



The background of the poster is a textured, dark gray surface. It features several white, six-pointed snowflake shapes of different sizes scattered across the upper half. In the center, there is a faint, stylized illustration of a mountain range with dark peaks and some red highlights. The overall aesthetic is cold and wintery.

WISS 2023

第31回インタラクティブシステムとソフトウェアに関するワークショップ (WISS)



WISS2023 開催にあたって

志築 文太郎 *

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、2023年11月29日（水）から12月1日（金）にかけて、ロイヤルホテル八ヶ岳（山梨県北杜市大泉町）で開催される「第31回インタラクティブシステムとソフトウェアに関するワークショップ」（Workshop on Interactive Systems and Software: WISS2023）での発表をまとめたものです。

WISSは、伝統的に2泊3日の泊まり込み形式での開催を特徴のひとつとしてきており、毎年約100名～150名の参加者が朝から深夜まで深い議論を行ってきました。ただ、Covid-19のために、WISS2020は開催形態を完全オンライン、WISS2021とWISS2022はハイブリッドとしてきた中で、今回のWISS2023は完全現地開催とすることにしてみました。

今回のWISS2023では発表形態として、従来からの登壇発表（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）に加えて、国際学会・国際論文誌採択論文発表（Non-archival）を用意しました。WISS2022までは、国際学会・国際ジャーナル論文誌等で発表済みの内容でも参加者にとって有益であると思われる場合には登壇発表として採択していました。これに対して、WISS2023では、未発表論文の発表機会を更に増やしつつ、難関の国際会議・国際論文誌に採択された論文の発表機会とノウハウを広めることを狙って本カテゴリを設けることにしました。

登壇発表については、それぞれの論文に対して4名の査読者（うちメタ査読者1名）を割り当てて査読を行った結果、27件の投稿に対して15件の論文を採択しました（ロング採択2件、ショート採択11件、議論枠採択2件）。なお、幾つかの論文については条件付き採録としました。これらについてはいずれも提出いただいた論文を本論文集に掲載しております。

国際学会・国際論文誌採択論文発表については、7件の申し込みがあり、7件のいずれにも利害関係の無い3名の審査員による審査を経て、全件を採択しました。また、ロングティザーフォーマットについては5件を、デモ・ポスター発表については78件を採択しました。

なお、デモ・ポスター発表への申し込みが盛況である一方で、課題も改めて浮き彫りになりました。まず、デモ・ポスター発表については、昨年と同様に、締切前に枠が埋まりました。また、参加者の登録についても、昨年と同様に、早期登録締切前に枠が埋まりました。多数のデモ・ポスター発表の需要があり、また、多数の参加希望者がいることは、WISSに関わる人の数と力とが強まっている現れなので、将来へ向けてとても明るい材料です。今後は、この明るい材料を活かすよう、デモ・ポスター発表の募集方法、参加登録の方法に加えて、WISSそのもののあり方（発表形態、会場、日程、プログラム編成等）も議論しながら改良していくことにより、WISSやヒューマンコンピュータインタラクション分野が更に伸びていくことと思われます。

さて、WISSは今回で31回目を迎えます。前回（第30回）のオープニングにおいて会場に尋ねたところ、初回（WISS'93）に参加されていた方もいらっしゃいました。WISS2023で初回から30年。今回、様々な感染症が流行している中、ワクチン接種が出来ない方もいらっしゃる中、また感染症への防疫システムが充分でない中、学会をどのように開催したら良いのか甚だ不確かな状態で、完全現地開催という開催形態を採用しました。このあたりについては、参加できなかった方々に評価を求める必要があると感じています。ただ、今回のWISSが、参加者にとって「…また来るぞ…」と思える場となっていたら、学生時代から参加して毎回楽しいと思ってきた自分としても（初参加はWISS'96），そこにプログラム委員／運営委員の皆様と関われたことはとても幸せなことです。

最後に、登壇発表論文を投稿いただいた皆様、査読なし発表に申込みいただいた皆様、スポンサー企業の皆様、協賛学会関係者の皆様、今回も会計業務をお手伝いしてくださっている林康恵様と賀集優里様、そして、実行委員長の中野さんを始めとするWISS2023プログラム委員／運営委員の皆様に深く感謝いたします。

WISSは続くよどこまでも

* 筑波大学、WISS2023 プログラム委員長

WISS 2023 プログラム委員長
志築 文太郎 (筑波大)

WISS 2023 実行委員長
中野 倫靖 (産総研)

WISS 2023 副実行委員長
渡邊 恵太 (明治大)

WISS 2023 運営委員

会計	丸山 一貴 (明星大) 松村 耕平 (立命館大)
出版・ デザインコンペ	池松 香 (LINE ヤフー, 東北大) 加藤 邦拓 (東京工科大) 石井 綾郁 (NTT) 越後 宏紀 (ソフトバンク)
広報・ 学会リエゾン	横窪 安奈 (東大) 門本 淳一郎 (東大)
WISS チャレンジ	高田 峻介 (神戸高専) 藤田 和之 (東北大)
動画生中継	土田 修平 (お茶大) 塚田 浩二 (はこだて未来大) 高橋 治輝 (立命館大) 渡邊 拓貴 (北海道大)
デモ・ポスター	奥本 隼 (TwoGate) 米田 英正 (Helpfeel)
ネットワーク	福里 司 (早大)
Web・ サーバー	山本 和彦 (ヤマハ)
投票	鳴海 紘也 (東大)
懇親会	倉本 到 (福知山公立大) 五十嵐 悠紀 (お茶大)
表彰	栗原 一貴 (津田塾大, クーリード) 三上 浩司 (東京工科大)
企業 スポンサー	秋山 耀 (チームラボ) 清木 昌 (ほぼ日)
学生 ボランティア	宮藤 詩緒 (東工大) 大西 鮎美 (神戸大)
表紙デザイン	村松 明幸 (北陸先端大)

WISS 2023 プログラム委員

秋山 耀 (チームラボ)
五十嵐 悠紀 (お茶大)
池松 香 (LINE ヤフー, 東北大)
石井 綾郁 (NTT)
越後 宏紀 (ソフトバンク)
大西 鮎美 (神戸大)
奥本 隼 (TwoGate)
梶本 裕之 (電通大)
加藤 邦拓 (東京工科大)
門本 淳一郎 (東大)
倉本 到 (福知山公立大)
栗原 一貴 (津田塾大, クーリード)
小池 英樹 (東工大)
坂本 大介 (北海道大)
志築 文太郎 (筑波大)
清木 昌 (ほぼ日)
高田 峻介 (神戸高専)
高橋 治輝 (立命館大)
塚田 浩二 (はこだて未来大)
土田 修平 (お茶大)
寺田 努 (神戸大)
中野 倫靖 (産総研)
鳴海 紘也 (東大)
福里 司 (早大)
藤田 和之 (東北大)
松村 耕平 (立命館大)
丸山 一貴 (明星大)
三上 浩司 (東京工科大)
宮藤 詩緒 (東工大)
山本 和彦 (ヤマハ)
横窪 安奈 (東大)
米田 英正 (Helpfeel)
渡邊 恵太 (明治大)
渡邊 拓貴 (北海道大)

