

WISS
2021



第29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)

WISS2021 開催にあたって

寺田 努 *

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、2021年12月8日(水)～10日(金)にかけて現地とオンラインのハイブリッド開催される「第29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: WISS2021)での発表をまとめたものです。

昨年度のWISS2020は完全オンライン・無料開催というかなりイレギュラーな開催形態となりました。無料開催により、WISSという会議の面白さを多くの人に伝えられたのではないか、という満足感がある一方で、やはり現地でわいわいデモを体験したり、ディスカッションや雑談することの価値を再確認した開催でもありました。WISS2021の開催方法を検討するにあたって、未だ多くの会議がオンライン開催をしている状況ではありますが、我々はハイブリッド開催、デモ発表は現地発表のみ募集・デモはオンライン配信なし、というかなり現地を優遇した開催を行うことにしました。

現地参加者が極めて少なくなり大赤字の可能性も高いと予測していましたが、開催直前の状況ですと赤字にならなくて済みそうな状況になりました。予想以上の投稿・参加登録をいただき、感謝とともにWISS現地開催の力を感じているところです。

WISS2021の発表形態としては、登壇発表(Archival), ロングティザーフォーマット(Non-archival), デモ・ポスター発表(Non-archival)のカテゴリを用意し、登壇発表は35件の投稿に対してロング採択5件、ショート採択15件となりました。ロングティザーフォーマットは10件、デモ・ポスター発表は48件を採択しました。デモ・ポスターは現地必須としたにもかかわらず、多数の投稿をいただき投稿者の皆さんに感謝します。

開催日程を大幅に動かしたり、開催形態や論文募集形態を頻繁に見直すなど、WISSはフットワークの軽さが特徴ではありますが、そのために実行委員の皆さんには多大な負荷をかけています。今回も、現地で開催する以上、ワクチン接種の確認や現地での感染対策を含めて多くの作業が必要となりました。

今回の開催に関しては、実行委員長の真鍋さんや会計の瀬川さんを始めとした委員のみなさんに多大な負荷をかけました。ここに深く感謝したいと思います。私はWISS2020とWISS2021で委員長を務めさせて頂きましたが、とてもイレギュラーな2年間だったと思います。私が初めてWISSに参加したのは学生時代のWISS1999なのですが、以降、自分にとって特別なワークショップとなりました(宴会芸もたくさんやりました)。この思い入れのある会議で委員長を務め、なんとか次の委員長までWISSを引き継ぐことができそうだ、ということで、ホッとすると同時に引き続きWISSに貢献していきたいと考えています。

最後に登壇発表論文を投稿頂いた皆様、デモ・ポスター発表に申込み頂いた皆様、議論に参加し会議の成功に貢献して下さる参加者の皆様、スポンサー企業の皆様、協賛学会関係者の皆様、査読者の皆様、そして、WISS2021プログラム委員/運営委員の皆様に深く感謝いたします。

* 神戸大学、WISS2021プログラム委員長

WISS 2021 プログラム委員長
寺田 努 (神戸大)

WISS 2021 実行委員長
真鍋 宏幸 (芝浦工業大)

WISS 2021 副実行委員長
村尾 和哉 (立命館大)

WISS 2021 運営委員

会計	細部 博史 (法政大) 瀬川 典久 (京産大)
出版・ デザインコンペ	池松 香 (ヤフー) 加藤 邦拓 (東京工科大)
参加者管理	樋口 啓太 (Preferred Networks, 醸造酒類研究所)
照明・音響	川口 一画 (筑波大)
広報・ 学会リエゾン	横窪 安奈 (青学大)
WISS チャレンジ	入江 英嗣 (東大)
チャット	福地 健太郎 (明治大)
動画生中継	阪口 紗季 (東大)
デモ・ポスター	土田 修平 (神戸大) 塚田 浩二 (はこだて未来大)
ネットワーク	原 貴洋 (ヤマハ) 奥本 隼 (TwoGate)
Web・ サーバー	小山 裕己 (産総研) 福里 司 (東大)
投票	鳴海 紗也 (東大)
懇親会	倉本 到 (福知山公立大)
表彰	宮下 芳明 (明治大) 栗原 一貴 (津田塾大, クーリード)
企業 スポンサー	岩本 拓也 (サイバーエージェント) 開原 沙文 (チームラボ)
学生 ボランティア	櫻井 翔 (電通大) 中野 倫靖 (産総研)

表紙デザイン

越後 宏紀 (明治大)

WISS 2021 プログラム委員

池松 香 (ヤフー)
入江 英嗣 (東大)
岩本 拓也 (サイバーエージェント)
沖 真帆 (はこだて未来大)
奥本 隼 (TwoGate)
開原 沙文 (チームラボ)
梶本 裕之 (電通大)
加藤 邦拓 (東京工科大)
川口 一画 (筑波大)
倉本 到 (福知山公立大学)
栗原 一貴 (津田塾大, クーリード)
小池 英樹 (東工大)
小山 裕己 (産総研)
阪口 紗季 (東大)
坂本 大介 (北海道大)
櫻井 翔 (電通大)
志築 文太郎 (筑波大)
瀬川 典久 (京産大)
塚田 浩二 (はこだて未来大)
土田 修平 (神戸大)
寺田 努 (神戸大)
中野 倫靖 (産総研)
鳴海 紗也 (東大)
原 貴洋 (ヤマハ)
樋口 啓太 (Preferred Networks, 醸造酒類研究所)
福里 司 (東大)
福地 健太郎 (明治大)
細部 博史 (法政大)
真鍋 宏幸 (芝浦工大)
宮下 芳明 (明治大)
村尾 和哉 (立命館大)
築瀬 洋平 (Unity)
横窪 安奈 (青学大)

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: 寺田 努
神戸大学 工学部 電気電子工学科
兵庫県神戸市灘区六甲台町 1-1
E-mail: tsutomu (at) eedept.kobe-u.ac.jp

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

もくじ

まえがき: WISS2021 開催にあたって
寺田 努 i

セッション 1：現地発表 1

SingDistVis: 多数の歌声から歌い方の傾向を可視化できるインタフェース
伊藤 貴之, 中野 倫靖, 深山 覚, 濱崎 雅弘, 後藤 真孝 1

本研究は「ある程度条件が統一された膨大なデータを効率的に分析するための可視化方法の模索」を目的とし、その第一歩として、youtube やニコニコ動画に公開されている“歌ってみた動画”の歌い手の傾向を対象とする手法を提案している。アイディア自体はシンプルではあるものの、いずれの査読者も WISS で議論すべき論文として、採録を推薦しました。手法の発展性などについて活発な議論ができると期待している。

ダンスフォーメーション作成のための候補提示型インタフェース
遠藤 輝貴, 盧 承鐸, 中島 一崇, 楊 溪, 五十嵐 健夫 9

本研究では、グループダンスのフォーメーションのデザインを支援する「候補提示型」ユーザインタフェースを提案しています。提案するシステムはグループダンスでのフォーメーション作成で有益であると考えられ、WISSにおいてはルールを増やした場合の想定やコンピューターアニメーションエディットシステムとの関連、ユーザテストの評価方法などについて議論できればと思います。

歌唱時に高音の発声ができるようになる VR 映像の提案と評価
坂名 純太, 土田 修平, 寺田 努, 塚本 昌彦 16

本当に有用かどうか判断できない段階の研究ではありますが、斬新な着眼点に対しての評価は高く、本ワークショップにおいて議論に足るショート採録と判定されました。

セッション 2：オンライン発表 1

Human-in-the-Loop 型適応によるインタラクティブな音楽的拍節解析
山本 和彦 23

本論文では、楽曲に対して拍の情報を付与するタスク（拍節解析）を実施するための、深層学習による自動化技術とユーザーの対話的入力に基づく適応的制御機構を組み合わせた技術的枠組みを提案している。認識の個人差ゆえに定義が難しい問題を human-in-the-loop なアプローチによって解決する点は面白い。また、提案技術には別の問題への応用可能性がある。以上よりロング発表での採択となった。WISS では人間を組み込むことで問題解決を図るという思想についての議論も期待する。

色の修飾語を用いた対話的なイラストの色編集
住田 桃子, Nolwenn Maudet, 福里 司, 五十嵐 健夫 30

論文の質、及びシステムの完成度に基づき、ショート採録としてふさわしいと査読者全員の意見が一致しました。査読では、新規性がありまた有用性が部分的に確認されている点も評価されました。今後の研究のご発展を期待しています。

Knowledge ComposTer: 知識断片の再活用による継続的な創造活動支援システム 生田 泰章, 高島 健太郎	ショート採録	ショート発表
		35

査読者間および PC 委員会での議論の結果、当該システムには [13][14] からの拡張に対して進歩性があることが評価されました。したがって、評価が欠如している欠点はあるものの、その運用の適切性や意義について議論することには参加者にも一定の価値がある結論となりました。これに鑑み、判定はショート採録となりました。

セッション3：現地発表2

実世界と順序型コンテンツを連携した IoT プログラミング学習支援システム 川谷 知寛, 塚田 浩二, 栗原 一貴	ロング採録	ロング発表
		43

動画教材や PDF 教材に連動して、遠隔地にいる学習者の手元にあるマイコンが動くというシステム面での新規性は大きいです。また有用性の可能性も感じられます。ただし、その検証は行われていません。しかしながら、WISS の査読方針としてユーザ実験は必須ではありません。また、フィジカルなデバイスであるマイコンの遠隔教育というとても重要なチャレンジングな課題に取り組んでいます。

本研究のシステム面での新規性と、今後更に有用かつ実用的なシステムの構築を進めるための議論を WISS で深めるためにも、ロング採録が妥当であると判定されました。

部屋の3次元縮小モデルを用いた MR 空間ににおける仮想オブジェクト配置 井原 圭一, 川口 一画	ショート採録	ショート発表
		50

本論文は Worlds in Miniature (WiM) 手法を AR でのオブジェクト操作に応用したものであり、この点での新規性は認められる。一方で、査読では技術的に新しい点はないこと、評価方法についての懸念があること、また提案手法の具体的な内容についても明確ではないことなどが示された。ただ、評価実験で得られた内容に有益な内容も含まれていることの指摘もあり、ショート採録と判定する。

Color-Path: 手本動作の AR 表示によるヘアアレンジ再現支援システム 松井 菜摘, 大西 鮎美, 寺田 努, 塚本 昌彦	ショート採録	ショート発表
		57

AR を用いてヘアアイロンを用いたヘアアレンジを支援する点に新規性があり、ユーザ実験では手本通りの操作軌道の再現が達成され、一定の成果があった点が評価されました。ただし、髪型の見た目の再現については良い結果ではなく、有用性が限定的であることからショート採録と判定されました。

時間歪曲機能を用いた VR スキートレーニングの拡張 松本 高, Erwin Wu, 小池 英樹	ショート採録	ショート発表
		64

本研究では、スキーシュミレータと HMD を組み合わせた VR スキートレーニングシステムにおいて、時間経過速度を遅くする時間歪曲を用いた練習手法の提案を行っています。また 4 パターンの時間歪曲方法の比較実験を行い、その効果の検証を行っています。WISS においては、本研究の特徴的な提案である動的倍率変更手法の改善に関する議論や、トレーニングシステムとしての長期的な効果に関する議論が出来ればと思います。

セッション4：混在発表1

ドライブレコーダのデータから旅行の思い出動画を自動生成する web サービスの実装

ショート採録

ショート発表

尾頭 花奈, 石黒 祥生, 椎尾 一郎, 武田 一哉 71

システムの完成度、有用性、新規性ともに認められたが、インタラクティブシステムとしての言及や記述が必要と判断し、ショート採録とする。

TSUGITE: インタラクティブな木工ジョイントの設計・加工システム

ショート採録

五十嵐 健夫, 梅谷 信行, 吉田 博則, ラルスン マリア 78

木造建築において複数の部材を組み合わせる「継手」技法について、インタラクティブなシミュレーションとファブリケーションを可能にするシステムを提案／実装し、製作事例を紹介している。UIST2020にフルペーパーとして採録されている論文でもあり、全ての査読者が新規性／有用性ともに高く評価した。一方、投稿時の原稿は記述の質が低く、また既発表論文であり内容が既に認知されていることを勘案して、ショート採録と判定した。

空間ラベリング: 非専門家による画像注釈のラベル品質を改善するための空間レイアウトの活用

ショート採録

ショート発表

張 家銘, 李 佳憲, 五十嵐 健夫 85

CHI の採択論文であることもあり、記述も明確で研究としても重要なテーマであると評価されています。

しかしフルペーパーを 6 ページに圧縮された内容となっており、今回の投稿論文から読み取れる情報からはいくつか疑問点が指摘されていることから、ショート採録が妥当であると判定しました。

しかし、機械翻訳をしたような分かりづらい記述が含まれているとの指摘もあり、最低限読者に誤解を与えないよう記述の修正が必要です。

対戦アクションゲームにおけるプレイヤーの挙動観察のためのシーン検索

ショート採録（採録条件あり）

ショート発表

三ツ井 慧太郎, 岡部 誠 91

キャラクタの位置関係に着目した深層学習ベースの映像検索システムによって過去の事例を参照させることで対戦格闘ゲームのプレイスキル上達を支援する新規の手法を提案しており、有用性は荒削りであるものの、WISS コミュニティにおいて共有する価値のある研究であると考えられる。以上の結果から、ショート採録と判断された。

Corridor-Walker: 視覚障害者が障害物を回避し交差点を認識するためのスマートフォン型屋内歩行支援システム

ロング採録（採録条件あり）

ロング発表

栗林 雅希, 粟川 青汰, Jayakorn Vongkulbhaisal, 浅川 智恵子, 佐藤 大介, 高木 啓伸, 森島 繁生. 97

Lider による周辺形状把握と画像認識による交差点検出、音声と振動による情報提示により視覚障害者の歩行支援を行うシステムの提案・評価がされています。事前の地図を必要とせずにスマートフォン 1 台で、障害物回避と交差点の把握を支援するシステムであり、視覚障害者の実際のニーズに合致した実用的な研究であること、また被験者も十分な数の視覚障害者によって構成されていることから、社会的貢献度の高い研究です。

セッション5：現地発表3

CircWood: CW レーザー加工機による炭化を利用した木製回路基板 ショート採録 ショート発表
石井 紗郁, 加藤 邦拓, 池松 香, 川原 圭博, 椎尾 一郎 105

一般的なレーザーカッターでの加工時に意図的にレンズフォーカスをずらすことで、木材などの表面の一部を効率的に炭化させ、センサや回路として利用できる造形手法を提案している。各査読者が共通して、新規性／有用性／正確性ともに十分高いと評価した。一方、支援ツール等は実装されておらず、パラメータ設定等の性能面も一機種に限定され汎用性が不明瞭であるため、アイデアと制作例等を中心としたショート発表が適すると判定した。

ペイント可能なタイムラインによる対話的なハイパーパラメータ最適化 ショート採録 ショート発表
樋口 啓太, 佐野 正太郎, 五十嵐 健夫 113

機械学習のパラメータチューニングにペインティング手法を応用したシステムの提案で、シンプルなアイデアであるが他の新規性があり、また機械学習専門家による評価実験により有効性を示している。アルゴリズムに関する説明不足や、論文での不十分な記述を修正することを条件として、ショート採録と判定した。

液体噴霧混合式の味ディスプレイの試作 ショート採録 ショート発表
宮下 芳明 121

すべての査読者が実装と潜在的な有用性を高く評価しており、4~6と採択を推すスコアをつけています。実際に発表をしていただいた際には、今後の評価や応用の発展の方向性について面白い議論が可能だと思います。

その一方で、評価とみなせる事項としては著者の過去の研究との比較があるだけで、評価実験や過去の関連研究がほとんど記載されていません。そのため、現段階の原稿に対する発表としては、ロング発表の時間は必要ないと判断しました。そのため、採録条件のないショート採録という判定とします。

評価のないこの段階ではデモこそがとても輝く研究だと思いますので、ぜひ可能なレベルで実機を紹介していただければ大変有意義な発表になると思います。

UPLIGHT: 全周囲ディスプレイと携帯型ゲームデバイスの融合 ショート採録 ショート発表
中村 俊勝, 加藤 達也, 伊藤 思音, 佐藤 俊樹 128

本論文は、1人のユーザを対象に、全周囲ディスプレイの見えない面を能動的に覗き込む行為のエンタテイメント性に着目したデバイス「UPLIGHT」を提案している。本デバイスのエンタテインメント性の議論が弱いという問題はあるものの、本デバイスのデザイン戦略には一定の新規性が認められる点を踏まえ、本論文はショート枠採録と判断する。

能動的音響計測に基づくスタイラスの把持状態識別手法の検討 ショート採録 ショート発表
高倉 礼, 鈴木 健介, 國分 晴利, 志築 文太郎 135

本論文は、能動的音響計測をスタイラスに適用し、スタイラスの把持状態を識別する手法を提案しています。把持状態認識結果に応じてタブレットの機能（鉛筆、筆、消しゴムなど）を切り替えることを狙っています。能動的音響計測を用いて物体の種類や形状、把持状態を認識する研究はこれまでにもありますが、スタイラスに適用し、スタイラスを3Dプリントして充填率を変えるという着眼点には一定の新規性があると判断しました。また、評価実験からも実用的な精度が得られ、深い考察もなされていると判断しました。

SingDistVis: 多数の歌声から歌い方の傾向を可視化できるインターフェース

伊藤 貴之^{*} 中野 倫靖 深山 覚 濱崎 雅弘 後藤 真孝[†]

概要. 人々が自身の歌った動画を動画共有サービスへ投稿する機会が増え、同一楽曲を異なる人々がそれぞれ歌った「同曲異唱コンテンツ」を楽しむことができるようになった。このような歌唱群の傾向を理解する手段として、歌い方としての基本周波数 (F0) 推移の分布をインラクティブに可視化できる歌声情報可視化インターフェース SingDistVis を提案する。SingDistVis は、曲全体を時間-周波数平面上で概略的に可視化する Overview 機能と、その局所部分をズームアップして歌唱力とともに可視化する Detail 機能の組み合わせにより、F0 分布のマルチスケール可視化を実現する。本論文では、1,000 人の歌唱者による同曲異唱を可視化し、被験者実験による可視化パラメータの最適化と有効性評価を行った結果を報告する。

1 はじめに

大規模データの分布や傾向を理解するための情報可視化技術は従来から活発に研究されている。近年では単に新しいデータ描画手段を開発するだけでなく、可視化への要求やタスクの定義、またそのタスクを実現するためのシステムデザインやインラクションデザイン、さらには可視化結果の可読性や視認性の評価なども重要な研究課題である。

そこで本論文では、大規模データの情報可視化の事例として、同一楽曲に対する多数の歌唱データである「同曲異唱コンテンツ」を対象に、歌い方の分布や傾向を可視化するインターフェース SingDistVis を提案する。人々が自身の歌った既存曲の歌唱（動画）を動画共有サービスやソーシャルカラオケサービス¹へ投稿した同曲異唱コンテンツは、オリジナル曲の鑑賞の楽しみ方を拡張し、新たな好みのコンテンツやアーティストの発見につながる[6]。さらに、歌唱における声質と歌い方はそれぞれ、歌手の個性や類似性を特徴づける重要な要素であることから[26]、同一曲に対する多様な声質と歌い方の大規模データの分析は、人の歌声知覚・生成や歌唱スタイル等特性の違いを分析・理解する上でも重要である。

本論文では、歌唱における歌い方に着目し、音高として基本周波数 (F0) 推移の傾向・分布を可視化する。従来、大規模な楽曲データを対象として、そこから 1 曲を選択するために楽曲単位で可視化するインターフェースは存在したが（例えば、[30, 6] 等）、曲の内容も含めた可視化は研究されてこなかった。例えば、1 曲分の単独歌唱において F0 推移は折れ線で可視化されることが多いが、多数の F0 を単純に重ね合わせてしまうと図 1 のように違いや傾向を観察できず、また 1 曲を選び出すことは難しい。

そこで SingDistVis では、「Overview+Detail」[3]

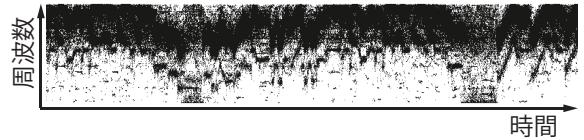


図 1. 多数の F0 を単純に折れ線集合で表示すると、過密による Visual Cluttering 問題が発生する。

という可視化デザインに沿った、以下の 2 画面の構成によるマルチスケールな可視化を実現する。

Overview: 大域的な F0 分布をヒートマップで概略的に表現する画面（例：曲の 1 番を全て表示）。

Detail: 局所的な F0 分布を細かな粒度で可視化するため、折れ線の集合で表現する画面。歌唱力の違いも色の違いで可視化する。

SingDistVis の操作手順を図 2 に示す。ユーザはまず Overview 画面から、F0 推移及びその各時刻での分散の 2 種類のヒートマップ表示を用いて、異なる音高で発声している歌唱者がいるフレーズや、F0 が幅広く分散するフレーズが、楽曲の中でどのように存在しているかを観察する。続いて、特徴的な歌唱の発見等のため、Overview 画面で特定のフレーズを矩形領域として指定することで、Detail 画面でその F0 分布を折れ線集合として詳細に観察できる。さらに Detail で特定の歌唱を指定すると、その F0 推移を Overview 上で重畠させて観察できる。

本論文は我々自身の先行発表[27]をインターフェース提案の観点からまとめ直している。歌唱力と F0 推移に関する知見から要求定義を検討し、パラメータ決定と評価のための被験者実験を追加している。

2 関連研究

同曲異唱を扱う研究事例として、Hamasaki *et al.* は、ニコニコ動画の歌ってみた動画を対象に、各歌唱の再生回数や男女比を可視化しながら、各歌唱を

* お茶の水女子大学

† 産業技術総合研究所

¹ ニコニコ動画、YouTube、Sing!カラオケ (Smule) 等。

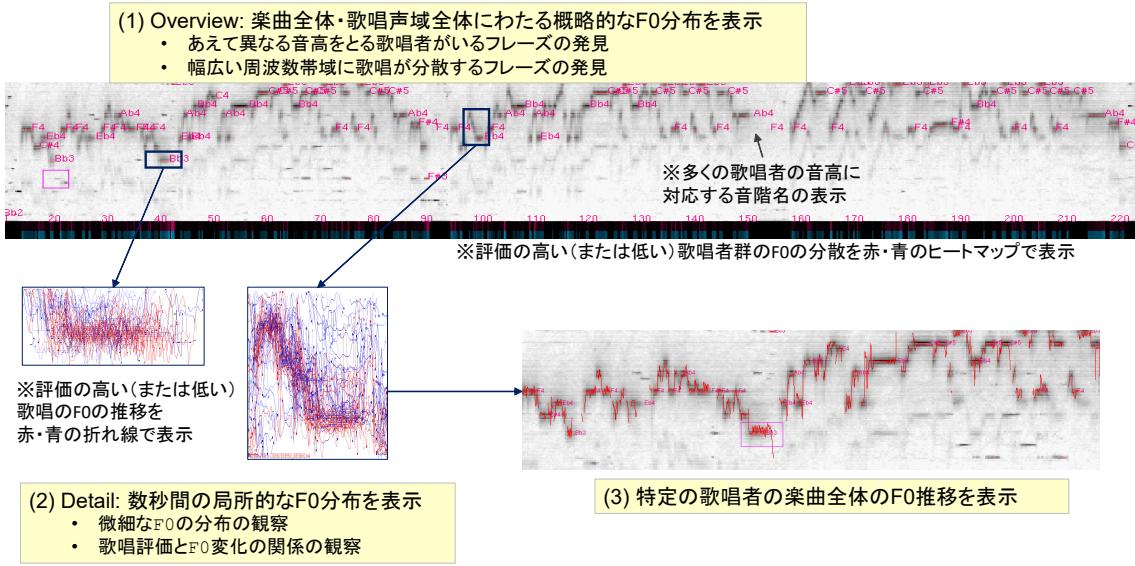


図 2. SingDistVis のスナップショットおよび操作手順. (1) Overview: 1 曲全体の F0 群と各時刻におけるその分散をヒートマップで可視化. (2) Detail: Overview で指定した矩形領域の F0 分布を, 歌唱力の違い (赤=高評価, 青=低評価) とともに, 折れ線の集合で可視化. (3) Detail で指定した F0 推移の Overview での重畠表示.

自動的につないで一つの楽曲としてリレー再生する Songrium RelayPlay を提案した [6]. 都築らは, 複数歌唱を同時再生して仮想的な合唱を制作できる Unisoner を提案し, 4,524 人の F0 を活用して F0 推定精度を向上させる技術, F0 分布を局所的にヒートマップで可視化する機能も提案した [28]. また近藤及び著者らは, 2,024 人の F0 に対し, ヒートマップを用いた大域的な可視化を提案した [25].

歌唱分析に F0 可視化を用いた研究事例として, 歌唱力向上等を目的として, 単一歌唱の F0 推移の可視化は多く対象とされてきた (例えば, [7, 29, 16, 13, 19] 等). ここで, F0 推移を音符単位にまとめ, 音高の楽譜からのずれを色の違いで可視化する研究もある [15, 24]. また, 歌唱スタイルの違いを可視化・識別することを目的として, 可視化の次元 (軸) を増やし, F0 と音量を 2 軸とした可視化 [11], F0 と F0 差分を 2 軸とした可視化 [9] が報告されている.

これらの研究に対して SingDistVis では, 数千人規模の F0 をマルチスケールで同時可視化する点で独自性を有する. 近藤及び著者らの研究 [25] を拡張し, Overview+Detail デザインを導入して, 概略と詳細のインタラクティブな可視化を可能とする.

3 大規模 F0 データの可視化方策

歌唱の F0 推移は時系列データとして扱うことができる. K 個の標本が, それぞれ L 個の時刻 t_l における実数値 $p_k(t_l)$ を有する場合, 以下の 3 種類の可視化手法の適用が可能である.

1. 一方の座標軸を t_l , 他方の座標軸を $p_k(t_l)$ とした折れ線グラフ [22, 18, 23] や散布図 [23].

2. 1. の折れ線や点群をその密度で表し, それを各画素の明度や色相に変換したヒートマップ.
3. 一方の座標軸を t_l , 他方の座標軸を K 個の標本番号 k とした上で, $p_k(t_l)$ を各画素の明度・色相で表現したヒートマップ [8, 17], もしくは線分の太さ [10] で表現する可視化手法.

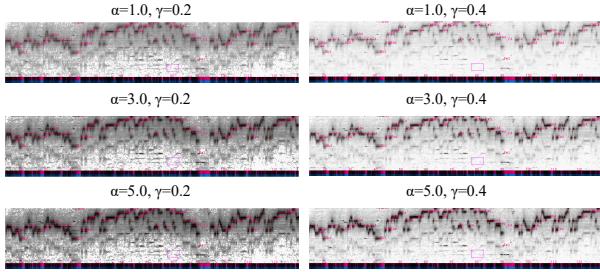
このうち 1. の折れ線グラフや散布図では, Visual Cluttering と呼ばれる視認性の低下が避けられない (図 1). 3. のヒートマップは, 色の識別能力の限界により数値を正確に読み取ることが難しく, 他方, 線分の太さの利用では描画に必要な画面サイズに限界があることから大規模データの可視化が難しい.

そこで, SingDistVis の Overview 画面には 2. の「密度のヒートマップ」を採用した. 一方で Detail 画面には, 微細な違いを視認しやすい 1. の「折れ線グラフ」を採用し, 表示対象となる折れ線の本数を制御することで Visual Clutter の影響を抑える.

従来, 時系列データの大局的及び局所的な時間変化に注視するマルチスケールな可視化としては, 時系列データを同じ視覚表現で拡大縮小するか [4, 20], 異なる時系列データを視覚表現を切り替えて組み合わせていた [5, 2]. これに対して SingDistVis では, 同一の時系列データ (F0) に対して, 異なる視覚表現を適用して歌唱分布を表現した点に特徴がある.

4 SingDistVis の実装

前章での議論に従い, 我々が開発した F0 分布の可視化手法の処理手順・操作手順を述べる. なお我々の先行発表 [27] では, より詳細な説明を含む.

図 3. α および γ の調節に対する可視化結果の変化.

4.1 前提となるデータ: F0 と歌唱評価結果

本論文では、 K 個の歌唱で構成される集合 S において、その要素 s_k を以下のように定義する。

$$\begin{aligned} S &= \{s_1, s_2, \dots, s_K\} \\ s_k &= \{p_{k1}, p_{k2}, \dots, p_{kL}, e_k\} \end{aligned} \quad (1)$$

ここで L は標本化された時刻の総数であり、 p_{kl} は k 番目の歌唱の l 番目の時刻における F0 の対数、 e_k は k 番目の歌唱への評価値（5 章で説明）である。無音部分の対数 F0 は、便宜上 $p_{kl} = 0$ とした。

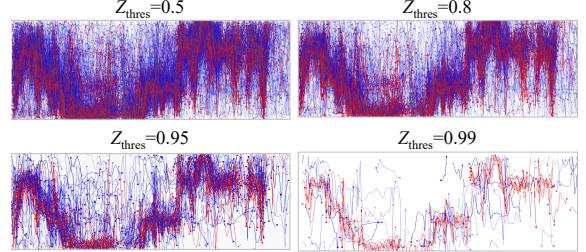
このようなデータを対象として、「Overview」「Detail」の可視化ではそれぞれ以下のタスクを支援する。

Overview: 異なる音高で発声している歌唱のあるフレーズや、幅広い周波数帯域に分散して歌唱が分布するフレーズが、楽曲全体の中でどのように存在しているかを観察する。そして注目に値するフレーズの F0 分布を対話操作によって指定し、Detail にて表示させる。

Detail: 楽曲中の数秒単位の短時間を対象として、歌唱間の微小な F0 の差異や、その分布と歌唱評価との関係を観察する。そして注目に値する特定の歌唱者を対話操作によって指定し、その歌唱者の楽曲全体にわたる F0 推移を Overview にて表示させる。

4.2 Overview: ヒートマップでの大域表示とインタラクション

Overview では F0 推移の密度分布をヒートマップで表現する。時刻を横軸、周波数の対数を縦軸とした長方形領域を設定し、これを格子状に分割する。時間方向は、歌唱の開始時刻および終了時刻を長方形領域の左右端に割り当てて N 個に分割する。また周波数領域の上限と下限を上下端にわりあて、これを M 個に分割する。分割してきた各領域を通過する F0 の個数を集計することで、2 次元ヒストグラムを生成し、これを $M \times N$ 画素のヒートマップ画像として扱う。左から u 画素目、下から v 画素目の明度 I_{uv} は、その画素に存在する p_{kl} の個数 r_{uv}

図 4. Z_{thres} の調節に対する折れ線の表示本数の変化.

から、以下のように求める。

$$I_{uv} = 1.0 - (\alpha r_{uv})^\gamma \quad (2)$$

ただし α および γ はユーザが調節可能なパラメータとし、その違いによる見た目の変化を図 3 に示す。 α が大きいと r_{uv} が大きい部位での明度が低く（＝黒が濃く）なり、 γ が大きいと r_{uv} が大きい画素と小さい画素のコントラストが大きくなる。

Overview ではさらに、各時刻における F0 の分散もヒートマップで表示する。ここでは、歌唱評価 e_k の高い（低い）歌唱群の F0 の分散を赤（青）のヒートマップとした。これにより、「高評価な歌唱者が異なる歌い方をしているフレーズ」、「低評価な歌唱者が音高を正しくとれない傾向があるフレーズ」などを発見しやすくなり、Detail で注目する矩形領域を決定する際の有力なヒントにもなる。

Overview には、時刻と F0 の範囲を示す矩形領域を重ねて表示する。ユーザは、マウスポインターの移動とともにこの矩形領域を移動させ、興味がある領域でクリックすることで、後述する Detail において局所的に詳細な可視化を観察できる。

4.3 Detail: 折れ線集合での局所表示とインタラクション

Detail では Overview の矩形領域内部に対応する F0 推移を折れ線の集合で表現するが、前述した Visual Cluttering を抑制するために、同時に描画する折れ線の本数を制御する。

サンプリングおよびクラスタリングの 2 種類に基づく本数制御を実装したが [27]、本論文ではサンプリングによる折れ線の本数制御のみを説明し、後述の実験でもこれのみを用いる。これは、各々の折れ線 p_k に対して、ユーザ指定のタイミングで再計算できる一様乱数 z_k ($0.0 \leq z \leq 1.0$) を割り当て、以下を満たす場合のみ描画する。

$$\beta_{e_k} z_k > Z_{thres} \quad (3)$$

ここで、 β_{e_k} は歌唱評価 e_k に応じた係数、 Z_{thres} は GUI スライダーで調整する閾値であり、いずれもユーザが調節可能なパラメータである。例えば高（低）評価な p_k のみを表示したい場合は、その β_{e_k}

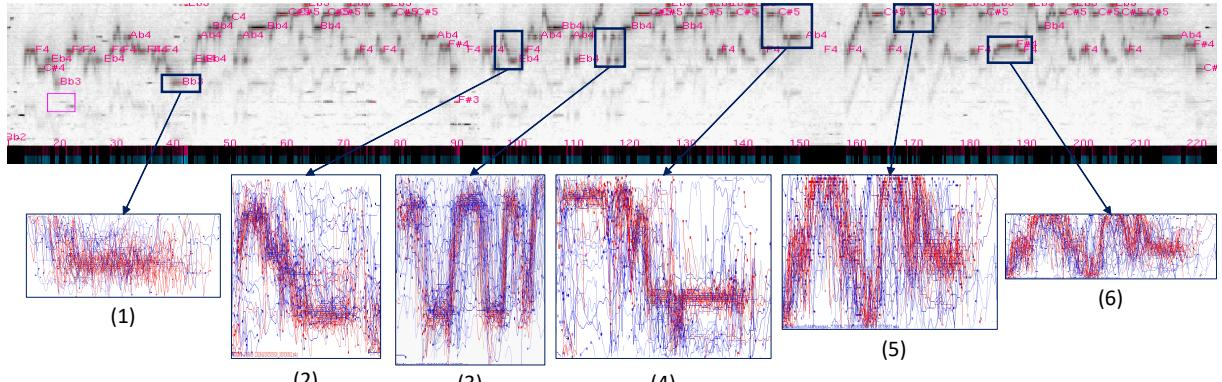


図 5. Overview と 6 箇所の Detail 例. Detail では、歌唱評価 e_k が高いほど赤く、低いほど青い折れ線で表示される.

が大きな値を取るように設定する。 $e_k = 1$ は歌っていない歌唱等に付与されたため、歌唱力が高い $e_k = 5$ と低い $e_k = 2$ の F0 が強調されるように、 $\beta_{e_k=2,5} = 1.0$, $\beta_{e_k=4} = 0.8$, $\beta_{e_k=1,3} = 0.5$ とした。

さらに、折れ線の色と不透明度を e_k に応じて変更[27]することで、歌唱力の違いも可視化した（高いほど赤く、低いほど青い）。また、発声タイミングについても、F0 系列の開始と終了を強調した。図 4 に、 Z_{thres} の値と折れ線の可視化結果を示す。

また、局所的な観察結果を大域表示と関連付けるために、Detail で特定の F0 推移をクリック操作で指定すると、Overview にも折れ線で重畠表示される。

5 SingDistVis の実行例

SingDistVis の実行結果を紹介する。著者らは本可視化手法を Java 1.12.0 および JOGL (Java binding for OpenGL) 2.3.2 で実装した。実行例には、DAMP (Let It Go Vocal Performances) [1] に収録された “Let It Go” を 231,278 人が歌った同曲異唱の無伴奏音響データから、1,000 人分を無作為抽出して適用した。これらの歌唱に対して pYIN [12] を用いて音響データから F0 を算出し、その相互相関最大化によって歌唱間の時間ずれを補正した。

Overview 可視化の画素数は $N = 1000, M = 480$ とし、対象となる周波数を 110Hz～1760Hz (音階で A1 から A5) の 4 オクターブとした。

Detail で用いる歌唱力評価結果 e_k は、1人の評価者が各歌唱を 5段階のリッカート尺度に基づいて主観評価した結果とした。評価者は 12 年間ピアノの指導を受けた経験があり、対象楽曲（英語原曲）を必要に応じて聴取しながら、各歌唱 1 コーラス目のサビ直前からそのサビ終了までを聴取して評価した。

5.1 Overview: Detail 表示箇所の探索

楽曲全体を表す Overview と、矩形領域を移動させた 6 箇所の Detail 例を図 5 に示す。図上部の Overview では、F0 分布ヒートマップが黒いほど多くの歌唱者が同一音高で発声していることを示す。

ユーザは、Overview の下端に表示される赤と青のヒートマップ（高・低評価の F0 分散）も参考にしながら、Detail を表示する領域を決定する。例えば、分散が大きい箇所は「歌い方が異なる」「難しいので多くの人が間違える」可能性があり、歌唱力が高くて異なった歌い方をしている歌唱者や、重点的に練習すべき箇所の発見等につながると考える。

5.2 Detail: 歌唱評価 e_k と F0 推移の詳細を観察

図 5(1)(2) は、いずれも音階が下降した後にロングトーンがある。図 5(1) は赤い折れ線が太い帯を形成しており、高評価な歌唱であっても F0 に幅があった。それに対して図 5(2) は、少数の高評価の歌唱者が半音単位で異なる F0 で発声し、何人かの低評価の歌唱者がそれ以上に大きく F0 を外していた。

図 5(3)(4)(5) は、いずれも音階が大きく上下するフレーズである。高評価な歌唱と低評価の歌唱とは F0 の正確さが異なることが観察できた。また図 5(5)(6) では図 5(1) と同様に、フレーズの末尾において、高評価な歌唱者であっても F0 に幅があった。

5.3 Detail と Overview の重畠表示

図 5(1) に示す Detail から、特定の折れ線をクリックして、Overview にて重畠表示した例を図 6 に示す。

図 6(1) は図 5(1) 中で赤い折れ線が集中している部分から抽出しており、楽曲全体にわたって正確な F0 で歌唱していた。高評価な歌唱の多くに同様な F0 推移が見られた。それに対して少数ながら、図 6(2) のように、多数の歌唱者とは異なる F0 推移を有しながらも高評価を得ている歌唱もあった。このような「多数の歌唱者と異なる F0 推移をとりながらも高評価な歌唱者」の例をいくつか聴取したところ、歌唱技法（ビブラート等）のスタイルや癖、タイミングのずれ、F0 誤推定、等の理由があった。

一方で低評価な歌唱の中には、多数の歌唱者 F0 推移に近い図 6(3) と、多数の歌唱者の F0 推移から大きく離れた図 6(4) が見られた。前者は F0 推移には表れない声質の問題（声が細い、舌足らず）や、

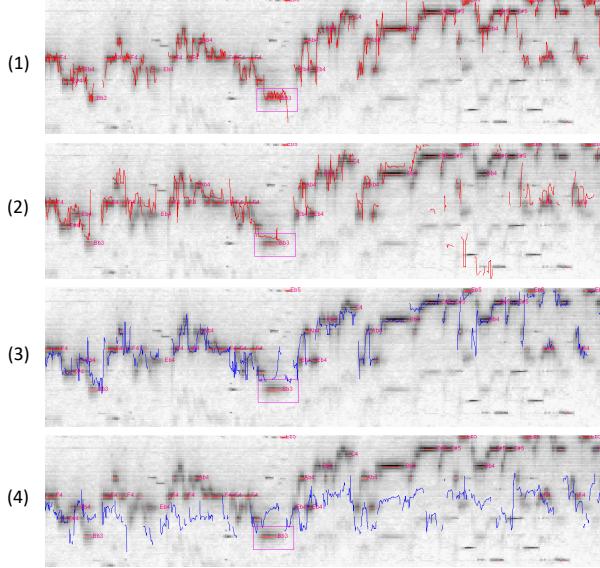


図 6. 特定歌唱者の F0 推移の例. (1)(2) 高評価と, (3)(4) 低評価の歌唱例. また F0 推移が多数の歌唱者と (1)(3) 類似するか, (2)(4) 異なる歌唱例.

声域の狭さ（高音部・低音部に難がある）等が原因で、低評価につながったと考えられる。

6 実験：被験者主観評価

SingDistVis の可視化パラメータの選定と、可視化デザインの有効性を被験者実験により評価する。条件統制のため、いずれの評価においても、被験者はインターフェース操作をせず、可視化結果の画像のみがディスプレイ上で呈示されて質問に回答した。

被験者は、ディスプレイによる閲覧が可能で色覚に不安がない、情報科学を専攻する学部学生 18 名であった。音楽経験は問わなかったが、実験前に可視化結果の見方を教示した。

6.1 可視化パラメータ値の選定

可視化結果の視認性向上には、適切なパラメータの設定が重要であることが知られている [23, 21, 14]。

本実験では、Overview におけるヒートマップの明度とコントラストに関する α 及び γ (式(2)) と、Detail における折れ線の本数制御に関する Z_{thres} (式(3)) について、パラメータを振って生成した複数画像と質問を被験者に呈示する。後述の質問 1~6 と画像の組合せをランダムに呈示した。

6.1.1 Overview : ヒートマップパラメータ

Overview については、後述する図 7 に示す α と γ の組み合わせで 20 通りの画像を生成し、左から右へ α が増加する順、下から上へ γ が増加する順で全てを同時に配置した 1 枚の画面を被験者に呈示

$\gamma=0.4$	2 6 17 9 12	$\gamma=0.4$	4 10 9 8 8
0.3	3 15 26 22 14	0.3	3 8 17 17 16
0.2	8 5 4 2 4	0.2	6 10 14 10 9
0.1	0 1 5 3 4	0.1	4 4 3 1 0
	$\alpha=1.0 \ 2.0 \ 3.0 \ 4.0 \ 5.0$		$\alpha=1.0 \ 2.0 \ 3.0 \ 4.0 \ 5.0$

図 7. 質問 1(左) と質問 2(右) に対する回答分布。

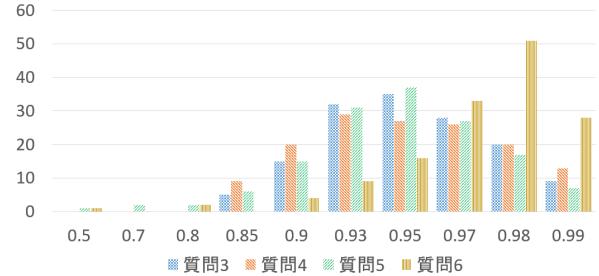


図 8. 質問 3~6 に対する回答分布。

した。全パラメータ値の可視化を同時に呈示するため、また曲中の場所に応じた影響の考察につなげるため、80 秒分のヒートマップを無作為に切り出した 3 箇所とした。3 箇所それぞれについて、20 通りの画像の中から以下の 2 つの質問に最も該当する画像を J_1 個選択してもらった ($J_1 = 3$)。

質問 1: 歌唱の全体表示として回答者が最も使いたい可視化はどれか。

質問 2: ズームアップして確認したい部分を選ぶのに向いた可視化はどれか。

3 箇所の可視化結果に対する被験者の選択結果の集計値を図 7 に示す。この結果から $\alpha = 3.0$, $\gamma = 0.3$ が良好なパラメータ値であると判断し、6.2 節での評価実験でもこの値を採用した。これらは全質問の全 3 箇所中、常に最多か 2 番目の回答数であった。

6.1.2 Detail : 折れ線の本数制御パラメータ

Detail については、後述する図 8 に示す Z_{thres} の値によって、10 通りの画像を生成し、 5×2 の配置で左上から右下まで値が増加する順に配置した 1 枚の画面を被験者に呈示した。これを無作為に選んだ曲中の 4 箇所について、以下の 4 つの質問に最も該当する画像を J_2 個選択してもらった ($J_2 = 2$)。

質問 3: 高評価な歌唱者は正確な音高で歌唱しているか、高評価だけどあえて音高をずらしている歌唱者がいるか、正確な音高なのに低評価な歌唱者がいるか、といった点を発見しやすい可視化結果はどれか。

質問 4: 全体からずれた音高で歌唱している人が、意図的な歌唱技法として音高をずらしているのか、技量不足や練習不足により意図せずに音

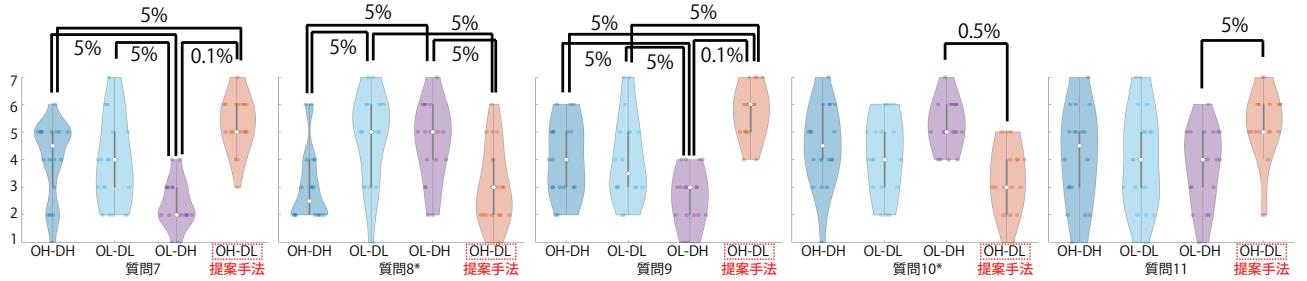


図 9. ポジティブな質問 7,9,11 とネガティブな質問 8*,10*に対する回答分布と仮説検定の結果（数字は有意水準）。

高がずれているのか、といった点を議論しやすい可視化結果はどれか。

質問 5: 歌唱評価結果と音高の正確さを比較しながら分析しやすい可視化はどれか。

質問 6: 注目すべき歌唱者を適切に発見できそうな可視化はどれか。

4箇所の可視化結果に対する被験者の選択結果の集計値を図 8 に示す。6.2 節での評価実験では、全集計で最多であった $Z_{thres} = 0.95$ を採用した。

ここで、特定の歌唱を全体傾向と比較する際（質問 3~5）では、 $Z_{thres} = 0.9$ から 0.97 のいずれかが最多回答となり、選択箇所に応じて多少ばらつきがあった。また、特定の歌唱を発見する際（質問 6）では、より折れ線数の少ない $Z_{thres} = 0.98$ が好まれる傾向にあり、全 4 個所で常に最多の回答を得た。

6.2 可視化デザインの正当性評価

SingDistVis の Overview (O) と Detail (D) について、ヒートマップ (H) と折れ線 (L) を用いて、視覚表現の組合せ方と個別の評価を実施する。後述の質問 7~13 と画像の組合せをランダムに呈示した。

6.2.1 組合せ方の評価

4通りの組合せ OH-DH, OL-DL, OL-DH, OH-DL (提案手法) を生成して比較した。被験者は、以下のポジティブもしくはネガティブ (*) な質問に、それぞれ 7段階のリッカート尺度で回答した。

質問 7: 頻繁に使いたい

質問 8*: 不必要に複雑である

質問 9: 使いやすい

質問 10*: 使用するには技術的な支援が必要だ

質問 11: 大量の歌唱の同時可視化にふさわしい

回答の分布をバイオリンプロットで図 9 に示す。また Wilcoxon の順位和検定を行い、帰無仮説が棄却されたペアを有意水準とともに図中に示す。

この図から、提案手法 (OH-DL) は、ポジティブな質問 7,9,11 について他の組合せよりも数値が大きく、ネガティブな質問 8*,10* については数値が小さかった。例外として質問 8 については、OH-DH が高評価（最も低い中央値）であった。

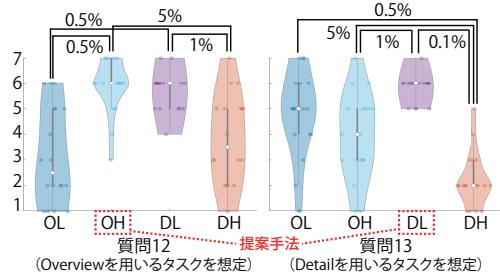


図 10. 質問 12~13 の回答分布と仮説検定結果。

6.2.2 個別の評価

OL, OH, DL, DH 4種類の可視化を生成して比較した (OH と DL が提案手法)。

被験者は、以下の質問について 7段階のリッカート尺度で回答した。

質問 12: 多くの歌手が同じ音高で歌唱している部分を発見しやすい

質問 13: 歌手によって歌い方の差が大きい部分を発見しやすい

回答の分布をバイオリンプロットで、図 10 に示す。また前項と同様、Wilcoxon の順位和検定の結果も示す。質問 12 は Overview を用いるタスクを想定した質問であり、提案手法 (OH) が高評価を得た。DL も高評価となつたが、追加ヒアリングの結果、折れ線の密度で判断しやすいとの回答が多かった。

質問 13 は Detail を用いるタスクを想定した質問であり、これも提案手法 (DL) が高評価を得た。

7 まとめ・今後の課題

本報告では大規模な同曲異唱コンテンツの F0 分布を「Overview+Detail」という可視化デザインに基づいて探索的に可視化する手法を提案し、その実行例と被験者主観評価結果を示した。今後の課題として、6.1 節で選定した可視化パラメータ値の一般化（異なるデータ数への適用等）や、インタフェース操作に関する評価実験を進めたい。またさらに、先行発表 [27] にも述べた通り、歌唱データ処理の改善、歌唱支援の実用シナリオに沿った可視化手法の拡張、歌唱以外のデータへの適用も検討していきたい。

参考文献

- [1] WELCOME TO STANFORD'S DAMP: Stanford Digital Archive of Mobile Performances, a repository of geo-tagged mobile performances to facilitate the research of amateur practices. <https://ccrma.stanford.edu/damp/>.
- [2] P. André, M. L. Wilson, A. Russell, D. A. Smith, A. Owens, and M. C. Schraefel. Continuum: Designing Timelines for Hierarchies, Relationships and Scale. In *Proc. ACM Symposium on User Interface Software and Technology (UIST 2007)*, pp. 101–110, 2007.
- [3] A. Cockburn, A. Karlson, and B. B. Bederson. A review of overview+detail, zooming, and focus+context interfaces. *ACM Computing Surveys*, 41(2):1–31, 2009.
- [4] E. Cuenca, A. Sallaberry, F. Y. Wang, and P. Poncelet. MultiStream: A Multiresolution Streamgraph Approach to Explore Hierarchical Time Series. *IEEE Trans. on Visualization and Computer Graphics*, 24(12):3160–3173, 2018.
- [5] M. Goto, K. Yoshii, H. Fujihara, M. Mauch, and T. Nakano. Songle: A Web Service for Active Music Listening Improved by User Contributions. In *Proc. 12th International Society for Music Information Retrieval Conference (ISMIR 2011)*, pp. 311–316, 2011.
- [6] M. Hamasaki, K. Ishida, T. Nakano, and M. Goto. Songrium RelayPlay: A Web-based Listening Interface for Continuously Playing User-generated Music Videos of the Same Song with Different Singers. In *Proc. International Computer Music Conference (ICMC 2020)*, pp. 1–4, 2020.
- [7] D. Hoppe, M. Sadakata, and P. Desain. Development of Real-time Visual Feedback Assistance in Singing Training: A Review. *Journal of computer assisted learning*, 22(12):308–316, 2006.
- [8] M. Imoto and T. Itoh. A 3D Visualization Technique for Large Scale Time-Varying Data. In *Proc. International Conference on Information Visualisation (IV10)*, pp. 17–22, 2010.
- [9] T. Kako, Y. Ohishi, H. Kameoka, K. Kashino, and K. Takeda. Automatic Identification for Singing Style Based on Sung Melodic Contour Characterized in Phase Plane. In *Proc. 10th International Society for Music Information Retrieval Conference (ISMIR 2009)*, pp. 393–398, 2009.
- [10] M. Krstajic, E. Bertini, and D. A. Keim. Cloud-Lines: Compact Display of Event Episodes in Multiple Time-Series. *IEEE Trans. on Visualization and Computer Graphics*, 17(12):2432–2439, 2011.
- [11] K. W. E. Lin, H. Anderson, N. Agus, C. So, and S. Lui. Visualising Singing Style Under Common Musical Events Using Pitch-Dynamics Trajectories and Modified TRACLUS Clustering. In *Proc. 13th International Conference on Machine Learning and Applications (ICMLA'14)*, pp. 237–242, 2014.
- [12] M. Matthias and S. Dixon. pYIN: A fundamental frequency estimator using probabilistic threshold distributions. In *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2014)*, pp. 659–663, 2014.
- [13] O. Mayor, J. Bonada, and A. Loscos. Performance Analysis and Scoring of the Singing Voice. In *Proc. AES 35th International Conference*, pp. 1–7, 2009.
- [14] L. Micallef, G. Palmas, A. Oulasvirta, and T. Weinkauf. Towards Perceptual Optimization of the Visual Design of Scatterplots. *IEEE Trans. on Visualization and Computer Graphics*, 23(6):1588–1599, 2017.
- [15] F. Moschos, A. Georgaki, and G. Kouroupetroglo. FONASKEIN: An Interactive Software Application for the Practice of the Singing Voice. In *Proc. Sound and Music Computing Conference (SMC) 2016*, pp. 326–331, 2016.
- [16] T. Nakano, M. Goto, and Y. Hiraga. MiruSinger: A Singing Skill Visualization Interface Using Real-Time Feedback and Music CD Recordings as Referential Data. In *Proc. IEEE International Symposium on Multimedia (ISM 2007) Workshops*, pp. 75–76, 2007.
- [17] G. Oliveira, J. Comba, R. Torchelsen, M. Padilha, and C. Silva. Visualizing Running Races through the Multivariate Time-Series of Multiple Runners. In *Proc. Conference on Graphics, Patterns and Images (SIBGRAPI 2018)*, pp. 99–106, 2013.
- [18] C. Perin, F. Vernier, and J.-D. Fekete. Interactive Horizon Graphs: Improving the Compact Visualization of Multiple Time Series. In *Proc. ACM CHI Conference on Human Factors in Computing Systems (CHI '13)*, pp. 3217–3226, 2013.
- [19] M. Shiraishi, K. Ogasawara, and T. Kitahara. HamoKara: A System that Enables Amateur Singers to Practice Backing Vocals for Karaoke. *Journal of Information Processing*, 27:683–692, 2019.
- [20] K. Sjölander and J. Beskow. Wavesurfer - An Open Source Speech Tool. In *Proc. International Conference on Spoken Language Processing (ICSLP 2000)*, pp. 464–467, 2000.
- [21] S. Smart and D. A. Szafrir. Measuring the Separability of Shape, Size, and Color in Scatterplots. In *Proc. ACM CHI Conference on Human Factors in Computing Systems (CHI '19)*, pp. 669:1–14.
- [22] Y. Uchida and T. Itoh. A Visualization and Level-of-Detail Control Technique for Large Scale Time Series Data. In *Proc. 13th International Conference on Information Visualisation (IV09)*, pp. 80–85, 2009.
- [23] Y. Wang, F. Han, L. Zhu, O. Deussen, and B. Chen. Line Graph or Scatter Plot? Automatic Selection of Methods for Visualizing

- Trends in Time Series. *IEEE Trans. on Visualization and Computer Graphics*, 24(2):1141–1154, 2018.
- [24] C. Weiβ, S. J. Schlecht, S. Rosenzweig, and M. Muller. Towards Measuring Intonation Quality of Choir Recordings: A Case Study on Bruckner's Locus Iste. In *Proc. 20th International Society for Music Information Retrieval Conference (ISMIR 2019)*, pp. 276–283, 2019.
 - [25] 近藤 芽衣, 伊藤 貴之, 中野 倫靖, 深山 覚, 濱崎 雅弘, 後藤 真孝. 同一楽曲に対する多数の歌唱と正解歌唱の音高推移分布の可視化. 情報処理学会 インタラクション 2021 (インタラクティブ発表), pp. 745–750, 2021.
 - [26] 斎藤 毅, 柳原 健一. 物真似歌唱の音響特徴とその知覚への影響の調査. 日本音響学会 秋季研究発表会 講演論文集, 2011, pp. 571–574, 2011.
 - [27] 伊藤 貴之, 中野 倫靖, 深山 覚, 濱崎 雅弘, 後藤 真孝. 同一楽曲に対する多数の歌唱の基本周波数分布のマルチスケール可視化. 情報処理学会 研究報告 音楽情報科学 (MUS), 2021-MUS-132-5, pp. 1–8, 2021.
 - [28] 都築 圭太, 中野 倫靖, 後藤 真孝, 山田 武志, 牧野 昭二. Unisoner: 様々な歌手が同一楽曲を歌った Web 上の多様な歌声を活用する合唱制作支援インターフェース. 情報処理学会論文誌, 56(12):2370–2383, 2015.
 - [29] 平井 重行, 片寄 晴弘, 井口 征士. 歌の調子外れに対する治療支援システム. 電子情報通信学会論文誌, J84-D-II(9):1933–1941, 2001.
 - [30] 魚田 知美, 伊藤 貴之. GRAPE: グラデーション画像によるポータブルビジュアルプレイリスト. 第 41 回可視化情報シンポジウム講演論文集, pp. 299–304, 2013.

未来ビジョン

本論文が提案する SingDistVis は、大規模な同曲異唱データの中から特徴的な歌唱を対話的に発見する新しいインターフェースである。本論文ではその研究の初期段階として、インターフェースの根幹となる可視化手法を示し、ユーザ評価実験によりその可視化デザインの妥当性を検証した。一方で、現状では、同曲異唱データにおける F0 の分布を概観し、特徴的な F0 推移を有する歌唱を発見するための汎用的な手法を開発した段階であり、今後は歌唱支援のための具体的なタスクを想定した拡張が重要であると考える。

本研究が目指す未来の一つとして、多数の歌唱者の中から自分がお手本としたい歌唱者・歌唱技法を探すことで多様な歌唱技法を理解し、その結果として発見した歌唱にならうことで歌唱力を向上させる、という一連のタスクを支援するインターフェースとしての拡張を考えら

れる。あるいは、同一楽曲を反復練習した過程を可視化することで、反復練習によってどのフレーズにおいて歌唱者がどの程度上達したか、あるいはどのフレーズにおいてなかなか上達していないか、といった点を理解することも可能となる。

さらに大きなビジョンとして、本研究の概念を拡張することで、歌唱に限らず多くのパフォーマンス（例えば楽器演奏、ダンス、スポーツなど）において、お手本としたい「自分だけの先生」の発見、あるいは各自の上達過程の観察などを可視化によって支援するフレームワークを提案し、「レッスン支援可視化 (LessonVis)」という研究ジャンルを確立したい。多様なパフォーマンスに取り組むユーザが各自のパフォーマンスをビジュアルに理解し、着実な上達につなげることで、パフォーマンスの楽しみ・夢・充実感をさらに向上させる、という環境を提供することが本研究の未来ビジョンである。

ダンスフォーメーション作成のための候補提示型インタフェース

遠藤 輝貴 * 盧 承鐸 * 中島 一崇 * 楊 溪 * 五十嵐 健夫 *

概要. 集団でのダンス作品において複数人で構成されるフォーメーションや移動は重要な要素であり、振付師は紙に図を書いて様々な案を検討したり説明したりするが、静的な図ではその理解が難しい。解決策としてフォーメーションや移動を可視化するアニメーションを生成することが挙げられるものの、各ダンサーの動きを一つ一つ指定するには時間と手間がかかってしまう。そこで本研究では、ダンスのフォーメーションや移動経路を効率的に設計するための「候補提示型」ユーザインタフェースを提案する。具体的には、ユーザが少数のダンサーの移動経路を入力すると、本システムは残りのダンサーの移動経路をルールベースで複数予測する。これらの候補は計算された優先度の高いものから順にサムネイル形式で提示され、ユーザはその中から一つを選択することで残りのダンサーの経路をまとめて生成する。本研究では4人の参加者に対してユーザテストを実施し、提案システムがフォーメーション移動の作成に役立つことを確認した。

1 はじめに

複数人でのダンスパフォーマンスにおいて、フォーメーションやその移動は観客を惹きつけるのに重要である。ダンス作品を作る振付師はフォーメーションや移動を紙に書いて考えることが多いが、紙に書かれた静的な図だけでは各ダンサーがいつ、どのように動くかの理解が難しい。その結果、振付師は各ダンサーの動きを把握しきれず、ダンサー同士の衝突などを含んだ移動を作ってしまう可能性がある。また、紙の図だけでは作った移動を各ダンサーに説明することも難しく、全員が移動を理解するために何度も練習をする必要がある。

この問題の解決策として、フォーメーションや移動のアニメーションを作るということが挙げられる。移動をアニメーションで可視化することにより、振付師やダンサーが移動の仕方を理解することが容易になり、ダンサー同士の衝突の有無も視覚的に確認できるようになる。このような移動アニメーションの作成をするためのアプリケーション [1, 18, 19] や研究 [7, 17] はいくつか存在するが、それらは基本的にユーザが移動を一つ一つ作っていく必要がある。ChoreoGraphics [16] ではダンサーの移動経路を一括で生成できる機能があるが、生成できる移動は単純なものに限られており、さらに移動のルールをユーザが明示的に指定する必要があった。

そこで我々はダンスのフォーメーションや移動の作成を支援するための、候補提示型インタフェースを備えたシステムを提案する。本システムでは、ユーザはまず移動前後のフォーメーションを指定し、少数のダンサーの移動経路を作成する。すると本システムはヒューリスティックなルールに基づき、移動

前後の位置の対応関係とその間の移動経路を自動で予測する。ユーザの入力に曖昧性があることを考慮して、本システムは移動の候補を複数予測し、それらをサムネイル形式のリストにしてユーザに提示する。ユーザはその中から一つを選択することで、全体のフォーメーション移動をまとめて作成できる。

本システムに備えられた候補提示型のインタフェースには以下の利点がある。まず、ユーザはシステムが提示した移動の候補を選択するだけで簡単に全体の移動を作成できる。これは移動するダンサーの数が多くなった時に特に有効である。次に、本システムはユーザが描いた経路から残りのダンサーの移動経路を自動で予測する。そのため、候補生成のためにユーザが追加で操作を行う必要はなく、簡単に利用できる。また、本システムは複数の移動候補を生成して提示するため、ユーザである振付師の意図していなかった移動も提示される可能性がある。このことは振付師が今まで思いつかなかった新しい移動のアイデアを得るのに役立つ。最後に、本システムは単に移動の候補をユーザに提示するだけであり、その候補を採用することをユーザに強制しない。そのためシステムがユーザの自由な制作を妨げることはなく、ユーザは必要な時だけ提示された候補を使うことができる。

提案システムの評価のため、4人の振り付け経験者に対してユーザテストを行い、本システムを用いてフォーメーションや移動を作成してもらった。実験で作られた移動経路のうち約63%は本システムによって提示されたものであり、さらに参加者から移動のアニメーションを簡単に作成できて便利だという回答を得られたことから、本システムの有効性、有用性を確認した。

なお、本稿は国際会議で採択された内容 [2] をより詳細に記述したものである。

Copyright is held by the author(s).

* 東京大学

2 関連研究

2.1 ダンスフォーメーション作成

グループダンスのフォーメーションとその移動を作成するためのアプリケーション [1, 18, 19] や研究 [7, 17] はいくつか存在するが、これらはユーザが各ダンサーの位置や移動経路を一つ一つ指定する必要があり、特にダンサーの人数が増えると作成に手間がかかる。ChoreoGraphics [16] ではユーザが移動前後のフォーメーションを作成してその間の移動ルールを指定すると、各ダンサーの移動経路をまとめて作成できる機能がある。ただしどの移動ルールを使用するかはユーザが明示的に指定する必要があります、ルールの種類も限られていて直線的な移動しか作成できない。一方で Lee と Lee [8] はダンスのフォーメーションと移動を全て自動で生成するシステムを提案・開発しているが、このシステムではユーザが移動を直接制御することはできない。我々のシステムではユーザが自由に移動経路を作成でき、候補の提示によりその作成を支援することを目的としている。ユーザは少数のダンサーの移動経路を作成するだけでシステムが残りのダンサーの経路を自動で予測して作成するため、移動のルールを指定しなくとも簡単に移動経路を作成できる。また本システムは移動の候補を一度に複数生成するため、ユーザの意図していない移動候補も提示される可能性があり、新しい移動のアイデアを得ることにも役立つ。

また、集団の動きを制御するための研究はクラウドシミュレーションの分野でも行われている [3, 4, 6, 9, 15, 20]。これらの研究は大人数の集団の全体的な動きを制御することに主眼が置かれているが、我々が扱うダンスのフォーメーション作成では比較的少ない人数を対象とし、各ダンサーの移動を個別に、より精細に制御することを目的としている。

2.2 候補提示型インターフェース

候補提示型インターフェースとはユーザから与えられたヒントをもとにユーザの望む操作を予測して複数の候補を生成しユーザに提示するインターフェースである。これは Igarashi と Hughes [5] によって提案され、2D [13] や 3D [5] の図形描画、グラフィックデザイン [14]、家具の配置 [10] などに応用されている。これらのシステムはユーザに複数の候補をサムネイル形式で提示するため、ユーザはその中から望みの候補や自分では考えつかないような候補を簡単に見つけることができるという特長があり、本システムでもこの候補提示型の手法を採用している。

3 提案システムのインターフェース

3.1 フォーメーションの作成

本システムのインターフェースを図 1 に示す。本システムでは PowerPoint [11] のスライドのようにそ

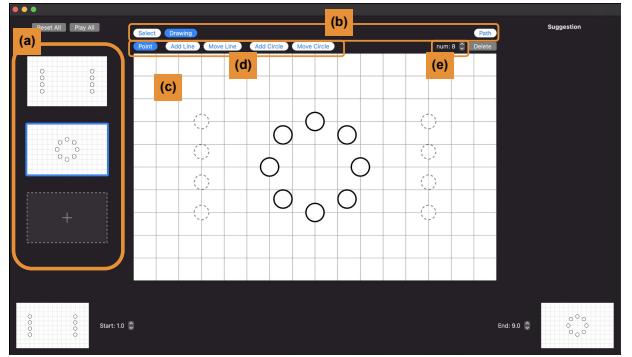


図 1. Drawing モードの UI. (a) シーンリスト. (b) モード切り替えボタン. (c) メインキャンバス. (d) ツールボタン. (e) ダンサー数変更ステッパー.

それぞれのフォーメーションを別々のシーンに分けており、ユーザは各シーンに対してダンサーの位置を配置してフォーメーションを作成する。それぞれのシーンはダンスのステージを上から見た図となっており、シーン上にある実線の丸は現在のシーンでのダンサーの位置を、点線の丸は1つ前のシーンでの位置を表す。なお、現在のシステムではダンサーの向きについては考慮していない。ユーザは左のシーンリスト（図 1a）から編集するシーンを選び、上部のボタン（図 1b）で Select, Drawing, Path の3種類のモードを切り替えて、真ん中のメインキャンバス（図 1c）でシーンを編集していく。

まず、ユーザは Drawing モードでダンサーを配置していく。ここでは、前後のシーンでのダンサーの位置の間に関係ではなく、ユーザは位置の対応関係を意識する必要はない。このような実装としたのは、振付師が作品を作る際に移動前後のフォーメーションは決まっているがその間の対応関係や移動を決めていないという場合があり、そのような場合でもシステムを利用できるようにするためにある。Drawing モードには図 1d に示すツールボタンがあり、Point ツールではクリックした位置にダンサーを一つ一つ配置できる。Add Line ツールではドラッグによって直線を引き、新たなダンサーの位置をその直線上に等間隔に配置することができる。描いた直線は Move Line ツールで位置や長さを変更でき、また図 1e に示すステッパーで直線上に配置するダンサーの数の変更も可能である。同様に、Add Circle, Move Circle ツールを使うとドラッグによって楕円を描き、その上に等間隔に新たなダンサーの位置を配置できる。また、Select モードでは既に配置したダンサーの位置をドラッグにより個別に動かして微調整ができる。

3.2 移動経路の作成

ダンサーの位置を配置して移動前後のフォーメーションを作成したら、ユーザは Path モードに切り

ダンスフォーメーション作成のための候補提示型インタフェース

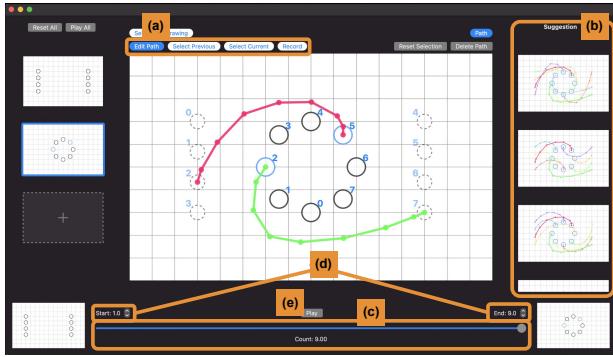


図 2. Path モードの UI. (a) ツールボタン. (b) 候補リスト. (c) シークバー. (d) 移動カウント指定用ステッパー. (e) アニメーション再生ボタン.

替えて間の移動を作成する。このモードでは点線の丸で示される1つ前の位置から実線の丸で示される現在の位置への各ダンサーの移動を編集する。Path モードの UI は図 2 のようになっており、上部にツールボタン (図 2a)，右側に候補リスト (図 2b)，下部にシークバー (図 2c) が表示される。

ユーザは以下の手順で移動経路を作成する。まず、シークバーの端にあるステッパー (図 2d) で移動の開始と終了のカウントを指定する。カウントとはダンスにおける音楽の拍を意味し、現在のシステムでは1カウントが0.5秒に対応している。ダンサーは8カウントを1つの単位として扱うことが多いため、デフォルトでは移動にかかるカウント数を8としている。ユーザは次にダンサーの移動経路を描く。図 3a のようにユーザが1つ前の位置の中の1つから現在の位置の中の1つまでドラッグすると、描いた軌跡をシステムが記録して図 3b のように経路を作成する。経路は1カウント毎に頂点を持つ折れ線で表され、識別のために経路毎に別の色がついている。また、折れ線の頂点は端の頂点に近ければその頂点に自動で重なるようになっており、これによって移動開始や終了のタイミングをズラすことができる。経路の作成後に図 2e の Play ボタンを押すと、移動のアニメーションが再生される。このアニメーションでは各ダンサーは1カウント毎に今いる頂点から次の頂点へと折れ線上を移動する。また、経路上の頂点の位置はドラッグによる微調整が可能である。

以上のようにしてユーザは移動経路を作成することができるが、この方法でダンサー全員の経路を作成するのは時間と手間がかかる。その代わりに、本システムではコンピュータに経路を予測させることで複数ダンサーの経路をまとめて作ることを可能にしている。これを行うためには、ユーザはまず移動前後で同じ数だけのダンサーの位置を選択する。システムはここで選択された位置に対して経路の予測を行う。図 2a に示す Select Previous, Select Current

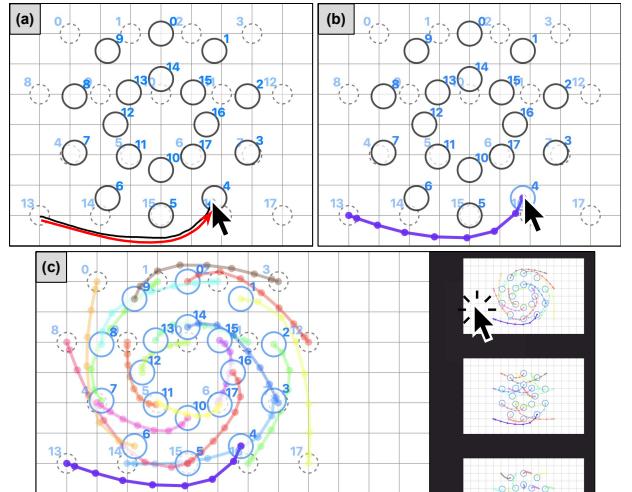


図 3. 本システムにおける移動経路作成の様子。(a) ユーザはドラッグでダンサーの移動経路を描く。(b) システムはドラッグの軌跡から経路を作成する。(c) システムによって提示された候補を選択するとその候補の移動がメインキャンバスに反映される。

ツールを使うと、それぞれ移動前、後のダンサーの位置の選択と解除をクリックで切り替えられる。選択された位置には0から始まる連続した番号がつけられるが、この番号が移動前後の位置の対応関係を表している訳ではなく、対応関係は番号順と関係なくシステムが自動で計算する。候補生成時に番号の順に位置を対応させる特別な場合もあるが、それについては第 4.2 節で述べる。なお、初期状態では配置された全てのダンサーの位置が選択されており、全員分の移動をまとめて作る場合は新たに位置を選択する必要はない。

その後、選択された位置の中の1つまたは複数に対してユーザが移動経路を描くと、システムは選択された残りの位置に対して移動経路を予測する。本システムの経路予測は2段階で行われ、まず最初に移動前後の位置の対応関係を計算し、次にダンサーが移動前の位置から対応する移動後の位置へと動く軌道を計算する。システムは移動の候補を複数生成し、図 3c のようにサムネイル形式のリストで提示する。これらの候補はアニメーションで動かすこともできるため、ユーザは各候補の移動の仕方を視覚的に確認できる。最後にユーザはこれらの候補から1つを選択することでその候補をメインキャンバスに反映し、移動の作成を完了する (図 3c)。ここで、システムによって予測された経路はユーザが描いた経路よりも色が薄くなっている、この経路が未確定であることを表している。ユーザがメインキャンバス上で経路を編集すると、システムは未確定の経路に対して予測を更新し、新たな候補を生成する。例えば図 4a ではオレンジ色の経路がユーザの描いた

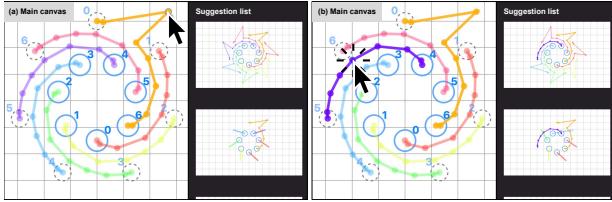


図 4. 移動経路修正の様子. (a) ユーザがオレンジの経路を編集すると右側のリスト中で残りの未確定の経路に対する予測が更新される. (b) ユーザが紫の経路を編集すると経路が確定し、右側のリストで予測が更新されなくなる.

経路、残りの経路がシステムによって予測された経路であり、ユーザがオレンジの経路を編集するとそれに合わせて右側の提示候補も更新されている。図 4b のようにユーザが未確定である紫色の経路を編集するとその経路は確定状態となり、予測の中で紫の経路は変更されなくなる。

4 候補生成アルゴリズム

4.1 アルゴリズムの概要

本システムはユーザが描いた経路を入力として受け取り、残りのダンサーの移動経路を予測してユーザに提示する。システムが経路を予測する際はまず移動前後のダンサーの位置の対応関係を計算し、次いで移動前の位置から対応する移動後の位置へと移動する軌道を計算する。経路の予測は互いに独立な5つのヒューリスティックなルールを用いて行われ、各ルールによって1つまたは複数の候補が生成される。最後にシステムは生成した候補を整理してサムネイル形式のリストとしてユーザに提示する。

4.2 ヒューリスティックルール

本システムでは R1 から R5 の5つのヒューリスティックなルールを用いて移動経路を予測している。各ルールはそれぞれ異なる移動の仕方を想定しており、それぞれの想定に従ってダンサーの移動前および移動後の位置を変換し、変換後の位置の間の距離の2乗和を最小にするマッチングを求める。その後、マッチング結果に基づいて軌道を求める。以降では経路を予測するダンサーの数を N 、番号 i の移動前の位置を p_i 、移動後の位置を q_i とし、それぞれの変換後の位置を $f(p_i)$, $g(q_i)$ とする。さらに $\sigma : \{0, \dots, N-1\} \rightarrow \{0, \dots, N-1\}$ を $p_{\sigma(i)}$ から q_i への移動を表すマッチングとする、各ルールが計算する最小マッチング m はコスト関数を $d(p, q) = |p - q|^2$ として

$$m = \arg \min_{\sigma} \sum_{i=0}^{N-1} d(f(p_{\sigma(i)}), g(q_i)) \quad (1)$$

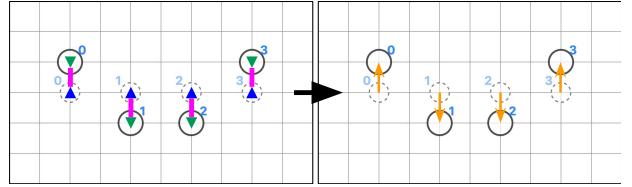


図 5. R1 による候補生成例. 青い上三角が $f(p_i)$ 、緑の下三角が $g(q_i)$ を表し、紫の線が最小マッチング m を、オレンジの矢印が作成された経路を表す.

と表せる。コスト関数に距離の2乗を使うと単純な距離を使うよりもマッチング中の最大移動距離を小さくでき、1人のダンサーのみが長い距離を動くような移動を防ぐことができる。なお、実装では最小コストマッチングの計算はハンガリアン法 [12] を用いて効率的に求めており、各候補は対話的な操作に差し支えない速度で生成できる。

R1(図5)は5つの中でも最も単純なルールであり、各ダンサーの移動コストの総和が最小になるように $f(p_i) = p_i, g(q_i) = q_i$ としてマッチングを求める。次にシステムは求めたマッチング m を用いて、各 $i = 0, \dots, N-1$ に対して $p_{m(i)}$ から q_i への経路を直線で作成する。R1だけはユーザの描いた経路を使用していないため、ユーザが経路を1つも描いていない場合でも候補を生成できる。

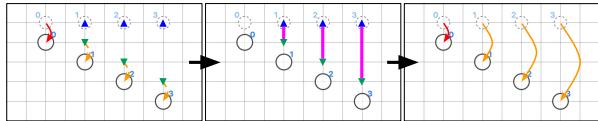
R2からR4(図6)はユーザが描いた経路を使用して候補を生成する。もしユーザが複数の経路を描いた場合、システムはそれらの経路を1つずつ選び、各経路に対して候補を生成する。このとき $f(p_i) = p_i$ であり、 $g(q_i)$ についてはユーザが描いた経路を残りの移動後の位置に対してコピーし、図6のオレンジ色の点線の矢印のように経路の終点が移動後の位置に重なるようにした時のコピーした経路の始点となる。システムはマッチングを求めた後、最初にコピーした経路を回転・拡大・移動することで、経路の始点と終点が移動前と移動後の点に一致するように変換する。以上のようにして、R2からR4はユーザが描いた経路の1つ1つに対して1つの候補を生成する。

R2(図6a)の場合はユーザが描いた経路の始点と終点をそれぞれ p_u, q_u として

$$g(q_i) = q_i - (q_u - p_u) \quad (2)$$

となる。これは直感的にはダンサー全員が同じ方向に移動することを想定しており、 $g(q_i)$ はユーザが描いた経路をそのまま平行移動して残りの移動後の位置にコピーした時の始点に対応する。R3(図6b)はダンサーの位置の中心を通り水平または垂直な軸に対して対称的な移動をすることを想定しており、ユーザが描いた経路をそのまま平行移動するか、軸を跨ぐ場合は反転してそれぞれの移動後の位置にコピーする。ダンサーの位置の中心 c は全ての p_i と

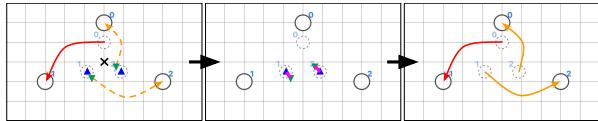
ダンスフォーメーション作成のための候補提示型インターフェース



(a) R2. 同一方向への移動を想定する.



(b) R3. 黒い太線の軸を対称とした移動を想定する.



(c) R4. 黒い×印を中心とした回転移動を想定する.

図 6. R2 から R4 による候補生成例. 赤い矢印がユーザの描いた経路を、左側のオレンジの点線の矢印がその経路のコピーを表す. その他は図 5 と同様.

q_i の平均を取ることで得られ、この点を通る軸を ℓ とし、 ℓ に関して p_u , q_u と対称な点をそれぞれ p'_u , q'_u とすると、 $g(q_i)$ は q_i と q_u が ℓ に関して同じ側のとき(3)、反対側のとき(4)となる。

$$g(q_i) = \begin{cases} q_i - (q_u - p_u) & (3) \\ q_i - (q'_u - p'_u) & (4) \end{cases}$$

R4(図 6c) はダンサーが中心 c の周りを回転するような移動を想定しており、ユーザが描いた経路を c の周りに回転、拡大して移動後の各位置にコピーする。このとき $g(q_i)$ は

$$g(q_i) = kR(p_u - c) + c \quad (5)$$

と表される。拡大係数 k と回転行列 R は以下の式から求められる。

$$k = \frac{|q_i - c|}{|q_u - c|}, \quad q_i - c = kR(q_u - c). \quad (6)$$

以上の4つのルールはダンサーの位置に振られた番号の順と関係なくマッチングを求めるので、例えば図 6c のように2番目の移動前位置から0番目の移動後位置へ移動することも可能である。

一方、R5(図7)は各位置に振られた番号の順に一列になって動くような移動を想定するので、移動前後で位置の番号は同じになる。この場合、 p_0 から q_0 へと動くダンサーが列の先頭になるので、ユーザはこの先頭のダンサーの移動経路を描く。するとシステムはこの経路の後ろに、番号順に並んだ p_i をつなげることで折れ線 C を作り、 C を残りのダンサーの移動経路とする。ただし、経路の終点だけは C 上でなく各 q_i に重なるようにする。このように、R5



図 7. R5 による候補生成例. オレンジの点線が C を表す. その他は図 5, 6 と同様.

