、照射回数は 2 回とし、照射距

離はシートが 35mm 離れた位置で行った。また、溝の深さは 1mm 未満であり、筆記に支障のない深さとした。この方法であれば 2 次元平面であればある程度自由に溝を付与することが可能である。

シートに施す溝間隔に規則性を付与し、例えば、溝の間隔幅を 1.0mm で等間隔に設置すると、筆記時の筆圧変化も 1.0mm の間隔で変化する。また、溝間隔を 1.0, 2.0, 1.0, 2.0 と規則的に変化する間隔幅で溝を施せば、その間隔規則と同じような筆圧変化が起こり、異なる ID として識別可能になる。この特徴を生かし、溝を用いた識別を可能にする。

本稿では、上記条件を満たす 4 種類の溝領域 (図 2) を対象に筆記データを取得し、提案手法の識別精度を検証する。また、その結果を踏まえたシート型 TUI の設計指針および活用可能性について考察する。

2.3 特徴量抽出と前処理

シート推定には、筆圧値のピークの発生回数と発生周期を特徴量として利用する。溝の位置は、ペン先が溝を通過した際に発生する局所的な筆圧変化を利用することで検出する。具体的には、ストローク中の各点において、左右の隣接点よりも筆圧が高く、かつ周囲の点との筆圧差が所定の閾値以上である場合、その点をピークとして検出する。検出されたピークの総数を発生回数と定義し、隣接ピーク間の x 軸方向距離から得られる間隔幅を発生周期と定義する。さらに、発生周期の出現頻度分布を解析することで、溝間隔に依存する特徴を捉え、シートの識別に用いる。なお、ピーク検出、特徴量算出およびシート推定結果を返す API 実装は、GitHub¹に公開している。

¹ https://github.com/henkakyu/mizo_classifier

3 精度検証

3.1 データセット構築

シートを用いた筆圧データセットを構築するため、特徴量と正解ラベルを同時に収集可能なシステムを Processing により実装した。本システムでは、ペン入力により取得される座標情報、筆圧値、ストローク ID、および使用した溝領域のラベルを記録する。筆記距離を識別特徴量として利用するため、同一座標の連続データは不要と判断し、座標が変化した場合にのみ筆圧を記録する仕様とした。取得したデータは CSV 形式で自動保存し、特徴量抽出および分類モデルの学習、分類式の作成に利用する。

本稿で提案するシートは、先行研究 [17] を参照しつつ初回の試作として構造を単純化するため、縦方向にのみ溝を施した。ここで予備実験 [17] において、1.0mm 未満の溝間隔の場合に筆圧の変化がほとんど観測されなかったことから、溝間隔としては 1.0mm, 1.5mm, 2.0mm, 2.5mm の幅を選定した。なお、溝間隔を規則的に変化させるものも考えられるが、今回は単純化のため等間隔のものに限定した。

なお、縦方向のみ溝を施した制約から、本データセット構築では横方向に一定距離以上の筆記を含むストロークを対象とした。データ収集は先行研究に倣い、表示された横直線を筆記しながらストロークに伴う筆圧データを取得することで行った。

データ収集時には、タブレット上に配置したシートを非利き手で固定し、画面に表示された横直線をなぞるタスクを筆頭著者自身が実施した。ここでは各溝条件につき 100 回の筆記を行い、全 5 種類の溝領域（うち 1 種類は溝無し）で計 500 ストローク分のデータを取得した。筆記には、PC に接続した Wacom One 12 (DTC121) を拡張ディスプレイとして使い、Processing で実装したウィンドウサイズ 2000 × 1000px のシステム上で、付属のデジタルペンを使用した。

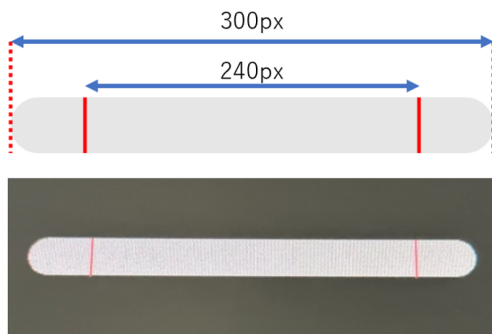


図 3. データセット構築の際の筆記ガイドライン

3.2 推定方法

シートのテクスチャによる筆圧変化を利用して筆記領域の溝の種類を推定するため、収集した筆圧データから特徴量を算出し、ランダムフォレストを用いた分類を行う。

ここで 1.0mm の溝が等間隔で並んでいる場合に、すべての溝でペンの筆圧の変化を適切に検知できるとすると、コンピュータの内部では、この 1.0mm の幅に該当するディスプレイのピクセル数ごとに筆圧が変化することになる。そのため、ピークとピークとの間のピクセル数がどの程度の頻度で出現するかを利用することで、シートの認識が可能となる。