ISS 研究会連絡先

日本ソフトウェア科学会
「インターラクティブシステムとソフトウェア研究会」
主査: 志築 文太郎
筑波大学 システム情報系

〒305-8573 茨城県つくば市天王台 1-1-1
E-mail: shizuki@cs.tsukuba.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: WISS2023 開催にあたって
志築 文太郎 i

セッション1：身体

VR 百人一首かるたにおける札取り動作の数理モデル解析とそのゲームスキル調整への応用 ショート採録
ショート発表
丸山 礼華, 栗原 一貴, 山中 祥太 1

VR 環境での百人一首を題材に、プレイヤー間の難易度を平滑化することを目指し、個々のプレイヤーに与えられている視覚情報（手の大きさ、札の大きさ）を変動させることによってゲームの難易度を調整する手法を提案し、その効果を実証的に検証しています。
当該手法による難易度の変化を丁寧に評価し、かつユーザが不満や不適切に思わない範囲でも十分な難易度調整が可能であることを示しており、有用性の極めて高い知見を提供していることを踏まえると、採択にふさわしい論文であると思います。

VR リダイレクションを用いたゴルフスイングトレーニングシステム ショート採録 ショート発表
余 志皓, 廖 振傑, 小池 英樹 10

本研究では、VR リダイレクション技術を活用したゴルフスイング矯正システムを提案しています。VR 内の仮想のゴルフボールが打者のスイングに応じて前方へ移動し、打者はこれに合わせてスイングを調整します。結果として頭の上へと振り上げる over-the-top スイングが改善し、クラブが正確にボールに当たるように矯正されていきます。このアプローチはユーザ実験で有用性が確認され、論文中では現システムの課題と展望も示されていました。論文の記述に改善の余地はありますが、採録（ショート）で問題ないと判定しました。

拡張現実を用いた姿勢可視化によるスキートレーニングシステム ロング採録 ロング発表
松本 高, Erwin Wu, Chen-Chieh Liao, 小池 英樹 17

本稿は、拡張現実を用いたスキートレーニングにおける姿勢を可視化する手法とその効果についての論文です。手法に強烈な目新しさはありませんが、キーという対象に合わせて効果を有することが期待される表現をシンプルにまとめ、納得度合いの高い評価を重ねて知見を得ていることから、採択に足る論文であると判断しました。

セッション2：学習

Smart Replay:e ラーニング動画における視覚的・時間的アクセシビリティの向上 ショート採録（採録条件あり）
ショート発表
SECHAYK Yotam, SHAMIR Ariel, 五十嵐 健夫 25

e ラーニングのようなビデオコンテンツによる学習時の資料とチャプターの概要、自らのペースに合わせてビデオの再生をコントロールできる等の機能が搭載されたシステムを提案しており、システムの技術的な内容は一定の評価ができます。事前調査によってシステムの要件の洗い出しをしていることも好感を持てます。一方で、提出された論文については出版前に十分な推敲が必要であると判断し、ショート採録（採録条件あり）と判定します。

議論枠発表

岡 大貴, 大西 鮎美, 西田 健志, 寺田 努, 塚本 昌彦 34

具体的なプログラミング言語でのコーディングよりも前に、自然言語で解決方針・手順を立案することが重要であり、少ない選択肢から選び、誤りに気づいたらすぐ手戻りするという試行錯誤を行う、というアイデアは興味深く、新規性も認められます。

しかしながら、試行錯誤の高速化がどのように達成されているのか、学習者がどのような気づきを得て選択肢を修正するのかが不明瞭なことと、最初に自然言語で検討することをかえって難しく感じてしまう初学者もいるのではないかといったことが査読者間および判定会議で議論となりました。

結果として、本論文は議論枠での採録となりました。発表に当たっては、学習者がどのように選択肢を操作し、どのような表示を見て誤りに気づいて修正するのか、具体的な操作が分かるようにご留意ください。

AR ハンドと EMS による手指の同期を用いた遠隔でのピアノ演奏指導

ショート採録（採録条件あり）

ショート発表

守 新太郎, 井原 圭一, 川口 一画 41

ピアノ学習者に対する遠隔指導手法として一定の新規性がある点、提案手法や実装を中心に論文の記述が明瞭である点が評価されました。一方、「EMS の性能や設計意図の記述不足」「タイミング制御手法の記述不足」「一部参考文献の不足」等が指摘されたため、ショート採録（条件付き）と判定します。

セッション3：表現

EmoCast: 顔への表情変化情報投影による話し相手の印象向上システム

議論枠採録

議論枠発表

浅野 恭志, 大西 鮎美, 寺田 努, 塚本 昌彦 48

頭部に装着したウェアラブルデバイスにより表情の画像認識し、プロジェクトを行なうことで表情の改変を行う技術について提案、検討したものである。

有用性の評価手法に不明瞭な点があるものの、提案の新規性は認められる。

その「荒削り」さにより、「根本的な課題解決になっているのか」、「装備や技術の進歩によって将来的には表現力が高まり不自然さが軽減されるのか」等、活発な議論が想定される。

以上の結果から、議論枠採録と判断された。

Taste Time Machine : 飲食物を過去や未来の味に変える装置の実現に向けて

ショート採録

ショート発表

宮下 芳明 55

著者らの味ディスプレイの応用として、食品の時間を制御した味にする（熟成させる、味を若く戻す）といった目的に対しての手法提案と実験を行っています。論文の記述や読みやすさには少し難がありますが、味の時間変化を制御するという事実自体が明らかに面白いため、査読者も全員ポジティブな評価をしていました。一方で、これによってどのように世界が変わるのが、どのようなインタラクティブシステムに応用可能なのか、という点を是非深掘りして議論していただきたいというのがレビューの求めるところです。

シャボン膜凍結式全周囲スクリーンの提案	ショート採録（シェファーディングあり）	ショート発表
橋本 侑樹, 西村 晶太郎, 寺澤 真一郎, 大塚 真柊, 佐藤 俊樹.....		62

Ephemeral インタラクションのデバイス研究として十分な新規性を有すると思われます。凍結シャボン膜ならではのインタラクション提案としての触れることによって開口が可能である点などを挙げられております。このあたりはもうすこし議論したいと思いました。

導電性・強磁性・及び土台の可撓性を持つ毛構造の造形手法の提案	ショート採録（採録条件あり）	ショート発表
鎌田 航誠, 高橋 治輝, 塚田 浩二.....		69