はユーザが p_0 から q_0 への経路を描いたときのみ、候補を1つ生成する。R5の場合にはダンサーの位置のマッチングは番号によって既に決まっているが、それでもシステムは4.4節で述べる候補整理のためにマッチングコストに相当するものを求める必要がある。この場合は $g(q_i) = q_i$ とし、 $f(p_i)$ は折れ線 C 上にあって以下の式を満たすものとして求める。

$$f(p_0) = q_0, \quad (7)$$

$$|f(p_i) - f(p_{i-1})| = |q_i - q_{i-1}|. \quad (8)$$

マッチングコストは $d(f(p_i), g(q_i))$ の総和とする。

4.3 移動タイミングの予測

システムはこれらの5つのルールで生成された経路に対し、各ダンサーの移動開始や終了のタイミングも同時に予測する。ユーザは移動経路描画後、タイムカウントを表す経路上の頂点の場所を変更することで、移動の開始や終了のタイミングを設定することができる。システムは移動カウントが指定された経路をもとに、残りの経路の移動カウントを予測する。移動カウントの指定された経路が1本の場合には、そのカウントがそのまますべての経路に適用される。カウントの指定された経路が複数ある場合、移動後の位置の番号 i から移動カウント c_i を計算する一次関数 $c_i = ai + b$ を、指定されたカウントを使った最小二乗線形回帰によって求め、番号 i の位置の移動カウントを $ai + b$ に最も近い整数とする。例えば図 8a ではユーザが0番と1番のダンサーの経路を描いており、移動開始はどちらも1カウント目で、移動終了はそれぞれ2, 3カウント目としている。システムはこれをもとに残りの i 番目のダンサーに対して移動開始が1カウント目、移動終了が $(i+2)$ カウント目であると予測し、経路を生成する。

4.4 候補の整理

あまりにも多くの候補を提示されるとユーザが困惑してしまうため、システムは最後に各ルールによって生成された候補を整理する。まず候補の中で同じ移動の仕方を生成しているものを1つにまとめ、その重複度が高い順に候補を並び替える。重複度が同じ候補については、計算したマッチングコストが小さい方をより高い順位とする。その後順位が高い5つの候補をリストにし、ユーザに提示する。

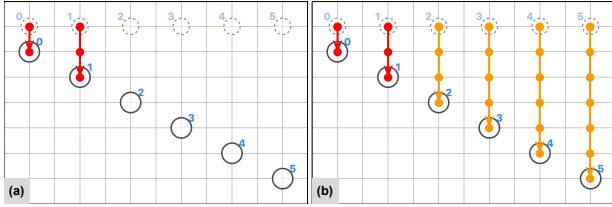


図 8. 移動タイミングの予測例. 矢印の色の意味は図 6 と同様であり、1 カウント毎に通過する点を丸い頂点で示している. (a) ユーザの描いた経路. (b) システムが (a) をもとに予測した経路.

5 ユーザテスト

5.1 実験手順

本システムの有効性と有用性を調べるために、P1からP4の4人の参加者に対してユーザテストを行った。参加者は全員が10人から30人程度で踊るダンス作品の制作経験を持ち、実験前のアンケートではP2とP4は主に紙を、P1とP3は紙とPowerPoint [11]などのアプリケーションを利用してフォーメーションや移動を考えていると答えた。本システムはSwiftで実装したアプリケーションであり、参加者に事前にアプリケーションを送って各々のパソコンから起動してもらい、使用している様子をビデオ会議ツールで録画することによって、オンラインで実験を行った。所要時間は1人当たり約1.5時間であった。

実験では参加者に対して本システムの使い方を説明し、練習用に用意したフォーメーションや移動を本システムで作ってもらったのちに、図9に示す2種類の課題の再現と参加者自身が考えた自由な作品の制作を本システムを使って行ってもらった。自由制作については最低でも10人以上のダンサーが2回以上移動するような作品を作るよう依頼した。各作品制作は参加者が満足した時点で終わりとし、実験後にインタビューをしてシステムの感想を聞いた。

なお実験時の実装ではコスト関数 d に距離の1乗と2乗との2種類を用意し、生成する候補数を倍にしていたが、候補整理の処理が複雑な割に距離の1乗を使った場合の候補はほとんど提示されなかったため、本稿では2乗のコスト関数のみを基準として説明している。

5.2 実験結果

参加者が制作したフォーメーションの移動における各経路の数を表1に示す。表の中の「手描」はユーザが描いた経路、「提示」は本システムによって提示され、そのまま採用された経路であり、「修正」はシステムに提案された経路を採用したのち、ユーザが編集を加えた経路を指す。この表より、実験で作成された全経路のうち提示経路の割合が63%を超えていたことから、本システムはユーザの望む経路をあ

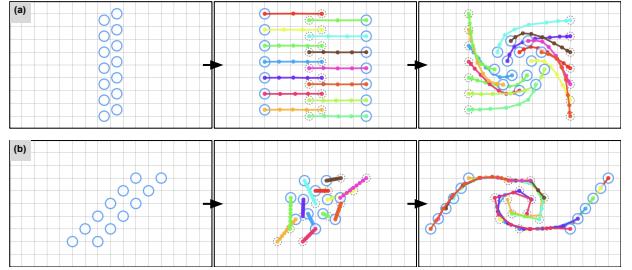


図 9. ユーザテストで用いた課題作品. (a) 課題 A. (b) 課題 B.

表 1. 参加者の制作したフォーメーションの移動における各経路の内訳.

	手描	修正	提示	合計
経路数	71	14	149	234
割合	30.34%	5.98%	63.68%	100%

る程度予測、提示できていることが分かる。

また、実験後のインタビューでは全ての参加者が本システムの候補提示機能について、移動経路を手で描く手間が省けて便利だと評価した。個別では、「候補がサムネイル形式で提示されるのが分かりやすい（P2）」、「複数の経路が提示されるので自分では思いつかないような移動経路も見つけられる（P4）」などの意見があった。一方で分かりやすさにおいては「移動前後の位置に振られた番号が移動の対応を表しているように見える（P2）」という意見があり、他の参加者もそのように誤解して操作しているように見えたため、改善が必要だと考えられる。提示された候補の精度についてはP1とP2は満足していたが、P3とP4は思い通りの結果を得られなかつたと答えた。P3はフォーメーションの右半分の移動経路を作成後にその構造を保ったまま左半分に経路をコピーするような候補の生成を望んでいた。

6 まとめと今後の課題

本研究では候補提示型インターフェースを備えた、ダンスのフォーメーションや移動を作成するためのシステムを提案した。提案システムはユーザが描いた一部のダンサーの移動経路をもとに残りのダンサーの経路を予測し、候補を生成して提示するため、振付師が移動経路を作成する手間と時間を節約することができる。現在の実装では候補の予測をルールベースで行っており、予測できる移動の種類や精度が不十分で改善の余地があると考えられる。ダンスのフォーメーションや移動のデータを大量に集めることができれば、機械学習の手法を使うことでよりユーザの意図に合った候補の生成、提示が可能になるとと考えており、今後の方針として検討している。

謝辞

本研究は JST CREST JPMJCR17A1 の助成を受けたものである。

参考文献

- [1] choreotechnology. Dance Designer Choreography Software (最終閲覧日：2021年9月12日) . <https://youtu.be/q4k4MAoaFZo>, 2009.
- [2] K. Endo, S.-T. Noh, K. Nakashima, X. Yang, and T. Igarashi. A Suggestive Interface for Designing Dance Formations. In *ACM SIGGRAPH 2021 Posters*, SIGGRAPH '21, New York, NY, USA, 2021. Association for Computing Machinery.
- [3] Q. Gu and Z. Deng. Generating Freestyle Group Formations in Agent-Based Crowd Simulations. *IEEE Computer Graphics and Applications*, 33(1):20–31, 2013.
- [4] J. Henry, H. P. H. Shum, and T. Komura. Interactive Formation Control in Complex Environments. *IEEE Transactions on Visualization and Computer Graphics*, 20(2):211–222, 2014.
- [5] 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 '01, pp. 173–181, New York, NY, USA, 2001. Association for Computing Machinery.
- [6] T. Kwon, K. H. Lee, J. Lee, and S. Takahashi. Group Motion Editing. *ACM Trans. Graph.*, 27(3):1–8, Aug. 2008.
- [7] J.-s. Lee. An Intuitive Mobile Application for Notation of Group Dance Floor Plan. In C. Stephanidis ed., *HCI International 2014 - Posters' Extended Abstracts*, pp. 349–354, Cham, 2014. Springer International Publishing.
- [8] J.-S. Lee and S.-H. Lee. Automatic path generation for group dance performance using a genetic algorithm. *Multimedia Tools and Applications*, 78(6):7517–7541, Mar. 2019.
- [9] C. D. T. Mathew, B. Benes, and D. Aliaga. Interactive Inverse Spatio-Temporal Crowd Motion Design. In *Symposium on Interactive 3D Graphics and Games*, I3D '20, New York, NY, USA, 2020. Association for Computing Machinery.
- [10] P. Merrell, E. Schkufza, Z. Li, M. Agrawala, and V. Koltun. Interactive Furniture Layout Using Interior Design Guidelines. *ACM Trans. Graph.*, 30(4), July 2011.
- [11] Microsoft. Microsoft PowerPoint (最終閲覧日：2021年9月12日) . <https://www.microsoft.com/ja-jp/microsoft-365/powerpoint>.
- [12] J. Munkres. Algorithms for the Assignment and Transportation Problems. *Journal of the Society for Industrial and Applied Mathematics*, 5(1):32–38, 1957.
- [13] S. Murugappan, S. Sellamani, and K. Ramani. Towards Beautification of Freehand Sketches Using Suggestions. In *Proceedings of the 6th Eurographics Symposium on Sketch-Based Interfaces and Modeling*, SBIM '09, pp. 69–76, New York, NY, USA, 2009. Association for Computing Machinery.
- [14] P. O'Donovan, A. Agarwala, and A. Hertzmann. DesignScape: Design with Interactive Layout Suggestions. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pp. 1221–1224, New York, NY, USA, 2015. Association for Computing Machinery.
- [15] S. Patil, J. van den Berg, S. Curtis, M. C. Lin, and D. Manocha. Directing Crowd Simulations Using Navigation Fields. *IEEE Transactions on Visualization and Computer Graphics*, 17(2):244–254, Feb. 2011.
- [16] A. Schulz, W. Matusik, and L. Velho. Choreo-Graphics: An Authoring Tool for Dance Shows. *Journal of Graphics Tools*, 17(4):159–176, 2013.
- [17] A. Soga. and I. Yoshida. SIMULATION SYSTEM FOR DANCE GROUPS USING A GAMEPAD. In *Proceedings of the International Conference on Computer Graphics Theory and Applications and International Conference on Information Visualization Theory and Applications - Volume 1: GRAPP, (VISIGRAPP 2012)*, pp. 365–368. INSTICC, SciTePress, 2012.
- [18] StageKeep. StageKeep (最終閲覧日：2021年9月12日) . <https://stagekeep.com>.
- [19] UGOKU. MO-SIM (最終閲覧日：2021年9月12日) . <https://mo-sim.com>.
- [20] B. Ulicny, P. d. H. Ciechomski, and D. Thalmann. Crowdbrush: Interactive Authoring of Real-Time Crowd Scenes. In *Proceedings of the 2004 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, SCA '04, pp. 243–252, Goslar, DEU, 2004. Eurographics Association.

歌唱時に高音の発声ができるようになる VR 映像の提案と評価

坂名純太 * 土田修平 * 寺田 努 * 塚本昌彦 *

概要. 歌唱には、音程やリズムなど様々な要素が関わっており、発声もその一つである。良い発声を行うためには、音程に合った姿勢を保ちつつ、特定の筋肉に力を入れて他の筋肉はリラックスさせることができるとなる。しかし、発声する音の種類によって適切な姿勢は変わるために、姿勢への意識が強くなると発声への意識が薄れてしまい効率が下がる恐れがある。そのため、専門知識をもたない人にとって歌唱時の発声練習は難しく、これまでの歌唱指導では「額から声を出すように」などのイメージで歌い方を伝えるような工夫が行われている。そこで本研究では、視覚刺激によって発声者の姿勢やメンタル状態を誘導すれば効果的な発声練習を実現できると考え、VR(Virtual Reality)映像を用いた視覚刺激を活用し、歌唱時の発声を暗黙的に支援するシステムの構築を試みる。発声者の高音の発声と発声時の姿勢に注目し、高音の発声がしやすくなると考えられる映像を作成し、映像の内容によってユーザの高音の発声のしやすさ及び発声時の姿勢に生じる変化について調査を行った。

1 はじめに

歌唱には、音程やリズムなど様々な要素が関わっており、発声の仕方もその一つである。良い発声は、声に柔軟性を与え、声帯を効率的に動かすことで少ない力で大きな声を発することができ、歌唱中の喉の負担を軽減できる。フースラーら [1] は、声楽的な発声法について生理学的・解剖学的知見を取り入れ研究し、歌唱中的人体の構造を細部にわたって説明している。その中で、発声器官の能力を十分に發揮するためには、発声器官のすべてが的確な神経支配を受け、またその器官の筋肉が敏速で伸縮自在な張力を最大限に獲得する必要があると述べている。このことから、正しい発声には特定の筋肉に力を入れて、その他の筋肉はリラックスしていることが同時に求められるなどいくつもの発声に関わる器官を適切に動かすことが求められる。

しかし、発声の際に重要な筋肉は、普段の生活で使う機会が少ない箇所が多く、単に意識するだけでは正しく筋肉を使いこなすのは困難であり、専門的知識のない人には正しく使っているかの判断は難しい。そのため、発声に重要な筋肉を正しく使うことを補助する目的で、特定の姿勢や特定のイメージを持って練習する手法がよく用いられている。例えば高音を出すためには「額から声を出すイメージで」や「自分が階段を降りているようなイメージで」といった説明がされることがある。ただし、求められる姿勢やイメージは発声パターンごとに異なるため、歌唱時に姿勢やイメージをその都度切り替える必要があり、発声者への負荷が大きい。また、発声者自身の頭の中でリアルなイメージを想起しながらの練習では、発声への意識が薄れてしまい効率が下がる

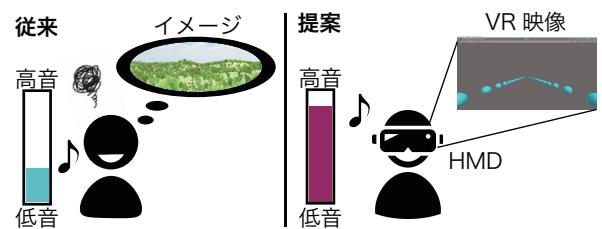


図 1: システム概念

恐れがある。

そこで本研究では、発声者が意識せざとも発声に重要な筋肉を正しく使うことを促せれば、より効果的な発声練習を実現できると考え、VR(Virtual Reality)映像によって視覚刺激を与えることで姿勢等に影響を与えられる [13] ことに着目し、VR 映像を見ることで意識下でユーザの発声時における身体機能に影響を与え、良い発声を促す手法を提案する(図 1)。本稿では、発声練習時に見る VR 映像によって重心位置および発声することのできる音程の上限に差異があるかについて調査した。このような映像効果と歌唱能力の関係が明らかになると、高音が出やすいカラオケ背景が作成できるなどさまざまな応用が可能になると考えられる。

2 関連研究

2.1 声楽的な観点からの発声研究

声楽的な観点から見て、歌唱時には発声器官の各役割を理解し、その多くをコントロールすることが求められる。石野 [2] によれば、地声や裏声といった声区は筋肉や呼吸法によって調整でき、声量、母音等によっても声区は変動すると述べている。歌唱

* 神戸大学大学院工学研究科

者は声区を意識しなくなることが理想であるが、そのためにも筋肉のメカニズムを知ることは必須であると述べている。また、理想的な発声は身体全体の筋肉の自然な連携を前提としており、その際に身体の様々な部位は無意識的に連動するとされ、意識した身体動作ではぎこちない発声となってしまう恐れがあると石野は述べている。しかし、現在行われている発声のための姿勢や各器官を適切に動かす練習は手本を見たり聞いたりしてそれを真似るといった手法が用いられており、練習者は意識的に身体動作を行わなければならない。野口[3]は意識と無意識という観点から、良い発声に大切なのは発声に関わる筋肉が良好に運動することであるが、発声者からすればそのような客観的な視点を意識すればするほど、筋肉は自然な運動性を失っていくと述べている。つまり正しい発声には、身体の発声に関わる諸器官が無意識のうちに適切に機能することが必要である。本研究では、無意識のうちに発声に関わる器官が正しく機能することを実現するための手法について提案する。

2.2 発声支援手法

歌唱時の発声支援システムとして、発声者の音声特徴量をフィードバックする方法が多く用いられている。斎藤ら[4]は、歌声に聴感的な響きを与える要因と考えられている2.4~3.2kHz付近の周波数帯域に着目し、歌唱指導前後でその周波数帯域の歌唱データがどう変化しているか比較することで習熟度を検討できることを示唆した。また、片岡ら[5]は強弱のある発声について、声楽経験の異なる被験者から音声データを取得し分析することで、習熟度をフィードバックするシステムを提案している。さらに、中野ら[6]は音声の音高を軌跡として可視化することで歌唱への理解を深めさせ、過去のデータやお手本となる歌唱データをゴースト表示することで練習効率を向上させる手法を提案している。

しかし、これらの手法はその時点での発声者の歌声を評価するものであり、直接的に発声練習の効率を上げるようなシステムではない。また、これらのデータをフィードバックしたとしても、有効に利用するためには専門的な知識が必要となり、練習者はこれらのデータから得たアドバイスを意識しながらの発声を行わなければならず、発声そのもののへの意識が薄れてしまう恐れがある。本研究では、声楽的な知識を必要とせずに歌唱力の向上が可能となるシステムの構築を目的とする。

2.3 視覚情報が与える影響

VR 映像以外による視覚刺激

杣田ら[7]は拡張現実技術を用いた周辺視野が人間の重心知覚に与える影響を調べた。この実験では、被験者に実物のアタッシュケースを持たせ、その位

置に合わせて拡張現実を用いてバーチャルなアタッシュケースを重ねて表示し、形状・大きさを変化させることで重心がどの位置にあると感じられるか調査を行った。その結果、実際のアタッシュケースに比べ取っ手の位置がずれている表示をした場合、ユーザはずらした方向に重心が傾いていると感じる傾向が見られた。岡野ら[8]が行った実験では、LEDを用いた周辺視野ディスプレイを被験者に装着させ、ルームランナー上でランニングをさせたところ、LEDの光がランナーに向ってくるよう点滅させた際の体感速度は被験者の走行時の速度よりも速くなる傾向が見られた。また、吉川ら[9]は断面が半円状のレンズを無数に並べられたシートを用いた歩行誘導の実験を行った。レンズに少しずつ絵柄をずらした画像を格納することで、レンズに対して視線を任意方向に動かすアニメーションを呈示でき、被験者は絵柄をずらした方向に誘導されて歩行した。

VR 映像による視覚刺激が与える影響

VR 技術を用いて、身体機能の向上や特定の運動機能の促進を狙うシステム構築に関する研究は多数行われている。小貫ら[10]は生体センシングを用いた VR ゲームの利用が高齢者の運動能力にどう影響するか調査した。その結果、VR ゲームを用いることで被験者の歩行能力やバランス能力が有意に向上し、アンケート調査でも満足度が高かった。遠藤ら[11]は VR による歩行映像、モニターによる歩行映像、静止立位をランダムで被験者に提示した後に歩行させ、その際の重心動搖を計測したところ、前後方向の重心の移動について、VR 映像とモニター映像を提示した時の移動量は静止時の移動量に比べ有意に大きかった。石井ら[12]は HMD の情報処理を用いて人間の焦点に影響を与え、ユーザの歩行時の進行方向を誘導することを可能にした。このように VR 技術を用いて運動機能を向上させることや姿勢や重心に影響を与えることが可能であり、本研究ではこれを歌唱時の発声に応用することで練習効率が上がるのではないかと考えた。

3 提案システム

我々は、ユーザーが良い発声のために意識すべきことを無意識のうちにに行わせることで、自然で良い発声が可能になると考えた。そこで本研究では、VR 映像による視覚刺激により身体や知覚を制御し、重心誘導を含めた意識すべきことを意識下で行わせ、ユーザの発声を支援する手法を提案する。提案手法では、被験者に VR 映像を提示可能な頭部装着型ディスプレイ (HMD: Head Mounted Display) を装着させ、作成した VR 映像を見せながら歌唱を行わせる。本論文で提示する映像は主に移動感覚を引き起こすようなものを用いた。没入感のある移動感覚提示により、高音の発声に適しているとされるつま先

重心(前方重心)の姿勢を無意識に促すことを意図している。つまり、本研究では良い発声の例として「高い音を発声できること」、意図する映像視聴効果を「移動感覚の付与による体勢制御」とした。VR 映像の作成には Unity を使用し、VR 提示可能な HMD には HTC 社の Vive(視野角: 110°, 重量: 640g, 以下 HMD)を用いた。

4 評価実験

本実験では、VR 映像を見ながら発声練習を行う際に、高音の発声と姿勢にあらわれる影響について調査する。

4.1 映像の選定

発声に適した姿勢は、発する声の音程や強弱などによってわずかに異なる。高音を発声する際はつま先に体重をかける姿勢が適しているとされる[1]。この知見をもとに、高音の発声に適した状態になるよう促す映像を調査するために 4 種類の映像を用意した。具体的には、高音の発声に適しているとされるつま先重心、前傾姿勢をとるよう促す効果が期待できる映像として、相対的にユーザが上または前に移動しているかのような感覚を与える映像と、その逆の感覚を与える映像を計 4 種類用意した。具体的には、視界いっぱいに配置された球形オブジェクトが鉛直下向きにスクロールされ、相対的にユーザが上昇しているかのような感覚を覚えることを狙った映像(下方向)、球形オブジェクトが鉛直上向きにスクロールされ、相対的にユーザが下降しているかのような感覚を覚えることを狙った映像(上方向)、ユーザの両脇に球形オブジェクトが線上に配置されており、それらがユーザの前方から後方へと向かってくるように移動することで、相対的にユーザが前進している感覚を覚えることを狙った映像(後方向)、後方向の映像とは逆方向にオブジェクトが前方に移動し、ユーザが後退している感覚を覚えることを狙った映像(前方向)である。ここで、作成した映像の種類を表 1 に、上下方向のイメージ画像を図 2 に、前後方向のイメージ画像を図 3 に示す。なお、球形オブジェクトは半径 50cm の大きさに見えるように設定し、イメージ画像中の矢印は球形オブジェクトの移動方向を示している。また、上方向と下方向の映像では被験者とスクロールする球形オブジェクトとの距離は 11m となっており、後方向と前方向ではユーザの左右に配置された球形オブジェクトとの距離は 5m となっている。これはユーザがオブジェクトの接近により回避行動をとるような状況が起きないよう設定されている。

4.2 計測項目

映像が表示されてから各パートが終了するまでの間、Wii Balance Board[15]を用いて被験者の足圧

表 1: 作成した映像

映像名	説明と想定される効果
上方向	オブジェクトが上方向に移動 後傾姿勢となり高音が出なくなる
下方向	オブジェクトが下方向に移動 前傾姿勢となり高音が出せる
後方向	オブジェクトが向かってくる 前傾姿勢となり高音が出せる
前方向	オブジェクトが遠ざかってゆく 後傾姿勢となり高音が出なくなる

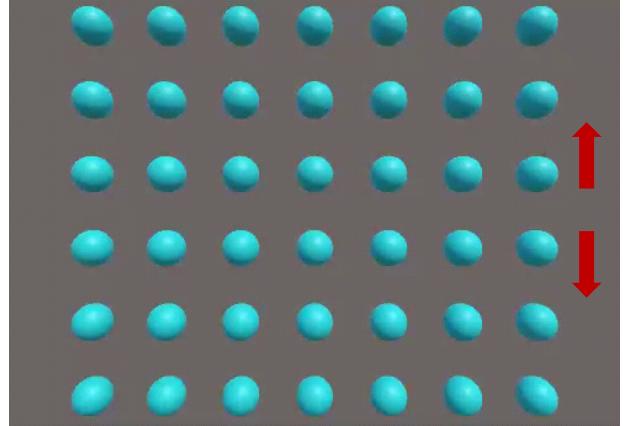


図 2: 上下方向のイメージ画像

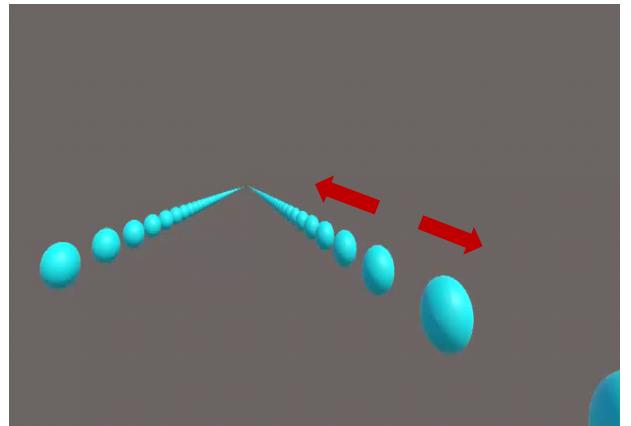


図 3: 前後方向のイメージ画像

中心(COP: Center of Pressure)を 100fps で取得する。COP は静止立位姿勢の重心動搖の測定において、重心位置として代替可能である[14]。Wii Balance Board を用いた足圧中心の測定には、公開されている Wii Balance Board Stabilometry System Ver 2.1[16]を使用する。

被験者に装着した無線ピンマイク (SYNCO-G1 2.4 GHz ワイヤレスオーディオ)[17]から、実験中

に被験者が発声した音声を WAV ファイル形式で保存する。その後、被験者の音声データから F0(基本周波数)を求め、被験者が地声で発声することできた音程を音声データと F0 から判定する。また、実験中の被験者の様子をビデオカメラで撮影する。計測用ソフトウェアは Python を用いて作成した。

4.3 実験手順

計測時のシステム構成を図 4 に示す。被験者は実験環境に入室した後、発声練習に用いるピアノ音源を聞く。その後、Wii Balance Board の上に乗り、HMD とイヤホン、ピンマイクを装着する。HMD は PC に接続されており、被験者の動作のトラッキングのために HMD の位置を計測するためのベースステーションを 1 台被験者の正面、2.5m の距離に設置する。また Wii Balance Board 上では、被験者にとって自然な直立姿勢を取ってもらう。被験者に取り付けたデバイスと実験をしている様子を図 5 に示す。

被験者は HMD を額に載せて映像を見ない状態と、オブジェクトが移動する 4 種類の映像を見ながらの発声練習をランダムな順番で行う。なお、発声練習にはドミソミドの音を 1 セットとしたピアノの音源を使用し、音程は mid2C(262Hz) から始める。被験者はピアノの音源を 1 セット聞いた後に、同じ音程のドミソミドを発声する。ピアノの音程は段階的に高くなり、被験者が明らかに発声出来ていない、もしくは裏声を用いなければ発声できない音階まで達した場合はそのパートの発声練習は終了とする。この判断は本論文の第一著者が行った。発声練習 1 回あたりにかかる時間は約 3 分であり、各発声練習の間には 5 分間の休憩を取る。被験者は 20 代男性 9 名(声楽経験、VR 経験なし)である。最後に、口頭もしくはチャットツールを用いてシステムに対するコメントを回答してもらった。

また、今回作成した映像を用いた発声練習に関して、高音の発声に与える影響がその後の本システムを用いない発声においても影響を与えるかどうかを調査するため、1 回目の実験から 1 週間後の同時刻に被験者 1 人当たり 3 回、同じ内容の発声練習をデバイスを装着していない状態で行わせた。なお、この実験は神戸大学大学院工学研究科の研究倫理審査委員会の承認を得て行った(受付番号: 03 - 12)。

4.4 結果と考察

全被験者における、練習パターンごとの足圧中心の平均座標(左右方向、前後方向)を箱ひげ図で表したものを作成したものをそれぞれ図 6, 7 に示す。これは Wii Balance Board の中心を 0 とした時の値であり、単位はいずれも mm である。また、左右方向では右が正、左が負の値であり、前後方向では前が正、後ろが負の値となっている。左右方向、前後方向におい

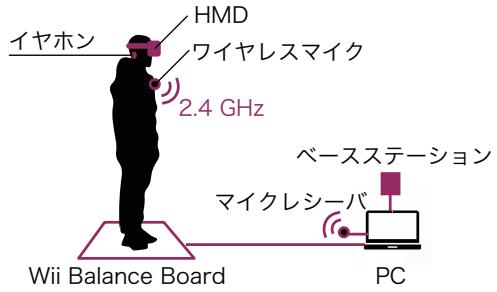


図 4: 計測時のシステム構成

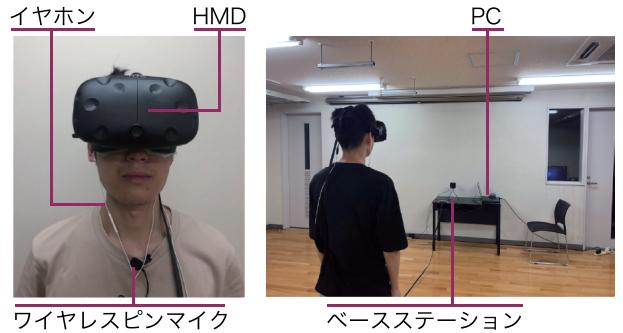


図 5: 実験環境

てそれぞれ一要因分散分析を行ったところ、提示する映像間に有意な差は見られなかった($p > .10$)。左右方向について、映像なしと前方向での発声練習では被験者間で足圧中心のばらつきが小さく、下方向で特にばらつきが大きい結果となった。また、前後方向については、後方向の映像を用いた発声練習中の足圧中心のばらつきが小さく、残りの 4 条件ではばらつきがみられた。この値は、実験における体勢の変化を計測するため行ったものであるが、実験中は VR 空間を把握するため首を振る被験者が多く見られ、その際の重心変化が大きかったため今回の実験では姿勢の違いを表すような値が計測できなかった可能性がある。

また、地声で発声することのできた音声の最大周波数に対して、提示映像ごとに全被験者の平均値を求めた結果を図 8 に示す。なお、被験者の発声した地声と裏声の判定は、声楽経験者に音声データを聞いてもらい、主観的に判断した。提示映像について一要因分散分析を行った結果、有意差が見られた($F(4, 32) = 5.02, p < .05$)。また Holm 法による多重比較を行なったところ、後方向と映像なし、後方向と下方向、後方向と前方向においてそれぞれ有意差が見られた($p < .05$)。このことから、オブジェクトが前方からこちらへ向かってくる映像を見ながらの発声では、仮説通り被験者が高音の発声に適した姿勢となり、より高い音の発声が出来るよう

歌唱時に高音の発声ができるようになる VR 映像の提案と評価

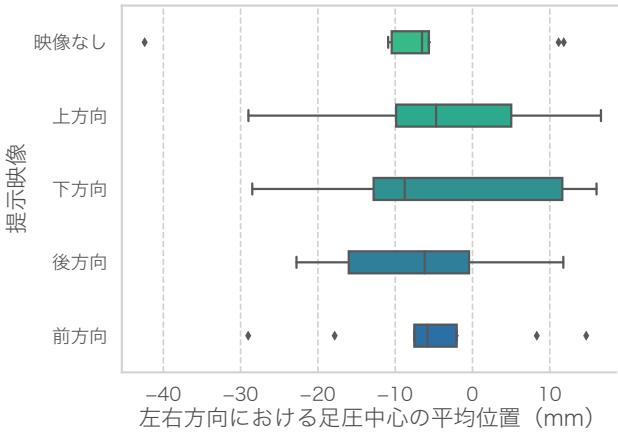


図 6: 各練習パターンの足圧中心の平均位置 (左右方向)

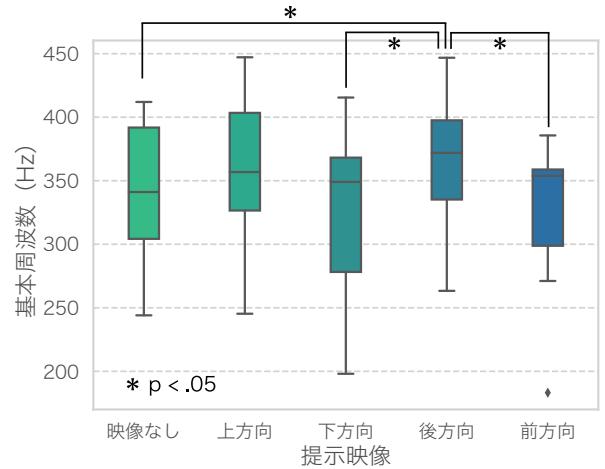


図 8: 各練習パターンにおける基本周波数の平均値

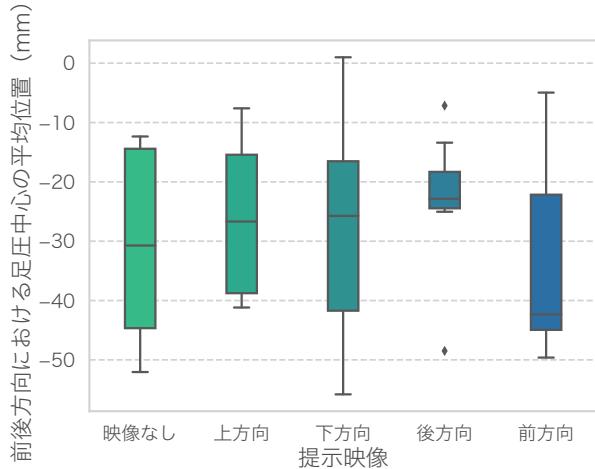


図 7: 各練習パターンの足圧中心の平均位置 (前後方向)

なった可能性が示唆された。また前方向の映像よりも後方向の映像で高音の発声ができていたことから、高音の発声にはユーザが前方へ移動している感覚を覚える映像が好ましいと考えられる。以上の結果から、VR 映像を見ることにより高音の発声を促すことができるという可能性が示唆されたが、今回用いた手法では、正しい発声に関わる要因のうち重心移動以外に影響を与える効果が存在する可能性が高いと考えられる。今後は他の要因に焦点を当て、提案手法により発声が可能となる原理を解明することが課題である。

4.5 映像を用いた練習効果の持続性

VR 映像を用いた発声練習の効果の持続性を評価するための実験を追加で行った。VR 映像を見なが

らの発声練習を行ってから 1 週間後、同被験者には HMD を装着せずに、同じ内容の発声練習を 3 回行ってもらい、音声データから最も高い周波数を取得し VR 映像を見ながらの発声練習での最高周波数と比較した。各被験者ごとの、1 回目の実験と 2 回目の実験での最高基本周波数を図 9 に示す。2 回目の発声練習にて、1 回目の発声練習よりも高い音を発声できた、もしくは 1 回目の発声練習時と同程度の音程が発声できた被験者は 4 名 (被験者 E,F,G,I) であった。長期的な効果が期待できると判断できた被験者数が全被験者数の半数に満たないことから、本システムは繰り返し練習を行うことで長期的な効果が見込める、または映像を見ているときにのみ効果が現れるものであると考えられる。

5 まとめ

本稿では、高音の発声に適した映像見せることで、発声時に望ましい身体状態を無意識のうちにに行わせるシステムの設計とその評価を行った。今回の実験では、高音に着目し低重心やつま先に重心を持ってゆく姿勢を誘導する効果が期待できる映像と、逆の効果が想定できる物を用意し、それぞれ映像なしの発声練習と比べどのような違いがあるのか調査した。その結果、オブジェクトが前方からこちらへ向かってくる映像を見ながらの発声では、より高音の発声が出来るようになったという結果が得られた。また、本システムを用いた練習効果の持続性については、

今後の課題として、今回用いた映像よりも重心の移動に影響を与え発声を支援する映像の作成と、実際の歌唱に本システムを導入し、音程に合わせて適切な効果のある映像を切り替えて表示させることが挙げられる。そのためには、提案手法を用いた練習によって発声の質が向上した場合に、発声に関わるどの要因に影響を与えたのか解明することも重要で

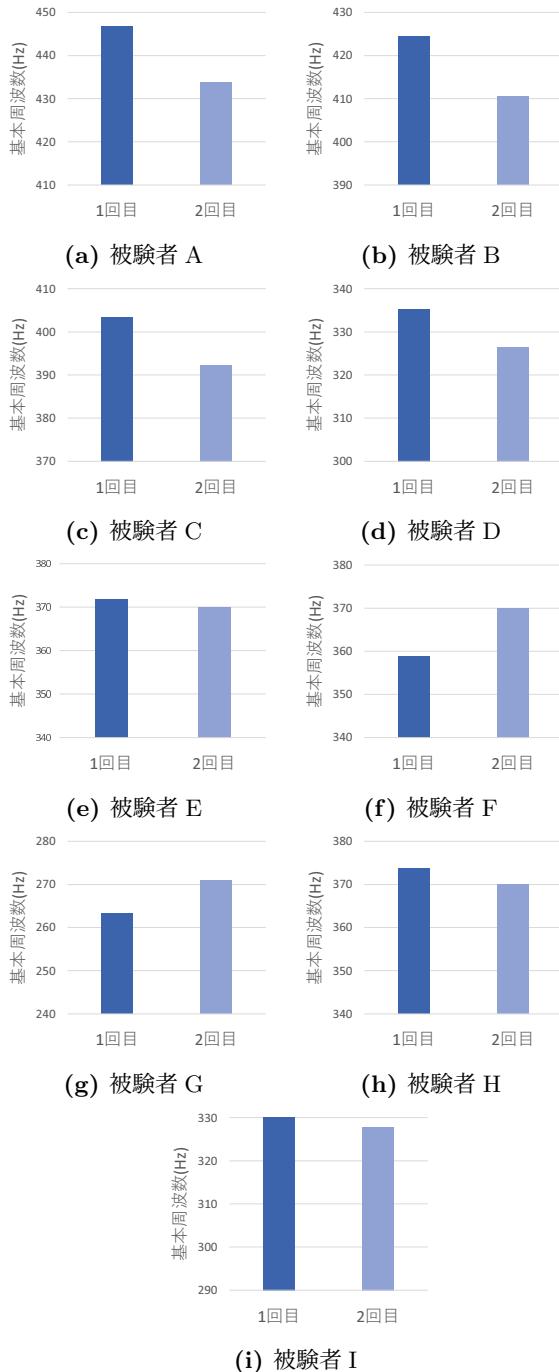


図 9: 各被験者が発声出来た音程上限の比較

ある。また、1人の被験者から上下方向のオブジェクト移動により浮遊感や下降感覚を覚え少し気分が悪くなつたという感想があった。この被験者は軽度の高所恐怖症を患つており、これらの感覚に抵抗を感じたと考えられる。2章であげたLEDによる視覚刺激に比べ、VR映像による手法ではVR酔いや、ある状況や感覚に不快感や恐怖感を覚えるユーザを考える。

慮した映像の作成も課題としてあげられる。さらに、実験後に被験者から得た感想としては、「オブジェクトに合わせて視線が移動してしまつ」といったものがあり、作成した映像が狙っている重心移動の妨げになつていたということが考えられる。この対策としては、VR空間内にポインターを表示させ、被験者の視線を一点に固定するなどの処置を行つて、自然な重心移動を促すことが有効であると考えられる。加えて、さらに多様な映像、条件を用いてさらに高音が発声しやすくなる映像や、高音以外の歌唱の要素を向上させる映像に必要な要因を調査する。また、複数の映像の順番と効果の関係を調べることや、各映像での長期的な練習による上達の程度に違いがあるのかも明らかにし、最終的にはカラオケにおける歌唱時の映像などにこれらの映像を組み込み、歌唱者にとって最も良い歌唱が可能となる環境の構築を目指す。

謝辞

本研究の一部は、JST CREST (JPMJCR18A3) 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] フレデリック・フースラー, イブォンヌ・ロッド=マーリング: うたうこと:発声器官の肉体的特質, 音楽之友社 (Jun. 1987).
- [2] 石野健二: 声楽発声のメカニズム, 宇都宮大学教育学部紀要. 第1部, No. 1, pp. 105–119 (Mar. 2015).
- [3] 野口雅史: 意識と無意識からみた歌唱の発声法, 新潟青陵大学短期大学部研究報告, Vol. 41, No. 41, pp. 151–161 (May 2011).
- [4] 斎藤毅, 後藤真孝: 歌唱指導による歌声中の音響特徴の変化: 歌唱ホルマントとF0動的変動に着目した音響分析, 日本音響学会講演論文集, pp. 457–458 (Sep. 2008).
- [5] 片岡靖景, 伊藤一典, 池田操: 歌唱支援システム構築のための歌声の分析と評価, 情報処理学会研究報告音楽情報科学 (MUS), Vol. 1998, No. 74, pp. 23–30 (Aug. 1998).
- [6] 中野倫靖, 後藤真孝, 平賀譲: MiruSinger: 歌を「歌って/聴いて/描いて」見る歌唱力向上支援インターフェース, 情報処理学会 インタラクション2007論文集, Vol. 2007, pp. 195–196 (Aug. 2007).
- [7] 杣田明弘, 溝口晃太, 木村朝子: 複合現実型視覚刺激が重心知覚に与える影響, 情報処理学会全国大会講演文集, Vol. 71, No. 4, pp. 115–116 (Mar. 2009).
- [8] 岡野裕, 雜賀慶彦, 橋本悠希: 速度感覚増強のための周辺視野への刺激提示手法の検討, 情報処理学会研究報告ヒューマンコンピュータインタラクション (HCI), Vol. 2008-HCI-127, No. 11, pp. 145–150 (Feb. 2008).

歌唱時に高音の発声ができるようになる VR 映像の提案と評価

- [9] 吉川博美, 蜂須拓, 福島政期: ベクション場による歩行誘導手法の提案, 情報処理学会 インタラクション, Vol. 2011, pp. 457–460 (Mar. 2011).
- [10] 小貫睦巳, 有田元英, 井上悦治: 生体センシング技術を使った仮想現実によるゲームが高齢者の運動機能に及ぼす影響について, 理学療法科学, Vol. 30, No. 6, pp. 811–815 (June 2015).
- [11] 遠藤佳章, 鈴木暁, 糸数昌史: ヘッドマウントディスプレイを用いた Virtual Reality 映像とモニター映像が立位重心動搖に及ぼす影響の違い, 理学療法科学, Vol. 33, No. 3, pp. 457–460 (Jan. 2018).
- [12] A. Ishii, I. Suzuki, S. Sakamoto and K. Kanai: Optical Marionette: Graphical Manipulation of Human's Walking Direction, Proc. of the 29th Annual Symposium on User Interface Software and Technology, pp. 705–716 (Oct. 2016).
- [13] 市川真澄, 渡邊悟: 直立姿勢に対する視覚情報の影響 ((特集) 姿勢制御と感覚情報), バイオメカニズム学会誌, Vol. 15, No. 2, pp. 59–64 (Feb. 1991).
- [14] 藤原勝夫, 池上晴夫: 足圧中心位置と立位姿勢の安定性との関係について, 体育学研究, Vol. 26, No. 2, pp. 137–147 (Aug. 1981).
- [15] Wii Balance Board, <https://www.nintendo.co.jp/wii/rfnj/balance/index.html>.
- [16] バランス Wii ボードの重心動搖計ソフトウェア, https://researchmap.jp/blogs/blog_entries/view/112495/5a9c28ee606c50172546b83b5ea9939a?frame_id=513522.
- [17] SYNC-G1(A1)-2.4 GHz ワイヤレスオーディオ, <https://www.syncaudio.com/products/sy-g1-a1-wair>.

Human-in-the-Loop 型適応によるインタラクティブな音楽的拍節解析

山本 和彦 *

概要. 音楽情報処理において拍（ビート）の抽出は最も基礎的で重要なタスクのひとつである。しかし、この拍の捉え方や解釈は個人によって大きく異なり唯一の正解というものが存在しない。そのため、あらかじめ完璧なアルゴリズムを設計しておくことは非常に困難である。本稿ではこの問題を解決するため、Human-in-the-Loop 型のビートトラッキングインターフェースを提案する。本システムではユーザは解析結果の気に入らない箇所をインタラクティブに修正するが、すべての修正箇所を手動で修正する必要はなく、最初に気付いたごく一部分のみを修正してシステムにフィードバックする。システムはそれをもとに内部の計算モデルをユーザ適応させ、ユーザが直接編集していない残りの部分もユーザの意図に沿うように自動修正する。このユーザとシステムのインタラクションを繰り返すことによってユーザは効率的に所望の結果を得ることができる。これを実現するため、我々は適応型実行時自己注意機構を導入する。適応型実行時自己注意機構のアルゴリズムはユーザインターフェースと密接に統合されており、ユーザの局所的な修正の影響を楽曲全体へ効率的に伝搬させることを可能にする。本稿では提案手法が所望の解析結果を得るためにユーザの労力を劇的に削減することをシミュレーション実験とユーザ実験の両方で示す。

1 はじめに

音楽において拍（ビート）は最も重要な要素のひとつである。多くの音楽情報処理システムが最初にビート解析をしたのちそれを後段の処理のための一単位として採用している。しかし、機械学習の発展によって音楽情報処理アルゴリズムの性能は劇的に向上したにもかかわらず、完璧なシステムを実現することは困難であり、どうしてもエラーは発生してしまう [1]。さらに、ビートの捉え方や解釈は個人によって異なる。たとえユーザの知覚する拍が楽譜上では作曲者が制作時に想定したものと一致していたとしても、その正確なタイミングは人によってわずかに前や後ろにずれていることが多い。同一のフレーズであってもどこを拍とすべきかは状況によってどのようにでも変化するということを意図的に示した音楽も多く存在する [2]。このため拍の解析には唯一の正解というものが存在せず、一意の入力から状況に応じて異なる結果を出力することが求められる。このような問題に事前学習された機械学習システムだけで対応することは原理的に困難である。

本稿ではこの問題を解決するため、Human-in-the-Loop のアプローチによって特定のユーザと楽曲に適応した結果を出力することのできるインタラクティブなビートトラッキングインターフェースを提案する（図 1）。本システムでは、ユーザはシステムの一時的な出力結果に対して気に入らない部分を修正しフィードバックする。システムはそのフィードバック情報を利用して内部のニューラルネットワークモデル（DNN）をこのユーザに対して適応し、再

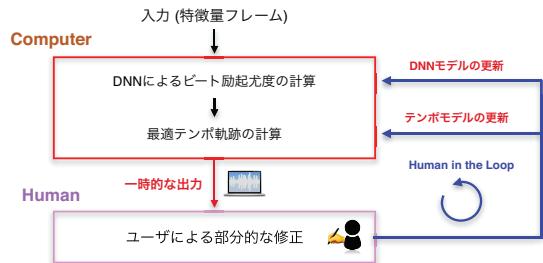


図 1. システム概要。ユーザはシステムの一時的な出力に対して一部分だけを修正してフィードバックする。システムはそれをもとに内部モデルを更新しより所望の結果を出力できるようになっていく。

度出力を更新する。このユーザとシステムのインタラクションを反復することで、システムはよりユーザにとって所望の結果を出力できるようになっていく。ここで重要な点はユーザはすべての修正箇所を手動で修正する必要はなく、ごく少数の最初に気付いた一部分のみの修正をおこなうだけでよいということである。システムはこの少数のユーザ修正から、残りの直接編集されていない箇所をユーザの意図に沿うように自動で修正できるように内部モデルを適応させる。これにより、ユーザが所望の解析結果を得るための労力を劇的に削減ができる。

このオンラインのユーザ適応を実現するため、本稿では新しいDNNモデルである適応型実行時自己注意機構（Adaptive Runtime Self-Attention, ARSA）を導入する [3]。ARSAの内部パラメータはインタラクティブなユーザインターフェースを密接に統合されている（5.3章）。ARSAは一般的な注意機構 [4] の拡張であり、他の注意機構と同様にある時点の出力を得るために入力系列全体の情報を集約する。ARSA

* ヤマハ株式会社

と従来の注意機構との大きな違いとして、ARSA はまったく事前学習をおこなわない。ARSA は実行時にのみ、別に用意された事前学習済み DNN の中に組み込まれ、実行時ユーザ適応のためだけに利用される。つまり、我々の DNN モデルは事前学習時と実行時でまったく異なる構造をとる。

我々の DNN モデルでは、事前学習時には楽曲の局所的な情報のみをつかってそれぞれの時刻でのビート励起尤度（その時刻がどのくらいビートらしいか）を学習する。一方、実行時には ARSA を組み込むことによってある時点の出力を得るために楽曲全体の情報を集約して計算をおこないユーザ適応をおこなう。この局所的な事前学習は学習時の計算コストを削減し、かつ時系列全体を利用した実行時適応はユーザの局所的なエラー修正の影響を楽曲全体へ効率的に伝搬させることができるという利点がある。この局所学習-全体適応の戦略は、ユーザが現実的に修正可能な局所的で少量のフィードバック情報量だけでは楽曲全体のユーザ適応をおこなうには不十分である、という仮定にもとづいている。本稿ではシミュレーション環境上での性能評価に加えて、音楽編集ソフトウェアに普段から慣れ親しんでいる被験者によるユーザスタディをおこない、提案手法がユーザの所望する解析結果を得るために労力を劇的に削減できることを示す。

本稿は、国際学会ですでに発表済みの内容 [3] を簡潔にまとめたものである。

2 関連研究

Human-in-the-Loop のコンセプトは、単にインタラクティブであるだけでなく、人間をシステムの反復の一部として取り込むことで、システム自体がより望ましい結果を出力できるように更新していくことがある。中野ら [5] は複数パートがミックスされた音源から歌声を分離するために、事前学習済みの DNN のファインチューニングをインタラクティブにおこなう手法を提案した。このシステムでは、ユーザが一旦推定された歌声の基本周波数 (F0) の軌跡を GUI 上で修正しファインチューニングのための新たな学習データとすることで、楽曲とユーザへの適応をおこなっている。これは我々のアプローチに近いが、前述したようにユーザの労力を考えたときに現実的に与えることのできる局所的かつ少量の修正データだけでは十分なモデル適応をおこなうには不足である可能性が高いという問題がある。

Human-in-the-Loop のシステムでは取り込む人が個人とは限らない。HumanGAN [6] では音声合成のための機械学習の品質を向上させるために、学習中に生成された音声が機械が生成したものなのか実際の人間の声の録音かどうかをクラウドソーシングを使って実際に人に判別させて評価関数として利用した。しかしながらこうした不特定多数の人

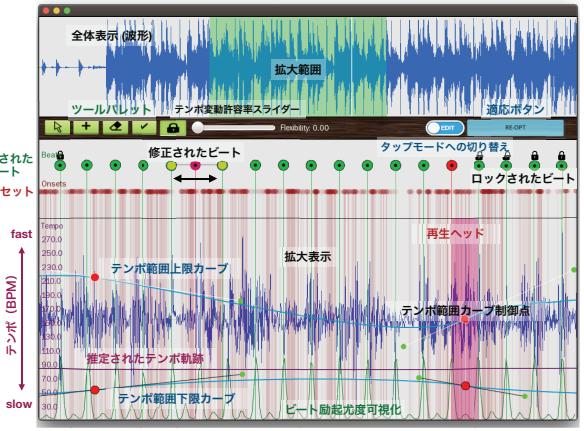


図 2. システム GUI. 楽曲が入力されると波形とともに推定されたビート情報を表示する。システムはユーザに複数種類のビート修正ツールを提供する。

間を利用するアプローチでは多くの人々の平均的な「正解」は得られるが、本稿の目的とする特定個人の所望の結果を得ることには向かない。

Bongjun ら [7] は音響イベントを効率的にアナテーションするための手法を提案した。ユーザがある音響イベントを一箇所指定してシステムに与えると、システムは時系列全体の中から類似した音響イベントの可能性が高い場所を候補場所としてユーザにいくつか提示する。ユーザはそれをチェックし、実際に求めている音響イベントかどうかを判断してシステムにフィードバックする。システムはそれを利用して内部の類似探索モデルを更新し、より精度良くユーザの求めている候補点を提示できるようになっていく。このアプローチは一部我々の提案手法にも取り入れられており、ユーザの局所的な修正の影響が高い可能性が高い場所を推薦するために利用している。

3 ワークフロー

解析対象の楽曲が入力されると、システムはまず事前学習済みの DNN で拍節の解析をおこない、初期状態として画面に結果を表示する（図 2）。楽曲を再生すると、システムは再生ヘッドとともに推定された拍の位置でクリック音を重ねて鳴らすことでユーザに拍位置を提示する。ユーザはそれを聴き、その結果が望ましいかどうかを判断し、もし望ましくなかった場合にはそれを修正する。この修正作業のためには内容に応じて何種類かのツールが用意されている。特定の拍の位置をマウスで移動したり、追加/削除できる基本的な操作に加えて、DNN の出力から最終的なビート位置を確定するためのテンポ軌跡推定（4 章）をする際の上下限の範囲をベジェ曲線で設定したり、テンポ変動がどれだけ許容されるかをスライダーで設定することができる。テンポ上下限の設定は、推定された拍が明らかに所望のものか

ら外れている場合（例えば2倍や半分のテンポだったとき）にまとめて全体を修正するのに有効である。またテンポ変動許容率はロックやポップスなどほぼテンポが一定の場合は低く、逆にクラシックのように大きく変動する楽曲に対しては高く設定すると所望の結果が得られやすくなる。また、ひとつずつ拍位置を修正するのではなく、再生しながら同時にキーボードをタップすることでその部分の拍をまとめて直接指定するツールも用意されている。

ユーザはこれらのツールを使ってごく少数の気に入らない箇所のみを修正し、残りの部分に修正箇所があるかどうかは閲知する必要はない。ユーザが修正したのち、図2右上の適応ボタンを押すことで内部モデルの適応プロセスが開始され、残りのユーザが直接編集していない箇所にも自動修正がおこなわれていく。この適応は反復ごとに経過が画面に表示されるので、ユーザは任意のタイミングで適応プロセスをストップし、結果を聴いて確認する。この反復作業をユーザは楽曲全体の推定結果が望ましいものになるまで繰り返す。

ここで想定し得る問題として、単純にモデル適応をおこなった場合、本当は修正してほしくないところまで修正されてしまう可能性がある。これを防ぐため、提案手法には修正が必要な可能性が高い場所には大きな影響を、逆にそうではない場所にはより小さな影響のみを及ぼすようなインタラクティブな適応プロセスを組み込んでいる（5.3章）。ユーザがある拍を修正すると、システムはその修正によって影響が大きい可能性がある場所を何箇所か提示する（図3の？マーク）。ユーザはその場所を再生して、

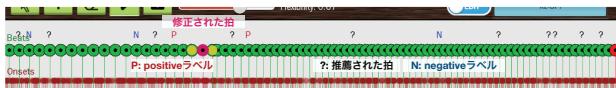


図3. 修正の影響が大きい場所のラベリング

本当に修正が必要かどうかを判断し、必要な場合はpositive、そうでない場合はnegativeラベルを割り当てる。もし提示された拍が修正が必要だったとしてもユーザはその可否を判断するだけではなく、直接手で修正する必要はない。これはユーザの労力を削減するうえで重要な点である。なぜなら拍の位置の修正作業は、単に音を聴いて可否を判断するだけよりずっとコストが高いからである。このラベリング作業は必須ではなく、またおこなう場合も提示されたすべてのビートにやる必要もない。しかし適応ボタンを押す前にわずかな追加タスクとして目についた推薦ビートに対しておこなうことでユーザは不都合なモデル適応を避けることができる。

4 ビートトラッキング

本稿で扱うビートトラッキングのアルゴリズムは基本的にTemporary Convolutional Network (TCN)

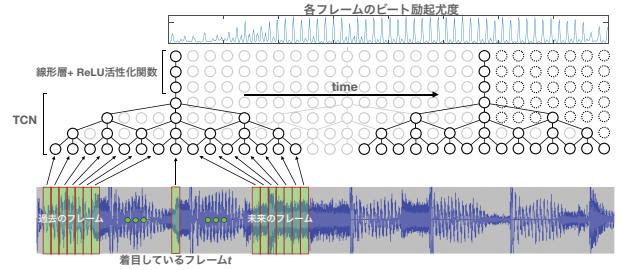


図4. 事前学習時にはある時刻周辺の局所的な情報を入力してその時刻のビート励起尤度を出力する。

をもちいた手法に基づいている[1]。この手法は2段階のプロセスから成り立つ。まず、楽曲を細かな時間フレームに分割（サンプリング周波数44.1kHzで窓長4096サンプル、Hanning窓、ホップサイズ2048サンプルを採用）しフレームごとの特徴量（48次元MFCC時間差分を採用）を計算し、これをDNNに入力する。このDNNは各時刻の前後約2.4秒分を含めた局所的な91フレームを入力として、その時刻におけるビート励起尤度 $\in [0, 1]$ を出力する。次に2段階目として、求まったビート励起尤度の系列をもとに楽曲全体を通して最も尤もらしいテンポ軌跡を計算することで実際の離散的なビート位置を確定する[8]。我々のDNNアーキテクチャは3段のTCNブロックと直列に接続された3層のReLU活性層付き線形層から成る（図4）。また、ビート励起尤度系列からの最尤テンポ推定には隠れセミマルコフモデル[9]によるViterbiアルゴリズムによって計算をおこなう。なお本稿の実験でのDNNの事前学習にはデータセットとしてGTZAN[10]、RWC Popular, Genre[11][12]とAmazon Musicから購入した幅広いジャンルの楽曲400曲から自前で作成したものを使用した。

5 適応アルゴリズム

5.1 ユーザ修正に対する適応目的関数

ユーザが拍位置を修正すると、システムは現在の解析結果との誤差からモデル適応をおこなうための目的関数を作成し、それを確率的勾配法で最小化する。ここでは例としてある拍の位置を移動した場合を想定して説明するが、他のツールでの修正についても同様の考え方を適用できる。適応後に目指すべき状態として、ユーザが修正後のビートを b_i 、その前後のビートを b_{i-1}, b_{i+1} とすると、DNNの出力が b_i の位置で1に近く（ビート励起尤度が高く）、逆に拍と拍の中点 $b_i^- = \frac{b_i+b_{i-1}}{2}$, $b_i^+ = \frac{b_i+b_{i+1}}{2}$ で0に近く（ビート励起尤度が低く）なるようにする（図5）。よって最小化すべき損失 $L(t)$ は修正位置周辺の各フレーム t （図5のオレンジのフレーム）上で、

$$L(t) = \left| DNN(t) - \left(1 - \min(1.0, \frac{|b^c - t|}{W}) \right) \right|_2 \quad (1)$$

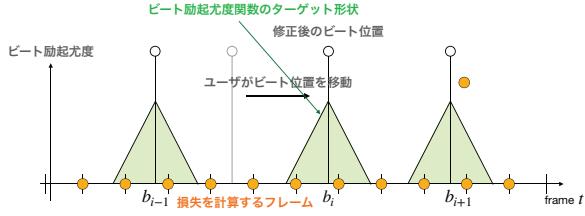


図 5. ユーザ修正から DNN へ与える損失の作成。

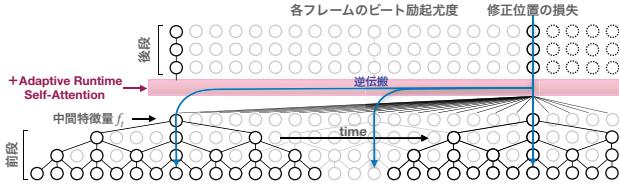


図 6. 実行時には事前学習された DNN を分割し、新たに ARSA ブロックを組み込む。

のように定義される。ここで、 b^c と W はそれぞれフレーム t から最近傍のビートと窓長、 $DNN(t)$ はフレーム t での DNN の出力である。

5.2 適応型実行時自己注意機構 (ARSA)

我々の DNN モデルは事前学習時にはある時間フレームでのビート励起尤度を出力するためにその周辺の局所的な情報のみを利用する構造をしている(図 4)。一方実行時には、その事前学習した DNN を前半ブロック(3 ブロックの TCN)と後半ブロック(TCN ブロック以降の出力段側)の 2 つに分割して、その間に ARSA を挿入する(図 6)。つまり我々のモデル構造は学習時と実行時で異なる。ARSA は学習時の局所的な視野の適用とは異なり、楽曲全体を通した中間特徴量(前半ブロックの出力)を集約し、後半ブロックのための入力を出力する。この学習時と実行時で異なる構造をとる新しいタイプの DNN モデルは、学習時の計算コストの削減とユーザの局所的なフィードバックの影響を楽曲全体へ効率良く分配する実行時適応をバランス良く両立する。

前半ブロックの出力である中間特徴量を $f^t \in \mathbb{R}^D$ とすると、ARSA はまず通常の注意機構と同様に各時刻 t のクエリ $Q^t = \mathbf{W}_Q \cdot f^t$ 、キー $K^t = \mathbf{W}_K \cdot f^t$ 、バリュー $V^t = \mathbf{W}_V \cdot f^t$, $\mathbf{W}_Q, \mathbf{W}_K, \mathbf{W}_V \in \mathbb{R}^{D \times D}$ を計算する。これらのベクトルから ARSA のフレーム t におけるフレーム i への注意マップは、

$$d(t \rightarrow i) = (w^t \otimes Q^t) \cdot (P_e(|t - i|) \otimes K^i), \quad (2)$$

で計算される。ここで、 $w^t \in \mathbb{R}^D$ はユーザのラベリングによって決まる重み、 $P_e(\delta t) \in \mathbb{R}^D$ は相対位置埋め込みである [4]。この $d(t \rightarrow i)$ を i についてソートし値の大きなものから N フレーム(ここでは $N=512$ を採用)サンプリングした集合を C_t とすると ARSA の出力は、

$$y_t = (1 - \alpha) \cdot V^t + \alpha \sum_{i \in C_t} \frac{e^{d(t \rightarrow i)}}{\sum_{j \in C_t} e^{d(t \rightarrow j)}} \cdot V^i, \quad (3)$$

となる。 α は重み係数である($\alpha = 0.5$ を使用)。なお、 $\mathbf{W}_Q, \mathbf{W}_K, \mathbf{W}_V$ については単位行列で初期化することによりこの ARSA は非適応時ほぼ恒等関数として機能する。

5.3 インタラクティブな注意マップ最適化

式 (2) の w^t は ARSA 独特のパラメータでありユーザのインタラクティブなラベリングによって決まる。3 章で述べたように、ユーザが拍を修正するとシステムはそこからモデル適応をおこなったときに影響が大きい可能性が高い場所がいくつか提示する。この提示箇所は注意機構本来の仕組みに基づいて決定される。注意マップ $d(t \rightarrow i)$ が大きいということはフレーム t はよりフレーム i に対して注目しているということである。逆に言えば、フレーム t 上で与えられた損失はフレーム i により多く逆伝搬するということを示している。よって単純に注意マップの値が大きな箇所 C_t 付近の拍が影響を受けやすいとみなすことができるため、それらを提示する。

提示された拍に対してユーザは本当に修正が必要かどうかを判断し、positive/negative ラベルを割り当てる。ラベル自体は拍ごとに割り当てられるが、システム内部ではそれらの拍周辺のフレームに対してラベル付けがおこなわれる。これらの関係から w^t は Fisher's criterion [7] と同様に、

$$w^t = \frac{(avg(A^{t, pos}) - avg(A^{t, neg}))^2}{std(A^{t, pos})^2 + std(A^{t, neg})^2}, \quad (4)$$

のように求められる。ここで、 $A^{t, pos}$ と $A^{t, neg}$ はそれぞれフレーム t に対する positive/negative フレーム i での $Q^t \otimes P_e(|t - i|) \otimes K^i$ である。この w^t はユーザが新しいラベルを割り当てるたびに再計算され、ユーザの修正によって影響を受けてほしくない場所への影響を減らし、逆に影響を受けてほしい場所には増やす効果がある。

6 評価

6.1 シミュレーション実験

提案手法の性能を評価するため、まずシミュレーション環境下での実験をおこなった。この実験では ARSA 有りと無しのときのモデル適応を比較する。ここで ARSA 無しの場合とは、単純に事前学習した DNN 構造をそのまま使い、ユーザが修正したデータを新たな学習データとしてファインチューニングすることを意味し、これは中野ら [5] の手法とほぼ同様のものとなる。検証用データとしては事前学習データには一切含まれていない SMC MIREX [13] を使用した。このデータセットにはビート位置のアノテーション情報とともに様々なジャンルの 217 曲が収録されている。またすべての曲の長さは 40 秒に揃えられている。

実験の手続きとしては、まずそれぞれの曲において解析をおこない、最も大きなエラー（拍周辺のフレームの平均二乗誤差）を観測した拍を曲中から探し出し、それを自動的にアノテーションされている正解位置へ修正する。次にシステムに対して 100 反復のモデル適応を実行する。この自動での拍修正とモデル適応の組み合わせを 5 回繰り返し、計 500 回の適応反復を得る。ARSA のラベリングについては自動で各修正ごとにランダムに選択された 5 箇所ずつおこなう。ここで可否の判断については、提示された拍がアノテーションから ± 70 ms 以内に入っているれば positive、それ以外を negative とみなした。

図 7 に結果を示す。上段に各楽曲全体を通した平均二乗誤差（各曲ごとに拍の数で正規化後）を示している。明らかに提案手法はほとんどの楽曲において、追加修正が与えられる以前に小さなエラーへ速く収束しているのがわかる。一方、比較手法では 100 反復ごとにおこなわれる拍修正のタイミングで階段状にエラーは下がっているが、曲全体としてはあまり誤差が小さくならないことがみてとれる。これは局所的な修正でのファインチューニングだけでは楽曲全体に対して適応をおこなうには不十分であり、ARSA の導入によってそれを大きく改善できることを示している。図 7 下段はどちらの手法のほうがどれだけ適応が速く進行するかの比較を示しており、F1 値（拍節解析において曲全体を通した正解率として広く使われている指標 [14]）の値が初めて 0.8 を超えたときの反復回数を N^A （提案手法）と N^B （比較手法）とすると、それらの差 $N^B - N^A$ を示している。つまり大きな正の値（青）になるほど提案手法のほうがより速く精度の高い状態に到達したことを示しており、ほとんどの曲において提案手法のほうが比較手法を上回っていることがみてとれる。実際、統計的にも 156/217 曲においてカイ二乗判定で有意 ($p < 0.05$) に提案手法のほうが速く収束している結果となった。また、計算コストとしては、このシミュレーション実験において 217 曲すべてを処理するのに提案手法で 5.5 時間、比較手法で 5 時間を要した。つまり平均して一曲 40 秒あたり 8 秒の差であり、これはユーザが実際に修正をおこなう時間に比べて実用的に無視できるオーバーヘッドと言える。

6.2 ユーザスタディ

6.2.1 セットアップ

提案手法がどれだけユーザの修正効率を向上させるかを評価するため、ユーザスタディをおこなった。この実験では、対象楽曲の解析結果に含まれるエラーを修正するタスクでの作業時間を計測する。この作業時間の基準として、我々は対象楽曲の総再生時間を設定した。これは以下の理由によるものである。既存の方法で最も速くビート位置をアノテーションする方法は、曲全体を通して一度で正確にタッ

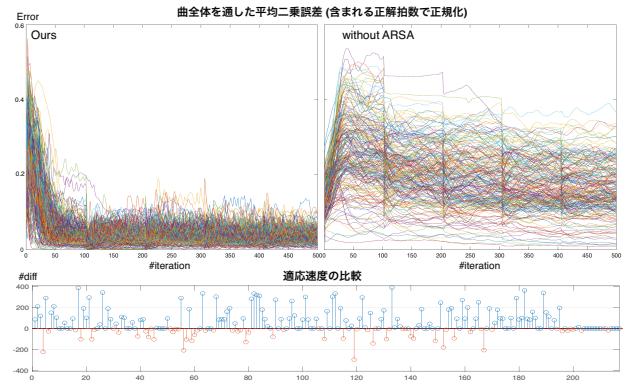


図 7. シミュレーション実験結果。上段: 反復回数にともなう各楽曲全体を通した平均二乗誤差の推移。下段: 提案手法と比較手法間での収束速度の差分。

普することであり、この最短時間は曲の長さ分である。よってこの時間より短時間で被験者が修正を終えることができれば、提案手法の有効性を示すことができると考えられる。

被験者として普段から音楽編集ソフトウェアに慣れ親しんでいる 5 名を集めた。まず最初にそれぞれの被験者に我々のインターフェースの使い方を説明し、実際に触って練習してもらった (45 分)。次に本番タスクとして、できる限りすばやくテスト楽曲に含まれる解析エラーを提案インターフェースを使って修正してもらった。このテスト楽曲には RWC Jazz データセット [11] から 2 曲 (No.40 - 7 分 41 秒, 正解ビート数 481 個, No.41 - 6 分 6 秒, 正解ビート数 637 個) を選んだ。どちらの楽曲も初期状態での解析では手動ですべて修正するには現実的ではない多くの数のエラーが含まれていた。すべての被験者がこの 2 曲を修正した。また、被験者依存による楽曲への解釈の差異を避けるため、我々は正解ビートの位置でクリック音を重ねた「正解音楽ファイル」をあらかじめ作成しておき、このタスクを始める直前に必ずすべての被験者に聴いてもらい、その「正解」を目指して修正をおこなうように指示をした。修正するためには 3 章で述べたあらゆるツールを使用可能としたが、タップ指定のツールだけは連続して 5 秒以上使用することを禁じた。これは、曲全体を一度だけ通してタップした場合のベースラインと明確に差別化するためである。一箇所修正するごとに必ず適応ボタンを押すことも義務付けた。このタスクをおこなっている間、我々は時間を計測するとともに F1 値の時間変化を記録した。それぞれのタスクは F1 値が 0.9 を超えた段階 (3 秒間以上継続。適応途中で瞬間に値が高くなる状態を除外するため) でたとえ被験者が曲全体を聴いて最終チェックをしていなかったとしても打ち切って終了させた。最後にこの本番タスクを終えたのち、インターフェースの使用感についての簡単なインタビューをおこなった (10 分)。

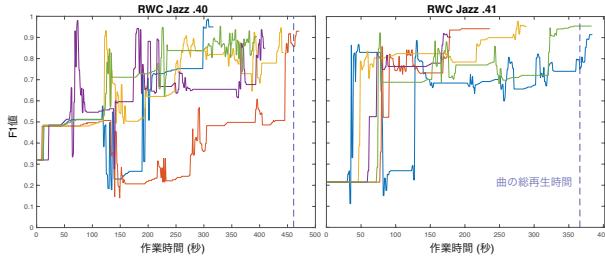


図 8. 作業時間経過(横軸)にともなう各被験者ごとのビート正解率(縦軸, F1 値)の推移.

6.2.2 結果

図 8 に本番タスクでの各被験者の F1 値の推移結果を示す。垂直の点線は楽曲の総再生時間を示している。それぞれの楽曲において 4/5 人と 3/5 人の被験者が曲の再生時間より早くタスクを終了することができた。また、残りの被験者もほぼ再生時間と同等の時間でタスクを終えている。これは、ほとんどの被験者が曲全体を聴いてはおらず、タスクが完了したことを気付く前(全体をチェックしていないため)に所定の精度に到達して打ち切られているということを意味する。もちろん、現実的に提案インターフェースを使用する場合は最後に曲全体を聴いてチェックをする必要があるので曲の総再生時間より作業時間が短くなることはありえない。それゆえあくまで条件付きではあるが、提案手法は効率的なエラー修正をおこなうことができたと結論付けることのできる結果を得たと言える。

インタビューについては、我々はあらかじめ 2 つの質問を用意した。ひとつめは「音楽制作ソフトにこのインターフェースが実装されたら使ってみたいですか?」である。この質問に対してはすべての被験者が Yes と返答した。2 つめの質問は、「このシステムを使用していた上で何か問題がありましたか?」である。この質問に対しては、何人かの被験者は A: 「様々な種類の修正ツールが用意されているため、発見したエラーに対してどのツールを使えばよいのかわからなかった」と答えた。また、B: 「適応ボタンを押したあと、反復を自分で止めなければいけないが、どのくらい待って止めればよいのか分からなかつた」という回答もあった。

6.2.3 議論

インタビューにおけるコメント A に対しては、被験者がインターフェースにより時間をかけて慣れ親しむことで解決していく問題だと考えられる。しかしながら、もっとより良い解決法としてシステムが現在のユーザのおかれている状況を把握し、最適な修正ツールを推薦することは今後の課題である。コメント B に対しては、数値計算の反復という概念自体が一般的のユーザには馴染みが薄く、そうしたユーザに反復回数を任意で決定させるのは現実的ではない

と考えられる。それゆえにシステムが自動的に反復をストップさせることができるような指標を設計することが求められるが、我々の適応プロセスは一意に収束するものではないためそれは簡単ではなく今後の課題である。似たような問題として、我々の解いている最適化問題では楽曲全体のエラー率が単調に減少せず、ときに反復の最中に増大することもある。これはユーザにとってみればエラーの修正作業をしたにもかかわらず曲全体のエラーが一時的にでも増えたように見える、という直感に反する事態となり、見せ方には今後工夫が必要だと考えられる。

最後に、この実験群だけでは本稿の大きな目的のひとつである、ユーザによって複数の解釈ができる楽曲に対して効率的に対応できる、ということを十分に評価できていない可能性が残っている。ユーザ実験でこれをおこなう場合、公平性を確保するため同一楽曲を同じ判断時間で異なる解釈をするような被験者を集めることができず困難であり、現状シミュレーション実験に頼らざるを得なくなっている。例えばシミュレーション実験で用いた SMC MIREX [13] は実際に様々な解釈ができる自動解析だけでは困難な楽曲に対して集団からランダムに選ばれた被験者の解釈でアノテーションしたものであり、その点では部分的に評価できていると言える。しかしながらこれでは厳密に同一楽曲へ対しての評価とは言えないため、ユーザスタディの実験方法を工夫しより詳細な分析をおこなっていくことは今後の課題である。

7 まとめと今後の展望

本稿では特定のユーザと楽曲へ Human-in-the-Loop のアプローチで適応できるインタラクティブなビートトラッキングインターフェースを提案した。これを実現するため、学習時の計算コストを削減でき、かつユーザの局所的なフィードバックの影響を入力系列全体へ伝搬させて効率的な実行時適応をおこなうことのできる適応型実行時自己注意機構を導入した。我々はいくつかの実験をおこない提案手法の実行可能性と性能を検証した。こうした機械学習モデルの事前学習は局所的におこない、逆に特定対象への実行時適応は系列全体の文脈を集約するアプローチは、ビートトラッキングだけではなく、コード認識、歌声合成、音源分離などといった他の音楽情報処理、さらには 3D アニメーションの生成や動画編集、といった多くの時系列データを扱う領域で有効であると今後期待される。

参考文献

- [1] M. E. P. Davies and S. Bock. Temporal Convolutional Networks for Musical Audio Beat Tracking. *EUSIPCO*, 2019.

Human-in-the-Loop 型適応によるインタラクティブな音楽的拍節解析

- [2] S. Reich. Clapping music (1972), Pianophase (1967)
- [3] K. Yamamoto. Human-in-the-Loop Adaptation for Interactive Musical Beat Tracking. *ISMIR*, 2021.
- [4] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention is All You Need. *NeurIPS*, 2017.
- [5] T. Nakano, Y. Koyama, M. Hamasaki, and M. Goto. Interactive Deep Singing-Voice Separation Based on Human-in-the-Loop Adaptation. *IUI*, 2020.
- [6] K. Fujii, Y. Saito, S. Takamichi, Y. Baba, and H. Saruwatari. Humangan: Generative Adversarial Network with Human-Based Discriminator and Its Evaluation in Speech Perception Modeling. *ICASSP*, 2020.
- [7] B. Kim and B. Pardo. A Human-in-the-Loop System for Sound Event Detection and Annotation. *IUI*, 2018.
- [8] F. Krebs, S. Bock, and G. Widmer. An Efficient State-Space Model for Joint Tempo and Meter Tracking. *ISMIR*, 2017.
- [9] R. Chen, W. Shen, A. Srinivasamurthy, and P. Chordia. Chord Recognition using Duration-Explicit Hidden Markov Models. *ISMIR*, 2012.
- [10] G. Tzanetakis and P. Coo. Musical Genre Classification of Audio Signals. *IEEE Transactions on Audio and Speech Processing*, 2002.
- [11] M. Goto, H. Hashiguchi, T. Nishimura, R. Oka. RWC Music Database: Popular, Classical, and Jazz Music Databases. *ISMIR*, 2002.
- [12] M. Goto. AIST Annotation for the RWC Music Database. *ISMIR*, 2006.
- [13] A. Holzapfel, M. E. P. Davies, J. R. Zapata, and J. L. Oliveira, A. Srinivasamurthy, and P. Chordia. Selective Sampling for Beat Tracking Evaluation. *IEEE Trans. Audio, Speech, and Language Processing*, 2012.
- [14] S. Dixon. Evaluation of the Audio Beat Tacking System Beatroot. *New Music Res*, 2007.