本データセット構築において、溝間隔としては 1.0mm, 1.5mm, 2.0mm, 2.5mm を採用したため、この値を基準として各溝間隔に対応するピクセル範囲を設定した。選定した特徴量は、ストローク内の総ピーク総数と、4つの幅（1.0mm, 1.5mm, 2.0mm, 2.5mm）それぞれに相当するピーク数の計 5 つである。

それぞれのピーク間隔における閾値は、基準となる間隔幅 1.0mm をもとに等間隔となるように設定した。具体的には、1.0mm に対応するピクセル数を Δ と定義し、1.5mm, 2.0mm, 2.5mm に対応するピクセル数をそれぞれ 1.5Δ , 2.0Δ , 2.5Δ とした。今回は基準間隔幅 1.0mm の半分の距離、すなわち $\Delta/2$ の間隔で溝を配置した。そのため、本研究では各間隔幅に対して、その半分の距離 $\Delta/2$ を閾値の基準とした。したがって、任意の間隔幅 $n\Delta$ に対するピーク間隔の判定範囲は、

$$n\Delta - \frac{\Delta}{4} \leq interval_n < n\Delta + \frac{\Delta}{4}$$

として定義した。今回用いた Wacom One 12 における 1.0mm がディスプレイ上で 6.53px であったため、 $\Delta = 6.53$ でそれぞれの値を求めた。

ランダムフォレストによる推定では、各条件のデータをランダムに 8 割を学習用、2 割をテスト用に分割し、5 分割交差検証を実施した。いずれの推定においても、1 ストロークあたりのデータのうち、横直線の前後 30px を除外した $x = 280 \sim 520$ の範囲を対象とし、書き始めおよび書き終わりに伴う急激な筆圧変化の影響を排除した（図 3）。ランダムフォレストの実装には Google Colab 上の Python 3.10 および scikit-learn 1.4.0 を使用し、主要パラメータは `n_estimators=100`, `random_state=42` とした。交差検証には StratifiedKFold（分割数 5, `shuffle=True`）を用いた。

3.3 推定結果

各シートの溝条件に対する推定分類精度を表 1 に示す。また、全体の平均正解率は 0.78 であった。

表 1 より、全てのシートの溝条件において 0.65 以上の精度で分類できていることがわかった。特に、

表 1. 各溝の分類精度

		推定溝間隔 (mm)				
		None	1.0	1.5	2.0	2.5
正 誤 (mm)	None	0.93	0.00	0.00	0.00	0.06
	1.0	0.00	0.83	0.11	0.03	0.01
	1.5	0.00	0.11	0.65	0.19	0.03
	2.0	0.04	0.00	0.18	0.66	0.10
	2.5	0.06	0.00	0.00	0.09	0.82

溝間隔 1.0mm では 0.83, 2.5mm では 0.82 の分類精度が得られ, 他条件と比較して安定した識別が可能であった。

3.4 考察

シート全体の識別精度は約 0.78 であり, 本手法が一定の精度で識別可能であることが明らかになった。単回 (シート半往復分のストローク) での識別精度が 0.78 の場合, 往復により識別精度が高まるため, シート 2 往復以上で 90% を超え, 3 往復で 95% を超える。このことから, 筆記を複数回往復させることで実用的に十分高い精度を達成することも可能であると考えられる。

本手法の主要な利点は以下の通りである。

1. 追加のセンサや外部デバイスを必要とせず動作する点
2. 溝をなぞる筆記行為そのものを入力として利用できる点
3. 筆圧変化の座標に基づき, オブジェクトの位置推定が可能である点
4. 実用に耐える水準の識別精度を既存のペン入力環境で実現できる点