本研究では、熱溶解積層方式により毛構造を3D プリントする先行研究を発展させ、通常の PLA とは異なる機能性を持つフィラメントによる毛の造形、従来とは異なる柔軟な土台への定着方法の提案、毛構造の形状による触感の違いのユーザ評価が行われました。議論されている内容は、複数の先行事例と区別され新規性があります。一方で、現段階では様々な特性が実現できることが示されたという状態です。そのような特性がインタラクションまたは他の文脈でどれほど有用であるかはこれから明らかにされていくものと感じます。

セッション5：誘い

MusicCommentVisualizer: 音楽動画へのコメントの可視化に基づく音楽推薦・鑑賞インターフェース	ショート採録
ショート発表	
倉持 友哉, 濱崎 雅弘, 中野 倫靖.....	76

本研究では音楽動画へのコメントを利用した音楽推薦・鑑賞インターフェースを提案されています。ソーシャルコメントを元に感情と状況の2つの側面で分類しユーザーの状態に合わせて音楽を推薦するシステムは一定の新規性が認められます。有用性が確認できない点やアルゴリズムの詳細説明が不足している点など懸念はあるものの、提案システムのクオリティは高く、不採録とするほどの際立った問題はないと判断し、ショート採録と判定されました。

SwipeGANSpace: 潜在空間の関心次元探索によるスワイプ操作に基づく嗜好画像生成	ロング採録
ロング発表	
中島 柚斗, 楊 明哲, 馬場 雪乃.....	84

GAN による嗜好画像生成におけるユーザの嗜好を、スワイプ操作を用いて効率的に行う手法を提案し、その手法に沿った生成アルゴリズムとして多腕バンディットアルゴリズムを適用することで、シンプルなペイズ最適化手法よりもよりユーザの嗜好性を反映した画像を生成できることを示している論文です。

動いている人の視線だけを惹きつけるモザイク投影手法	ショート採録	ショート発表
幸谷 有紗, 覆山 英高, 浦垣 啓志郎, 宮藤 詩緒, 小池 英樹.....		92

高速のプロジェクタでモザイク状の模様を高速に変化させて投影することで、ユーザ側に特殊な装置を必要とせず、動いている人にだけ注目を集めるような情報提示手法を提案している。有用性の立証が不十分であるものの、新しい手法の提案としては WISS で議論する意義があるという結論になった。以上の結果から、ショート採録（条件なし）と判断された。

セッション6：入力手法

LensTouch: スマートグラスのレンズ面を使った入力手法 ショート採録 ショート発表

川崎 竜也, 真鍋 宏幸 99

本研究の新規性は、スマートグラスのレンズ面にタッチパッドを付けることでタッチ入力を実現した点にあると考えられます。既存研究としても、VR HMD では Gugenheimer らの FaceTouch がありますが、See-through HMD あまり見られません。また一般的にはレンズへのタッチは指がぼやけるため入力が不正確となると思われますが、本論文では必ずしもそうではなく比較的精度が高いという結果が出ている点は評価されます。一方、有用性に関して、本手法そのものの精度は述べられていますが、他の方式、例えばバーチャルキーボードとの入力精度・速度の比較が必要と思われます。

スマートウォッチの加速度センサを用いた Thumb-to-finger ジェスチャ認識に基づく片手文字入力手法

ショート採録 ショート発表

角田 陸, 志築 文太郎 106

不明瞭な点はあるものの、ジェスチャのデザイン、予備評価による適切なジェスチャセットの選定、評価実験と丁寧に書かれており、スマートウォッチ内蔵センサのみを用いたハンドジェスチャの提案として充分採録に値するとすべての査読者に判断されました。

VR 百人一首かるたにおける札取り動作の数理モデル解析とそのゲームスキル調整への応用

丸山 礼華* 栗原 一貴* 山中 祥太†

概要. 本論文では百人一首かるたにおける新しいゲームスキル調整方法について検討する。VR 空間内で実装された百人一首かるたゲームにおける、札を取るプロセスの一部がいわゆるポインティングタスクとみなせることから、被験者実験によって得られたデータについて Fitts の法則を用いた数理モデル解析を行った。その結果、札までの距離、手の大きさ、札の大きさを VR 空間内で変更することが札を取る所要時間に与える影響を精度良く予測することができた。また、その際の各種パラメータ変更に対してプレイヤが適切だと感じられる範囲を抽出した。これらの結果から、VR 空間内で札までの距離、手の大きさ、札の大きさを適切な範囲で変更することにより、札を取る所要時間を制御し対戦相手との拮抗に近づける手法の有効性が示唆された。またこのような変更がプレイヤ自身や対戦相手の知覚に与える影響について考察した。

1 はじめに

インクルーシブな対人ゲームの実現において重要な観点の一つは、プレイヤ間のスキル差の適切な調整である。対人ゲームにおいて、あるプレイヤが他のプレイヤと異なるスキルレベルである場合に、パフォーマンスや勝敗判定が拮抗するようにゲームに調整を施すことは一般的である。調整を行うことで、異なるスキルレベルのプレイヤ同士であっても楽しく互角の勝負をしやすくなり、より多様なプレイヤの参入とモチベーション維持につながることが期待できる。

しかしそのような調整の結果、プレイヤが実力を十分に発揮することが難しくなったり、満足感が損なわれたりすることは避けられるべきである。たとえばゲームの調整の主な方法として、ゲームルールの変更、有利な側のプレイヤへの行動の制限や不平等なスコア評価の実施などが挙げられるが、ルールを大幅に変更する調整は、プレイヤたちに普段とは異なる身体の使い方や戦略を強いるため、やりづらさを感じさせてしまう可能性がある。また、より強いプレイヤの実力やその結果としての正当なスコア評価をあからさまに制限してしまうと、プレイヤたちはプレイ中に制限の存在をことさらに意識してしまいかねない。これは申し訳なさや遠慮といったネガティブな感情につながる可能性がある。これでは、互角の勝負ができるとしても、良い調整であるとは言えないだろう。

そこで、プレイヤが身に付けているスキルの運用

にあまり影響を与えず、理想的には知覚することすら困難な形でゲームスキルを調整することを検討する。我々はこれまでに、対人ゲームである百人一首かるたを対象とし、「VR を用いた互いに知覚困難なスキル調整」を提案している[1]。これはプレイヤが互いに知覚しにくい、VR 内でのスキル調整である。図 1 はその様子を表した図である。支援の相手の環境に与える影響が小さくなるよう注意深くデザインすることで、支援が相手のスキル運用に与える影響を小さくし、プレイヤにあまり意識させることなくスキル調整を行うことを意図する。