色の修飾語を用いた対話的なイラストの色編集

住田 桃子 * Nolwenn Maudet[†] 福里 司 * 五十嵐 健夫 *

概要. イラスト制作の際、色の編集は非常に重要な工程である。しかし、色を調整するには複数のパラメータ（例：RGB 値や HSV 値）を設定しなければならず、特に初心者ユーザにとって、頭の中で想像する色を見つけることは非常に難しい。そこで本研究では、“Lighter”等の色に関する単語（修飾語）を用いて、効率的に色空間内を探索する手法を提案する。具体的には、ユーザは修飾語を指定することで、修飾語に対するカラースライダを自動生成するものである。また、本システムは複数の領域に対して同時に色の編集を行うことができるため、画像全体の色の関係を維持したまま編集することができる。更に初心者ユーザとアマチュアアーティストを対象としたユーザテストを実施し、本研究の有用性を検証した。

1 はじめに

アーティストやイラストレータは、カラースライダ等を用いてイラストの色編集を行っている。しかし、色は3つ（以上）のパラメータで構成されているため、特に初心者にとってパラメータ値に対する色を把握することは未だに困難である（例：画像加工においてどのパラメータを変更すれば良いのかの判断や、シーンにあった新色を追加する際の判断）。

これらの問題を解決するために、パラメータ空間（色空間）を効率的に探索する技術が複数提案されている。Shugrina ら [14] は、アナログの絵具パレットのように、PC 画面上で複数色のブレンドを直接行うためのデジタルパレットを開発した。更に、Shugrina ら [15] は、ユーザが指定した色に（RGB/HSV 空間で）近い色を可視化する拡張機能も考案している。これらの手法は、ブレンド操作の時にパラメータ値の直接的な操作（例：スライダ操作）を行う必要がないものの、ユーザは適切な色サンプルを用意しなければならない。更に、「今の色よりもっと薄くしたい」等の色の編集には不向きである。更に、入力イラストの色情報（初期条件）やイラスト全体のバランスを考慮しながら色を調整することも難しい。

そこで本稿では、“lighter”や“greener”等の色に関する単語（修飾語）を用いて、イラストの色を効率的に編集するための「修飾語スライダ」を提案する（図1）。提案手法では、入力イラスト中で「編集したい領域」に対する修飾語スライダを逐次構築作成することで、適切な色を設定することができる。

提案する修飾語スライダを定性的に評価するために、16名の被験者（素人ユーザ8名とアマチュア・アーティスト8名）を対象としたユーザテストを実施した。その結果、提案手法は、ユーザに新しい色をイメージさせやすくする点や、既存のツールと組

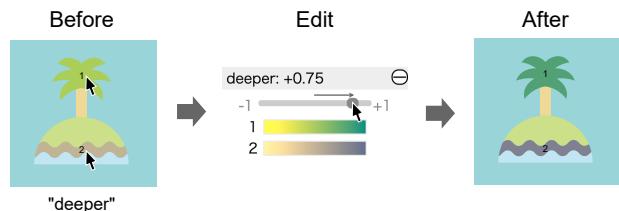


図 1. システムの概要。ユーザは画像領域と修飾語を選択し（左）、生成された修飾語スライダを用いて選択領域の色値を調整する（中・右）。イラストは FLAT ICON DESIGN [17] から引用。

み合わせることで、より色の探索を効率化できる点といったポジティブな回答が得られた。

2 関連研究

2.1 色の編集

市販のペイントツールには、カラースライダや、ブレーン、ホイール、パレットといった色の設定機能が多数用意されているが、RGB/HSV 色空間（パラメータ値）の中から適切な色を見つけることは依然として難しい。このような問題を解決するために、色設定をサポートするための手法が数多く提案されている。これらは大きく三つのアプローチに分類される。

一つ目のアプローチは、ユーザが描きたい絵の構図を考える工程自体を支援する方法である。Meier ら [7] や Jalal [5] は、色付きのシンプルな図形（例：円や四角形）を画面上に配置することで、描きたい絵の構図（イラスト全体の色のバランス）を事前に確認するシステムを考案している。しかし、これらは各図形に対する色の設定は自力で行わなければならない課題がある。二つ目のアプローチは、事前に用意された複数の色を基に新たな色を生成する手法である。Shugrina ら [14] は、アナログの絵の具パ

* 東京大学

† The University of Strasbourg

レットを参考に、ユーザが画面上で複数の色を直接ブレンドできるデジタルパレットを考案した。しかし、この手法は、ブレンドするための適切なサンプル点（初期の色）を複数用意しなければならず、RGB空間内の色の設定を効率化することはできない。三つ目のアプローチは、RGB/HSV空間内での探索をサポートする方法である [5][8]。例えば、Shugrina ら [15] は、任意の色に対する（RGB/HSV 空間上の）隣接色を視覚化する機能を設計した。しかし、これらの手法はイラストの色編集を行うタスクにおいて、入力イラストの色情報を考慮することができない。これらを踏まえ、我々は入力イラストの色情報を考慮しつつ、より直感的な色の編集手法を検討する。

2.2 自然言語処理分野における色の表現

任意の色に対して特定の単語（例：“赤”や“青”）が用いられている背景から、自然言語処理の分野では「色情報」と「言語情報」の関係に関する研究が盛んに行われている（例：色記述の生成、色修飾語の役割）。例えば、McMahan ら [6] や Monroe ら [11] は色を説明するための単語や文章をモデル化することで、任意の色に対する説明文を自動生成する手法を提案している。また、Monroe ら [12] は、言語情報を利用し、類似する色の中から各色を識別する方法を提案した。更に、Winn と Muresan [18] は色の修飾語（例：lighter）に対する色の変換ルールを提案し、Han ら [3] によって、このルールが一般化されている（詳細は第 3.1 節参照）。本研究では、色の修飾語による色の変換ルールを利用し、イラストの色の編集を行うタスクに特化した手法を検討する。

2.3 自然言語の修飾語を用いたパラメータ操作

言語情報を用いてユーザーのデザインプロセスを支援する手法が提案されている [2]。Yumer ら [19] は、「よりファッショナブルに」などの自然言語の修飾語を用いて、3D 形状を変形させる手法を提案した。また、Streuber ら [16] は、クラウドソーシングによって得られたデータを用いて 3D アバターを生成するツールを設計した。

言語情報を用いた色情報の編集技術としては、主にカラーパレットの設計 [1] や線画イラストへの色付け [20] 手法が挙げられる。また、Heer ら [4] は、Web アンケートによって集めたデータを基に色の修飾語の確率モデルを構築し、カラーデザイン向けのインターフェース（例：修飾語によるピクセル選択機能）を構築した。これらの手法は、言語から直接色を生成するタスクに適している反面、入力画像の色情報を考慮しながら色値の編集は困難である。そこで本研究では、上記の手法を参考に、自然言語情報を用いたパラメータ操作を検討する。



図 2. 提案システムのスクリーンショット。

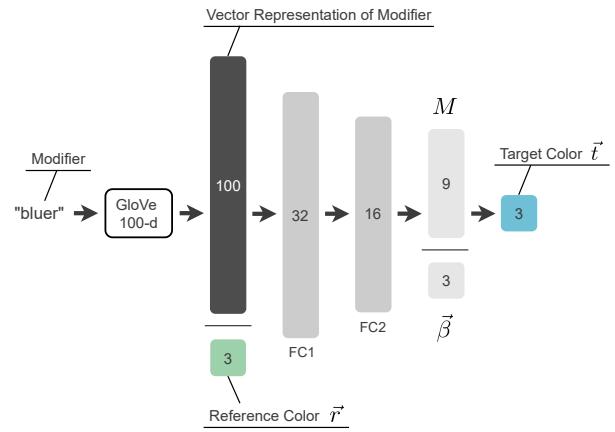


図 3. 色の修飾語を基に入力の色 $\vec{r} \in \mathbb{R}^3$ を変換するネットワークの構造。

3 提案手法

本システム（図 2）は、Python 3 と JavaScript をベースに、Flask フレームワークを用いた Web アプリケーションとして実装している。

3.1 色変換モデル

Han ら [3] は、“lighter”などの色の修飾語を用いて入力の色 $\vec{r} \in \mathbb{R}^3$ を変換するモデルを提案した。具体的には、Winn と Muresan [18] が提案するルールを RGB 色空間上で一般化するために、以下のモデルを提案した。

$$\vec{t} = M\vec{r} + \vec{\beta} \quad (1)$$

$\vec{t} \in \mathbb{R}^3$ を変換後の色、 $M \in \mathbb{R}^{3 \times 3}$ は変換行列、 $\vec{\beta} \in \mathbb{R}^3$ は変換ベクトルを意味する。

次に、Winn と Muresan のデータセット [18] を用いて、色の修飾語から M と $\vec{\beta}$ を推定するネットワークを構築する（図 3）。但し我々は、色の修飾語を特徴ベクトルに変換する際、Han らが用いた 300 次元の Word2Vec [9][10] の代わりに、GloVe

表 1. 色の修飾語のリスト.

acidic	brownish	deeper	faded	hotter	muddier	pastel	rich	truer
bluer	burnt	dirtier	goldener	lighter	murkier	pinker	richer	uglier
bluish	cooler	duller	grayish	medium	off	pinkish	rosier	violet
brighter	darker	dustier	greenish	mid	orange	redder	sandier	weird
browner	darkish	electric	grosser	more	paler	reddish	softer	yellower

100-d [13] を用いることで計算コストの削減を行った。本ネットワークには、任意の色の修飾語をテキスト入力することができるものの、精度が不十分な修飾語が存在する可能性がある。そこで我々は、良好な結果が得られる修飾語を選別し、リスト化した（表 1）。但し、任意の修飾語を指定できるように、本システムにテキストボックスを用意している（図 2(右側)）。

3.2 対話的な色調整

本システムを用いたイラストの色の編集の手順は、入力イラストの(1)領域を選択、(2)修飾語の指定、(3)修飾語に対応したスライダによる選択領域の色の調整、を繰り返すものである。

3.2.1 領域選択

入力イラスト（SVG 形式）の領域を直接マウスクリックすることで、一つの領域を選択することができる（過去に選択された領域情報は上書きされる）。また、Shift キーを押しながらマウスクリックをすることで、複数の領域を同時選択することができる。視認性の向上のため、選択された領域にラベル ID ($i \in \{1, \dots, N\}$) を付与し、1.5 秒に 1 回点滅する機能を追加している。但し、この点滅は、ユーザがスライダ操作を行う際は自動的に OFF になる。その一方で、Clear Selection ボタンを押すことで、選択された領域情報をリセットすることができる。

3.2.2 色の修飾語に対するスライダの生成

本システム下部に表示された修飾語のリスト（表 1）をクリック（またはテキスト入力）することで、色変換モデル（式 1）に基づく修飾語スライダを生成する。スライダ操作による色の値 $\vec{c} \in \mathbb{R}^3$ は以下のように計算する。

$$\vec{c} = (1.0 - w)\vec{r} + w\vec{t} \quad (2)$$

$w \in [-1.0, 1.0]$ は重み係数、 \vec{r} は選択された領域の色、 \vec{t} は色変換モデルによって推定された色である。つまり、スライダ操作によって重み係数を変更することで、ユーザは自由に選択した領域の色を変更することができる（図 4）。スライダの右側 (= +1.0)

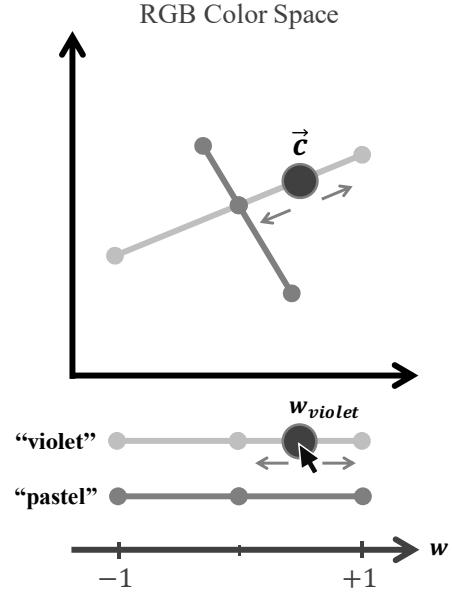


図 4. 色の修飾語スライダを用いた RGB 空間内の色探索の一例。

は修飾語に対応する色に近づき、スライダの左側 (= -1.0) は修飾語から離れた色を意味する。

また、右上部にあるマイナスボタンをクリックすることで、一度生成した修飾語スライダを削除することができる（図 2）。

3.2.3 サポート機能

スライダ操作による色の編集を効率化するために、三種類のサポート機能を実装している。

a) **プレビュー機能**： 各スライダの下部に、選択された領域（ID 番号）に対応する色を直接表示する。また、ユーザがスライダ値を調整した時、画像上の選択領域の色もリアルタイムに変更する。

b) **スライダの範囲拡張機能**： Extend ボタンを押すことで、スライダの値 w の範囲を拡張することができる（最大で $-3.0 \leq w \leq +3.0$ まで拡張可能）。

c) **スライダの更新機能**： 修飾語スライダは、指定した領域の色情報 \vec{r} を基にスライダを生成する。そこで、スライダ操作後の色 \vec{c} を新たな領域の色 $\vec{r}' = \vec{c}$



図 5. ユーザテストで使用した SVG 形式のイラスト.

表 2. ユーザテストで用いたアンケート内容.

#	質問項目
Q1	色空間を把握できた。
Q2	適切な色を見つけやすかった。
Q3	言葉に合う新しい色を想像しやすかった。
Q4	色の微調整がしやすかった。
Q5	イラストの色編集が十分にできた。

に代入することで、修飾語スライダの方向（変換行列 M と変換ベクトル $\vec{\beta}$ ）を更新することができる。

4 ユーザテスト

本手法の有用性を評価するために、16名の被験者（デジタルイラスト作成経験のあるアマチュアアーティスト：8名と素人ユーザー：8名）を対象とし、従来の RGB/HSV スライダ（ベースライン）との比較実験を行った。

4.1 手順

SVG 形式のイラスト（図 5）と編集内容を説明する文章（例：森の中で悲しんでいるような色合い、洞窟の中で怒っているような色合い）を被験者に与え、各手法を用いてイラストの色を編集してもらうタスクを行った。タスク完了後に、7段階のリッカート尺度（1 = “全然満足していない” - 7 = “とても満足している”）を用いたアンケート（表 2 参照）に回答してもらうことで、各手法を評価してもらった。更に、簡単なインタビューを実施し、ユーザ体験に関するコメントや今後どのような拡張が必要なのかといったコメントを収集した。所要時間は、一人当たり約 1 時間である。

4.2 結果

図 6 にイラストの色編集結果の一例、表 3 にアンケート結果のスコアと Wilcoxon 符号順位検定の結果を示す。Q3 は通常の RGB/HSV スライダと比較し、提案手法のスコアが高いことがわかる。この結果から、修飾語に合うような「新しい色」をイメージするときに有用であったことが予想される。その一方で、Q4 では、通常の RGB/HSV スライダが高いスコアを示している。このことから、RGB/HSV スライダは色の微調整を行う際に適していることがわかる（但し、アーティストが RGB/HSV スライダによる色編集操作に慣れていた点が影響している



図 6. 修飾語スライダを用いた色編集結果の一例。左側：“洞窟の中で怒っているような色味”，右側：“海岸で陽気な雰囲気の色味”。

表 3. アンケート調査の結果.

#	平均 ± 標準偏差		p 値
	ベースライン	提案手法	
Q1	4.06 ± 1.83	4.94 ± 1.51	1.72×10^{-1}
Q2	4.88 ± 2.00	5.18 ± 1.41	5.86×10^{-1}
Q3	3.44 ± 1.41	5.69 ± 1.36	2.63×10^{-3}
Q4	5.63 ± 1.28	3.44 ± 1.93	4.88×10^{-4}
Q5	5.38 ± 2.49	5.25 ± 2.39	7.54×10^{-1}

可能性がある）。その他の質問項目については、平均スコアに多少の差があったものの、いずれも有意差は得られなかった。

また、被験者から、修飾語スライダは新しい色を考える時や、色空間の探索に優れているといった回答が得られた。例えば、4名の被験者からは、「RGB/HSV スライダは色の原理（計算方法）を知らないユーザにとって非常に扱いにくい。その一方、修飾語スライダは色値の変化の方向を簡単に設定することができるため、初心者にとても向いている」との意見があった。その他のコメントとして、3名の素人ユーザーからは「修飾語スライダ自体が楽しい」、2名のアーティストからは「色と言葉を結びつける作業自体が新たな楽しみになるかもしれない」「ユーザの好みに合わせてスライダをカスタマイズできること」との意見があった。しかしその一方、6名の被験者からは「RGB/HSV スライダは色の微調整に適しているため、修飾語スライダと RGB/HSV スライダを併用するとより便利になる」といったシステム自体の拡張に対する意見も得られた。

5 むすび

本研究では、色の修飾語に対するスライダを生成し、入力イラストの色を効率的に編集する手法を提案した。現実装では、色の修飾語の種類や言語が限定されているため、将来的には修飾語の種類の拡張（例：海、英雄など）や、様々な言語のカスタマイズ可能な機能を検討している。また、修飾語と色の対応についても、まだ不自然な結果が出てくることが多く、改善の余地があると考えられる。任意の修飾語とスライダの組み合わせは、色の調整以外の様々なコンテンツ制作支援に利用することが考えられるので、他の応用先についても検討する予定である。

謝辞

本研究はJST CREST JPMJCR17A1の助成を受けて実施されたものである。

参考文献

- [1] H. Bahng, S. Yoo, W. Cho, D. K. Park, Z. Wu, X. Ma, and J. Choo. Coloring with Words: Guiding Image Colorization Through Text-Based Palette Generation. In V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss eds., *Proceedings of 15th European Conference on Computer Vision (ECCV)*, pp. 443–459, Cham, Switzerland, 2018. Springer.
- [2] S. Chaudhuri, E. Kalogerakis, S. Giguere, and T. Funkhouser. Attribit: Content Creation with Semantic Attributes. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology, UIST ’13*, p. 193–202, New York, NY, USA, 2013. Association for Computing Machinery.
- [3] X. Han, P. Schulz, and T. Cohn. Grounding learning of modifier dynamics: An application to color naming. 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)*, pp. 1488–1493. ACL, 2019.
- [4] J. Heer and M. Stone. Color Naming Models for Color Selection, Image Editing and Palette Design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI ’12*, p. 1007–1016, New York, NY, USA, 2012. ACM.
- [5] G. Jalal, N. Maudet, and W. E. Mackay. Color Portraits: From Color Picking to Interacting with Color. In *Proceedings of the 2015 CHI Conference on Human Factors in Computing Systems, CHI ’15*, p. 4207–4216, New York, NY, USA, 2015. ACM.
- [6] B. McMahan and M. Stone. A Bayesian Model of Grounded Color Semantics. *Transactions of the Association for Computational Linguistics*, 3:103–115, 2015.
- [7] B. J. Meier, A. M. Spalter, and D. B. Karelitz. Interactive color palette tools. *IEEE Computer Graphics and Applications*, 24(3):64–72, 2004.
- [8] N. Mellado, D. Vanderhaeghe, C. Hoarau, S. Christophe, M. Brédif, and L. Barthe. Constrained Palette-Space Exploration. *ACM Transactions on Graphics (TOG)*, 36(4):60:1–60:14, 2017.
- [9] T. Mikolov, K. Chen, G. S. Corrado, and J. Dean. Efficient Estimation of Word Representations in Vector Space. In *Proceedings of the International Conference on Learning Representations (ICLR 2013) Workshop Track*, 2013.
- [10] T. Mikolov, W.-t. Yih, and G. Zweig. Linguistic Regularities in Continuous Space Word Representations. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 746–751. ACL, 2013.
- [11] W. Monroe, N. D. Goodman, and C. Potts. Learning to Generate Compositional Color Descriptions. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 2243–2248, Austin, Texas, 2016. ACL.
- [12] W. Monroe, R. X. Hawkins, N. D. Goodman, and C. Potts. Colors in Context: A Pragmatic Neural Model for Grounded Language Understanding. *Transactions of the Association for Computational Linguistics*, 5:325–338, 2017.
- [13] J. Pennington, R. Socher, and C. Manning. GloVe: Global Vectors for Word Representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1532–1543. ACL, 2014.
- [14] M. Shugrina, J. Lu, and S. Diverdi. Playful Palette: An Interactive Parametric Color Mixer for Artists. *ACM Transactions on Graphics (TOG)*, 36(4):61:1–61:10, 2017.
- [15] M. Shugrina, W. Zhang, F. Chevalier, S. Fidler, and K. Singh. Color Builder: A Direct Manipulation Interface for Versatile Color Theme Authoring. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI ’19*, p. 1–12, New York, NY, USA, 2019. ACM.
- [16] S. Streuber, M. A. Quiros-Ramirez, M. Q. Hill, C. A. Hahn, S. Zuffi, A. O’Toole, and M. J. Black. Body Talk: Crowdshaping Realistic 3D Avatars with Words. *ACM Transactions on Graphics (TOG)*, 35(4):54:1–54:14, 2016.
- [17] TopeconHeroes. FLAT ICON DESIGN, Online accessed: 25-january-2021.
- [18] O. Winn and S. Muresan. ‘Lighter’ Can Still Be Dark: Modeling Comparative Color Descriptions. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, Vol. 2, pp. 790–795. ACL, 2018.
- [19] M. E. Yumer, S. Chaudhuri, J. K. Hodgins, and L. B. Kara. Semantic Shape Editing Using Deformation Handles. *ACM Transactions on Graphics (TOG)*, 34(4):86:1–86:12, 2015.
- [20] C. Zou, H. Mo, C. Gao, R. Du, and H. Fu. Language-Based Colorization of Scene Sketches. *ACM Transactions on Graphics (TOG)*, 38(6):233:1–233:16, 2019.

Knowledge ComposTer: 知識断片の再活用による継続的な創造活動支援システム

生田 泰章^{*} 高島 健太郎[†]

概要. ワードプロセッサやプレゼンテーションツールは、創造活動をするうえで欠かせないものとなっており、これらのツールを使うことで、多くの人は複数のファイルを日々継続的に創造している。このような創造活動において、過去に作り出されたファイルの一部（知識断片）が新たな創造活動時に再活用されることが頻繁に行われている。本稿では、知識断片の再活用を支援することで、継続的な創造活動を支援する Knowledge ComposTer を提案する。Knowledge ComposTer は、メモの蓄積、思考の整理、プレゼンテーションの作成、文書作成をそれぞれ行うことが可能な 4 つのサブシステムが相互に連携し、各サブシステムで作られた知識断片を再活用可能なように一元的に蓄積・管理されている。また、各サブシステムで作り出された知識断片は一度生成されたら漏れなく収集される。さらに、Knowledge ComposTer は知識断片に対する操作をトラッキング可能なようにログを収集するため、知識断片の生成から活用までの一連のプロセスを分析することができる。

1 はじめに

ワードプロセッサやプレゼンテーションツールなどのいわゆるオフィススイートは、オフィスワークの業務遂行に限らず、学生の課題作成や発表などにも使用され、多くの人にとって創造活動をする上で欠かせないものとなっている。人々はこれらのツールを使いこなし、資料作成やプレゼンテーションなどの創造的な活動を日々継続的に行っている。中でも、文書やプレゼンテーション資料の創造活動では、一から目的のファイルが創造されているだけでなく、製作者が自己または他者が過去に作った文書ファイルやプレゼンテーションファイルの一部（以降、知識断片と呼ぶ）を再活用して作成するということが頻繁に行われている [6][7]。本研究では、このような文書およびプレゼンテーションに関わる創造活動を対象とし、経時的に複数のファイルを創造する行為を継続的な創造活動と呼ぶ。

従来、これらの創造活動における知識断片の再活用の現状を支援すべく、過去の異なるファイルで作り出された知識断片の再活用を支援するシステムが提案されている [9][11]。また、近年の Microsoft PowerPoint には、異なるプレゼンテーションファイルに含まれるスライドを再活用可能な機能が実装されている。Google ドキュメントには、Google Keep で作成されたメモを再活用できる機能が実装されており、実サービスにおいても、過去に創造されたファイルに含まれる知識断片をユーザが再活用することで、創造活動を支援する機能が実装されている。

これらの取り組み以外にも、創造活動時の知識断片の再活用について、検討すべき事項が多く残されている。例えば、筆者らはこれまで、創造活動の過程には試行錯誤のプロセスが存在することに着目し、文書作成過程で一旦は書き出されたが、最終的に削除された知識断片（不用知）が、別の文書作成に再活用可能であるという実験結果を得ている [13]。

これらの結果から、あるファイルの創造活動過程中に、一旦は構成要素として作成された知識断片が、その後削除されたかどうかによらず、再活用可能であると推察される。しかし、不用知の活用については文書以外のファイルの種別で検証が行われていない。また、創造活動で再活用される知識断片の特徴（例えば、生成元となったファイルの内容との関係や文章量など）について明らかとなっていないことが多いがあるが、従来、創造活動の種類によって用いられるメディアは異なるため、異なる種類のファイルで作成された知識断片の再活用について、実態を把握することは困難であった。

これらの問題を解決すべく、本研究において創造活動支援システム Knowledge ComposTer を提案する。Knowledge ComposTer は 3 つの特徴を有する。一つ目は、創造活動の種類に応じたメディアが複数連携して構成されている。本稿では、メモの蓄積、思考の整理、プレゼンテーション作成、文書作成が可能なように、複数のサブシステムが連携している。二つ目は、各創造活動で生成された知識断片をすべて蓄積し、別の創造活動時に活用可能に構成している。三つ目はユーザが知識断片に対して行った操作をログとして記録可能に構成している。これらの特徴を有することで、異なる創造活動間で生成された知識断片を活用することができ、かつ、操作

Copyright is held by the author(s).

* サイボウズ・ラボ株式会社

† 北陸先端科学技術大学院大学 先端科学技術研究科

ログが記録されることで、これまで実態を把握することが困難であった、知識断片の生成から活用までの一連のプロセスを分析することが可能となる。

2 関連研究

2.1 ファイルの再活用状況の調査

文書ファイルやプレゼンテーションファイルの創造活動時において、別ファイルの一部が再活用されることが調査によって明らかになっている。Mejovaら[7]は、企業内におけるプレゼンテーションの再活用に関する調査を多角的に行っている。自己や他者のプレゼンテーションが再活用されることを明らかにしたうえで、プレゼンテーション中の何が再活用されるかや、再活用を妨げる要因等の調査を行った。この調査では、プレゼンテーション中のどのパートが再活用されるかについて調査を行っているが、プレゼンテーションと異なるファイルへの再活用については対象外である。Jensenら[6]は、9人の実験参加者について、平均8.6週間でMicrosoft Office製品をどのように使ったかを調査し、異なるアプリケーション間で情報がやり取りされていることを明らかにした。文献[6]では、各Office製品のファイルに対する操作をトレース可能なTaskTracerというシステムを導入することで、異なる製品間での情報のやり取りをトレースしているが、やり取りされる情報が何であったかの詳細は分析できていない。本研究では、異なるファイル間における知識断片の再活用をより詳細にトラッキングできるようにシステムの設計・実装を行う。

2.2 ファイルの再活用によって創造活動を支援するシステム

創造活動の過程で、別のファイルの内容を再活用する支援によって、創造活動を支援するための取り組みがいくつかある[3][9][10][11]。文献[3]では、新たな文書作成のために、過去の文書を断片化している。ALOCOMフレームワーク[11]は、文書、SCORM、プレゼンテーション作成のために、過去のこれらの成果物を断片化している。プレゼンテーションの再活用の機能としては、Microsoft PowerPointのスライド再利用機能や、ConRep[9]がある。これらは、すでに完成しているファイルを断片化して同種のファイルの創造活動時に再活用を促すシステムである。Scraps[10]は、文書作成のために別の文脈で作成された画像付きメモを再活用する仕組みを提案し、Googleドキュメントは、あらかじめ蓄積されたGoogle Keepのメモ書きをインポートすることができ、文書作成の意図の有無を問わず作成されたメモを、文書作成に再活用する機能である。また、Microsoft PowerPointには、Microsoft Wordで作成したアウトラインをインポートする機能も存

在する。これは、プレゼンテーションの作成のために、Wordで作成したアウトラインを活用する機能である。

本研究では、これらの研究を包含すべく、システムの設計・実装を行う。つまり、本研究では、上記のシステム同様、ファイル作成時に異なる種類・内容のファイルで生成された知識断片を余すことなく活用可能な要件を検討する。

2.3 不用知の活用に関する研究

文献[13][14]では、文書作成時において、一旦は成果物の候補になったが、最終的には不採用となって削除された文章の活用可能性が検討された。その結果、たとえあるファイルの創造活動時で不採用になった知識断片（不用知）であっても、別のファイルの創造活動時には再活用できることが示唆された。この不用知は、2.1節、2.2節のどの従来研究においても見過ごされてきており、再活用の実態を把握することができれば、不用知を新たな知的資源とすることができる。しかしながら、文献[13][14]で提案のText ComposTerは、单一文書で作成された不用知を収集する機能は存在するが、別文書作成時に活用する機能はなく、より包括的に再活用の実態を把握するためには、不用知を含む知識断片を複数文書間で再活用可能なシステムが必要となる。そのため、本研究では、不用知を積極的に収集し、再活用可能とするように、一旦生成された知識断片を全て蓄積し、ユーザが後に再活用可能のようにシステムの設計・実装を行う。

3 Knowledge ComposTer

3.1 設計方針

本節では、提案システムであるKnowledge ComposTerを実装するにあたり、複数種のファイルの創造活動時に知識断片を漏れなく収集し、異なるファイルの創造活動時に過去に創造された知識断片を再活用可能となるようなシステムの設計について検討を行う。

まず、創造活動時に、その種類に応じて適切なメディアを用意し、創造活動を行う者の思考の外在化を支援することで、より多くの知識断片を収集することができ、後の新たな創造活動時に再活用されやすくなる。そのため、提案システムは、複数のメディアによって外在化を支援し、各メディアで作られた知識断片を収集するというアプローチをとる。また、本研究では、知識断片の構造化の度合いに応じた次の3つの創造活動に焦点を当て、それぞれに適したメディアを連携させる。

- (1) 文書・プレゼンテーションの創造活動
- (2) 思考の整理
- (3) アイデアや思考の外在化

Knowledge ComposTer: 知識断片の再活用による継続的な創造活動支援システム

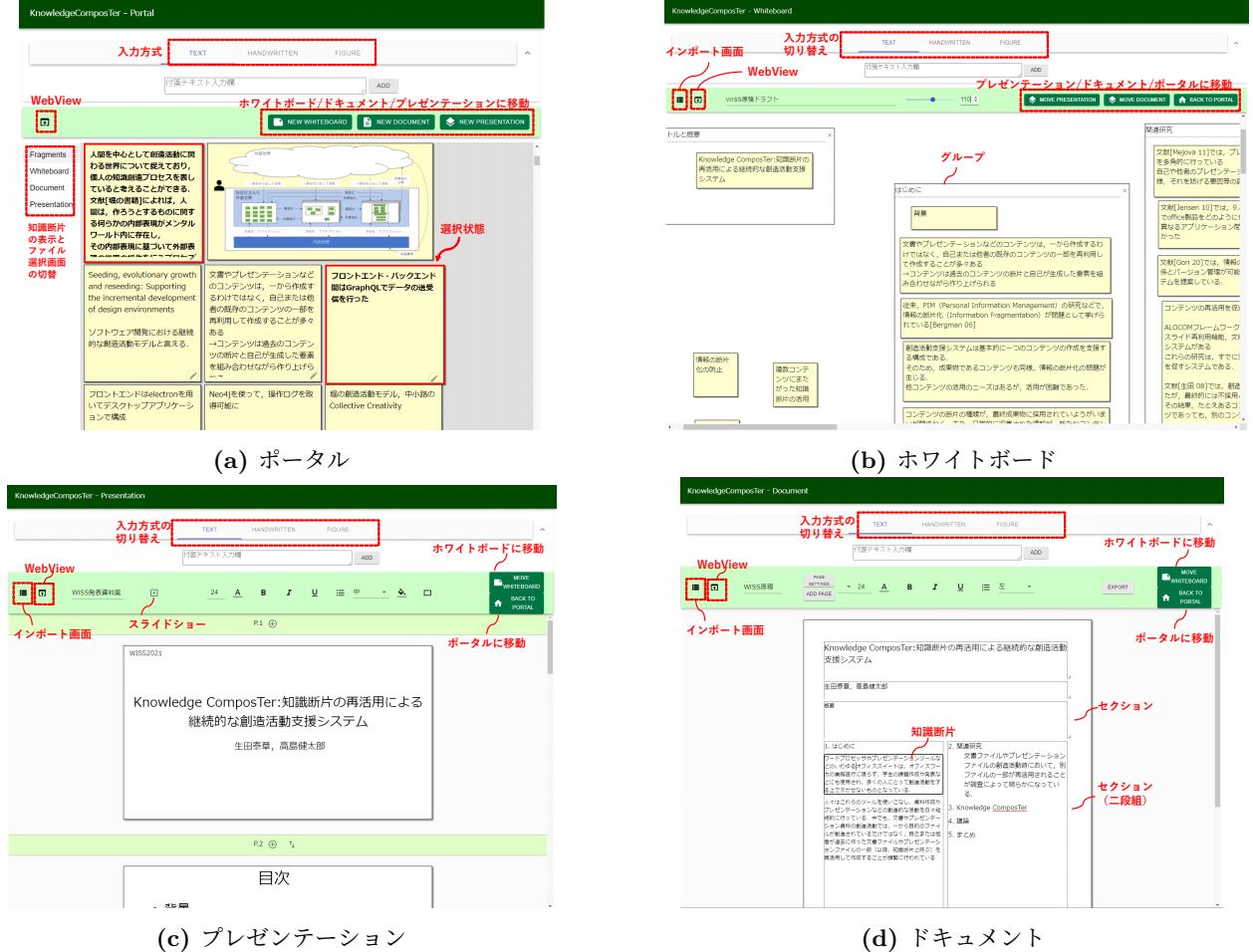


図 1: Knowledge ComposTer スクリーンショット

(1) 文書・プレゼンテーションの創造活動は、文書またはプレゼンテーションの規定のフォーマットに従うように、知識断片の創造・加工・整形等を行う創造活動である。本稿では、中小路ら [12] のアプローチによってメディアを構築する。つまり、これらの創造活動においては、文書やプレゼンテーションファイルを様々な粒度の一塊からなる知識断片で構成されているとみなし、ユーザが一塊ごとに編集、再編集したり、順序や全体の置き場所を直感的に操作可能なメディアとする。

(2) 思考の整理は、個人が自己のメンタルワールド内にある思考を知識断片として外在化し、外在化された知識断片の関係性を見出すことによって思考を整理することを指す。(1)の文書・プレゼンテーションの構造化とは異なり、所定のフォーマットに従うように知識断片を整形する必要はない。むしろ、知識断片間の関係を自由に表現できるメディアが望ましい。この思考の整理においては、従来、デジタルノートやマインドマップ、アイデアプロセッサ、KJ 法支援システム [15] など、知識断片を二次元空間上に自由配置可能なツールがメディアとして用いられ

てきた。本研究も、これらに従う。

(3) アイデアや思考の外在化は、特段の構造化を意図しない、個人のメンタルワールド内にある思考の外在化を指す。ブレインストーミングのようなアイデア創造を意図した活動であってもいいし、ふと思いついた散発的な思考の外在化であってもいい。この創造活動においては、知識断片間に関係を見出さないインターフェースが用いられてきた。例えば、メモアプリや To Do リストのような一覧表示可能なインターフェースである。本研究も、これに従う。

また、上記の各創造活動において知識断片を削除する操作が行われると、収集する知識断片に漏れが出てしまう。そこで、本研究では、(1)-(3) の創造活動で一旦外在化された知識断片を全て収集するというアプローチをとる。知識断片に対する各種の操作(編集・分割・統合等)についても、操作前の知識断片とは別の知識断片とすることで、より包括的に知識断片を収集する。

次に、創造活動時に用いられる知識断片の活用方法について検討する。収集された知識断片の蓄積場所が、生成元のファイル別に管理されていたとき

に、再活用したい知識断片を含むファイルを見つける必要がある。しかしながら、多くのファイルが生成された場合、情報の断片化 [4] が生じ、再活用したい内容を見つけるのは非常に困難となる。そこで、本研究では、Content ReuseにおけるSingle Sourcing[1][2]に倣い、共通の知識断片の保存領域（以下、共通リソース）に一元的に蓄積・管理するアプローチをとる。また、創造活動時に、共通リソースに蓄積された知識断片を参照可能に構成することで、再活用を促進するアプローチを採用する。

以上の設計方針で Knowledge ComposTer の実装を行うことで、知識断片を漏れなく収集し、収集した知識断片を創造活動時に活用可能にすることを目指す。

3.2 システムの概要

Knowledge ComposTer は、テキスト、手書き画像、ファイルシステムからインポートした画像を、知識断片として生成・蓄積し、これらを組み合わせることで、二次元空間への配置を可能にし、さらにプレゼンテーションおよび文書を作成可能なシステムである（図 1）。本研究においては、Knowledge ComposTer は、Electron を用いたデスクトップアプリケーションとして構成した。また、ユーザが作成した知識断片のデータと、知識断片やファイルに対する操作ログとが、グラフデータベースである Neo4j に蓄積されるように実装した（3.4 節参照）。

Knowledge ComposTer は、ポータル、ホワイトボード、プレゼンテーション、ドキュメントの4つのサブシステムが相互に連携して構成されている（図 2）。ポータルサブシステム（図 1a）では、テキスト、手書き画像、画像の入力・蓄積をすることでき、付箋状の知識断片を一覧することができる。3.1 節における、(3) アイデアと思考の外在化のためのメディアに相当する。また、本研究では、ポータルを共通リソースとしても機能させるように実装した。すなわち、ホワイトボード、プレゼンテーション、ドキュメントで作成・編集された知識断片が全て表示される。また、ポータルは、ホワイトボード、プレゼンテーション、ドキュメントで作成された各ファイルを一覧、移動することができる。

ホワイトボードサブシステム（図 1b）では、付箋状の知識断片を生成し、それらを二次元平面上に自由配置、グルーピングすることができる。3.1 節における(2) 思考の整理のためのメディアに相当する。ユーザはホワイトボードを用いることで、思考を外在化し、それらを空間上に配するインタラクションを行うことで、思考の整理を行うことができる。

プレゼンテーションサブシステム（図 1c）では、プレゼンテーションファイルの構成要素として知識断片を生成し、スライド内に配置し、そのスライドを線形に並べることが可能である。つまり、プレゼン

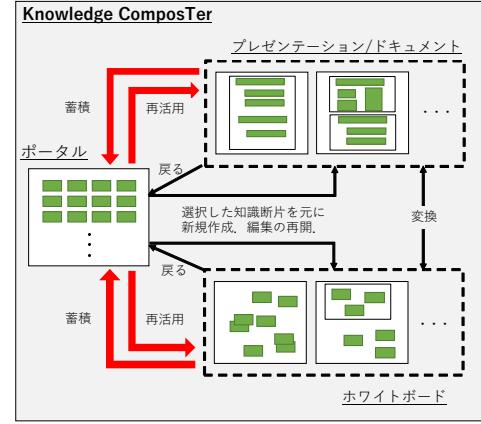


図 2: Knowledge ComposTer の各サブシステム間の関係（黒矢印はサブシステムの遷移、赤矢印は知識断片の流れ）



図 3: ホワイトボードサブシステムでのインポート画面の表示

ンテーションサブシステムは、複数の知識断片をプレゼンテーションのフォーマットに整形し、プレゼンテーションファイルを作成することを支援するメディアであり、3.1 節の(1) 文書・プレゼンテーションの創造活動のためのメディアに相当する。なお、本稿では、作成したプレゼンテーションファイルについて、スライドショー機能を用いることで、プレゼンテーションを行うことができる実装（Reveal.jsを利用）を行った。

ドキュメントサブシステム（図 1d）では、文書ファイルの構成要素として知識断片を生成し、ページ内に配置し、そのページを線形に配置することで文書ファイルを作成することができる。本研究では、知識断片を配置可能な複数のセクションをページ内に配置し、そのセクション間で知識断片を移動させようとするドキュメントサブシステム構成することで、ナレッジインタラクション [12] を実現可能にしている。プレゼンテーションサブシステム同様、ドキュメントサブシステムは、複数の知識断片を文書のフォー

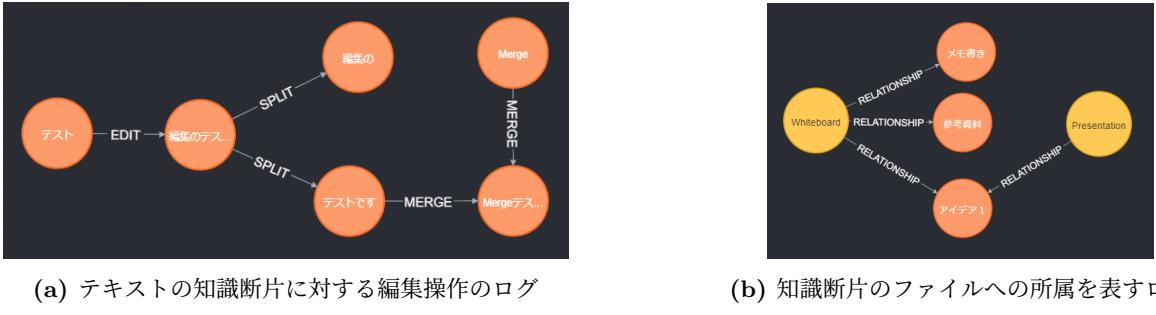


図 4: 操作ログを表すグラフ

マットに整形し、文書作成を支援することを可能にするメディアであり、3.1節の(1)文書・プレゼンテーションの創造活動のためのメディアに相当する。

また、Knowledge ComposTer は、知識断片を扱う中小路の ART#001[8] や Text ComposTer[14] と同様に、生成された知識断片の編集、分割、統合する機能を、各サブシステム共通で備えており、各作業空間における創造活動を支援している。ここで、ホワイトボード、プレゼンテーション、ドキュメントの各サブシステム内の知識断片は削除可能であるが、ポータルは上述のように共通リソースとして機能させているため、知識断片を削除できないように構成している。

3.3 サブシステム間の連携

Knowledge ComposTer は、図 2 に記載のように、対応するサブシステム間で連携し、知識断片を相互にやり取りすることを可能にしている。

ポータルサブシステムは、ホワイトボード、プレゼンテーション、ドキュメントの各サブシステムと連携している。ポータルサブシステムにおいて、ユーザは知識断片を複数選択した状態で（図 1a 参照）、各サブシステムに移動することができる。この移動によって、ポータルで選択された複数の付箋をもとに、新たなホワイトボード、プレゼンテーション、文書の作成が可能になる。また、ホワイトボード、プレゼンテーション、ドキュメントの各サブシステムは、インポート画面を重畠表示可能である（図 3）。インポート画面には、ポータル内の知識断片のうち、表示元のサブシステムに配置されていない全ての知識断片が表示されている。ユーザは、インポート画面内のインポートボタンによって、ポータル内の知識断片を再活用することが可能となる。また、上述のように、ホワイトボード、プレゼンテーション、ドキュメントの各サブシステムでは、新たに生成された知識断片は全て、共通リソースとして機能しているポータルサブシステムに蓄積される。

Knowledge ComposTer では、ホワイトボードサブシステムで二次元配置された知識断片を、プレゼンテーションサブシステム、ドキュメントサブシステムそれぞれに移行させることができる。また、プ

レゼンテーションサブシステム、ドキュメントサブシステムからそれぞれ、ホワイトボードサブシステムに移行させることができる。現段階では、ホワイトボード内のグループが、プレゼンテーションにおけるスライドに、ドキュメントにおけるセクションに相当するように移行させている。

3.4 操作ログの記録

Knowledge ComposTer は、知識断片に対する操作のログをグラフデータとして記録している。具体的には、知識断片、ファイル、グループ、スライド、ページ、セクションをノードとし、知識断片と知識断片を含むその他のノードとの関係をエッジとして表現している（図 4）。図 4(a) は、テキストの知識断片が複数回にわたって編集操作が行われたログの例であり、「テスト」という文字列から「Merge テストです」という文字列に知識断片が編集された過程が示されている。具体的には、一番左側のノードが起点となり、一度内容が編集（EDIT）され、その後、2つの知識断片に分割（SPLIT）されている。そして、分割された一方の知識断片と別の知識断片とが、一つに統合（MERGE）される。

図 4(b) は、ホワイトボードファイルとプレゼンテーションファイル内にある知識断片の関係を表したグラフである。この例では、ホワイトボードに3つの知識断片が所属しており、そのうち一つがプレゼンテーションにも所属していることを表している。つまり、このグラフからは、プレゼンテーション（またはホワイトボード）で生成された知識断片がホワイトボード（またはプレゼンテーション）で再活用されたことが分かる。実際は、各知識断片のノードには、複数のプロパティが保存されており、どのファイルで生成されたかが記録されているため、どのファイルで生成された知識断片がどのファイルで再活用されたかということが、ログから判断できる。

以上のようなログを収集することで、例えば、「不知として削除された知識断片がその後、どのファイルに活用されたか」や、「知識断片それぞれについて、いくつのファイルで再活用されたか」など、知識断片の再活用の状況をトラッキングすることができる。そのため、Knowledge ComposTer を用いて

継続的に創造活動が行われる設定の実験を行うことで、創造活動時における知識活用について、明らかでなかった知見（例えば、再活用されやすい知識断片の内容の特徴）を得ることが期待できる。

4 議論

4.1 継続的な創造活動を支援できるか

3章で述べたように、Knowledge ComposTerは、創造活動の種類によって異なるメディアを有するサブシステムが相互に連携し、各サブシステムで外在化した知識断片を漏れなく収集している。継続的な創造活動では、人間は異なるメディアを使って複数のファイルを創造する。例えば、研究活動を行っていると、研究室内や所属部署内などで進捗報告や関連論文に関する勉強会の資料をプレゼンテーションで作成することはよくある。その後、研究の成果がまとまってきたときに、論文にまとめる作業を行う。このようなケースにおいて、Knowledge ComposTerを使わない場合は、文書作成ツールで論文を執筆しながら、関連するプレゼンテーションファイルを開き、内容を再活用することが想定される。一方、Knowledge ComposTerを使用する場合、ドキュメントサブシステムで文書作成を行いながら適宜インポート画面を開き、過去にプレゼンテーションサブシステムで作った知識断片のうち、再活用したいものを探し出してインポートを行う。この例において、Knowledge ComposTerを使用していない場合は、再活用したい内容を含むファイルを探し出す必要があるが、Knowledge ComposTerであれば、その内容をインポート画面で即座に確認することができる。

ここで、従来、Gitによるファイルの差分管理やFileWeaver[5]のようなファイル管理システムが提案されており、これらは逐次進捗があるファイル群の関係性を把握可能とし、継続的な創造活動を支援する仕組みであると言える。しかしながら、これらの仕組みであっても、過去の数バージョン前にどのような内容が含まれていたかは、そのファイルを直接参照する必要があり、結局は目的のファイルを探し出す必要がある。また、これらのファイル管理システムでは、各ファイルの編集中に削除された内容が記録されることはない。数バージョン先の未来において、削除された知識断片を活用したいというニーズには応えることができない。つまり、これらの仕組みでは、知識断片の活用機会を失っているかもしれない。Knowledge ComposTerは、一旦生成された知識断片は全て蓄積しており、さらに、バージョン間の違いによらず、過去の知識断片を一覧表示可能なことから、従来のファイル管理システムでは難しかった知識断片の活用に関する要求に応えることができる。以上の議論から、Knowledge ComposTerの仕組みとしては、継続的な創造活動支援が可能で

あると思われる。

4.2 想定される課題

知識断片の再活用において、共通リソースによって知識断片を一元管理することがメリットであることを述べた。しかしながら、Knowledge ComposTerは、現状、知識断片に対する編集・分割・統合を行った場合、それぞれ別の知識断片が生成されたとして、それらを全て蓄積し、ポータルにて表示する実装を行っている。そのため、蓄積・表示すべき知識断片の数が多くなり、活用したい知識断片を探し出すのが困難になるおそれがある。例えば、本稿第一著者が、Knowledge ComposTerを使って社内の勉強会用に作成したプレゼンテーションファイルについて、スライド40枚に対して知識断片の数は121個あった。そして、削除された知識断片は96個あった。第一著者は、この資料を作成する際に、あまり試行錯誤を要さず、一気通貫して資料作成を行ったという印象を持っているが、このファイル作成時に生成された全知識断片のうち、削除された知識断片が約4割を占めていた。試行錯誤が伴うファイルの作成に関しては、削除された知識断片の方が数が多くなることも想定される。

のことからも、Knowledge ComposTerに蓄積される知識断片数が非常に多くなることが考えられる。その場合、現状のポータル画面やインポート画面に表示される知識断片数の数が多くなり、再活用が阻害される可能性がある。そのため、今後、知識断片の検索機能や似通った内容の知識断片をまとめて表示するなどのインターフェースの工夫が必要になってくると思われる。

5 まとめと今後の課題

本稿では、創造活動時に一度でも外在化した知識断片を漏れなく収集し、かつ新たな創造活動時に過去に収集した知識断片を再活用可能にすることで、継続的な創造活動を支援するシステム Knowledge ComposTerを提案した。創造活動の種類に応じて適したメディアを相互連携したKnowledge ComposTerを使用することで、ユーザは、各メディアによって創造活動自体が支援されながら、後に活用可能な知識断片を自然に収集・蓄積することが可能となる。そして、ユーザが新たな創造活動時に蓄積された知識断片を活用することで、その創造活動が支援される。

今後は、実証実験を行いながら、創造活動に資する知識断片とはどのようなものかを明らかにしていくとともに、4章で議論したような、潜在的な課題を検討し、Knowledge ComposTerを改良することで、知識創造と知識活用のサイクルをより円滑にし、人間の継続的な創造活動を支援するメディアの創造を目指す。

参考文献

- [1] K. Ament. *Single Sourcing: Building Modular Documentation*. William Andrew Publishing, 2003.
- [2] C. C. Ann Rockley. *Managing Enterprise Content: A Unified Content Strategy, Second Edition*. New Riders, 2012.
- [3] D. Barta and J. Gil. A System for Document Reuse. In *Proceedings of the 7th Israeli Conference on Computer-Based Systems and Software Engineering, ICCSSE '96*, p. 83, USA, 1996. IEEE Computer Society.
- [4] O. Bergman, R. Beyth-Marom, and R. Nachmias. *The Project Fragmentation Problem in Personal Information Management*, p. 271–274. Association for Computing Machinery, New York, NY, USA, 2006.
- [5] J. Gori, H. L. Han, and M. Beaudouin-Lafon. FileWeaver: Flexible File Management with Automatic Dependency Tracking. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology, UIST '20*, p. 22–34, New York, NY, USA, 2020. Association for Computing Machinery.
- [6] C. Jensen, H. Lonsdale, E. Wynn, J. Cao, M. Slater, and T. G. Dietterich. *The Life and Times of Files and Information: A Study of Desktop Provenance*, p. 767–776. Association for Computing Machinery, New York, NY, USA, 2010.
- [7] Y. Mejova, K. De Schepper, L. Bergman, and J. Lu. *Reuse in the Wild: An Empirical and Ethnographic Study of Organizational Content Reuse*, p. 2877–2886. Association for Computing Machinery, New York, NY, USA, 2011.
- [8] K. Nakakoji, Y. Yamamoto, S. Takada, and B. N. Reeves. Two-Dimensional Spatial Positioning as a Means for Reflection in Design. In *Proceedings of the 3rd Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques, DIS '00*, p. 145–154, New York, NY, USA, 2000. Association for Computing Machinery.
- [9] M. Sharmin, L. Bergman, J. Lu, and R. Konuru. On Slide-Based Contextual Cues for Presentation Reuse. In *Proceedings of the 2012 ACM International Conference on Intelligent User Interfaces, IUI '12*, p. 129–138, New York, NY, USA, 2012. Association for Computing Machinery.
- [10] A. Swearngin, S. Iqbal, V. Poznanski, M. Encarnación, P. N. Bennett, and J. Teevan. Scraps: Enabling Mobile Capture, Contextualization, and Use of Document Resources. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, CHI '21*, New York, NY, USA, 2021. Association for Computing Machinery.
- [11] K. Verbert, X. Ochoa, and E. Duval. The ALOCOM framework : towards scalable content reuse. *Journal of Digital Information*, 9(1):1–24, 2008.
- [12] 中小路 久美代, 山本 恭裕. 創造的情報創出のためのナレッジインタラクションデザイン. 人工知能学会論文誌, 19(2):154–165, 2004.
- [13] 生田 泰章, 高島 健太郎, 西本 一志. 廃却文章断片の創造的文章作成時における活用可能性の検証. 情報処理学会研究報告. HCI, 研究報告ヒューマンコンピュータインタラクション, 2018-HCI-176(17):1–8, 2018.
- [14] 生田 泰章, 高島 健太郎, 西本 一志. 文書作成過程で削除された文章断片の効率的収集手段と活用可能性に関する考察. 情報処理学会論文誌, 59(12):2299–2314, 2018.
- [15] 由井薗 隆也, 宗森 純. KJ法支援グループウェアの現状と今後(ヒューマンコミュニケーション基礎). 電子情報通信学会技術研究報告 = IEICE technical report : 信学技報, 115(237):61–68, 2015.

未来ビジョン

(この未来ビジョンは、本稿第一著者が見て
いるビジョンです。)

創造活動支援システム研究の方向性として、創り出すことのハードルを下げるための仕掛けを考案し、実装・評価するというのが王道ではないかと思う。一方で、創造活動時の生みの苦しみは、創造活動のメディアが変わっても、多かれ少なかれ存在し続けるのではないかとも思っている。その苦しみの過程で創り出されたものは、成果物であろうが、成果物に反映されない中間生成物であろうが、せっかく生み出されたのだから余すところなく活用されてほしい（もったいない！）。その活用を支援することで、今は大変でも、いつかは未来の自分を含む誰かの役に立つはずだと、挫折することなく創造活動を続ける人が増えてほしいという想い

を持っている（著者がそうありたい）。本研究のKnowledge ComposTerという名は、知識を創り出す（Compose）だけでなく、過去の創り出す過程をたい肥（Compost）に変えて、創造活動の土壌を豊かにしたいという想いを込めて名付けた。本稿で実装したKnowledge ComposTerは完成形ではなく、研究の進展によって、適宜改良していく。今後は、創造活動のドメインを増やす方向と、活用対象者を増やす方向を検討している。本稿では、文書とプレゼンテーションの成果物を生み出す環境を構築したが、音楽やスケッチなどの創造活動のドメインを増やした場合に、創造活動がどう変化するのか、もしくはしないのかを明らかにしていきたい。また、個人だけでなく、研究室や、組織単位で創造活動のための知識活用がどのようにになっているかをKnowledge ComposTerを使用することで、明らかにしていきたい。

IoTeach: 実世界と順序型コンテンツを連携した IoT 学習支援システム

川谷 知寛 * 塚田 浩二 * 栗原 一貴 †

概要. 近年, IoT が一般化している中で, Arduino や M5Stack などの IoT デバイスの動作や仕組みを他者と共有する場面が増えてきている。さらにコロナ禍に伴い, 各教育機関においても, チュートリアル動画等を用いたオンデマンド型や, ビデオ会議システム等でのスライド共有を中心とするライブ型のオンライン講義の導入が進められている。一方, こうした手法は実世界のデバイスやソースコードとのリンクに乏しい等の問題があった。そこで本研究は, Web 上で共有される動画やスライドといった順序型コンテンツにスクリプト言語を同期させ, 実世界の IoT デバイスや編集可能なソースコードと連携させることで, IoT プログラミング学習を支援するシステム「IoTeach」を提案・開発する。本稿では, システムのコンセプトと実装を説明し, コンテンツ事例や課題と展望について議論する。

1 はじめに

IoT (Internet of Things) という言葉に代表されるように, 身の回りのモノとインターネットを連動したシステムが一般化しつつある。さらに, IoT 機器のプロトタイピングを支える Arduino, Raspberry pi, M5Stack などの要素技術も充実してきており, その教育ニーズも日増しに高まっている。しかし, 試作した IoT デバイスの振る舞いや仕組みを, 空間的・時間的にその場にいない他者と共有することは難しく, 実世界での振る舞いをわかりやすく示すために, チュートリアル動画がしばしば採用されている。例えば, M5Stack 社は自社 IoT デバイスのチュートリアル動画を多数公開している¹。

一方, チュートリアル動画は従来のドキュメントに比べると, ソースコード等の実際にデバイスを動作させる要素とのリンクに乏しいという問題がある。ユーザは動画を見てデバイスの振る舞いを理解することはできても, 手元の実機で動作を再現するのは容易ではない。ソースコードのどの部分がその挙動を起こしているかは動画では表現しにくいため, 従来のドキュメントを改めて参照してソースコードを探し, 専用の開発環境から手元のデバイスに転送する必要がある。すなわち, チュートリアル動画は特に初心者のユーザにデバイスの振る舞いを伝える「概要把握」に優れる反面, 実機での「動作確認」や仕組みの「詳細理解」には, 従来型ドキュメントやソースコードをユーザが自力で参照する必要があり, 学習プロセスとしてみるとギャップがあった(図1)。

さらに, コロナ禍に伴い各教育機関ではプログラミング演習のオンライン化が急速に進められており, オンラインでのプログラミング教育支援の必要性が増加している。オンライン授業は主に, (チュートリ

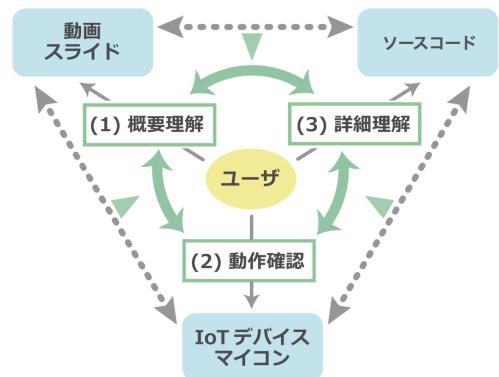


図 1. 本研究の狙い

アル動画の例に近い) 動画等を学習者の好きなタイミングで視聴するオンデマンド型と, ビデオ会議システム等を活用したライブ型に分けられる。ライブ型の講義においても, 講師は学習者の手元にある計算機や IoT デバイスの状態の把握や制御が難しい等の問題を抱えている。

そこで本研究では, Web 上で共有される動画・プレゼンテーション資料などの順序型コンテンツに対して, 進行に対応して実行されるスクリプト言語を同期させることで, スクリプトと視聴者の手元の IoT デバイスとを連携可能なシステム「IoTeach」を構築する。例えば, 図2のような動画上のデバイスの動きに対応して, 実世界のデバイス操作とソースコードの提示を行うことができる。IoTeach を用いることで, 視聴者のデバイスに関する理解度の上昇やモチベーションの向上といった効果を期待する。

* 公立はこだて未来大学

† 津田塾大学

¹ <https://www.youtube.com/c/M5Stack/>

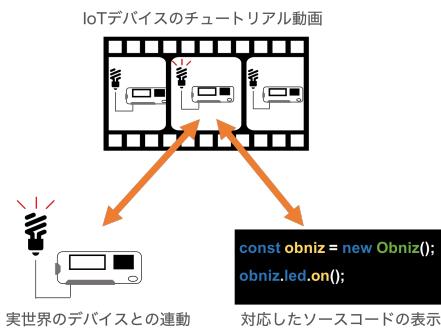


図 2. IoTeach の概要. Web 上の動画やスライド等の順序型コンテンツの進行に対応したスクリプトを同期させることで、実世界の IoT デバイスの操作や関連するソースコードの提示を行う.

2 関連研究

2.1 コンテンツにスクリプトを組み込む事例

順序型コンテンツの特定時刻にスクリプトを実行することは、Adobe Flash の時代から広く用いられてきたが、その用途は主にアニメーションへのインタラクティブ性の付与等に限定されていた。Sikuli[5] は、プログラムに画像を使用して、自動化補助や検索を行えるようにしたシステムである。

Text Alive[4] は、解析された音楽情報を基にした動画アニメーションの作成を行う Web サービスである。作成された動画はプログラミング要素を含んでおり、字幕の表示設定や図形の描画などをプログラムを書いてユーザ自身が実装できる。

ぴぴつい [6] は PowerPoint のスライドショーに連動して、事前に発表者ノートに記述した内容をツイートする PowerPoint のアドインである。

2.2 コンテンツと実世界デバイスを連動する事例

Songle Sync[3] は、音楽を元に作成された動画を実世界の様々なデバイスで制御して、音楽と物理空間が融合した体験を生成する。

また、IoT デバイスと Web を連携する仕組みとしては、obniz Board²といったプラットフォームがある。これは電気回路と直接接続可能なマイコンであり、クラウドを介して、Web 上からソースコードの記述や実行が可能なソフトウェアを備えている。

2.3 オンラインプログラミング教育支援の事例

Progate[1] やドットインストール [2] は、オンラインで様々なプログラミング言語を学ぶことができる Web サービスである。スライドや動画での学習をした後に、サイト上でコードを書いて実践的に学ぶことができる。しかし、これらのサイトで IoT デバイスについて学習することはできない。

² <https://obniz.com/>

また、コロナ禍に伴い大学でのプログラミング演習のオンライン化も急速に進められており、オンラインでのプログラミング教育支援の必要性も増えている。Typing.run[9] は、プログラミング学習のタイピング練習と基本命令や関数の理解と定着の支援を行う Web システムである。このシステムによって、利用者のタイピングスコアの上昇やプログラムの理解や定着に良好なことが確認された。Amadeus[8] は、学習者側だけではなく講師側の支援も行ったプログラミング初心者のためのオンラインコンパイラシステムである。学生側の環境構築の手間を省き、同時に講師は学生が困っている時にすぐに気づくことができるようになっている。

2.4 本研究の特徴

このように、順序型コンテンツへのスクリプトの同期や、Web と実世界のデバイスを連動させる仕組みは存在していたが、両者を統合した学習支援の試みはほとんど行われてこなかった。また、オンラインプログラミング学習支援システムは多数存在するが、IoT デバイスに焦点を当てた支援はあまり行われてこなかった。本研究では、動画やスライド等の順序型コンテンツにスクリプトを同期させ、コンテンツ内のデバイスの動作を実世界のデバイスやソースコードと連携させることで、IoT プログラミング学習を支援する点で独自性がある。

3 提案

3.1 コンセプト

図 1 をもとに、IoTeach のコンセプトを説明する。大きく分けて、Web 上の要素である順序型コンテンツ（動画／スライド）とソースコード、実世界の要素である IoT デバイス（マイコン）の 3 つの要素がある。ここで、コンテンツの順序（例：動画の時間軸）に基づいて、ソースコードの表示や実世界の IoT デバイスの挙動を連携して制御する。ユーザは、動画内のデバイスの動作と連動したソースコードを見る能够となるため、制御の詳細を理解しやすくなる。また、動画を見るだけで手元のデバイスが動作する体験は、特に初学者ユーザの学習意欲の向上等に繋がると考える。

システムの利用例を図 3 に示す。ここでは、ユーザは PC 等で IoT マイコンのチュートリアル動画を視聴している。動画内のマイコンの LED が光ると、ユーザの持つマイコンの LED も連動して光る。さらに、対応するソースコードが画面右側のエディタ上に提示される。

3.2 システム要件

本研究では、順序型コンテンツ（動画／スライド）の視聴時において、コンテンツとデバイスの連動、

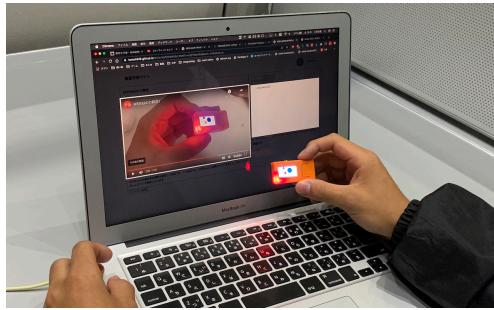


図 3. IoTeach の利用例

及び、コンテンツとソースコードのスムーズな連携ができるシステムの構築を目指す。そこで、以下の3点のシステム要件を設定した。

ユーザの手元でデバイスが動作

動画／スライド上でデバイスの挙動を見るだけでなく、実際に手元のデバイスが同じ挙動をしたり、ユーザによる入力で動くことで、理解を深める。例えば、センサの感度やアクチュエータの駆動範囲等を体験的に理解することができる。また、動画を見るだけで手元のデバイスが動くという体験は、特に初学者の学習意欲の向上につながると考える。

デバイスの挙動に合わせたソースコードの表示

前述した通り、動画／スライド上のデバイスの動作がソースコードのどの部分で実現されているか確認するのは難しい。そこで、動画上等でデバイスが動いたタイミングで、Web 上のテキストエリアに対応するソースコードをコピー等の編集操作が可能な状態で表示することにより、動画とソースコードのスムーズな連携を行う。

注目箇所の誘導

IoT プログラミング学習では、コンテンツ（動画／スライド）、ソースコード、手元のデバイス等、複数の要素を意識して学習する必要があり、どこを見ればよいのか、混乱が生じやすい。そこで、システムで注目すべき要素を強調表示することで、ユーザの注意を誘導する仕組みを構築する。

3.3 動画版システムとスライド版システム

本研究では、オンデマンド教育／ライブ教育を見据えて、動画／スライドの2種類の順序型コンテンツに対応したシステムを構築する。なお、学習者は主に、マイコンや IoT 開発の経験がない初学者を想定する。

動画版システムは、YouTube 等のオンライン動画を Web アプリ上に埋め込み、動画の時間軸に沿って処理を実行するスクリプトを連動させる。学習者

は原則一人でシステムを利用するため、一つのアプリ（インスタンス）に対して、一つの IoT デバイスが対応する。オンデマンド教材以外にも、個人やハッカソン等で制作した IoT 作品の公開／共有にも活用できると考える。

スライド版システムは、PDF 等のページ単位を持つスライドをオンラインに配置し、ページ遷移に沿って処理を実行するスクリプトを連動させる。主にライブ型オンライン授業でビデオ会議システムと併用し、一人の講師が複数の学習者に対して、スライドを画面共有等で見せながら進行する形態を想定している。よって、一つのアプリ（インスタンス）に対して、各学習者の持つ複数の IoT デバイスが対応する必要がある。

4 実装

3.3 節で述べたとおり、本研究では動画版システムとスライド版システムの2種類を開発する。本システムのシステム構成図を図 4 に示す。両システムとも、Web アプリとして実装した。大まかな実装技術としては、HTML, CSS, JavaScript 等を用いてフロントエンドを構築し、Firebase³を用いて、ホスティングとデータベースの管理を行っている。以下、利用する IoT デバイスとスクリプトについて紹介した後に、動画版／スライド版システムについて述べる。

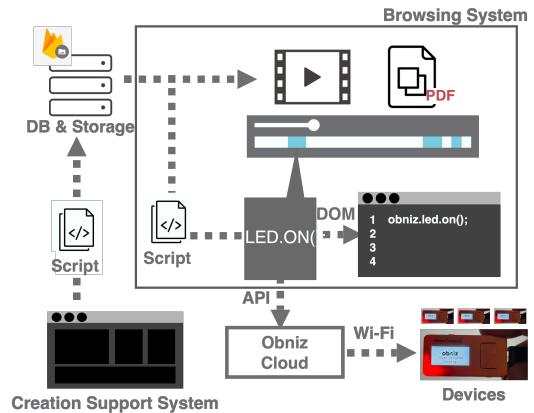


図 4. システム構成図

4.1 IoT デバイス

システムで利用する IoT デバイスの要件としては、

1. 外部のセンサやアクチュエータを接続する拡張性を持つ。
2. Wi-Fi 等でインターネットに接続できる。
3. インターネット経由でプログラムを実行／書き換えできる

³ <https://firebase.google.com/>

といった点が上げられる。これらの要件は、ESP32マイコンや Arduino+Wi-Fi シールド等でも達成可能であるが、3 点目の実現にはファームウェア開発等の手間が掛かる。

そこで、本研究では obnizOS が導入されたマイコン⁴を利用することにした。obnizOS は、IoT マイコン上で動作するファームウェアであり、デバイス毎にユニークな ID を備えている。クラウドの API から ID を指定することで、特定のデバイスに対して、内蔵センサ／アクチュエータの制御や、ポートの入出力制御を行うことができる。また、obniz 系マイコンの開発言語は、Web アプリで広く利用される JavaScript であり、次節で紹介する動画と同期して実行されるスクリプト言語「srt.js[7]」との相性が良い点も配慮した。

学習者は対応する obniz 系マイコンさえ持っていれば、複雑な準備を必要とせずに、本システムを利用できる。

4.2 コンテンツと同期するスクリプト

ここでは、動画版／スライド版システムと同期して動作するスクリプトについて説明する。どちらのスクリプトも、基本的には標準的な JavaScript と obniz ライブライを用いるが、動画版では「動画上の経過時間」、スライド版では「スライド上のページ遷移」に対する記述方式を備える点が特徴である。

具体的には、動画版では「srt.js」という JavaScript のフレームワークを使用する。これは、動画の字幕情報を記述するフォーマットである srt 形式のファイルを拡張したものであり、動画上の経過時間に対応して特定のコードを実行できる。例えば、図 5 上部のように動画上の経過時間（例：10 秒→15 秒）を記述して、次にその時間に実行するコードを記述する。ここで、経過時間の直前の通し番号から空行までが一ブロックとして認識される。

スライド版では、図 5 下部のように、関数の配列を用意し、スライドのページ番号をインデックスとして指定して、関数内にコードを記述する。関数は、各ページに遷移した（画面に表示された）タイミングで実行される。

これらのスクリプトは、コンテンツと紐付けて Firebase 上のデータベースに保存する。

4.3 動画版システム

動画版システムは、大きく Web アプリ、動画共有サービス（例：YouTube）、IoT マイコン（obniz）から構成される。Web アプリとしては、閲覧アプリと作成支援アプリの二つを開発した。以下、詳細について述べる。

⁴ <https://obniz.com/ja/products/obnizos>

```

0
00:00:00,000 -> 00:00:01,000
const obniz = new Obniz("input id here");
displayCaption("Hello World");

1
00:00:10,000 -> 00:00:15,000
obniz.display.clear();
obniz.diplay.print("Hello World");

const obniz = new Obniz("input id here");

slideFunction[2] = () => [
  obniz.display.clear();
  obniz.diplay.print("Hello World");
]

```

図 5. スクリプトの記述例（上：動画版、下：スライド版）



図 6. 閲覧アプリのスクリーンショット

4.3.1 閲覧アプリ

閲覧アプリのスクリーンショットを図 6 に示す。大きく、(1)YouTube から埋め込まれた動画表示部分、(2)ソースコード表示部分、(3)字幕表示部分、(4)出力コンソールの 4 つの要素から構成される。

ソースコード表示部分は、関数／変数等の色付けに対応したエディタライブラリである、CodeMirror を用いて実装した。コード表示部分は、「リアルタイム」「全体」という二つのタブを備えている。「リアルタイム」では、動画の時間軸（現在のデバイスの挙動）に対応したコードのみを表示する。「全体」では、一連の処理を行うコードをまとめて表示する。

字幕表示部分は、デバイスの動作の補足説明やユーザに注目してほしい箇所等を表示する。作成者は、srt.js の記法に沿って、コードと交えて字幕を記述することができる。出力コンソールは、手元のデバイスの動作（例：ボタンや加速度センサ）に対応したセンサデータ等を、リアルタイムに出力する。

また、これら 4 つの要素（動画／ソースコード／字幕／コンソール）について、注目させたい箇所をハイライト表示させることができる。ハイライト表示の例と、対応するコードを図 7 に示す。

4.3.2 作成支援アプリ

動画版システムでは、スクリプトに動画の経過時間を入力する必要がある。これを手作業でやる場合、動画とエディタを並べて開いておいて、動画上でデバイスの動作場面を探し、経過時間をエディタに記

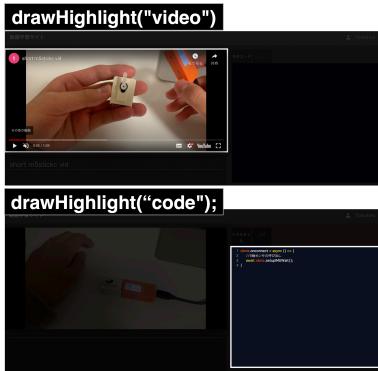


図 7. ハイライト表示の例（上：動画のみ，下：ソースコードのみ）

述する必要があり、手間が掛かる。

そのため、動画版システムのコンテンツ作成を支援するシステムを開発した。作成支援アプリのスクリーンショットを図 8 に示す。大きく、(1) 動画表示部分、(2) タイムライン、(3) コードエディタ（オリジナル）、(4) コードエディタ（タイムライン）の 4 つの要素から構成される。

タイムラインは動画の時間軸と対応しており、タイムライン上の任意の箇所をダブルクリックすることで、動画の経過時間⁵を自動的に挿入したコードブロックを作成できる。さらに、コードブロックをダブルクリックすることで、右上のコードディタ（タイムライン）で開いて編集する。コードエディタ（オリジナル）には事前に作成したコードを読み込み、コピー＆ペースト等を用いることで、srt.js 形式のコードを効率的に編集できるように配慮している。

最後に、「アップロード」ボタンを押すことで、Firebase 上のデータベースに srt.js 形式のファイルがアップロードされる。アップロード後には閲覧アプリの URL が提示され、即座に動作を確認できる。

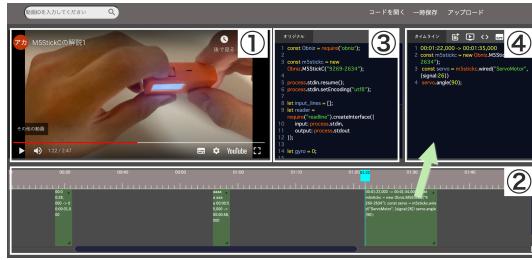


図 8. 作成支援アプリのスクリーンショット

4.4 スライド版システム

スライド版システムの作成にあたり、ページ遷移を API 経由で検出できるサービスが見つからなかっ

⁵ 開始時間を動画の現在再生時刻、終了時間を現在再生時刻から 3 秒後としている

たため、reveal.js⁶というライブラリを利用して、独自のスライドショーシステムを構築することにした。

図 9 にシステムのスクリーンショットと利用の流れを示す。まず、作成者（主に講師）は初期画面で、スライド（PDF）、スクリプト（JavaScript），講師／学習者のデバイス ID の一覧（CSV）を Web アプリに読み込む。前述したように、スライド版システムでは講師の PC と全ての学習者の obniz 系マイコンを接続するため、CSV ファイルでまとめて読み込めるように配慮した。

3 つのファイルを読み込んだら、プレビュー画面に遷移する。プレビュー画面では、各スライドへの遷移時に実行されるコードを確認することができる。問題がある場合、「戻る」ボタンを押すことで、特定のファイルのみを差し替えることができる。

プレビュー画面から、「プレゼンテーション開始」ボタンを押すとスライドが全画面表示され、デバイスとの連動が開始される。ここで、特定のスライドが表示される（ページ遷移する）と、ページ番号に対応した関数が実行され、事前に読み込んだデバイス ID に対応する全ての obniz 系マイコンが連動して動作する。講師の負担を軽減するために、コード上では複数のマイコンに対する繰り返し処理は隠蔽され、自動的に補完されるように配慮した。

なお、アップロードされた PDF / JavaScript / CSV は Firebase 上のデータベースで管理され、PDF ファイル名とタイムスタンプをキーとして後から呼び出すことができる。

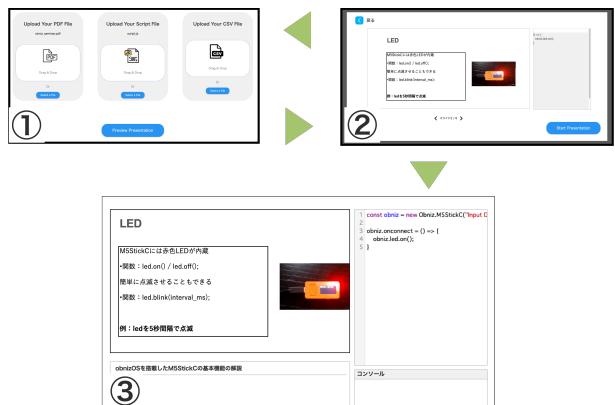


図 9. スライド版システム（1. 初期画面、2. プレビュー画面、3. プrezentation画面）

5 議論

5.1 コンテンツ事例

ここでは、主に先行して実装した動画版システムを中心としたコンテンツ事例について述べる。まず、

⁶ <https://revealjs.com/>

コンテンツを作成するにあたり、基本的な撮影方法を検討した。通常／システム利用中の様子の2種類の撮影方法でコンテンツを作成した(図10)。通常の撮影方法については、動画とソースコードをわかりやすく見ることができる反面、動画と手元のデバイスがいつ連動するのかや、ユーザの入力が必要な場面がわかりにくいといった問題があった。システム利用の様子を撮影した場合については、手元のデバイスがいつ動くのかはわかりやすくなつたが、肝心の動画が見づらいという問題点があった。また、動画を二重に撮影する必要があるため、作成者の手間も増えてしまう。このように、それぞれの撮影方法は一長一短であるが、手間の少なさと動画の見やすさを考慮して、1点目の撮影方法を基本的に利用し、手元のデバイスに注目させる方法を整理／実装していく方針とした。

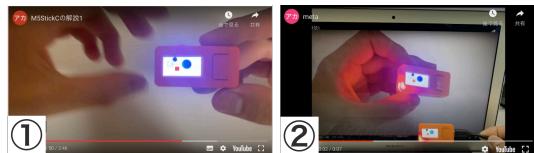


図 10. 2 つの撮影方法 (1. 通常, 2. システム利用中)

また、m5stickC のような加速度センサ／ジャイロセンサが搭載されているマイコンであれば、センサでデバイスの状態を検出して、システムの挙動に反映できる。一例として、ユーザがデバイスを手に取っている時に動画を停止し、手から離すと動画を再生するような機能を実装した。一方、この機能はセンサ搭載デバイスでしか利用できず、意図せぬ動作を引き起こす可能性もあるため、汎用的な手法も併せて検討する必要がある。

さらに、「LED を点滅させる」、「LCD で図形を描画する」、「加速度センサの値をコンソールに出力する」等といったマイコンの基本的な機能を紹介するコンテンツや、「ジャイロセンサの値に応じてサーボモーターを制御する」、「加速度センサやマイコンのボタンを用いてリモコンのように YouTube 動画を操作する」といった簡単なアプリケーションを提案システムを用いて作成した。内容の一部については、デモビデオでも紹介する。

5.2 システムの使用感と課題

本格的な評価実験はまだ行っていないが、これまでに著者らにより複数のコンテンツを作成し、所属研究室内で、動画版システムの予備実験とスライド版システムを用いた簡易ワークショップを実施した。詳細は紙面の都合で割愛するが、これらの経験に基づくシステムの印象や課題について議論する。

まず、コンテンツ閲覧時の使用感として、動画版システムでは、予備実験のアンケートからコンテン

ツに合わせてデバイスが連動する点や、動作に関連するソースコードを提示する点については好印象を得られた。スライド版システムを用いてライブ型の体験をしてもらった際には、スライドの切替に応じて各ユーザの手元のデバイスが連動する際に、「おおっ」という感嘆の声が複数上がっていた。

課題としては、動画版システムでは、現状では時系列通りに動画を閲覧しないと、変数の初期化等が正しく実施されず、予期せぬ動作を起こす可能性がある。よって、体験者にはスキップ等の動画の操作をなるべく利用しないように依頼したが、ストレスであるという声も上がった。対策として、全てのスクリプトの実行前に自動的に実行される初期化用コードを用意したり、スキップした間のコードを自動的に実行するような仕組みを検討している。また、スクリプトの埋め込んである時刻をタイムライン上で可視化するような機能の実装も予定している。

次に、著者らがコンテンツ作成を通じて感じた印象や課題について述べる。動画版では、当初手動で通し番号や動画の経過時間を記述していたため、作成支援システムを開発したことで、こうした労力は大分減ったと感じた。一方、ユーザにコードを提示する場合は実行用／表示用で2度同じコードを書く必要があるため冗長であると感じた。また、動作確認の際に閲覧アプリと作成支援アプリを何度も切り替える必要があり、少々面倒であった。今後は、コードを手軽に表示できる記法や、作成支援システム上の文法チェックや簡易的なデバッグを行えるように改良していきたい。

スライド版では、一人の講師が複数の学習者に対して利用するが、学習者側のデバイスの動作状況が分かりにくく問題があった。今回は、音声やチャット等でコミュニケーションをとったが、今後は学習者側のデバイスの動作状況(例: 入出力ポートの状態)等を手軽に把握できるような視覚化機能の実装を検討していく。また、現時点では講師のスライドと全ての学習者のマイコンが連動するため、個別のトラブルシューティング等を行う際には不便である。今後は、スライドとマイコンの接続関係を手軽に変更できるような機能と UI の実装を進めていきたい。

6 まとめ

本研究は、動画やスライド等の順序型コンテンツにスクリプト言語を同期させ、実世界の IoT デバイスや編集可能なソースコードと連携することで、IoT プログラミング学習を支援するシステム「IoTeach」を提案・試作した。さらに、動画版を中心としたコンテンツ事例や、動画版／スライド版システムの使用感と課題を議論した。今後は議論で述べたシステムの改善を行った上で、システムの有無での学習効果の変化やモチベーション向上について検証を進めていく。

参考文献

- [1] Progate. <https://prog-8.com>.
- [2] ドットインストール-3分動画でマスターできるプログラミング学習サービス. <https://dotinstall.com/>.
- [3] K. Jun, O. Masa, I. Takahiro, and G. Masataka. Songle Sync: A Large-Scale Web-based Platform for Controlling Various Devices in Synchronization with Music. In *Proceedings of the 26th ACM International Conference on Multimedia*, pp. 1697–1705, 2018.
- [4] K. Jun, N. Tomoyasu, and G. Masataka. TextAlive: Integrated Design Environment for Kinetic Typography. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 3403–3412, 2015.
- [5] Y. Tom, C. T. Hsiang, and M. R. C. Sikuli: Using GUI screenshots for search and automation. In *UIST 2009 - Proceedings of the 22nd Annual ACM Symposium on User Interface Software and Technology*, pp. 183–192, 2009.
- [6] 栗原一貴. 学会イベント支援 2) 放送化の時代のプレゼンテーション支援システム. 56(5):465–471, 2015.
- [7] 栗原一貴, 橋本美香. srt.js:映像コンテンツへの IoT 指向拡張プログラム埋め込みフレームワーク. 日本ソフトウェア科学会 WISS2016 論文集, 2016.
- [8] 森田浩平, 松本慎平. 大学でのプログラミング基礎講義における初学者を支援するためのオンラインコンペイラの開発. 人工知能学会 第 82 回 先進的学習科学と工学研究会 (SIG-ALST), Vol. B5, No. 03, pp. 63–66, 2017.
- [9] 又吉康綱, 中村聰史. typing.run: 初学者のプログラミング学習を支援するプログラムタイピングシステムの提案と実践. 研究報告ヒューマンコンピュータインターフェース (HCI) , Vol. 2020-HCI-189, pp. 1–8, 2020.