以上の利点から, 本手法は TUI 研究における汎用性の課題に対して一つの有効なアプローチとなり得る。この位置づけをさらに明確にするため, 4 章ではシート型 TUI を用いたアプリケーションの応用例を示し, 5 章では関連研究との比較を行うことで, 本研究の実用性および位置づけを行う。

4 アプリケーション例

本研究で提案した溝間隔に基づくシート推定手法は, シートに ID となる溝を彫り, その ID をペンでなぞることでソフトウェア的に読み取り, 活用することができる。そのため追加の電子デバイスを必要とせず, 筆記面の物理的パターンを入力として利用できる点に特徴がある。また, シート自体に色を塗ったり, コンテンツと対応づく絵を描くことにより, ソフトウェアによる操作を意識せずとも機能の切り替えや, コンテンツの対象を指定および操作することが可能となる。そのため, コンピュータの操

作に不慣れなユーザでも抵抗を覚えず操作できるようになると期待される。

4.1 物理カラーパレットインタフェース

筆記という自然な行為をそのまま入力として利用する応用例として, アクリル板の表面に複数種類の溝パターンを施し, 各領域を特定の色に対応させることで物理的なカラーパレットを実現した (図 4A)。この応用例では, ユーザはペン先で該当する溝領域をなぞることで, 溝間隔に基づいて領域を識別し, 筆記色を操作することができる。また, パレット領域をなぞった長さや回数はそのままシステムに入力として渡され, 「なぞった分だけ色が付く」「混ぜた分だけ新しい色が生まれる」といった実際の絵の具のような挙動を再現できる。例えば, 赤と青の領域を続けてなぞると紫が得られ, その際に赤の領域を長くなぞれば, 赤みの強い紫色が生成できる。これにより, 従来の GUI カラーピッカーでは難しかった混色や比率調整を, 実際の絵の具のように触覚的かつ直感的に行うことができる。

4.2 学習支援シートインタフェース

触覚的な行為を通じて, 物体から情報が得られることは認知的理解を高めるとされており, TUI 分野でも多くの研究がなされている [9, 7]。また, 教育現場でスタイラスペンを用いることは一般化されつつある [19]。そこで本応用例では, シート自体を特定の対象物を象った形状とし, ユーザが該当部分をペンでなぞることで, その部分に関する説明および図解を提示する (図 4B)。このように, 教材を触覚的に操作しながら情報を得ることで, 視覚的理解を補完しつつ触覚的体験を伴った教育的インタラクションを提供できる。

4.3 エンタテインメントシートインタフェース

本応用例では, 溝領域や溝そのものを入力として利用することで, 筆記操作を音や映像に変換するエンタテインメント性の高いインタラクションを実現した。図 4C では, 一つ一つの溝領域をピアノの鍵盤に対応させることで, ペンでなぞるたびに鍵盤のように音が鳴り, まるで演奏しているかのように楽しめる。本応用例では, シートにおいて物理的にプリントされた鍵盤上をなぞることで, 連続的な音の変化を楽しむことができる。なお, 鍵盤は本来叩いて音を鳴らすものであるため, よりなぞるに特化した音の鳴らし方にするには, グラスハープなどの楽器を採用することも可能である。また, 本手法では, 溝領域の識別だけでなく, 溝の位置を検知した際にユーザにフィードバックを返すことも可能である (図 4D)。この例では, 筆圧変化を検知した箇所から火花のアニメーションが打ち上がる。これらの仕組みにより, 子供は遊びながらペンを走らせる体験を得