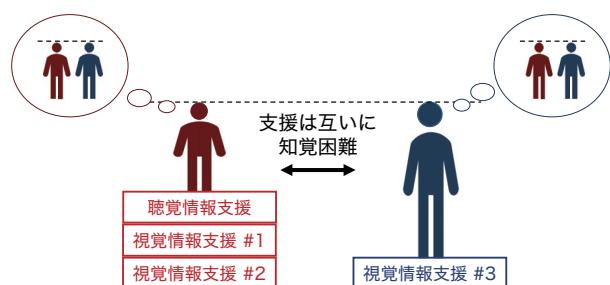


図 1. VR におけるスキル調整。異なるスキル支援を受けることで、プレイヤは対戦相手も自分と同程度の実力であると錯覚できる。

本論文ではこのような百人一首かるたにおける VR を活用したゲームスキル調整方法の新たな可能性について検討する。VR 百人一首かるたにおける札を取るプロセスは、いくつかの要素に分解できる。(1)和歌が読まれ始めてから、記憶をたよりに取るべき札を想起するプロセス、(2)配置されている複数の札から取るべき札を検索するプロセス、そして(3)その札に手を動かして触れるプロセスである。このう

Copyright is held by the authors.

* 津田塾大学, † ヤフー株式会社

ち最後の(3)のプロセスは、ヒューマンコンピュータインターラクション研究分野においてしばしば論じられる、入力インターフェースによるオブジェクトへのポインティングタスクとしてモデル化できるため、Fitts の法則などを応用したタスク完了時間の予測手法の適用が見込める。実験を通じた数理モデル解析により、VR 空間内での札のサイズ、手の大きさ、札までの距離を適切な範囲内で調整することにより、札を取る動作に必要な時間を制御する手法について検討する。またこのような変更がプレイヤ自身や対戦相手の知覚に与える影響について考察する。

本論文の構成は次の通りである。2 章で関連研究について述べ、3 章では実装した VR 百人一首かるたシステムについて説明する。次に 4 章で実験と解析について述べ、5 章で実験結果の意義、および今後の課題と展望について議論し、6 章でまとめを行う。

2 関連研究

2.1 インクルーシブなゲーム

対人ゲームをよりインクルーシブにするために様々な方法が検討されている。1 つはゲームのルールを変える、あるいは作る方法である。世界ゆるスポーツ協会が提案する「ゆるスポーツ」[2]は、様々なスポーツを年齢、性別、運動神経、障がいの有無に関わらず楽しめるようにしたものである。また、視覚情報を付加することで、スキルを補う方法もある。例えば、「きまり字かるた」[3]は、初心者にも使いやすい、各札に物理的に決まり字が印刷された百人一首の札である。またスポーツの分野でも、情報技術を用いてボールの軌道を予測し、可視化することで、初心者を支援する例がある[4][5][6]。藤原ら[7]は、オンラインの対人ゲームにおける、色覚多様性者のプレイヤの不利を軽減するための色の利用について論じた。これらの方法はより多様な人々がゲームを楽しめるようになることを目標とする点で我々の研究と共通しているが、支援がプレイヤ全員に影響を及ぼすという点で我々の研究と異なる。本研究では、これらの支援も取り入れながら、VR を用いてプレイヤのスキル運用に影響を与えていく支援を検討する。

プレイヤのスキル運用に影響を与えていくスキル調整の例として、前川らの綱引きのシステム[8][9]がある。これはシステムから外力を介入させることでスキル調整を行うシステムである。本研究では VR を用いることで、さらに身体や物理法則の変更といった現実ではできない支援を行えるようにする。また、プレイヤ間の支援の違いをより知覚しにくくなり、プレイヤの意識がスキル調整に向きにくくなる

ことが期待される。

築瀬らの研究[10][11]は、一人用のゲームにおける知覚困難な難易度調整の可能性を探究したものである。本研究では類似のコンセプトを対人ゲームのひとつである百人一首かるたに応用する。百人一首かるたのスキル調整においては、各プレイヤーの札を取る速度を定量化し、それらが拮抗するように調整する必要があるので、本研究で探求するような数理モデルが特に有用である。一方で、他の対人ゲームや一人用のゲームにおいてもそのような精緻な数理モデルが得られればスキル調整に活用することは可能かもしれない。

2.2 MR を活用したトレーニングと人間拡張

MR を用いることで、視聴覚情報を付加する支援が容易になる。例えば、体の使い方のイメージを視覚的に表現すること[12][13]や、プレイヤの動きに対してフィードバックを行うこと[14]、環境の中の必要な情報を強調すること[15]、対戦相手の身体の動きを予測し、可視化すること[16]ができる。さらに VR では、現実世界では変更することができない設定も変更することができるため、技術を身に付けるためのトレーニングに利用されることも多い。例えば、重力などの力を操作したり[17][18][19][20]、時間を歪曲させたり[21][22]、速度認識を誤認させたり[23]することができる。これらは、スキルを身につけるためのトレーニングを行うことを目的としており、我々の研究とは目的が異なる。しかし、これらの研究で用いられている技術は、今後、プレイヤ間のスキル差を一時的に軽減するためにも用いることができると言える。

また VR を用いた人間拡張の例として、小川らの指の長さを変える研究[24]や、SUN らの腕の長さを変える研究[25]がある。これらも、研究の目的は異なるが、プレイヤが違和感を感じない範囲内でこれらの技術を活用することは、一時的なスキル向上にも役立つと考える。

2.3 身体の動きを外部から制御する技術

身体の動きを外部から制御する技術は、対人ゲームにおけるスキル調整にも応用できる可能性がある。なぜなら、プレイヤはトレーニングすることなく、一時的にゲームに必要な能力を発揮することが期待できるためである。EMS を用いて身体をコントロールする研究は複数存在する[26][27][28]。蛭子らの研究[29][30]もその一つである。これは、初心者でもすぐにドラムを演奏できるようにするものである。これらは学習を目標としている点で我々の研究とは異なるが、これらのシステムで利用されている仕組みを用いることで、初心者でもすぐに理想的な動作を行えるようにすることは、我々の研究においても

将来の展望として有望である。

2.4 情報技術による百人一首かるたの支援

情報技術による百人一首かるたの支援には、いくつかの方法が提案されている。北川ら[31]は競技かるたにおける払いの動作を、PoseNetを用いた姿勢推定によって可視化した。山田ら[32]は、手首に加速度、角速度センサを装着することで、現実の競技かるたの試合中の札取得タイミングの情報を得た。都丸ら[33]のように、決まり字を可視化することで、試合前の戦略立てを支援する研究もある。また百人一首の読み上げの音声を解析することで、それぞれの和歌の決まり字までの音響的な違いを示し、和歌の判別をより早く行える可能性を検討した研究[34][35]もある。これらは、かるたの選手が現実世界のかかるたのスキルを向上させることを支援するものであり、経験者であるか初心者であるかを問わず、VRを用いたゲーム中にリアルタイムで支援を行う本研究とは目的が異なる。しかし、これらの研究の結果や示唆は、VRを用いたスキル調整としての支援にも役に立つ可能性があり、これから研究の方向性の1つとして考えられる。徳島ら[36]はパソコン上で動作する百人一首かるたの学習システムを提案した。このシステムでは、ゲームが進行するにつれて徐々に変化する決まり字の動的特性を学習することができる。徳島らのシステムは1人用の学習用システムであるのに対し、本研究は可視化を利用したVRゲームのスキル調整である点が異なる。