部屋の3次元縮小モデルを用いたMR空間内における仮想オブジェクト配置

井原 圭一* 川口 一画†

概要. 本研究では、深度カメラを用いて部屋の3次元縮小モデルを構築し、MRにおいてWiM手法を利用可能にする。これにより、オブジェクトを3次元空間に自由に配置でき、配置を変更したオブジェクトの位置を即座に確認できるシステムを提案する。我々はまず、システムの基本設計を示すためにプロトタイプを作成した。その後、今後のシステム設計の指針を得るために、実尺オブジェクトのレイキャスト操作と縮小オブジェクトの直接操作を比較する実験を実施し、それらの操作手法の特性とその特性に影響を与えるMRデバイスの特性を調査した。その結果、縮小オブジェクトの直接操作は、オブジェクトの探索、移動、および身体的負担において有効であるが、スケーリング操作が困難であることが示された。また、これらはMRデバイスの視野角およびハンドトラッキングの精度による影響が示唆され、今後のシステム設計における指針が得られた。

1 はじめに

Microsoft社のHoloLens[9]に代表されるMixed Reality(MR)デバイスは、実空間上に仮想オブジェクトを重畳表示することが可能である。さらに、ユーザはコントローラやハンドトラッキングを用いた操作によって表示された仮想オブジェクトに対してインタラクションを行うことも可能である。近年では、NrealLight[6]のような比較的安価な製品も販売されており、今後MRデバイスが広く普及し日常的に用いられるようになると考えられる。MRデバイスを日常的に用いる場合、自身の周辺環境に仮想的な情報を重畳し拡張していく用途が考えられる。例えば、ニュースや新着メッセージ、自身の予定等を提示する情報ウィンドウを部屋のいたるところに分散配置することが考えられる。このような用途を想定する場合、周辺環境の任意の位置に、任意の大きさで仮想オブジェクトを配置できる必要がある。この場合、既存のPCのような2次元のインターフェース内でオブジェクトを移動させるのとは異なり、奥行き方向の操作が必要となる。しかし、仮想オブジェクトの正確な位置の認識は困難であることが分かっている[4]ため、操作に時間がかかり、身体的負担が大きいと考えられる。

これまでの研究では、操作時間の短縮および身体的負担の軽減のため、実空間の特徴(壁やエッジ)を利用した仮想オブジェクトの配置手法[5][13]や、部屋の奥行き方向のスケールを縮小・復元することによる仮想オブジェクト配置手法が提案されてきた[2]。しかし、前者では3次元空間内の特徴点が無い場所には仮想オブジェクトを配置できない。また後者では、部屋の奥行き方向のスケールを縮小した状

態で仮想オブジェクトを操作するため、実空間中の仮想オブジェクトの位置を即座に確認することができない。本研究では、3次元空間の任意の位置に自由に仮想オブジェクトを配置でき、さらに配置位置を即座に確認できるWorlds in Miniature(WiM)手法[16]を利用する。

WiM手法とは、Virtual Reality(VR)における仮想オブジェクト操作手法の1つである。WiM手法では、仮想環境全体を縮小コピーしたモデル(ミニチュア)を構築し、ユーザの近くに配置する。そしてユーザがミニチュア内で仮想オブジェクトを操作すると、それと連動して縮小前の実尺の仮想環境内でも対応するオブジェクトが操作される。MRを用いて実空間に情報を重畳提示する際、その対象となる空間はユーザの目の届く範囲、すなわち1つの部屋の中で完結すると考えられる。そのため、部屋のミニチュアを用いるWiM手法が有効であると考えた。MRにおいて、WiM手法を実現するために、MRデバイスに搭載されている深度カメラから得られる深度データを利用する。得られた深度データより、部屋の3次元縮小モデルを生成し、ミニチュアとして利用する。ミニチュアは、ユーザの手元に配置し、その中の仮想オブジェクトを実空間の仮想オブジェクトと連動させる。これにより、ユーザはミニチュア内の仮想オブジェクトを操作することで、実空間の仮想オブジェクトを移動させることができる。実装したプロトタイプの概要を図1に示す。

本論文では、まず関連研究について説明し、そこで得られた知見に基づき提案システムの設計指針を示す。次に提示した設計指針に基づき実装したプロトタイプについて述べる。その後、MRにおけるミニチュアを用いたオブジェクト操作の特性を調査するために実施した評価実験について述べる。最後に、実験結果に基づき、今後のシステム設計の方針

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 システム情報系

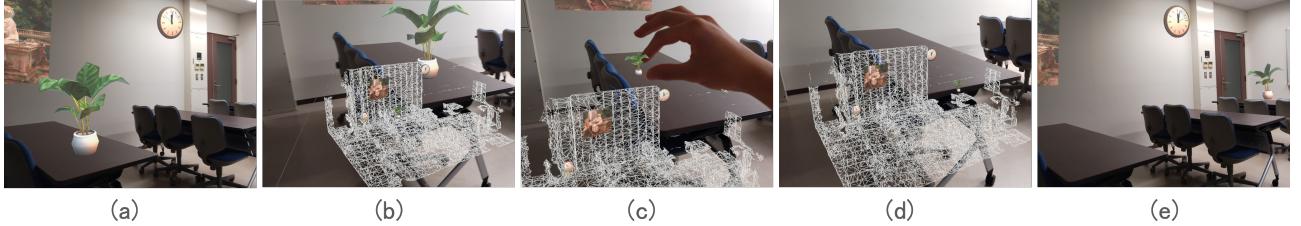


図 1. プロトタイプの概要.(a) 部屋に仮想オブジェクト（植木鉢、時計、および絵画）が配置されている。(b) ミニチュアを表示し、仮想オブジェクトの位置を確認する。(c, d) 直接操作によってミニチュア内にある縮小されたオブジェクトの配置を変更する。(e) ミニチュア内のオブジェクトの位置はリアルタイムに実尺のオブジェクトに反映される。

について述べる。

2 関連研究

本章では、まずMRにおけるオブジェクト操作の既存手法および課題について述べる。次にVRにおけるオブジェクト操作手法について述べる。最後に、関連研究における知見に基づき、本研究におけるシステムの設計指針について述べる。

2.1 MRにおけるオブジェクトの操作手法

本研究で着目するMR空間内でのオブジェクト操作に対しては、レイキャストを用いる操作が最も一般的である。しかし、近年、操作時間の短縮および身体的負担の軽減のため、深度情報を利用し、オブジェクト操作を行う研究が行われている。Nuernbergerらは、実物体のエッジや平面を利用し、オブジェクトの配置を行った[13]。Chaeらは、壁に重畠表示した仮想の壁を前後に移動させ、仮想的に部屋のスケールを縮小・復元することにより、オブジェクトの配置を補助した[2]。しかし、実物体のエッジや平面を用いた配置では3次元空間にオブジェクトを自由に配置できず、壁の移動を行う場合は配置を即座に確認することができないという課題がある。

2.2 VRにおけるオブジェクトの操作手法

MRではなく、VR空間内でのオブジェクト操作に関する研究が多く実施されている。VRにおけるオブジェクト操作は、一人称視点による操作と、三人称視点を用いた操作に分けられ、さらに、一人称視点による操作は手のメタファを用いた操作と、ポインタのメタファを用いた操作に分けられる[1]。

手のメタファを用いた操作であるGo-Go Interaction[14]は、ユーザーの実際の腕と仮想の腕を非線形にマッピングすることで腕を伸ばし、離れた仮想オブジェクトを操作することを可能にしている。ポインタのメタファを用いた操作であるレイキャスト操作は、ユーザーの手から放射されるレイを用いて、離れた仮想オブジェクトの操作を行う。手の伸縮やレイキャストを用いる手法は、手をオブジェクトに向けることが必要となるため、ユーザーの身体的負担が

大きいことが考えられる。また、MRにおける利用において、一人称視点のみを用いる場合、オブジェクトの位置の正確な認識が困難である[4]。

三人称視点を用いた操作であるWiM手法[16]は、仮想環境のコピーを縮小させたモデルを生成し、そのモデル内のオブジェクトを操作することにより、離れた仮想オブジェクトの操作を行う。WiM手法は、操作における手の移動が小さいことから、身体的負担の軽減が可能であると考えられる。また、新たな視点を加えることにより、オブジェクトの位置の認識の補助が可能であると考えられる。

WiM手法は、VRにおけるオブジェクト操作だけでなく、ナビゲーション[12]やIoT操作[15]、空間デザイン[17]等においても用いられている。WiM手法は、MRにおけるオブジェクト操作においても、有効であると考えられる。しかし、MRにおけるWiM手法の効果は明らかにされていない。

2.3 本研究における設計指針

MRにおけるオブジェクト操作のための既存手法[13][2]は、操作時間の短縮および身体的負担の軽減に有効である一方で、3次元空間にオブジェクトを自由に配置できない、またはオブジェクトの配置を即座に確認することができないという課題があった。これらの課題の解決のために、本研究ではVRにおけるオブジェクト操作のための手法であるWiM手法をMRで実現するシステムの実装を行う。そのための設計指針は以下の通りである。

- ユーザ周囲の環境を3次元モデルとして取得する
- 取得した3次元モデルを縮小したミニチュアを生成し、ユーザーの手元に配置する
- ミニチュア内に仮想オブジェクトを配置する機能を実装し、さらにミニチュア内での配置に対応して実空間内でも仮想オブジェクトが配置されるようにする

これらの設計指針によって、WiM手法をMRで実現するシステムを実現することにより、仮想オブジェクトを3次元空間に自由に配置すること、およ

部屋の3次元縮小モデルを用いたMR空間内における仮想オブジェクト配置

び配置を変更したオブジェクトの位置を即座に確認することを可能にする。また、このシステムを用いることにより、操作時間および手の移動距離を減らすことが可能であると考えられる。

3 プロトタイプ

本章では、2.3節で示した設計指針に基づき実装したプロトタイプについての説明を行う。

3.1 システム構成

ハードウェアはHololens2[7]を、ソフトウェアはMixedRealityToolkit (MRTK, バージョン2.6.2)とUnity (バージョン2019.4.22.f1) [10]を利用した。

3.2 ミニチュアの生成

Hololens2では、内蔵された深度カメラを用いた実空間の検知機能である空間マッピング[11]を利用できる。本実装では、空間マッピングにおいて生成される部屋のメッシュを取り得し(図2)，大きさを変更した部屋の縮小モデルをミニチュアとして利用した(図1(b))。そのため、本実装では追加の装置を必要とせず、またリアルタイムにミニチュアを生成できる。



図2. Hololens2により取得した部屋のメッシュ。

3.3 仮想オブジェクトのリンク

ミニチュア内において仮想オブジェクトを操作すると、実空間内の対応する仮想オブジェクトも運動して操作が反映される機能を実装した。部屋の仮想オブジェクト(実尺オブジェクト)Aを縮小させた仮想オブジェクト(縮小オブジェクト)Amの位置および大きさは以下のように計算される。実空間上に設定された原点の位置を X_r 、実空間上に配置された実尺オブジェクトAの位置を Pr 、大きさを Sr 、ミニチュアの配置位置を Xm 、ミニチュア内に配置された縮小オブジェクトAmの位置を Pm 、大きさを Sm とおく。ミニチュアの縮小率は t とおく。

$$XrPm = XrXm + t \times XrPr \quad (1)$$

$$Sm = t \times Sr \quad (2)$$

縮小オブジェクトへの操作が検知される度に、実尺オブジェクトの位置 Pr および大きさ Sr が、縮小オブジェクトの位置 Pm および大きさ Sm に基づいて更新される。

3.4 仮想オブジェクトの操作方法

本実装では、縮小オブジェクトの移動操作およびスケーリング操作に、直接操作を用いる。直接操作とは、Hololens2のデフォルトで利用できる操作であり、人差し指と親指を用いて直接オブジェクトをつまむ動作を用いた操作である。具体的には、移動操作は片手で縮小オブジェクトをつまみ、ミニチュア内の任意の位置に移動させる。スケーリング操作は両手で縮小オブジェクトをつまみ、両手間の距離を広げることにより、縮小オブジェクトを拡大させ、狭めることにより、縮小させる。

4 評価実験

MRにおけるミニチュアを用いたオブジェクト操作手法の効果を検証するため、評価実験を行う。ここでは、オブジェクト操作の基礎的特性(操作時間、身体的負荷、ユーザビリティ)を確認するため、空間内に提示された操作対象オブジェクトをターゲット位置に移動するという単純な操作を実験タスクとする。

4.1 実験条件

実験条件は以下の2つの条件とした。

C1. 実尺オブジェクトのレイキャスト操作 部屋に配置された実尺オブジェクトをレイキャストを用いて操作する。

C2. 縮小オブジェクトの直接操作 手元に配置された縮小オブジェクトを直接つまむことで操作する。

参加者は、大学生および大学院生12名(男性10名、女性2名)であり、右利き11名、左利き1名であった。参加者内配置の実験計画とし、参加者はそれぞれ2つの条件でタスクを実施した。条件の順序はカウンターバランスをとり、6名はC1・C2の順序で、残りの6名はC2・C1の順序でタスクを実施した。

4.2 実験タスク

部屋の中に操作対象となる白い球体(以後、白球)と、ターゲットとなる赤い球体(以後、赤球)をそれぞれ1つずつ表示し、白球を移動およびスケーリングさせ、ターゲットの赤球に一致させるタスクを実施した。実験環境は部屋の一画に設けられた5m×5m×2.5mの空間とした。なお本実験は、オブジェ

クト操作の基礎的特性を評価することを目的としたことから、実験環境内には操作対象物とターゲットのみを提示することとし、机などの他のオブジェクトは配置しなかった。そのためプロトタイプで実装した空間マッピングを用いたメッシュ取得とミニチュアの作成は行わず、実験環境の範囲を示す床のみを提示した。白球と赤球は実験環境全体をグリッド上に分割した48個の候補点（図3）のうち、タスク毎に無作為に配置された。白球と赤球はどちらも48個の候補点に重なることなく一回ずつ出現させ、計48回の試行を行った。白球、赤球の大きさは（0.5m, 0.4m, 0.3m）の3通りであり、それぞれを16回ずつ出現させた。白球と赤球の中心距離が赤球の半径以下、かつ白球の大きさが赤球の大きさの±10%となれば、タスクを完了したとみなし、次のタスクに移行する。縮小モデルは部屋の1/10スケールであり、部屋の中心（図3（右）2つの青点の中点）に固定する。C1およびC2どちらの条件も、移動操作は片手で行い、スケーリング操作は両手で行う。各条件の操作の様子を図4に示す。タスク終了後、参加者に対し、ユーザビリティ調査のためのSystem Usability Scale (SUS) [1]、作業負荷調査のためのNASA-TLX[3]、および自由記述アンケートの3つを実施した。

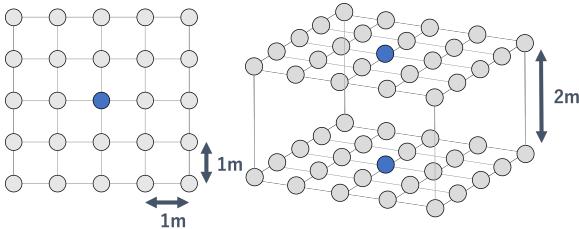


図3. 球の配置候補点（青点はユーザの初期位置のため表示しない）。(右) 下段は床から0.25mの高さに位置する。

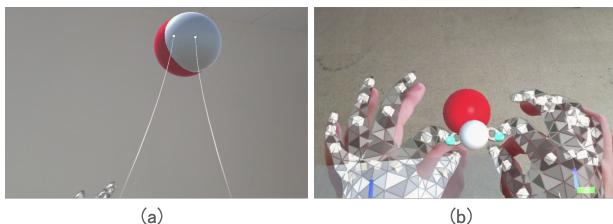


図4. (a) C1条件の操作の様子。(b) C2条件の操作の様子。

4.3 実験手順

実験では、最初に書面により研究の目的・内容等について説明した後、参加者の同意を取得した。

次に、本実験のタスク内容について説明を行った。続いて、参加者にピンチによるオブジェクト操作に慣れてもらうため、Hololensヒント[8]およびオブジェクトの基本操作の練習を満足できるまで行ってもらった。その後、各条件でタスクを実施した。タスクにあたっては、より速くより正確な操作を行うように実験参加者に指示した。各条件のタスク終了後、SUS、NASA-TLX、および自由記述アンケートを実施した。

4.4 取得するデータ

我々は、オブジェクト操作が、オブジェクトを見つけるフェーズ、オブジェクトを移動させるフェーズ、オブジェクトをスケーリングさせるフェーズに分かれると考えた。そのため、我々は、1タスクの開始から終了までの時間（全操作時間）だけでなく、探索時間、移動時間、スケール時間の取得も行った。ここで、探索時間とはタスクが開始されてから、球の位置を把握し球に触るまでの時間として計算した。また、手の移動距離を取得するため、タスク中の手の位置のログデータを取得した。手の位置はHololens2のハンドトラッキング機能によって取得した。

4.5 結果

まず、タスクの全操作時間、探索時間、移動時間およびスケール時間に対して、第三四分位数に四分位範囲の1.5倍を足した値より大きい値もしくは、第一四分位数から四分位範囲の1.5倍を引いた値より小さい値を外れ値と見なし、破棄した。

4.5.1 操作時間

各条件の平均操作時間を図5に示す。各条件の全操作時間の平均は、C1条件20.34分（SD=3.42）、C2条件8.57分（SD=1.58）であった。ウィルコクソンの符号順位検定を実施した結果、有意差が見られた（ $V=78$, $p<0.01$ ）。各条件の探索時間の平均は、C1条件11.21分（SD=2.23）、C2条件2.93分（SD=0.73）であった。ウィルコクソンの符号順位検定を実施した結果、有意差が見られた（ $V=78$, $p<0.01$ ）。各条件の移動時間の平均は、C1条件7.20分（SD=1.92）、C2条件3.82分（SD=1.35）であった。ウィルコクソンの符号順位検定を実施した結果、有意差が見られた（ $V=78$, $p<0.01$ ）。各条件のスケール時間の平均は、C1条件1.92分（SD=0.63）、C2条件1.82分（SD=0.56）であった。ウィルコクソンの符号順位検定を実施した結果、有意差は見られなかった（ $V=46$, $p=0.622$ ）。

部屋の3次元縮小モデルを用いたMR空間内における仮想オブジェクト配置

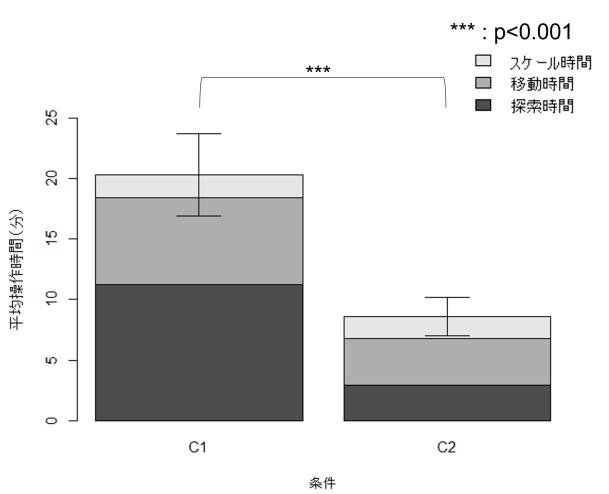


図 5. 各条件の平均操作時間。エラーバーは標準偏差を示す。なお、エラーバーおよび記号 (*) は全操作時間の平均値についての値を示す。

4.5.2 手の移動距離

各条件の手の平均移動距離は、C1 条件 223.7m ($SD=49.4$) , C2 条件 86.1m ($SD=13.3$) であった(図 6)。対応のある t 検定を実施した結果、有意差が見られた ($t (12) = 11.34, p < 0.05$)。

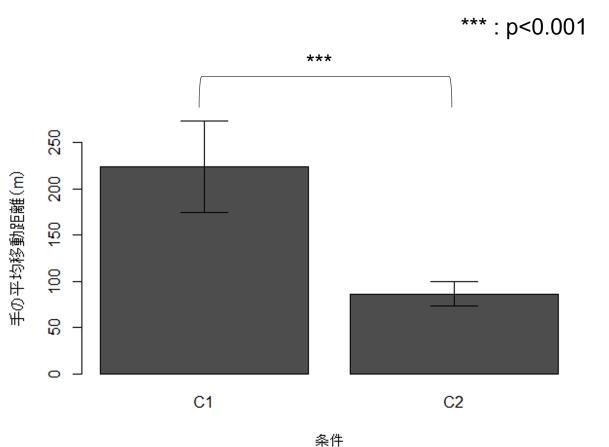


図 6. 各条件の手の平均移動距離。エラーバーは標準偏差を示す。

4.5.3 SUS

各条件のSUSの平均得点は、C1 条件 67.7 ($SD=15.94$) , C2 条件 77.7 ($SD=9.85$) であった。対応のある t 検定を実施した結果、有意差は見られなかった ($t (12) = -1.97, p=0.074$)。

4.5.4 NASA-TLX

各条件の総合的な作業負荷の平均は、C1 条件 72.5 ($SD=17.7$) , C2 条件 39.9 ($SD=15.3$) であった。ウィルコクソンの符号順位検定を実施した結果、有意差が見られた ($V=78, p<0.05$)。各条件の身体的要因に関する作業負荷の平均は、C1 条件 84.6 ($SD=17.2$) , C2 条件 36.3 ($SD=25.1$) であった。ウィルコクソンの符号順位検定を実施した結果、有意差が見られた ($V=78, p<0.05$)。

4.5.5 自由記述アンケート

C1 条件の探索に関して、「レイキャストは、視野に対して物体の置かれる範囲が広くて、探すのが一番大変だった」、「画角が狭いために、歩き回ってオブジェクトの位置を把握する必要があった」という回答が得られた。それに対し、C2 条件の探索に関しては、「ミニチュアは部屋全体を俯瞰的に観察できるため、操作に集中しやすく、とてもつかいややすかった」という回答が得られた。また、C2 条件のスケーリング操作に関して、「すでに小さいものを両手でつかみ、縮小させるという操作が困難だった。枠などのハンドルがあった方が操作しやすかったと思う。」「スケーリング（縮小）の時に両手の指がぶつかり、少々手間取った。」「ハンドトラッキングの精度とミニチュアのサイズのトレードオフが存在すると感じた」という回答が得られた。

5 考察

5.1 オブジェクトの探索

まずオブジェクトの探索に関して、C1 条件より C2 条件の方が探索時間は短かった。また、自由記述アンケートにおいて、C1 条件では MR デバイスの視野角の狭さから探索が困難であり、C2 条件では部屋を俯瞰できるため、操作が簡単であったという回答が得られた。そのため、MR デバイスの視野角の狭さが、探索に影響を与えた可能性が考えられる。関連して、VR 用デバイスでは視野角 120 度程度の製品が多く存在するのに対して、現在市販されている MR デバイスは視野角が 60 度未満である (Hololens2 : 54 度)。そのため、MR では広い空間内のオブジェクト探索が VR よりも困難となることから、ミニチュアを用いる本研究のアプローチは妥当性を有すると考えられる。一方で、このような MR デバイス特有の視野角の狭さを考慮すると、ミニチュアを用いる場合にも、そのサイズや位置について十分な検討が必要になると考えられる。

5.2 オブジェクトの移動

次にオブジェクトの移動に関して、C1 条件より C2 条件の方が移動時間は短かった。また、手の移動距離において C1 条件より C2 条件の方が移動距離が

短いという結果が得られた。そのため、手の移動距離の差によって、移動時間に差が生じたことが考えられる。関連して、身体的負担については、NASA-TLX の結果から、C1 条件より C2 条件の方が作業負荷が小さいことが示された。これに関しても、手の移動距離の差によって、身体的負担に差が生じたことが考えられる。

5.3 オブジェクトのスケーリング

最後に、オブジェクトのスケーリングに関して、C1 条件と C2 条件のスケーリング時間に差は見られなかった。また、自由記述アンケートにおいて、C2 条件において両手の指の衝突が生じたという回答や、ハンドトラッキングの精度が縮小オブジェクトの操作に影響を与えていた可能性があるという回答が得られた。本実装では、ミニチュア内のオブジェクトを操作しようとする場合、操作の有効領域（タッチイベントを受け付ける範囲）もミニチュアと同等の倍率で縮小されていた。そのため両手を同時に用いた複雑な操作を要するスケーリングを行う際に、ハンドトラッキングの精度の影響が生じていたと考えられる。今後は、ミニチュア内のオブジェクトの操作を行う際の有効領域を広げる、もしくは両手を用いない別の操作方法を採用する等の検討を行う必要があると考えられる。

5.4 今後のシステムの設計指針

本実験の結果をまとめると、縮小オブジェクトの直接操作は、実尺オブジェクトのレイキャスト操作と比較し、オブジェクトの探索時間が短くなることが示された。また操作のための手の移動距離および移動時間が短くなり、身体的負担も軽減されることが示された。そのため、MR でのオブジェクト操作にミニチュアを用いる本研究のアプローチは妥当性を有すると考えられる。なお、オブジェクトの探索にあたっては MR デバイス特有の視野角の狭さが影響を与えていた可能性が示された。そのため今後のシステム設計においては、ミニチュアが可能な限り視野角内に収まるようにするため、サイズや位置について十分な検討を行っていく。

次にスケーリング操作に関しては、スケーリング時間に差がなく、また縮小したオブジェクトを直接操作する際にハンドトラッキングの精度が課題となる可能性が示唆された。これはミニチュアを作成する際に、操作の有効領域もミニチュアと同等の倍率で縮小されていたことによるものと考えられる。そのため、今後のシステム設計において、スケーリングを含む両手を用いた複雑な操作に対応するため、ハンドトラッキングの精度に応じて、縮小率の下限を決定する必要がある。もしくは、縮小オブジェクトに UI を追加する、もしくは縮小オブジェクトより大きなハンドルを設ける等の工夫を行う必要がある。

5.5 今後の課題

本論文で実施した実験では、比較を「実尺オブジェクトのレイキャスト操作」と「縮小オブジェクトの直接操作」の 2 条件でしか行っていない。本手法の優位性をより正確に示すためには、既存の類似手法との比較を行う必要がある。また、本実験ではオブジェクト操作の基礎的特性（操作時間、身体的負荷、ユーザビリティ）を確認するため、空間内に提示された操作対象オブジェクトをターゲット位置に移動するという単純な操作を実験タスクとした。そのため既存手法における課題、すなわち 3 次元空間にオブジェクトを自由に配置できること、およびオブジェクトの配置を即座に確認できることに対する提案手法の効果を検証できていない。そのため今後はこれらの課題に対する提案手法の効果を評価するための実験を行っていく必要がある。具体的には、オフィスやリビング等でシステムを使用する場面を想定し、ユーザが自分の自由意思で周囲の環境に仮想オブジェクトを配置するタスクを設定して評価を行う。

上記の実験を実施した後、その結果に基づいて特定の用途に特化したアプリケーションの実装を行い、その効果の検証を行っていく。例えば、オフィス環境の拡張に特化したアプリケーションや、リビング等のインテリア検討に特化したアプリケーションを実装し、それらを用いた場合の効果を検証していく。

6 まとめ

我々は、深度カメラを用いた部屋の 3 次元縮小モデルを用い、MR において WiM 手法を利用可能にすることにより、オブジェクトを 3 次元空間に自由に配置でき、配置を変更したオブジェクトの位置を即座に確認できるシステムを提案した。また、MR において WiM 手法を利用することにより、ユーザの操作時間の減少および身体的負担の軽減が可能であると考えた。我々はまず、システムの基本設計を示すためにプロトタイプを作成した。その後、今後のシステム設計の指針を得るために、実尺オブジェクトのレイキャスト操作と縮小オブジェクトの直接操作を比較する実験を実施し、それらの操作手法の特性とその特性に影響を与える MR デバイスの特性を調査した。その結果、縮小オブジェクトの直接操作は、オブジェクトの探索、移動、および身体的負担において有効であるが、スケーリング操作が困難であることが示された。今後は実験によって得られた指針に基づき、示された欠点を補完するための機能を追加し、システム設計を完了させる。また、そのシステムを用いたアプリケーションの開発および評価実験を実施する。

参考文献

- [1] J. Brooke. SUS: A quick and dirty usability scale. *Usability Eval. Ind.*, 189, Nov. 1995.
- [2] H. J. Chae, J.-i. Hwang, and J. Seo. Wall-Based Space Manipulation Technique for Efficient Placement of Distant Objects in Augmented Reality. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*, UIST '18, pp. 45–52, New York, NY, USA, 2018. Association for Computing Machinery.
- [3] 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., *Human Mental Workload*, Vol. 52 of *Advances in Psychology*, pp. 139–183. North-Holland, 1988.
- [4] E. Kruijff, J. E. Swan, and S. Feiner. Perceptual issues in augmented reality revisited. In *2010 IEEE International Symposium on Mixed and Augmented Reality*, pp. 3–12, 2010.
- [5] J. H. Lee, S.-G. An, Y. Kim, and S.-H. Bae. Projective Windows: Bringing Windows in Space to the Fingertip. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–8, New York, NY, USA, 2018. Association for Computing Machinery.
- [6] N. Ltd. Nreal Light - ウェアラブル複合現実グラス. <https://www.nreal.ai/light/>. (Accessed on 09/20/2021).
- [7] Microsoft Corporation. HoloLens 2—Pricing and Options — Microsoft HoloLens. <https://www.microsoft.com/en-us/hololens/buy>. (Accessed on 09/20/2021).
- [8] Microsoft Corporation. HoloLens ヒントを入手 - Microsoft Store ja-JP. <https://www.microsoft.com/ja-jp/p/hololens-%E3%83%92%E3%83%B3%E3%83%88/9pd4cxkk1c47?activetab=pivot:overviewtab>. (Accessed on 09/20/2021).
- [9] Microsoft Corporation. Microsoft HoloLens — Mixed Reality Technology for Business. <https://www.microsoft.com/en-us/hololens>. (Accessed on 09/20/2021).
- [10] Microsoft Corporation. MRTK-Unity Developer Documentation - Mixed Reality Toolkit — Microsoft Docs. <https://docs.microsoft.com/en-us/windows/mixed-reality/mrtk-unity/?view=mrtkunity-2021-05>. (Accessed on 09/20/2021).
- [11] Microsoft Corporation. 空間マッピング - Mixed Reality — Microsoft Docs. <https://docs.microsoft.com/ja-jp/windows/mixed-reality/design/spatial-mapping>. (Accessed on 09/20/2021).
- [12] A. Mulloni, H. Seichter, and D. Schmalstieg. Indoor Navigation with Mixed Reality World-in-Miniature Views and Sparse Localization on Mobile Devices. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, AVI '12, pp. 212–215, New York, NY, USA, 2012. Association for Computing Machinery.
- [13] B. Nuernberger, E. Ofek, H. Benko, and A. D. Wilson. SnapToReality: Aligning Augmented Reality to the Real World. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 1233–1244, New York, NY, USA, 2016. Association for Computing Machinery.
- [14] I. Poupyrev, M. Billinghurst, S. Weghorst, and T. 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*, UIST '96, pp. 79–80, New York, NY, USA, 1996. Association for Computing Machinery.
- [15] D. W. Seo, H. Kim, J. S. Kim, and J. Y. Lee. Hybrid reality-based user experience and evaluation of a context-aware smart home. *Computers in Industry*, 76:11–23, 2016.
- [16] R. Stoakley, M. J. Conway, and R. Pausch. Virtual Reality on a WIM: Interactive Worlds in Miniature. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '95, pp. 265–272, USA, 1995. ACM Press/Addison-Wesley Publishing Co.
- [17] Y. Sugiura, H. Ibayashi, T. Chong, D. Sakamoto, N. Miyata, M. Tada, T. Okuma, T. Kurata, T. Shimamura, M. Mochimaru, and T. Igarashi. An Asymmetric Collaborative System for Architectural-Scale Space Design. In *Proceedings of the 16th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry*, VRCAI '18, New York, NY, USA, 2018. Association for Computing Machinery.

Color-Path: 手本動作の AR 表示によるヘアアレンジ再現支援システム

松井菜摘^{*} 大西鮎美^{*} 寺田 努^{*} 塚本昌彦^{*}

概要. ヘアアイロンは髪に熱を加えることで自由に髪形を変えることのできる道具である。しかし、ヘアアイロンの操作には回転や前後へ傾ける動きが混在しているため、手本映像を見ながら練習するのみではその操作を理解し再現するのが難しい。そこで本論文では、手本とする巻き髪を作るためのヘアアイロンの動かし方をユーザに直観的に理解させるスマートミラー型のヘアアレンジ支援システム Color-Path を提案する。提案システムでは、カメラとヘアアイロンに装着した加速度センサを用いてユーザのヘアアイロンの操作軌道を取得し、システムが保持する手本の操作軌道をスマートミラー上で AR 表示する。評価実験より、提案システムを使った際にヘアアイロンの操作軌道が再現できることを確認した。しかしシステムを使った際に手本の見た目を再現できるとはいはず、ヘアアイロンで髪に熱を加える時間が見た目の再現に関係する可能性があることがわかった。

1 はじめに

ヘアアイロンは髪に熱を加えることで自由に髪形を変えることのできる道具であり、パーマをかけずにカールスタイルを作ることができる。ユーザはヘアアイロンのクリップ部分で髪を挟み、本体を回して加熱部に髪を巻きつけながら熱を加えて髪の形を変える。近年では男性も使うようになり、ユーザ層が広がりつつある[1]。

ヘアアイロンを用いたヘアアレンジは難しい。ヘアアイロンの操作は回転や前後へ傾ける操作が混在しており複雑で、ユーザが髪を上手く巻くためには練習が必要である。ヘアアレンジの練習ではユーザは鏡を見つつ、解説動画や本を見ながらその動きを真似る方法が一般的である。しかし、映像や画像のみで複雑な操作を理解し再現するのは難しく、巻き髪の見た目が不自然になってしまうことが多い。このようにヘアアイロンを用いたヘアアレンジは難しいが、スタイリングを支援するシステムは提案されていない。

そこで本研究では、ヘアアイロンの操作を再現できれば手本のヘアスタイルを再現できるという仮説をもとに、ヘアアイロンの複雑な操作をユーザに直観的に理解させるスマートミラー型のヘアアレンジ支援システム Color-Path を提案する。提案システムでは、PC のカメラとヘアアイロンに内蔵した加速度センサを用いてヘアアイロンの通過位置やヘアアイロンの角度、回転方向といった操作軌道の情報を算出し、図 1 のように直観的に理解できるように単純化してミラー型システムの画面上に提示する。

本論文の貢献は以下の 2 つである。

- 美容師のヘアアレンジの観察結果を設計に取り入れ、美容師の巻き方を手本としてヘアア



図 1: 提案システムの提示画面と使用時の様子

イロンの操作軌道を単純化して提示し、ユーザに直観的に理解させるスマートミラー型のヘアアレンジ支援システム Color-Path を提案した。

- 提案システムの提示によって手本のヘアアイロンの操作軌道を再現できることを確認した。

本論文では以降、2 章で関連研究を紹介し、3 章で提案システムについて述べる。4 章で評価実験について説明する。5 章で実験結果と考察を述べ、最後に 6 章で本論文をまとめる。

2 関連研究

2.1 美容に特化したスマートミラー

美容に特化したスマートミラー製品は数多くある。HiMirror 社のスマートミラー HiMirror Mini には、鏡に映った顔から毎日の肌の状態を分析し、最適な化粧品を提案する機能が実装されている[2]。また、Mirror Roid 社のスマートミラー Mirror Roid はヘアスタイルやヘアカラーをシミュレーションできるスマートミラーであり、美容室などすでに導入さ

^{*} 神戸大学大学院工学研究科

れている[3]。また、ユーザの化粧技術を支援するスマートミラーの研究も存在し、コーネル大学のスマートミラー SnowBeauty では、理想の顔に近づくための最適なファンデーションの色味や質感、テクニックなどを提案し、ユーザに化粧に関する知識を提供する[4]。Iwabuchi らはメイク箇所への自動ズームや照明の変化をシミュレーションする機能を実装した Smart Makeup Mirror を提案している[5]。

これらのようにスマートミラー製品は美容分野やヘアスタイルの提案といった様々な分野に普及しているが、ヘアアレンジの具体的な方法は提示されていない。また、システム使用後のユーザの見た目の変化やヘアアレンジの出来栄えについて評価されていない。そこで本研究では、ユーザにヘアアイロンの操作軌道を提示するスマートミラーシステムによって、ユーザにヘアアイロンの操作軌道を提示し手本のヘアアレンジの見た目も再現できるようにヘアアレンジ支援を行う。

2.2 動作習得支援の研究

スポーツやリハビリにおける目標とする動作を単純化し、動作軌跡を表示しユーザ自身の動きに重ね合わせて提示することでユーザの自主練習を支援するシステムの研究は数多くある。Anderson らは YouMove という全身の動作習得支援 AR ミラーを提案している[6]。システム内では単純で認知負荷が低くなるようにデザインされた提示が複数用いられている。例として、ユーザの静止時の姿勢が手本の姿勢と異なるときにフィードバックを与える姿勢ガイド、動きが止まらない部分はそのまま動画を再生し、止まったところから次の動きのルートをリボン型のガイドで提示する動きガイドというような提示が提案されている。Sousa らは SleeveAR という、腕に装着したマーカをトラッキングしてリハビリーション時の腕動作を単純なアイコンで表現し、それらを床に映し出すことでリアルタイムで動作のフィードバックを行うシステムを提案している[7]。従来のビデオを用いたフィードバックと比較した結果、提案システムを用いた方が手本のリハビリーション動作をうまく再現でき、アンケート調査においても提示が分かりやすかったという意見が多数得られた。

これらの研究から、複雑な動作を単純な提示に置き換えることでユーザが動作を直観的に理解できると考えられる。本論文で提案するシステムでは、ヘアアイロン操作にこれらの手法を適用し、ヘアアイロンの回転動作や姿勢を単純化して提示することで動作再現支援を行う。さらに提案システム使用時に巻き髪の見た目を再現できているかを評価する。

3 提案システム

本研究では、手本のヘアスタイルを再現させるためにヘアアイロンの動かし方をユーザに直観的に理解させ、ヘアアレンジを支援することを目的とする。ユーザは自分の姿と手本を見ながらアレンジを進めるため、鏡のように自分の姿を見ながら使用できるデザインやユーザのアレンジの進み具合に合わせて次に何をすればよいかが分かる指示が行われる必要がある。手本の映像を見ることは手順の大まかなイメージをつかむために重要であると考えられるため、映像も参考にできた方が良いと考えられる。毎朝のヘアセットでの使用を想定しており、ユーザがシステム使用時に理想の髪形を再現することと未経験の巻き方でも何度も練習せずとも髪形が再現できることを目的とし、これによってユーザが新たなヘアスタイルに気軽に挑戦できる環境を提供することを目指す。したがって、ユーザのヘアアイロン技術の向上やそれによるシステムからの離脱を目的としない。

3.1 指示内容の決定のための予備調査

指示の内容を決定するため、ヘアアレンジに熟練しており、日常的にヘアアイロンを使って髪をセットしている女性美容師 2 名に自分自身の髪をアレンジしてもらい、そのときの動画を観察し、美容師の意見を聴取した。髪の巻き方は普段から行っている慣れた髪形を形成してもらった。また、美容師でない一般のヘアアイロン使用者 3 名に対して、手本の動画を参考にしながらヘアアレンジを行ってもらい、そのときの様子をビデオ撮影して観察し、よくある失敗を聴取した。

ヘアアレンジ中の動画では、美容師は 2 名とも常に鏡に映る自分の姿を見てヘアアイロンを操作していた。さらに、美容師は顔の位置に対して常にほぼ一定の距離でヘアアイロンを回転させており、それが回転の安定性などに寄与して見た目の再現性を高めていると考えられたため、顔位置に対するヘアアイロンの相対位置を提示する必要があるとわかった。

一般的のユーザからは髪をアイロンで挟んだ直後に回転方向を間違えてしまうという意見が多数得られたため、回転方向の指示も必要であるとわかった。加えて、美容師はヘアアイロンの姿勢を常に一定に保って髪を巻いていた。それに対して一般ユーザは髪を挟む位置および角度と回転方向が一致していない場合が多く、巻いている途中でヘアアイロンの姿勢が変化して髪がたわんでしまうことがあった。髪の巻き方は、同じ位置を巻いてもヘアアイロンの姿勢によって形成される髪の見た目が大きく異なる。ヘアアイロンの姿勢が横向きであれば外はねや内はねが形成され、姿勢が縦向きであれば螺旋状の巻き髪が形成される。よって、ヘアアイロンの姿勢も見た目の再現に重要であると考えた。

以上より、ヘアアイロンの動かし方は巻き髪の見

Color-Path: 手本動作の AR 表示によるヘアアレンジ再現支援システム

た目に影響を及ぼすと考えられ、特にヘアアイロンの位置、回転、姿勢の情報が重要であることがわかったため、本論文で提案するシステムではこれらをユーザが直観的に理解できるように指示を与える。

3.2 システム要件

想定環境と予備調査の結果をふまえ、以下の要件を満たすシステムを提案する。

- ヘアアレンジ中の自分自身の姿と指示を同時に見ることができる
- 次に動かすべきヘアアイロンの位置、回転、姿勢の情報が直観的に理解でき、動かし方が再現できる

システム要件に基づき、本研究ではヘアアイロンの操作軌道をユーザに提示するスマートミラー型のヘアアレンジ支援システムを提案する。PC画面を鏡のようにしてAR表示で指示を与えることで、ユーザは自分の姿を見つつ、画面上の提示によってヘアアイロンの動かし方の指示を受けられる。

提示する操作軌道であるヘアアイロンの位置、回転、姿勢の情報は、関連研究[6], [7]を参考に単純化することで提示によって認知負荷が高まることを避けて直観的に理解可能なデザインにする。具体的には、ヘアアイロンにカラフルなシールを貼り、同じ色の提示に合わせていくことで位置や回転が合うようにする。姿勢は小型な加速度センサを用いて取得し、ヘアアイロンの傾きに合わせて変化する表示を用意し、それを目標に合わせるようにして一致させる。提示に合わせて動かせたときに提示を次に切り替える。

3.3 システム構成

図2に提案システムの構成を示す。システムはヘアアイロンとPCで構成され、PC上で鏡のように画面にユーザの顔を映しながら手本の操作軌道を提示する。ヘアアイロンには、図3のように先端1周に4色の位置認識用シールを貼り、持ち手に加速度センサを取り付けてある。PCではWebカメラでシールの色検出を行いヘアアイロンの回転情報と位置を、無線で受信した加速度センサ値からヘアアイロンの傾きを表す姿勢角をそれぞれ算出する。提案システムでは、あらかじめ手本とするヘアアレンジにおけるヘアアイロンの回転や位置などの操作軌跡を保持しておく。システムがユーザの操作軌道と手本の操作軌道との一致を検出し提示を切り替えることで、ユーザのヘアアレンジ進度に合わせてPC上に手本の操作軌道を提示する。システム画面はUnity[9]で作成し、ヘアアイロンはI-ne社のセラミックカーラーアイロン32mm[10]、マイコンはM5Stack社のM5StickC[11]、加速度センサは、M5StickCに搭載されたmpu-6886[12]を用いた。

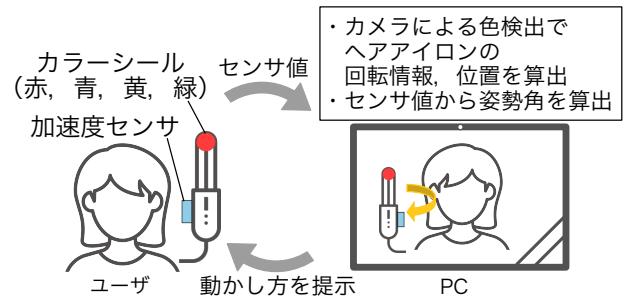


図 2: Color-Path のシステム構成



図 3: 提案デバイス

3.4 ヘアアイロンの位置と回転および姿勢角の算出方法

ヘアアイロンの位置は、ヘアアイロンの先端に貼ったカラーマーカの画面内の位置とする。カラーマーカの位置は画面からマーカの色のRGB値を満たす画素を検出し、その座標から算出する。ヘアアイロンの回転は検出するマーカの色の変化から算出する。

ヘアアイロンの姿勢角は、加速度センサの値から算出する。提案システムで算出する姿勢角はPCの画面に対するヘアアイロンの左右の傾きと前後の傾きである。

3.5 提示情報

ヘアアイロンの位置と回転および姿勢角の提示

提案システムの画面を図4に示す。2種類の提示を組み合わせて操作軌道を提示する。

ゲート:ヘアアイロンの位置と回転を提示する。色のついた枠に同色のカラーシールが通るようヘアアイロンの先端を合わせるための円状の目印である。**スティック:**ヘアアイロンの姿勢角を提示する。ヘアアイロンの目標姿勢を示す棒状の提示である。左右の傾きは、スティックの傾きで示す。図5のように、ヘアアイロンの左右の傾きが手本に近いとスティック内の色が変化する。また、前後の傾きは、スティック内部の着色部分の増減で示す。ヘアアイロンの左右の前後の傾きが手本に近いとスティックの横にある三角形のアイコンの色が変化する。

提示の切り替えは、ユーザのヘアアイロンが目標位置と目標姿勢角に一致しているとシステムが判定したときに行う。本システムでは、一連のヘアアレンジ

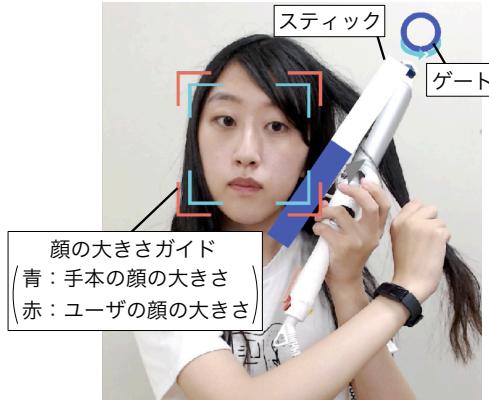


図 4: 提案システムの提示画面



図 5: スティックの変化で傾きの一致度を表す提示

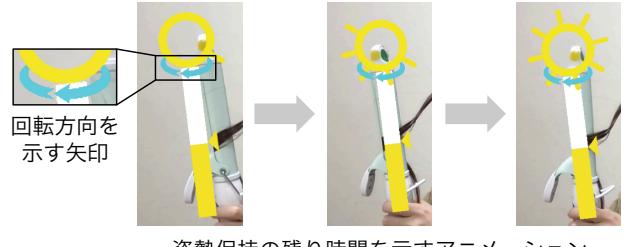
シジ動作を分割し、動作の指示を1つずつシステム画面に提示する。それぞれの動作にヘアアイロンの目標位置と目標姿勢角が設定されており、ユーザのヘアアイロンがそれらに一致しているとシステムが判定したときに次の動作の提示に切り替える。

ヘアアイロンと顔の相対位置を再現するための提示

提案システムでは提示とユーザの顔の相対位置が手本動作のヘアアイロンと顔の相対位置と同じになるように、映像から算出したユーザの顔位置に合わせて提示を移動させる。また画面と顔との距離も手本と同じにするために、図4に示す顔の大きさガイドを作成した。青色の枠で提示される手本の顔の大きさに、赤色の枠で提示されるユーザの顔の大きさをユーザ自身が合わせることで、画面と顔との距離を調整する。画面に映るユーザの顔の大きさが手本の顔の大きさに合っている場合のみゲートとスティックの提示が画面に表示される。

ヘアアイロンの姿勢保持の時間と回転方向の提示

髪に巻き髪の形をつけるには、ヘアアイロンを一定時間静止させ、巻く毛束に熱を加える必要がある。そこでユーザに髪の加熱時間を示すために、図6のようなゲート周辺の強調線の本数でヘアアイロンの姿勢を保持させる時間を示すアニメーションを作成した。また、ゲート下部に示す矢印でヘアアイロンの回転動作とその方向をユーザに提示した。



姿勢保持の残り時間を示すアニメーション

図 6: 姿勢保持の時間と回転方向の提示



図 7: 実験の様子

4 評価実験

4.1 実験方法

本実験では、システムによって被験者が手本の操作軌道を再現できているか、手本の操作軌道を再現できたときに見た目も再現できているかを評価した。被験者はシステムによる提示がない場合とある場合でヘアアレンジを行った。手本の髪形は2つ用意した。それらは美容師が髪全体をヘアアレンジしたものであり、それぞれ異なる髪形である。

被験者は20代女性4名で、日常的にヘアアイロンを使用している。ヘアアレンジする髪の長さ、髪質を統一するために、被験者はウィッグを装着してヘアアレンジを行った。被験者はシステムの提示なしの場合と提示ありの場合で手本の巻き方を再現するように髪全体をヘアアレンジした。各被験者の提示なしとありの順番と手本の髪形は以下のようにランダマイズした。

被験者 A 提示なし髪形1 → 提示あり髪形2

被験者 B 提示あり髪形2 → 提示なし髪形1

被験者 C 提示あり髪形1 → 提示なし髪形2

被験者 D 提示なし髪形2 → 提示あり髪形1

実験の様子を図7に示す。実験中はシステム画面を録画し、実験後に被験者の巻いたウィッグの写真を4方向から撮影した。またシステムに関する質問紙調査を行った。なお、本実験は神戸大学大学院工学研究科倫理委員会の承認を得て行った。

4.2 評価指標

提案システムによって被験者がヘアアイロンの操作軌道を上手く再現できているときに巻き髪の見た目が良くなれば、提案システムはヘアアレンジの支援を行うインターフェースとして有効であるといえる。

システムを用いて手本の操作軌道を再現できたかは、各被験者の手本との操作軌道のズれをシステムの提示がない場合とある場合で比較して評価する。手本との操作軌道のズレは、顔位置に対するヘアアイロンの相対座標とヘアアイロンの姿勢角の手本との差から評価する。顔位置に対するヘアアイロンの相対座標の差とヘアアイロンの姿勢角の差がシステムの提示がない場合よりもある場合の方が小さければシステムの提示が動作の再現に有効であるといえる。

軌道のズレの算出方法について、カラーマーカのいずれかの色が真正面に向いたときを通過点とし、その通過点での顔位置に対するヘアアイロンの相対座標とヘアアイロンの姿勢角の値の手本との差から軌道のズれを算出する。実験で用いた2つの手本では巻く毛束の数や1束ごとの巻きの種類がそれぞれ異なりヘアアレンジ全体の操作軌道の比較は難しい。そのため今回は2つの手本動画間で巻き方の種類、毛束の左右の位置、ヘアアイロンを回す方向が同じである髪1束を選ぶ。その1束を巻くときの顔位置に対するヘアアイロンの相対座標とヘアアイロンの姿勢角の手本との差の1通過点あたりの平均を算出して比較を行う。システム画面の座標空間と実験環境でのシステム画面に映る範囲を図8に示す。

被験者が手本の操作軌道を再現できたときに見た目も再現できたかは、被験者の操作軌道の再現度と見た目の良さの関係から評価する。巻いた髪の見た目の再現度がシステムの提示がない場合よりもある場合に高くなれば、提案システムの提示が見た目の再現に有効であるといえる。見た目の評価は手本の髪形と実験で被験者が巻いた髪形の写真を見比べて、被験者が巻いた髪形が手本の髪形の再現として許容できるかできないかをヘアアイロンを使ってヘアアレンジをしたことのある、被験者を除く4名の女性に判定させた。

また、提案システムの提示がユーザにとって理解しやすいことも重要であるため、質問紙調査の結果から提示に関する主観的な分かりやすさも評価する。被験者にはゲート、スティック、顔位置の提示がそれぞれヘアアレンジで役立ったか、提示が顔に合わせて動くことに違和感を感じてヘアアレンジはやり辛かったかをそれぞれ回答させた。

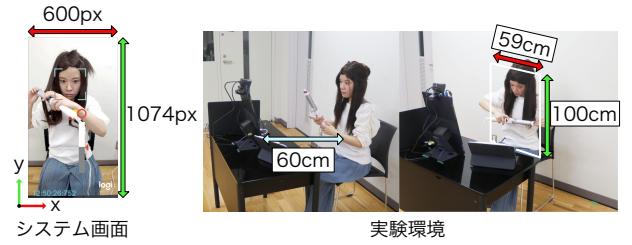


図 8: 座標空間の定義

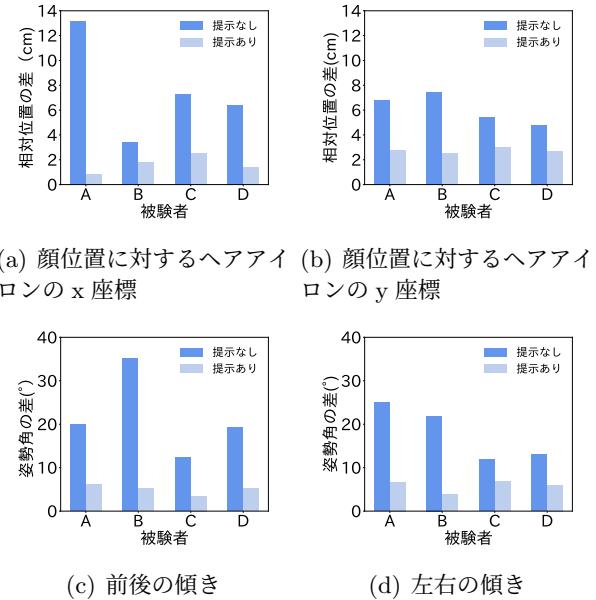


図 9: ヘアアイロンの操作軌道の手本との差

5 実験結果と考察

5.1 操作軌道の再現

顔位置に対するヘアアイロンの相対座標とヘアアイロンの前後左右の傾きの手本との差を図9に示す。被験者4名の全員が全ての評価項目でシステムの提示ありの場合に手本との差が小さくなっていること、システムの提示なしよりも提示ありの方が手本の操作軌道を再現できていた。よって、システムの提示が操作軌道の再現に有効であったと考えられる。

5.2 操作軌道と見た目の関係

被験者が手本の操作軌道を再現できたときに見た目が再現できたかを確かめるために、手本と被験者の試行における操作軌道の差、見た目の差の関係性を調べた。各被験者のそれぞれの試行で巻いた髪形が手本の再現として許容できるかどうかを実験の被験者と異なる女性4名に判定させた。判定する人はそれぞれの試行で巻いた髪形とその手本の髪形の後ろ姿の写真2枚を提示し、手本の再現として許容できるかどうかを逐次回答させた。結果を表1に示

表 1: 髮形の写真と許容できるか判定した結果

		提示なし		提示あり	
	手本	被験者A	被験者B	被験者C	被験者D
見た目の許容(人)	できる	2	4	3	3
	できない	2	0	1	1
被験者	手本	被験者C	被験者D	被験者A	被験者B
	髪形1				
見た目の許容(人)	できる	3	0	1	0
	できない	1	4	3	4
被験者	髪形2				

す。髪形1では提示なしと提示ありで手本として許容できるという回答が同数であった。提示ありの被験者C, 被験者Dの見た目に関して、「緩く巻かれているという全体的な雰囲気は一緒だと感じる」「巻き具合が手本に似ている」という意見があった。しかし髪形2では提示ありよりも提示なしの方が許容できるという回答が多かった。提示ありの被験者A, 被験者Bの見た目に関して、「全体的に強く巻きすぎているように感じる」「全体のシルエットが違う」という意見があった。

そこで、巻きの強さに影響する要素である髪に熱を加える時間について分析した。操作軌道の再現度を調べるために用いた髪1束を巻くときの手本よりも超過した時間を図10に示す。髪形1を提示ありで再現した被験者C, Dより髪形2をシステムの提示ありで再現した被験者A, Bが巻き終わるまでの時間が手本よりも長かった。これは被験者が提示に合わせるのに時間がかかったことで手本動作よりも髪に熱を加える時間が長くなり、これによって巻きが強くなったと考えられる。したがって、今後システム使用時に髪を1束巻く所要時間が手本の所要時間に近くなるように提示方法を工夫する必要がある。

5.3 提示に関する質問紙調査

提示に対する主観的な評価としてゲートと顔位置の提示が役に立ったかという質問に対しては4名とも全ての提示が役立ったと回答した。しかし、スティックの提示が役立ったかという質問に対して4名は役立ったと回答したが、1名が提示に合わせるのが難しかったため役立たなかったと回答した。

5.4 議論

提案システムでは髪全体のヘアアレンジの操作軌道を1束ずつに分けて提示したが、それぞれの毛束の位置や毛量をユーザーに提示できていない。これらが手本と異なるとき、毛束におけるカールの位置も

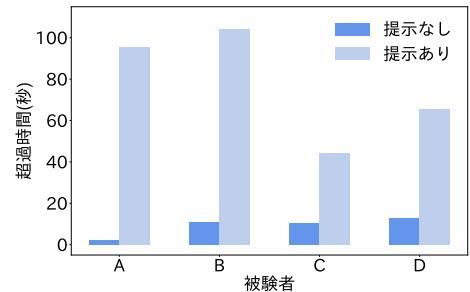


図 10: 髪1束を巻くときの手本よりも超過した時間

手本と異なるため再現度が低くなる可能性がある。また、見た目の許容について、全体の雰囲気で判定するという意見が多くかった。提案システムの提示は髪1束の見た目の再現度を高めるようなものであり、全体の雰囲気やバランスの再現は考慮できていない。毛束の位置や毛量は巻き髪の仕上がりに影響する要素であり、これらが全体の雰囲気やバランスに影響する可能性がある。今後これらの提示の実装についても検討する。加えて見た目の再現度をより詳細な基準を設けて評価し、髪形のどの部分が雰囲気やバランスの良さに影響するのかを評価する。

6 まとめ

本論文ではヘアアレンジの手本動作を再現できるように、ユーザーにヘアアイロンの操作軌道を提示するスマートミラー型のシステムColor-Pathを提案した。提案システムでは端末のカメラと加速度センサからヘアアイロンの位置座標と姿勢角を取得する。取得した値に基づき、提案システムはユーザーのヘアアレンジ進度に合わせて提示を適切に切り替えてヘアアイロンの操作軌道をスマートミラーの画面上にAR表示する。操作軌道の提示にはヘアアイロンの位置と回転を示すゲート、ヘアアイロンの姿勢角を示すスティックを用いた。ユーザーの顔に対するヘアアイロンの相対座標が手本と同じになるように、操作軌道の提示はユーザーの顔に追従させて表示した。

提案システムの提示の有効性を評価した結果、ユーザーは提案システムを用いてヘアアイロンの操作軌道を再現できた。しかし、提案システムを用いて見た目も再現できるとはいえたかった。今後は、ヘアアイロンの操作軌道と見た目の再現度の関係についてより詳しく調査し、見た目の再現度を高めるためにシステムを改良していく。

謝辞

本研究の一部は、情報処理推進機構(IPA)の2020年度未踏IT人材発掘・育成事業、及びJST CREST(JPMJCR18A3)の支援を受けて実施されたものである。ここに記して謝意を表す。

参考文献

- [1] 花王: ヘアスタイルとスタイリング技術の歴史, url<https://www.kao.com/jp/haircare/history/14-2/>.
- [2] HiMirror: HiMirror Mini, <https://www.himirror.com/ja-JP/product/HiMirrorMini>.
- [3] Mirror Roid: Mirror Roid, <https://www.mirrorroid.co.jp/products/>.
- [4] KOSÉ Corporation: 肌分析機能内蔵の「Snow Beauty Mirror」でデジタル技術を活用した最先端のパーソナライズ提案を開発パナソニックと協業へ向けた実証実験, <https://www.kose.co.jp/company/ja/content/uploads/2019/12/191206.pdf>.
- [5] E. Iwabushi, M. Nakagawa, and I. Sii: Smart Makeup Mirror: Computer Augmented Mirror to Aid Makeup Application, *Proc. of the 13th International Conference on Human-Computer Interaction (HCI 2009)*, pp. 495–503 (July 2009).
- [6] F. Anderson, T. Grossman, J. Matejka, and G. Fitzmaurice: YouMove: Enhancing Movement Training with an Augmented Reality Mirror, *Proc. of the 26th Annual ACM Symposium on User Interface Software and Technology (UIST 2013)*, pp. 311–320 (Oct. 2013).
- [7] M. Sousa, J. Vieira, D. Medeiros, A. Aeseno, and J. Jorge: SleeveAR: Augmented Reality for Rehabilitation using Realtime Feedback, *Proc. of the 21st International Conference on Intelligent User Interfaces (IUI 2016)*, pp. 175–185 (Mar. 2016).
- [8] CREATE ION: ヘアアイロンメーカー公式コレクション・アイロン巻き方テクニックまとめ, <https://createion.jp/column/kote-technique>.
- [9] Unity Technologies: Unity, <https://unity.com/ja>.
- [10] I-ne: SALONIA セラミックカールヘアアイロン 32mm, <https://salonia.jp/product/hair/iron/curl/>.
- [11] M5Stack: M5StickC, <https://shop.m5stack.com/collections/stick-series/products/stick-c?variant=17203451265114>.
- [12] InvenSense: MPU-6886, https://m5stack.oss-cn-shenzhen.aliyuncs.com/resource/docs/datasheet/core/MPU-6886-000193%2Bv1.1_GHIC_en.pdf.

時間歪曲機能を用いたVRスキートレーニングの拡張

松本 高* Erwin Wu* 小池 英樹*

概要. 近年、VRを用いた様々なスポーツのシミュレーションが作成され、仮想環境でのスポーツのトレーニングを行うことが可能になっている。VRを用いる利点として、非現実的な機能を実装できるということが挙げられる。時間の経過速度を実際よりも遅くすること(時間歪曲と呼ぶ)によって、動作に余裕を持たせる、という機能も非現実的な機能の一つである。本論文では、屋内用のスキーシミュレータとVR機器を組み合わせたスキーシミュレーションシステムを作成し、時間歪曲機能を実装した。時間歪曲を行う方法にも、一定の倍率で静的に遅くする、ユーザの状況に合わせて動的に遅くする、というように複数の手法が考えられる。時間歪曲を行わないベースライン条件と4つの時間歪曲条件を比較する評価実験を行ったところ、時間歪曲がスキーにおいても有効であることが示された。また、静的な条件の方が動的な条件と比べてより好ましい条件であるという結果が得られた。

1 はじめに

スキーは人気の高いスポーツであるが、季節と場所に対する制限が多く、トレーニングを行うための障壁が多い。この問題を解決するため、スキーのシミュレーションシステムを作成し、屋内でもトレーニングを行えるようにした研究は数多く行われている。近年では特に、スキーのシミュレーションマシンとVR機器を組み合わせることで、トレーニングへの没入感を高める研究が盛んである[10, 11, 14]。

スポーツのトレーニングとして、お手本となる熟練者の動作をよく観察し、その動作を模倣するという練習方法は一般的によく用いられている。ただし、初心者が熟練者の動作を模倣する場合など、両者の実力の差が大きすぎる場合には、模倣を効果的に行えない可能性が考えられる。そこで、様々なスポーツのシミュレーションにおいて、仮想環境内の時間の経過速度を遅くすること(以下、時間歪曲)によって模倣を行いやすくした研究が行われており、練習の効果が高まったという実験結果が報告されている[12, 13]。

これらのこと踏まえ、本論文では図1に示すようなスキーのシミュレーションにおいても、時間歪曲を行うことがスキーの上達の手助けになることを検証する。また、速度の変更方法にも一定の倍率で遅くする場合や、ユーザの状況に合わせて、動的に倍率を変更する方法が考えられる。評価実験において、倍率の異なる2つの静的な条件、および2つの動的な条件で練習を行うことによる、スキー熟練者の滑走との差分や、ユーザの主観評価、シミュレータ酔いの観点で比較を行った。

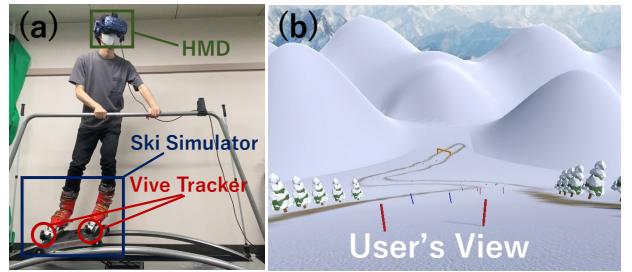


図 1. システムの概要 (a), 仮想スキーコースのユーザ視点 (b).

2 関連研究

2.1 スキーシミュレーション

スキーの技術向上させるための、様々なシミュレーションシステムが提案されている。Kobeissiら[8]は、バランスボードにモーションセンサを取り付け、検知されたバランスボードの回転によって操作を行うスキートレーニングゲームを提案した。Aleshinら[2]は、SkyTechSportのスキーシミュレーションマシンと仮想のスキーコースを表示するスクリーンを組み合わせたトレーニングシステムを提案し、このシステムが初心者だけでなく、プロのアスリートのトレーニングにも用いられると述べた。Koら[7]は、スキーのエクササイズゲームにおいて、VRと非VR環境でのユーザの運動能力や集中力がどのように影響されるかを調査した。野澤ら[14]は、VR空間においてユーザの前にスキー熟練者のモーションデータをコーチとして表示し、ユーザがその動きを模倣することによってスキーのトレーニングを行うことができるシステムを提案した。このシステムでは、様々な方法を用いてコーチとユーザとの間の差分を可視化している。一方で、熟練者の動きが速

Copyright is held by the author(s).

* 東京工業大学

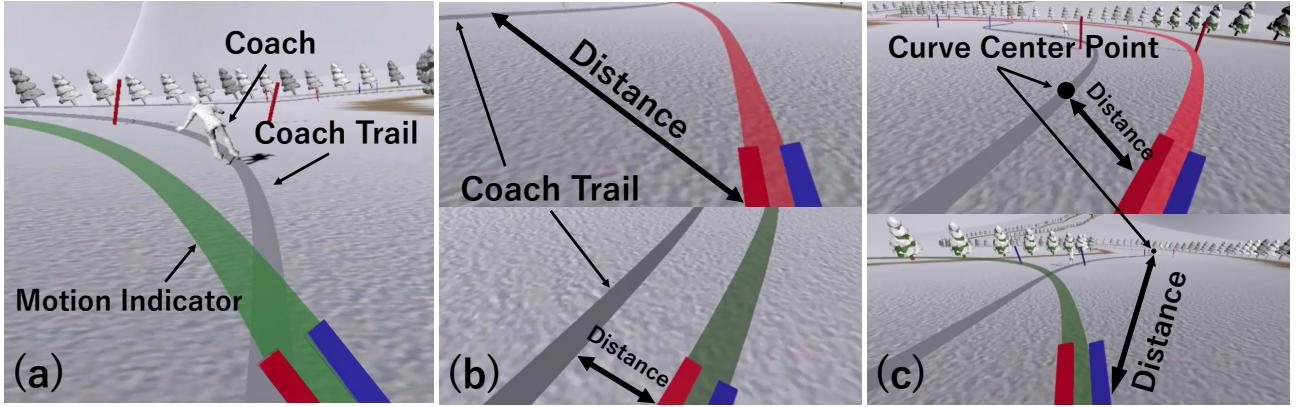


図 2. HMD に表示される情報 (a), 動的な時間歪曲条件 2 つ (b)(c).

(b) では Coach Trail との距離が大きくなるほど, (c) では Curve Center Point に近づくほど時間の経過速度が遅くなっている (経過速度が遅いほど, MotionIndicator の色が緑から赤に変化する).

すぎることによってユーザが付いていけなくなる問題が指摘されている。Wu ら [10, 11] はこのシステムを拡張し、差分の可視化機能を追加した上で、どの機能がより高い効果を得られるかを定量的に評価した。Wu らの実験によれば、差分を画面上のグラフにプロットするような複雑な可視化手法は、ユーザが提示された情報を上手く扱えず、効果が低くなつたことが示唆されている。

本論文のシステムは、VR を用いてコーチの動作を模倣する支援を行っている点で野澤、Wu らのシステムと類似している。一方で、ユーザが情報を処理する余裕がないという問題点を、時間歪曲機能を実装することにより解決を図っている点で異なる。

2.2 時間歪曲を用いたトレーニングシステム

様々なスポーツにおいて、時間歪曲を用いて動作の学習を支援する研究が行われている。川崎ら [13] は、VR 空間で再現したけん玉において玉の速度を実際よりも遅くし、徐々に速くしていくことによってけん玉の技を練習できるシステムを提案した。川崎らの実験により、多くのシステム体験者が 5 分間程度の練習で今までに成功したことのない技を習得できることが示された。Adolf ら [1] は、VR 環境でのボールジャグリングにおいて、重力加速度を小さくすることによるユーザのパフォーマンスやモチベーションへの影響を調査した。濱西ら [5] は、テニスボールなどの高速で移動する物体を VR 空間では認識しにくくなる問題を、ユーザの動きに応じて時間の経過速度を制御することにより解決している。Bansal ら [3] は、自身の動きに応じて時間の経過速度が変化するような VR 体験をする前後において、マウスカーソルをお手本と同様に動かす課題をユーザに課す実験を行った。その結果、ユーザに動作時間を短く見積もるような時間認識の調整が行われることが示され、この現象をリハビリテーションへと

適用できる可能性を示唆した。Wu ら [12] は、VR 空間での卓球のスピンのかかったボールを打ち返すトレーニングにおいて、単に時間の経過速度を遅くするだけでなく、様々な視覚的フィードバックを提示するシステムを提案している。

これらの関連研究は、時間歪曲がトレーニングの効果を高める可能性を示している。一方で、これらのほとんどの研究は、単一の時間歪曲機能のみを検証しており、複数の時間歪曲機能を比較した研究は少ない。

3 システムデザイン

本章では、検証に用いたシミュレーションシステムの構成および実装について説明する。

3.1 ハードウェア

本論文で作成したシミュレーションシステムは、図 1(a) に示すように、屋内用のスキーシミュレータ¹ (PRO-SKI SIMULATOR) 及び HMD (HTC VivePro), 2 つの Vive トラッカー, 2 つのベースステーションから構成される。Vive トラッcker をシミュレータの左右のスキー板の先端に取り付けることで、スキー板の回転を検出できるようにする。

3.2 Unity 上での実装

VR を用いてトレーニングを行うために、Unity にて図 1(b), 図 2(a) に示すようなスキーの仮想コースを作成した。ユーザは HMD を通してこのコースを見て、シミュレータのスキー板を動かすことにより仮想コース内を滑走できる。

操作性を単純にするため、仮想コース内の移動は図 3 に示すように、スキー板の地面と垂直な方向に

¹ <https://www.ski-simulator.com/product/power-ski-simulator/>

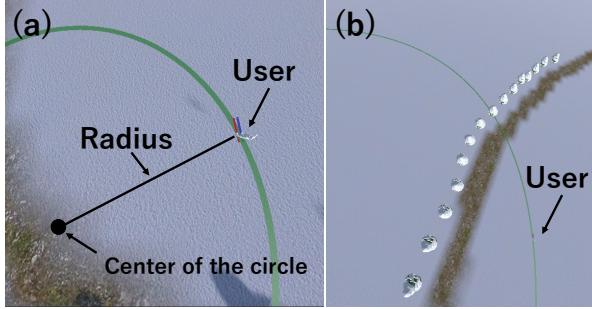


図 3. スキー板のロール角が大きい場合は急なターン
(a), 小さい場合は緩やかなターン (b) となる.

対するロール角が大きくなるほど回転半径が小さくなるような水平面での円運動とした。より移動をわかりやすくするために、ユーザが現在の足の角度を固定した場合に進んでいく軌道を図 2(a) のように画面上に表示した(以下, MotionIndicator)。

実際のスキーに近づけるために、ターンの前半部分(谷回り)では加速し、ターンの後半部分(山回り)では減速が起こるようにした。後述するスキーの熟練者の意見を参考にし、仮想コース内での移動の初速度は 63km/h とし、初速度の $0.8\sim1.2$ 倍の範囲で速度が変動するようにした。

3.3 熟練者データの取得

熟練のスキーヤー 1 名にこのシミュレーションシステムを使用してもらい、コーチデータとして記録した。前述の HMD、左右のスキー板の Vive トラッカーに加えて、両手、腰にトラッcker を装着し、計 6 点のモーショントラッキングを行った。記録したモーションデータを 3D モデルとして再生し、コーチが滑走したルート上を動かすことによって、コーチのリプレイデータとして再現ができる。

実際のスキーでは人々が走ったルート上の雪がかき分けられ、跡が生じる。初心者は熟練者の跡をなぞることにより、適切なルートを滑走することができる。これを仮想コース内にて再現するため、コーチが通過するルートを灰色の曲線(以下, Coach Trail)として表示した(図 2(a) 参照)。

3.4 時間歪曲機能の実装

Unity のパラメータである TimeScale の値を変更することにより、シミュレータ内の時間の経過速度を変更することができる。通常の速度の TimeScale の値を 1 として、1 よりも小さい値に変更することで時間の経過速度が遅くなる。これにより、ユーザの移動速度やコーチの姿勢データの再生される速度が遅くなるため、ユーザの動作に余裕が生まれることが期待される。HMD の画面上では時間の経過速度が変化していることを視覚的にもわかりやすくするために、MotionIndicator を TimeScale の値が 1 の

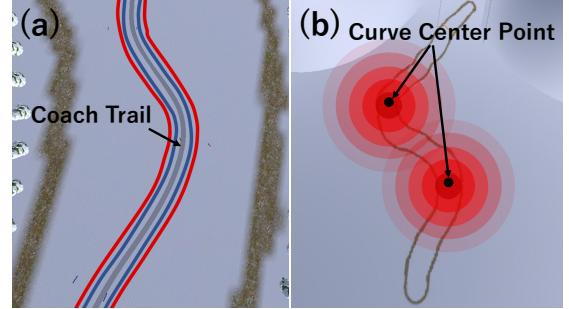


図 4. (a) では青線の内側を滑走している場合は上手く模倣が行えているとみなされ時間の速度が増加、赤線の外側では速度が減少する。(b) では同心円の中心に近づくほど時間の速度が減少する。

時に緑色で表示し、TimeScale の値が小さくなるにつれて赤色に変化していくようにした(図 2 参照)。

時間の経過速度を変更する方法として、一度の滑走中には一定の倍率とする方法や、ユーザの状況に合わせて、滑走中にも動的に速度を変更する方法など、様々な手法が考えられる。本論文では、以下の 4 つの時間歪曲条件を設定した。

Constant0.6 / Constant0.8

それぞれ時間の経過速度を常に 0.6 倍、0.8 倍にする。なお、速度倍率は、川崎らの”けん玉できた!VR”[13] の例を参考にし、0.2 倍刻みとした。ただし、0.4 倍速以下では一度のカーブに掛かる時間が長すぎることによって、ユーザがカーブをするための姿勢を長時間維持する必要があり、大きな負担が生じることが予備実験でわかったため、今回は 0.6 及び 0.8 倍速の 2 条件と限定した。

DistanceToCoachTrail

ユーザと Coach Trail との間の距離が大きい場合、すなわちコーチの滑りを上手く模倣できていない場合に時間の経過速度を遅くする(図 2(b), 図 4(a) 参照)。毎フレーム時間の経過速度を変更するとシミュレータ酔いを激しく引き起こすことが予想されるため、100 フレーム毎(90Hz)にユーザと Coach Trail との間の距離の平均を計算するようしている。

DistanceToCurvePoints

スキーのコースにおいて、カーブしている部分は直線的な部分と比べてより複雑な操作が必要となり、コーチの動作を模倣する難易度も高くなることが予想される。そのため、この条件では仮想コースのカーブの中心部分(Curve Center Point)に近づくほど時間の経過速度を遅くする(図 2(c), 図 4(b) 参照)。時間の経過速度の倍率が一定である前者 2 つが静的な条件であるのに対し、後者 2 つは滑走中に倍率が変化していくため、動的な条件であるといえる。また、動的な条件 2 つの TimeScale は 0.5 から 1.0 を 0.1 刻みで変動するようにした。

4 評価実験

4.1 仮説

本論文では、以下の仮説を検証するために評価実験を行った。

H1: 時間歪曲による練習効果

時間の経過速度を遅くして練習することによりスキーの上達が見込める。

H2: より低い速度倍率による練習効果

時間の経過速度が遅いほどコーチの動作を模倣しやすくなり、練習効果が高まる。

H3: 動的な倍率変更による練習効果

ユーザの状況に応じて動的に時間の経過速度を変更することにより、練習効果が高まる。

4.2 実験手順

被験者に実験の説明をした後、スキーブーツを履きシミュレータに乗ってもらった。その後、以下の手順にて実験を行った。

1. 事前にVRやシミュレータの動かし方に慣れてもらい、学習率による影響を緩和するための、5分程度の練習を行う。この練習においてのみ、コーチの姿勢とCoach Trailを表示しないことで、被験者がシミュレータの動かし方に慣れるための練習に集中できるようにした。
2. ベースライン条件にて計測を行う。表示されるコーチの姿勢を参考にしながら、できる限りCoach Trailと同じルートを滑るように被験者に指示した。
3. ベースライン条件終了後に、図5に記す5つの質問(7段階のリッカート尺度)及び任意の自由記述でのコメント、シミュレータ酔いの程度を計測する質問(SimulatorSicknessQuestionnaire(SSQ)[6])に回答する。
4. 上述した4つの時間歪曲条件のうち、1つの条件下での練習を2分程度行う。
5. 再びベースライン条件の計測を行う。
6. 練習を行った条件について、3.と同様の質問に回答する。
7. 手順4.~6.を、条件を変えながら4回繰り返す。練習を行う条件の順番は被験者ごとにBalancedLatinSquare[4]に従ってカウンターバランスを行った。

手順2.及び5.にて計測したベースライン条件の結果を比較することにより、時間歪曲条件による練習効果を検証することができる。また、4つの時間歪曲条件の順番を被験者間でカウンターバランスを行い、結果を比較することにより、異なる速度倍率での練習効果や、動的な倍率変更による練習効果の検証が可能となる。

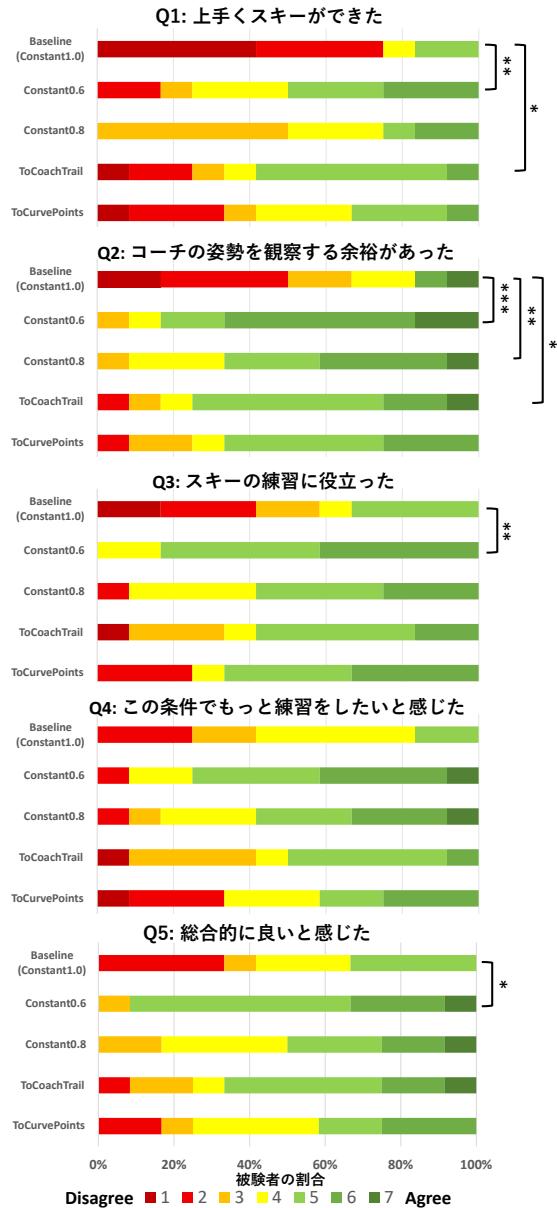


図 5. ユーザの主観評価による5つの質問への回答、グラフは7段階評価の分布を表す(1:「強く否定する」, 7:「強く同意する」)。右の括弧は各条件間の有意差を表す: * ($p < 0.05$), ** $p < 0.01$, *** $(p < 0.001)$

他の条件で行った練習の影響を軽減するため、手順3.と6.においては質問への回答が早く終わった場合でも、5分以上の間隔を空けてから次の条件での練習に移るようにした。また、被験者は実験中の任意のタイミングで休憩及び水分補給を行うことができた。

4.3 被験者

主に研究室内の学生12人(年齢21-32歳、平均23.6、標準偏差3.15、女性3名、男性9名)からデー

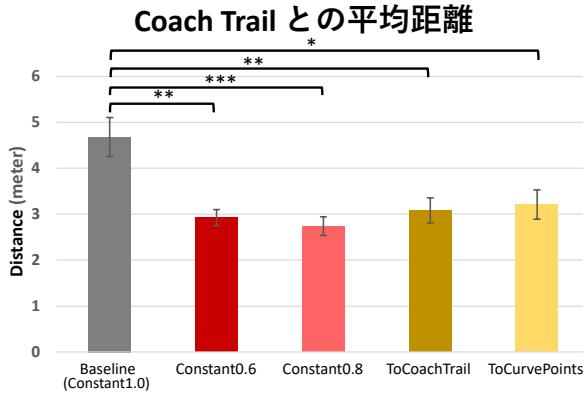


図 6. 各条件での練習後に計測したベースライン条件におけるユーザと Coach Trail とのフレーム毎の平均距離。上の括弧は各条件間の有意差を表す: * ($p < 0.05$), ** ($p < 0.01$), *** ($p < 0.001$)

タを収集した。被験者のうち、6人は実際のスキーをした経験が一度もなく、4人は2年未満の経験があり、2人は5年以上の経験があった。また、VR の使用頻度についても尋ねたところ、3人は年に1回程度、7人は月に1回程度、2人は週に1回以上の頻度で VR の使用経験があった。

4.4 実験結果

4.4.1 Coach Trail との距離

最初に計測したベースライン条件および4つの条件で練習した直後に計測したベースライン条件における被験者と Coach Trail との距離の、フレーム毎の平均を図 6 に示す。one-way ANOVA を実行したところ、条件間に有意差があることがわかった ($F_{4,55} = 6.4275, p < .001$)。そこで、どの条件間で有意差があるかを示すために Tukey HSD[9] を行った。結果は図 6 に示す。どの時間歪曲条件においても、ベースライン条件と比べて有意に高いパフォーマンスが得られ、Constant 0.8 条件では最も強い有意差が得られた (Constant 0.6: $p < .01$, Constant 0.8: $p < .001$, ToCoachTrail: $p < .01$, ToCurvePoints: $p < .05$)。一方で、時間歪曲条件の間では有意差が見られなかった。

4.4.2 被験者の主観評価によるアンケート

被験者の主観評価によるアンケートの回答結果を図 5 に示す。one-way ANOVA を実行したところ、Q1, Q2, Q3, Q5において有意差があることがわかった (Q1: $F_{4,55} = 3.6396, p < .05$, Q2: $F_{4,55} = 5.4573, p < .001$, Q3: $F_{4,55} = 3.7942, p < .01$, Q4: $F_{4,55} = 2.4962, p = .05318$, Q5: $F_{4,55} = 2.7626, p < .05$)。そこで、これらの質問のどの条件間で有意差があるかを示すために Tukey HSD を行った。

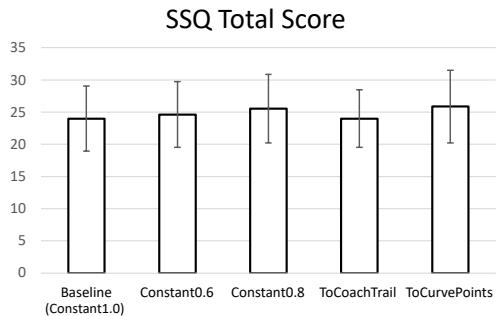


図 7. 各条件後の Simulator Sickness Questionnaire の Total Score の値

結果は図 5 に示す。Constant 0.6 条件については Q4 以外の質問においてベースライン条件と比べて優位に高い評価が得られており (Q1: $p < .01$, Q2: $p < .001$, Q3: $p < .01$, Q5: $p < .05$)、強く否定する回答 (評価 1 と 2) の割合も他の条件と比べて低い。動的な条件については DistanceToCoach-Trail 条件ではパフォーマンスに関する Q1 と Q2 でベースライン条件と比べて有意に高い評価が得られている (Q1: $p < .05$, Q2: $p < .05$)。一方、DistanceToCurvePoints 条件では有意な差は得られなかった。

4.4.3 シミュレータ酔い

シミュレータ酔いの程度を表す指標となる SSQ の TotalScore の結果を図 7 に示す。one-way ANOVA を実行したところ、各条件間には有意差が得られなかった ($F_{4,55} = 0.0263, p = .9986$)。

4.4.4 自由記述によるコメント

静的な条件について、Constant 0.6 は「コーチの姿勢の観察がしやすい」(P4,P7) というコメントがあった一方で、「簡単すぎて練習にならない」(P1) というコメントがあった。Constant 0.8 は「ちょうどよいスピードだと感じた」(P1,P3,P7) というコメントが多く得られた。動的な条件では、「滑走中に速度が変化することに対応がしにくい」(P1,P7) や「速度変化によって酔いを感じた」(P2,P6) といったコメントがあった。また、条件に関わらずシミュレータ酔いについて言及したコメントが多く得られた (P2,P5,P8,P10,P12)。

5 議論

5.1 仮説の検証

5.1.1 H1: 時間歪曲による練習効果

図 6 に示すように、最初に計測したベースライン条件に対し、4つの条件下にて練習を行った直後のベースライン条件では被験者と Coach Trail との

距離が有意に小さくなつたことから、どの条件下でも、よりコーチと近い滑走が行えていたことがわかる。図5に示すアンケートについて、Constant0.6では練習へのモチベーションに関するQ4以外の質問において有意差が得られている。特にQ5の総合評価については評価5~7の回答が91.7%を占めたことから概ね好評であったことがわかる。これらのことから時間の経過速度を遅くすることが被験者のスキーの上達に繋がつたと考えられるため、仮説H1は支持される。

5.1.2 H2:より低い速度倍率による練習効果

図5に示すアンケートについて、Q4以外の質問において、Constant0.6はConstant0.8と比べて、ベースライン条件とより強い有意差が得られている。一方で、この2つの条件間では有意差が得られておらず、図6のCoach Trailとの距離についてはConstant0.8の方がベースライン条件に対し、より強い有意差を得られていることから、時間の経過速度を遅くするほど練習効果が高まるかどうかの検証は不十分であり、仮説H2は支持されない。

5.1.3 H3: 動的な倍率変更による練習効果

図6のCoach Trailとの距離について、動的な条件2つはベースライン条件に対して有意差が得られてはいるものの、その有意差は静的な条件2つに対して弱いものとなっている。また図5に示すアンケートについては、いずれの質問でも強く否定する回答(評価1と2)が静的な条件よりも多い傾向にあり、被験者に不評であったことがわかる。よって動的に時間の経過速度を変更することが練習効果を高めるとは言えないため、仮説H3は棄却される。

5.2 考察

実験の結果をもとに、4つの時間歪曲条件の有用性について考察を行う。我々の予想に反し、動的な条件の結果が静的な条件に対して良くなかった。滑走している間に時間の経過速度の変更を行うことで、被験者は変化後の速度に対応する必要があり、負担が生じてしまった。これにより練習やコーチの姿勢の観察に集中できなかつた可能性が考えられる。

動的な条件では滑走中に速度の変化が生じるため、比較的強いシミュレータ酔いが発生することが予想されていた。しかし、図7に示すように、SSQの結果としては有意差が得られなかつた。一方で、どの条件においてもSSQのスコアの平均が20を超えており、自由記述のコメントにおいてもシミュレータ酔いについて言及したものが多く得られている。そのため、動的な条件が酔いを増加させなかつたのではなく、条件に関わらず一定程度の酔いが引き起こされていたため、有意差が生じなかつたことが可能性として考えられる。

被験者の主觀評価によるアンケートではConstant0.6が最も高い評価であったのに対し、図6によれば、Constant0.8の方が被験者がよりコーチと近い滑走を行えていたのは興味深い結果である。速度倍率が低いほどコーチの動作の模倣は行いやすくなるものの、コーチの動作自体が実際のものと大きく変わってしまうため、練習としての効果は低くなつてしまつた、ということが原因として考えられる。

5.3 今後の課題

本論文で検証に用いたスキーシミュレーションシステムは、実際のスキーを忠実には再現ができるおらず、これがユーザのパフォーマンスやシミュレータ酔いなどに影響を与えた可能性が考えられるため、システムの全体的な改善が必要である。仮想コース内のカーブを緩やかにして直線部分を増やすなど、より現実に近いコースに変更する。スキー板を傾けた場合のユーザの曲がりやすさが、より現実的になるようにパラメータを調整する。などの改善点が挙げられる。

今回の実験では動的な条件がユーザに不評であったが、速度を変更する方法を改善することによって、静的な条件よりも良い結果が得られる可能性がある。滑走中に速度変更を行うのではなく、一度滑り終わつた後にその滑走のパフォーマンスを評価し、評価に応じて次回の速度倍率を決定するような動的条件が考えられ、今後実装していくことを考えている。

今回の実装では、ユーザとコーチとの間の距離が広がりすぎた場合に、コーチの姿が見えにくく、模倣が行えなくなる問題があった。コース上だけでなく、HMDの画面上の固定された位置にコーチの姿勢を表示することにより、改善が可能であると考えられる。

6 まとめ

本論文では、スキーのシミュレータと、VR機器を組み合わせたシミュレーションシステムを作成し、時間の経過速度を変更する複数の時間歪曲機能を実装した。評価実験の結果より、他のスポーツと同様にスキーにおいても時間歪曲が上達の手助けになることが確認できた。一方で、ベースライン条件との比較以外では有意差を得ることは出来なかつたが、ユーザに好まれやすい時間歪曲条件の傾向をつかむことができた。

謝辞

本研究はJST CREST JPMJCR17A3およびJSTムーンショット型研究開発事業JPMJMS2012の支援を受けている。

参考文献

- [1] J. Adolf, P. Kán, B. Outram, H. Kaufmann, J. Doležal, and L. Lhotská. Juggling in VR: Advantages of Immersive Virtual Reality in Juggling Learning. In *25th ACM Symposium on Virtual Reality Software and Technology, VRST '19*, New York, NY, USA, 2019. Association for Computing Machinery.
- [2] V. Aleshin, S. Klimenko, M. Manuilov, and L. Melnikov. Alpine skiing and snowboard training system using induced virtual environment. *Science and Skiing IV*, pp. 137–144, 01 2008.
- [3] A. Bansal, S. Weech, and M. Barnett-Cowan. Movement-Contingent Time Flow in Virtual Reality Causes Temporal Recalibration. *Scientific Reports*, 9:4378, 03 2019.
- [4] A. L. Edwards. Balanced Latin-Square Designs in Psychological Research. *The American Journal of Psychology*, 64(4):598–603, 1951.
- [5] N. Hamanishi and J. Rekimoto. TTT: Time Synchronization Method by Time Distortion for VR Training Including Rapidly Moving Objects. In *25th ACM Symposium on Virtual Reality Software and Technology, VRST '19*, New York, NY, USA, 2019. Association for Computing Machinery.
- [6] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *The International Journal of Aviation Psychology*, 3(3):203–220, 1993.
- [7] J. Ko, S.-w. Jang, H. Lee, H.-K. Yun, and Y. S. Kim. Effects of Virtual Reality and Non-Virtual Reality Exercises on the Exercise Capacity and Concentration of Users in a Ski Exergame: Comparative Study. *JMIR Serious Games*, 8:e16693, 10 2020.
- [8] A. H. Kobeissi, H. Chible, R. Berta, A. De Gloria, and F. Bellotti. Design and implementation of an IoT system for enhancing proprioception training. In *2017 29th International Conference on Microelectronics (ICM)*, pp. 1–4, 2017.
- [9] J. Tukey. Comparing individual means in the analysis of variance. *Biometrics*, 5:2:99–114, 1949.
- [10] E. Wu, T. Nozawa, F. Perteneder, and H. Koike. VR Alpine Ski Training Augmentation using Visual Cues of Leading Skier. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 3836–3845, 2020.
- [11] E. Wu, F. Perteneder, H. Koike, and T. Nozawa. How to VizSki: Visualizing Captured Skier Motion in a VR Ski Training Simulator. In *The 17th International Conference on Virtual-Reality Continuum and Its Applications in Industry, VRCAI '19*, New York, NY, USA, 2019. Association for Computing Machinery.
- [12] E. Wu, M. Piekenbrock, T. Nakamura, and H. Koike. SPinPong - Virtual Reality Table Tennis Skill Acquisition using Visual, Haptic and Temporal Cues. *IEEE Transactions on Visualization and Computer Graphics*, 27(5):2566–2576, 2021.
- [13] 川崎 仁史, 脇坂 崇平, 笠原 俊一, 齐藤 寛人, 原口 純也, 登嶋 健太, 稲見 昌彦. けん玉できた! VR: 5 分間程度の VR トレーニングによってけん玉の技の習得を支援するシステム. エンタテインメントコンピューティングシンポジウム 2020 論文集, pp. 26–32, 2020.
- [14] 野澤 崇行, E. Wu, F. Perteneder, 小池 英樹. 熟練者の動作の可視化による VR スキートレーニングシステム. WISS2019 論文集, pp. 19–24, 2019.