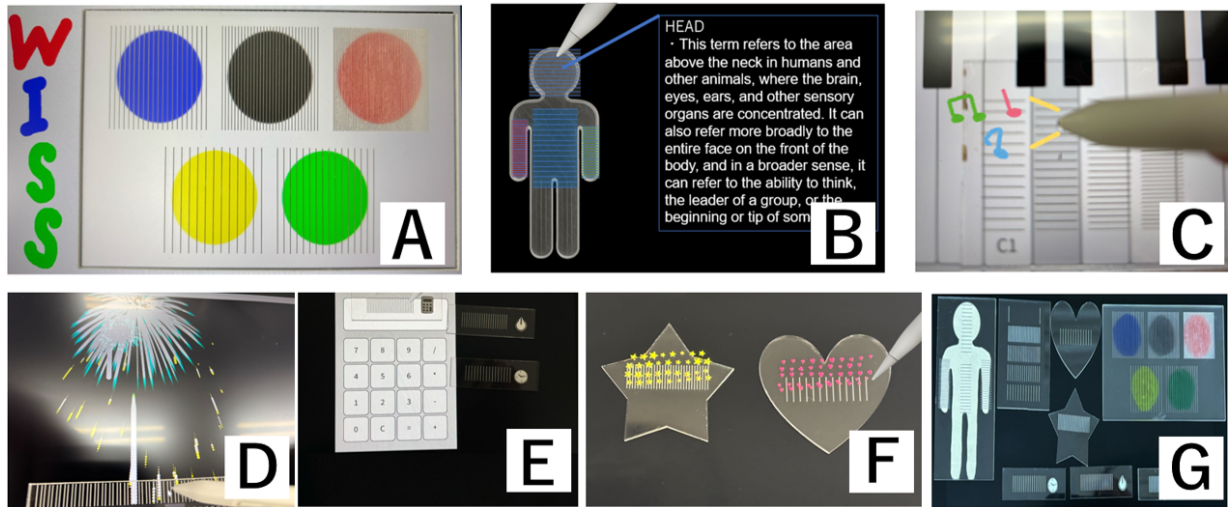


図 4. シート型 TUI を用いたアプリケーション例. (A：物理カラーパレット, B：学習支援, C, D：エンターテインメント, E：UI 格納, F：視覚支援, G：作成シート一覧)

られ、フィードバックを通じて、文字や線を書き始める動機づけとして楽しむことができる。

4.4 認証シートインタフェース

本応用例では、溝領域をタグのように扱い、暗証番号入力に応用するシステムを実装した。ユーザは特定の溝領域を決められた順序でなぞることでロックを解除できる。本応用例は筆圧変化そのものを単独の生体認証と見なすには限界があるものの、ユーザが意図的に設定できる領域のなぞり順序や回数といった行動的シーケンスを認証情報として用いることで、高い組合せ数を持つパスコードを構成でき、単独の生体認証ではなく行動的パスコードや多要素認証の一要素としての応用が期待される。さらに、シート自体が鍵として機能するため、一見すると入力手段であることすら認識されにくいという匿名性を持つことも期待できる。本応用例は、単独の認証方式として利用できるだけでなく、従来のログイン方式（パスワードや PIN コード、生体認証など）と組み合わせることで、多要素認証の一部としてセキュリティを強化できる可能性を持つ。

4.5 UI 格納型シートインタフェース

本応用例では、溝領域を UI の呼び出しタグとして利用し、特定の溝をなぞることで対応するウィンドウやメニュー画面を表示する仕組みを実装した（図 4E）。ユーザは必要に応じて溝領域を操作することで、任意の UI 要素を呼び出し、自由に再配置することができる。また、使用していないときにはシートをディスプレイ外に退避できるため、常に画面上に UI を配置する必要がなく、作業空間を圧迫しないという利点がある。

4.6 視覚支援シートインタフェース

本応用例では、シート自体をユニークな形状に加工し、触覚によって識別できるインタフェースとして利用した。このような形状の違いは、ユーザが視覚に頼らずにシートを区別する手がかりとなり、視覚にハンディキャップをもつユーザに対する操作支援になる。これにより、ディスプレイ上の GUI が提示できない情報を、物理的な形状を介して直感的に伝達できる。またこうしたシートは見た目でも機能が把握しやすいため、図 4F のように、星やハートのような形状のシートの上で手描きをすると、星やハートのブラシで手描きができるといった表現にも応用可能である。

5 関連研究

5.1 物理デバイスを用いた操作支援に関する研究

タブレット上に物理デバイスを配置して操作を支援する TUI 研究には、Bricks[5] や TUIIC[16] などがある。Bricks は、物理的なブロックを動かして画面内の図形を操作する手法であり、有線の磁気センサを用いて物理ブロックを識別している。TUIIC は、マルチタッチパターンによって人間の指との区別が可能な TUI を実現している。この手法では、オブジェクトの認識により他の操作を妨げず、かつ既存センサのみを用いて認識することを目指しているが、オブジェクト自体に電子回路を組み込んでいるという特徴がある。また、Ohmic-Touch[8] は、オブジェクト内部に抵抗素子やセンサを組み込むことでタッチ面との間の電流変化を利用し、識別や連続入力を可能にしている。音響認識を用いている手法もあり、TapSense[6] や AugmentingTouch[10] は指の異なる部位や物体からのタップを表面にマイクを付与し