2.5 ポイントティングタスク

PCやスマートフォンを利用するうえで、オブジェクトの選択は最も頻繁に行われる操作であり、所要時間の推定モデルや操作を支援する手法が多く提案されている。よく知られた時間推定モデルには以下のFittsの法則がある。

$$MT = a + b \cdot ID, \quad ID = \log_2 \left(1 + \frac{D}{W} \right)$$

MT(Movement Time)は所要時間、*a*と*b*は実験的に決まる回帰定数、*ID*(Index of Difficulty)はタスクの難度指標、*D*と*W*はオブジェクトまでの距離とサイズである。このモデルはVR空間内でコントローラ操作でポインティングをする場合にも適合することが知られている[37]。

Fittsの法則によれば、オブジェクトまでの距離*D*を短縮するか、サイズ*W*を大きくすることで難度が下がり、より短時間で操作できる。これに基づいて、PC画面上のターゲット選択時間を短縮する手法が多く提案されている。例えばオブジェクトをカーソル付近まで自動的に近づける手法[38]やカーソルの近くにあるオブジェクトを拡大する手法[39]がある。またカーソルがオブジェクトにスナップする手法

[40]やカーソルがオブジェクトに吸い寄せられるように重力をかける手法[41]もある。上記のいずれも、VR内における百人一首かるたの難易度調整に活用しうるが、例えば手を自動的に移動させるスナップや重力を使う方法では、プレイヤー視点では支援されていることが明確になる懸念がある。またVR空間において遠くのオブジェクトに手を伸ばすのを支援する手法として代表的なのがGo-go手法[42]であり、手が自然に届く範囲を超える位置にあるオブジェクトを操作したい場合に、二次関数的に手が伸びることでタスク難度を低減している。本研究で我々が採用するのは*W*を拡大することで難度を調整する方法である。先行研究では、オブジェクトのサイズを大きくする方法[39]と、カーソルのサイズを拡大する方法[43]のいずれも所要時間低減になると報告されている。VR内における百人一首かるたでは、それぞれ札と手の大きさを変更することに該当する。

3 システムの実装

我々は、これまでにUnityを用いて、2人用のVR百人一首かるたシステムを作成している[1]。VRを用いた互いに知覚困難なスキル調整を行うためのいくつかの支援を導入することで、百人一首を覚えていない、また札を取る動作に慣れていないプレイヤーから、百人一首を覚えており、日常的にかるたをプレイしているプレイヤーまで、一緒に楽しめるゲームとすることを意図した。

実装では、Meta Quest2を使用した。VR空間では、Meta Quest2のコントローラがある位置に手のオブジェクトが表示され、キャリブレーションにより、実際の手の位置と手のオブジェクトの位置が同じになるように調整することができる。図2は実際のゲーム画面である。

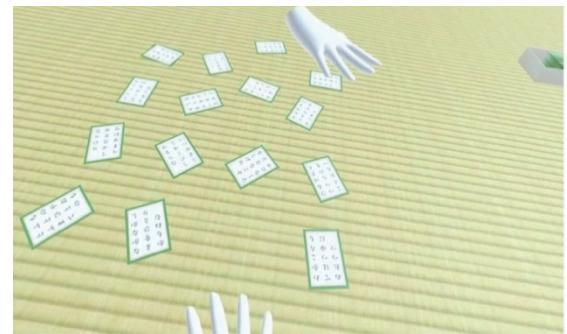


図2 プレイ画面

4 実験

4.1 概要

VR百人一首かるたにおける札を取るプロセスの一部である、札に手を動かして触れるプロセスは、

Fitts の法則で分析可能な、オブジェクトへのポインティングタスクとみなすことができる。本章では VR 百人一首かるたシステム内で条件を変えながら札を取るのにかかった時間を測定することで、札を取るまでの時間に与える影響を検討し、以下の 2 つの仮説について検証する。(1)手で札を取る動作にかかる時間は、Fitts の法則のモデルに高い適合度を示す。すなわちターゲットである札の大きさや手の大きさ、札までの距離を VR 空間内で変えることが、札を取る早さに与える影響は予測可能である。(2)VR 内において、実物とは異なる大きさの手や札のオブジェクトに対しても、プレイヤはかるたゲームを行う上で適切であると感じる。

4.2.1. 概要

実験では、VR 空間ににおける手の大きさと札の大きさをそれぞれ変えながら、札を取るまでの時間とプレイヤの適切感、すなわち手や札のサイズに関してプレイヤの感じた適切さの度合いを検証する。今回は、それぞれ 5 通りの大きさの手と札を用いた。これらはそれぞれのオブジェクトにおいて 1.0 倍と規定する基準の大きさを定め、そのオブジェクトを拡大縮小したものである。具体的な拡大縮小の倍率は、0.8 倍、1.0 倍、1.3 倍、1.5 倍、1.75 倍の 5 通りである。図 3 は手の倍率を変えた様子である。このようにオブジェクトの倍率を変更することで、手の当たり判定の大きさも拡大縮小する。また、図 4 は札の倍率を変えた様子である。このようにオブジェクトの倍率を変更することで、札の当たり判定の大きさも拡大縮小する。

ここでは手の大きさについて 1.0 倍と規定する基準を縦 20cm、横 10cm、同様に札の大きさについて、1.0 倍と規定する基準値を横 5.2cm、縦 7.3cm、に設定する。手の大きさの基準値は成人の平均的な手の大きさ[44]、札の大きさの基準値は、全日本かるた協会が定める競技かるたの札の大きさに基づいて定めた。



図 3 手の倍率を変えた様子



図 4 札の倍率を変えた様子

実験協力者として、かるた経験の多様な 28 名を集めた。実験では手と札の倍率を変えながら、1 枚の札が表示されてから、札を取るまでの時間を計測するタスクを複数回繰り返した。また、主観評価としてその際の手の大きさ・札の大きさがかるたゲームを行う上で適切であると感じたかについての 7 段階評価とその理由を述べる自由記述のアンケートを行った。図 5 は実験の様子と実験の画面である。