未来ビジョン

本論文では、場所や季節への制限が多いスキーを題材として、VR 機器を用いた仮想トレーニングシステムの開発を行った。with コロナ, after コロナの時代では、現地で直接コーチから指導を受けるような他の様々なスポーツのトレーニングの代用として、このような仮想トレーニングの需要が増していくと考えられ、本論文の結果を応用できる可能性が高い。また、本論文では、時間歪曲というシミュレーションならではの非現実的な機能を実装することによって、VR を単なる実際の環境でのトレーニングの代用とする以上の効果が得られる可能性についての模索も行った。

本論文では簡易的なスキーシミュレータマシンを用いたため、足の角度のみを基準として仮装コース内の移動を決定していたが、今後はより高度なシミュレータを用いることで、スキー板を踏み込む強さ等の計測が可能であり、それに応じた仮想コース内の動きを制御すればより現実のスキーに近づけることができると考えられる。

今回の実装では、時間の経過速度を遅くして練習を行いやすくする、という初心者向けの実装をしたが、逆に、経過速度を実際よりも早くすることによって、トレーニングの負荷をより高め、通常の速度におけるパフォーマンスを向上させる、という中級者や上級者向けの実装も可能である。

ドライブレコーダのデータから旅行の思い出動画を自動生成する web サービスの実装

尾頭 花奈 * 石黒 祥生 *† 椎尾 一郎 ‡ 武田 一哉 *†

概要. 本研究では、ドライブ旅行の思い出を楽しく振り返ることを目的に、ドライブレコーダのデータを活用した思い出動画自動生成システムを提案する。最近のドライブレコーダは、車内外の様子を高解像度動画として記録するため、紀行ムービーの素材として最適である。ドライブレコーダで撮影された映像データは、SD メモリカードから簡単に入手できるが、膨大なデータを編集するには多くの時間と労力が必要なため、ほとんど利用されていない。そこで、思い出動画を自動生成するシステムを開発した。提案システムは名所、会話の盛り上がりなど、旅行において重要なシーンを判定して動画を要約し、文字テロップ・地図・合成音声の追加を行う。本稿では、自動編集システムの設計に加え、広く一般に使用してもらうために開発した web サービスの実装と評価について述べる。本研究は、今まであまり注目されてこなかった自動車乗車後の UX 向上に貢献する。

1 はじめに

スマートフォンやウェアラブルカメラの登場により、個人の体験を映像として簡単に記録できるようになった [9, 14]。また、SNS の普及により、体験を他者に共有し、コミュニケーションをとる機会が増えている。自動車にも多数のカメラやセンサ類がついており、自動車乗車中の体験が細かく記録されている。しかし、これらのセンサ情報は自律走行や交通安全以外の目的にはあまり利用されていない。

ここ数年、車両から得られる映像データやセンサ情報をエンターテイメントシステムに利用しようという試みがなされている [12, 13, 21]。しかし、これらは乗車中の UX 向上を目的としており、乗車後の UX 向上に焦点を当てた研究は極めて少ない。そこで本研究では、旅行の振り返りに焦点を当て、乗車後の UX 向上を目的とした。

本研究では、近年急速に普及しているドライブレコーダに着目した [16]。ドライブレコーダの主な用途は交通事故や煽り運転車との遭遇のような非常事態への対策である。そのため、ドライブレコーダに記録されたデータは通常はほとんど活用されず、記録媒体の容量を超えた内容から廃棄される。最近のドライブレコーダは GPS 機能を搭載したモデルや複数台の高品位カメラを搭載したモデルも多く [10]、ホームビデオや SNS で共有する動画の素材として最適である。

しかしながら、長時間の走行動画を単に再生するだけでは冗長なシーンが多いため、楽しい動画にするには適切な編集が必要である。ドライブレコーダ

から得られる大量の動画ファイルから印象的なイベントが記録された箇所を選択し、結合し、エフェクトを追加するには多くの時間と労力が必要であるため、これら全てを毎回手作業で行うのは困難である。

そこで本研究では、これらの課題を解決するため、ドライブレコーダのデータからドライブ時に特化したイベントを検出し、車旅行の思い出動画を自動生成するシステムを開発した。今回作成したシステムは、自動車から得られる映像データの新しい活用アイデアをサポートし、これまであまり検討されてこなかった自動車乗車後の UX 向上に貢献する。

本システムのプロトタイプについては、第 69 回 UBI 研究発表会 [20] と AutomotiveUI '21[1] で発表した。本稿ではこれを発展させ、一般公開用に開発した web サービス全体の設計について述べる。

2 関連研究

2.1 ドライブレコーダの活用

ドライブレコーダのデータを交通安全に活用する試みは多数存在する。例えば、ドライブレコーダを交通事故発生予測に活用する研究や [3, 19]、路上駐車のリアルタイム検知に活用する研究 [15]、注意散漫なドライバーを自動検出する研究 [4] などがされている。本研究では、これらの取り組みとは違い、ドライブレコーダのデータをエンターテインメントに活用することを目的とする。

2.2 動画の自動要約技術

動画の自動要約に関する研究は多数してきた。例えば特異値分解 (SVD) とクラスタリングによる映像要約技術や [7]、シーンの変化をグラフのモデル化により検出する動画要約技術 [17]、ビデオフレー

* 名古屋大学

† 株式会社ティアフォー

‡ お茶の水女子大学



図 1. 構築した web システムの概要図。

ムからの色特徴抽出と k-means クラスタリングアルゴリズムに基づいた動画要約技術 [5] などがある。これらの研究では主にシーンの変化に着目し、要約を行っている。しかし、車内映像に関しては、自動車に乗車中は人の動きは少ないため、視覚的なシーンの変化はほとんどない。また、旅行においては、全てのシーンを網羅的に切り取る必要はなく、例えば、「目的地や観光名所付近の映像は必要だが高速道路や住宅街の映像は不要である」など、走行位置や周囲の状況によってシーンの重要度が大きく異なる。そのため、これらの手法はドライブ旅行の映像の要約には適さない。

2.3 身近な製品を利用した体験要約

身近な製品を使用することで、生活シーンの切り取りや要約、記憶支援を実現する研究も多い。Comic-Diary は、携帯情報端末 (PDA) やキオスク端末を利用し、個人のプロフィールや行動履歴、他のユーザーとの交流記録などから、学術会議参加や博物館見学の日記を漫画形式で自動生成する [18]。Video-Recording Your Life では、ウェアラブルカメラである GoPro と、スマートフォンの加速度センサを利用し、日常生活を記録した映像から利用者が興味を引くと思われるシーンを自動抽出する [2]。マイクロソフトのプロジェクト「MyLifeBits」で使用されたウェアラブルカメラ SenseCam には多数のセンサが組み込まれており、身に附いているだけで光量の著しい変化やカメラの前の体温などを自動で検出して写真を撮影する [11, 6]。身近な製品を利用するることは、多くの人に使ってもらう一つのきっかけになると考えられる。本研究では、近年急速に普及しており、手軽に購入可能なドライブレコーダーの活用に着目し、要約の対象を自動車乗車中とした。

また、世に広く普及している旅行の自動記録システムとして、Google Maps のタイムライン機能がある。このサービスは、移動経路や滞在場所を地図上にプロットし、滞在時間などの情報を合わせた旅行の記録を自動でまとめる。また、Google photo に保存されている写真を合わせて表示することもできる [8]。本研究でも、移動経路、滞在場所、時刻な

どの情報や、ユーザが旅行中に撮影した写真を自動で動画に組み込むように設計した。

3 web システム概要

提案する自動編集システムは、ドライブレコーダーから得られた動画ファイル、GPS のログファイル、ユーザが旅先で撮影した写真をもとにハイライトシーンを抽出し、サウンドエフェクトやトランジション効果を加えながら旅の要約動画を自動生成する。また、自動編集システムを広く使ってもらうための web システムを構築した。本 web システムの概要を図 1 に示す。提案システムは <https://bitokana.com> から誰でも利用可能である(図 2)。対応ブラウザは Safari, Chrome, Firefox であり、PC から閲覧することを前提としている。ユーザは本サイトに必要事項を入力し、ドライブレコーダーのデータをアップロードするだけで思い出動画を手に入れることができる。web サイトの構築のためにさくら VPS を使用し、SSL 証明書には Let's Encrypt を用いた。VPS の CPU は仮想 4Core、メモリは 4GB、SSD は 400GB、OS は CentOS を選択した。本システムは Python の Web フレームワークである FastAPI を利用している。

3.1 システム使用手順

アップロード機能は web サイトのトップページに配置した。ユーザは本システムを利用するにあたって以下の操作をする必要がある。

3.1.1 対応機種

ドライブレコーダーは機種によってカメラの撮影範囲、保存されるデータの拡張子、1 ファイルあたりの撮影時間などが異なるため、使用する機種によって自動編集システムの処理が異なる。本稿では、水平方向に 360° 撮影可能であり、GPS 測位機能が搭載されている Jupiteru 社 Q20-P を使用した際の処理について述べる。このドライブレコーダーは動画データと位置情報を、それぞれ MP4 ファイルおよび NMEA (National Marine Electronics Association) ファ

ドライブレコーダのデータから旅行の思い出動画を自動生成する web サービスの実装



図 2. 作成した web サイトを通してデータをアップロードする様子。データ送信前の画面 (左), アップロード中の画面 (中央), データ送信後の画面 (右)。

イルとして SD カードに記録する。いずれも約 1 分間のデータを 1 ファイルとして (以下、1 分録画ファイル), 複数のファイルが生成される。NMEA ファイルには, GPS により検出された位置情報が 1 秒毎に書き込まれる。MP4 ファイルの映像の記録形式は魚眼と 2 分割 (上半分が車外, 下半分が車内の様子) が選べるが, 本システムでは 2 分割の形式で記録していることを前提とした。なお, 現段階では Yupiteru 社 Q20, Q20-P 以外の機種では検証していないため未対応となっているが, 今後は他の機種にも対応する予定である。

3.1.2 事前準備

ユーザは以下の手順によりアップロード用のファイルを事前準備する。

- (1) “movie” という名前の空のフォルダを作成する
- (2) microSD カードに入っている全てのフォルダを movie フォルダの中にコピーする
- (3) movie フォルダを圧縮する

また, 旅先で撮影した写真 (但し Exif 情報があるものに限る) を要約動画に組み込みたい場合は上記の手順 3 の前に追加で以下の処理を行う必要がある。

- (a) “photo” という名前の空のフォルダを作成する
- (b) photo フォルダの中に要約動画に組み込みたい写真を入れる
- (c) photo フォルダを movie フォルダの中に移動する

3.1.3 データの送信

ユーザは上記の手順で作成した zip ファイルに加え, 以下の 3 つの情報をシステムに送る必要がある。一つ目は, 完成した動画をユーザに送信するためのメールアドレスである。本システムが生成する要約動画はデータサイズが大きいため, メールに動画をそのまま添付することはできない。そこで, 本システムは動画の共有に GoogleDrive のリンクを利用

する。そのため, 本システムの利用には Google アカウントが必要であり, G メールアドレスの入力を必須とした。二つ目はドライブレコーダの機種である。前述のように, ドライブレコーダは機種によって映像の記録形式や保存データのディレクトリ構造が異なる。そこで, ドライブレコーダの機種をユーザに選択してもらうように設計した。三つ目は旅行期間である。通常ドライブレコーダから取り出した microSD カードの中には要約したい旅行期間以外のデータが含まれている。そこで, アップロードしたデータ全てを要約するのか, それとも一部を要約動画の素材として使用するのかを選択できるようにし, 要約動画に使用したい動画データの期間を設定できるようにした。

以上の 3 項目を入力し送信ボタンを押すと, データがサーバに送信される。データの送信が完了するとユーザに受け付け完了のメールが自動送信される。

3.2 動画処理

サーバにアップロードされた zip ファイルは, 以下の手順で要約動画に変換される。まず, アップロードされた zip ファイルを解凍する。そして, システムがデータの管理をしやすいように, 解凍したファイルのファイル名をユーザのメールアドレスに変更する。ここで, アップロードされたデータのうち一部の期間のデータのみを要約動画の素材として利用する場合は, NMEA ファイルから取得した日時の情報を用いて, ユーザが指定した時間外のデータを削除する。こうして出来上がったファイルを自動編集システムに読み込むことで, 要約動画が生成される。

自動編集システムは, Python3 によって記述され, コマンドラインツールである FFmpeg を利用して動画処理を行なう。自動編集システムの詳細な設計については 4 章で説明する。



図 3. 自動編集システムによる明るさの調整. 調整前(左)と調整後(右).

3.3 自動返信機能

作成された要約動画は以下の手順で Google ドライブのリンクとしてユーザに共有される。まず、自動編集システムが作成した動画を Google Drive API を用いて Google ドライブにアップロードし、動画の共有リンクを取得する。この時、ユーザがリンク先の動画を閲覧できるように、ユーザが入力した G メールアドレスを利用してユーザに動画の共有権を付与する。そして、取得した共有リンクをメールの本文に貼り付け、Gmail API の自動送信機能により、ユーザのメールアドレス宛に自動送信する。

以上の処理により、ユーザは要約動画を取得する。また、全ての処理を終えたら、アップロードされた全データを自動で消去し、サーバ上に個人情報を残さないようにした。

4 自動編集システム

本システムは以下の 7 つの処理を行う。

4.1 動画のサイズの調整

Q20-P ドライブレコーダから得られた動画データは、横縦が 2048×1536 画素、28fps である。1 フレームは上下に 2 分割され、それぞれに車内外の映像が記録されている。車外の映像を使用するときは上半分中央部の 1360×765 画素を、車内の映像を使用するときは下半分中央部の 1360×765 画素を切り出し、 1280×720 にリサイズして使用する。なお、本システムが最終的に出力するムービーの画素数も 1280×720 画素である。

4.2 動画の明るさの調整

ドライブレコーダの主な目的は車外の様子の撮影であるため、360° 全天球カメラで撮影された映像は窓から見える景色の光量に合わせた露出となっている。そのため、昼間でも車内の映像は暗く、夜間に至っては暗くてほとんど何も見えない状態であった。そこで、車内の映像に対しガンマ補正、彩度調整、コントラスト調整を行った(図 3)。

4.3 ハイライトカットの自動判定と切り取り

本システムは、ドライブ中に注目すべき事象が発生した際に撮影されるカット(以下、ハイライトカッ

ト)を自動抽出する。現在実装している事象は以下に示す 6 種類である。

(1) **出発時**: 車両の ACC(アクセサリー)電源が ON になったタイミングを出発時と判定し、そのときの車外の映像を 7 秒、車内の映像を 10 秒切り取った。車外の映像には NMEA ファイルから取得した時刻の情報を用いて、文字テロップと合成音声の挿入を行った。なお、ドライブレコーダは ACC 電源の供給に連動して自動的に記録を開始/終了するため、運転中だけが記録対象である。本研究ではこれを利用し、録画時間が 10 分以上開いた場合、休憩地点もしくは目的地に到着したと判断する。

(2) **到着時**: ACC 電源が OFF になったタイミングを到着時とし、車内の映像を 10 秒、車外の映像を 7 秒切り取った。車外の映像には、NMEA ファイルから取得した時刻と座標の情報、場所情報 API を用いて、到着時の時刻と住所を取得し、文字テロップと合成音声の挿入を行った(図 4)。

(3) **県境を通過したとき**: NMEA ファイルから取得した座標情報と場所情報 API を用いて都道府県名を取得した。そして、取得した都道府県名の変化から県境を判定し、県を跨いた瞬間の車外の映像を 3 秒切り取った。また、都道府県名を文字テロップの挿入と合成音声の挿入により示した(図 4)。

(4) **名所付近を通過したとき**: NMEA ファイルから取得した座標情報と場所情報 API を用いて周辺の大規模施設や地名などの地域情報を取得した。そして、取得した情報から名所付近を通過したことを判定し、そのときの車外の映像を 3 秒切り取った。また、地域情報を文字テロップの挿入と合成音声の挿入により示した。このとき、情報を表示するのは場所情報 API の Score(取得した地域情報の確信度)が 99.9 以上の場合に限定し、名所のすぐ近くを通過するときだけ表示を行った(図 4)。

(5) **高速道路の出入り口**: NMEA ファイルから速度を取得し、1 分間の平均速度を計算した。本研究では、1 分間の平均速度が 60km/h 以上なら高速道路と判断し、 60km/h 未満なら一般道であると判断する。また、ETC (Electronic Toll Collection System) ゲートは 20km/h 以下の徐行が求められていることから、高速道路と一般道の切り替わる箇所のうち、 $10\text{km/h} \sim 20\text{km/h}$ の箇所を高速道路の出入り口であるとし、その部分を 4 秒切り取った。

(6) **会話が盛り上がっているとき**: MP4 ファイルから取得した音声データにバンドパスフィルターを適用し、日本人の音声周波数帯域とされる $300 \sim 3400\text{Hz}$ の周波数帯域を抽出した。そして、抽出した音声データの振幅が一定の閾値を超えた箇所を会話が盛り上がっているシーンと判定し、そのときの車内の映像を 6 秒切り取った。また、Python の SpeechRecognition ライブラリを利用して会話に字幕を表示した(図 4)。



図 4. 自動編集システムが切り出すハイライトカット. 左から順に到着時, 県境を通過したとき, 名所付近を通過したとき, 会話が盛り上がっているとき.

4.4 旅先で撮影した写真の追加

スマートフォンやデジタルカメラで撮影された写真には、撮影日時や位置情報を含む Exif (Exchangeable image file format) 情報が記録されている。そこで、旅先で撮影した写真に Exif 情報が埋め込まれていた場合、そこから撮影日時を取得し、作成動画の適切な位置に表示させる処理を行なった。画像の表示時間は 1.2 秒に設定し、縦長の写真の場合は左右に余白をつけることで 16:9 の比率に合わせた。

4.5 地図の生成

車旅行では、移動経路や滞在場所の地理情報も重要である。そこで、ハイライトカットの対象にならなかった 1 分録画ファイルに対しては、地図を伴った表示を行なった。このために、Python の Folium ライブラリを用いて取得座標を地図上に可視化する処理を行った。出発してから目的地に到着するまでに通過した道を赤色の線で示し、目的地に到着したときには赤色のマーカーをプロットした。そして、再出発時には、これまでに通過した道を赤色の線からピンク色の線に変更し、再出発してから次の目的地に到着するまでに通過した道を新たに赤色の線で示した。この地図を車外映像と組み合わせるために、2 種類の表現方法を採用した。まず、それぞれの 1 分録画ファイルの 1 フレーム目を取り出し、一つ前の 1 分録画ファイル第 1 フレームの対応画素の RGB 値の差分の絶対値の総和を計算する。この総和を、設定した閾値と比較して以下の処理を行なう。

差分が閾値以上だった場合: 周囲の景色の変化が大きいと判断する。地図上に、移動経路、車外の映像から切り取った静止画、車のイラストを載せた静止画を、1 分録画に対して 1 枚作成しこれを 0.3 秒程度表示する。車外映像領域は 1 秒のコマ落としにした(図 5. 以下、地図主体表示)。

差分が閾値未満だった場合: 景色の変化が少ないと判断する。1 分録画から切り出した車外映像を 100 倍速にして、移動経路を表示した地図をアルファブレンディングする(図 5. 以下、100 倍速動画)。

4.6 処理の優先順位の調整

切り取り箇所重複の回避: ここまで手順により作成された動画ファイルを時系列順に繋げることで、

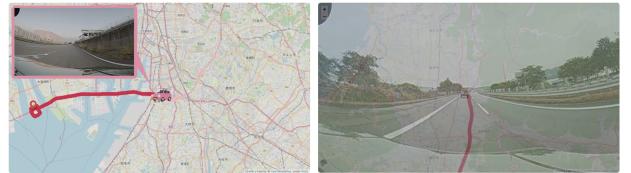


図 5. ハイライトシーンがなかった 1 分録画ファイルに対する 2 種類の処理. 景色の変化が大きいときの処理(左)と景色の変化が小さいときの処理(右).

思い出動画を作成することができる。しかし、このままでは同じシーンが、複数のハイライトカットの対象となり、重複して切り取られてしまう可能性がある。そこで、ハイライトカットに優先順位を設定し、同一箇所が複数回切り取られることがないように調整した。

地図表示手法の調整: 作成したハイライトカットや地図表示をそのまま繋げると、画面が過剰に切り替わり、見辛い動画になってしまう。そこで、1 分録画ファイルに対する処理を指定する配列に対して以下の処理を行い、変化を抑えた。

1. 地図主体表示が 100 倍速動画に挟まれている場合、地図主体表示の代わりに 100 倍速動画を表示する。
2. 100 倍速動画が 3 連続未満なら、それら全てを地図主体表示に変更する。
3. 地図主体表示が続かない場合、その地図情報は使用しない。

4.7 動画の連結とエフェクトの追加

以上の処理によって生成された素材に対して、以下の視覚音響効果とファイルの結合処理を行うことで思い出動画が完成する。

吹き出しの拡大・縮小: 地図主体表示とその他の動画ファイルの切り替えをスムーズに行うため、吹き出しのエフェクトを使用した(図 6)。

フェードイン・フェードアウト: 出発時にはフェードインの処理を、到着時にはフェードアウトの処理を加えることで一旦画面を暗くし、目的地に到着したことを示した。

表 1. 使用したデータの情報と自動編集処理にかかる時間、手動編集の作業時間。

旅行期間	乗車時間	ファイルサイズ	ファイル圧縮時間	アップロード時間	自動編集時間	手動編集時間(P1)	手動編集時間(P2)
5時間18分	98分	12.42GB	4分	35分	36分	4時間47分	3時間48分
8時間24分	108分	15.61GB	5分	45分	36分	4時間37分	3時間08分
8時間54分	159分	20.1GB	6分	58分	50分	3時間50分	3時間45分



図 6. 吹き出しの拡大エフェクトによるトランジション。



図 7. 放射状に広がるワイプエフェクト。

放射状に広がるワイプエフェクト: 通常速度の動画と100倍速動画の切り替えをスムーズに行うために、放射状のワイプエフェクトを加えた(図7)。
ジングル: 県境と名所付近を通過したとき、吹き出しのエフェクトによるトランジション、放射状に広がるワイプエフェクトにはそれぞれ効果音をつけた。100倍速動画にはアップテンポのジングル、地図主体表示にはミディアムテンポのジングルをつけた。
動画ファイルの結合: 以上の処理で作成された動画はすべて、切り取り箇所の該当時間に対応した名前がつけられ、完成動画用フォルダに保存される。たとえば、80ファイル目の30秒地点から4秒間の映像を切り取った場合、“00000080.30.mp4”という名前がつけられる。そこで、全ての処理完了後に、完成動画用フォルダにある処理済みの動画を名前順にソートし、全てのファイルを連結して一つの動画を生成した。

5 処理時間の評価

著者の1人が母親と2人で実施した3回のドライブ旅行のデータを本システムに適用した。また、自動編集と手動編集の違いを検証するため、双方とも動画編集初心者である、母親(P1)と兄弟(P2)に手動で動画編集をしてもらった。

P1: 50代女性、旅行に参加

P2: 20代男性、旅行に不参加

編集にはiMovieを利用した。本システムに適用した際にかかった時間と2人の被験者が手動編集にかかった時間を表1にまとめる。ドライブレコーダーの

データは、1時間半程度のドライブで約10GBと大きいため、アップロードと自動編集システムの処理に時間がかかる。加えて、ユーザはデータのアップロード中はブラウザを閉じることはできない。しかし、今回の実験では、送信ボタンを押してから要約動画の受信までにかかった時間は、実際の乗車時間よりも短く、手動編集する際にかかる時間に比べて大幅に短かった。また、送信ボタン押下後はユーザの作業は一切不要であるため、実用的であると言える。

6 議論

多くのドライブレコーダーは車外のみを撮影し、車内の様子は記録できないため、今回は旅行ルートの要約と景色の切り取りに重点を置いたシステム設計にした。本システムは、手動編集では難しい周辺地域情報の取得やルートの可視化が可能である。一方、会話の内容に基づく動画の切り取りは難しく、現在は実装されていない。会話の内容による自動要約を実現するには、システムが会話の内容を適切に理解する必要があるのに加え、人によって面白いと思う会話や思い出に残したいシーンが大きく異なるといった課題がある。今後、会話の内容に基づくシーンの抽出を行う際には、例えば、システムが表情認識や音声感情認識を用いてハイライトカットの候補を絞り、その後、ユーザが動画に組み込みたいシーンを選択するなどの半自動化処理を検討したい。なお、自動編集システムと手動編集の詳細な比較評価、及び被験者アンケートについては[1]を参照されたい。

7 おわりに

本研究では、ドライブレコーダーのデータを活用し、旅のハイライトを自動検出して思い出動画を自動生成するシステムの提案と実装を行った。ユーザはwebサイトにデータをアップロードするだけで、ドライブ旅行の思い出を紀行ムービーとしてまとめることができる。今後はより多くのユーザが利用できるよう、対応機種を増やす予定である。

一方、自動運転の技術が成熟し、急速に普及しつつある。自動運転車に搭載されている多数の高画質カメラ、センサ類を利用して紀行ムービーを作成することで、現在の手法に比べてはるかに高いユーザエクスペリエンスを提供できると考えている。

参考文献

- [1] K. Bito, I. Sii, Y. Ishiguro, K. Takeda. Automatic Generation of Road Trip Summary Video for Reminiscence and Entertainment using Dashcam Video. In *13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, AutomotiveUI '21, New York, NY, USA, 2021. ACM.
- [2] D. Buschek, M. Spitzer, and F. Alt. Video-Recording Your Life: User Perception and Experiences. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '15, p. 2223–2228, New York, NY, USA, 2015. ACM.
- [3] F.-H. Chan, Y.-T. Chen, Y. Xiang, and M. Sun. Anticipating Accidents in Dashcam Videos. In S.-H. Lai, V. Lepetit, K. Nishino, and Y. Sato eds., *Computer Vision – ACCV 2016*, ACCV 2016, pp. 10114:136–153, Cham, 2017. Springer.
- [4] J. Cronje and A. P. Engelbrecht. Training Convolutional Neural Networks with Class Based Data Augmentation for Detecting Distracted Drivers. In *Proceedings of the 9th International Conference on Computer and Automation Engineering*, ICCAE '17, p. 126–130, New York, NY, USA, 2017. ACM.
- [5] S. E. F. de Avila, A. P. B. Lopes, A. da Luz, and A. de Albuquerque Araújo. VSUMM: A mechanism designed to produce static video summaries and a novel evaluation method. *Pattern Recognition Letters*, pp. 32(1):56–68, 2011.
- [6] J. Gemmell, G. Bell, R. Lueder, S. Drucker, and C. Wong. MyLifeBits: Fulfilling the Memex Vision. In *Proceedings of the Tenth ACM International Conference on Multimedia*, MULTIMEDIA '02, p. 235–238, New York, NY, USA, 2002. ACM.
- [7] Y. Gong and X. Liu. Video summarization using singular value decomposition. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, CVPR 2000, pp. 2:174–180. IEEE, 2000.
- [8] Google. Google Maps Timeline. <https://www.google.com/maps/timeline>. (accessed 2021-09-18).
- [9] GoPro. HER10 Black. <https://gopro.com/en/us/shop/cameras>, 2021. (accessed 2021-09-18).
- [10] Grand View Research. Dashboard Camera Market Size, Share & Trends Analysis Report By Technology (Basic, Advanced, Smart), By Product, By Video Quality, By Application, By Distribution Channel, By Region, And Segment Forecasts, 2020 - 2027, 2020.
- [11] S. Hodges, L. Williams, E. Berry, S. Izadi, J. Srinivasan, A. Butler, G. Smyth, N. Kapur, and K. Wood. SenseCam: A Retrospective Memory Aid. In *Proceedings of the 8th International Conference on Ubiquitous Computing*, UbiComp '06, p. 177–193, Berlin, Heidelberg, 2006. Springer-Verlag.
- [12] Holoride. Holoride. <https://www.holoride.com>, 2019. (accessed 2021-09-18).
- [13] Honda. Honda Dream Drive. <https://www.honda.co.jp/CES/2017/detail/007/>, 2017. (accessed 2021-09-18).
- [14] Insta360. Insta360 Go 2. <https://www.insta360.com>, 2021. (accessed 2021-09-18).
- [15] A. Matsuda, T. Matsui, Y. Matsuda, H. Suwa, and K. Yasumoto. A Method for Detecting Street Parking Using Dashboard Camera Videos. *Sensors and Materials*, pp. 33(1):17–34, 2021.
- [16] Mordor Intelligence. DASHBOARD CAMERA MARKET - GROWTH, TRENDS, COVID-19 IMPACT, AND FORECASTS (2021 - 2026). <https://www.mordorintelligence.com/industry-reports/dashboard-camera-market>. (accessed 2021-09-18).
- [17] C.-W. Ngo, Y.-F. Ma, and H.-J. Zhang. Automatic video summarization by graph modeling. In *Proceedings Ninth IEEE International Conference on Computer Vision*, ICCV 2003, pp. 1:104–109. IEEE, 2003.
- [18] Y. Sumi, R. Sakamoto, K. Nakao, and K. Mase. ComicDiary: Representing Individual Experiences in a Comics Style. In *Proceedings of the 4th International Conference on Ubiquitous Computing*, UbiComp '02, p. 16–32, Berlin, Heidelberg, 2002. Springer-Verlag.
- [19] Y. Takimoto, Y. Tanaka, T. Kurashima, S. Yamamoto, M. Okawa, and H. Toda. Predicting Traffic Accidents with Event Recorder Data. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Prediction of Human Mobility*, PredictGIS '19, p. 11–14, New York, NY, USA, 2019. ACM.
- [20] 尾頭 花奈, 椎尾 一郎. ドライブレコーダを活用した思い出ムービー自動編集システム. 研究報告ユビキタスコンピューティングシステム（UBI）, 2021-UBI-69(19), p. 1–8, 2021.
- [21] 石黒 祥生, 山田 献二朗. 自動運転車両内インフォテインメントのためのVRゲーム. エンタテインメントコンピューティングシンポジウム 2017 論文集, pp. 210–216, 2017.

TSUGITE: インタラクティブな木工ジョイントの設計・加工システム

Maria Larsson 吉田博則 梅谷信行 五十嵐健夫[†]

概要. 本研究では、フレーム構造用の木材ジョイントを設計・製作するためのインタラクティブなシステム「Tsugite」を提案する。フレーム構造用のジョイントの設計、手作業での製作は容易ではなく、時間を要す。本システムは、モデリング・インターフェースとコンピュータ数値制御（CNC）による加工を組み合わせることで、カスタムジョイントの作成を容易にする。設計空間には、効率的な幾何学的解析と組み合わせ検索を可能にするボクセルの3Dグリッドを採用している。ユーザーはリアルタイムのグラフィカルなフィードバックを受けながらジョイントを編集し、スライド方向、加工性、接合方向に対する耐久性などの性能指標に基づいた提案を受けることができる。接合部の設計が完了すると、3軸のCNCフライス盤を用いて、コーナーの丸みをつけて接合性を確保する専用のパス設計アルゴリズムを用いて接合部を製造することができる。

1 はじめに

釘を使わない木材の接合方法は、伝統的な工芸技術である。建築や家具の製作に応用されており、見た目の美しさ、品質の高さ、組み立てやすさといった点で評価されている。伝統的な接合方法には、ほぞつき、相欠きつき、組みつきなど様々な種類があり、これらの形状は、機能性と見た目の美しさを兼ね備える。このようなジョイントの設計と解析は、幾何学的な複雑さと同時に考慮しなければならない点が複数あり、容易ではない。さらに、手工具を使っての製作は、時間と労力を多く要する。そこで、本研究では、木工ジョイントを作成するためのインタラクティブなシステムを提案する。このシステムでは、実用的な多数の接合特性を解析し、CNCフライス加工によって接合部をデジタル処理する。システム名「Tsugite」は、日本語の「継手」に由来する。設計空間には、各ボクセルが一意の木材に属する、規則的に配置された3Dグリッドを採用した。この設定は効率的な計算に適しており、リアルタイムのフィードバック、提案、組み合わせ検索を可能にしている。本稿はACM UIST2020で発表済みの論文をまとめたものである[23]。

2 関連研究

Yaoらは、自由な形状の装飾ジョイントをデザインするためのツールを提案した[21]。このツールは、幾何学的自由度の高い装飾的なジョイントの設計をサポートする。しかし、このツールはリアルタイム

に提案やフィードバックを提供するものではない。

近年の複数の研究では、ジョイントや3Dパズルのトポロジカルな最適化を行い、さまざまな動作を実現している[2, 13, 14, 18]。これらでは、我々の研究と同様に、ボクセルの3Dグリッドのデザイン空間内にジョイントを作成する。しかし、ユーザーインターフェースがなく、CNCによる加工性や木目の方向に対する耐久性などの実用的な制約が考慮されていない。

組立可能な構造物を設計・製作するための計算機システムを提案している研究は、木材からCNCマシンでジョイントを削り出すシステム[7, 15, 22]や、レーザーカッターや3Dプリンタでオーダーメイドのジョイントを作るシステム[1, 4, 6, 10]など、多数ある。さらに、標準的な木工工具や、標準的な工具とデジタルファブリケーションマシンを組み合わせて実行できる製作プランを作成する方法を提案している研究も存在する[8, 19]。さらに別の研究では、家具の物理的な妥当性を分析し、モデリング中にユーザーに提案を提供している[16]。これらのシステムの対象は大局的な最適化であり、ジョイントそのものの自由度は高くなく、いくつかの標準ジョイントをパラメータで調整する程度となっている。本研究は、これらのシステムとは異なり、大局的な構造ではなく、ジョイントそのもののデザインに焦点を置いている。

CNCファブリケーションに特化したジョイント形状に着目した研究も複数ある。Grosは、3軸のCNCマシンに適したジョイント形状を設計し、「50 Digital Wood Joints」というオンラインライブラリをアップロードして自由に使えるようにした[3]。金崎と楊も同様に、日本の伝統的な形状のジョイント

Copyright is held by the author(s).

[†] 東京大学

をデジタルで加工可能な形状に変換することを目指している[5, 20].しかし、彼らの成果は、直交する2本以下の材木の間の1つか2つのジョイント形状に限られている。

また、CNC フライスのツールパス設計は1970年代から研究されている。ツールパス最適化問題は、一般に、意図した形状に切り込みますに除去される材料の体積を最大化することとして定式化される[11].本システムの設定では、相手の木材とぴったり合うように鋭い外角を切り落とす等、意図した形状に切り込む必要がある。到達できない領域があるジョイントのパス設計の問題は、分かっている範囲ではこれまでに検討されていない。

異方性材料の対象物の向きを活用した例として、Li と Barbic [9]は、木材、植物、筋肉などの異方性材料の挙動をシミュレートするモデルを提案している。Umetani と Schmidt [17]は、フィラメントを用いた3D プリントでは、xy 面での連続した材料の結合よりも z 方向での層間の材料の結合が弱いことを観察し、これを用いて物体の向きを適宜変更している。

3 ユーザインターフェース

本システムは、長方形の断面を持つ木材を接続するための単一の木材ジョイントを設計するツールとして実装されている。図1にその概要を示す。ユーザーは、スライド方向、交点の位置、材木の数などのジョイントタイプの変数を指定する(図1a).これらの情報は、構造物や家具の全体的な設計から得られるものであり、その自動設定は本研究の範囲外である。マニュアル編集モードでは、グラフィカルなフィードバックや提案を受けながらジオメトリを操作する(図1b).ギャラリーモードでは、ユーザーは有効なジオメトリを見て、その中から1つを選ぶことができる(図1c).最後に、Tsugite システムは、フライスピットの半径を入力とし、ツールパスを CNC フライス盤にエクスポートする。

(図1d)

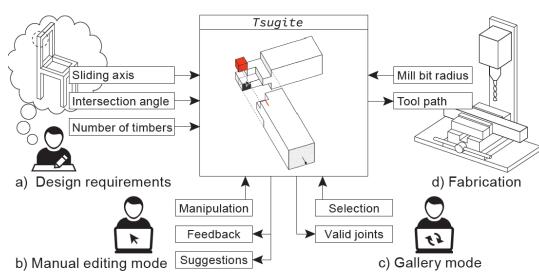


図 1. 提案システムの概要.

3.1 基本的な操作

ユーザーは、スライド方向の面を押したり引いたりすることで、ジョイントを編集することができる(図2a).さらに、材木の本体をクリックしてドラッグすることで、材木の位置や向きを変更することができる(図2b, 2c).また、x, y, z キーを押すことで、直交するスライド軸を選択することができる(図2d). スタックのように、すべての木材が1つのスライド軸に沿ってスライドするデザインにも対応している。交差する木材の数は、理論的には6本まで対応可能だが、実際の加工場面を考慮すると、4本以下の木材の接続に適している(図2e). デフォルトのボクセル解像度(ジョイントの要素分割数)は、 $3 \times 3 \times 3$ であり、2~5の間で変更可能になっている(図2f).また、ユーザーは、非直交の交差角度(図2g), 木材の断面の高さと幅(図2h)を設定することができる。

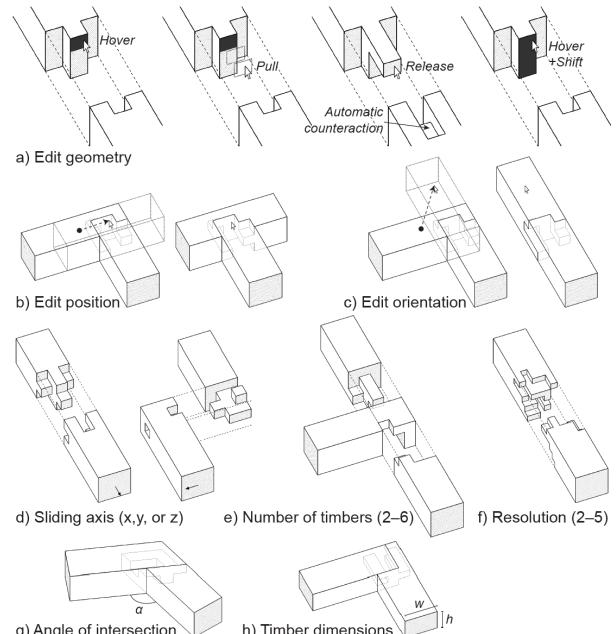


図 2. 基本的な操作.

3.2 フィードバック

Tsugite は、以下の8つの性能指標に基づいて、ユーザーに視覚的なフィードバックを提供する(図3).

- a) **Connectivity:** 木材の本体から切り離されたボクセルは赤で表示される
- b) **Bridging:** 木材の中ほどにジョイントがある場合(Tジョイント)，その木材のジョイントの両端は接続(ブリッジ)されていなければならぬ。ブリッジされていない場合部分が二色で表示される。
- c) **Milling direction:** 切削は一方向からのみ行われるので、そのような条件で加工できない形状の場

TSUGITE: インタラクティブな木工ジョイントの設計・加工システム

合は、木材の本体がオレンジ色で表示される。このような状況は、接合部に3本以上の木材がある時の、間に挟まれた木材にのみ発生する。

- d) **Checkerboard:** 対角上に2種類の木材が存在しているパターンは禁止されており、中央に太い赤い縦線で示される。禁止されている理由の詳細については、実装の項を参照。
- e) **Slidability:** ジョイントは一方向からのみスライドできることを望ましい。現状のデザインでスライド可能な方向が、各木材の端にある矢印によって示される。複数のスライド方向がある場合は、その木材のアウトラインが赤く表示される。
- f) **Durability:** 木目の方向に対して垂直に突き出ている部分は簡単に折れてしまう可能性がある。従来のジョイント形状でもこののようなパーツは避けられている。そのためそのようなボクセルは黄色で表示される。形状からの単純な判定であり物理的な負荷の計算を行っているわけではない。
- g) **Contact area:** 接着を目的とした接合部では、異なる木材の素材同士が接触する接触面が大きいほど望ましい。接触面はテクスチャで表示される。
- h) **Friction area:** 摩擦で固定することを目的とした接合部では、摩擦面積が大きい方が好ましい。摩擦面とは、接触面の中で面の法線がスライド方向と垂直なものである。摩擦面はテクスチャで表示される。形状からの単純な判定であり物理的な摩擦の力を計算しているわけではない。

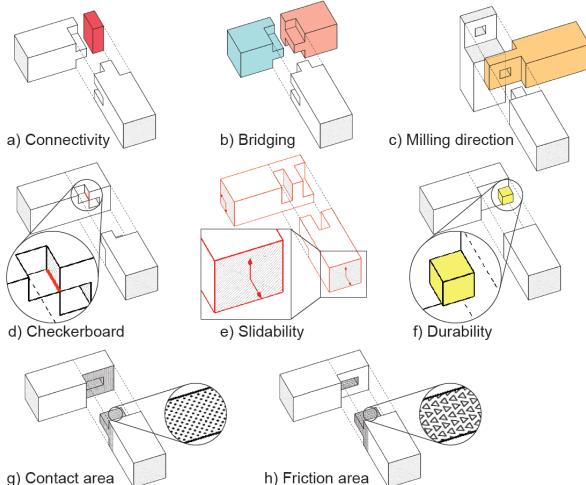


図 3. 視覚フィードバック。

3.3 加工

デザインを決定した後、ユーザーはミリングパスをプレビューし、それを CNC マシンにエクスポートすることができる。ジョイントを製作するには、CNC マシンのベッドに材料をセットし、マシンの起點を設定し、マシンを稼働させる。

4 実装

本システムは、3軸の CNC マシンによって加工可能なジョイントの設計を支援する (4.1)。支援にあたっては、8種類の性能指標の評価を行う (4.2)。設計が完了すると、ジョイントの組立性を確保するようにパス設計を行う (4.3)。このシステムは Python で実装され、OpenGL と GLFW が主な依存関係にある。

4.1 加工可能性

ジョイントの幾何学的基準は、製作方法によって異なる。本研究では、手頃な価格で普及している標準的なフライスピットを装備した3軸 CNC マシンを用いてジョイントを製作する。この機械のセットアップには、2つの大きな制約がある。まず、内側からみて凹んだ角を円筒状のフライスピットで内側から削ることは不可能である (図 4a)。このような角は、フライスピットの半径で丸みを帯びていた形状となる。第二に、機械は材料の上からしかアプローチできない。つまり、上からのアクセスが妨げられているような形状を加工することはできない (図 4b)。

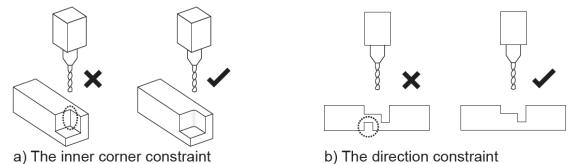


図 4. 3軸 CNC を用いた際の加工の制約。

内側の角の制約(図 4a)は、システムにとって2つの意味を持つ。一つ目の問題は、ボクセル化された形状を、ジョイントの各木材を個別に考慮する従来のパスプランニングツールで加工しようとした場合、1本の木材の内側の角の余剰材料が、他の木材の対応する鋭い外側の角と衝突するため、ピースを組み立てることが不可能となる (図 5a)。この問題に対する一般的な解決策は、内側の角からさらに材料を取り除く (図 5b) か、外側の角を丸くする (図 5c) かの二通りとなる。この2つの解決策は、既存のデジタルジョイントにも見られる。本研究では、エアポケットがない方が見た目が良く、摩擦面積と強度が増すため、後者の解決策 (外側の角を丸める) を採用した。角を丸めることを自動化することは、本研究の特殊なパスプランニングアルゴリズムの重要な機能である (詳細は「パス計算」の項を参照)

a) Problem	b) Solution 1	c) Solution 2 (ours)

図 5. 通常のパス計画の問題.

二つ目の問題は、対角上に 2 種類の材木が存在する Checkerboard に関するものである（図 6）。もし、追加のルールなしに Checkerboard を認めるとき、材木を組み立てることができなくなる（図 6a）。代替ルールとしては、1 本の木材の角を外側の角のように丸めることができると考えられる（図 6b）。しかし、2 本目の木材の 2 つの突出した部分の間の隙間 (d) は、フライスの直径よりも狭くなり、製作することができなくなる。3 つ目の可能性は、この隙間をフライスピットが通過できるまで大きくすることである（図 6c）。この場合は、形状を製作して組み立てることができるが、システムが複雑になるため、あまり良い解決策ではない（どの木材にどのルールを適用するか、ユーザーが判断する必要がある）。さらに、片側の材料がもう片側に比べてかなりの量が削られるため、隣り合う Checkerboard の場合、フライスピットが届かないような狭い部分ができてしまう（図 6d）。そこで、Checkerboard を全面的に禁止することにした。

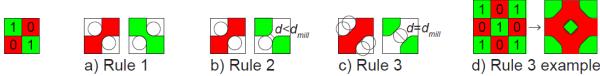


図 6. Checkerboard Pattern.

方向制約（図 7）として、各木材が单一の方向のみから加工されるという制限を導入している。さらに、単一の方向から加工されたジョイントは、必ずその方向にスライドさせることができる（図 7a）。この性質を利用して問題を簡単にするために、ジョイントのすべての木材のスライド軸は 1 つだけであり、製作軸とスライド軸は共有されているという制限を設けた。これにより、組み立てられた状態で 2 本の木材（最初の木材と最後の木材）だけがスライドして出てくるような形状を実現することができる（図 7b）。

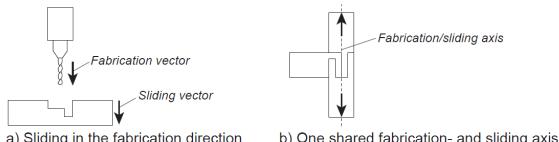


図 7. 加工方向制約.

4.2 性能指標

8 つの性能指標は以下のように計算される。

a) **Connectivity.** 木材のすべての固定端から始まる塗り潰しアルゴリズムを実行し、木材に属するすべてのボクセルをカバーしているかどうかを判断す

る。カバーされたボクセルは接続され、カバーされていないボクセルは接続されない。

b) **Bridging.** ブリッジングは、2 つの固定端を持つ木材について評価される。木材の 1 つの固定端から始まる塗り潰しアルゴリズムを実行する。2 番目の固定端に隣接するボクセルが含まれる場合、ジョイントはブリッジされる。

c) **Milling direction.** 2 本以上の木材がある場合、中間の木材が製作可能かどうかを検証する。中間材は、スライド軸に沿った 2 つの方向のうちの 1 つ（上からか下からか）からミリングする必要がある。システムは自動的に適切な加工方向を決定する。中間材を両方の方向から加工する必要がある場合、その形状は無効であると判断される。

d) **Checkerboard.** 内側のすべての頂点について、製作軸に垂直な平面上の 4 つの隣接する頂点を分析する。交差する対角線が 2 本ある場合、市 checkerboard と判定され警告が表示される（図 8a）。対角線が 1 つしかなく、他の 2 つのボクセルが異なる材木に属している場合は、checkerboard ではないと判定される（図 8b）。

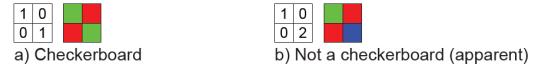


図 8. Checkerboard の解析.

e) **Slidability.** 直交する 6 つのスライド方向をそれぞれ個別に評価する。各木材 ID と、テストされている方向に整列した各ボクセルの列について、その方向で異なる ID のボクセルの後に、同じ ID のボクセルがあるかどうかを確認する。同じ ID のボクセルがあった場合ボクセルは衝突を起こすため、木材がその方向にスライドするのを防ぐ。

f) **Durability.** 材の纖維方向を分断するような加工は、材の耐久性を低下させると判断されユーザに提示される。なお材の纖維方向と長手方向は一致していると仮定する（図 9）。

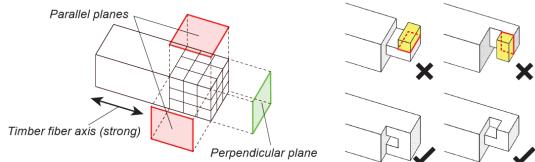


図 9. 耐久性の解析.

g) **Contact area.** 各木材について、他の木材に隣接する外表面の面積の合計を計算する（図 10a）。

h) **Frication area.** 各木材について、他の木材に隣接する外表面の面積と、遮断されていないスライド方向に垂直な面にない外表面の面積の合計を計算する

TSUGITE: インタラクティブな木工ジョイントの設計・加工システム

(図 10b). 接触面積と同じ原理で、摩擦面積が最も小さいスライド可能な木材が接合部の摩擦面積を表す。

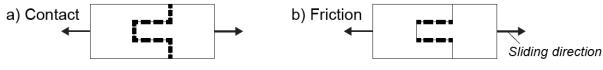


図 10. 接触面積と摩擦面積.

4.3 パス計算

ツールパスは各木材の 3 次元バイナリマトリクスから直接生成される。ジョイントは、ボクセルの最上層から層ごとに削り取られる。まず、ボイド領域が検出される(図 11a)。ボイド領域とは、接続されたボクセルのうち、除去すべき領域のことを指す。ボイド領域を囲む材料に沿って、内側にオフセットした最終パスを作成する(図 11b)。ミーリングビットの中心はパスに沿って移動し、その半径の長さに応じて両側の材料を除去する。したがって、パスは材料の端からフライスピットの半径分だけオフセットされる。



図 11. パス計算アルゴリズム.

そして、最終パスの外側の角を条件付きで丸める(図 11c)。これが、ジョイントの組立性を確保するアルゴリズムの重要な機能である。外側の角は、相手材の内側の角に位置している場合、つまり、隣接する 2 つの空のボクセルが共通の材で占められている場合に丸められる(図 12a)。そうでない場合、2 つの隣接する空のボクセルが異なる木材で占められていれば、角は丸められない(図 12b)。最終パスを作成した後、最終パスでカバーされていない空ボクセル領域の範囲は、共線的な振動荒加工パスで追加切削される(図 11d)。ボイド領域の最終的なツールパスは、最終パスと粗加工パスを組み合わせたものである(図 11e)。そして、アルゴリズムは、すべての空のボクセルがカバーされるまで、次の領域、そして次の層へと進む。

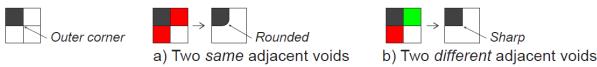


図 12. 接触面積と摩擦面積.

5 結果

本システムの有用性と能力を示すために、実際に製作された、もしくはグラフィックとして展示された機能的な家具やその他の組立品を紹介する。図 13

は、本システムを用いて製作した椅子である。この椅子には 9 つのジョイントがあり、その内訳は 5 つの 2 本から成るジョイントと 4 つの 3 本からなるジョイントである。ジョイントは摩擦のみで固定されており、人が座っても安定している。図 14 はテーブルで、L 字型、斜め T 字型、X 字型など 14 個のジョイントがある。



図 13. 椅子の製作例.



図 14. テーブルの製作例.

様々な継ぎ手の種類、パラメータ、技術的特性を示すために、いくつかのサンプルを設計し、製作した。図 15 は、2 本から成るティンバージョイントの向きや位置がすべて異なる接合部の例である。さらに、3 本、4 本の木材が交差するジョイントも製作した(図 16)。また、他の解像度を試すために、デフォルトではない設定でジョイントを設計、製作した。最も低い解像度である $2 \times 2 \times 2$ では、実現可能性が非常に低く、有効なジョイントもほとんどない(図 17a)。最高解像度の $5 \times 5 \times 5$ では、面数の多いモデルでは少々手間はかかるが、バラエティに富んだ高摩擦の関節を作ることができた(図 17b)。また、斜めの接合部を作ることで交差角のパラメータを実証し(図 18a)、正方形ではない断面の木材でジョイントを作ることで高さと幅の寸法パラメータを実証した(図 18b)。

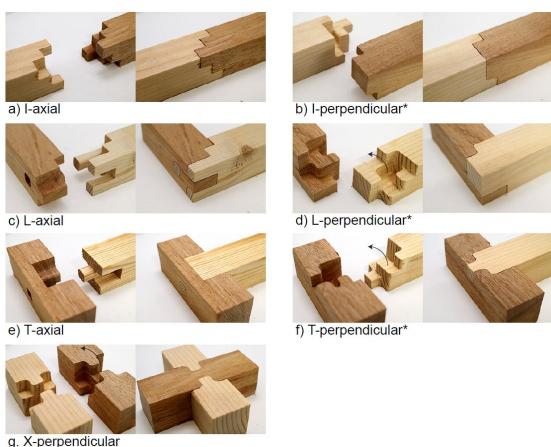


図 15. 様々な 2 本のジョイントの製作例.

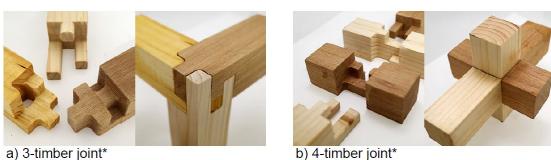


図 16. 3 本および 4 本のジョイントの製作例.

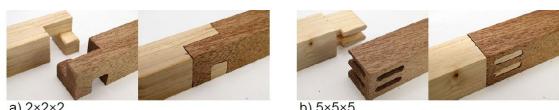


図 17. 異なる解像度の製作例.



図 18. 斜めおよび非正方形ジョイントの製作例.

組立品は、45x45mm の木の棒を 10mm のフライスピットで加工した。接合サンプルは、30x30mm の明るい色と暗い色の木の棒を 90mm の長さで加工したもので、一部は塗装した。特に理由のない限り、フライスは直径 6mm のものを使用した。使用したのは、Pro Spec Tools 社の 3 軸 CNC フライス、モデル PSF240 である。フライス加工に要した時間は、接合部の形状と木材の寸法に応じて、1 本あたり 12 ~ 25 分であった。機械の原点を手動で設定する際の誤差は、実際の材料に比べてやや小さめの幅を入力することと、ジョイントを組み立てた後に側面をやすりがけすることで対処した。

6 まとめ

フレーム構造用の木材ジョイントを設計・製作す

るためのインタラクティブなシステム「Tsugite」を提案した。将来的には、ジョイント単体のデザインだけでなく、構造体全体の設計プロセスを統合することが有効であると考えられる。

謝辞

本研究は、JST ACT-I JPMJPR17UT, JSPS KAKENHI 17H00752, JST CREST JPMJCR17A1による支援を受けて実施したものである。

参考文献

- [1] Patrick Baudisch, Arthur Silber, Yannis Kommana, Milan Gruner, Ludwig Wall, Kevin Reuss, Lukas Heilman, Robert Kovacs, Daniel Rechlitz, and Thijs Roumen. 2019. Kyub: A 3D Editor for Modeling Sturdy Laser-Cut Objects. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19). ACM, New York, NY, USA, Article 566, 12 pages. DOI:<http://dx.doi.org/10.1145/3290605.3300796>
- [2] Chi-Wing Fu, Peng Song, Xiaoqi Yan, Lee Wei Yang, Pradeep Kumar Jayaraman, and Daniel Cohen-Or. 2015. Computational Interlocking Furniture Assembly. ACM Trans. Graph. 34, 4, Article 91 (July 2015), 11 pages. DOI:<http://dx.doi.org/10.1145/2766892>
- [3] Jochen Gros. 2020. 50 Digital Wood Joints by Jochen Gros. (2020). <http://winterdienst.info/50-digital-wood-joints-by-jochen-gros/>
- [4] Alec Jacobson. 2019. RodSteward: A Design-to-Assembly System for Fabrication using 3D-Printed Joints and Precision-Cut Rods. Computer Graphics Forum (2019).
- [5] Kenji Kanasaki and Hiroya Tanaka. 2013. Traditional Wood Joint System in Digital Fabrication. Computation and Performance - Proceedings of the 31st eCAADe Conference 1 (2013), 711–717. <http://resolver.tudelft.nl/uuid:d36152ad-7fcf-44b6-bdfe-654f159a3e65>
- [6] Robert Kovacs, Anna Seufert, Ludwig Wilhelm Wall, Hsiang-Ting Chen, Florian Meinel, Willi Müller, Sijing You, Yannis Kommana, and Patrick Baudisch. 2017. Demonstrating TrussFab: Fabricating Sturdy Large-Scale Structures on Desktop 3D Printers. (2017), 445–448. DOI:<http://dx.doi.org/10.1145/3027063.3050427>
- [7] Maria Larsson, Hironori Yoshida, and Takeo

TSUGITE: インタラクティブな木工ジョイントの設計・加工システム

- Igarashi. 2019. Human-in-the-loop Fabrication of 3D Surfaces with Natural Tree Branches. In Proceedings of the ACM Symposium on Computational Fabrication (SCF '19). ACM, New York, NY, USA, Article 1, 12 pages. DOI: <http://dx.doi.org/10.1145/3328939.3329000>
- [8] Danny Leen, Tom Veuskens, Kris Luyten, and Raf Ramakers. 2019. JigFab: Computational Fabrication of Constraints to Facilitate Woodworking with Power Tools. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19). ACM, New York, NY, USA, Article 156, 12 pages. DOI: <http://dx.doi.org/10.1145/3290605.3300386>
- [9] Yijing Li and Jernej Barbić. 2014. Stable Orthotropic Materials. In Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation (SCA '14). Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, 41–46. <http://dl.acm.org/citation.cfm?id=2849517.2849524>
- [10] Shiran Magrisso, Moran Mizrahi, and Amit Zoran. 2018. Digital Joinery For Hybrid Carpentry. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18). ACM, New York, NY, USA, Article 167, 11 pages. DOI: <http://dx.doi.org/10.1145/3173574.3173741>
- [11] Saigopal Nelaturi, Gregory Burton, Christian Fritz, and Tolga Kurtoglu. 2015. Automatic Spatial Planning for Achining Operations. In Proceedings of The eleventh annual IEEE International Conference on Automation Science and Engineering (IEEE CASE 2015). Gothenburg, Sweden. nel-bur-fri-kur-case2015.pdf
- [12] Robert J. Ross. 2010. Wood handbook—Wood as an engineering material. Department of Agriculture, Forest Service, Forest Products Laboratory.
- [13] Peng Song, Chi-Wing Fu, and Daniel Cohen-Or. 2012. Recursive Interlocking Puzzles. ACM Transactions on Graphics (SIGGRAPH Asia 2012) 31, 6 (December 2012), 128:1–128:10.
- [14] Peng Song, Chi-Wing Fu, Yueming Jin, Hongfei Xu, Ligang Liu, Pheng-Ann Heng, and Daniel Cohen-Or. 2017. Reconfigurable Interlocking Furniture. ACM Trans. Graph. 36, 6, Article 174 (Nov. 2017), 14 pages. DOI: <http://dx.doi.org/10.1145/3130800.3130803>
- [15] Rundong Tian, Sarah Sterman, Ethan Chiou, Jeremy Warner, and Eric Paulos. 2018. MatchSticks: Woodworking Through Improvisational Digital Fabrication. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18). ACM, New York, NY, USA, Article 149, 12 pages. DOI: <http://dx.doi.org/10.1145/3173574.3173723>
- [16] Nobuyuki Umetani, Takeo Igarashi, and Niloy J. Mitra. 2012. Guided Exploration of Physically Valid Shapes for Furniture Design. ACM Trans. Graph. 31, 4, Article Article 86 (July 2012), 11 pages. DOI: <http://dx.doi.org/10.1145/2185520.2185582>
- [17] Nobuyuki Umetani and Ryan Schmidt. 2013. Cross-sectional Structural Analysis for 3D Printing Optimization. In SIGGRAPH Asia 2013 Technical Briefs (SA '13). ACM, New York, NY, USA, Article 5, 4 pages. DOI: <http://dx.doi.org/10.1145/2542355.2542361>
- [18] Ziqi Wang, Peng Song, and Mark Pauly. 2018. DESIA: A General Framework for Designing Interlocking Assemblies. ACM Trans. Graph. 37, 6, Article 191 (Dec. 2018), 14 pages. DOI: <http://dx.doi.org/10.1145/3272127.3275034>
- [19] Chenming Wu, Haisen Zhao, Chandrakana Nandi, Jeffrey I. Lipton, Zachary Tatlock, and Adriana Schulz. 2019. Carpentry Compiler. ACM Trans. Graph. 38, 6, Article Article 195 (Nov. 2019), 14 pages. DOI: <http://dx.doi.org/10.1145/3355089.3356518>
- [20] Yuan Chieh Yang. 2016. Classical Through Digital. (2016). <https://space10.com/project/classical-through-digital/>
- [21] Jiaxian Yao, Danny M. Kaufman, Yotam Gingold, and Maneesh Agrawala. 2017. Interactive Design and Stability Analysis of Decorative Joinery for Furniture. ACM Trans. Graph. 36, 2, Article 157a (March 2017). DOI: <http://dx.doi.org/10.1145/3054740>
- [22] Hironori Yoshida, Maria Larsson, and Takeo Igarashi. 2019. Upcycling Tree Branches As Architectural Elements Through Collaborative Design and Fabrication. In Proceedings of the Thirteenth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '19). ACM, New York, NY, USA, 589–593. DOI: <http://dx.doi.org/10.1145/3294109.3295639>
- [23] Maria Larsson, Hironori Yoshida, Nobuyuki Umetani, Takeo Igarashi. Tsugite: Interactive Design and Fabrication of Wood Joints. UIST 2020.

空間ラベリング：非専門家による画像注釈のラベル品質を改善するための空間レイアウトの活用

張家銘* 李佳憲† 五十嵐健夫*

概要. 非専門家による画像注釈において、ラベルの品質は重要な問題である。本研究では、非専門家の画像注釈のラベル品質を改善するために空間レイアウトを活用することを提案する。本システムでは、アノテーターは最初にターゲット画像を空間的にレイアウトし、関連するアイテムと一緒に配置しながらオープンスペースでラベル付けを行う。これは、暫定的なラベリングのための作業スペース（空間組織）として機能する。作業中に、アノテーターはアイテム間の類似点と相違点を観察して整理する。最後に、アノテーターは空間レイアウトの結果に基づいて画像に最終的なラベルを割り当てる。画像ラベリングタスクを用いて本システムと従来の非空間レイアウトを比較するユーザー調査を行った。結果として、アノテーターは非空間レイアウトインターフェイスよりも空間レイアウトインターフェイスを使用することで、より正確にラベリングタスクを完了できることを示した。

1 はじめに

非専門家による画像のラベル付けは非常に困難であり、多くのエラーが含まれている可能性がある。したがって、プロセスをより効率的にし、エラーを減らすためのサポートを提供することが重要である。例えば、犬の品種に関する十分な知識のないアノテーターが、犬の品種のラベルを犬の画像に割り当てるよう求められた場合、ラベルに関連付けられたサンプル画像を参照することができるが、サンプル画像では類似した品種間の微妙な違いを捉えるには不十分な場合がある。このような場合、アノテーターは、サンプル画像とラベルを割り当てる画像を比較することにより、微妙な違いを観察および整理できる。この観察と整理のプロセスは、非専門家によるラベリングにおいて重要な役割を果たしていると考えられるが、従来の画像ラベリングツールでは十分にサポートされていない。本研究では、アノテーターの空間レイアウトを活用して、画像とラベルの類似点と相違点の観察および整理することをサポートする。この空間編成プロセスは、暫定的なラベリングとして機能する。

図1は、プロトタイプの実装のスクリーンショットである。概念的な類似性を表す画像とラベルを空間的にレイアウトするためのオープンスペースがアノテーターに提示される。アノテーターはターゲット画像をオープンスペースにドラッグし、可能なラベルを近くに配置する（つまり、画像はラベルの1

つに属している可能性がある）。このプロセスにより、アノテーターは特定のドメインで画像のラベルをより正確に選択できるようになると考えられる。



図1. 空間ラベリングプロトタイプ実装のスクリーンショット。

本研究の空間レイアウトラベリングと従来の非空間レイアウトラベリングを比較するため、画像ラベリングタスクを用いてユーザー実験を行った。結果として、参加者は非空間レイアウトインターフェイスよりも空間レイアウトインターフェイスを介して特定のラベリングタスクをより正確に完了したことが示された。さらに、参加者は、非空間レイアウトラベリングタスクよりも空間レイアウトラベリングタスクで選択したラベルにより自信を持っていた。これらの調査結果は、空間レイアウトインターフェイスが、ラベル付けプロセス中にオープンスペースに画像とラベルをレイアウトすることにより、専門家以外の画像注釈のラベル品質を向上できることを示している。

本論文は、ACM CHI 2021 で発表された論文の内容をまとめたものである[17]。

Copyright is held by the author(s).

* 東京大学, †LeadBest Consulting Group

2 関連研究

手動の画像注釈（例：VIA [1]、LabelMe [2]、ESP [3] および階層的タスク割り当て [4]）を支援し、非専門家の注釈（例：Revolt [5]）のラベル品質を向上させるために、さまざまな注釈ツールが提案されている（Pairwise HITS [6]、2 ラウンドアノテーションフレームワーク [7]）。これらのツールのほとんどは、労働負荷の高いアノテーション作業のプロセスとラベルの品質を向上のために、より支援的で効率的かつ楽しむことができるシステムを提供している。しかし、これらのツールは通常、アノテーターが十分なドメイン知識を持っていることを前提としている。さらに、十分なドメイン知識が不足している専門家以外のアノテーターのサポートはまだ制限されており、改善する必要がある。クレシャラ [8] は、Web ページ分類のためのデータラベリングにおける概念進化という考えを導入し、データラベリング中にアノテーターが概念を定義および改良するのに役立つ 2 つの構造ラベリングソリューションを提示した。本研究とは、データラベリングにおけるアノテーターのコンセプト編成をサポートするという同じ動機を共有している。ただし、クレシャラ [8] は主に Web ページ分類のコンテキストでの時間の経過に伴う概念の変化に焦点を当てている一方、本研究は画像分類のコンテキストの単一セッションでのアイテムの観察と概念の整理に焦点を当てている。

空間レイアウトの概念は、情報ナビゲーション用のズーム可能なインターフェイスなど、さまざまな研究で使用してきた[9, 10]。パウアーラ [11]は、ズーム可能なインターフェイスの概念と「パイル」メタファーに基づいて個人情報収集を管理するための空間ツールを提案した。渡辺ら[12]は、「バブルクラスター」という名前のインターフェイスを導入したが、これはオブジェクトのグループ化とグループ化解除を通じて、グラフィックオブジェクトの空間的集約を操作するための手法である。人々が空間レイアウトインターフェイスを介して情報やドキュメントをより効果的に管理できる理由の1つとして、空間記憶（つまり、環境内の何かを記憶する能力）が挙げられる[13]。空間レイアウトの概念は、アノテーターが画像とラベルの類似点と相違点を観察および整理するためのラベリングインターフェイスデザインにも使用できると考えられる。さらに、空間レイアウトは、キャンバス上でのユーザーインターラクションを活用して検索プロセスを容易にする検索システムでも使用されている[14, 15]。対照的に、本研究では、ユーザーが手動で画像に注釈を付ける際に考えを整理するためのワークスペースとして、空間レイアウトインターフェイスを使用する。空間レ

アウトは、システムによって作成または解釈はされない。

3 提案手法

本研究のラベリングインターフェイスは、空間レイアウトの概念に基づいて設計された。本システムは、ラベリングプロセス中に概念的な類似性を表す画像とラベルを空間的にレイアウトするためのオープンスペースを提供する。図2に、インターフェイスの初期状態を示す。ラベル（各ラベルにはテキストで表示される名前とサンプル画像が含まれている）がインターフェイスの両側に一覧表示され、ラベル付けされるターゲット画像が中央に1つずつ表示される。

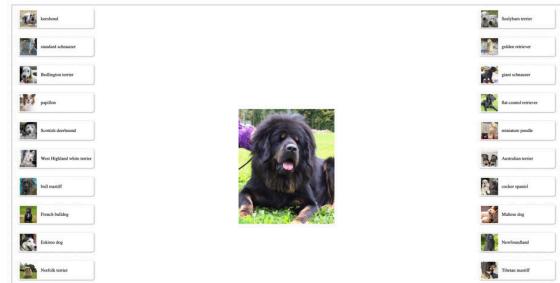


図2. 空間ラベリングインターフェイスの初期状態.

図3は、インターフェイスの動作状態を示している。アノテーターは、画像のラベルを選択する前に、オープンスペースに画像とラベルを空間的に配置する(たとえば、関連するアイテムと一緒に配置する)。



図3. 空間ラベリングインターフェイスの動作状態. 空間レイアウトインターフェイスには、アノテーターがラベル付けタスクを完了するための4つの主要な機能が含まれている：(a) 画像とラベルを空間的にレイアウトする、(b) 画像にラベルを割り当てる、(c) 確信度を示す、(d) 割り当てられたラベルを変更する（図4を参照）。

(a) 画像とラベルを空間的にレイアウトする
オープンスペースでは、アノテーターは概念的な類似性を表すために画像とラベルをレイアウトする。

空間ラベリング：非専門家による画像注釈のラベル品質を改善するための空間レイアウトの活用

より類似した画像がグループ化され、候補となるラベルが近くに表示される。

(b) 画像へのラベルの割り当て

オープンスペースに画像とラベルを配置した後、アノテーターは空間レイアウトの結果に基づいて画像に最終的なラベルを割り当てる。

(c) 信頼状態を示す

画像にラベルを割り当てる直後（ドラッグアンドドロップ操作）、ポップアップウィンドウが表示され、アノテーターにラベル選択の確信度を示すように求められる。

(d) 割り当てられたラベルの変更

ラベルをクリックするとそのラベルが割り当てられたすべての注釈付き画像を表示される。アノテーターは、画像を別のラベルまたはオープンスペースにドラッグアンドドロップすることで、既に割り当てられているラベルを変更できる。

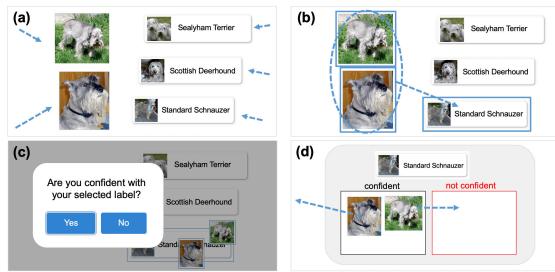


図 4. 空間レイアウトラベリングインターフェイスのユーザーインターラクション

4 ユーザー実験

16人の参加者（18～49歳の男性8人と女性8人）が、ユーザー実験に参加した。参加者は全員、画像ラベリングまたは画像分類に関する専門知識の経験はなかった。また、全参加者がペットとして犬を飼っておらず、犬のカテゴリーについての専門知識を持っていなかった。

ラベル付けタスクには、ImageNet (ILSVRC 2012) [16]の画像データセットを使用した。まず、犬のラベル（品種）ごとに100枚の画像をランダムに選択し、合計12,000枚の画像を基本データセットとして使用した。次に、同等の難易度の2つの互いに素なデータセット（データセットAとデータセットB）を手動で作成した。各データセットには、20のラベル（20の犬種）と50の画像（選択した20の犬種に属する50の犬の画像）が含まれていた。2つのデータセットのラベルと画像の間に重複はなかった。図5は、3つの難易度レベルでデータセットAとデータセットB用に選択された20の犬種（ラベル）を示している。



図 5. データセットAおよびデータセットBで使用される20の犬種（ラベル）。レベル1：明確。レベル2：あいまい。レベル3：非常にあいまい。

ユーザー評価のために2つのオンライン画像ラベリングシステムがReact.jsで開発された。（a）非空間レイアウトラベリングシステムと（b）空間レイアウトラベリングシステムである。

参加者内比較が採用され、各参加者は、2つのラベリングシステムを使用して2つのラベリングタスク（NSタスクとSタスク）をするように求められた。

・NSタスク：非空間レイアウトイインターフェイスと条件

ベースライン条件となっている。図6(a)は、非空間レイアウトラベリングシステムのスクリーンショットである。インターフェイスの左側には、ラベルがテキストとサンプル画像の両方で一覧表示されている。インターフェイスの右側には、ラベル付けするターゲット画像が表示される。参加者は、20の犬種（20のラベル）のリストから適切なラベルを選択し、50の犬の画像にラベルを付けるように求められた。この非空間レイアウトイインターフェイスでは、一度に1つの画像しか表示されない。参加者は、左側のリストをスクロールしてターゲット画像に適切なラベルを探し、暫定的な割り当てとしてラベルを選択した（選択したラベルがターゲット画像の下に表示された）。次に、「次へ」をクリックして割り当てを完了した。その後、選択したラベルについての自信を示すための「選択したラベルに自信があるか？」という質問に答えた。その後、「次へ」をクリックして次の画像に移動した。以前に選択したラベルを変更したい場合は、「前へ」をクリックして前の画像に戻ることも可能であった。

・Sタスク：空間レイアウトイインターフェイスと条件

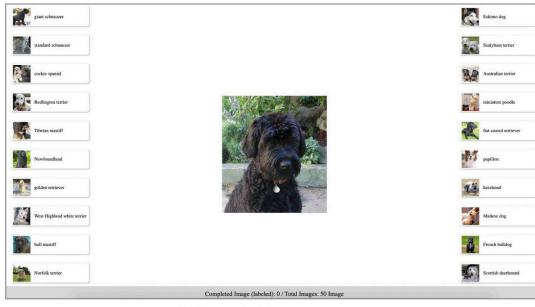
本研究の提案手法である。図6(b)は、空間レイアウトラベリングシステムのスクリーンショットを示している（空間レイアウトイインターフェイスについてはセクション3で説明した）。ラベル付けタスク（50匹の犬の画像にラベルを付ける）には、非空間レイアウトイインターフェイスに割り当てられたも

のと同じタスクが含まれていた。Sタスクでは、最終的なラベルを画像に割り当てる前に、最初にすべての画像と概念的な類似性を表すラベルをレイアウトするように参加者に指示した。

本ラベリングシステムは、参加者がラベリングタスクを完了するまでに要した時間とエラー率（画像に適切なラベルを選択できなかった）を自動的に記録および測定した。タイマーは、参加者が「開始」をクリックすると開始し、「終了」をクリックすると停止した。システムは、参加者がラベル付けに費やした時間も記録した。また、自信のあるラベル選択率（参加者がつけたラベルに自信を持っていた画像）も測定した。



(a) non-spatial layout system



(b) spatial layout system

図6. 非空間および空間レイアウトラベリングシステムのスクリーンショット.