たタッチスクリーンで識別している。他にも、実物体に付与した識別パターンをカメラで読み取ることによって識別する StructCode[3] や赤外光とカメラを使って認識する IR Rings[13] や Lumino[2] などがある。

これらの研究は、オブジェクトの識別精度が高いという利点がある一方で、オブジェクト側に電子部品や電子回路の組み込みや外部カメラやマイクを用意する必要があり、コストがかかるという課題がある。本手法は、シートに溝を彫るだけでよく、低コストで実現可能である。

5.2 溝の入ったシートを用いた TUI に関する研究

溝を用いた TUI 研究もいくつか行われている。Acoustic Barcodes[7] は、刻み目を持つ物体をなぞる際に生じる振動音をバイナリ ID に変換し、オブジェクトの識別を可能にしている。しかし、得られるのは ID 情報にとどまり、オブジェクトの位置を取得できないため配置が固定されてしまう点や、ユーザにとって溝をなぞる行為自体に強い意味が付与されにくいという課題がある。また、Rekimoto らの DataTiles[12] では、平面ディスプレイ上に RFID タグを埋め込んだ透明なタイルを配置し、電磁波によって識別する手法である。この方式はスタイラスペンでの筆記操作と組み合わせることができ、溝に沿った操作行為に意味を持たせられる点が特徴である。しかし、タイルごとに RFID タグを付与する必要があるほか、RFID センサコイルが配置された領域でしか利用できないという制約を伴う。

本手法は、Acoustic Barcodes の「溝を掘るだけで識別できる」という低コスト性と、DataTiles が示した「筆記操作を入力行為として利用できる」という特徴を備えている。本手法で DataTiles のタイルに溝を彫り、それをなぞることで DataTiles の ID 識別に活用することが可能となるため、DataTiles をより安価で実現可能になると期待される。

5.3 筆記操作を利用した TUI に関する研究

Elkin ら [4] や Song ら [14] は、ペンに複数の入力インタフェースを組み込み、回転や持ち替え、グリップや指のジェスチャといった操作を利用可能にしたが、いずれも入力が入力の物理構造に依存するという制約がある。McGee ら [11] は、紙地図と付箋をセンサーと結びつけることで、紙に描いて貼るという自然な作業をそのまま入力に転用する TUI システムを開発した。Asselborn ら [1] は、小型ロボットを用いて文字の形や書き順を触覚的に提示し、子どもの筆記学習を支援する手法を示している。

本研究も同様に、筆記操作という日常的かつ自然な行為に着目する。また、ペン側に機能を追加するのではなく、書かれるシート側に物理的な溝を設けることで入力空間を拡張し、新たな TUI の可能性を提示するものである。

6 議論

本手法の精度検証および応用例の実装から気づいた点、さらなる改良について議論する。

まず、識別精度は平均 0.78 程度であり、一定の有効性を示す一方で、さらなる改善の余地も残されている。これは、筆記方法やデータ取得方法を工夫することで精度向上を行うことが必要である。また、現状では interval 特徴量の識別力が十分に発揮されていない可能性がある。この点については、筆圧取得デバイスの分解能の影響が大きく、今後のデバイス性能向上により改善されることが考えられる。

次に、本研究ではレーザカッターを用いてアクリル板に溝を加工したが、現状ではアクリル素材のみを対象としている。しかし、Acoustic Barcodes の知見が示すように、浅い溝であれば素材を問わず付与可能であり、本手法はアクリルに限らず多様な素材へ展開できる発展性を持つと考えられる。また、素材ごとにペン先と表面との摩擦や硬さが異なるため、筆圧変化の特徴も変化すると考えられる。この点は単なる識別精度の改善にとどまらず、「素材特性に応じた手書き入力インタラクション」という新しい観点を提供する可能性があり、今後もさらなる調査を行う価値があると考えられる。