図 5 左:実験の様子、右:実験の画面

4.2.2. 手順

まず、実験協力者にタスクの流れおよび操作方法を説明した。タスクは全部で 9 セッションあり、1 セッションごとに手の大きさと札の大きさが変わる。実験協力者ごとにその提示順はランダムに変化させた。実験協力者は 1 セッションが終わるごとに手の札の大きさについての主観評価を行う。それぞれのセッションで実験協力者は、ランダムな場所に表示された 1 枚の札に触る動作を 10 回繰り返す。実験協力者には、札が表示されたらすぐに VR 内の手で札に触ること、無駄な手の動きをしないことを依頼した。なお、札の表示範囲は、無理なく札を取ることができる範囲として、全日本かるた協会の定める競技線の範囲である横 87cm、縦 50.8cm の長方形内を採用した。実験は、実験協力者が Meta Quest2 のヘッドマウントディスプレイを装着したのち、実験協力者自身が開始ボタンを押して、開始した。5 セッションが終わった時点で任意の時間の休憩を挟んだ。

4.2.3. 結果と分析

2.5 節で述べた Fitts の法則について、手の中心の初期位置から札の中心までの距離を D 、札の横幅 W_{card} と手の横幅 W_{hand} の合計値を W [43][45] として回帰直線を求める。初期状態では手は床についており、札は表示していない。床に札を表示した瞬間に MT を計測開始する。ここでは手は縦長の楕円に近似して考え、その縦横比は 2:1 とした。

まず前処理として、実験で得られたデータのうち、四分位数の 1.5 倍以上外れたデータを外れ値として除去した[46]。[47]にならい、IDについて 0.2 ずつ区間を取り、各区間における ID と MT の平均値をその

VR 百人一首かるたにおける札取り動作の数理モデル解析とそのゲームスキル調整への応用

区間の代表値としてプロットした。このとき 1 区間あたりのサンプルサイズが 9 以下のものは除外した。これらのデータから、最小二乗法で a および b を決定し、相関係数 R^2 を求めた。図 6 はそのグラフである。 $R^2 = 0.9896$ となり、Fitts の法則のモデルに高い適合度を示している。

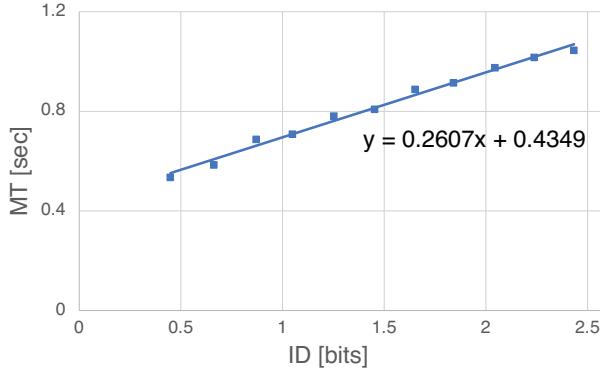


図 6 手と札の大きさを変えながら札を取った時の ID と MT

表 1、表 2 は手および札の倍率を 0.8 倍、1.0 倍、1.3 倍、1.5 倍、1.75 倍にした時の、手と札それぞれに対して実験協力者が適切であると感じた度合いの評価の 7 段階評価リッカート尺度（1:適切でない、7:適切である）の平均値を示した表である。いずれも極端に低い数値ではなく、このような手や札の大きさの変更はある程度適切だと評価されることが見て取れる。さらに、より違和感のない保守的な変更の範囲を明らかにするため、実物大（1.0 倍）を基準として、同等あるいはそれ以上に高評価を得たもののみを抽出することを以下のように試みた。手の評価を示した表 1 では手の実物大である 1.0 倍への評価を基準とし、それよりも評価が良好であるものに色をついた。札の評価を示した表 2 では札の実物大である 1.0 倍への評価を基準とし、それよりも評価が良好であるものに色をついた。表 3 は表 1 において手の評価が 1.0 倍よりも良好であり、かつ表 2 において札の評価が 1.0 倍よりも良好であるものを示した表である。

表 1 手と札の大きさがそれぞれ 0.8 倍、1.0 倍、1.3 倍、1.5 倍、1.75 倍のときの手の評価。実物大である 1.0 倍よりも高評価のセルには色をついた。

手 \ 札	0.8倍	1.0倍	1.3倍	1.5倍	1.75倍
0.8倍	4.6	5.0	5.9	4.8	4.8
1.0倍	4.3	5.1	5.6	6.0	5.8
1.3倍	3.5	5.2	5.9	5.7	5.5
1.5倍	4.7	4.0	5.5	6.0	5.3
1.75倍	3.6	6.2	5.3	5.8	5.2

表 2 手と札の大きさがそれぞれ 0.8 倍、1.0 倍、1.3 倍、1.5 倍、1.75 倍のときの札の評価。実物大である 1.0 倍よりも高評価のセルには色をついた。

手 \ 札	0.8倍	1.0倍	1.3倍	1.5倍	1.75倍
0.8倍	5.2	5.0	4.2	4.8	4.8
1.0倍	4.7	5.3	4.9	5.0	4.6
1.3倍	5.5	6.2	5.8	5.0	5.8
1.5倍	4.5	5.8	4.8	5.3	5.6
1.75倍	4.7	5.0	5.0	4.8	5.2

表 3 手の大きさも札の大きさも実物大よりも高評価であるセルに色をついた。

手 \ 札	0.8倍	1.0倍	1.3倍	1.5倍	1.75倍
0.8倍					
1.0倍		■			
1.3倍			■		■
1.5倍				■	
1.75倍					

4.2.4. 考察

実験より、手で札を取る動作にかかる時間は、Fitts の法則のモデルに高い適合度を示した。よって、手の大きさやターゲットである札の大きさ、札までの距離を VR 空間内で変えることによる、札を取る時間への影響は予測可能であることが示唆され、仮説(1)は支持された。手および札の倍率を変更した時の、手と札それぞれに対する実験協力者の適切感の評価より、表 1 および表 2 で示されたような多様な条件下で極端には適切感は低くならないという結果が得られた。特に表 3 に示した倍率の組み合わせについて手の大きさ、札の大きさとともに実物大の条件と同等あるいはそれ以上の適切感が得られており、保守的にはこの条件下において仮説(2)は支持されたといえる。

なお、適切感の理由を述べる自由記述アンケートでは、回答として自分の記憶する実物の大きさとの一致度、操作のしやすさ、札の文字の見やすさが挙げられた。これらは適切感を考える上で重要な観点であると考えられる。記憶する実物の大きさとの一致度を評価基準とする人の中には、そもそも実物の大きさを正しく思い出せていないケースも複数あった。これにはかるたを行う際の目と手の距離や、板口ら[48]の研究のように VR 空間内での物の大きさの捉え方が影響している可能性があり、更なる調査が必要である。さらに、今回手の大きさの基準値である 1.0 倍は成人の平均的な大きさを元に決めたが、それぞれの被験者の実際のサイズにすることで、よ