5 結果

図7は、参加者が平均16分47秒と17分55秒を費やして、非空間および空間レイアウトインターフェイスを使用して50枚の画像にそれぞれラベルを付けたことを示している。タスク完了時間に関して、対応のあるt検定の結果は、非空間レイアウトインターフェイスと空間レイアウトインターフェイスの間に有意差がないことを示した ($p > 0.05$)。

図8は、非空間および空間レイアウトインターフェイスを使用して参加者が完了したラベリングタスクのエラー率（画像に適切なラベルを選択できない）を示している。結果として、エラー率が非空間レイアウトインターフェイスで43.50%、空間レイアウト

トインターフェイスで37.63%であることがわかった。対応のあるt検定を使用したエラー率の分析では、2つのラベリングインターフェイス間に有意差 ($p < 0.05$) があることが示された。

この結果は、非専門家のアノテーターが、タスク完了時間を大幅に増加させることなく、従来の非空間レイアウトインターフェイスよりも空間レイアウトインターフェイスを使用することで、特定のラベリングタスクをより正確に完了できることを示した。

画像データセットの3つの難易度レベルでエラー率を分析した（図9を参照）。結果は、レベル1とレベル3で非空間レイアウトインターフェイスと空間レイアウトインターフェイスの間に有意差 ($p > 0.05$) がなかったのに対し、レベル2では有意差 ($p < 0.05$) があることがわかった。これは、空間レイアウトインターフェイスの利点が、ラベリングタスクにあいまいな画像が含まれている場合にのみ現れることを示している。この利点は、ターゲット画像があいまいでないか、あいまいすぎない場合には得られない。

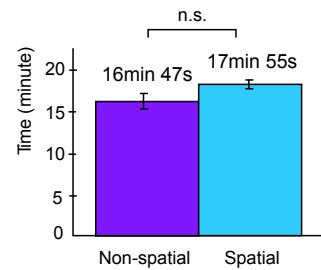


図7. ラベル付けタスクの完了時間.

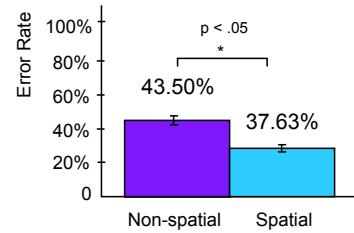


図8. ラベリングタスクのエラー率.

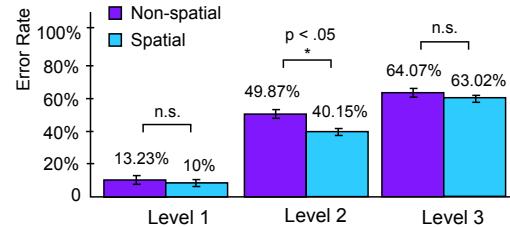


図9. 難易度のエラー率.

ユーザー実験では、参加者は、選択したラベルの確信度（自信があるかどうか）を示すように求めら

空間ラベリング：非専門家による画像注釈のラベル品質を改善するための空間レイアウトの活用

れた。図10は、参加者が非空間レイアウトタスクでラベル付けされた画像の47.13% (24/50 画像) に自信を持っていたのに対し、空間レイアウトタスクでは 59.63% (30/50 画像) であることを示している。対応のあるt検定の結果、非空間レイアウトインターフェイスと空間レイアウトインターフェイスの間に有意差 ($p < 0.01$) があることがわかった。これは、空間レイアウトインターフェイスが、手動の画像注釈に対する非専門家の主観的な自信を高めることができることを示している。

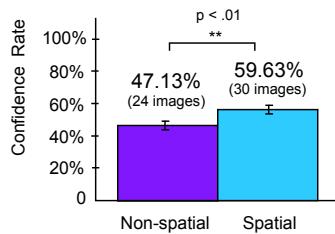


図10. ラベリングタスクの信頼率. Non-spatial: mean = 47.13; SE = 4.52; Spatial: mean = 59.63; SE = 4.12; p = 0.0008.

6 今後の課題

本研究では、専門家でない画像注釈のラベル改善の本当の理由は明確には示されなかった。改善の理由が自己学習によるものなのか、注釈付け中に空間レイアウトインターフェイスを介して慎重にラベル付けするためなのかは不明である。将来的には、ラベルの改善の本当の理由（学習または注意深いラベル付け）を調査したいと考えている。

7 結論

本研究では、専門家以外の画像注釈のラベル品質を向上させるために、空間ラベリングと呼ばれる空間レイアウトラベリングインターフェイスを提案した。このインターフェースは、ラベル付けプロセスでラベルを選択する前に画像とラベルをレイアウトするためのオープンスペース（つまり、空間編成プロセス）で構成されていた。本研究の空間レイアウトインターフェイスと従来の非空間レイアウトインターフェイスの比較のため、画像ラベリングタスクを用いてユーザー調査を実施した。結果は、非専門家のアノテーターは、従来の非空間レイアウトインターフェイスよりも本研究の空間レイアウトインターフェイスを使用した時に、画像のラベルをより正確に選択することを示した。さらに、空間レイアウトインターフェイスにより、手動の画像ラベリング中に専門家ではないアノテーターの確信度が向上することが観察された。空間ラベリングは、空間編成を介して非専門家の画像注釈のラベル品質を改善す

るための代替ソリューションを提供する。本研究の結果は、注釈ツールの将来の開発に使用できる重要な洞察を示している。.

謝辞

本研究は JST CREST JPMJCR17A1 の支援を受けたものである。

参考文献

- [1] Abhishek Dutta, and Andrew Zisserman. 2019. The VIA Annotation Software for Images, Audio and Video. In Proceedings of the 27th ACM International Conference on Multimedia, pp. 2276-2279.
- [2] Bryan C. Russell, Antonio Torralba, Kevin P. Murphy, and William T. Freeman. 2008. LabelMe: A Database and Web-Based Tool for Image Annotation. International journal of computer vision 77, no. 1-3: 157-173.
- [3] Luis Von Ahn, and Laura Dabbish. 2004. Labeling Images with a Computer Game. In Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 319-326.
- [4] Chia-Ming Chang, Siddharth Deepak Mishra, and Takeo Igarashi. 2019. A Hierarchical Task Assignment for Manual Image Labeling. In 2019 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC), pp. 139-143.
- [5] Joseph Chee Chang, Saleema Amershi, and Ece Kamar. 2017. Revolt: Collaborative Crowdsourcing for Labeling Machine Learning Datasets. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pp. 2334-2346.
- [6] Takeru Sunahase, Yukino Baba, and Hisashi Kashima. 2017. Pairwise HITS: Quality Estimation from Pairwise Comparisons in Creator-Evaluator Crowdsourcing Process. In Thirty-First AAAI Conference on Artificial Intelligence.
- [7] Yi-Li Fang, Hai-Long Sun, Peng-Peng Chen, and Ting Deng. 2017. Improving the Quality of Crowdsourced Image Labeling via Label Similarity. Journal of Computer Science and Technology 32, no. 5: 877-889.
- [8] Todd Kulesza, Saleema Amershi, Rich Caruana, Danyel Fisher, and Denis Charles. 2014. Structured Labeling for Facilitating Concept Evolution in Machine Learning. In Proceedings of the SIGCHI Conference on Human Factors in

- Computing Systems. ACM, 3075–3084.
- [9] Ken Perlin, and David Fox. 1993. Pad: An Alternative Approach to the Computer Interface. In Proceedings of the 20th annual conference on Computer graphics and interactive techniques, pp. 57-64.
- [10] Benjamin B. Bederson, and James D. Hollan. 1994. Pad++: A Zooming Graphical Interface for Exploring Alternate Interface Physics. In Proceedings of the 7th annual ACM symposium on User interface software and technology, pp. 17-26.
- [11] Daniel Bauer, Pierre Fastrez, and Jim Hollan. 2005. Spatial Tools for Managing Personal Information Collections. In Proceedings of the 33th IEEE Annual Hawaii International Conference on System Sciences, pp. 104b-104b.
- [12] Nayuko Watanabe, Motoi Washida, and Takeo Igarashi. 2007. Bubble Clusters: An Interface for Manipulating Spatial Aggregation of Graphical Objects. In Proceedings of the 20th annual ACM symposium on User interface software and technology, pp. 173-182.
- [13] George Robertson, Mary Czerwinski, Kevin Larson, Daniel C. Robbins, David Thiel, and Maarten Van Dantzig. 1998. Data Mountain: Using Spatial Memory for Document Management. In Proceedings of the 11th annual ACM symposium on User interface software and technology, pp. 153-162.
- [14] Nan-Chen Chen, Jina Suh, Johan Verwey, Gonzalo Ramos, Steven Drucker, and Patrice Simard. 2018. AnchorViz: Facilitating Classifier Error Discovery through Interactive Semantic Data Exploration. In 23rd International Conference on Intelligent User Interfaces, pp. 269-280.
- [15] Yanir Kleiman, Joel Lanir, Dov Danon, Yasmin Felberbaum, and Daniel Cohen-Or. 2015. DynamicMaps: Similarity-based Browsing through a Massive Set of Images. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, pp. 995-1004.
- [16] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang et al. 2015. ImageNet Large Scale Visual Recognition Challenge. International journal of computer vision 115, no. 3: 211-252.
- [17] Chia-Ming Chang, Chia-Hsien Lee, and Takeo Igarashi. 2021. Spatial Labeling: Leveraging Spatial Layout for Improving Label Quality in Non-Expert Image Annotation. In Proceedings of the ACM Conference on Human Factors in Computing Systems. ACM.

未来ビジョン

通常、高品質のデータは専門家のアノテーターに依存している。ただし、優れた注釈ツールが提供されれば、専門家でない注釈者も高品質のデータを生成できると考えている。この研究は、ラベリングインターフェースの設計が非専門家のデータ注釈のラベル品質を改善できるという成功例を示している。今日、AIの時代では、新しいインテリジェントシステムには常に膨大な数のトレーニングデータが必要である。専門家ではないアノテーターがアノテーション中にドメイン知識を構築するのに役立つアノテーションツールを提供できれば、それは大きな貢献になる可能性がある。将来的には、難しいアノテーション作業でも誰もがアノテーターになれるよう願っている。

対戦アクションゲームにおけるプレイヤーの挙動観察のためのシーン検索

三ツ井 慧太郎¹ 岡部 誠¹

概要. 我々は対戦型格闘ゲームのプレイヤーが強くなるための支援に取り組んでいる。強いプレイヤーになるための重要なアプローチの1つに、他の強いプレイヤーの挙動を観察し、強い行動パターンを学ぶことがある。そこで我々はYouTubeなどの動画共有サイトに上がっている多くの対戦動画から、強いプレイヤーの挙動を効率よく観察するために、ゲームシーンの検索手法を提案する。提案手法は、対戦中のある瞬間の画像をクエリーとして入力すると、その画像と似通った状況のシーンを対戦動画の中から複数検索して提示する。ユーザは検索結果を再生することで、強いプレイヤーならクエリーの状況からどういった行動をとるのかについて観察できる。このシーン検索を実現するために、対戦ステージや背景を無視し、キャラクターの位置関係のみを考慮した特徴ベクトルを生成できるディープニューラルネットワークを学習する。シーン検索の精度については主観評価を行い、約8割の精度でシーン検索ができるることを確認した。また、ユーザスタディを行い、動画で上級プレイヤーの挙動を観察しながらCPU対戦を繰り返したとき、提案手法を用いることでより高いレベルのCPUに勝てるようになることを確認した。

1 はじめに

近年、競技性の高いビデオゲームをスポーツ競技として捉えて”eSports”と呼ぶほど、ビデオゲームに注目が集まっている。eSportsはファースト・パーソン・シューティング(FPS)や、対戦型格闘ゲーム、サッカーやバスケットボールなどのスポーツゲーム、レーシングゲームなど、多種多様なジャンルがあり、それぞれのジャンルに複数のタイトルがある。eSportsで活躍するためには、プレイヤーたちは各タイトルに特化してゲームを練習し強くなる必要がある。我々は毎年世界大会が開かれているニンテンドースイッチの大乱闘スマッシュブラザーズSPECIAL(スマブラ)[1]を題材にプレイヤーが効率よく強くなるための支援に取り組んでいる。

スマブラのような対戦型格闘ゲームにおいて、強いプレイヤーになるために重要なことが2つある。1つはゲームの操作技術を磨くことである。これはキャラクターの移動操作をミス無くこなしたり、技を出す際のコマンド入力を正確にできるようになるなどだが、これらはプレイヤー自身が時間をかけて練習して習得する必要がある。もう1つの重要なことは対戦相手を意識した練習であることである。eSportsはビデオゲームを媒体にしているが、要は人ととの試合であり、相手をだますようなフェイントを掛けたり、相手のミスを誘発するような行動

を仕掛けたりする駆け引きに面白さがある。対戦相手を意識した練習には、YouTubeなどの動画共有サイトに上がっている対戦動画を見て、強いプレイヤーの挙動を観察し、強い行動パターンを学ぶことが有効である。しかしこれは簡単ではない。動画共有サイト上の動画数が多いため、自分の見たい対戦動画がなかなか見つからなかったり、また、対戦動画が見つかっても、その中のどこに自分の見るべきシーンがあるかが分からぬ、といった問題がある。

そこで我々は、多くの対戦動画から強いプレイヤーの挙動を効率よく観察するために、ゲームシーンの検索手法を提案する。提案手法は、対戦中のある瞬間の画像をクエリーとして入力すると、その画像と似通った状況のシーンを対戦動画の中から複数検索して提示する(図1)。ユーザは検索結果を再生することで、強いプレイヤーならクエリーの状況からどういった行動をとるのかについて観察できる。

この目的を実現するために、既存の画像/動画検索手法について調査したが、我々の要求に対応するものは存在しなかった。既存の検索手法の大きな問題の1つは、我々はゲーム内のキャラクター同士の位置関係のみに興味がある。しかし、既存の画像/動画検索手法は画像全体の見た目が似たものを検索してしまうため、キャラクター同士の位置関係よりもむしろ、戦っているステージや背景が似たものを検索してしまう。そこで我々は入力画像をキャラクター

Copyright is held by the author(s).

1 静岡大学大学院総合科学技術研究科工学専攻
数理システム工学コース

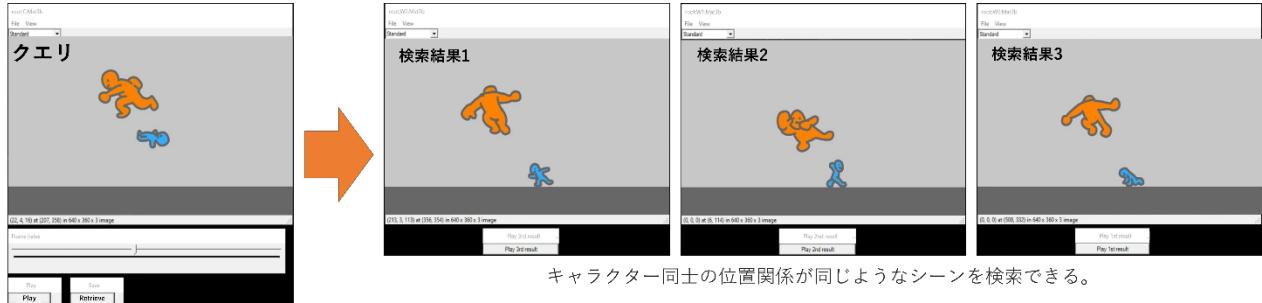


図1. 提案手法のユーザインタフェース。ユーザは自分の対戦のリプレイ動画を見ながら（左のウィンドウ）、気になるシーンで一時停止し、「検索」ボタンを押す。すると、他の対戦動画の中から、類似した3つのシーンが検索される（右の3つのウィンドウ）。

同士の位置関係のみの情報を持った特徴ベクトルに変換できるようなディープニューラルネットワークを学習することで、シーン検索手法を実現する。

シーン検索の精度については主観評価を行い、約8割の精度でシーン検索ができる事を確認した。また、ユーザスタディを行い、動画で上級プレイヤーの拳動を観察しながらCPU対戦を繰り返したとき、提案手法を用いることで平均的に0.714回多く勝てる事が示された。言い方を変えると、より高いレベルのCPUに勝てるようになることが確認できた。

2 関連研究

類似画像/動画の検索手法については数多くの関連研究が存在する[2]。基本的な考え方は、画像を大規模な畠み込みニューラルネットワーク(CNN)によってエンコードし、得られた特徴ベクトルの類似度によってデータベース内の画像にランキング付けをすることである[3]。同一クラス内の画像の特徴ベクトル同士は類似するように、逆に、異なるクラス間の画像の特徴ベクトル同士は異なるようにしたく、そのような特徴ベクトルを出力できるCNNを学習するために、ネットワークの構造やloss関数に様々な工夫が提案されてきている。例として、Revaudらは画像検索学習モデルの評価指標の1つであるmean Average Precision(mAP)を直接最適化できるランキング学習手法を提案した。その結果、画像集合に対し、そこに写った建物毎にラベル付けしたデータセットを学習することで、クエリー画像に写ったものと同じ建物が写っている画像を高い精度で検索できる手法を提案した[4]。こういった画像検索技術の発展形としては、Voらは画像と文をクエリーとし、画像を文によって変化させたような画像を検索できる手法を提案した[5]。

しかし、上記の既存手法のいずれも、我々の目的であるキャラクター同士の位置関係のみに着目したシーン検索という問題にそのまま適用できる手法は存在しなかった。そこで、まず動画のセグメンテー

ション手法を適用し、動画からキャラクターのみを抽出してから上記の検索手法を適用するというアプローチを試みた。動画からキャラクターのみを抽出するために、SiamMask[6]とSTM[7]という2つの手法を実験した。共に動画の最初のフレームで追跡したい物体を指定すると、2フレーム目以降を自動的に追跡し、物体のマスクを生成する手法である。しかし、実際にスマブラのキャラクターを追跡させてみると、最初の数秒は上手く追跡できても、キャラクター同士の衝突やエフェクト（煙や炎など）が起こってキャラクターが一瞬隠れたり、キャラクターのポーズが大きく変化するなどすると、その後のフレームを追跡できなくなったり。特にSTMは動画のセグメンテーションの分野では最も精度の高い手法の1つであるため、STMが上手くキャラクターを抽出できないならば、このアプローチは不可能と判断した。

スマブラに特化した研究として、強化学習によって強いAIプレイヤーを育てる手法がある[7]。こういったAIプレイヤーとの対戦も、人がスマブラに強くなるための1つのアプローチかも知れない。

3 ユーザインタフェース

図1に提案手法のユーザインタフェースを示す。図1-左のウィンドウでは、ユーザは自分の対戦のリプレイ動画を見ているとする。そして、図1-左のウィンドウに表示されているようなシーンで「こんなふうにマリオの頭上にドンキーコングが浮いている状況のとき、強いプレイヤーならどういった行動をとるのだろう？」と気になったなら、ここでリプレイを一時停止し、「検索」ボタンを押す。すると、強いプレイヤーの対戦動画の中から、類似した3つのシーンが検索されて表示される（図1-右の3つのウィンドウ）。検索結果はいずれもマリオの頭上にドンキーコングが浮いている状況である。ユーザはこれらの動画を再生することで、強いプレイヤーがここからどういった行動をとるのかを観察できる。

「検索」ボタンを押してから結果が表示されるま

対戦アクションゲームにおけるプレイヤーの挙動観察のためのシーン検索

での時間は、現在の我々の実装では 10 秒未満を要する。また、表示される検索結果の数は、現在は類似性の高いものから順に 4 つまで、としている。

尚、検索対象となる強いプレイヤーの動画は、ユーザが予め収集しておくものとする。動画の収集方法としては、①「世界戦闘力」という任天堂が公式に定めるプレイヤーの強さの指標に基づいて集める、②ゲーム大会におけるトーナメント上位の対戦動画を集める、③実績のある有名なプレイヤーの動画を集め、などが考えられる。

4 シーン検索手法

関数 f は画像 x を入力すると特徴ベクトル $y = f(x)$ を出力するニューラルネットワークであるとする。入力画像 x_i が図 2-a のような画像であって $y_i = f(x_i)$ であるとする。このとき、背景は異なるがキャラクター同士の位置関係が図 2-a と同じような画像集合（図 2-b、正例と呼ばれる）の中の画像 1 つを x_p とすれば、その特徴ベクトル $y_p = f(x_p)$ は y_i と似たものであってほしい。逆に、キャラクター同士の位置関係が図 2-a と異なるような画像集合（図 2-c、負例と呼ばれる）の中の画像 1 つを x_n とすれば、その特徴ベクトル $y_n = f(x_n)$ は y_i と異なったものあってほしい。

このような関数 f を実現できれば、特徴ベクトル y の類似度を使うことで、ステージや背景を無視しキャラクター同士の位置関係のみに着目したシーン検索が実現できる。類似度には内積を用い、検索対象の全対戦動画の各フレームに対し、特徴ベクトルの類似度に基づいたランキング付けをすることでシーン検索ができる。

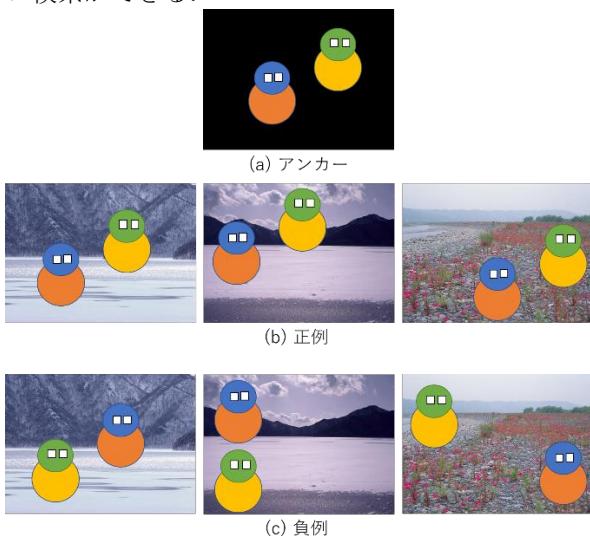


図 2. Supervised Contrastive Learning におけるアンカー画像(a)とそれに対する正例(b)と負例(c)。

関数 f に ResNet-50[9] を用いた。画像 x のサイズは

256×144 ピクセル $\times 3$ チャンネルであり、特徴ベクトル y のサイズは 2048 次元である。上記のような関数 f を実現するために Supervised Contrastive Learning[10] を用いる。損失関数は

$$\sum_{i \in I} \frac{-1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp(y_i \cdot y_p / \tau)}{\sum_{a \in A(i)} \exp(y_i \cdot y_a / \tau)}$$

である。 I は全ての画像の添え字の集合、 $P(i)$ は正例の画像の添え字の集合、 $A(i)$ はアンカー画像以外の全ての画像の添え字の集合、 τ は温度と呼ばれるハイパーパラメータである。 $a \cdot b$ はベクトル a とベクトル b の内積である。上記の損失関数を最小化することによって、 y_i と y_p の内積が y_i と y_n の内積よりも大きくなるよう関数 f を学習することができる。この関数 f によってステージや背景を無視しキャラクター同士の位置関係のみの情報を持った特徴ベクトルを得ることができる。

4.1 学習

図 2 に示したように、関数 f の学習にはアンカー画像に対する正例集合と負例集合が大量に必要となる。動画中のあるフレームをアンカーに選んだとき、それ以外のフレームは全て負例となり得る（キャラクターが移動しており位置関係がアンカーと異なるため）ので、負例集合はアンカー以外のフレームを選ぶことで簡単に用意できる。しかし、正例はキャラクターの位置関係がアンカーと同じで、かつ、背景が異なる画像であるため、通常は簡単に用意できない。そこで我々はスマブラの特徴を生かした正例の大量生成手法を以下に提案する。



図 3. ステージ「75m」を利用した正例の大量生成

データセットの作成にあたり、スマブラの対戦動画から 2000 フレームを抽出した。この動画は「75m」というステージで対戦されたものだが、このステージに関しては背景に現れるピクセルの色数が少ない（図 3-アンカー画像）。そのため容易に背景部分

のピクセルを特定することができる、即ち、キャラクター部分のマスクを作ることができる（図 3-マスク）。このマスクを用いて背景を別の画像に入れ替えることで、アンカー画像の正例を大量生成することができる（図 3-下段）。

このようにして背景を入れ替えたバージョンを 20 種類作ったため、合計で $2000 \times 21 = 42000$ 枚の画像からなるデータセットとなった。

学習は 2000 フレームから 5 フレームをランダムに抽出し、それらの背景を入れ替えたバージョン（合計で $5 \times 21 = 105$ 枚）を 1 つのバッチとして学習を行う。即ち、このバッチの中であるアンカー画像を 1 枚選んだ時、そのアンカー画像に対する正例がこのバッチには 20 枚、負例がこのバッチには 84 枚存在することになる。このような学習を 400 回繰り返すことで 2000 フレーム全てに対する学習を行ったとし、これを 1 エポックとする。400 回の損失関数の値の平均の推移（図 4）を見ると、学習が上手く進行していることがわかる。

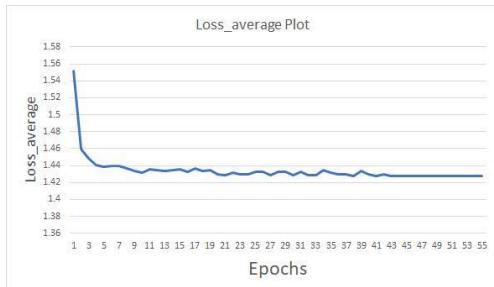


図 4. 損失関数の値の変化。

4.2 シーン検索の精度評価

ステージや背景を無視しキャラクター同士の位置関係のみを考慮した検索が上手くできているかを調べるために主観評価を行った。今回の実験では 9 人の被験者（工学系の大学または大学院に所属する 20 代の男性）に協力してもらった。

各被験者には、クエリー画像 1 枚と、そのクエリー画像を入力として提案手法で検索を行って得られた上位 4 枚の検索結果の画像を提示する。クエリー画像と検索結果を見比べてもらい、以下のレベル 1 ~ 3 の質問に対し、Yes/No で答えてもらった。

- レベル 1 : 4 枚の検索結果の中に、クエリー画像とキャラクターの位置関係が似ているものが 1 枚でもあるかどうか
- レベル 2 : 4 枚の検索結果の中に、レベル 1 に加え、エフェクト（煙や炎など）までもが似ているものが 1 枚でもあるかどうか
- レベル 3 : 4 枚の検索結果の中に、レベル 1 & 2 に加え、キャラクターのポーズまでもが似ているものが 1 枚でもあるかどうか

各被験者には上記の評価を 97 枚のクエリー画像に対して行ってもらった。

97 枚のクエリー画像には、青のマリオと白のドンキーコングの対戦動画からランダムに選出したフレームを用いた。97 枚中の 49 枚は「75m」というステージでの対戦動画から選出されたフレームとなり、残りの 48 枚は「75m」以外のステージでの対戦動画から選出されたフレームとなった。検索対象は全て「75m」での対戦動画である。図 5 に主観評価の結果を示す。

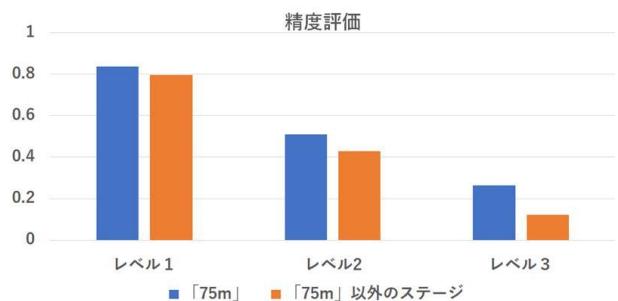


図 5. 主観評価の結果

図 5 を見るとレベル 1 の質問においては「75m」と「75m」以外のステージの両方で 8 割の精度を達成できており、つまり、提案手法は上位 4 件の中にキャラクターの位置関係が似ているシーンを約 8 割の精度で検索できることが分かる。また、「75m」と「75m」以外のステージの間でほとんど差が無いことから、ステージや背景を無視した検索に成功していることが分かる。レベル 2 の質問の結果からは、4 ~ 5 割の精度でエフェクトも似ているシーンが検索できていることが分かる。レベル 3 の質問の結果からは、キャラクターのポーズの類似性までを考慮した検索については精度が低いことが分かる。また、クエリー画像について、背景が単純な「75m」と背景が複雑な「75m」以外のステージの間で精度に 2 倍程度の開きがあることから、キャラクターのポーズの特徴を捉える際に背景の複雑さがノイズとなって影響を与えていることが分かる。

5 ユーザスタディ

提案手法を用いて強いプレイヤーの挙動を観察したとき、それがプレイヤーの成長にどう影響を与えるかを調査するため、ユーザスタディを行った。

5.1 実験方法

今回の実験では 14 人の被験者（大学または大学院に所属する 20 代の男性 13 名と女性 1 名）に協力してもらった。14 人のうち 7 人は提案手法を使って強いプレイヤーの挙動を観察するグループ、残りの

対戦アクションゲームにおけるプレイヤーの挙動観察のためのシーン検索

7人は提案手法を使わずに強いプレイヤーの挙動を観察するグループとした。

各被験者にはスマブラのCPU戦をプレイしてもらつた。プレイするキャラクターは青のマリオとし、敵であるCPUのキャラクターは白のドンキーコングとした。CPUのレベルは比較的弱い3から対戦を始め、勝った場合はレベルを1つ上げ、次のレベルのCPUと対戦してもらう。負けた場合は、負けたレベルのCPUと強いプレイヤー（任天堂公式の「世界戦闘力」の値がランキング上位5~6%に入つており、VIPの称号を持つ）が対戦している動画が予め用意してあるので、それを見て強いプレイヤーがどうやってそのCPUを倒しているかを観察してもらい、その後、再び同レベルのCPUと対戦してもらった。以上を負けた回数が5回になるまで繰り返してもらい、最終的にどのレベルのCPUが倒せるかを調べた。尚、CPUのレベルの上限は9であり、レベル9に勝った場合はそれ以上レベルを上げることができないため、再びレベル9と対戦してもらうこととした。

強いプレイヤーの対戦動画を見るとき、提案手法を使うグループの被験者は、まず自分の負けた戦いのリプレイ動画を見る。任意のフレームで一時停止し、「検索」ボタンを押すことで、強いプレイヤーの対戦動画から4つの類似シーンが検索されるので、各シーンを再生してもらい、強いプレイヤーの挙動を観察してもらった。強いプレイヤーの対戦動画を見るとき、提案手法を使わないグループの被験者は、動画再生ソフト(Windows Media Player)で強いプレイヤーの対戦動画を自由に見るだけとした。どちらのグループも対戦動画を見て挙動観察をする時間は5分間とした。

各被験者の対戦は全て動画で記録するとともに、勝敗やゲーム内のスコアなども記録した。

5.2 実験結果

提案手法を使ったグループと使わなかつたグループの対戦結果を図6と図7に示す。

	A	B	C	D	E	F	G	合計	平均
当初のレベル	3	6	8	5	8	4	9	43	6.142857
最後に負けたレベル	8	9	9	7	9	9	9	60	8.571429
勝った数	5	3	1	2	1	5	1	18	2.571429

図6. 対戦結果：提案手法を使ったグループ

上図において、提案手法を使ったグループのメンバーはAさん～Gさんとし、提案手法を使わなかつたグループのメンバーはHさん～Nさんとする。

「当初のレベル」とは、各被験者が最初に負けたCPUのレベルのことである。その平均値を見てみると、提案手法を使ったグループでは6.14、提案手法を使わなかつたグループでは6.29となっており、ほぼ同じ値となっている。つまり、片方のグループに初めから強い人または弱い人が集まつた、ということはなかつたことを示している。

	H	I	J	K	L	M	N	合計	平均
当初のレベル	6	8	5	7	9	6	3	44	6.285714
最後に負けたレベル	9	9	9	9	9	6	4	55	7.857143
勝った数	3	1	5	2	1	0	1	13	1.857143

図7. 対戦結果：提案手法を使わなかつたグループ

「勝った数」とは、「当初のレベル」が決まつた後（つまり、CPUに1回負けた後）、実験が終わるまでの間にCPUに勝つた数を示している。即ち、ほとんどの被験者で「勝った数」は「最後に負けたレベル」から「当初のレベル」を引いた値になつてゐる。ただし、被験者GとLについては、彼らが強いプレイヤーであつて「当初のレベル」が上限の9であつたため、異なる値くなつてゐる。

「勝った数」を比較すると、提案手法を使ったグループの平均値は2.571、使わなかつたグループの平均値は1.857となつており差がついた。これは、提案手法を使った場合の方が、使わなかつた場合より、平均的に $2.571 - 1.857 = 0.714$ 回多く勝つてゐることを示してゐる。言い方をえれば、0.714分の高いCPUレベルに成長できたことを示してゐる。

以上より、グループの間に当初の能力の差はなかつたが、提案手法を使った場合では勝つた数が多かつたことより、提案手法の有用性が示せたといえる。

5.3 アンケート結果

上記の実験終了後に、各被験者に下記の5つの質問に5段階評価で答えてもらつた（図8）。

「提案手法を使って、見たいシーンが検索できましたか？」という項目について、提案手法を使ったグループのみに答えてもらつたが、6割程度の賛同が得られた。シーン検索の精度評価（4.2章）においては約8割の精度で検索ができるという結果が得られていたにも関わらず、「見たいシーン」の検索精度となると6割程度の精度になつてしまつたということになる。この理由としては、キャラクターの位置関係が似ていても、再生してみると、その後の動画の展開が「見たいシーン」のものではなかつた、という場合があるためである。今後は動画の展開も考慮した検索手法を模索していきたい。

「スマブラは面白いと感じましたか」という質問

について、提案手法を使ったグループの方がやや高い評価を得た。提案手法を使うことで、被験者にはやや複雑な検索操作を強いたが、それによってスマブラを面白いと感じなくなる、ということは無いことが分かった。

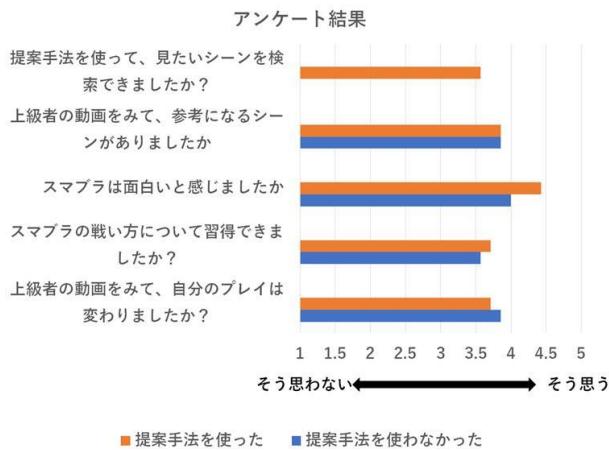


図 8. アンケート結果。棒グラフは平均点を示す。

提案手法を使ったグループに「システムの使いやすかった点、使いにくかった点、改善点」という項目で、文章を書いてもらうアンケートも実施した。

「自分が見たい展開と検索されたシーンが異なっていたので展開まで指定したいと感じた。」「自分のみたいシーンが出てこないときがあった。」という意見から、画像レベルでの検索はできいても、ユーザの求める動画が出力できていないことを示すフィードバックが得られた。他には「どこで一時停止して『検索』ボタンを押せば参考になる動画を検索できるのかがわからないので教えてほしい。」や「動画で強いプレイヤーが入力しているコマンドを知りたい。」というような、より高度なサポートをシステムに希望するような指摘も頂いた。このような問題にも対応できるよう、今後も開発を進めたい。

6 今後の課題

今回のアンケートを受けて、初心者のプレイヤーに対しては操作方法の補助が有効であるように思われた。例えば、強いプレイヤーの動画に合わせて使っているコマンドを推定して表示するなどができる、便利ではないかと思われる。また、対戦動画を見て挙動観察をするといつても、見る人のレベルによって、何を見たらいいのかがそもそも分からず、という問題があることが分かった。システム側からもっと積極的に「このシーンは真似する価値があるよ」というような提案ができるようになると面白いのではないかと思う。一方で、eSportsに参加しているような上級プレイヤーに提案手法を使ってもらら

うようなユーザスタディも今後はやっていきたい。

今後は扱える検索対象を増やしていく。今回の実験は青のマリオと白のドンキーコングのみを対象に行ったが、スマブラでは他にも多種多様のキャラクターが存在し、更にそのカラーとスキンも変更可能である。それらも検索対象としたいため、より多くのデータセットを用意し、ニューラルネットワークを学習し直したい。検索対象を増やすと精度が低下するかどうかを調査したり、低下した場合にはニューラルネットワークのサイズを大きくしたり、状況毎に異なるニューラルネットワークを構築したりすることで対応していく。

参考文献

- [1] 大乱闘スマッシュブラザーズ SPECIAL, https://www.smashbros.com/ja_JP/
- [2] Shiv Ram Dubey, “A Decade Survey of Content Based Image Retrieval using Deep Learning”, IEEE Transactions on Circuits and Systems for Video Technology.
- [3] A. Babenko, A. Slesarev, A. Chigorin, and V. Lempitsky, “Neural codes for image retrieval,” in ECCV, 2014, pp. 584–599.
- [4] Jerome Revaud, Jon Almazán, Rafael Sampaio de Rezende, César Roberto de Souza, “Learning with Average Precision: Training Image Retrieval with a Listwise Loss”, ICCV 2019.
- [5] Nam Vo, Lu Jiang, Chen Sun, Kevin Murphy, Li-Jia Li, Li Fei-Fei, James Hays, “Composing Text and Image for Image Retrieval -An Empirical Odyssey”, CVPR 2019.
- [6] Q. Wang, L. Zhang, L. Bertinetto, W. Hu, P. H. Torr, “Fastonline object tracking and segmentation: A unifying approach”, CVPR 2019.
- [7] S. Wug Oh, J.-Y. Lee, N. Xu, and S. Joo Kim, “Video objectsegmentation using space-time memory networks”, ICCV 2019.
- [8] Vlad Firoiu, William F. Whitney, Joshua B. Tenenbaum, “Beating the World’s Best at Super Smash Bros. with Deep Reinforcement Learning”, arXiv:1702.06230 [cs.LG].
- [9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. “Deep residual learning for image recognition”, CVPR 2016.
- [10] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, Dilip Krishnan, “Supervised Contrastive Learning”, arXiv:2004.11362 [cs.LG].

Corridor-Walker: 視覚障害者が障害物を回避し交差点を認識するためのスマートフォン型屋内歩行支援システム

栗林 雅希^{*} 粥川 青汰^{*†} Jayakorn Vongkulbhaisal[‡] 浅川 智恵子[‡] 佐藤 大介[§]
高木 啓伸[‡] 森島 繁生[¶]

概要. 本研究では屋内の通路を歩く視覚障害者がスマートフォン1台を用いて障害物を回避し、交差点を認識するためのシステム、Corridor-Walkerを提案する。屋内の廊下を歩く視覚障害者は、自らの手や白杖の触覚だけを頼りに箱や椅子などの障害物を安全に回避し、曲がるべき交差点を発見する必要がある。提案システムはスマートフォンのLiDARセンサを用いた周囲の床領域の検出結果に基づいて地図を作成し、障害物を回避する経路の生成および交差点の位置と形状の検出を行う。提案システムは骨伝導イヤホンの左右からのビープ音、音声、振動シグナルを用いてユーザーへ生成された経路に沿って歩くように指示し、検出された交差点の位置と形状を伝える。14名の視覚障害者を対象に評価実験を行った結果、提案システムを用いることでユーザーは壁伝いに歩かずに障害物を回避し、目的の交差点でその形状を認識したうえで曲がることができた。

1 はじめに

視覚障害者が病院やホテルなどの建物内の通路を歩く際に直面する困難として、障害物との衝突回避と交差点の把握の2つがある。視覚障害者が屋内の通路を壁伝いに歩こうとすると、消化器やゴミ箱などの壁沿いに配置された障害物と衝突する危険性がある[13]。そのため、視覚障害者は白杖の感覚を頼りに障害物を検出し、回避するための経路を自身で探索する必要がある[17]。また、視覚障害者は壁伝いに歩くことで通路の交差点を発見できるが、壁沿いの障害物を回避しなければいけない状況では曲がるべき交差点を通り過ぎてしまう可能性がある[8]。

そこで本研究は、屋内の通路を歩く視覚障害者に障害物を回避する経路と交差点の位置および形状を案内するスマートフォン型支援システム(*Corridor-Walker*)を提案する。壁沿いに配置された障害物を回避するため、システムは周囲の壁や障害物と一定の距離を確保する経路を生成し、ユーザーが経路からずれること(ベアリング[veering]と呼ぶ)が無いようにその経路に沿って案内する。前方に障害物が現れた場合はその障害物を迂回するような経路を生成する。また、システムはユーザーが交差点を通り過ぎないように交差点の位置を事前に伝え、曲がるべき交差点かを確認するためにその形状も伝える。システムはiPhone 12 ProのLiDARセンサを用いた周囲の床領域の検出と占有格子地図の作成、および

A*経路計画アルゴリズム[10]を用いた障害物を回避する経路の生成をリアルタイムに行う。同時にシステムは占有格子地図の画像と物体検出手法[23]を組み合わせることで交差点の位置と形状を検出する。

本研究では14名の全盲の視覚障害者を対象にシステムの評価実験を行った。実験では交差点を曲がるタスク、複数の障害物が配置された真っすぐな通路を歩行するタスク、複数の交差点と障害物が存在する通路を歩行するタスクの3種類のタスクを行い、それをおいて、システムと白杖を併用した時と白杖のみを使用した時でタスク完了時間、交差点形状の正答率、白杖と壁や障害物との接触回数を比較した。実験から、システムを併用することで白杖のみを使った時と比較して、1) タスク完了時間が伸びたもののユーザーはスムーズに障害物回避と歩く方向の修正ができたと評価したこと、2) 壁や障害物と白杖の接触回数が減少したこと、3) ユーザは交差点の形状をより正確に把握できることを確認した。

2 関連研究

視覚障害者向け屋内歩行支援システムとして、事前に地図の準備が必要なもの[9, 15, 18, 25]と必要としないもの[4, 14, 21]が存在する。地図を用いるシステムでは、地図情報とユーザーの自己位置推定結果を基に交差点の情報の案内[18, 25]、また地図情報とセンサを用いたリアルタイムセンシング結果を基に障害物を回避する経路を生成、案内[9, 15, 18]が可能となる反面、多くの建物では地図が用意されていない問題点がある[26]。またこれまで提案されてきた地図を必要としない歩行支援システム[4, 14, 21]は障害物回避を主目的としており、交差点の認識を

* 早稲田大学

† 日本科学未来館

‡ IBM Research

§ Carnegie Mellon University

¶ 早稲田大学理工学術院総合研究所

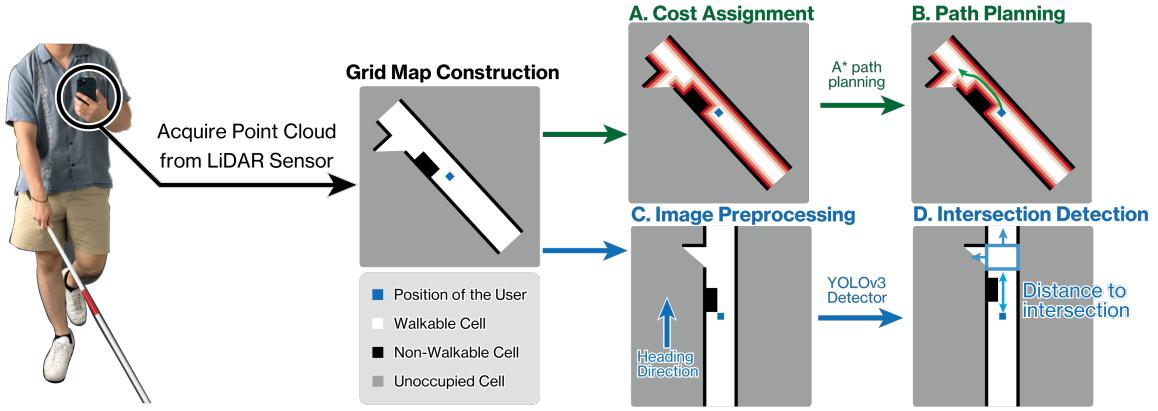


図 1. システムの実装の概要図.

支援するシステムは提案されていない。そこで本研究では、事前に地図の準備を必要としない、障害物回避と交差点検出の実現を目指す。

障害物の回避を目的として、障害物の位置や種類などをユーザへ伝えるシステム [2, 22] が提案されてきたが、これらはシステムからの指示を基にユーザ自身が障害物を回避する経路を設計する必要があり、ユーザ側の負担が大きくなる。そのため、障害物を回避する経路をシステムが生成しユーザに提示するシステムが研究されている [9, 15, 17, 18, 21]。しかしながらこれらのシステムはウェラブルデバイス [17, 18, 21, 28], スーツケース型デバイス [15], ロボット [9]などを使用する必要があり、システムを普及させる面で課題がある [12]。そこで本研究は、多くの視覚障害者が使用しているスマートフォン [19] 1 台のみを用いて、障害物を回避する経路に沿ってユーザを案内するシステムを提案する。

屋内の交差点検出手法としてロボティクス分野ではレーザセンサを用いた手法 [16, 27] が提案されているが、これらの手法は周囲 180 度をスキャンするレーザセンサを必要とするため、スマートフォンに搭載された画角の狭い LiDAR センサから交差点を検出することは難しい。また、ロボットに固定された RGB カメラを使用する手法も提案されているが [1, 6, 7]、歩行中の視覚障害者が撮影する画像はモーションブラーがかかっている可能性があるため [24]、交差点を撮影した画像から直接交差点を検出することは難しい。そこで提案システムは、スマートフォンの LiDAR センサを用いて周囲の地図を作成し、その地図に対して交差点検出を行う。

3 Corridor-Walker の実装

本研究では屋内の通路において視覚障害者が障害物を回避し交差点を認識するためのシステムを提案する。システムは、1) 病院やホテルのような道幅が 1.5m から 3.5m 程度の通路、2) ユーザは目的地に行くために交差点を曲がる回数が分かっている、と

いう状況での支援を想定している。

3.1 占有格子地図と経路計画

本研究では、ロボティクス分野で広く使われているコストマップを用いた経路計画アルゴリズム [20] を参考に、スマートフォンの LiDAR センサデータから障害物を回避する経路を生成するシステムを実装した。システムはまず、スマートフォンの LiDAR センサから三次元点群を取得し、各点の法線ベクトルを計算する [11]。次に、平面検出アルゴリズム (RANSAC) [5] を用いて床面の高さを計算し、取得した点群の各点において、高さが床面から閾値 ε m 以内かつその法線ベクトルが重力加速度ベクトルと平行な点を全て “walkable”，そうでないものを “non-walkable” として分類する。一辺 ω m の占有格子地図上において各点がどの格子（ピクセル）に属するかを計算し、各点の分類結果に基づいて格子を walkable cell (図 1 の白いピクセル)， non-walkable cell (黒) あるいは未占有 (灰色) に分類する。

次に、システムは周囲の壁や障害物から一定の距離を確保した経路を生成するために、占有格子地図と同じサイズで各ピクセルにコストを割り当てるコストマップを作成し (図 1-A) [20]、コスト最小の経路を計画する (図 1-B)。各 walkable cell i に割り当てるコスト $cost_i$ は以下のように定義する。

$$cost_i = \begin{cases} \beta(1 - \frac{\delta_{i-1}}{\alpha}), & 1 \leq \delta_i \leq \alpha, \\ 0, & \delta_i > \alpha. \end{cases} \quad (1)$$

δ_i は walkable cell から最も近い non-walkable cell とのピクセル距離、 α は walkable cell が正のコストを持つ上限のピクセル距離、 β は割り当てるコストの最大値である。壁および障害物の判定精度やスマートフォン上で実行速度を考慮し $\varepsilon = 0.1$, $\omega = 0.15$, $\alpha = 3$, $\beta = 50$ として実装した。

システムはコストマップを基に、スマートフォンが向いている方向 3.5m 先の通路の中央を目的地と

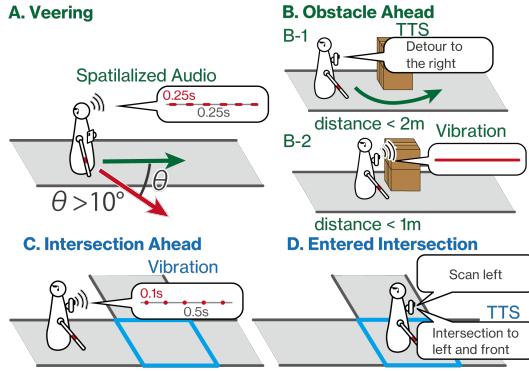


図 2. システムのインターフェースの概要図。

して障害物を避ける経路を A*経路計画アルゴリズム [10] を用いて生成する。ユーザが歩行中、システムは常に前方に障害物があるかを判断する。具体的にはまず、作成した占有格子地図の non-walkable cell において壁に属している cell を RANSAC アルゴリズム [5] を用いて検出する。システムの前方、扇形（中心角 30°、半径 2m）の範囲に、壁に属さなかった non-walkable cell の個数が 30%以上含まれる場合、前方に障害物があると判断する。システムは障害物を検出した場合もしくはユーザが経路の半分を歩いた場合に新しい経路を計画しなおす。

3.2 交差点検出

システムは占有格子地図の画像を入力として物体検出手法 (YOLOv3[23]) を用いて交差点の検出を行う。占有格子地図は交差点を検出する際に必要な前後左右の情報を持たない。そのため、前処理として現在から 2 秒前までのユーザの平均位置と 2 秒前から 4 秒前までのユーザの平均位置を結んだベクトルをユーザの進行方向として計算し、その方向と画像の上方向が一致するように占有格子地図を回転させ、ユーザの位置が画像中央に位置するように画像変換を施す（図 1-C）。本研究では交差点検出モデルを学習させるために、前処理された交差点の占有格子地図の画像を集め、アノテーションおよびモデルの学習を行った。アノテーション時は交差点領域を示す矩形（図 1-D の青い四角形）および、交差点から伸びる通路の方向を全て列挙したラベルを付与した。例えば図 1-D に示した交差点の場合は「Left, Front, Back」のラベルを与えた。ラベルの組み合わせは全部で 9 通りである。またシステムでは、矩形領域の底辺とユーザの距離をユーザから交差点までの距離と定義する（図 1-D）。また、矩形領域がユーザの位置を内包する場合、ユーザは交差点内部にいると判断する。学習させた交差点検出モデルを 1215 枚のデータを用いて評価したところ、ユーザから 2.25m 先にある交差点は F 値 0.95, 3.25m 先にある交差点は F 値 0.63 の精度で検出できた。

今回学習させた交差点検出モデルは、屋内の通路にある壁の窪みを交差点と誤認識してしまう可能性があるため、システムはユーザが交差点に入った際に検出された交差点が本当に交差点かを確認する。具体的には占有格子地図において、検出された交差点から右、左または前方へ 1.5m 以上離れた場所に walkable cell が存在している場合、提案手法はその方向に通路が伸びていると確定する。

3.3 インタフェース

システムは骨伝導イヤホンのビープ音および音声とスマートフォンの振動を用いて生成された経路に沿って歩くようにユーザを案内する。ユーザがペアリングしている間はユーザが向くべき方向から 0.25 秒間隔のビープ音 (400Hz のサイン波) が鳴り、ユーザが正しい方向を向くとビープ音が鳴り止む。例えば、ユーザが進むべき方向よりも右に向いている場合には、左側の骨伝導イヤホンからビープ音を鳴らすことで、ユーザへ左を向くように促す（図 2-A）。ペアリングの判定基準はスマートフォンの向きが目的方向よりも 10° 以上ずれたこととした。

ユーザの前方 2m 以内に障害物が検出された場合は迂回すべき方向を指示する。例えば、ユーザの進行方向左側に障害物が検出された場合には「右迂回開始」と指示する（図 2-B-1）。ユーザ前方 1m 以内に壁を含む障害物が検出された場合は、緊急停止シグナルとしてスマートフォンが強く長く振動する（図 2-B-2）。

前方に交差点が検出された場合は、スマートフォンが周期的に振動することによって交差点が近いことを伝える（図 2-C）。ユーザが交差点に入った際は、交差点の曲がりうる方向をスキャンするように音声を用いて指示する。スキャンの結果実際に通路が伸びていることをシステムが確認できた場合、システムは交差点の形状を音声を用いて伝える（図 2-D）。

4 評価実験

4.1 実験方法

Corridor-Walker の有用性を検証するために 14 名の全盲の視覚障害者（年齢：21–71（平均 43.9）歳、男性 5 名、女性 9 名）を対象に評価実験を行った。最初に参加者へシステムの使い方の説明および練習を 30 分行った後、本番タスクで白杖のみを使った場合と白杖とシステムを併用した場合の 2 つのコンディションを比較した。システムの交差点検出機能と障害物回避機能を評価するために本番タスクでは、1) 交差点（図 3 の 4 種類の交差点）を事前に指示された方向へ曲がるタスク、2) 複数の障害物が配置された直線の通路（図 4 のルート 2-1 とルート 2-2）において歩行するタスク、3) 複数の障害物および交差点を含んだ 40m 程のルート（図 5 のルート 5-1 とルート 5-2）において歩行するタスクを行った。

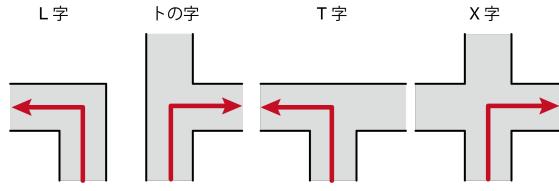


図 3. タスク 1 の交差点.

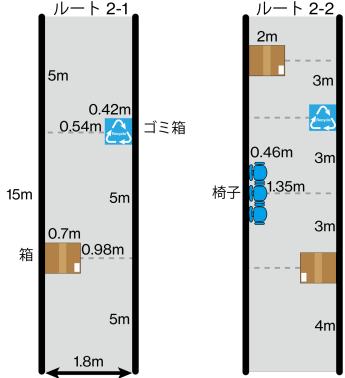


図 4. タスク 2 の通路と障害物の配置.

ト 3-1 とルート 3-2) を歩行するタスク、の 3 種類のタスクを行った。最初の 7 名の参加者 (P01–07) は図 3–図 5 に示した交差点および通路を白杖のみで歩行し、図 3–図 5 を左右反転した交差点および通路をシステムを併用して歩行した。残りの 7 名の参加者 (P08–14) は図 3–図 5 の交差点および通路をシステムを併用して歩行し、図 3–図 5 を左右反転した交差点および通路を白杖のみで歩行した。

4.2 評価指標

全てのタスクでタスク完了時間を測定した。タスク 1 では交差点の 4m 手前から 1m 後の区間を歩く時間を測定し、タスク 2 とタスク 3 ではルートの開始地点から終了地点の間を歩く時間を測定した。参加者が交差点の形状を把握したうえで曲がれたかを評価するために、タスク 1 では参加者に交差点を曲がった後にその交差点の形状を答えてもらい、その正答率を計算した。また、全てのタスクにおいて障害物を回避し、壁伝いに歩かずに歩行できているかを評価するために白杖と壁および障害物との接触回数を数え、システムを評価した。本番終了後、システムと白杖を比較するために参加者はそれぞれに関して 7 段階 (1: 全く同意しない, 7: 強く同意する) の主観評価項目 (図 6, Q1–6) に回答した。

5 結果

5.1 タスク完了時間

タスク 1 で各交差点を歩くのにかかった時間は、白杖のみで平均 9.3 秒、システム併用では 15.4 秒で

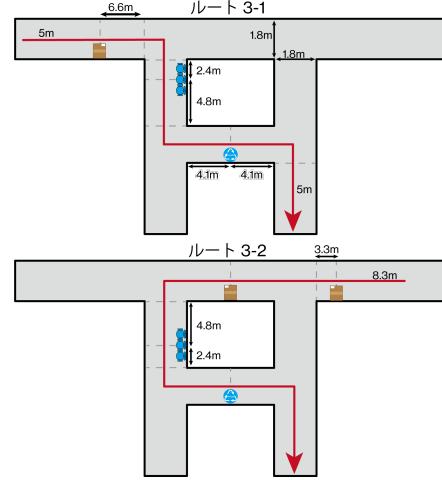


図 5. タスク 3 の通路と障害物の配置.

あった。タスク 2 のルート 2-1 では、白杖のみで平均 18.6 秒、システム併用で 23.5 秒、ルート 2-2 では、白杖のみで 20.9 秒、システム併用で 28.5 秒かかった。タスク 3 のルート 3-1 では、白杖のみで平均 50.7 秒、システム併用で 69.3 秒、ルート 3-2 では、白杖のみで 63.0 秒、システム併用で 85.7 秒かかった。Mann-Whitney の U 検定 (有意水準 1%) を用いて白杖のみの場合とシステムを併用した場合を比較したところ、全てのタスクにおいてシステムを併用した方が有意に時間がかかるという結果が得られた。これはシステム併用時にユーザがシステムの指示の確認や自身の歩く方向の修正に時間がかかることが原因として考えられる。

5.2 交差点の形状の正答率

交差点の形状の正答率は、白杖のみを使用した時は 30.4%，システムを併用した時は 83.9% だった。システムを併用した際に誤回答した回数は合計 9 回 (L 字の交差点で 1 回、T 字で 1 回、X 字で 7 回) であり、理由は以下の 3 つに起因する：1) システムは交差点の形状を正しく伝えたが、ユーザがシステムのフィードバックに集中していなかったため、誤った形状を答えた (5 回)。2) ユーザがスマートフォンを安定して持つことが出来ず、占有格子地図の構築が不十分であったため、システムが交差点の形状を誤認識した (1 回)。3) システムは X 字の交差点を正しく検出し、左右両方をスキャンするように指示したが、ユーザは曲がるべき方向のみをスキャンしたため、システムがトの字の交差点と誤認識した (3 回)。また、白杖のみを用いた時に交差点を気付かずに通り過ぎてしまう現象が 3 回見られた。

5.3 白杖と壁および障害物との接触回数

表 1 に各タスクにおける白杖と壁および障害物との接触回数を示した。Mann-Whitney の U 検定を

Corridor-Walker: 視覚障害者が障害物を回避し交差点を認識するためのスマートフォン型屋内歩行支援システム

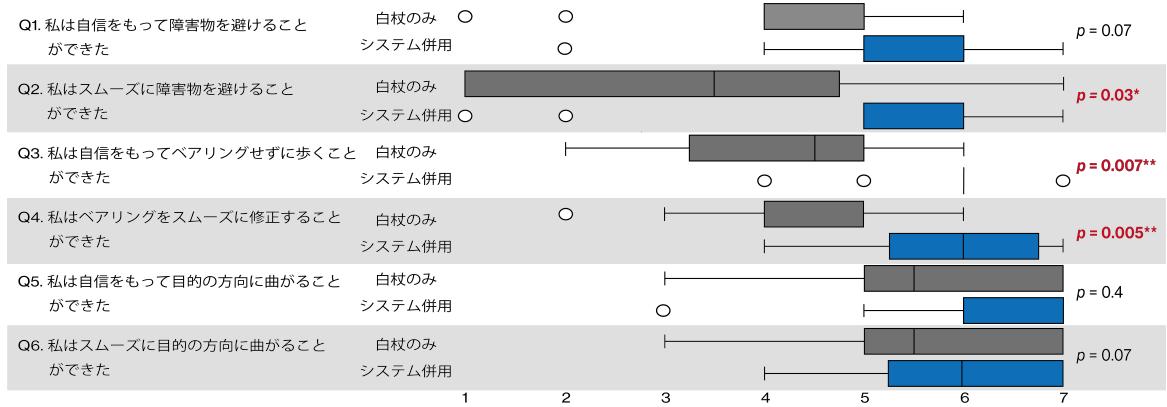


図 6. 白杖のみを用いた場合とシステムを併用した場合を比較した質問項目. *は有意水準 5%, **は有意水準 1%で認められた有意差を示す.

表 1. 白杖と壁および障害物との接触回数の詳細. *は有意水準 5%, **は有意水準 1%で認められた有意差を示す.

タスク	コンディション	接触対象	接触回数 (白杖のみ)		接触回数 (システム併用)		p 値
			平均 ± 標準偏差	中央値	平均 ± 標準偏差	中央値	
1	L 字	壁	3.86 ± 2.35	4.5	0.14 ± 0.36	0	0.00007**
	T 字	壁	3.57 ± 2.31	4	0.29 ± 0.47	0	0.0003**
	トの字	壁	3.42 ± 2.28	4.5	0.14 ± 0.36	0	0.0001**
	X 字	壁	3.71 ± 2.20	4.5	0.14 ± 0.36	0	0.00007**
2	ルート 2-1	障害物	1.28 ± 0.73	1	0.50 ± 0.52	0.5	0.0004**
		壁	3.14 ± 3.61	1.5	0.57 ± 0.94	0	0.04*
	ルート 2-2	障害物	2.21 ± 1.42	2.5	1.35 ± 1.00	1	0.1
		壁	1.86 ± 3.18	0.5	0.62 ± 1.01	0	0.4
3	ルート 3-1	障害物	3.07 ± 1.49	3.5	1.28 ± 1.32	1	0.005**
		壁	12.21 ± 9.67	10.5	1.07 ± 1.27	1	0.0002**
	ルート 3-2	障害物	3.71 ± 2.34	3.5	0.85 ± 1.29	0	0.0009**
		壁	15.21 ± 12.75	14	1.43 ± 2.10	0	0.0008**

用いて白杖のみを用いた場合とシステムを併用した場合を比較したところ、タスク 2 のルート 2-2 以外ではシステム併用時に白杖と壁および障害物との接触回数が有意に減少した。ルート 2-2 では障害物が互いにわずか 3m しか離れていないため（図 4）、他のタスクと比べ難しかったことが原因だと考えられる。表 1 の壁との接触回数の結果から、システムを併用することで参加者は壁伝いに歩かずに通路を移動できたことが分かった。

5.4 主観評価結果

図 6 に主観評価結果を示した。システムの各機能について、白杖のみを使用した時 (only cane) とシステムを併用した時 (system) を Wilcoxon の符号順位検定（有意水準 5%）を用いて比較したところ、Q2-4 において有意差が認められた。

5.5 ユーザフィードバック

インタビューを通して、参加者それぞれがシステムの異なる側面を評価していることが判明した。12

名の参加者は障害物回避機能を評価し、以下のようなコメントが得られた：A1 「迂回シグナルでスムーズに障害物を回避できた。白杖を使わなくても通路を歩ける安心感があった。」(P05), A2 「指示を聞くためにゆっくり歩く必要があったが、障害物を回避出来て嬉しい。」(P01) 特に 9 名は以下のように、システムが生成した経路に沿ってペアリングせずに歩くように支援する機能を評価した：A3 「このシステムはまっすぐ歩くことの補助になる。最初は必要ないと思っていたが、壁に沿って歩けない所では通路の真ん中をまっすぐ歩く必要があるので、助けになる。」(P09) また、13 名の参加者は交差点検出機能を評価し、特に 9 名はこの機能をメンタルマップ作成時に使いたいとコメントした：A4 「もうすぐ交差点であるとわかることで、交差点を突っ切ってしまう心配がない。交差点の伸びている方向を全て確認することで、実は他の方向にも道が伸びていたと知れて、新しい発見になる。」(P01)

13 名の参加者は、システムがスマートフォン 1 台

で完結している点を高く評価した：A5「スマートフォンは必ず持ち歩いているので、スマートフォン1台で完結していることは良い。」(P08) 一方で11名はスマートフォンを持つことによって片手がふさがる事を指摘し、特に4名の参加者はスマートフォンの持ち方に否定的なコメントをした：A6「スマートフォンを体と平行に持つのが難しい。」(P08)

システムを使いたい場所として、病院やショッピングモールがあげられた。本研究では目的地への道順をユーザが把握している状況での使用を想定してシステムをデザインしたが、6名の参加者はシステムを他社のオフィスなど目的地への道順が分からぬ場所でも使いたいと答えた：A7「このシステムがアリストア等でリリースされたら視覚障害者によって（実験時の用途に限らず）色々な使用方法が提案されると思う。」(P13)

6 議論

6.1 Corridor-Walker の有用性

実験の結果、参加者はシステムを使うことで壁伝いに歩かずに障害物を回避し（表1）、交差点の形状を把握したうえで曲がることができた（5.2節）ことが分かった。システムを併用した場合、システムからの指示の確認や方向修正によってタスク完了時間が白状のみの場合よりも長くなったが（5.1節）、主観評価において参加者は、システムを併用することで白杖のみの時よりもスムーズに障害物を回避し（Q2、A2）、自信を持ってスムーズに歩く方向を修正できた（Q3、Q4）とシステムを高く評価した。12名の参加者からは白杖を使わずに障害物を避けたこと（A1、A2）、9名から壁に頼らず歩けたこと（A3）を歓迎するコメントが得られた。また、交差点検出機能に関する質問事項（Q5、Q6）に関して有意差は認められなかったものの、13名から交差点検出機能は交差点の通り過ぎを防止し、特に9名からはメンタルマップ作成に役立つと評価された（A4）。さらに、P03以外の全ての参加者はシステムがスマートフォン1台で完結していることは強みであると同意した（A5）。

6.2 システムの持ち方と訓練方法

スマートフォンを用いる利点として、ユーザがスマートフォンを手にもって、周囲をスキャンするというインタラクションが可能になる点があげられる。一方で11名の参加者はスマートフォンを持つことで片手がふさがってしまうことを欠点と感じ、特に4名はシステムを持つ方法が難しいとコメントした（A6）。評価実験では30分ほどの練習時間を確保したが、タスク1ではユーザが安定してスマートフォンを持てなかつたためにシステムが交差点の形状を誤認識した場合があった（理由2）。今回のシステ

ムの持ち方は視覚障害者の間で一般的な持ち方でないため、システムの普及のためには視覚障害者の歩行訓練（O&Mトレーニング）にどのように組み込んでいくか検討する必要がある。

6.3 課題と今後の展望

システムの課題としては、交差点検出モデルが通路の直交する交差点のみを対象としている点が挙げられる。提案モデルは交差点の通路に対して「Left, Right, Front, Back」の4種類のラベルを推定する多クラス分類のタスクとして交差点の種類を判定しているが、実世界においては、五差路やY字の交差点など通路が直交せず上述の4種類のラベルを付与できない複雑な交差点も存在する。これらの複雑な交差点にも対応するためには、通路の角度を直接推定する交差点検出方法を考案する必要がある。

またシステムは幅が3.5m以上ある広い通路やオープンスペースで安定して使用することが難しいという課題がある。これはシステムが目的地を設定するためには通路の両側の壁がLiDARセンサの検出範囲内に入っている必要がある反面、iPhone 12 Proに搭載されたLiDARセンサの検出範囲が5mに限定されるためである[3]。システムはスマートフォン1台で完結していることが利点である一方、使える場所がデバイスのセンサ上の制約によって限定されることがある。

本研究ではユーザが目的地への道順を把握している状況を想定してシステムを開発したが、6名の参加者から道順がわからない場所でもシステムを使いたいというコメントが得られた。ユーザが日々の生活でシステムを試す中で、我々が意図しなかった様々な使用方法を産み出す可能性があるため（A7）、今後はシステムをアプリとして広く配布し、ユーザからのフィードバックを集めていきたい。

7 まとめ

本研究は屋内の通路を歩く視覚障害者が障害物を回避し、交差点を認識するためのスマートフォン型歩行支援システムを提案した。システムはスマートフォンのLiDARセンサから周囲の占有格子地図を作成することで、障害物を回避する経路を生成し、交差点の位置と形状を認識する。システムは骨伝導イヤホンの左右からのビープ音、音声、振動シグナルを用いて生成された経路に沿ってユーザを案内し、交差点の位置と形状を伝える。実験からシステムを併用した参加者は壁伝いに歩かずに障害物を回避し、交差点の形状を認識したうえで曲がることができた。また、システム併用時にタスク完了時間は伸びたものの、参加者はスムーズに障害物回避と歩く方向の修正ができたとシステムを高く評価した。今後はシステムの配布に向けて、より一般的な交差点の検出方法を検討していきたい。

謝辞

本研究は、JSPS 科研費 (JP20J23018) と JST 未来社会創造事業 (JPMJMI19B2) の助成を受けた。

参考文献

- [1] M. Adachi, S. Shatari, and R. Miyamoto. Visual navigation using a webcam based on semantic segmentation for indoor robots. In *2019 15th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, pp. 15–21, San Francisco, CA, USA, 2019. IEEE.
- [2] A. Constantinescu, K. Müller, M. Haurilet, V. Petrausch, and R. Stiefelhagen. Bring the Environment to Life: A Sonification Module for People with Visual Impairments to Improve Situation Awareness. In *Proceedings of the 2020 International Conference on Multimodal Interaction*, pp. 50–59, New York, NY, USA, 2020. ACM.
- [3] A. Developer. Displaying a Point Cloud Using Scene Depth. Retrieved in September, 2021 from https://developer.apple.com/documentation/arkit/environmental_analysis/displaying_a_point_cloud_using_scene_depth, 2021.
- [4] A. A. Díaz-Toro, S. E. Campaña-Bastidas, and E. F. Caicedo-Bravo. Vision-Based System for Assisting Blind People to Wander Unknown Environments in a Safe Way. *Journal of Sensors*, 2021, 2021.
- [5] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.
- [6] A. Garcia, E. Mattison, and K. Ghose. High-speed vision-based autonomous indoor navigation of a quadcopter. In *2015 international conference on unmanned aircraft systems (ICUAS)*, pp. 338–347, Los Alamitos, CA, USA, 2015. IEEE.
- [7] A. Garcia, S. S. Mittal, E. Kiewra, and K. Ghose. A convolutional neural network vision system approach to indoor autonomous quadrotor navigation. In *2019 International Conference on Unmanned Aircraft Systems (ICUAS)*, pp. 1344–1352, Los Alamitos, CA, USA, 2019. IEEE.
- [8] J. a. Guerreiro, E. Ohn-Bar, D. Ahmetovic, K. Kitani, and C. Asakawa. How Context and User Behavior Affect Indoor Navigation Assistance for Blind People. In *Proceedings of the 15th International Web for All Conference*, New York, NY, USA, 2018. ACM.
- [9] J. Guerreiro, D. Sato, S. Asakawa, H. Dong, K. M. Kitani, and C. Asakawa. CaBot: Designing and Evaluating an Autonomous Navigation Robot for Blind People. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*, pp. 68–82, New York, NY, USA, 2019. ACM.
- [10] P. E. Hart, N. J. Nilsson, and B. Raphael. A formal basis for the heuristic determination of minimum cost paths. *IEEE transactions on Systems Science and Cybernetics*, 4(2):100–107, 1968.
- [11] D. Holz, S. Holzer, R. B. Rusu, and S. Behnke. Real-time plane segmentation using RGB-D cameras. In *Robot Soccer World Cup*, pp. 306–317, New York, NY, USA, 2011. Springer.
- [12] D. Jain. Path-guided indoor navigation for the visually impaired using minimal building retrofitting. In *Proceedings of the 16th international ACM SIGACCESS conference on Computers & accessibility*, pp. 225–232, New York, NY, USA, 2014. ACM.
- [13] W. Jeamwatthanachai, M. Wald, and G. Wills. Indoor navigation by blind people: Behaviors and challenges in unfamiliar spaces and buildings. *The British Journal of Visual Impairment*, 37:140 – 153, 2019.
- [14] S. Kayukawa, K. Higuchi, J. Guerreiro, S. Morishima, Y. Sato, K. Kitani, and C. Asakawa. BBeep: A sonic collision avoidance system for blind travellers and nearby pedestrians. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, New York, NY, USA, 2019. ACM.
- [15] S. Kayukawa, T. Ishihara, H. Takagi, S. Morishima, and C. Asakawa. Guiding Blind Pedestrians in Public Spaces by Understanding Walking Behavior of Nearby Pedestrians. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(3):1–22, 2020.
- [16] G. Lacey and S. MacNamara. Context-aware shared control of a robot mobility aid for the elderly blind. *The International Journal of Robotics Research*, 19(11):1054–1065, 2000.
- [17] Y. H. Lee and G. Medioni. Wearable RGBD indoor navigation system for the blind. In *European Conference on Computer Vision*, pp. 493–508, New York, NY, USA, 2014. Springer.
- [18] B. Li, J. P. Munoz, X. Rong, J. Xiao, Y. Tian, and A. Arditi. ISANA: wearable context-aware indoor assistive navigation with obstacle avoidance for the blind. In *European Conference on Computer Vision*, pp. 448–462, New York, NY, USA, 2016. Springer.
- [19] J. Morris and J. Mueller. Blind and deaf consumer preferences for android and iOS smartphones. In *Inclusive designing*, pp. 69–79. Springer, New York, NY, USA, 2014.
- [20] M. Nakamiya, Y. Kishino, T. Terada, and S. Nishio. A Route Planning Method Using Cost Map for Mobile Sensor Nodes. In *2007 2nd International Symposium on Wireless Pervasive Computing*, Los Alamitos, CA, USA, 2007. IEEE.
- [21] V. Pradeep, G. Medioni, and J. Weiland. Robot vision for the visually impaired. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*, pp. 15–22, San Francisco, CA, USA, 2010. IEEE.

- [22] G. Presti, D. Ahmetovic, M. Ducci, C. Bernareggi, L. Ludovico, A. Baratè, F. Avanzini, and S. Mascetti. WatchOut: Obstacle sonification for people with visual impairment or blindness. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*, pp. 402–413, New York, NY, USA, 2019. ACM.
- [23] J. Redmon and A. Farhadi. YOLOv3: An Incremental Improvement, 2018.
- [24] P. R. Sanketi and J. M. Coughlan. Anti-blur feedback for visually impaired users of smartphone cameras. In *Proceedings of the 12th international ACM SIGACCESS conference on Computers and accessibility*, pp. 233–234, New York, NY, USA, 2010. ACM.
- [25] D. Sato, U. Oh, J. Guerreiro, D. Ahmetovic, K. Naito, H. Takagi, K. M. Kitani, and C. Asakawa. NavCog3 in the wild: Large-scale blind indoor navigation assistant with semantic features. *ACM Transactions on Accessible Computing (TACCESS)*, 12(3):14, 2019.
- [26] J. Striegl, C. Lotisch, J. Schmalfuss-Schwarz, and G. Weber. Analysis of Indoor Maps Accounting the Needs of People with Impairments. In *ICCHP*, Vol. 12377, New York, NY, USA, 2020. Springer.
- [27] J. Yuan, S. Zhang, Q. Sun, G. Liu, and J. Cai. Laser-based intersection-aware human following with a mobile robot in indoor environments. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2018.
- [28] 森 大輝, 入江 英嗣, 内原 正一, 荒川 明宏, 坂井 修一. リアルタイム経路生成と振動通知による視覚障害者の歩行支援システム. Retrieved in November, 2021 from <https://www.wiss.org/WISS2018Proceedings/oral/04.pdf>, 2018.