最後に、本手法が幅広い応用可能性があることを 4 章において示したが、筆記方法や利用条件に関する制約が今後の課題である。本研究では溝を縦方向に加工し、横方向の筆記に限定して識別を行ったため、縦書きや斜め方向の筆記利用は十分に検証していない。また、溝の形状や深さによって筆圧変化の現れ方が異なり、識別精度に影響を与える可能性がある。今後は、円形の溝や斜線型の溝、さらには深さを変化させた溝など、より多様なデザインを導入することで、筆圧変化に多彩な特徴を与えられる可能性がある。また、筆圧変化を利用する特性上、ユーザの筆記習慣や筆圧特性に大きな個人差が生じると考えられる。そのため、事前キャリブレーションによる補正や個人差を活用したパーソナライズ化した応用例について検討が必要である。

7 むすび

本論文では、溝間隔の違いによる筆圧変化を活用した、追加デバイスを必要としないシート埋め込み型 ID 認識手法を提案した。またシートの識別精度が実用可能性を持つことを明らかにし、実際にシート型 TUI を利用した複数のアプリケーション例を実装し、幅広い応用可能性を示した。今後の課題として、識別精度の向上がある。精度が向上することにより、応用の幅が広がると期待される。また、アクリル板以外の多様な素材への展開や溝構造の多様化を行う予定である。

参考文献

- [1] T. Asselborn, A. Guneyasu, K. Mrini, E. Yadollahi, A. Ozgur, W. Johal, and P. Dillenbourg. Bringing letters to life: handwriting with haptic-enabled tangible robots. In *Proceedings of the 17th ACM Conference on Interaction Design and Children*, IDC '18, pp. 219–230, 2018.
- [2] P. Baudisch, T. Becker, and F. Rudeck. Lumino: tangible blocks for tabletop computers based on glass fiber bundles. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1165–1174, 2010.
- [3] M. D. Dogan, V. H. Chan, R. Qi, G. Tang, T. Roumen, and S. Mueller. StructCode: Leveraging Fabrication Artifacts to Store Data in Laser-Cut Objects. In *Proceedings of the 8th ACM Symposium on Computational Fabrication*, SCF '23, 2023.
- [4] L. A. Elkin, J.-B. Beau, G. Casiez, and D. Vogel. Manipulation, Learning, and Recall with Tangible Pen-Like Input. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, pp. 1–12, 2020.
- [5] G. W. Fitzmaurice, H. Ishii, and W. A. S. Buxton. Bricks: laying the foundations for graspable user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '95, pp. 442–449, 1995.
- [6] C. Harrison, J. Schwarz, and S. E. Hudson. TapSense: enhancing finger interaction on touch surfaces. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, UIST '11, pp. 627–636, 2011.
- [7] C. Harrison, R. Xiao, and S. Hudson. Acoustic barcodes: passive, durable and inexpensive notched identification tags. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology*, UIST '12, pp. 563–568, 2012.
- [8] K. Ikematsu and I. Siio. Ohmic-Touch: Extending Touch Interaction by Indirect Touch through Resistive Objects. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pp. 1–8, 2018.
- [9] H. Ishii and B. Ullmer. Tangible bits: towards seamless interfaces between people, bits and atoms. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, CHI '97, pp. 234–241, 1997.
- [10] P. Lopes, R. Jota, and J. A. Jorge. Augmenting touch interaction through acoustic sensing. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, ITS '11, pp. 53–56, 2011.
- [11] D. R. McGee, P. R. Cohen, R. M. Wesson, and S. Horman. Comparing paper and tangible, multimodal tools. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '02, pp. 407–414, 2002.
- [12] J. Rekimoto, B. Ullmer, and H. Oba. DataTiles: a modular platform for mixed physical and graphical interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '01, pp. 269–276, 2001.
- [13] V. Roth, P. Schmidt, and B. Güldenring. The IR ring: authenticating users' touches on a multi-touch display. In *Proceedings of the 23rd Annual ACM Symposium on User Interface Software and Technology*, UIST '10, pp. 259–262, 2010.
- [14] H. Song, H. Benko, F. Guimbretiere, S. Izadi, X. Cao, and K. Hinckley. Grips and gestures on a multi-touch pen. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pp. 1323–1332, 2011.
- [15] M. Weiss, J. Wagner, Y. Jansen, R. Jennings, R. Khoshabeh, J. D. Hollan, and J. Borchers. SLAP widgets: bridging the gap between virtual and physical controls on tabletops. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pp. 481–490, 2009.
- [16] N.-H. Yu, L.-W. Chan, S. Y. Lau, S.-S. Tsai, I.-C. Hsiao, D.-J. Tsai, F.-I. Hsiao, L.-P. Cheng, M. Chen, P. Huang, and Y.-P. Hung. TUIC: enabling tangible interaction on capacitive multi-touch displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pp. 2995–3004, 2011.
- [17] 瀬崎夕陽, 関口祐豊, 中村聡史. 微細な筆圧変化を用いたシート判別手法の提案. ヒューマンコンピュータインタラクション (HCI), 2025-HCI-211(19):1–8, 2025.
- [18] 青木良輔, 宮下広夢, 井原雅行, 大野健彦, 千明裕, 小林稔, 鏡慎吾. くるみる: 複数導電部もつ枠型物理オブジェクトを用いたタブレット操作. 研究報告ヒューマンコンピュータインタラクション (HCI), 6:1–8, 2011.
- [19] 文部科学省. 令和 6 年度学校における教育の情報化の実態等に関する調査結果 (概要)【速報値】. https://www.mext.go.jp/content/20250829-mxt_shuukyo01-000044325_001.pdf, 2025. 最終アクセス日: 2025-10-17.