り適切な範囲を得ることができるだろう。

5 議論

5.1 実験結果のスキル調整への応用

以下に実験結果を踏まえ、VR 百人一首かるたにおけるスキル調整へと活用する展望を示す。検討に値するのは、札までの距離、札の大きさ、手の大きさを変更することによってMTを制御し、プレイヤ間における札に触れるまでにかかる時間差を小さくするスキル調整方法である。

まず、札までの距離を変更することでMTを調整できることから、札の散らばり具合、すなわち個々の札までの距離を制御することが考えられる。例えば、札の配置については対戦相手と同一なまで、プレイヤの座る座標を調整することで札までの距離を変化させれば、スキル調整が可能であろう。また、札のサイズと手のサイズについても、実験で得られた適切だと感じられる条件下で変更することでMTを調整できると考えられる。

本実験で得られた回帰式は次のようにになった。

$$MT = 0.4349 + 0.2607 \cdot \log_2 \left(1 + \frac{D}{W_{\text{card}} + W_{\text{hand}}} \right)$$

平均的なスキルを持ったプレイヤが、競技かるたの競技線を参考に定める自然に手を伸ばせる範囲の中で最も遠い $D = 66.9\text{cm}$ 離れた位置にある札を取るとき、通常 ($W_{\text{card}} = 5.2\text{cm}$, $W_{\text{hand}} = 10.0\text{cm}$) であれば $MT = 1.069$ 秒かかるが、札の大きさを 1.75 倍 ($W_{\text{card}} = 9.1\text{cm}$)、手の大きさを 1.5 倍 ($W_{\text{hand}} = 15.0\text{cm}$) にすれば $MT = 0.934$ 秒になると予測される。この結果から、我々は札の取得時間について 0 から 0.135 秒程度調整を可能であるとわかる。我々の別の調査では、様々な条件下で和歌が読まれてから札を取るまでにかかる時間は経験者の場合で、平均 0.909 秒、標準偏差 0.301 秒であった [1]。これを百人一首かるたにおける札を取るまでにかかる時間についての一種の概算値として捉えると、これらの平均値および標準偏差に対し、我々が平均的なスキルを持ったプレイヤに対し調整できる 0~0.135 秒という値は無視できない割合を占める。これは提案手法のスキル調整手法としての有効性を示唆するものである。

なお、実際の活用時は、競技の事前に対戦相手とのスキル差を計測し、どの程度だけMTを調整すればスキルが釣り合うかを算出してゲームに反映させることを想定している。その際、今回の実験で調査して適切だとみなした手と札の大きさの倍率条件だけでなく、より多様に倍率条件を設定しMTを調整できることでスキル調整上、より都合がよい。そのためにはより多様な倍率条件における適切感の追加調査が必

要である。これは今後の課題である。

5.2 スキル調整とプレイヤの知覚

前節で述べたスキル調整は、VR 空間において個々のプレイヤごとに別の環境として提供することができることから、プレイヤ間で相互に存在が知覚されにくい状況での運用が期待できる。また、プレイヤ自身にも適切だと感じられる範囲で、かつ目立たない範囲でのパラメータ変更を行えば、プレイヤ自身にも存在が知覚されにくいスキル調整が可能かもしれない [10][11]。あるスキル調整を行うとき、それによって調整されたプレイヤが相互に、さらにはプレイヤ自身にも知覚が難しいような状態が実現できるのであれば、それはスキル調整に対する気まずさや違和感の軽減につながるため、より望ましい状態といえる。このようなスキル調整の相互知覚困難性、および自己知覚困難性についての詳しい検証は今後の有望な研究課題である。

5.3 総合的な百人一首かるたのスキル調整へ向けて

百人一首かるたで札を取る 3 つのプロセス、すなわち(1)和歌が読まれ始めてから、記憶をたよりに取るべき札を想起するプロセス、(2)配置されている複数の札から取るべき札を検索するプロセス、そして(3)その札に手を動かして触れるプロセスのうち、本研究では(3)を検討した。(1)については先行研究 [1] で検討されている。(2)については今後の課題であるが、例えば VR を用いて、表示する札の枚数を一時的に少なく制限すれば、その中から正しい札を検索する作業は高速化できるだろう。このような取り組みを通じて、百人一首かるたにおける総合的なスキル調整方法として取りまとめていく予定である。

6 まとめ

本論文では百人一首かるたにおける新しいゲームスキル調整方法を検討した。協力者数 28 名の実験データの解析の結果、VR 百人一首かるたにおける手で札を取る動作は、Fitts の法則のモデルに $R^2 = 0.9896$ という高い適合度を示した。また、プレイヤが適切だと感じられるパラメータ変更の条件を得た。この範囲で、手の大きさ、札の大きさ及び札までの距離を変更することは、百人一首かるたの試合に十分影響を与える調整量となり、札を取る所要時間を制御し対戦相手との拮抗に近づける手法として活用できる可能性が示唆された。パラメータ変更がプレイヤ自身や対戦相手の知覚に与える影響についてさらに探求することで、より望ましいスキル調整を行えることが期待できる。

謝辞

本研究の一部は、情報処理推進機構（IPA）の2023年度未踏IT人材発掘・育成事業の支援を受けて行われた。また、ユニティ・テクノロジーズ・ジャパン株式会社の築瀬洋平氏には有益な示唆を頂いた。謹んで感謝の意を表する。