未来ビジョン

これまで視覚障害者のための移動支援システムが数多く提案されてきたが、多くの研究が地図情報などの事前準備が必要であったり、扱い慣れていない追加デバイスを持ち歩く必要があったりと、視覚障害者が日常的に使用する上でハードルが高いため、実際に視覚障害者の間で使われているものは極めて少ない。そのため本研究では、視覚障害者に日常的に使ってもらってこそ支援システムを提案する意義が出るという思いから、システムが普及するためのハードルを下げるために「スマートフォン1台」かつ「事前準備を必要としない」という2つの条件にこだわり Corridor-Walker を提案した。将来的に実際の視覚障害者のコミュニティでシステムが広く使われることで、A7 のコメントでも指摘されたように、ユーザ側でもシステムの使い方や新しい使い道に関して検討や議論が行われることも期待している。

また、「スマートフォン1台」という条件にこだわった本研究の究極的な目標として、我々は一般的に普及しているデバイスが視覚障害者にとって白杖に代わるユビキタスな支援システムとして広く普及する可能性を期待している。白杖は一步先の危険（障害物や段差など）を確実に検出できるため視覚障害者に広く利用されている。しかしながら、白杖は視覚障害者しか所有しない特殊なツールであり、そのような白杖を携帯して周囲から注目を集めることに抵抗感を感じるユーザも存在している。本研究で提案した「スマートフォン1台」という条件にこだわる方針の先には、白杖のような視覚障害者専用の支援ツールではなく晴眼者含め誰もが持ち歩いているデバイスだけを用いて視覚障害者も周囲に溶け込みつつ晴眼者さながら自立して移動できるような未来が実現できると期待している。

CircWood: CW レーザー加工機による炭化を利用した木製回路基板

石井 紗郁^{*} 加藤 邦拓[†] 池松 香[‡] 川原 圭博[§] 椎尾 一郎^{*}

概要. 本研究では、レーザー加工機で木材表面を部分的に炭化させ、木材に直接センサや回路を作成する手法を提案する。一般的なプリント基板（PCB）の配線と同様、炭化部分は導電性を有する配線として機能する。先行研究として、フェムト秒（fs）レーザーをラスタスキャニングし小規模なグラフエンを作成する手法が提案されているが、fs レーザーは連続波（CW）レーザーと比較するとパーソナルファブリケーションに不向きである。また、ラスタスキャンでは配線が主体の回路を作成する場合多くの時間を要する。そこで我々は、CW レーザーによるベクタスキャニングを利用し、炭による回路の作成に必要なコストと時間を低減する。提案手法では、パーソナルファブリケーションで頻繁に使用される CW レーザー加工機を用いて、タッチセンサ、ダメージセンサ、荷重センサなどのさまざまなセンサや回路を木材表面に直接作成することができる。さらに、作成したセンサや回路は木工 DIY で一般的に使用される金属製のネジや釘を使用し PCB やマイコンなどに容易に接続できる。

1 はじめに

木は有機的な構造を持ち熱伝導率が低く、独特の温もりを有するため、家具や壁掛けボードなどの日常的に触れるようなさまざまな製品の材料として使用されてきた。木材を加工する伝統的な各種工具に加えて、最近ではレーザー加工機による加工が普及しつつある。レーザー加工機ではレーザー光を利用して木材を加工するが、切断と彫刻以外の加工方法は広く探求されてこなかった。我々は、連続波（CW）レーザー加工機で木材を部分的に炭化させ、木材表面に直接さまざまなセンサや回路を作成する手法を提案する。一般的なプリント基板（PCB）と同様に、炭化した部分は導電性のある配線として機能する。

ラスタスキャン方式（主に彫刻に用いられる方法）のフェムト秒（fs）レーザーを用いて木や葉の表面にグラフエンを生成し、これを用いた回路を作成する手法が提案されている [11]。fs レーザー加工機は CW レーザー加工機に比べて高価であり、パーソナルファブリケーション（以下、ファブリケーション）に不向きである。また、ラスタスキャンでは、配線を主体とした回路パターンの作成に多大な時間を要する。本論文では、CW レーザーによるベクタスキャニング（主に切断に用いられる）を利用することで先行研究の手法を拡張する。導電性を有する炭を十分に生成させるため、レーザーの焦点をぼかし、繰り返し（8~15 回）ベクタスキャニングを行う。本手法により、炭による回路の作成時間を先行研究の手法と比べて 1/600 に短縮することができる。また、回路以外にも、タッチセンサ、ダメージセン

サ、荷重センサなどのさまざまなセンサを実装した。これらのセンサや回路は、一般的な PCB やマイコンなどに、木工 DIY で使用される金属のネジや釘を使って容易に接続することができる。本手法は木に導電性塗料などの人工物を適用することなく、木そのものから生成する物質を利用しセンサや配線パターンを作成する。よって、素材を活かしたインターフェースによる自然なインタラクションが可能である。なお、本研究の一部について ACM UIST2021 でデモ発表を行った [6]。

2 関連研究

2.1 回路のファブリケーション

回路やセンサのファブリケーションのために、さまざまな手法が提案されている。その一つに、導電性インクを用いインクジェットプリンタで導電回路パターンを印刷する手法がある。Kawahara らは、銀ナノ粒子インクを装填したインクジェットプリンタを使用し、紙に電子回路を印刷するラピッドプロトタイピング技術を提案した [8]。低コストなプロトタイピングが可能なため、その後の多くの研究でこの手法を利用したセンサ [3, 12, 7, 22, 14], NFC アンテナ [16], 積層回路 [21], ディスプレイ [15] などが作成されている。この手法では焼結などの後処理を行わないため、印刷された回路は耐久性が低いとされている [14]。そこで、印刷面を高温で硬化させることで、伸縮可能な材料の上に堅牢な回路を作成する Soft Inkjet Circuits が提案された [9]。

プリンタの印刷可能領域を超える大きな平面へ回路を実装するために、導電性塗料やスプレーを利用する技術がいくつか提案されている。FlexTouch [24], Electrick [27], Sprayable User Interfaces [25] は、導電性塗料を壁や家具に塗ったり吹き付けたりするこ

* お茶の水女子大学

† 東京工科大学

‡ ヤフー株式会社

§ 東京大学

とで、静電容量方式によるタッチ入力を可能にする。

また、抵抗値が低く、容易に切断、折り曲げができる低コストである銅テープや銅箔を使用する手法も提案されている。Midasは、銅テープをカッティングプロッタでカットし、静電容量式のタッチセンサや配線を作成するツールである[18]。Perumalらは、汎用のインクジェットプリンタやレーザープリンタと汎用のインクを使用したPCBの作成を実現するPrintemフィルムを開発した[1]。これらの手法には、切削後に不要な材料を除去するのが面倒で時間がかかること、除去する際に細い配線が断線する可能性があることなど、いくつかの欠点がある[18]。上記以外にも、金箔[20, 17]、水圧転写技術[4]、導電性材料を塗布したナイロン繊維[10]、カーボンコート紙[28]などが回路の作成に用いられている。

以上のように、実験室環境での迅速で低成本な回路のファブリケーション手法が多数提案されている。一方、本研究では、レーザー加工によって木材表面に炭を生成し、これを用いて回路やセンサを作成する。炭による回路配線の抵抗値は既存手法での配線の抵抗値よりも高いが、我々の手法では木そのものから生成する物質を利用してセンサや配線パターンを作成できる。従来手法で木材表面に回路を作成するには、木材に導電性材料を印刷／塗布する必要があり、木の自然な風合いが損なわれてしまう。

2.2 レーザー加工によるファブリケーション

レーザー加工による3Dオブジェクトのファブリケーション手法が盛んに研究されている。例えば、2Dの板材に熱を加えて折り曲げる[13]、または溶接して積層する[23]などして、3次元のオブジェクトをラピッドプロトタイピングする手法がある。これらの手法では、我々と同様、レーザーの焦点を意図的にぼかしてベクタスキャニングすることで、素材を傷つけず特定の領域を加熱し、折り曲げと溶接を実現している。レーザー加工機を利用して電子回路を作成する技術も提案されている。LASEC[5]では、一般的なレーザー加工機で透明な導電シートを切断し、伸縮性を持つ回路を作成している。一方我々は、導電材料を新たに塗布することなく、木材表面に炭による回路を作成する。

レーザーを用いて材料の表面を炭化し、小規模(最大 $3 \times 3 \text{ cm}^2$)のグラフェンを作成する手法がいくつか提案されている[26, 2, 11]。最初期の研究[26]では、不活性ガスを使用して材料の燃焼を防ぐことで、木材表面にグラフェンを作成する手法が提案されている。その後、CWレーザーの焦点をぼかして複数回照射する手法[2]とfsレーザーを1回照射する手法[11]が提案され、不活性ガス不使用でグラフェンを作成可能になった。我々の手法では、木の表面に電子回路やセンサを作成することを目的とし、ファブリケーションへの適用を目指している。また、先行研究[26, 2, 11]では、ラスタスキャ

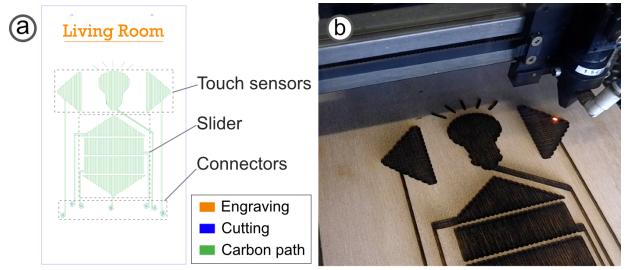


図 1. (a) タッチスイッチとスライダ搭載の壁掛けボードの設計例。(b) レーザーで木の表面に炭を生成し回路を作成する。木材の温度が上がりすぎないよう、レーザーの焦点をぼかしている。

方式のレーザーを使用しているが、我々はベクタスキャニング方式のレーザーを使用した。配線を主体とした回路パターンを作成する場合、ラスタスキャニング方式だとパターンの全領域を走査する必要がある。一方で、ベクタスキャニング方式は配線に沿ってレーザーを照射するため、回路作成に必要な時間が短くなる(詳細な比較は7節に示す)。fsレーザーを用いた手法の研究[11]では、小型の回路や温度センサの試作を行っているが、fsレーザーは高価であり¹、ファブリケーションには向きである。CWレーザー加工機を利用する我々の手法により、DIYユーザでも容易に回路を作成することができる。

3 ファブリケーション手順

本節では、提案するファブリケーション手法の手順について説明する。使用したレーザー加工機は株式会社コマックスのVD7050-60Wであり、パワーを30–50%(=18–30 W)、スピードを25–30%(=381–457 mm/s)に設定した。

3.1 回路パターン作成

まず、一般的な2Dグラフィクスエディタ(Adobe Illustratorなど)やCADを使用し回路パターンを作成する。図1(a)は複数のタッチスイッチとスライダーを搭載した壁掛けボードの設計例である。提案手法ではレーザーをベクタスキャニングするため、導電部のパターンは全て線で構成される。タッチスイッチやPCBとの接続部などの面積の大きい部分には櫛形パターンを使用する。櫛の間隔は、炭配線を生成した際、線同士に隙間や過剰な重なりが生じないよう決定する必要がある。これは焦点をぼかすためにレーザー加工機のステージを下げる距離に依存し、焦点位置から5 mm下に下げる場合は1 mm間隔、6mm下げる場合は1.25 mm間隔であった。

3.2 レーザーによる炭化

回路パターン作成が完了したら、レーザー加工機にデータを送信する。一般的なラスタスキャニングによるレーザー彫刻では、スキャンライン間に隙間が生じ、導電性の高い炭を生成することができない。

¹ CWレーザー加工機の価格は数十万円程度であるが、fsレーザー加工機は一千万円程度である。

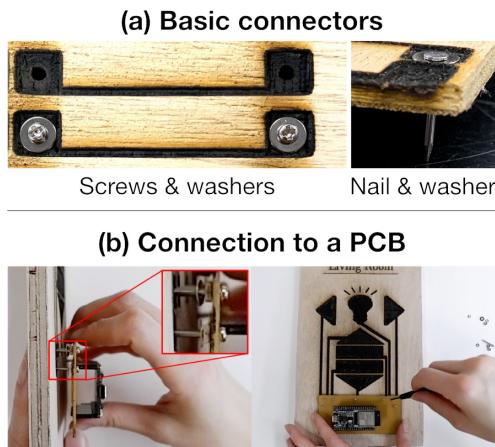


図 2. (a) 一般的なレーザーによる切断で炭配線に穴を開け、ねじとワッシャーで接続する。釘とワッシャーを使用することもできる。(b) マイコンを搭載した PCB と炭配線との接続。

一方、一般的なベクタスキャニングによるレーザー切断を使用すると木は焼けて切断されてしまい、十分な量の炭が生成しない。炭を十分に生成させ導電性のある配線を作成するためには、適度な熱を繰り返し与える必要がある。そこで、先行研究 [2] と同様にレーザー加工機のステージを下げ、焦点をぼかす方法を用いる(図 1 (b))。適切なぼかし距離は焦点距離から 5–6 mm 下であった。また、レーザー加工機のパワーを通常の切断時よりも低く、スピードを速く設定し繰り返しレーザーを照射した。これにより、導電性の高い炭を十分に生成できることを確認した。レーザーの繰り返し照射回数は 8–15 回が適当であった。以上のレーザーパラメタは、5.2 節で述べる技術評価の結果に基づいて決定した。

3.3 PCB との接続

最後に、一般的な金属ねじや釘を使用し、炭配線を電子部品や PCB に接続する。レーザー加工機やボール盤を使用し、炭配線の端子部分の中央に穴を開け、ねじや釘を差し込む。このとき、炭とねじとの接触面積を増やすため、ワッシャーやナットを利用している(図 2)。本論文の試作では、主に M2 のねじとナット、直径 6 mm のワッシャーを使用した。

4 CircWood

本節では、提案手法を用いることで可能となるセンサ機能と入出力について述べる。これにより、木の特性を生かした様々なインターラクションが実現できる。また金属ねじやヒンジなどの木工部品を利用し、通常の電子部品や PCB との接続が可能である。

4.1 センサ機能

タッチ／近接センシング、スライダー。 提案手法は、従来の静電容量式タッチセンサと同様の仕組み

でタッチセンシングを実現できる。炭で作られた電極を用意し、これに人体が近づくことで変化する静電容量値を利用し、人のタッチ検出ができる。

ダメージセンシング。 木の板は大きな衝撃を受けるとひび割れを起こす。この性質を利用し、割れなどの損傷を検出することができる。試作では、Vcc を固定抵抗と炭配線で分圧し、Arduino のデジタル入力端子に接続した。衝撃が加わり炭配線が断線すると抵抗値が無限大になるため、破損によるダメージを検出できる。

荷重／曲げセンシング。 木板が歪むと板上の炭配線の抵抗値が増減する。固定抵抗と炭配線とを直列接続した分圧回路を用意し、Arduino のアナログ入力端子に接続することで、炭配線にかかる歪みを検出できる。また、レーザ加工機により作成するリビングヒンジの裏に炭配線を作成し、その抵抗値の変化を測定することで、曲げの検出も可能である。

回転角センシング。 2 枚の木板に炭の電極を作成し、それらを向かい合わせに重ねることで、2 個の電極間の静電容量値から、重なり合った部分の大きさを検出できる。このとき、2 枚の木板をねじ止めすることで、回転角度の検出が可能である。また、2 枚の木板をスライドするよう取り付けることで、摺動量の検出も可能である。例えば、木製の引き戸や窓とその戸袋の縁に炭による電極を作成することで、ユーザが引き戸を開閉する動作を検出できる。

4.2 木工部品との接続

炭配線のコネクタとして、ねじ、釘、ヒンジ、鋸など、DIY で頻繁に使われるさまざまな金属製の木工部品が使用できる。3.3 節で述べたように、PCB を炭配線に接続するためにねじ、釘、インサートナットなどを使用することができる。木板の両面にある 2 本の炭配線は、ねじとハトメを使用し接続できる。絶縁ステープルは、他の炭配線との不要な接触を防ぐためのジャンパワイヤとして使用できる。また、ヒンジや鋸、キャッチなどを使用し、2 枚の木板上の炭配線同士を接続可能である。引き戸に付いている金属のローラと戸袋の金属レールは、引き戸と戸袋に作成された炭配線同士を接続するために使うことができる。さらに、ドアノブ、取手、ドアノッカーは、これらの部品に人が接触したことを検出するために使用することができる。

5 炭化の最適化

5.1 適用可能な木材

提案手法を適用可能な木材を調査するため、さまざまな種類の木材を使用し 3 章の手順で炭を作成した。具体的には、ラワン合板と 8 種類の無垢材(ラワン、ヒノキ、桐、スギ、シナ、ブナ、ナラ、ウォールナット)を使用した。また、燃焼を防ぎ炭素の発生を促す難燃剤を木材表面に塗布した場合の効果に



図 3. 木材表面に生成した炭配線の例. (a) 本手法が適用可能な木材. ラワンや難燃剤を塗布したヒノキは導電性を持つ炭配線を生成できる. (b) 本手法に向きのない木材. 炭の粒同士に間隔があり配線が十分に黒くならず, 導電性が低い.

ついても検証した. 難燃剤の成分はポリリン酸アンモニウムと硫酸アンモニウムであり, 食品添加物なため環境にも人体にも無害である.

結果としては, ラワン材を用いて作成した炭の導電性が最も高く(面抵抗率: $30\text{--}75 \Omega/\text{sq}$)², 提案手法に適した素材であることがわかった. また, ヒノキの無垢材でも, 難燃剤を塗布することで導電性を有する($\sim 76 \Omega/\text{sq}$)炭配線を作ることができた. ラワンの無垢材に難燃剤を塗布するとさらに導電性が向上したが($\sim 25 \Omega/\text{sq}$), ラワン合板に難燃剤を塗布した場合, 導電性が低下した($\sim 150 \Omega/\text{sq}$). これは, 合板の層の間にある接着剤の影響であると推測される. 以上より, 最適な木材は, 難燃剤を塗布したラワン無垢材, 次いでラワン合板, 難燃剤を塗布したヒノキ無垢材となった. 材料によって木の風合いやさまざまな特性が異なるため, プロジェクトに適した材料を選択することができる.

図 3 に木材表面に生成された炭配線の例を示す. 図 3 (a) に示すように, 導電性を有する炭配線は全体的に黒く, 炭素の粒子で満たされている. 一方, 図 3 (b) に示すように, 十分な導電性が得られなかつた炭配線($450 \Omega/\text{sq}$ -非導通)は茶色っぽく, 炭の粒子の間に隙間が見られ, 十分な炭が生成できなかつた様子が伺われる. 導電性が低い炭配線は, ブナ, カシ, ウォールナットなどの硬い木や, スギなどの軟らかい木から生成される傾向があった. 一方, 導電性を有する炭配線は, ラワンやヒノキなどの中程度の硬さの木材から生成される傾向があった.

5.2 レーザーパラメタ

厚さ 4 mm のラワン合板を用いて, 導電性の高い炭を作成するためのレーザー加工機の最適な設定パラメタを調査した. この合板は表層の厚さが約 0.5 mm, 中間層の厚さが約 3 mm であり, 炭は主に中間層に発生した. この合板による炭は, 実験したラワン合板の中で最も高い導電性を示した. 前節で述べたように, 導電性が低下するため, 合板に難燃剤は塗布しなかった. 使用したレーザー加工機は 3 節で述べたコマックス社の VD7050-60W である.

まず, 焦点をぼかすためにレーザー加工機のス

テージを下げる距離の最適値を調査した. その結果, 焦点距離から 5 mm 下げたとき, 生成する炭配線の抵抗値が最も低くなることがわかった. レーザーの照射速度が低速な場合, レーザーのパワーを高くすると木が焼けてしまい, 低くすると炭が発生しないなど調整が困難であったため, 高速で繰り返し照射した. レーザーの最適な繰り返し照射回数を調査したところ, 15 回目までは炭の抵抗値が低下し続けたが, 16 回目には炭が焼け落ちて抵抗値が上昇することがわかった. 以上までの調査で判明した焦点ぼかし距離とレーザー照射回数を最適条件(焦点距離から 5 mm 下, 15 回繰り返し照射)として固定し, レーザーのパワーとスピードの最適値を調査した. 結果としては, 最も抵抗値が低かったのは $30 \Omega/\text{sq}$ で, このときのパワーは 50 % (30 W), スピードは 25 % (381 mm/s) であった. これらの設定値を最適パラメタとしてアプリケーションの実装時に使用した.

木材の種類が変わると最適パラメタにも若干の変化が見られた. 厚手のラワン合板(1 mm 厚のベニヤ 3 枚と 3 mm 厚のベニヤ 2 枚が交互に重ねられている)とラワン無垢材では, パワーとスピードの最適パラメタは上述の値と同じであったが, 最適なレーザーのぼかし距離は焦点距離から 6 mm 下であった. このとき, ラワン合板上に生成した炭の最小抵抗値は $75 \Omega/\text{sq}$ であった. ラワン無垢材は難燃剤を塗布するとさらに抵抗値が下がり, 最小抵抗値は $25 \Omega/\text{sq}$ であった. 難燃剤を塗布したヒノキを上述のパラメタで加工したところ燃えて切れてしまったため, 最適パラメタのさらなる調査を行った. その結果, 最適なぼかし距離は焦点から 5 mm 下であり, パワー 30 % (18 W), スピード 30 % (457 mm/s) で 8 回レーザーを繰り返し照射することが最適であることがわかった. このときの最小抵抗値は $76 \Omega/\text{sq}$ であった. 以上の調査に基づき, 表 1 に提案手法の最終的な推奨設定を示す.

6 応用例

6.1 タッチスイッチ付き壁掛けボード

提案手法はベクタスキャン方式のレーザーを使用しており, ラスタスキャン方式より大型の回路を高

² 作成した炭の抵抗値 (R) を測定し, 面抵抗率 (R_s) を $R_s = R \times (\text{線幅}/\text{線長})$ の公式に従い計算した.

表 1. 提案手法の推奨パラメタ. 抵抗値は測定できた最小値を示す.

木材	難燃剤	パワー [W]	スピード [mm/s]	繰り返し	焦点ぼかし [mm]	抵抗値 [Ω/sq]
ラワン無垢	有	30	381	15	6	25
ラワン合板	無	30	381	15	5–6	30–75
ラワン無垢	無	30	381	15	6	65
ヒノキ無垢	有	18	457	8	5	76

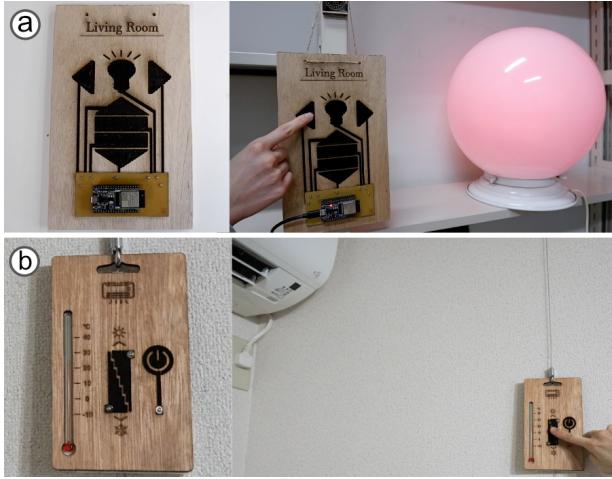


図 4. (a) 調光ボード (ボード全体のサイズ: 255 mm×155 mm). 矢印型のスイッチをタッチして光の色を制御できる. (b) エアコンの制御ボード (ボード全体のサイズ: 167 mm×103 mm). スライダーに触れることで温度調節が可能である.

速に作成できる. よって, 3 節で述べたような家電製品の制御ボードなど, スマートホーム向けのアプリケーション実装に適している. そこで, 図 4 に示すように, 室内の調光用とエアコンの制御用の 2 種類の壁掛けボードを実装した. それぞれのボードには Arduino を搭載した PCB が金属ねじで固定され, 同時にそのねじにより炭配線との接続が行われている. 調光用ボードでは, 電球型のタッチスイッチで RGB ランプを点灯・消灯し, 矢印型のタッチスイッチで光の色を変え, 上下のスライダーで明るさを調整できる. エアコン制御ボードには, 電源 ON/OFF 用のタッチスイッチと, 温度調節用のアナログタッチ式スライダーが搭載されている. タッチスイッチの実装には Arduino の Capacitive Sensing Library を使用した. 今回は壁掛けボードとして試作したが, 木製の壁全体に適用することで, より大規模な平面でのインタラクションも可能と考えている.

6.2 衝撃履歴検出ボックス

物流の用途に木箱が使われることがある. この木箱に提案手法を適用することで, 木箱自体に衝撃履歴を一度のみ検出する機能を付与できる. そこで, 木



図 5. 衝撃履歴検出ボックス. (a) 箱の底に炭配線が作成されている. 配線の破損はスマートフォンアプリで容易に検知できる. (b) 炭配線を介したタッチ入力が検出される. (c) 衝撃を加えたところ, 炭配線が破損し衝撃履歴が検出された.

箱の裏面に炭配線を作成し, 大きな衝撃の履歴を検出する応用例を実装した(図 5). Off-Line Sensing [19] の技術と同様に, スマートフォンのタッチスクリーンでの静電容量センシングを利用して, 炭配線の破損を検出した. 配線が切れていない場合は, 炭配線の一端を指で触れ, もう一端をタッチスクリーンに接触させると, タッチスクリーン上でのタッチ入力が発生する. この機能は木製の建材表面や梯子などにも応用でき, 亀裂を検出することで劣化の検知が可能であると考えている. 加わる衝撃とそれに応じて破損する配線パターンとの関係については今後調査したい.

6.3 インタラクティブチェア

提案手法により, 木製の椅子の座面や背面に荷重センサを作成可能である. 図 6 (a) は, 健康に悪影響を及ぼす長時間の静的な着座を防止するインタラクティブチェアを示す. この椅子の座面の裏側には炭配線が作成されており, 人の着座により座面が歪むと配線の抵抗値が増加する³. この抵抗値の変化を読み取ることで, 着座状態を検出することができる. 長時間の着座を検出すると PC アプリケーションを介してユーザに通知し, 姿勢を変えることを促す.

6.4 音量コントローラ

提案手法による回転角センサを利用し, ダイヤル式の音量コントローラを実装した(図 6(b)). 半円状の電極が作成された 2 枚の木板を向かい合わせに配置しネジで固定すると, 2 個の電極が平行板可変容量コンデンサとして機能する. ユーザが板を回転させると, 2 個の電極の重なる部分の面積が変化し, 静電容量値が変化する. よって, 板の回転による連

³ 初期状態では約 26k Ω であり, 420N 程度の座面への加圧で約 32 k Ω に変化した.



図 6. (a) インタラクティブチェア. 座面の裏に炭配線が作成されている. 長時間の静的な着座を検出すると, ユーザに通知を行い姿勢の変更を促す. (b) 木製の音量コントローラ. ダイヤルを回すことでスマートスピーカの音量を調節できる.

続的な静電容量値の変化を読み取ることで, スマートスピーカの音量を制御することができる.

6.5 木工部品を使用した木製品との接続

3.3 節で述べたように, 提案手法は多様な木工部品を炭配線のコネクタとして使用できる. そこで, マグネットキャッチ, ヒンジ, L 字金具をコネクタとして使用し, 木製の収納ボックスを実装した(図 7 (a)). これら 3 個の部品は箱の内側の板の境界部分に設置されており, 部品を介して異なる板上の炭配線が接続され, 閉回路が形成されている. ユーザがこの箱を開けると, マグネットキャッチが分離し, 閉回路が切断される. この電気的な断線を Arduino で検出し, 扉が開くと箱内部の LED が自動点灯するアプリケーションを実装した. この構成により, 扉の開閉をネットワーク経由で通知する機能も容易に実装できる. 本手法は配線材と電子部品を使用した回路作成と比較し, 木材に馴染んだ実装が可能であり, PCB のように複数生産にも向いている.

金属製のドアノブによるタッチセンシングを利用し, ドアの開閉を検知することもできる(図 7 (b)). 提案手法を木製のドアに適用することで, ドアに直接センシング機能を付加し, ドアノブを介した自然なタッチインタラクションを実現可能である.

7 議論と結論

ラスタスキャン方式のレーザーを用いてグラフェンを作成する既存技術 [2, 11] に比べ, 提案手法ではベクタスキャン方式のレーザーを用いることで, 大規模な炭による回路を短時間で作成することが可能である. このことを示すために, 合計 1,000 mm 長の炭配線がある 100 mm 四方の回路パターン例

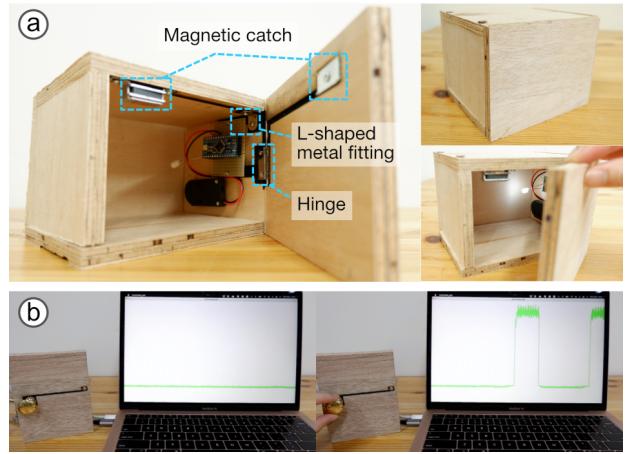


図 7. (a) 木製の収納ボックス. 箱の内面に, マグネットキャッチ, ヒンジ, L 字金具を用いて炭配線の閉回路を作成した. ユーザが扉を開けると閉回路が切断され, 箱内部の LED が自動的に点灯する. (b) タッチセンシングが可能なドアノブ.

を作成する場合に要する時間を計算した⁴. 論文に示されている情報によると, この回路を作製するためのレーザーのラスタスキャンラインの全長は, [2]⁵では 393,700 mm, [11]⁶では 236,200 mm である. [2] では, 導電性が最も高い配線の作成に必要なレーザー速度は 150 mm/s で, 3 回繰り返しスキャンをしている. よって, 回路作成の所用時間は 7,874 秒 ($393,700 / 150 \times 3$) である. また, [11] では, 導電性が最も高い配線の作成に必要なレーザー速度は 10 mm/s でスキャン回数は 1 回である. よって, 所要時間は 23,620 秒 ($236,200 / 10$) である. 提案手法では, 5.2 節で述べたように, 最適なレーザー速度は 381 mm/s で, スキャンは 15 回繰り返される. よって所要時間は 39.4 秒 ($1,000 / 381 \times 15$) となる. したがって, 提案手法はラスタスキャンを用いる既存手法 [2] と比べると 200 倍, [11] と比べると 600 倍の高速化を実現した.

以上のように, 本研究では, CW レーザー光の焦点をぼかし, 繰り返しベクタスキャンを行なう手法を提案し, レーザーによる炭の回路の作成を高速化した. また, DIY で使用される木工部品による接続手法も提案することで, 炭による回路やセンサのファブリケーションが可能であることを示した. さらに, 木材やレーザーパラメタなどの最適条件について調査し, ファブリケーションの際の指針を示した. そして, 多様なセンサや回路を使用した応用例を実装し, 生活空間内の木製品とのインタラクションを実現した. 今後は配線パターンの設計支援ソフトウェアを実装予定である. また, 異なるレーザー加工機での実現可能性, 長期の耐久性, スケラビリティについて調査を行う予定である.

⁴ レーザーヘッドの帰線移動時間などは除く.

⁵ 100 mm 四方を 1000DPI でラスタスキャン.

⁶ 100 mm 四方を 600DPI 以上でラスタスキャン (線幅 40 μm として報告されている).

謝辞

本研究の一部は JSPS 科研費 21J13365 の助成を受けたものである。

参考文献

- [1] V. P. C and D. Wigdor. Printem: Instant Printed Circuit Boards with Standard Office Printers & Inks. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software and Technology*, UIST '15, pp. 243–251, 2015.
- [2] Y. Chyan, R. Ye, Y. Li, S. P. Singh, C. J. Arnsch, and J. M. Tour. Laser-Induced Graphene by Multiple Lasing: Toward Electronics on Cloth, Paper, and Food. *ACS Nano*, 12(3):2176–2183, 2018.
- [3] N.-W. Gong, J. Steimle, S. Olberding, S. Hodges, N. E. Gillian, Y. Kawahara, and J. A. Paradiso. PrintSense: A Versatile Sensing Technique to Support Multimodal Flexible Surface Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pp. 1407–1410, 2014.
- [4] D. Groeger and J. Steimle. ObjectSkin: Augmenting Everyday Objects with Hydroprinted Touch Sensors and Displays. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 1(4):134:1–134:23, 2018.
- [5] D. Groeger and J. Steimle. LASEC: Instant Fabrication of Stretchable Circuits Using a Laser Cutter. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pp. 699:1–699:14, 2019.
- [6] A. Ishii, K. Kato, K. Ikematsu, Y. Kawahara, and I. Siio. Fabricating Wooden Circuit Boards by Laser Beam Machining. In *The Adjunct Publication of the 34th Annual ACM Symposium on User Interface Software and Technology*, UIST '21 Adjunct, pp. 109–111, 2021.
- [7] Ç. Karataş and M. Gruteser. Printing multi-key touch interfaces. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '15, pp. 169–179, 2015.
- [8] Y. Kawahara, S. Hodges, B. S. Cook, C. Zhang, and G. D. Abowd. Instant Inkjet Circuits: Lab-based Inkjet Printing to Support Rapid Prototyping of UbiComp Devices. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '13, pp. 363–372, 2013.
- [9] A. Khan, J. S. Roo, T. Kraus, and J. Steimle. Soft Inkjet Circuits: Rapid Multi-Material Fabrication of Soft Circuits Using a Commodity Inkjet Printer. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, UIST '19, pp. 341–354, 2019.
- [10] K. Klamka, R. Dachselt, and J. Steimle. Rapid Iron-On User Interfaces: Hands-on Fabrication of Interactive Textile Prototypes. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, pp. 93:1–93:14, 2020.
- [11] T.-S. D. Le, S. Park, J. An, P. S. Lee, and Y.-J. Kim. Ultrafast Laser Pulses Enable One-Step Graphene Patterning on Woods and Leaves for Green Electronics. *Advanced Functional Materials*, 29(33):1902771, 2019.
- [12] H. Manabe and W. Yamada. A Capacitive Touch Sensing Technique with Series-connected Sensing Electrodes. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, UIST '17, pp. 645–654, 2017.
- [13] S. Mueller, B. Kruck, and P. Baudisch. LaserOrigami: Laser-Cutting 3D Objects. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pp. 2585–2592, 2013.
- [14] S. Olberding, S. Soto Ortega, K. Hildebrandt, and J. Steimle. Foldio: Digital Fabrication of Interactive and Shape-Changing Objects With Foldable Printed Electronics. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software and Technology*, UIST '15, pp. 223–232, 2015.
- [15] S. Olberding, M. Wessely, and J. Steimle. PrintScreen: Fabricating Highly Customizable Thin-film Touch-displays. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, UIST '14, pp. 281–290, 2014.
- [16] W. Pachler, J. Grosinger, W. Bösch, G. Holweg, K. Popovic, A. Blümel, and E. J. W. List-Kratochvil. A silver inkjet printed ferrite NFC antenna. In *2014 Loughborough Antennas and Propagation Conference (LAPC)*, pp. 95–99, 2014.
- [17] K. Saito, T. Nakamura, K. Kamezawa, R. Ikeda, Y. Hashimoto, and B. Shizuki. Japanese Patterns as NFC Antennas for Interactive Urushiware. In *Proceedings of the Fourteenth International Conference on Tangible, Embedded, and Embodied Interaction*, TEI '20, pp. 443–451, 2020.
- [18] V. Savage, X. Zhang, and B. Hartmann. Midas: Fabricating Custom Capacitive Touch Sensors to Prototype Interactive Objects. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology*, UIST '12, pp. 579–588, 2012.
- [19] M. Schmitz, M. Herbers, N. Dezfuli, S. Günther, and M. Mühlhäuser. Off-Line Sensing: Memorizing Interactions in Passive 3D-Printed Objects. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pp. 182:1–182:8, 2018.
- [20] N. Segawa, K. Kato, and H. Manabe. Rapid Prototyping of Paper Electronics Using a Metal Leaf and Laser Printer. In *The Adjunct Publication of the 32nd Annual ACM Symposium on User Interface Software and Technology*, UIST '19, pp. 99–101, 2019.

- [21] T. Ta, M. Fukumoto, K. Narumi, S. Shino, Y. Kawahara, and T. Asami. Interconnection and double layer for flexible electronic circuit with instant inkjet circuits. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp '15*, pp. 181–190, 2015.
- [22] M. Tsuruta, S. Nakamae, and B. Shizuki. Root-Cap: Touch Detection on Multi-electrodes using Single-line Connected Capacitive Sensing. In *Proceedings of the 2016 ACM International Conference on Interactive Surfaces and Spaces, ISS '16*, pp. 23–32, 2016.
- [23] U. Umapathi, H.-T. Chen, S. Mueller, L. Wall, A. Seufert, and P. Baudisch. LaserStacker: Fabricating 3D Objects by Laser Cutting and Welding. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software and Technology, UIST '15*, pp. 575–582, 2015.
- [24] Y. Wang, J. Zhou, H. Li, T. Zhang, M. Gao, Z. Cheng, C. Yu, S. Patel, and Y. Shi. Flex-Touch: Enabling Large-Scale Interaction Sensing Beyond Touchscreens Using Flexible and Conductive Materials. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 3(3):109:1–109:20, 2019.
- [25] M. Wessely, T. Sethapakdi, C. Castillo, J. C. Snowden, O. Hanton, I. P. S. Qamar, M. Fraser, A. Roudaut, and S. Mueller. Sprayable User Interfaces: Prototyping Large-Scale Interactive Surfaces with Sensors and Displays. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, CHI '20*, pp. 122:1–122:12, 2020.
- [26] R. Ye, Y. Chyan, J. Zhang, Y. Li, X. Han, C. Kittrell, and J. M. Tour. Laser-induced graphene formation on wood. *Advanced Materials*, 29(37):1702211, 2017.
- [27] Y. Zhang, G. Laput, and C. Harrison. Elec-trick: Low-Cost Touch Sensing Using Electric Field Tomography. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, CHI '17*, pp. 1–14, 2017.
- [28] C. Zheng, H. Oh, L. Devendorf, and E. Y.-L. Do. Sensing Kirigami. In *Proceedings of the 2019 on Designing Interactive Systems Conference, DIS '19*, pp. 921–934, 2019.

ペイント可能なタイムラインによる対話的なハイパーパラメータ最適化

樋口啓太^{*} 佐野 正太郎^{*} 五十嵐 健夫^{†*}

概要. 本論文では、ハイパーパラメータの自動最適化システムにインタラクティブ性を持たせ、ユーザのパラメータ分布に関する事前知識を活用する手法を提案する。本手法では、ユーザは自動最適化の進捗状況を継続的に観察し、パラメータ空間のどこを探索するかを動的に指定する。本手法の実現可能性を示すために、最適化システムのためのインタラクティブダッシュボードのプロトタイプを実装した。インタラクティブなダッシュボードの主な特徴は「ペイント可能なタイムライン」であり、ユーザは標準的なタイムラインのように試行が行われた過去のパラメータ値を観察するだけでなく、簡単なペイント操作でパラメータの選択範囲を指定することができる。機械学習の専門家への聞き取りに基づき設計した、ユーザの介入によりパフォーマンスが向上すると考えられる3つのパラメータ最適化問題の例を紹介する。また、専門家を対象とした実験を実施し、対象となる問題のパラメータ分布に関する事前知識があれば、インタラクティブな最適化は完全な自動最適化と比較して良い結果を得ることができることを示した。

1 はじめに

ハイパーパラメータの最適化は、機械学習を用いたアプリケーションの開発において重要な作業である。ディープニューラルネットワーク(DNN)のハイパーパラメータを最適化するには、数時間から数日かかることがあるため、この作業を高速化する手法があれば、実用上大きな価値がある。従来の最適化手法は、手動最適化と自動最適化の2つのアプローチに分けられる。自動最適化によって、開発者は試行錯誤から解放される。しかし、機械学習技術の専門家への聞き取り調査では、いくつかの理由により、自動最適化が手動最適化に完全に取って代わるには至っていないことがわかった[12]。第一に、パラメータ最適化では、最適化対象のパラメータ選定や探索範囲など探索空間が大きくなる。専門家がパラメータ最適化をより効率的に行うためには、対象となる問題のパラメータ分布に関する独自のドメイン知識を活用する必要がある。また、自動最適化を行うためには、対象となる探索空間が明確に定義されている必要がある。しかし、現実の最適化問題では、最適化結果の質、計算時間、計算コストなど、複数のトレードオフが存在することが多く、適切な探索空間が探索前に明確に定義されているとは限らない。そのため、開発者は探索を実行した後、中間結果を観察して探索空間を修正する必要があることが多い。

このような自動最適化の限界を解決するために、本研究では自動最適化システムにインタラクティブ性を取り入れた対話型ハイパーパラメータ最適化を提案する。このアプローチは、ユーザがハイパー

パラメータ最適化プロセスを観察したり介入したりすることで、パラメータ分布に関するユーザのドメイン知識を活用することを目的としている。インタラクティブ最適化では、ユーザは自動最適化の進行状況を観察し、それに基づいて探索空間(どのパラメータ範囲を探索するか)を動的に変更できる。本研究では、対話型最適化の実現可能性と有効性を示すために、対話型ダッシュボードというプロトタイプを開発した。このダッシュボードは、最適化システムのプラウザベースのフロントエンドとして機能する。このシステムの主な特徴は、「ペイント可能なタイムライン」である。このタイムラインは、既存の最適化システムで最適化プロセスの進捗状況を示す典型的な可視化に似ているが、それに加えて探索空間をリアルタイムに設定できる機能を備えている。ユーザは、対象となる問題に関する事前知識があれば、過去の試行結果から最適解が存在する可能性の高い領域を推定し、今後の試行ではその領域に注目するようシステムに指示することができる。その結果、システムは有望な探索空間により多くの時間を費やすことができ、最終的な最適化結果をより向上させることができ得る。

2 関連研究

2.1 自動パラメータ最適化のフレームワーク

ハイパーパラメータ最適化とは、目的関数の値を最大化する最適なハイパーパラメータを求める問題である。自動最適化システムでは、パラメータのサンプリング戦略(ベイジアン、ランダム、進化計算など)を用いてユーザが定義したパラメータ空間を探索する[22, 4, 5]。また、近年の自動最適化システムでは、最適化コードを記述することなく、目標パラメータと目的関数を定義するだけで最適化を実行で

* Preferred Networks

† 東京大学

きるフレームワークが提供されている [1, 11, 6]. これらのフレームワークは、機械学習アプリケーションの開発だけでなく、シミュレーションや材料解析などにも利用されている [7, 20, 16, 19] しかし、これらのフレームワークでは、探索中のユーザの介入が制限されている。本研究では、探索中にユーザがパラメータ探索範囲を変更できるようにし、専門家のドメイン知識を活用して探索を効率化する手法を実現した。

2.2 AutoMLへの人間の介入

AutoML とは機械学習モデルの構築や学習を自動化しようとするものである。AutoML は、機械学習の専門家でなくても利用できることを目指して開発されており、先行研究では、AutoML システムへの人間の介入方法が検討されている。Gil らは Human-Guided 機械学習を提案し、それに適したタスクや要件を検討している [10, 9]. Santos らは、機械学習に不慣れな専門家のための対話型 AutoML システム Visus を開発した [21]. また、AutoML システムのヒューマンファクタを調査した結果、可視化の重要性やユーザと AI の協調の重要性が明らかになった。本研究では、ハイパーパラメータ最適化に着目し、そのプロセスに人間が介在する方法を提案している。

2.3 インタラクティブビジュアライゼーション

本研究は、機械学習システムの制御のために設計された対話的な可視化技術にヒントを得ている。ManiMatrix では、ユーザが混同行列を直接操作して分類器の内部パラメータを調整することができる [14]. Dis-function は、ユーザが散布図を直接操作して、距離関数を調整することができる [8]. 彼らは、本研究と同様に、システムの監視と制御のために可視化を使用している。また、既存のシステムがスナップショットの制御を主目的としているのに対し、本研究ではタイムライン型の可視化を活用することで、過去の試行のモニタリングと未来の試行の制御をシームレスに統合するアイデアを提案している。

本研究は Klau らによる Human-Guided Search の影響も受けており [15], HuGS では最適化問題への人間の介入の重要性を強調し、組み合わせ最適化を解くための視覚的なメタファーを提供した。彼らは、そのアプローチの有効性を 4 つの事例（グラフレイアウト、経路探索、タンパク質フォールディング、タスクスケジューリング）で実証した。本研究は同じ最適化への人間の介入に着目しているが、ハイパーパラメータ（ブラックボックス）最適化のための技術を提案している。

ペイントやスケッチは、データとのインタラクションによく使われる。Baudisch は時間間隔を指定するためにペイントを使用した [3]. Hochheiser と

Shneiderman はラバーバンドを、Mannino と Azza はフリーハンドのスケッチを時系列データへのクリエイとして用いている [13, 17]. Mannino と Azza は人工的なデータの合成にもスケッチを用いている [18]. Asai らは散布図へのスケッチとペイントをデータ合成と編集に用いている [2]. 本研究の提案インタラクションは、過去の最適化結果を観察しながら、未来のパラメータ探索範囲を制御するためにペイントを使用するという点で新規性がある。

3 提案システム

本研究では、専門家へのインタビューを行い、対話型パラメータ最適化の要件と望ましい機能を明らかにした [12]. その結果、パラメータの範囲に関する専門家の予備知識は重要であるが、それを活用して効率的なパラメータチューニングを実現することは容易ではないことがわかった。

このセクションでは、専門家へのインタビューで明らかになった完全自動最適化の限界を解決するために設計したインタラクティブダッシュボードについて説明する。専門家は、パラメータ分布に関する事前の知識を活用することが重要であると指摘したが、既存のフレームワークではパラメータ探索範囲を動的に変更する機能が限られていた。そこで本研究では、ハイパーパラメータ最適化フレームワークに対話的な機能を導入する。この本提案は、ユーザが最適化の進捗状況を観察するだけでなく、最適化プロセスをその場で制御することができる、対話的な最適化をサポートしている。ユーザは、探索の詳細なスケジュール（パラメータ空間のどの領域をいつ探索する必要があるか）を探索前に設定することができ、また、探索中にスケジュールを動的に変更することもできる。

3.1 可視化

インタラクティブダッシュボードは、ハイパーパラメータチューニングシステムのフロントエンドとして機能する。現在の実装は、特定のハイパーパラメータ探索フレームワーク（Optuna）用に設計されているが、基本的なアイデアは他のフレームワークにも適用できるはずである。図 1 に本システムのスクリーンショットを示す。主に、ノブ付きのスケジュールパネル、スコアパネル、パラメータパネルで構成されている。スコアパネルは、一般的なダッシュボードのタイムラインビューに似ている。過去の試行のスコアを可視化し、最適化の進捗状況を示す。スケジュールパネルは、過去と未来の最後までの試行回数を表示する。ノブを使って、ハイパーパラメータ最適化の試行回数を設定・追加することができる。パラメータパネルは、ペイント可能なタイムライン機能を持つ本システムの主要部分である。過去にどのようなパラメータ値が検討され、将来ど

ペイント可能なタイムラインによる対話的なハイパーパラメータ最適化

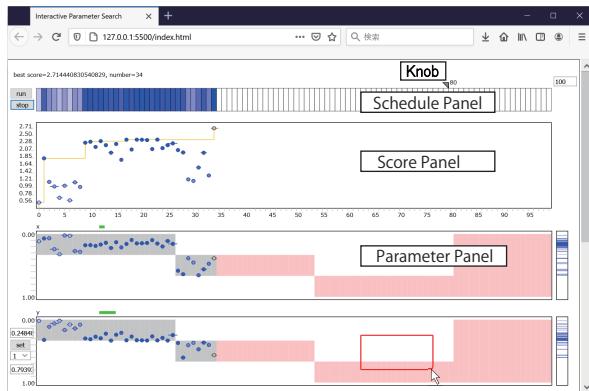


図 1. プロトタイプシステムのスクリーンショット。

のような値が検討されるかを可視化する。また、将来的にどのようなパラメータ値を検討しなければならないかを簡単に設定することができる。また、パラメータパネルには、集約されたビュー（右）とパラメータの推定重要度（左上）が表示される。パラメータパネルでは、高次元のパラメータも扱うことができる。各次元（パラメータ）は、下部の水平パネルで表現され、任意の数の次元を追加することができる。

ペイント可能なタイムラインは、左から順に、過去、現在、未来の試行を表示する。過去のトライアルでは、テストされたパラメータ値とそのスコアが可視化される。また、現在のトライアルでは、テストされているがスコアのないパラメータ値が可視化される。トライアルのスコア、パラメータの範囲、選択された値は、縦方向に配置される。スコアパネルはスコアを、パラメータパネルは選択されたパラメータを、それぞれ円で表現する。例えば、図 1 では、最初のトライアルで、スコアが 0.56、パラメータ x が 0.01、パラメータ y が 0.00 であることを示している。円の色は、各試行の相対的なスコアを示しており、色が濃いほどスコアが良いことを示している。最良のスコアを得た試行の結果は、黄色の円で表される。未来の試行では、検討すべきパラメータ値の範囲をピンクの帯で可視化している。スコアとパラメータのパネルでは、円の後ろにある短い線が、相対的な一試行あたりの計算時間を示している。

3.2 ユーザインタラクション

ペイント可能なタイムラインは、パラメータ探索範囲を可視化するだけでなく、ペイント操作により探索範囲を変更できる。現在の実装では、3種類のペイント操作をサポートしている。一つ目は、空白領域から始めるラバーバンド操作である（図 2 左）。ユーザがタイムライン上でドラッグ操作を開始すると、システムはラバーバンドを表示する。ユーザは、

輪ゴムの境界をドラッグするか、下限値と上限値を数値で入力して、ボックスの位置と大きさを調整する。次に、set ボタンを押すと、探索するパラメータの範囲を設定することができる。また、グリッド数を指定することで、指定した領域を分割してグリッド探索を行うこともできる。2つ目の操作は、パラメータ範囲の境界をドラッグすることである（Fig 2 中）。境界線にマウスカーソルを合わせると、境界線がハイライトされる。ユーザはマウスを下にクリックして境界をつかみ、任意の場所にドラッグする。その後、ユーザは“set”を押して、パラメータの範囲を更新する。3つ目は、過去の試行の選択である（図 2 右）。ユーザは、ラバーバンドを使って1つまたは複数の試行を選択する。すると、システムは、選択された試行から下界と上界をとったラバーバンドを表示する。これは、過去の試行に基づいて、スコアの高い領域や低い領域を指定することができる。

ペイント可能なタイムラインは、ユーザのドメイン知識や判断を活用して、探索を効率化するように設計されている。図 3 に提案システムを用いたパラメータ最適化の様子を示す。まず、最適化の際に、ユーザはペイント可能なタイムラインを利用して探索空間を徐々に狭めていくことができる。過去の試行結果やスコアの推定結果を見ながら、どの範囲を探索すれば効率的な探索ができるかを判断することができる。これは、分布が滑らかであり、与えられた範囲内に単一の最適値があることをユーザが知っている場合に有効である。次に、パラメータ値を変えてみる必要がなくなったと判断した場合には、探索中にパラメータ値を固定することができる。このようにして、自動最適化は他のパラメータに集中することができる。3つ目は、ユーザが、どのパラメータ空間をいつ探索するかを指定することで、今後の探索のスケジュールを立てることである。例えば、ある特定のパラメータ値の組み合わせが有望であることがわかっている場合（例：パラメータ X が高いときはパラメータ Y も高いはず），ペイント可能なタイムラインを使って特定の領域を探索させることもできるし、半自動のグリッド探索に使うことができる。ユーザは、ペイント可能なタイムラインを使って、グリッドセルを指定する。すると、システムは各グリッドセル内を探索する。これは、一様に探索を行う必要があることがわかっている場合に有効である。第 4 に、スコアパネルの過去の試行をラバーバンドで選択し、選択した試行のパラメータの範囲を設定することができる。これは、数回の試行を経て、有望なパラメータの範囲に探索を集中させるのに有効である。

3.3 実装

本研究では、ブラックボックス最適化フレームワークである Optuna [1] を用いてプロトタイプシステ

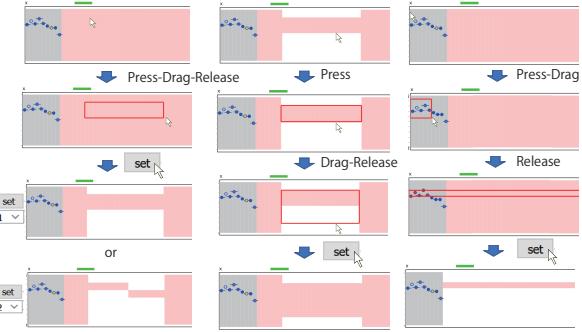


図 2. 実装されているペイント操作. 左: 矩形を描く. 中: 境界を移動する. 右: 過去の試行を選択する.

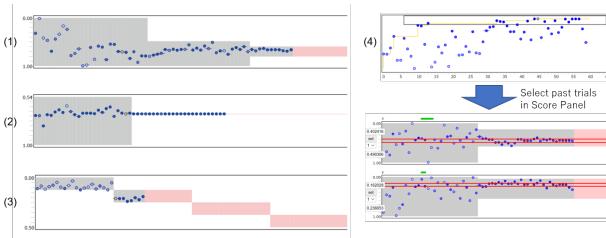


図 3. (1) 探索空間を狭める (2) パラメータの固定 (3) グリッド探索 (4) 過去試行の選択

ムを実装した。このフレームワークでは、ハイパーパラメータ候補を自動的にサンプリングするための複数のアルゴリズムを Sampler としてサポートしており、これを拡張してユーザ定義のサンプリングアルゴリズムを実装することができる。本研究ではインタラクティブなハイパーパラメータ最適化のバックエンドシステムとして、*interactive sampler*を開発した。フロントエンドとバックエンドは WebSocket 介してパラメータ範囲や最適化結果などを通信する。この実装では、ユーザがインタラクティブハイパーパラメータ最適化を利用するため、オリジナルのコードを大幅に変更する必要はないという利点がある。

4 利用シナリオ

このセクションでは、インタラクティブな最適化の利用例を3つ紹介する。専門家へのインタビューで語られた代表的な事例をもとに例題を作成した。スコア分布の事前知識を用いた対話的な最適化が完全な自動探索よりも優れていると考えられる3つのケースをピックアップし、事前知識を活用して探索を誘導するためのペイント可能なタイムラインの使用方法を説明する。3つの例とも、操作画面のスクリーンショットは実際のインタラクティブな最適化の結果を示している。

ここでは、ユーザが対象となる最適化問題のスコ

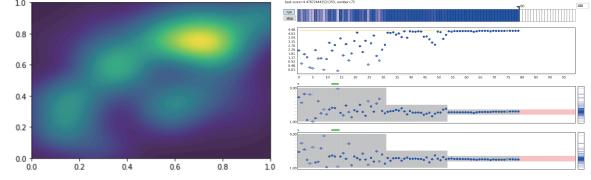


図 4. 例 1: 対角線上の分布

ア分布（最適解がどこに分布しているか）について何らかの事前知識を持っていることを前提としている。この事前知識は、問題の構造に関する既存のドメイン固有の知識から得られるものと、類似の問題に対する過去の最適化の経験から得られるものがある。また、ユーザは、スコア分布は共通しているが、実際の最適スコアやそのパラメータ値が異なる類似の問題に繰り返し取り組むことを想定している。ユーザは、インタラクティブな最適化の方法を徐々に学び、その経験を活用して、より効率的な探索を実現する。

4.1 例 1: 対角線上の分布

最初の例では、パラメータ間に相関があり、最適解が対角線領域の近くに位置する場合を示している。図 4). この例ではパラメータ x が小さければ、パラメータ y も小さくなるはずであり。パラメータ x が大きければ、パラメータ y も大きいはずであるという事前知識がある。この例は、DNN の学習では、学習率とバッチサイズに相関関係があることが知られていることから考案した（学習率が大きいときは、バッチサイズも大きくする必要がある）[23]。探索空間を対角線付近に限定することで、効率的な探索を行うことができる。図 4 にスコア分布とインタラクティブ最適化の結果を示す。ユーザはまず、自動最適化に任せてパラメータ空間全体を探索する。しばらくして、30回目くらいの試行で、 x と y の両方が高いところでは、おおむねスコアが高いことがわかる。続いて、ペイント可能なタイムラインを使って探索空間を絞り込む。スコアの高いサンプルが集中していることを観察しながら、さらに探索空間を絞り込むことができる。その結果、自動最適化よりも対話型最適化の方が、最も有望な領域を効率的に探索することができるため、高いスコアが得られることがある。

4.2 例 2: 支配的パラメータ

2番目の例は、パラメータ x が支配的で、パラメータ y のスコアへの影響が限定的な場合である（図 5）。この場合、まず最適なパラメータ x を探索し、次にパラメータ x を固定して最適なパラメータ y を探索するのが最適な戦略となる。図 5 にスコア分布と対話型最適化の結果を示す。ユーザはま

ペイント可能なタイムラインによる対話的なハイパーパラメータ最適化

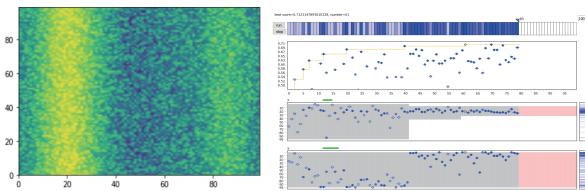


図 5. 例 2: 支配的パラメータ

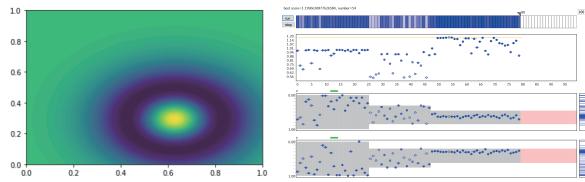


図 6. 例 3: 孤立領域

ず、自動最適化に任せてパラメータ空間全体を探索する。しばらくして、40回目くらいの試行で、あるパラメータ x の値でスコアが全般的に高くなることが明らかになった。その結果、自動最適化よりも対話型最適化の方が、有望でない領域を無駄に探さずに済むため、高いスコアが得られる可能性がある。

4.3 例 3: 孤立領域

3つ目の例は、最高スコアが低スコア領域の中央に位置するケースである（図 6）。このような場合、自動最適化では、低スコアの周辺にある高スコアの領域（平地）に焦点を当て、低スコアの領域を無視することが多いため、特に難しい。しかし、このような予備知識があれば、意図的に低スコア領域を探索することで最適解を得ることができる。図 6 にこの例のスコア分布と対話型最適化の結果を示す。ユーザはまず、自動最適化にパラメータ空間全体を探索させる。しばらくして、試行回数 30 回目くらいになると、ある領域でスコアが全体的に低くなっている。そこで、ペイント可能なタイムラインを使って、スコアの低い領域を探索するように指示する。最終的には、孤立した島を見つける。これは、完全な自動探索では難しいことである。

5 ユーザ実験

本研究では提案手法であるペイント可能なタイムラインを用いたインタラクティブな最適化の効果を検証するために、ユーザ実験を実施した。本実験の目的は、スコア分布に関する事前知識を用いたインタラクティブな最適化が、いくつかの特定ケースにおいて自動最適化よりも優れているかどうかを確認することである。

5.1 仮定

ユーザ実験のシナリオとして、2つの重要な前提を置いている。第一に、ユーザは対象となる問題タイプにすでに精通しており、何度か取り組んだことがあると仮定する。具体的には、スコアの全体的な分布特性や、効率的な探索のための対話型最適化の際の探索範囲の絞り込み方などをすでに知っている。理想的には、研究参加者に何度もその問題タイプに取り組んでもらい、自分で戦略を見つけてもらうのがベストである。しかし、ユーザ実験としてはかなりの時間を要するため、分布と探索戦略を参加者に教えている。

次に、探索には数時間から数日といった相応の時間がかかる想定している。自動探索で数秒から数分で最適解が見つかるのであれば、ユーザの介入は必要ないかもしれない。インタラクティブな最適化が有効になるのは、全体の探索時間が対話時間よりもはるかに長い場合である。理想的には、最適化に数時間から数日かかるような問題を研究対象とするのがベストである。しかし、ユーザ実験としてはかなりの時間を要するため、1分以内に終わるような小さな最適化問題を大きな問題の代用としている。参加者には、何時間も何日もかかる問題が前提なので、ユーザとの対話の時間を最短にしようとすると必要ないと伝えている。より良いスコアを見つけることは、インタラクションの時間を短くすることよりも重要であると伝えた。

5.2 実験条件

この研究には 7 人の機械学習の専門家が参加した。参加者の数が限られているのは、手動および自動最適化を用いたハイパーパラメータチューニングに十分な経験を持つ専門家が必要だからである。参加者は全員、著書の所属する社内から協力を得たエンジニアもしくはリサーチャーである、7 名のうち 5 名は仕事の一環としてハイパーパラメータチューニングに取り組んでいる。彼らは全員、事前の専門家インタビューに参加している。すべてのユーザ実験のセッションは、オンラインで行われた。各参加者は、3 種類の問題からなる 15 のインタラクティブ最適化タスクに取り組み、それぞれの問題には 5 つのバリエーションがあった。この 3 つのタイプは、4 章で示した例から取られている。各タイプの問題には、ランダムに 5 つのバリエーションを用意し、参加者全員に同じセットを使用した。参加者が与えられたタスクを完了するのに約 1 時間かった。

まず、参加者に提案システムの使い方を説明し、UI 利用の練習をしてもらった。その後、問題の種類ごとに 1 つずつ取り組んでもらった。問題の種類ごとに、探索の誘導方法をデモで説明し、参加者は練習として同じ問題に取り組んだ。説明は口頭で行い、スコア分布の可視化は見せなかった。

参加者は説明を受けた後、メインタスクとして15の課題に取り組んだ。参加者は課題中にいつでも説明を読むことができた。5つのバリエーションは、一般的な特徴は共通しているが、実際の最適なスコアや場所は参加者にはわからないようになっている。そのため、同じ探索戦略を用いても、試行の結果を見て、探索範囲をどこに設定するかを決める必要がある。

最適化の試行回数（パラメータのサンプリング回数）80回の内で最高スコアを獲得してもらい、その結果を提案手法の interactive 条件とした。7人の実験参加者が、1つの問題タイプにつき5つのバリエーションに取り組むため、3種類のタイプそれぞれに35の最適化結果が得られた。また、ベースラインとして、TPEアルゴリズムによりパラメータ探索を行った auto 条件を採用した。80回イテレーションの自動最適化を5つのバリエーションごとに7回行っているので、interactive 条件の場合と同様に、3種類のタイプごとに35個のスコア結果が得られた。両条件とも最初は同じパラメータ範囲を使用しているが、対話型最適化では探索中に参加者が自由にパラメータ範囲を変更した。参加者には、最初の範囲内にベストスコアがあることを伝えた。また、十分に細かいグリッド (1000×1000) でグリッドサーチを行い、その結果を真の最適値とした。手動探索と自動探索の性能は、それぞれのスコアを対応する真の最適値で割ることで測定する。また、interactive 条件では、80試行中の介入時間を測定した。

5.3 結果

図7は、ユーザ実験の最適化結果（得られたスコア/最適値）を示している。1つ目のタイプでは、auto 条件と interactive 条件の平均性能は、それぞれ 0.931 (SD: 0.134), 0.964 (SD: 0.104) となっている。2つ目のタイプでは、auto 条件と interactive 条件の平均性能は、それぞれ 0.990 (SD: 0.007), 0.993 (SD: 0.005) であった。3つ目のタイプでは、auto 条件と interactive 条件でそれぞれ 0.859 (SD: 0.044), 0.940 (SD: 0.070) となった。Shapiro-Wilk 検定ではすべての結果に正規性が認められなかったため、Mann-Whitney U 検定を用いて両条件での有意性を観察した。また、多重比較のために Benjamini-Hochberg 法を用いて p 値を調整した。検定の結果、すべての機能タイプで条件間に有意な差が見られた。interactive 条件における介入時間の平均は、第1, 第2, 第3のタイプで、それぞれ 100.8 秒 (SD: 58.0), 54.7 秒 (SD: 30.6), 114.9 秒 (SD: 129.6) であった。

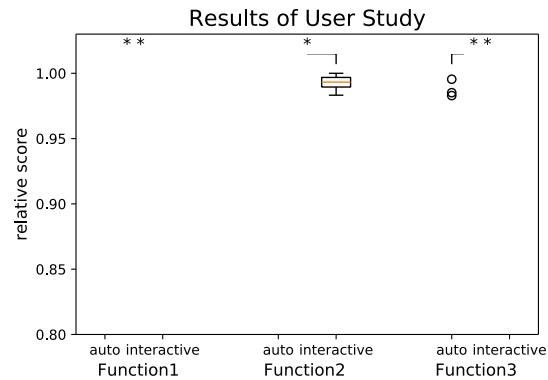


図 7. 実験結果。*($p < 0.05$) および **($p < 0.001$) は Mann-Whitney U test における統計的有意差を示す。ここに示す相対スコアは、最適スコアをグリッドサーチで探した最適値で割ったものである。

5.4 議論

ユーザ実験の結果、3つの具体的なケースにおいて、適切な探索戦略を用いたインタラクティブな最適化が自動最適化を上回ることが実証された。すべてのタスクにおいて、対話型の interactive 条件は自動最適化の auto 条件よりも高い平均性能を達成した。利用シナリオで説明したように、3つの機能タイプでは、事前のドメイン知識が役立つことが予想された。特に、1番目と3番目のタイプは、局所最適領域が大きいため、完全な自動最適化による探索には難しい問題だった。参加者は、説明から得た知識を利用して、パラメータの探索範囲を制御することに成功したことが観察された。このように、ユーザ実験では我々の予想が確認された結果となった。今回の結果は、インタラクティブ最適化が多くの問題において有効だと示したものではない。今後より広範な実験によりインタラクティブ最適化がどのような問題に有用であるかを確認したい。

6 むすび

本研究では、対話型のハイパーパラメータ最適化を提案し、インタラクティブダッシュボードのプロトタイプを開発した。機械学習の専門家へのインタビューをもとに、最適化の進行状況を観察したり、ペイント操作でパラメータ探索範囲に介入できる機能を設計した。また、インタラクティブな最適化が効果的であると思われる3つの例を示した。インタラクティブなハイパーパラメータ最適化の実現可能性を示すために、7人の専門家を対象にユーザ実験を実施した。その結果、3つの具体的な事例において、適切な探索戦略を用いた対話型最適化が完全な自動最適化よりも優れていることが示された。

参考文献

- [1] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama. Optuna: A Next-Generation Hyperparameter Optimization Framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '19, p. 2623–2631, New York, NY, USA, 2019. Association for Computing Machinery.
- [2] K. Asai, T. Fukusato, and T. Igarashi. Integrated Development Environment with Interactive Scatter Plot for Examining Statistical Modeling. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, p. 1–7, New York, NY, USA, 2020. Association for Computing Machinery.
- [3] P. Baudisch. Don't Click, Paint! Using Toggle Maps to Manipulate Sets of Toggle Switches. In *Proceedings of the 11th Annual ACM Symposium on User Interface Software and Technology*, UIST '98, p. 65–66, New York, NY, USA, 1998. Association for Computing Machinery.
- [4] J. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl. Algorithms for Hyper-Parameter Optimization. In *Proceedings of the 24th International Conference on Neural Information Processing Systems*, NIPS'11, p. 2546–2554, Red Hook, NY, USA, 2011. Curran Associates Inc.
- [5] J. Bergstra and Y. Bengio. Random Search for Hyper-Parameter Optimization. *J. Mach. Learn. Res.*, 13(null):281–305, Feb. 2012.
- [6] J. Bergstra, B. Komera, C. Eliasmith, D. Yamins, and D. D. Cox. Hyperopt: a Python library for model selection and hyperparameter optimization. *Computational Science & Discovery*, 8(1):014008, jul 2015.
- [7] J. Bergstra, D. Yamins, and D. D. Cox. Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures. ICML'13, p. I–115–I–123. JMLR.org, 2013.
- [8] E. T. Brown, J. Liu, C. E. Brodley, and R. Chang. Dis-function: Learning distance functions interactively. In *2012 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 83–92. IEEE, 2012.
- [9] Y. Gil, J. Honaker, S. Gupta, Y. Ma, V. D’Orazio, D. Garijo, S. Gadewar, Q. Yang, and N. Jahanshad. Towards Human-Guided Machine Learning. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*, IUI ’19, p. 614–624, New York, NY, USA, 2019. Association for Computing Machinery.
- [10] Y. Gil, K.-T. Yao, V. Ratnakar, D. Garijo, G. V. Steeg, P. Szekely, R. Brekelmans, M. Kejriwal, F. Luo, and I.-H. Huang. P4ML: A Phased Performance-Based Pipeline Planner for Automated Machine Learning. In *Proceedings of Machine Learning Research, ICML 2018 AutoML Workshop*, 2018.
- [11] D. Golovin, B. Solnik, S. Moitra, G. Kochanski, J. Karro, and D. Sculley. Google Vizier: A Service for Black-Box Optimization. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD ’17, p. 1487–1495, New York, NY, USA, 2017. Association for Computing Machinery.
- [12] K. Higuchi, S. Sano, and T. Igarashi. Interactive Hyperparameter Optimization with Paintable Timelines. In *Designing Interactive Systems Conference 2021*, DIS ’21, p. 1518–1528, New York, NY, USA, 2021. Association for Computing Machinery.
- [13] H. Hochheiser and B. Shneiderman. Dynamic Query Tools for Time Series Data Sets: Timebox Widgets for Interactive Exploration. *Information Visualization*, 3(1):1–18, 2004.
- [14] A. Kapoor, B. Lee, D. Tan, and E. Horvitz. Interactive Optimization for Steering Machine Classification. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’10, p. 1343–1352, New York, NY, USA, 2010. Association for Computing Machinery.
- [15] G. W. Klau, N. Lesh, J. Marks, M. Mitzenmacher, and G. T. Schafer. The HuGS Platform: A Toolkit for Interactive Optimization. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, AVI ’02, p. 324–330, New York, NY, USA, 2002. Association for Computing Machinery.
- [16] T. Ko. A Tendon-Driven Robot Gripper With Passively Switchable Underactuated Surface and its Physics Simulation Based Parameter Optimization. *IEEE Robotics and Automation Letters*, 5(4):5002–5009, 2020.
- [17] M. Mannino and A. Abouzied. Expressive Time Series Querying with Hand-Drawn Scale-Free Sketches. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI ’18, p. 1–13, New York, NY, USA, 2018. Association for Computing Machinery.
- [18] M. Mannino and A. Abouzied. Is This Real? Generating Synthetic Data That Looks Real. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, UIST ’19, p. 549–561, New York, NY, USA, 2019. Association for Computing Machinery.
- [19] K. Oono and T. Suzuki. Graph Neural Networks Exponentially Lose Expressive Power for Node Classification. In *International Conference on Learning Representations*, 2019.
- [20] Y. Ozaki, Y. Suzuki, T. Hawaii, K. Saito, M. Onishi, and K. Ono. Automated crystal structure analysis based on blackbox optimisation. *njp Computational Materials*, 6(1):1–7, 2020.
- [21] A. Santos, S. Castelo, C. Felix, J. P. Ono, B. Yu, S. R. Hong, C. T. Silva, E. Bertini, and J. Freire. Visus: An Interactive System for Automatic Machine Learning Model Building and Curation. In *Proceedings of the Workshop on Human-In-the-Loop Data Analytics*, HILDA’19, New York, NY, USA, 2019. Association for Computing Machinery.

- [22] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. De Freitas. Taking the human out of the loop: A review of Bayesian optimization. *Proceedings of the IEEE*, 104(1):148–175, 2015.
- [23] S. L. Smith, P.-J. Kindermans, C. Ying, and Q. V. Le. Don’t Decay the Learning Rate, Increase the Batch Size. In *International Conference on Learning Representations*, 2018.

液体噴霧混合式の味ディスプレイの試作

宮下芳明*

概要. 本稿では、液体を噴霧混合する味ディスプレイを試作した。塩化ナトリウムやスクロースなど、味を提示する液体が入ったタンク及びスプレーを 10 基搭載しており、透明なフィルム上に噴霧する。フィルムは手前に巻き取られ、その下には液晶画面で料理コンテンツが再生されている。これにより、画面を直接舐めることで味を体験できる「味ディスプレイ」を、電気味覚的な技術を一切利用せずに実現できた。使用する液体は電解質である必要がなくなったため、渋味、辛味、アルコール味、脂肪味など、これまで電気味覚技術で再現できなかった味の再現の可能性を秘めており、舐めたときのみならず、飲み込むまで味が継続される。混合液と再現したい料理の味の差を、味センサ自体で再測定することで、味の再現がうまくいったかどうかを評価できるメリットもある。

1 はじめに

筆者は、味を再現するイオン泳動式の味ディスプレイ Norimaki Synthesizer[1]を開発し、それを味センサと連動させ[2]、画面を舐めることで視聴覚のみならず味も追体験できる味ディスプレイとして提案した[3]。これは 5 本のゲル中に基本五味を呈する電解質が溶かされ、電極が刺さったもので、身体に電極を接触させながらゲルを舌に当てるとき、舌に触れるそれぞれの電解質の量を変えることができ、基本五味のバランスを調整して味を作り出すことができる。味センサで取得した値に基づいて制御することで、飲食物の味を、味蕾にある五種の受容体によって知覚される狭義の味覚の範囲で再現して味わうことができるようになった。(狭義と書いたのは、痛覚である辛味や、収斂作用である渋味、ときには、嗅覚である香りすらも一般社会において「味覚」とよばれることがあるからである)。

本稿は、それと同じく味センサと連動する味ディスプレイに関する研究である。しかし、電気味覚どころかイオン泳動も用いず、単に液体を噴霧混合して再現する仕組みとして提案するものである。類似のコンセプトとしては Edible bits[13] や SUSHI TELEPORTATION[14]がある。本稿は UIST2021 でポスター発表予定の原稿[15]に、説明や議論をより加えたものである。

Copyright is held by the author(s).