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

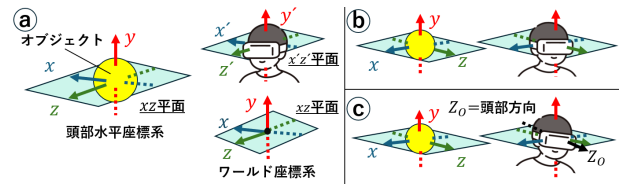


図 1. 頭部水平座標系. a: 初期状態. b: a の状態からユーザが左に向くと、 z 軸は左に向く. c: a の状態からユーザが左下を向いても、 z 軸は左に向く.

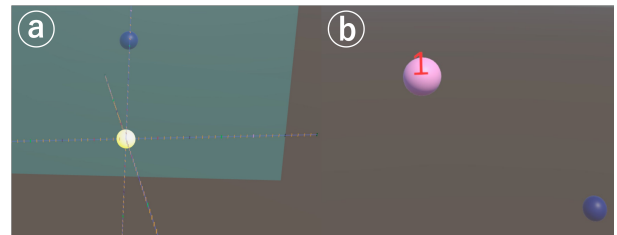


図 2. アプリケーションの画面. a: 頭部水平座標系および選択状態にあるオブジェクト（黄色の球）. b: 移動対象（桃色の球）および移動先（青色の球）.

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

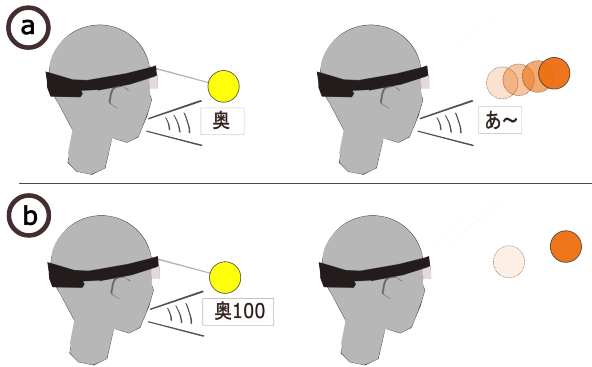


図 3. 各発声操作のイメージ。a：非言語的発声操作のイメージ。ユーザは移動方向を発声した後に非言語的な発声を行うことにより、発声を行っている間、オブジェクトをその方向へ連続的に移動できる。b：移動距離発声操作のイメージ。ユーザは移動方向と移動距離を発声することにより、オブジェクトをその方向へその距離だけ瞬時に移動できる。