* 明治大学

2 提案システム

2.1 機構

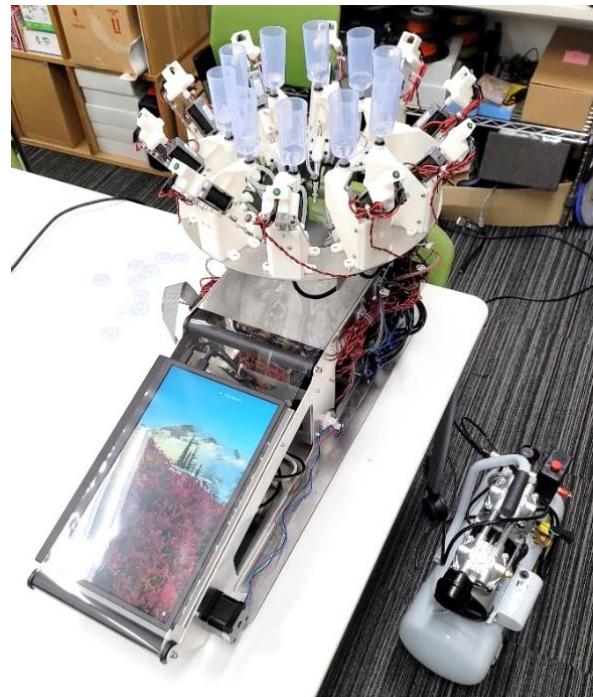


図 1. 提案する味ディスプレイ及びエアコンプレッサ

図 1 のように、提案する味ディスプレイは、横幅 420mm、高さ 500mm、奥行き 800mm のサイズで重量は約 20Kg である。システム上部には 10 個の噴霧装置がついており、透明 OPP シートに液体を噴霧・混合する。このシートはロール状になっており、手前に巻き取られていく。シートの下には液晶画面があり、映像が透けて見える。画面を舐めると、

シートに噴霧された混合液の味を感じることができ、映っているコンテンツに応じて違う味を感じることができ。これにより、画面に映っている食べ物の味を、画面を直接なめることで感じられるような体験を提供できる。図右下にあるのは空気圧を送るためにエアコンプレッサ(EARTH MAN ACP-13SLA)である。圧力がたまつていれば、味ディスプレイ動作時に作動させる必要はない。

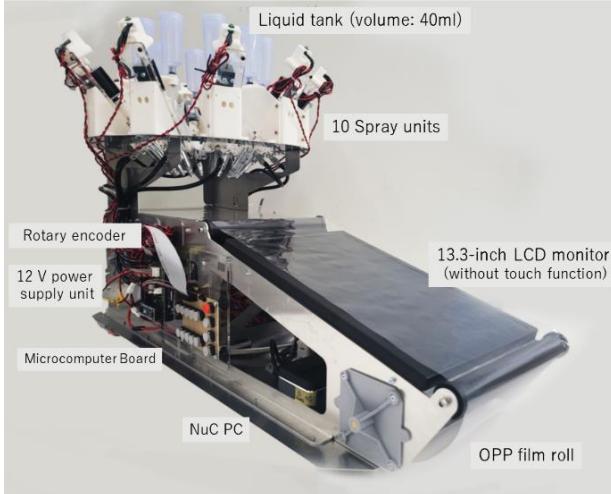


図 2. 味ディスプレイ側面図

図 2 はシステムを横から見たときの側面図である。カセットテープのように、送り出しと巻き戻し用のモータがあり、また、ロータリーエンコーダも搭載することでその送り・戻し量を得ることができる。側面にはスタンダードアロンでも動作・調整できるボタンがついており、シートの送り・戻し、各噴霧ユニットからの出力に加え、全噴霧ユニットからの出力ボタンなどがある。(これは特に、メンテナンスに有用であり、ぬるま湯を入れて出しちゃなしにすることで噴霧ユニット内を洗浄するときに用いる。)

図 3 は噴霧ユニットである。エアブラシとソレノイドを接合させた構造となっている。タンク容量は 1 種類につき 40ml である。エア圧を固定し、ソレノイドで噴霧時間を制御することによって吐出量を調節している(最小吐出量は 0.01ml 以下である)。10 種類のポンプを少量ずつスプレーすることにより、攪拌のプロセスを行わずとも意図した比率で混合することができている。各ソレノイドの尾部には LED がついており、駆動と同時に点灯する。



図 3. 噴霧ユニット

図 4 は OPP シートの巻取装置である。シートには厚さ 0.04mm, 30m 卷 OPP ロールを使っており、使用済みのロールは使い捨てる。噴霧装置および巻取装置は、Arduino Mega 2560 と本体内の小型 Windows PC のシリアル通信 (UART 115200bps) で制御している。PC には 13.3 インチの液晶モニタ(タッチ機能なし)を HDMI 接続し、シートの下に設置している。OPP シートの横幅は手作業で切って調整しているため、右側のシートが乱れているが、液晶画面の表示部よりも外側であるため特に影響はない。

液体を吹き付けた様子は図 5 のようになっている。用いる液体はすべて無色透明であるため、映像は透過し見えているが、吹き付けられた箇所においては水滴によって解像感が下がっている。また、水滴で光が屈折し虹色を生じさせてしまっているのがわかる。とはいって、この図のように濡れている飲食物の写真であれば、違和感は少ない。



図 4. OPP シート及び巻取装置

液体噴霧混合式の味ディスプレイの試作



図 5. 液体が噴霧されたシート下に映像を映したところ
(輝度など無加工の写真)

噴霧を行うスペースには、皿を置ける余地を設けた(図 6). シートを巻き取らずに動作させることで、混合噴霧のテストや試食デモに活用している。OPP ロールが巻き取られて廃棄されとはいえ、コロナ禍において複数の人に画面を舐めてもらうのは心理抵抗が大きいが、紙皿に噴霧し配るというやり方であれば、比較的受け入れられやすいと考えている。

また、クラッカーや可食紙などを置くことで、噴霧されたものをそのまま食べてしまうこともできる。ごはんやパンに、ふりかけやピザソースなどの味を吹き付けることもできる。



図 6. (上) ステージ上に皿を置いたところ
(下) クラッカーへのトッピング

2.2 味センサとの連携

タンクに充填する味提示溶液は、以下の表 1 にある 7 種類の溶液+精製水の合計 8 種類である。現時点ではタンクが 2 つ余っているが、その拡張については展望で述べる。

表 1. 味提示溶液

提示される味	噴霧する溶液
塩味	300mM 塩化カリウム水溶液
酸味	0.3mM 酒石酸水溶液
旨味	10mM グルタミン酸ナトリウム水溶液
苦味 (医薬品)	0.1mM キニーネ水溶液
苦味 (食品)	0.01vol% イゾα酸水溶液
渋味	0.05wt% タンニン酸水溶液
甘味	スクロース水溶液
(アルコール味)	(エタノール水溶液)
(辛味)	(カブサイシンポリビレングリコール溶液)
無味	精製水

これらの液体は、味センサ TS-5000Z[4][5][6]の各センサのキャリブレーションで用いられている基本味サンプル溶液と同一である。それぞれの味覚において人の知覚閾値のおよそ 10 倍の濃度となっており、人が口にするほとんどの料理や飲み物の味は、この味の濃さを超えることがない（調味料を直接なめる場合をのぞく）。したがって、これらの溶液を最大値とした組み合わせを行えば、原理上ほとんどの飲食物の味を再現できると言うことができる。味の基本味は 5 つだが、味センサ TS-5000Z はこれに含まれない「渋味」も測定できるうえ、さらに医薬品の苦味と食品の苦味の 2 種類を測定でき、合計 7 味のデータを出力する。提案システムは、そのデータの再現を 7 種の溶質の混合で担っている。

味センサとの連携は[2][3]と同様のやり方をとっている。味センサは比較対象の液体と比べたときの相対値を出力する。提案システム用のデータを得るために、7 つの溶質を全て溶かした水溶液を比較対象として、飲食物の味を測定すればよい。各 7 味の強度として測定された値から逆算すると、それぞれ上記の基準液に換算した濃度が出せる。提案システムは、この値に基づいた噴霧混合を行って味を再現することができる。

なお、佐藤らの研究[7]では原材料をブレンドする比率で遺伝的アルゴリズムを使う仕組みを提案しているが、本システムでは、それぞれのパラメータの値を再現するための濃度のみを出している。

2.3 コンテンツおよび提示方法

提案システムにおいては、混合比を噴霧時間で制御している。もしタンク 1~10 の混合比率を 1 : 2 : 3 : ... : 9 : 10 として噴霧したいなら、10ms, 20ms,

30ms, ..., 90ms, 100ms ずつ噴霧すればよい。このように記述・制御が簡単なのと、図6の皿によって簡単に出力と味見を行えることから、料理における調味のように、味センサに頼らず人力で味を作り出すことも可能である。塩味と酸味は塩分計とpHメータの値を参考に決めつつ、あとは少しずつ噴霧量を増減させながら、再現したい飲食物と交互に味見することでそれなりに近い味にまでもっていくことができる。こうした「レシピ」まで取り込むことができると、かなり多くの飲食物の再現データがそろうことになる。

提案システムでは、中にUSBマイクも取り付けることで、これらの味覚データを、スマートスピーカーのように音声入力で呼び出せるようにした。ソフトウェアの実装はProcessingで行っており、音声認識にはWindowsのSystem.Speech機能を用いた。登録されている飲食物の名前のみの文法ファイルを用意することで、誤認識を大幅に減らすことができる。味わいたい飲食物の名前を声に出すと、その飲食物に対応するフォルダ内の動画ファイルが再生され、同一名で保存されたテキストファイル内のデータ(10個の値、ミリ秒)に基づいて噴霧が行われる。飲食物データの追加や修正が容易であるうえ、dropbox等のサービスでフォルダを同期すればデータの共同編集もできる。

さらに、「濃いめで」「薄めで」といった修飾語によって全体の味の濃さを指示したり、「酸味を弱めで」といった指示によって味イコライザ[2][3]として機能させたりできるようにした。



図 7. 外装をつけた状態

さらに、外装(図7)を施し、多くのアプリケーションを試作している。味見できるメニュー、ソムリエ訓練アプリ(ワインの識別)、調理過程を途中で味見できる料理アプリ、味覚クイズ(お菓子の味がどちらかを当てる)、味覚相性診断(どちらの味が好みかをきくことによってその類似度を判定)、味覚ゲーム

(同じ味のアイテムをキャッチする)、味覚スロットマシン(そろった絵柄の味が楽しめる)、味覚ガチャ(図8. 各溶液の出力量を乱数で混合)、味覚ピアノ

(図9. 鍵盤で味を混ぜるもの)、食べているものが市販品の場合の味覚ビデオチャットなどである。



図 8. 味覚ガチャ

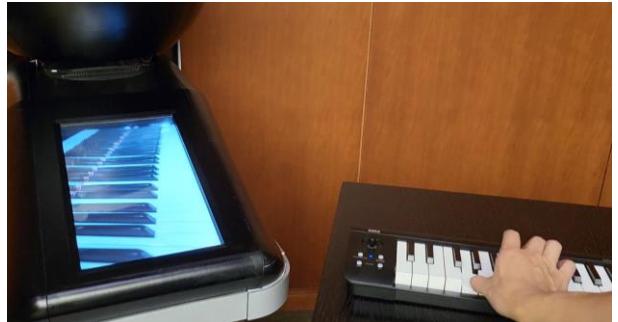


図 9. 味覚ピアノ

3 議論

本稿での提案システムは、Norimaki Synthesizer[1]を味センサと連動させたシステム[2][3](以下、イオン泳動式味ディスプレイ)と比較して、多くの違いがあり長所も多い。以下、それぞれの観点において議論を行う。

3.1 表現できる味の強度範囲

まず、表現できる味の強度範囲が異なる。イオン泳動式味ディスプレイは、ゲル内に溶けている電解質を電気で動かすことによって味の濃さを表現する。電気をかけていないときにはそのゲルの味の濃さを感じ、電気をかけるほど無味に近づいていく。しかし、ゲルの味を濃くしていくと、どれほど電流をかけてもある程度はイオンが舌に触れてしまうため、無味にもっていくことができず、味が残って感じられる。このため、イオン泳動式味ディスプレイは、ゲルの味の濃さに制限がある。実際[1][2]では塩味のゲルは1%のNaClを用いており、一般的な飲食物の塩味を再現するには物足りない塩味であるが、かといって、上記制約のためこれ以上濃くするのが難しいという現状がある。

これに対し、本稿で提案する噴霧混合式の場合、噴霧しなければ「無味」が表現でき、一方で濃い味についてもほとんど上限がない。表1に記した濃度

液体噴霧混合式の味ディスプレイの試作

がとりあえずの上限にはなるが、これも暫定で決めた値に過ぎず、もっと表現域を上げたいのならタンク内の溶液を濃くするだけでよい。飽和水溶液であっても問題ない。(ただし、結晶等ができないよう、使用後の洗浄は不可欠となる)。

また、噴霧混合式は原理上、味覚電気刺激が一切ないので、その違和感がない。電気味覚研究の分野ではノイズ除去やフェードインなどで刺激を軽減する研究がなされている[10]が、それでも原理上刺激を生じさせない本手法のほうが有利であろう。

3.2 表現できる味の種類の範囲

イオン泳動式では基本五味にとどまっていた味を、噴霧混合式では、渋味を含む7味に拡張することに成功している。イオン泳動式では「電解質を用いなくてはならない」という制約があり、たとえば甘味を表現するにはグリシンを用いるなどの工夫が必要であった。しかし噴霧混合式ではこうしたことが不要で、たとえば甘味の表現ではシンプルにスクロースを用いることができる。グリシンとスクロースの風味の違いを考えても、実際に飲食物によく使われるスクロースが有利であるし、甘味の強度の点でも格段の差がある。

3.3 食体験としての違い

イオン泳動式では、ゲルを舌にあてたときの接触面でしか味の制御ができなかったが、噴霧混合式では、実際に舐めたあと、「飲み込むところまで」その味がする。イオン泳動式でも首の近辺に電極を貼ることで嚥下時の味を多少制御できるが、それでも全ての味が混ざった状態であり、個別に制御することはできない。提示された味がそのまま飲み込めるのは、味覚の体験を超え、飲食体験に近づいている。

3.4 味再現の正確性

味の再現がより正確と考えられることも噴霧混合式の強みである。イオン泳動式では、人間がどのような濃さに感じるかという評価実験のデータを用いて味をそろえていた。しかもそれには個人差があり、誰がそのデバイスを用いるかによってパラメータを変更しないといけなかった。イオン泳動説を実証した青山らの論文[11]でも、データを正規化することによって線形回帰できているが、逆に言うと個人ごとの事前測定・事前設定が重要であることを示している。塩味のリアルタイムな伝送を試みた小林らの論文[12]には被験者ごとの事前キャリブレーションのデータが掲載されているが、いかに個人ごとに傾きや切片が異なることが見てとれる。

一方で、噴霧混合式の場合はそうした問題は起こらない。加えて、強力なメリットは「混合された液体の味を、再び味センサで測定できる」ことにある。

味を再現したい飲食物と、それを再現すべく噴霧混合された液体とともに味センサに入れ、その2つが同じ味になっているかどうかを客観的な数値として比較できる。これにより、さらなる再現精度の向上を実現する計算方法を考えることもできる。なお、味センサは脂質膜電極の電位差を測定するため、このようなことはイオン泳動式では行えない。

3.5 味の時間的変化

イオン泳動式と比べて、噴霧混合式のほうが劣る点もある。まず、噴霧混合式は、イオン泳動式と異なり、即時的な味変化ができない。噴霧混合式の場合、噴霧混合およびシートの巻き取りを含めると1秒程度の遅延が避けられない。また、「味のクロスフェード」など、動的な味変化を噴霧混合で行うことがやや難しい。将来的には、制御精度を高め、きれいな「味のグラデーション」を実現することで克服できる可能性はあるものの、この点においてはイオン泳動式のほうが優れていると考えられる。

3.6 味エフェクタ・味イコライザへの対応

イオン泳動式と噴霧混合式が同等な点もある。基本五味、あるいは7味のバランスを変える味エフェクタや味イコライザ[2][3]はどちらの方式でも機能する。個人ごとの嗜好に合わせた強弱調整を行うことも、方式によらず実現可能である。

3.7 画質への影響

画質に対する影響は、両方式ともにあると考えられる。噴霧混合式は、映像が曇り、水滴がついて解像感が下がる。イオン泳動式でもゲルを画面上に載せる必要があり、これによって映像が曇ったり解像感が下がったりする。両者とも今後改良が必要と考えられる。

4 展望

現在の提案システムは、一ヵ所に噴霧混合を行っているが、今後、インクジェットプリンタのような仕組みでヘッドを左右に動かせるようにしたい。そうすれば、映像の中の2次元的な配置や分布に基づく味提示が行えるようになる。

さて、表1に括弧付きで示しているように、噴霧混合式の味ディスプレイには、さらに、味センサでも測定できないアルコール味や辛味を加えることができる余地が備わっている。エタノール濃度計[8]を購入・使用したところ、精度の高い測定ができたので、同じ度数になるようにエタノールを噴霧混合することは可能である。たとえば、カクテルの味の再現や、ワインの味比べといったことができるようになるはずである。

また、辛味は学術的分類としては痛覚ということ

になっているが、食体験の再現という意味では味覚のうちに加えたいと考えている。実際、食べ物の辛味を測定できる Scoville Meter[9]が市販されているので、その値に応じて、カプサイシンを噴霧混合するように改良したいと考えている。エスニックな料理などの味の再現には不可欠だと考えているので、狭義の味覚の定義にこだわらずに改良していきたい。

このほかにも、脂肪味やコク味といった概念があり、これも狭義の味覚に入ってはいないが食体験としては重要であると考えている。粘性の低い油などを利用することを検討していきたい。

そして、もはやこれを行う場合には 10 タンクでは足りないが、嗅覚にうつたえる香料(フレーバー)を加えることも検討したい。実際、嗅覚の食体験に大きく影響を及ぼすと考えている。

この展望の章で記述されていることはどれも、電気味覚技術では非常に実現が難しい事柄であったが、噴霧混合であればそれほど難しいことではない。味ディスプレイの可能性を開拓するためにも、ぜひ挑戦したいと考えている。

謝辞

本研究は、総務省戦略的情報通信研究開発推進事業(SCOPE)の独創的な人向け特別枠「異能 vation」プログラムの支援を受けた。

参考文献

- [1] Homei Miyashita. Taste Display that Reproduces Tastes Measured by a Taste Sensor. Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology. Association for Computing Machinery, New York, NY, USA, 1085–1093, 2020.
DOI: <https://doi.org/10.1145/3379337.3415852>
- [2] Homei Miyashita. Taste Display that Reproduces Tastes Measured by a Taste Sensor. Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology. Association for Computing Machinery, New York, NY, USA, 1085–1093, 2020.
DOI: <https://doi.org/10.1145/3379337.3415852>
- [3] 宮下芳明. 画面に映っている食品の味を再現して味わえる味ディスプレイの開発, 第 28 回インテ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ピング (WISS2020) 論文集, 2020.
- [4] Yusuke Tahara and Kiyoshi Toko. Electronic tongues--A review, IEEE Sensors Journal, 13(8), 3001—3011, 2013.
DOI: <https://doi.org/10.1109/JSEN.2013.2263125>
- [5] Kiyoshi Toko. Taste sensor with global selectivity, Materials Science and Engineering: C, 4(2), 69-82, 1996.
DOI: [https://doi.org/10.1016/0928-4931\(96\)00134-8](https://doi.org/10.1016/0928-4931(96)00134-8)
- [6] Kiyoshi Toko. Taste sensor, sensors and actuators B: Chemical, 64(1–3), 205—215, 2000.
DOI: [https://doi.org/10.1016/S0925-4005\(99\)00508-0](https://doi.org/10.1016/S0925-4005(99)00508-0)
- [7] Masako Satoh, Hideki Satoh, Hidekazu Ikezaki, Optimization of Food Ingredients and their Blend Ratios Based on Taste Sensor Output, IEEJ Transactions on Sensors and Micromachines, 2015, Volume 135, Issue 2, Pages 57-64, Online ISSN 1347-5525, Print ISSN 1341-8939, 2015.
DOI: <https://doi.org/10.1541/ieejsmas.135.57>
- [8] ATAGO Digital Hand-held “PEN” Ethyl alcohol Refractometer PEN-Ethanol V
<https://prolab.id/product/atago-digital-hand-held-pen-ethyl-alcohol-refractometer-pen-ethanol-v/>
- [9] FoodSense, Hotness of Chillies (ChilliPot – Scoville Meter)
<https://www.tms-lab.com/product/foodsense-hotness-of-chillies-chillipot-scoville-meter/>
- [10] 鍛治慶亘, 宮下芳明. TasteSynth : 電気味覚のための刺激波形デザインシステム, エンタテインメントコンピューティングシンポジウム論文集, Vol.2021, pp.266 - 275, 2021.
- [11] Kazuma Aoyama, Kenta Sakurai, Satoru Sakurai, Makoto Mizukami, Taro Maeda, Hideyuki Ando. Galvanic Tongue Stimulation Inhibits Five Basic Tastes Induced by Aqueous Electrolyte Solution. Frontiers in Psychology. 8:2112, 2017.
- [12] 小林未侑, 宮下芳明. TeleSalty:リアルタイムで塩味を伝える通信システム, エンタテインメントコンピューティングシンポジウム 2021 論文集, Vol.2021, pp.276-280, 2021.
- [13] Maynes-Aminzade, D. Edible Bits: Seamless Interfaces between People, Data and Food, in the Ext. Abstracts CHI '05, 2005.
- [14] OPEN MEALS, SUSHI TELEPORTATION. SXSW2018, 2018.
- [15] Homei Miyashita. TTTV (Taste the TV): Taste Presentation Display for “Licking the Screen” using a Rolling Transparent Sheet and a Mixture of Liquid Sprays. In the Adjunct Publication of the 34th Annual ACM Symposium on User Interface Software and Technology (UIST '21 Adjunct), Virtual Event, USA, 2021.
DOI: <https://doi.org/10.1145/3474349.3480223>

未来ビジョン

電気味覚研究には 2 つのビジョンがある。ひとつは、食事の味を濃く感じさせることで、美味しさと健康を両立するというビジョン。もうひとつは、味覚の記録再生を行う味メディアとしてのビジョンである。

筆者も 10 年にわたり、電気味覚研究の意義や重要性を主張するためにこれらのビジョンを強調してきた。そうした中で最近気づいたことは、これらのビジョンは多くの共感を集めることほどに正しく、それぞれのビジョンが実現するならむしろ手段は電気味覚でなくてもかまわないのではないか、ということである。

テレビにはブラウン管や液晶など様々な方式がある。プリンタにもインクジェットやレーザーなど様々な方式がある。このように考

えたとき、味覚メディアというビジョンを実現する「電気味覚以外の方式」はないのだろうかと考えた。

「液体を透明なシートに噴霧し、ベルトコンベアながら液晶画面の上にスクロールさせる」という提案は、極めて原始的な試みのように思える。しかしながら、本稿でイオン泳動式の味ディスプレイと比較した議論を行っていくと優れた点も多く、その対比のおかげで、イオン泳動式の味ディスプレイに何が足りないのかをより浮き彫りにできたと考えている。

味ディスプレイに求められる重要な機能はは結局何なのか？イオン泳動式でもなく液体噴霧でもない他の方式はないだろうか？このような議論を、本ワークショップで行いたいと考えている。

UPLIGHT: 全周囲ディスプレイと携帯型ゲームデバイスの融合

中村 俊勝 * 加藤 達也 † 伊藤 思音 ‡ 佐藤 俊樹 *

概要. 我々は、立方体や球などの高さと側面を持つ構造物の隠れた部分に着目し、この隠れを身体を動かして能動的に見に行く行為がエンタテインメント性を持っていると仮説を立てた。本研究では、このエンタテインメント性を、携帯ゲーム機のプレイスタイルに落とし込むことで生まれた新たな全周囲ディスプレイデバイス「UPLIGHT」を提案する。本研究では内蔵する小型短焦点プロジェクタと超広角魚眼レンズを用いた全周囲投影技術に加え、自由に交換可能な様々な形状の全周囲ディスプレイで、全周囲連続的な映像表現のゲームコンテンツを実現した。更に全周囲ディスプレイの左右下部にコントローラを取り付けることにより、伝統的な携帯型ゲーム機のプレイスタイルを取り入れた。本論文では、開発したプロトタイプ及び操作可能なゲームコンテンツ、展示活動中に得た知見を基に行った考察について述べる。

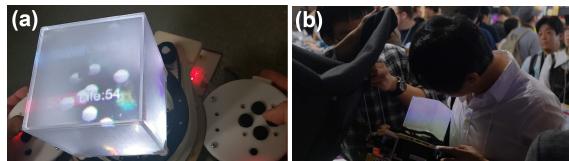


図 1. プロトタイプの外観とユーザによる試遊の様子
(a)UPLIGHT, (b)UPLIGHT を遊ぶ様子

1 はじめに

全周囲ディスプレイは、立方体や球体などの立体物の表面全周囲に映像を表示するデバイスである。全周囲ディスプレイには、壁や天井をディスプレイとし、映像を内側から見るもの[4, 1]と、直径6mの大きな球体[7]や、手のひらに収まる小さな筒をディスプレイとし、映像を外側から見るもの[5]がある。全周囲ディスプレイには、一般的な平面形状のディスプレイ内やHMD等を通して見た3Dグラフィックスや入力デバイスを介した間接的な対話手法とは異なり、映像を伴った立体的な形状を直接目で見て、直接手で触れることができるという高い現実感のある対話が可能な利点がある。

このような全周囲ディスプレイには、立体的な形状の全周囲に映像を表示させるがゆえに、ユーザの頭部位置に応じた「死角」が存在する。例えば、ユーザがある角度からディスプレイ手前側の側面に表示された映像を見ているとき、このユーザは反対側の側面に表示された映像を同時に見ることはできない。このようなディスプレイの死角を全周囲ディスプレイの制限とみなし、ディスプレイの一部に死角を補う映像を映したり、タッチ操作等で映像をスクロールさせることでこれを回避する手法[6]が研究されて

きた。また、この死角を複数人同時使用時の各ユーザーのプライベート領域として積極的に活用する研究[2]もある。しかし、従来の研究ではこの死角によって生じた見えない領域を、ユーザに身体動作を用いて能動的に見に行かせることに着目し、「見えない部分を見に行く」動作のエンタテインメント性について深く議論した研究は少ないといえる。

そこで本研究では、これまで全周囲ディスプレイの死角をユーザに積極的に「見に行かせる」ことで、ゲーム体験のエンタテインメント性を高める試みに着目する。これまでにも、大形の全周囲ディスプレイを複数人で囲んで用いるシステムの提案[12]を行ってきたが、本研究では新たにシングルユーザに着目し、両手で保持した全周囲ディスプレイを両手・上半身の動作を用いて様々な角度から閲覧する手法を提案する。さらに、これを誰もが親しんできたシングルユーザ向けの小型ゲームデバイスの操作メタファと融合させる試みも行い、新しく提案する全周囲ディスプレイ搭載のゲームデバイス「UPLIGHT」(図1)のインターフェース手法について述べる。

2 目的

本研究の目的は、両手で持つことが可能な小型の全周囲ディスプレイを用いて、全周囲ディスプレイの隠れた部分をユーザが自発的に見にいくインタラクションが持つエンタテインメント性を活かし、人が楽しむことができる新たな小型全周囲ディスプレイデバイスを開発することである。これを達成するうえで、本論文では、デバイスに適した形状や入力手法、必要な機能などの提案デバイスの設計についての考え方と、提案デバイスによってユーザが体験するであろう楽しさについての考え方、そして、ディスプレイ部の隠れへのユーザの興味を高めるために必要な機能についての考え方を以下に述べる。

Copyright is held by the author(s).

* 北陸先端科学技術大学院大学

† 電気通信大学

‡ Delft University of Technology



図 2. 上下が定まっている立体構造物の隠れの閲覧方法
(a) ひっくり返し, (b) 覗きこみ, (c) 手回し

2.1 閲覧方法の分析と UPLIGHT の要件

まず、我々は、両手で持てるサイズの人形や、側面に柄の描かれたコップ等の、ユーザにとって上下が定まっている立体構造物を両手で把持し、閲覧するスタイルに着目した。これは、机に置くときや、両手で把持するときの姿勢が、置きやすさや持ちやすさから、ほぼ一姿勢に定まる立体構造物を両手で把持し、様々な方向から眺める閲覧スタイルだといえる。このような構造物の隠れた部分を閲覧するときに、ユーザは主に以下の3つの見にいき方をする。

1つ目は、図2(a)のように、両手の手首をひねって、対象物の裏側を見るために、ひっくり返して見る方法である。この見にいき方は、両手の手首をひねるだけで行うことができ、極短時間で対象物の裏側の隠れを、ユーザの真正面に持ってきて見ることができるという特徴をもっている。また、この方法は、主に、ユーザが座って対象物を目線と同じ高さの水平な位置に保持していたり、対象物の隠れた部分の情報の反転や、対象物自体の姿勢の変更を気にする必要がなかったり、対象物の裏側の情報をすぐに観察したかったり、対象物自体がユーザにとって軽かったりするときに行われやすいと考えられる。

2つ目は、図2(b)のように、対象物のユーザが見たい部分の隠れが見えるように、ユーザ自身の上半身と対象物をそれぞれ逆向きに傾けて覗き見る方法である。この見にいき方には、ユーザ自身と対象物それぞれの姿勢の傾け方を変えることで、対象物の全周囲側面を連続的に見ることができるという特徴がある。また、この方法は、主に、ユーザが立って対象物を見下ろす位置で保持していたり、対象物の隠れの情報の反転を気にする必要がなかったり、対象物自体の姿勢を出来るだけ維持したかったり、対象物の隠れの情報を即座に、かつ、連続的に観察したかったりするときに行われやすいと考えられる。

3つ目は、図2(c)のように、利き手で対象物の下部を持って、利き手と逆の手で対象物の頭部を掴み回転させて、対象物の裏側の上下を反転させずにユーザの真正面に持ってきて見る方法である。この見にいき方には、対象物の隠れの裏側の情報を上下をひっくり返すことなく、じっくりと真正面から観察することができるという特徴や、対象物の側面全周囲を連続的に真正面から眺めることができるという特徴がある。また、この方法は、主に、対象物の隠れた部分の情報が上下が反転した状態では理解す

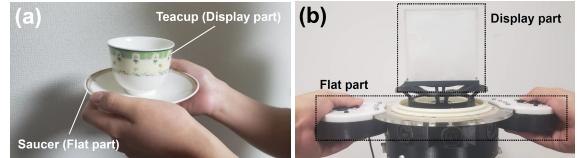


図 3. 全周囲ディスプレイに上下を作り出すための着目
(a) ティーカップ, (b) 把持可能な平面部の付加

ることや捉えることが難しいものであったり、隠れを観察する時間が十分に存在したり、隠れにユーザが興味を持っていたりするときに行われやすいと考えられる。

これらの閲覧方法は、身体的な動きを多く含み、能動的な情報探索行為の能動性をより高めることができると考えた。そして、高められた能動性は、能動的な情報探索行為の結果によってユーザにもたらされる驚きや興奮、発見といった感情の起伏をより大きなものに変えることができるとも考えている。そのため、提案デバイスに適した形状は、以上の閲覧方法を行いやすく、加えて、自然とユーザがこれらのインタラクションを行うような物である必要があると我々は考えた。以上の条件を満たすためには、主に、以下の要件を考慮した全周囲ディスプレイデバイスを設計する必要があると考える。

- 要件 1. デバイス全体の上下が定まっていて、その上下がユーザにとって認知しやすいこと。
- 要件 2. デバイスを両手で把持しやすくて、両手で持ちながらでも手首がひねりやすいこと。
- 要件 3. 全周囲ディスプレイは、どの角度から眺めても隠れが存在し続ける形状であること。
- 要件 4. デバイスを把持しながら入力を行えること。
- 要件 5. ディスプレイを見ながら入力を行えること。
- 要件 6. ディスプレイ部の底面以外で、全周囲連続的な映像が表示できること。
- 要件 7. ディスプレイ部が容易に片手で回せること。

2.2 UPLIGHT に適した形状

まず、我々は、全周囲ディスプレイに上下を作りだす必要がある。そこで、我々は、図3(a)のティーカップとソーサーのような、立体構造物と平面の関係に着目した。すなわち、図3(b)のように、全周囲ディスプレイの底面に横長の平面を付加することで、全周囲ディスプレイ部が上部、平面部が下部とユーザに認識されると考えたのである。これによって、要件1を満たすことができると考えた。さらに、その平面部は、両手で把持することができる十分な大きさとすることで、要件2を満たすことができると考えた。

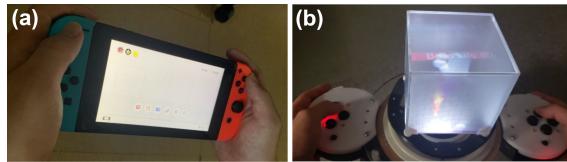


図 4. 携帯ゲーム機の伝統的なプレイスタイルの継承
(a)Nintendo Switch, (b)UPLIGHT

次に、全周囲ディスプレイ部の形状について考える。詳しくは考察の章にて述べるが、前提として、全周囲ディスプレイには、ユーザがどんな角度から見ても、隠れが存在し続ける必要がある。そのため、全周囲ディスプレイ部は、立方体や、球、三角柱などの柱体といった、どの角度からもその全周囲を一覧することができない形状である必要がある。また、全周囲ディスプレイ部のどの部分を底面として、両手で把持可能な平面部と接合するのかを考える必要がある。立方体や球といった形状であれば、どの部分を底部としようが問題ないが、三角柱といった形状であれば、側面を底部とすると、頭上から一覧できてしまう。そのため、柱体形状の場合は基本的には底面をそのまま底部とする必要がある。以上のように、ディスプレイ形状を決定することで、要件 3 を満たすことができると思った。

2.3 UPLIGHT に適した入力手法

まず、我々は、提案デバイスの平面部の左右を両手で把持し、全周囲ディスプレイ部を目前に保持して閲覧するというスタイルが携帯ゲーム機のプレイスタイルに類似していることに注目した。小型のゲーム機本体を両手で把持し、目の前で保持しながら、右手・左手用に分かれて配置されたボタンを両手で操作して遊ぶのが携帯ゲーム機のプレイスタイルである。この携帯ゲーム機のプレイスタイルは、1980 年前後に発売されたエポック社のゲームポケコンや、任天堂のゲーム&ウォッチ、ひいては 2017 年に発売された任天堂の Switch(図 4(a)) に至るまでほぼ一貫している。スマートフォンやタブレットが急速に普及しタッチスクリーンによる入力が浸透した今日においても、このスタイルの携帯型ゲーム機はゲームファンの間では根強い人気があり、もはや確立されたプレイスタイルであると言っても過言ではない。

そこで、本研究では、提案デバイスの入力手法として、図 4(b) のように、携帯ゲーム機と同様の左右に分かれた 1 組のボタンコントローラを取り入れることにした。このような伝統的な携帯ゲーム機の操作手法を取り入れることで、提案デバイスが携帯ゲーム機の延長線上にあるものと認識され、特別な操作説明を必要とせず、誰でも使用可能なデバイスとなることを期待している。

また、以上のような入力手法を採用することで、要件 4 と要件 5 を満たすことができると考えられる。

2.4 UPLIGHT に必要な機能

はじめに、ディスプレイ部の全周囲連続的な映像表現を実現する手法について考える。これには様々な実現方法が存在するが、本提案デバイスに適した手法を考える必要がある。まず、OLED を複数枚組み合わせた手法 [8] では、OLED 間に映像の断裂が起きてしまうため要件 6 を満たすことができない上に、様々なディスプレイ形状を実現するためのハンドルが高いという問題がある。また、深度カメラと外部プロジェクタを用いた手法 [11] では、ユーザが両手でデバイスを手持した状態で扱うため、上半身や両腕に投影を遮られ、ユーザの真正面の部分に投影が届かないことが起こる可能性がある上に、その解決にはプロジェクタを数多く必要とするという問題がある。加えて、赤外線カメラと小型プロジェクタをヘルメットと一体化させたデバイスを頭部にかぶる手法 [3] では、ディスプレイ部のユーザが見ている部分にしか投影ができない上に、ディスプレイ部を掴む際に手に投影が重なってしまうという問題があり、要件 7 を満たすことができない。更に、HMD を用いた手法にも言えるが、頭に装置を付けるため頭部が重くなってしまい、ディスプレイ部の全周囲を閲覧する際に身体を動かしにくくなってしまうという問題もある。そこで本研究では、プロジェクタをデバイスに内蔵し、ディスプレイ部の下部からのリアプロジェクションを行った。そのため、プロジェクタは小型で短焦点なものを採用し、ディスプレイ下部に配置した魚眼レンズによって投影範囲を広げた全周囲投影ユニットを用いた。これにより、全周囲連続的な映像表現が常に可能で要件 6 を満たせると考えられ、かつ、様々なディスプレイ形状を実現しやすいという利点や、投影対象であるディスプレイ部とプロジェクタを一体化していることで、ユーザが激しく動いても投影がぶれないという長所もある。

次に、ディスプレイ部の水平回転機構を実現する手法について考える。まず、前述のように、ディスプレイ部はリアプロジェクションによって実現するため、投影する映像の角度と実際のディスプレイ部の回転角を正確、かつ、リアルタイムに同期させる必要がある。そこで、詳しくは実装の章で述べるが、タイミングパリとタイミングベルト、ロータリエンコーダを用いた機構で、これを実現する。これにより、要件 7 を満たすことができると考えられる。

2.5 UPLIGHT がもたらす楽しさ

まず、楽しさには多くの種類があるが、我々が提案する手法によってもたらされる楽しさは、能動的な情報の探索行為による驚きや発見、興奮といった感情の起伏によって得られる楽しさが主なものであると考えている。また、その発見によって、ユーザ自身の知的好奇心が満たされるという楽しさもある。加えて、情報探索行為を身体を動かして行うために

発生する、身体を動かすことそれ自体による楽しさも得られると考えている。さらに、身体を動かしながら能動的に情報探索行為を行う場合、その能動性は身体を大きく動かすほど強くなると考えている。そして、能動性が高まった分だけ、能動的な情報探索行為による感情の起伏は大きくなり、その結果として、ユーザはより強く楽しさを感じられたり、より大きくなつた知的好奇心を満たせたりするのではないかと考えている。

2.6 隠れへ視線を誘導するための工夫や機能

ここでは、全周囲ディスプレイの見えない部分に対するユーザの好奇心をくすぐり、ユーザの能動性をより高めるために考案した三つの工夫や機能について述べる。

はじめに、ディスプレイ形状の工夫について述べる。まず、前述したように、本研究において、全周囲ディスプレイの形状には隠れが必要である。よって、制限はあれど、球や立方体や柱体などの様々な形状が採用可能である。しかし、よりユーザの視点を隠れへ誘導するには、曲面よりも、平面で構成された形状の方が適しているのではないかと考えている。実際に、平面で構成されたディスプレイ部の隠れを見にいった際に、一定角度を越えるとそれまで見えていなかった映像が一気に見えるようになる。そのため、一度に発見や驚きがより多く起こるようになり、感情の起伏が大きくなる。こういった理由から、立方体や柱体といった形状の方がより隠れへユーザの視点を誘導できるのではないかと考えている。

次に、アプリケーション側の制御によるディスプレイ部の自動回転を用いた視覚的な誘導機能について述べる。まず、詳しくは考察の章にて述べるが、ディスプレイ部の自動回転がユーザによって制御できる状態は本研究においては望ましくない。しかし、アプリケーション側の制御で自動回転機能を用いることで、ユーザが見たい情報がユーザの視界から隠れるように、ディスプレイ部自体が自動回転するといった要素を実装でき、ユーザの隠れへの好奇心を高める要素になるとを考えている。

最後に、異なる位置に設置した複数の音源を用いた聴覚的な誘導機能について述べる。これは全周囲ディスプレイ部下部に円周上に等間隔で外側に向けてスピーカを配置し、キャラクタがアクションを行ったディスプレイ上の位置に対応したスピーカから、アクションに応じた効果音を発生させるという手法である。これを活用することで、全周囲ディスプレイの隠れから聞こえてくる音にユーザが誘導されるといった効果が期待できる。

以上のようなコンセプトのもと、本研究では全周囲ディスプレイデバイス「UPLIGHT」の設計と開発を行った。本論文では以下に、本研究の立ち位置を示した後、実際に開発したプロトタイプの実装方

法や、有効だと考えられるアプリケーションの構成要素と、実際に開発したアプリケーションの詳細、展示活動などを経て得た知見と考察について述べる。

3 関連研究

これまでにも、全周囲ディスプレイにユーザから見て隠れた部分が存在するという特徴を活かした、エンタテインメントへの応用がなされてきた。まず、Julie は、球状全周囲ディスプレイの側面のプライベート性と上部のパブリック性を活用したマルチユーザによる対戦ゲームと協力ゲームを提案している[10]。そして、Ryouta は、円筒形全周囲ディスプレイの自分の見えていない部分は対戦相手に見えているという特徴を活かした、ワイヤレスコントローラを用いた対戦ゲームを提案している[9]。しかし、これらの全周囲ディスプレイの隠れの応用はどれもマルチユーザへのものであり、シングルユーザのインタラクションに着目したものはなかった。

また、平面ディスプレイ上で遊べるPolytron Corp. のFezは視点の横回転機能を持ち、これまでの視点だと見えていなかった情報が、視点の横回転によって見えるようになり、驚きや発見を生み出すゲームデザインが為されている。しかし、情報探索行為である重要な視点の横回転機能が指先を動かすことでの操作可能なボタンなどによって行われている。我々は、この情報探索行為を身体を動かしていればそれ自体がより楽しくなり、更に、情報探索行為によって発見や驚きがあったとき、能動的に身体を動かしているからこそ、より強い楽しさを生むことができると考えている。

4 実装

4.1 システム概要

図5にプロトタイプのハードウェア構成を示す。提案デバイスは小型プロジェクタ (Smart Beam Laser) と広角レンズから成る全周囲投影ユニットと、その上部に配置された10cm四方の取り替え可能な全周囲ディスプレイ、ステッピングモータとロータリエンコーダ (PKP214U06A-R2EL-L)、タイミングベルトおよびブリーリから成るディスプレイ回転機構、ボタンコントローラ、回転制御用のマイクロコントローラ、ディスプレイの下部に円周上に配置されたスピーカ、および映像投影用の計算機から成る。

4.2 広角レンズとプロジェクタを使った全周囲投影

広角レンズとプロジェクタを使った全周囲投影システムの概略を図5に示す。小型プロジェクタから照射された映像は直上の広角レンズを通り、全周囲ディスプレイの内部からディスプレイ上面および4つの側面にリアプロジェクションされる。

全周囲ディスプレイに沿った映像を投影するため

UPLIGHT: 全周囲ディスプレイと携帯型ゲームデバイスの融合

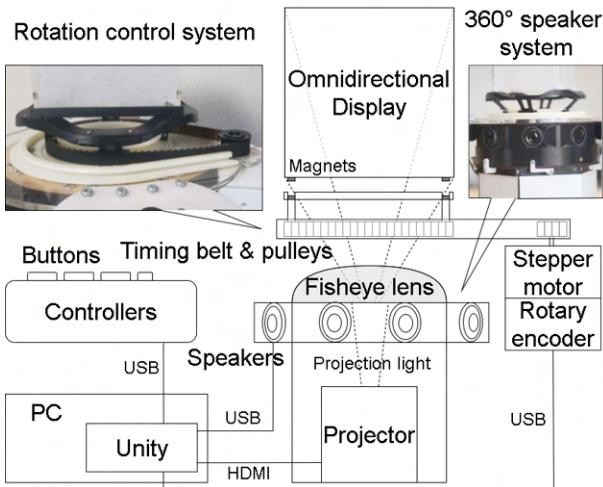


図 5. システム構成図

に、まずプロジェクトキャリブレーションを行いプロジェクトの内部・外部パラメータ、およびレンズの歪みパラメータを求めている。これらをUnity内のカメラに適用することで、3DCG空間上に再現したディスプレイと同形状のメッシュモデルに適用されたゲームのシーンを投影したテクスチャがそのまま全周囲ディスプレイに投影されるようになる。

全周囲ディスプレイは、今回の実装では立方体形(図6(a))に加え、直径10cmのくびれのある球体形状(図6(b))も用意した。立方体ディスプレイについては板厚2mmのアクリルキューブの内側からトレーシングペーパーを張り付けることで制作した。また球体ディスプレイについては直径10cmのLED電球のシェード部分を用いて制作した。

4.3 全周囲ディスプレイ回転システムの設計

全周囲ディスプレイ回転システムを図5(左上)に示す。ディスプレイは輪状のタイミングプーリに装着されステッピングモーターとタイミングベルトで接続されている。そのため、ディスプレイは手で掴んで直接回転させることができながら、必要に応じてマイクロコントローラを介してUnityから自動回転させることも可能である。ディスプレイの回転角度は、ロータリエンコーダでリアルタイムに読み取り、マイクロコントローラから計算機に送信される。Unity上でこれを受け取り、メッシュモデルに同じ回転を適用することで、回転に追従した映像を投影することができる。

4.4 360度スピーカシステムの設計

360度スピーカシステムを図5(右上)に示す。音声はプロジェクタに内蔵されたディスプレイ下部のスピーカに加え、ディスプレイの周囲に円周上に配置した8個のスピーカから出力する。音声の再生角度は、回転していない状態のディスプレイにおける

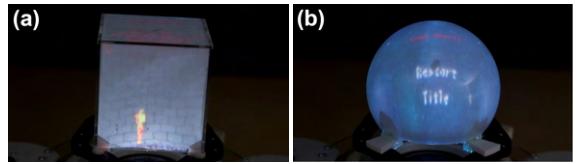


図 6. (a) 立方体形ディスプレイ、(b) 球形ディスプレイ

角度をUnity上で指定し、音声の再生は、指定された角度に最も近い角度にあるスピーカを使用する。また、ディスプレイの回転角度を取得し、再生角度を補正することで、ディスプレイの回転に追従した音声の出力が可能である。

5 アプリケーション

5.1 本提案を達成するうえで必要な要素について

本インタラクションを達成するうえで、ゲームアプリには以下の要素が必要だと考えられる。まず、ディスプレイ関連では、全周囲に映るゲーム映像の連続的な繋がり、ユーザから隠れている部分も含めた全周囲表面上の形状に沿ったキャラクターの移動といった要素が考えられる。そして、デザイン関連では、立体形状の隠れをユーザが見にいったときに驚きや発見が生まれるようなゲームデザイン、様々な面をユーザに沢山見たいと思わせるようなゲームデザイン、常にユーザに確認してもらうことが必要な情報の天面への配置といったUIデザイン、隠れがあることによるゲームの高難易度化の予測と適切なレベルデザインといった要素が考えられる。

以上の要素を複数持ったゲームアプリ例として、UPLIGHT用のアクションゲームとブロック崩しゲームを制作したため、その内容について述べる。

5.2 サイドビューアクション for UPLIGHT

これは、一般的なサイドビューアクションゲームを元に開発したものである(図7(a))。これは、側面で360度繋がったステージ上を、移動やジャンプ、弾の発射などが行えるキャラクタを操作しながら、隠れた部分に潜む敵や障害物を倒したり、かわしたりして、次のステージに続くはしごを目指し、奥に進んでいくゲームである。操作キャラクタは、ユーザから隠れた部分を含めてディスプレイの側面全周囲を移動でき、隠れた部分に移動した操作キャラクタを姿勢を変えて見に行き、そこに敵やジャンプで越えないといけない壁等の発見を用意することで、ユーザに発見や興奮、驚きを与えることができると考えている。更に、キャラクタは全周囲を短時間で移動するため、上記の体験を短時間で繰り返し体験できる良さがある。そして、本ゲームでは、天面も利用している。まず、ユーザのヒットポイント等の基本的なパラメータや、メニューなどのUIを表示している。また、敵がディスプレイの隠れた部分か

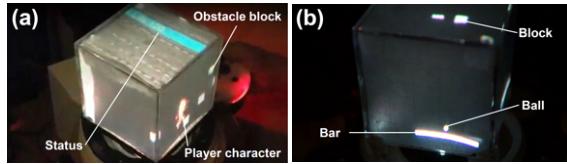


図 7. (a) アクションゲーム, (b) ブロック崩しゲーム

ら弾を発射したときに、側面だけではなく、天面を通過して操作キャラクタに降りかかるなどといったギミック上の利用も行っている。加えて、ユーザが操作するキャラクタも、向こう側に存在する敵を天面越しに狙うことができる。

5.3 ブロック崩し for UPLIGHT

これは、一般的なブロック崩しゲームを元に開発したものである(図7(b)). これは、全周囲ディスプレイに沿って飛ぶボールを、ディスプレイ底部に達しないように、全周囲側面を移動するバーを操作して跳ね返し、天面部に配置されたブロックをボールを当てて全て破壊することを目指すゲームである。ユーザはディスプレイ形状に沿って飛ぶ特殊な弾道のボールを用いたブロック崩しゲームが楽しめる。また、ゲーム中のボールは立体形状を変えるごとに異なる弾道となり、楽しみ方や体験を変えることができる。更に見えないところに飛んでいったボールの軌道を完全に予測することは難しく、ユーザは見えない部分を身体を動かして見にいき、意外な位置や軌道を描くボールに驚きや発見を見出し、全周囲ディスプレイの隠れがもたらす面白さを体験できる。

6 考察

6.1 ディスプレイ部の自動回転操作の問題点

本提案デバイスの、初期のプロトタイプでは、モータによるディスプレイ部の左右への回転をユーザの手元のボタン操作でも行えるようにしていた。しかし、これは展示活動を通して得た知見であるが、自動回転操作ができるなどを知らずに遊んでいたユーザはディスプレイの隠れを身体を動かして能動的に見にいきながらアプリを楽しんでいたにも関わらず、自動回転操作ができるなどを伝えた途端、身体を動かすことなく、自動回転操作のみで隠れを見るようになってしまった。これでは、本研究が提案する楽しさをユーザは体験することができないと考え、ディスプレイ部の自動回転機能は、ユーザの操作ではなく、アプリ側の制御のみで使用できるように現在は設定している。

6.2 隠れのないディスプレイ形状の問題点

隠れのないディスプレイ形状として、半球形状ディスプレイについての検討を行った。しかし、ユーザはディスプレイの映像を一覧できるため、ディスプレ

イ上方からの視界を維持したまま、能動的に動くことなく遊ぶようになってしまったと考えた。そのため本デバイスのディスプレイ形状には、立方体や、くびれのある球形状といった、どの角度から見ても一覧できないものが適切だと考えている。

6.3 操作における制限について

ユーザが反対側の面を覗き込んだ場合等で、キャラクタの左右の移動方向とユーザの操作が意図と異なってしまう問題がある。この問題は、固定型ゲームコントローラの特性上、慣れることによって解決すると考え、実際ある程度プレイを続けると、反転操作にも慣れ、問題なくプレイできる傾向がみられた。また、一般的なトップビューのゲームではフィールドの全体が見えており、かつ、キャラクタが平面上に存在することから、直線的な弾道予測は難しくない。うってかわって、全周囲ディスプレイ上においては立体形状や隠れた部分の存在によって弾道予測の難易度が高くなっている。しかし、適切なゲームバランスの設定を行えば、ゲームのやりがいに繋げられる長所だと我々は捉えている。

6.4 プロトタイプの制限

使用したプロジェクタの遅延により、自動回転中にディスプレイに対して映像が遅れ、表示がずれてしまう問題がある。これを解決するためには、低遅延の投影が可能なプロジェクタを用いるか、プロジェクタの投影遅延と回転速度を考慮した予測投影を行う必要があると考える。また、レンズ周辺部の解像度・画質の低下も顕著である。実装を行ったゲームでは、背景とキャラクタの色をコントラストの高い配色にし、キャラクタのグラフィックスもシンプルにすることで視認性の向上を図っている。またUPLIGHTは携帯ゲーム機のように、デバイスを目の前まで持ち上げて遊ぶプレイスタイルであり、加えて覗き込みなどの手や上半身を動かす運動を頻繁に行う必要がある。そのため、出来るだけ軽量で取り回しやすいのが望ましい。ただ、プロトタイプは2.5kgと、市販の携帯ゲーム機(0.2~0.4kg程度)と比べて重く、また、筐体外部の計算機とケーブルで繋がれ、取り回しにくい。そのため、より良い体験を実現するためには軽量化や無線化が必要だと考えられる。

7 結論と今後の展望

本論文では、全周囲ディスプレイの隠れに着目した、シングルユーザのインタラクションに基づいた新たな全周囲ディスプレイデバイスを開発・提案した。今後は、本インタラクションのエンタテインメント性を実験により検証していく。また、視覚や聴覚だけでなく、触覚や嗅覚を組み合わせるなど、ユーザの情報探索行為の能動性をより高める手法を検討していく。

参考文献

- [1] H. Benko and A. D. Wilson. *Pinch-the-Sky Dome: Freehand Multi-Point Interactions with Immersive Omni-Directional Data*, p. 3045–3050. Association for Computing Machinery, New York, NY, USA, 2010.
- [2] H. Benko, A. D. Wilson, and R. Balakrishnan. Sphere: Multi-Touch Interactions on a Spherical Display. In *Proceedings of the 21st Annual ACM Symposium on User Interface Software and Technology*, UIST '08, p. 77–86, New York, NY, USA, 2008. Association for Computing Machinery.
- [3] G. Cortes, E. Marchand, G. Brincin, and A. Lecuyer. MoSART: Mobile Spatial Augmented Reality for 3D Interaction With Tangible Objects. In *Frontiers in Robotics and AI*. Association for Computing Machinery, 2018.
- [4] C. Cruz-Neira, D. J. Sandin, T. A. DeFanti, R. V. Kenyon, and J. C. Hart. The CAVE: Audio Visual Experience Automatic Virtual Environment. *Commun. ACM*, 35(6):64–72, June 1992.
- [5] A. Gomes, L. L. Priyadarshana, A. Visser, J. P. Carrascal, and R. Vertegaal. Magicscroll: A Rollable Display Device with Flexible Screen Real Estate and Gestural Input. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '18, New York, NY, USA, 2018. Association for Computing Machinery.
- [6] Z. Li, S. Miyafuji, T. Sato, H. Kuzuoka, and H. Koike. OmniEyeball: Spherical Display Equipped With Omnidirectional Camera And Its Application For 360-Degree Video Communication. In *The 31st Annual ACM Symposium on User Interface Software and Technology Adjunct Proceedings*, UIST '18 Adjunct, p. 33–35, New York, NY, USA, 2018. Association for Computing Machinery.
- [7] T. Machida. GEO-COSMOS: World's First Spherical Display. In *ACM SIGGRAPH 2002 Conference Abstracts and Applications*, SIGGRAPH '02, p. 189, New York, NY, USA, 2002. Association for Computing Machinery.
- [8] P. Pla and P. Maes. *Display Blocks: Cubic Displays for Multi-Perspective Visualization*, p. 2015–2020. Association for Computing Machinery, New York, NY, USA, 2012.
- [9] K. Ryota. loopScape. Japan, 2003.
- [10] J. R. Williamson, J. Williamson, D. Sundén, and J. Bradley. Multi-Player Gaming on Spherical Displays. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '15, p. 355–358, New York, NY, USA, 2015. Association for Computing Machinery.
- [11] Y. Zhou, S. Xiao, N. Tang, Z. Wei, and X. Chen. *Pmomoo: Projection Mapping on Movable 3D Object*, p. 781–790. Association for Computing Machinery, New York, NY, USA, 2016.
- [12] 東海林 和, 中村 俊勝, 三村 巧, 佐藤 俊樹. ユーザの自発的移動を促進させる全周囲ディスプレイ. WISS2017 予稿集. 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, 2017.

能動的音響計測に基づくスタイラスの把持状態識別手法の検討

高倉 礼^{*} 鈴木 健介[†] 國分 晴利^{*} 志築 文太郎[‡]

概要. 本稿では、能動的音響計測および機械学習を用いて、ユーザによるスタイラスの把持状態を識別する手法を示す。特に、スタイラスの内部構造の違いにより音響信号の伝わり方が異なることに着目し、我々は3Dプリント時の充填率を変えることによりスタイラスの内部構造を変え、把持状態の識別精度がどのように変わるか調査した。結果、充填率が25%, 50%, および75%のスタイラスにおいて6つの把持状態を95%程度にて識別できること、および充填率100%のスタイラスでは把持状態の識別精度が89.8%となることがわかった。この結果に基づき、能動的音響計測に基づく把持状態が識別可能なスタイラスを作製する場合における設計指針を述べる。

1 はじめに

タブレット端末を用いてメモをとる時、およびイラストを描く時には主にスタイラスが用いられる。多くのスタイラスは筆圧および傾きを検知し、実世界における鉛筆や筆のように描く線の太さおよび濃さを調節できる。しかし、マーカや消しゴムといった描画ツールを変更する時、および描画色を変更する時には、アプリケーション内にある多くのメニュー群の中から目的の描画ツールおよび色を選択する必要があり、操作に時間がかかる。Apple Pencil [1] はスタイラスにてダブルタップジェスチャを行うことにより、Bamboo Fineline [22] はスタイラスに設けられているボタンを押すことにより、描画ツールを高速に切替えられる。しかし、ユーザは現在使用している描画ツールを把握しておく必要があるため、描画作業を再開する時には適宜メニューを確認しなければならない。

我々はこの課題に対して筆記用具の把持方法に着目した。筆記用具にはそれぞれ特有の把持方法がある。例えば、鉛筆を持つ時は1本掛け、習字にて筆を持つ時は2本掛け（双鉤法とも呼ばれる）といった把持方法が挙げられる。また、消しゴム付きの鉛筆を利用するときは逆さ持ちをすれば消しゴムが使える。これらの把持方法のメタファに基づいて、把持状態に応じた描画ツールを使用できれば、ユーザはアプリケーション内のメニューを確認せずとも現在使用しているスタイラスの描画ツールを把握できると考えられる。そこで我々は、ユーザによるスタイラスの把持状態に着目し、指の1本掛けでスタイラスを把持すれば鉛筆、2本掛けで把持すれば筆、逆持ちで把持すれば消しゴムといったように、描画ツールを把持状態によって切替えられる手法を考案



図 1. 把持状態およびその周波数応答。左から順に1本掛け、2本掛け、握り持ち。



図 2. スタイラス把持状態識別システム。

した。既存のスタイラスの筆圧および傾きを検知する機能をそのまま活かすために、我々はスタイラスの物理的な形状を制限しない能動的音響計測 [6, 16, 24] を利用することにより、把持状態を識別できるか検討した。本手法の概念実証を示すために、今回我々はオーディオインターフェースを用いた試作システムを作成した（図 1 および図 2）。

能動的音響計測とは、物体に振動を与える、物体を伝搬した振動を解析し、その周波数応答に基づき物体の状態を識別および推定する手法である。特に物体の材質および内部構造の違いにより周波数応答は異なる [12]。そのため、スタイラスの材質および内部構造が把持状態の識別精度へ影響を与えると考えられる。しかし、市販のスタイラスは材質および内部構造が多様なため、どの要因が把持状態識別へ影響を与えるのか特定することが難しい。したがって、

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 情報科学類

‡ 筑波大学 システム情報系

我々は同じ材質および大きさにて異なる内部構造の物体を作成できる3Dプリンタを使用してスタイラスを作成した。これにより、内部構造の違いのみに着目してスタイラスの把持状態識別への影響を検証できる。本稿ではまず、能動的音響計測を用いてスタイラスの把持状態を識別する手法を述べる。その後、充填率の異なるスタイラスにおける把持状態の識別精度の評価および結果を述べたのち、能動的音響計測に基づく把持状態識別を行う場合におけるスタイラスの設計指針を示す。

本研究の貢献を以下に示す。

- 3Dプリントしたスタイラスを用いて、能動的音響計測により最大95.7%の精度にて6種類の把持状態を識別できることを示した。
- 把持状態の識別精度が、充填率25%の時に最大95.5%，充填率100%の時に最大89.8%であることから、充填率の違いにより把持状態の識別精度へ影響があることを示した。
- 能動的音響計測により把持状態が識別できるスタイラスを作製する場合における設計指針を示した。

2 関連研究

本手法は能動的音響計測により、ユーザによるスタイラスの把持状態を識別する手法である。本節では関連研究として、能動的音響計測により物体の状態を識別する手法、およびスタイラスの入力を拡張する手法をそれぞれ述べる。

2.1 能動的音響計測により物体の状態を識別する手法

能動的音響計測とは、物体に振動を与え、物体を伝搬した振動を解析し、その周波数応答に基づき物体の状態を識別する手法である[6, 13, 16, 17, 23–28]。

Onoら[16]は能動的音響計測を用いて、物体上の触れている場所および把持力を識別できることを示した。岩瀬ら[24]は能動的音響計測を用いて、スピーカとマイクとの間に置かれた物体の位置を推定できることを示した。Hirakiら[6]は能動的音響計測によりタッチ点を推定する手法を示し、プロジェクタが壁に投影した映像に対してタッチ入力ができるシステムを示した。これらの研究から、能動的音響計測を用いることによりスタイラスに触れている指の位置が推定できると考えられるため、ユーザによるスタイラスの把持状態が識別できると考えられる。しかし、スタイラスの1本掛けおよび2本掛けは近い姿勢であり、また人によって指の掛け方は異なる。したがって本研究では、スタイラスの指の掛け方および人ごとの握り方の違いが、把持状態の識別へどのように影響を与えるか調査する。

Kuboら[12, 13, 25]は能動的音響計測を用いることにより、同じ外見を持つ内部構造が異なる物体を識別できることを示した。物体の内部構造の違いにより周波数応答が変化することから、物体の内部構造により周波数応答の把持状態ごとの変化も異なると考えられる。そのため、この周波数応答の変化の違いが把持状態の識別精度に影響を与えると考えられる。したがって、本研究はスタイラスの内部構造（本稿では充填率）の違いにより把持状態の識別精度がどのように変わるか調査する。

2.2 スタイラスの入力を拡張する手法

スタイラスに追加のセンサを搭載することにより、スタイラスの入力を拡張する手法が多く研究されてきた[2, 3, 5, 7, 8, 14, 15]。Matulicら[14]はスタイラスの後端に取り付けた魚眼カメラにて両手を含む周囲の様子を撮影し、ジェスチャ認識によりスタイラスの入力を拡張する手法を示した。Hwangら[7]はスマートフォンに搭載された磁気センサ、およびスタイラスに取り付けた永久磁石を利用し、Hinckleyら[5]は慣性センサおよび静電容量センサを用いて、タブレット端末およびスタイラスの状態を識別し、状況に応じたインタラクションができる手法を示した。これらの研究では、センサを決められた場所へ取り付ける必要があり、既存のスタイラスの物理的な形状を維持することが難しい。

一方、追加のセンサを用いず受動素子のみを用いてスタイラスの入力を拡張する手法が研究されてきた[9, 20]。Ikematsuら[9]は、スタイラスと静電容量式タッチディスプレイとの間に抵抗を介在させ、スタイラスの入力を拡張する手法を示した。Schmitzら[20]は、底面に導電性素材のパターンを形成した物体を作成し、異なるパターンを有する物体をタッチディスプレイ上にて識別する手法を示した。これらの研究は、追加のセンサを必要とせず安価にスタイラスの入力語彙を拡張できる。ただし、スタイラスの把持状態を識別する場合、受動素子のみを用いたスタイラスでは厳密な内部構造の設計が必要となり、スタイラスのデザインの自由度を制限する。

本節にて述べた研究では、把持状態を識別するにあたり、スタイラスの物理的な形状を制限する。能動的音響計測ではセンサを自由に配置できるため、これらの研究と比較して物理的な形状に対する制限が小さい点にて優れる。

3 実装

本節では、本手法を実現するシステムの実装方法を述べる。本システムは、音響信号を送波および受波するハードウェア、ならびに音響信号を生成および解析するソフトウェアからなる。システムの全体の構成図を図3に示す。なお、PythonおよびQtライブラリを用いてソフトウェアを開発した。ソフト

能動的音響計測に基づくスタイラスの把持状態識別手法の検討

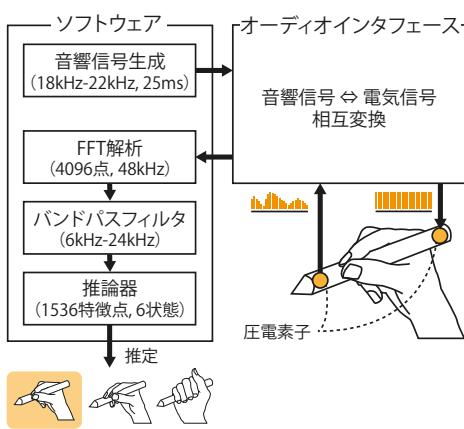


図 3. 能動的音響計測に基づくスタイラスの把持状態識別システムの構成図。

ウェアを実行するコンピュータには MacBook Air (CPU: Intel Core i7-1060NG7, RAM: 16GB) を用いた。

先行研究に倣い [10, 28], 音響信号には 18 kHz–22 kHz を 25 ms にて掃引するアップチャープ信号を用いた。使用する音響信号の周波数帯は人間が聞こえづらい周波数帯であり [21], またスタイラスを触った時の雑音および環境音 (4 kHz 以下) の影響を受けづらい。ソフトウェアにて生成された音響信号は、オーディオインターフェース (Steinberg UR24C) を介して、スタイラス後端に貼り付けられた圧電素子 (THRIVE K2512BS1) からスタイラスへ送波される。スタイラスの内部を伝播した音響信号はスタイラス先端に取り付けられた圧電素子にて受波され、オーディオインターフェースを介してソフトウェアへ送られる。得られた音響信号はサンプリング周波数 48 kHz, 4096 点の高速フーリエ変換 (FFT: Fast Fourier Transformation) により周波数パワースペクトルへ変換される。なお、抽出後の標本データには窓関数としてハニング窓を掛ける。その後、周波数パワースペクトルからサンプリング定理に基づき 24 kHz 以上の周波数特徴点を、外乱の影響を除くために 6 kHz 以下の周波数特徴点をそれぞれ取り除くバンドパスフィルタを適用する。最後に、得られた 1536 点の特徴点を持つ周波数パワースペクトルが機械学習アルゴリズムを用いた推論器にて処理され、把持状態の識別が行われる。本研究において、機械学習アルゴリズムは scikit-learn ライブライ [18] の Support Vector Classifier (SVC), および Random Forest (RF), ならびに LightGBM ライブライ [11] の Gradient Boosting Decision Tree (GBDT) から選択された。なお、上記の一連の処理は約 50 ms の時間にて完了する。

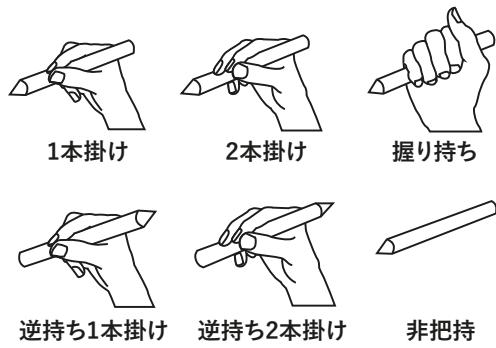


図 4. 提案手法にて識別する 6 つの把持状態。

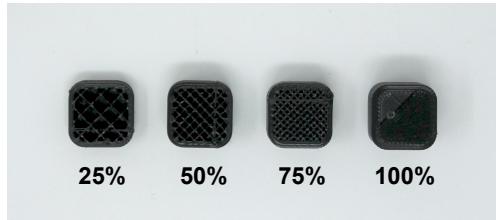


図 5. スタイラスの内部構造。左から充填率 25%, 50%, 75%, 100%。充填構造は全て Grid。

4 評価実験

提案手法による把持状態の識別精度を評価するために、データの収集を行った。把持状態には順持ちの 1 本掛け, 2 本掛け, 逆持ちの 1 本掛け, 2 本掛け, ならびに握り持ちおよび非把持の 6 つを選んだ (図 4)。実験には、4 名の大学生および大学院生 (22 歳–25 歳, 平均 23.5 歳, 全員男性, 右利き) がボランティアにて参加した。我々は 3D プリンタ (Ultimaker 3 Extended) を用いて、実験に使用するスタイラスを作製した。なおスタイラスを使用してタッチパネルの画面をタッチできるように、導電性フィラメント (ProtoPasta Conductive PLA [19]) を用いた。圧電素子を直接スタイラスへ貼り付けると電気的に圧電素子間が短絡してしまうため、圧電素子には厚さ 0.5 mm のアクリル板を貼り付けた。

スタイラスの内部構造の違いによる把持状態の識別精度への影響を調べるために、本研究ではまず充填率の違いによる内部構造の違いに着目した。そのため、我々は充填率の異なる 4 種類のスタイラス (図 5, それぞれ充填率 25%, 50%, 75%, 100%, すべて内部充填構造は Grid, 長さ 150 mm × 幅 10 mm × 奥行き 10 mm) を作成し、それぞれ評価を行った。

4.1 設計

参加者は実験全体を通して、椅子に着席した状態にて実験を行った。まず、実験者は 4 種類のスタイラスのうち 1 つのスタイラスを選んだのち、そのスタイラスの先端および後端へ両面テープ (3M, 2477P-

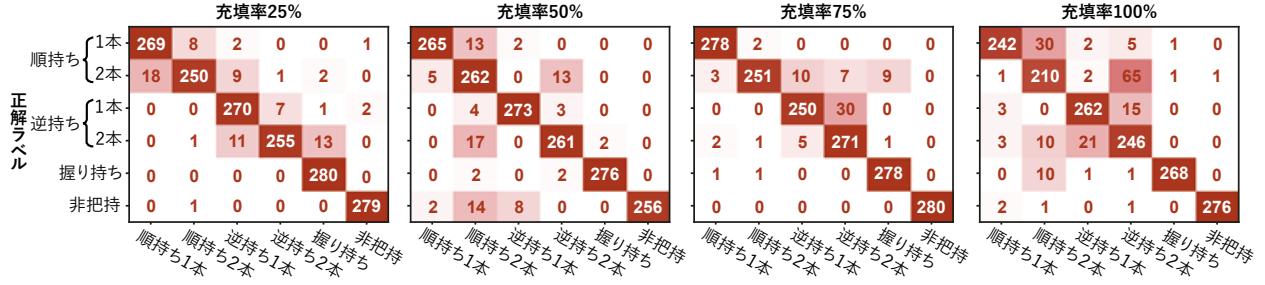


図 6. 充填率が異なるスタイラスそれぞれにおける、Gradient Boosting Decision Tree を用いた把持状態識別の混同行列。

表 1. スタイラスの充填率ごとの把持状態識別精度（試行間 5 分割交差検証）。

充填率	SVC	RF	GBDT
25%	84.3%	93.1%	95.5%
50%	92.5%	94.9%	94.8%
75%	89.1%	95.6%	95.7%
100%	69.2%	89.8%	84.5%

S) を用いて圧電素子を取り付ける。このとき、先端側がマイク、後端側がスピーカとなるように圧電素子を取り付ける。その後、実験者は参加者へ圧電素子を取り付けたスタイラスを手渡す。実験者は 6 種類の把持条件から 1 つの把持条件を選択し、その把持条件にてスタイラスを右手で把持するように参加者へ指示する。なお、非把持条件では、スタイラスを木製の机の上に横向きに置くように指示する。実験者は参加者が指定した把持条件にてスタイラスを把持したことを確認した後、スタイラスへ音響信号を送波し 3 秒間録音する。その後、実験者は参加者へスタイラスを一度机の上に置き、持ち替えるように指示する。これを同じ把持条件にて 5 試行行う。5 試行の計測後、実験者はスタイラスから圧電素子を取り外す。

上記の手順を 4 種類のスタイラスおよび 6 種類の把持条件それぞれ 2 セットずつ行ったため、参加者 1 人あたりスタイラス 4 種類 × 把持条件 6 種類 × 5 試行 × 2 セット = 240 試行分の波形データが得られた。なお、スタイラスおよび把持条件の選択順序は参加者ごとにランダム方格法を用いて決定され、順序効果が取り除かれた。圧電素子の貼替え時に 1 分程度の時間を要したため、実験全体にて参加者 1 人あたりが要した時間は 60 分程度であった。

4.2 結果

実験にて得られた波形データは 4096 点(約 85 ms)ごとにスペクトルデータへ変換されたため、各スタイラスの各把持姿勢ごとに 1400 個のスペクトル

表 2. スタイラスの充填率ごとの把持状態識別精度 (leave-one-participant-out 交差検証)。

充填率	SVC	RF	GBDT
25%	29.6%	24.9%	24.2%
50%	36.9%	39.8%	38.9%
75%	28.4%	22.9%	30.3%
100%	28.6%	23.5%	25.1%

データが得られた。このスペクトルデータを用いて、Google Colaboratory [4] にて把持状態の識別精度を評価した。なお、学習時にはライブラリのデフォルトのハイパーパラメタを使用した。それぞれのスタイラスごとに学習および推論を行い、識別精度を算出した。

能動的音響計測により把持状態が識別できるか評価するために試行間に 5 分割交差検証を行った。加えて、未知のユーザにて推論モデルが適用できるか評価するために leave-one-participant-out 交差検証を行った。各スタイラスにおける把持状態の識別精度を表 1 および表 2 に示す。

試行間の交差検証において、どの充填率のスタイラスにおいても高い精度にて把持状態を識別できることがわかった。しかし、充填率 75%以下のスタイラスと比較して、充填率 100%のスタイラスはどのアルゴリズムでも識別精度が低かった。一方で、leave-one-participant-out 交差検証においては識別精度が低く、既存のモデルを用いて未知のユーザの把持状態を識別することが困難であることがわかった。

試行間の交差検証において、GBDT アルゴリズムの最も精度が高かったモデルにおける混同行列を図 6 に示す。混同行列より、全てのスタイラスにおいて、把持状態が近いもの（例えば、1 本掛けと 2 本掛け）にてしばしば誤識別をしていることがわかった。加えて、充填率 100%のスタイラスでは、同じ指の本数の持ち方、特に 2 本掛けおよび逆持ち 2 本掛けにてよく誤識別していることがわかった。

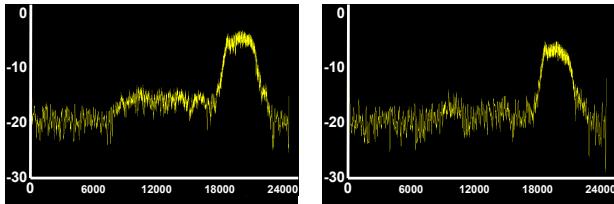


図 7. 充填率 25% のスタイラスにおける周波数パワー スペクトル。縦軸は音の強さ (dB : デシベル), 横 軸は周波数 (Hz : ヘルツ) を表す。左が 1 本掛け、 右が 2 本掛け。

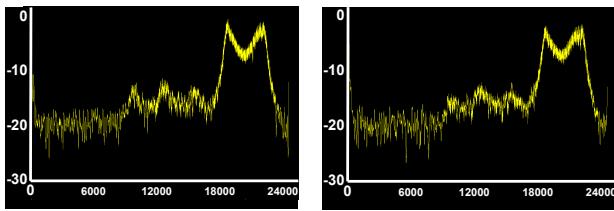


図 8. 充填率 100% のスタイラスにおける周波数パワー スペクトル。縦軸は音の強さ (dB : デシベル), 横 軸は周波数 (Hz : ヘルツ) を表す。左が 1 本掛け、 右が 2 本掛け。

5 議論および今後の課題

5.1 充填率の影響

実験の結果より、どの充填率のスタイラスにおいても高い精度にて把持状態を識別できることがわかった。特に、充填率が 75%以下のスタイラスは識別精度が 95%程度ととても高い。一方で、充填率が 100%のスタイラスは最大識別精度が 89.8%であり、他のスタイラスと比較して識別精度が低くなることが示された。そこで、充填率が 25%および充填率 100%のスタイラスにおいて、1 本掛けおよび 2 本掛けそれぞれの周波数パワースペクトルの変化を観察した(図 7 および図 8)。充填率 25%のスタイラスにおいては 1 本掛けの時に周波数 8 kHz–18 kHz にかけてなだらかな山があり、かつ周波数 18 kHz–22 kHz まで強さが一定の山が観察できる。一方、2 本掛けの時は周波数 8 kHz–18 kHz の山が全体的に小さくなり、かつ周波数 22 kHz 付近の強さが小さくなるという大きなスペクトルの変化が観察できた(図 7)。充填率 100%のスタイラスは、充填率 25%のスタイラスと比較し、1 本掛けと 2 本掛けのスペクトルの変化が小さいことが観察できた(図 8)。

これは充填率が高いほど内部空間の変形余地が小さくなるために、スペクトルの変化が小さくなつたからであると考えられる。内部空間がある場合、物体はしなりやすく変形余地がある。把持により押さえられた箇所は変形余地が小さくなり振動が伝わりづらくなる。その結果、充填率 25%のスタイラスにおいて、1 本掛けよりも 2 本掛けのスペクトルが全

表 3. スタイラスの充填率ごとのユーザ識別精度 (試行 間 5 分割交差検証)。

充填率	SVC	RF	GBDT
25%	88.1%	90.6%	89.2%
50%	99.6%	97.6%	98.0%
75%	99.2%	99.7%	99.4%
100%	99.6%	98.5%	97.9%

体的に大きく減衰するように変化したと考えられる(図 7)。一方で、内部空間が小さいまたは無い場合、物体は硬くなり変形余地が小さくなる。把持により押さえられた箇所はさらに変形余地が小さくなるものの、十分に内部空間がある場合よりもその変形余地の変化の程度は小さいものとなる。そのため、充填率 100%のスタイラスはスペクトルの変化が小さくなったと考えられる。

なお、本稿ではスタイラスを用いて描いている最中の把持状態は学習していない。描くときにペン先がタッチパネルに触ることにより境界条件が変化すること、および描画時の筆圧により指とスタイラスの接触状態および把持力が変化することから、音響スペクトルも変化すると考えられる。そのため、描画中の把持状態の変化も考慮した推論モデルを作成することは今後の課題である。

5.2 ユーザの識別

leave-one-participant-out 交差検証の結果から、学習したモデルから未知のユーザの把持状態を識別することは困難であることがわかった。これは、ユーザごとにスタイラスの持ち方が大きく異なるためであると考えられる。一方で、試行間の把持状態の識別精度が 95%程度と高いことから、利用するユーザの把持状態を学習させれば、スタイラスを把持しているユーザがどのユーザであるのかを識別できると考えられる。ユーザが識別できれば、ユーザごとにスタイラスの把持状態による描画ツールの割り当てを変更するなどの機能が実装できるため有用である。そこで、実験にて収集したデータを用いて、能動的音響計測にてスタイラスを把持しているユーザを識別できるか検証した。

試行間の 5 分割交差検証を行ったユーザ識別精度を表 3 に示す。充填率 25%以外では識別精度が 99%を超える結果となった。したがって、スタイラスを把持しているユーザの識別ができるといえる。一方で、充填率 25%のスタイラスにて識別精度が低く、充填率 100%のスタイラスにて識別精度が高いという把持状態識別とは異なる傾向となった。このことから、充填率 25%のスタイラスでは、ユーザごとの持ち方によるスペクトルの変化よりも把持状態によるスペクトルの変化が大きく反映されると考えら

れる。そのため、ユーザ識別および把持状態識別を両立させるためには、どちらも識別精度が高い充填率 50% および 75% の構造が良いと考えられる。しかし、今回は実験参加者が 4 人と少ないとことから、より多くのユーザのデータを収集し検証する必要がある。

5.3 スタイラスの設計

今回は 3D プリントした物体の充填率を変えた場合のみを検証した。5.1 節にて述べたように、能動的音響計測を用いてスタイラスの把持状態を識別する場合は、例えば、変形しづらい硬い素材を使うのではなく変形させやすいプラスチックを用いるといった、内部空間に変形余地を残すように設計することがより望ましいと考えられる。ただし、先行研究 [16, 17, 24, 28] にて述べられているように、振動を吸収してしまう素材には本手法は適用できない。そのため、スタイラスが変形しやすいうように外装を薄くするといった、設計上の工夫が必要であると考えられる。また、充填率だけでなく、スタイラスに穴を設ける、または一部外装を薄くするなどの工夫をすることにより、把持状態ごとのスペクトル変化が大きくできるため、より多くの把持状態を識別できる可能性がある。一方で、本手法はスタイラスの内部空間さえあれば内部構造を問わずに高い精度にて把持状態を識別できる。そのため、物体内部には市販のスタイラス同様の回路および機能を設けられる。したがって、本手法によりさらに多くの把持状態を探究するよりも、本手法と既存のスタイラス機能を組み合わせて利用することが望ましいと考えられる。

今回、提案手法の概念実証のために据え置き型のオーディオインターフェースを使用した。ただし、今回使用したオーディオインターフェースが為す役割は音響信号および電気信号の相互変換に留まっており、より小型な Analog Digital Converter および Digital Analog Converter でもこの役割をこなせる。また、音響処理に要する計算量は小さく、計算資源に乏しいマイクロコントローラでも特別な電子部品なしに実行可能である。さらに、本システムにて使用しているケーブルの役割は電気回路の導線の役割に留まっており、細いケーブルを使用する、または導線にて直接マイクロコントローラに接続することにより、ケーブル部をスタイラスに内蔵できる。これらのハードウェアの設計の工夫により、システムをスタンドアロン化できると考えられる。そのため、スタンドアロン化し実使用に即したシステム構成にすることは今後の課題である。

6 応用例

本手法の応用例として、ペイントアプリケーションが考えられる。我々はタッチパネルを有するコン



図 9. 本手法のペイントアプリへの応用。1 本掛けにて把持すれば細い線を描くことができ、2 本掛けにて把持すれば太い線を描くことができる。

ピュータ (ThinkPad T14 Gen2) にて本手法を導入したペイントアプリケーションを実装した (図 9)。このアプリケーションではスタイラスの把持状態に応じて、1 本掛けなら鉛筆のように細い線、2 本掛けなら筆のように太い線、逆持ち 1 本掛けなら消しゴムと描画ツールを切替えられる。試作システムは使用時にケーブルが干渉し描画を妨げてしまうことがあるため、使用に向かない。そのため、圧電素子およびケーブルをスタイラスの内部に入れるなどの工夫が必要である。加えて、より携帯性を高めるために、システムを小型化しタブレット端末にて使えるようにすること、およびユーザ実験から本アプリケーションの使用感を評価することが今後の課題である。

7 おわりに

本稿では、能動的音響計測および機械学習を用いて、ユーザによるスタイラスの把持状態を識別する手法を示した。加えて我々は、スタイラスの内部構造の違いにより音響信号の伝わり方が異なることに着目し、3D プリント時の充填率に基づきスタイラスの内部構造を変えることによって、把持状態の識別精度がどのように変わるか調査した。結果、充填率が 25%, 50%, および 75% のスタイラスにおいて 6 つの把持状態を 95% 程度にて識別できること、および充填率 100% のスタイラスでは把持状態の識別精度が 89.8% 程度と低くなることがわかった。このことから、内部に空間を設けられたスタイラスにて本手法を用いることが望ましいと考えられる。

なお、leave-one-participant-out 交差検証の結果、どのスタイラスにおいても識別精度が 30% 程度であったことから、未知のユーザに対する識別可能性が低いことが示された。一方、追加の調査にて、把持姿勢から 4 名のユーザを最大 99.7% の精度にて識別できた。このことから、本手法を用いて、スタイラスを利用するユーザごとに異なる描画ツールの割り当てを提供するなどのシステム設計が可能であると考えられる。

今後は本手法を用いて、市販のスタイラスにおける把持状態識別の精度を調査する予定である。

謝辞

本研究は JSPS 科研費 JP18H04104 の助成を受けたものです。

参考文献

- [1] Apple Inc. Apple Pencil - Apple, 2021. <https://www.apple.com/jp/apple-pencil/> (最終参照日 : 2021 年 11 月 16 日) .
- [2] D. Cami, F. Matulic, R. G. Calland, B. Vogel, and D. Vogel. Unimanual Pen+Touch Input Using Variations of Precision Grip Postures. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*, UIST '18, pp. 825–837. ACM, 2018.
- [3] N. Fellion, T. Pietrzak, and A. Girouard. FlexStylus: Leveraging Bend Input for Pen Interaction. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, UIST '17, pp. 375–385. ACM, 2017.
- [4] Google LLC. Colaboratory - Google, 2021. <https://colab.research.google.com/> (最終参照日 : 2021 年 11 月 16 日) .
- [5] K. Hinckley, M. Pahud, H. Benko, P. Irani, F. Guimbretière, M. Gavriliu, X. A. Chen, F. Matulic, W. Buxton, and A. Wilson. Sensing Techniques for Tablet+Stylus Interaction. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, UIST '14, pp. 605–614. ACM, 2014.
- [6] T. Hiraki, M. Fukumoto, and Y. Kawahara. Touchable Wall: Easy-to-Install Touch-Operated Large-Screen Projection System. In *Proceedings of the 2018 ACM International Conference on Interactive Surfaces and Spaces*, ISS '18, pp. 465–468. ACM, 2018.
- [7] S. Hwang, A. Bianchi, M. Ahn, and K. Wohm. MagPen: Magnetically Driven Pen Interactions on and around Conventional Smartphones. In *Proceedings of the 15th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '13, pp. 412–415. ACM, 2013.
- [8] S. Hwang, A. Bianchi, and K. Wohm. MicPen: Pressure-Sensitive Pen Interaction Using Microphone with Standard Touchscreen. In *Proceedings of Extended Abstracts of the 2012 CHI Conference on Human Factors in Computing Systems*, CHI EA '12, pp. 1847–1852. ACM, 2012.
- [9] 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*, pp. 1–8. ACM, 2018.
- [10] S. Katakura and K. Watanabe. ProtoHole: Prototyping Interactive 3D Printed Objects Using Holes and Acoustic Sensing. In *Proceedings of Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI EA '18, pp. 1–6. ACM, 2018.
- [11] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T. Liu. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In *Advances in Neural Information Processing Systems*, Vol. 30, pp. 1–9. Curran Associates, Inc., 2017.
- [12] Y. Kubo, K. Eguchi, and R. Aoki. 3D-Printed Object Identification Method Using Inner Structure Patterns Configured by Slicer Software. In *Proceedings of Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI EA '20, pp. 1–7. ACM, 2020.
- [13] Y. Kubo, K. Eguchi, R. Aoki, S. Kondo, S. Azuma, and T. Indo. FabAuth: Printed Objects Identification Using Resonant Properties of Their Inner Structures. In *Proceedings of Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI EA '19, pp. 1–6. ACM, 2019.
- [14] F. Matulic, R. Arakawa, B. Vogel, and D. Vogel. PenSight: Enhanced Interaction with a Pen-Top Camera. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, pp. 1–14. ACM, 2020.
- [15] F. Matulic, B. Vogel, N. Kimura, and D. Vogel. Eliciting Pen-Holding Postures for General Input with Suitability for EMG Armband Detection. In *Proceedings of the 2019 ACM International Conference on Interactive Surfaces and Spaces*, ISS '19, pp. 89–100. ACM, 2019.
- [16] M. Ono, B. Shizuki, and J. Tanaka. Touch & Activate: Adding Interactivity to Existing Objects Using Active Acoustic Sensing. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, UIST '13, pp. 31–40. ACM, 2013.
- [17] M. Ono, B. Shizuki, and J. Tanaka. Sensing Touch Force Using Active Acoustic Sensing. In *Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction*, TEI '15, pp. 355–358. ACM, 2015.
- [18] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- [19] Proto-pasta. Conductive PLA Technical Data Sheet Rev 1. - Proto-pasta, 2021. https://cdn.shopify.com/s/files/1/0717/9095/files/TDS_Conductive_PLA_1.0.1.pdf (最終参照日 : 2021 年 11 月 16 日) .
- [20] M. Schmitz, F. Müller, M. Mühlhäuser, J. Riemann, and H. V. V. Le. Itsy-Bits: Fabrication and Recognition of 3D-Printed Tangibles with Small Footprints on Capacitive Touchscreens. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, 2021.
- [21] The Physics Factbook. Frequency Range of Human Hearing - The Physics Factbook,

2003. <https://hypertextbook.com/facts/2003/ChrisDAmbrose.shtml> (最終参照日: 2021年11月16日).
- [22] Wacom Co., Ltd. Bamboo Fineline - Wacom, 2021. <https://www.wacom.com/ja-jp/products/stylus/bamboo-fine-line> (最終参照日: 2021年11月16日).
- [23] K. D. D. Willis and A. D. Wilson. InfraStructs: Fabricating Information inside Physical Objects for Imaging in the Terahertz Region. *ACM Transactions on Graphics*, 32(4):1–10, 2013.
- [24] 岩瀬大輝, 伊藤雄一, 秦秀彦, 尾上孝雄. SenseSurface : アクティブ音響センシングによる物体識別と位置推定. 情報処理学会論文誌, 60(10):1869–1880, 2019.
- [25] 久保勇貴, 江口佳那, 青木良輔, 近藤重邦, 東正造, 犬童拓也. 内部構造パターンの差異を利用した3Dプリントオブジェクト識別手法. 第27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 (WISS2019) 論文集, pp. 43–48. 日本ソフトウェア科学会, 2019.
- [26] 西井遙菜, 双見京介, 村尾和哉. アクティブ音響センシングを用いた食材認識カトラリー. 第28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 (WISS2020) 論文集, pp. 55–60. 日本ソフトウェア科学会, 2020.
- [27] 川崎祐太, 伊藤雄一, 藤田和之, 尾上孝雄. アクティブ音響センシングを用いた物体情報識別における環境温度変化に関する一検討. ヒューマンインターフェース学会研究会研究報告集, 22(6):55–60, 2020.
- [28] 大野誠, 志築文太郎, 田中二郎. アクティブ音響センシングを用いた把持状態認識. 第17回一般社団法人情報処理学会シンポジウムインタラクション 2013 論文集, pp. 56–63. 情報処理学会, 2013.

著者索引

- Jayakorn Vongkulbhaisal, 97
Nolwenn Maudet, 30
浅川 智恵子, 97
五十嵐 健夫, 9, 30, 78, 85, 113
生田 泰章, 35
池松 香, 105
石井 紗郁, 105
石黒 祥生, 71
伊藤 思音, 128
伊藤 貴之, 1
井原 圭一, 50
Erwin Wu, 64
梅谷 信行, 78
遠藤 輝貴, 9
大西 鮎美, 57
岡部 誠, 91
加藤 邦拓, 105
加藤 達也, 128
粥川 青汰, 97
川口 一画, 50
川谷 知寛, 43
川原 圭博, 105
栗林 雅希, 97
栗原 一貴, 43
小池 英樹, 64
國分 晴利, 135
後藤 真孝, 1
坂名 純太, 16
佐藤 大介, 97
佐藤 俊樹, 128
佐野 正太郎, 113
椎尾 一郎, 71, 105
志築 文太郎, 135
鈴木 健介, 135
住田 桃子, 30
高木 啓伸, 97
高倉 礼, 135
高島 健太郎, 35
武田 一哉, 71
張 家銘, 85
塚田 浩二, 43
塚本 昌彦, 16, 57
土田 修平, 16
寺田 努, 16, 57
中島 一崇, 9
中野 倫靖, 1
中村 俊勝, 128
盧 承鐸, 9
濱崎 雅弘, 1
樋口 啓太, 113
尾頭 花奈, 71
深山 覚, 1
福里 司, 30
松井 菜摘, 57
松本 高, 64
三ツ井 慧太郎, 91
宮下 芳明, 121
森島 繁生, 97
山本 和彦, 23
楊 溪, 9
吉田 博則, 78
ラルスン マリア, 78
李 佳憲, 85

日本ソフトウェア科学会 研究会資料シリーズ ISSN 1341-870X
No. 94 2021年12月8日発行

第29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 2021) 論文集