「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/>

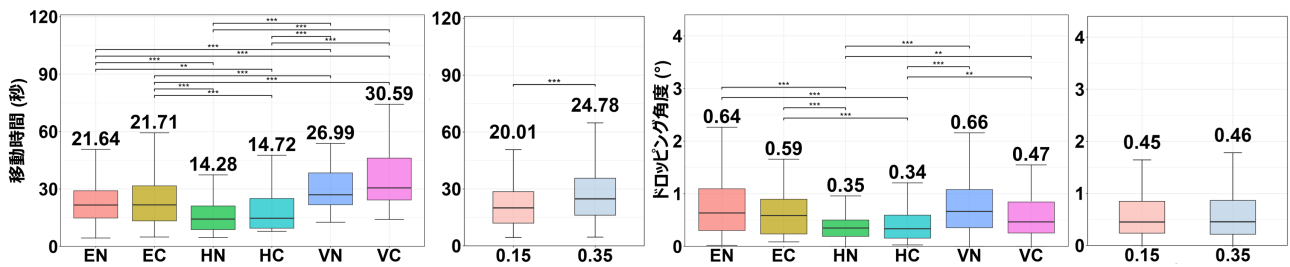


図 4. 実験結果. 左から移動時間 (手法条件別), 移動時間 (距離条件別), ドロッピング角度 (手法条件別), ドロッピング角度 (距離条件別). 有意差は * $p < .05$, ** $p < .01$, *** $p < .001$. ボックス上の数値は中央値.

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・

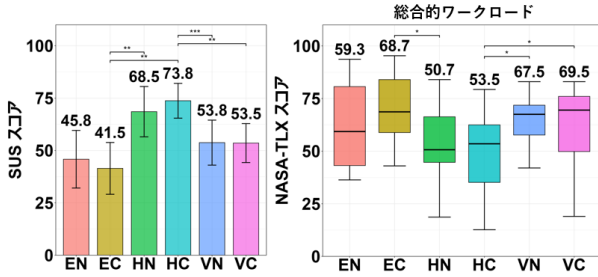


図 5. 実験結果. 左: 手法条件ごとの SUS スコア. エラーバーは 95%信頼区間. バー上の数値は平均値. 右: 手法条件ごとの NASA-TLX 総合的ワークロードスコア. ボックス上の数値は中央値. 有意差はそれぞれ * $p < .05$, ** $p < .01$, *** $p < .001$.

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, フリッ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.

著者索引

芦刈 治将, 59
阿部 優樹, 110
SHAMIR Ariel, 150

五十嵐 健夫, 15, 150
池谷 優希, 134
石井 行, 1
伊藤 貴之, 134
板東 恵理子, 59
岩井 望, 110
岩永 七海, 59

Hsueh Han Wu, 29
臼井 義人, 73

大城 亨太, 52
大津 陽翔, 44
大友 千宙, 22
大西 鮎美, 52
岡部 誠, 142
奥野 達也, 118
尾本 悦基, 67
折居 篤, 103

笠原 暢仁, 118
加藤 淳, 1

北野 和紀, 15

日下部 完, 126, 158

小池 英樹, 29, 36
與倉 陵太, 59
後藤 真孝, 95
小林 広夢, 126

崔 明根, 173
坂本 大介, 110, 126, 158

志築 文太郎, 103, 173
清水 春翔, 118

杉浦 裕太, 44
鈴木 悠真, 103, 173
鈴木 湧登, 126
鈴木 皇志, 173

関口 祐豊, 80, 166
瀬崎 夕陽, 166

高橋 洸汰, 142
高橋 治輝, 73, 88
竹内 大空, 29

Kelvin Cheng, 29

塚田 浩二, 67
塚本 昌彦, 52

寺田 努, 52

戸田 竜星, 8
豊田 真吾, 158

中川 由貴, 59
中澤 一輝, 95
中野 倫靖, 95
中村 聡史, 59, 166

佃 洸摂, 95
RAVE Hennes, 150

樋口 啓太, 80

福里 司, 1

Jana Hoffard, 29
本間 大一優, 118

COLLEY Mark, 150
牧野 倫太郎, 88
松村 耕平, 73, 88

宮崎 竜輔, 36
宮下 芳明, 8, 22, 118
宮藤 詩緒, 29, 36

山中 祥太, 118
山本 匠, 44

LI Yuan, 150

吉本 健義, 22
SECHAYK Yotam, 150

渡邊 果南, 59

日本ソフトウェア科学会 研究会資料シリーズ ISSN 1341-870X
No. 104 2025年12月3日発行

第33回 インタラクティブシステムと
ソフトウェアに関するワークショップ (WISS 2025) 論文集