参考文献

- [1] A. Maruyama and K. Kurihara. Mutually Imperceptible Skill Adjustment in VR for Making Hyakunin Isshu Karuta Inclusive. *Proceedings of IFIP ICEC'23*, 2023, in printing.
- [2] 世界ゆるスポーツ協会, YURU SPORTS. <https://yurusports.com> (2023/8/30 確認)
- [3] 株式会社 大石天狗堂, きまり字かるた. <https://www.tengudo.jp/100poems/competition/2528.html> (2023/8/30 確認)
- [4] E. Wu, H. Koike. FuturePong: Real-time Table Tennis Trajectory Forecasting using Pose Prediction Network. *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (CHI EA '20)*, pp. 1-8, Association for Computing Machinery, 2020.
- [5] H. Ishii, C. Wisneski, J. Orbanes, B. Chun, J. Paradiso. PingPongPlus: design of an athletic-tangible interface for computer-supported cooperative play. *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. pp. 394–401, 1999.
- [6] K. Sato, Y. Sano, M. Otsuki, M. Oka, K. Kato. Augmented Recreational Volleyball Court: Supporting the Beginners’ Landing Position Prediction Skill by Providing Peripheral Visual Feedback. In: *Proceedings of the 10th Augmented Human International Conference 2019*, pp. 1–9, 2019.
- [7] 藤原優花, 中村聰史. 色覚特性を考慮したゲームの有利不利制御の Among Us を用いた検証, エンタテインメントコンピューティングシンポジウム 2022 論文集, Vol.2022, 120-128, 2022.
- [8] A. Maekawa, K. Kasahara, H. Saito, D. Uriu, G. Ganesh, M. Inami. The Tight Game: Implicit Force Intervention in Interpersonal Physical Interactions on Playing Tug of War. *SIGGRAPH '20 ACM SIGGRAPH 2020 Emerging Technologies, Article No.10*, pp. 1–2, 2020.
- [9] A. Maekawa, H. Saito, D. Uriu, K. Kasahara, M. Inami. Machine-Mediated Teaming: Mixture of Human and Machine in Physical Gaming Experience. *CHI Conference on Human Factors in Computing Systems*, pp. 1-11, 2022.
- [10] 築瀬洋平. 誰でも神プレイできるシューティングゲーム. WISS2013, 2013.
- [11] 築瀬洋平, 鳴海 拓志. 誰でも神プレイできるジャンプアクションゲーム. 日本バーチャルリアリティ学会論文誌 21巻3号, pp. 415-422, 2016.
- [12] 鈴木湧登, 坂本大介, 小野哲雄. Gino.Aiki: 合氣道の身体の使い方の習得を支援する MR ソフトウェア. WISS2022, 2022.
- [13] 坂名 純太, 土田 修平, 寺田 努, 塚本 昌彦. 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. WISS 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. pp.16–22, 2021.
- [14] K. Masai, T. Kajiyama, T. Muramatsu, M. Sugimoto, T. Kimura. Virtual Reality Sonification Training System Can Improve a Novice’s Forehand Return of Serve in Tennis. *2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 845-849, 2022.
- [15] E. Wu, M. Piekenbrock, T. Nakamura, H. Koike. SPinPong - Virtual Reality Table Tennis Skill Acquisition using Visual, Haptic and Temporal Cues. *IEEE Transactions on Visualization and Computer Graphics*, pp. 2566–2576, 2021.
- [16] E. Wu, H. Koike. FuturePose - Mixed Reality Martial Arts Training Using Real-Time 3D Human Pose Forecasting With an RGB Camera. *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1384-1392, 2019.
- [17] H. Kawasaki et al. CanKendama: A System for Supporting Users to Master Kendama Tricks through a Few Minutes of VR Training. *Proceedings of IPSJ Entertainment Computing*, pp. 26–32, 2020.
- [18] H. Kawasaki, S. Wakisaka, H. Saito, A. Hiyama, M. Inami. A System for Augmenting Humans’ ability to Learn Kendama Tricks through Virtual Reality Training. *Proceedings of the Augmented Humans International Conference 2022 (AHS '22)*, pp. 152–161, 2022.
- [19] J. Adolf, P. Kán, B. Outram, H. Kaufmann, J. Doležal, L. Lhotská. Juggling in VR: Advantages of Immersive Virtual Reality in Juggling Learning. *25th ACM Symposium on Virtual Reality Software and Technology (Parramatta, NSW, Australia) (VRST '19)*, pp. 1-5, 2019.

- [20] S. Jiang, J. Rekimoto. Mediated-timescale learning: Manipulating timescales in virtual reality to improve real-world tennis forehand volley. *26th ACM Symposium on Virtual Reality Software and Technology*, pp. 1–2, 2020.
- [21] 松本 高, E. Wu, 小池 英樹. 時間歪曲機能を用いたVRスキートレーニングの拡張. WISS2021, pp.64–70, 2021.
- [22] T. Matsumoto, E. Wu, and H. Koike. Skiing, Fast and Slow: Evaluation of Time Distortion for VR Ski Training. *Proceedings of the Augmented Humans International Conference 2022 (AHS '22)*, pp. 142–151, 2022.
- [23] M. Löchtefeld, A. Krüger, H. Gellersen. DeceptiBike: Assessing the Perception of Speed Deception in a Virtual Reality Training Bike System. *Proceedings of the 9th Nordic Conference on Human-Computer Interaction (Gothenburg, Sweden) (NordiCHI '16)*. Article40, pp. 1-10, 2016.
- [24] 小川奈美, 伴祐樹, 櫻井翔, 鳴海拓志, 谷川智洋, 廣瀬通孝. えくす手: 変調バーチャルハンドへの即応的な身体所有感の生起による身体拡張システム, インタラクション 2016 論文集. pp. 1022-1027, 2016.
- [25] SUNHONG, 柴田史久, 木村朝子. 仮想空間における上肢伸長感覚の誘発要因の分析, 研究報告ヒューマンコンピュータインタラクション (HCI), 2018.22, pp. 1-6, 2018.
- [26] S. Faltaous, T. Winkler, C. Schneegass, U. Gruenefeld, S. Schneegass. Understanding Challenges and Opportunities of Technology-Supported Sign Language Learning. *Proceedings of the Augmented Humans International Conference 2022 (AHS '22)*, pp. 15–25, 2022.
- [27] S. Kasahara, J. Nishida, P. Lopes. Preemptive Action: Accelerating Human Reaction using Electrical Muscle Stimulation Without Compromising Agency. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*, Paper 643, pp. 1–15, 2019.
- [28] A. Takahashi, J. Brooks, H. Kajimoto, P. Lopes. Increasing Electrical Muscle Stimulation's Dexterity by means of Back of the Hand Actuation. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*, Article 216, pp. 1–12, 2021.
- [29] A. Ebisu, S. Hashizume, K. Suzuki, A. Ishii, M. Sakashita, Y. Ochiai. Stimulated Percussions: Techniques for Controlling Human as Percussive Musical Instrument by Using Electrical Muscle Stimulation. *SIGGRAPH ASIA 2016 Posters* (SA '16), Article 37, pp. 1-2, 2016.
- [30] A. Ebisu, S. Hashizume, K. Suzuki, A. Ishii, M. Sakashita, Y. Ochiai. Stimulated percussions: method to control human for learning music by using electrical muscle stimulation. *Proceedings of the 8th Augmented Human International Conference (AH '17)*, Article 33, pp. 1–5, 2017.
- [31] 北川リサ, 伊藤貴之. 競技かるたにおける払いの動作の可視化 , 研究報告ヒューマンコンピュータインタラクション (HCI) , pp. 1-7, 2022.
- [32] 山田浩史, 村尾和哉, 寺田努, 塚本昌彦. 競技かるたにおける手首動作に基づく札取得タイミング推定手法の提案, インタラクション 2015 論文集, pp. 969-971, 2015.
- [33] 都丸幸浩, 藤岡優, 安本匡佑, 羽田久一, 太田高志. 競技かるたの札の配置を行うインタラクティブコンテンツ, 第 75 回全国大会講演論文集 2013.1, pp. 255-256, 2013.
- [34] 松田孟留. 競技かるたの決まり字に関する統計的解析, 応用統計学 49.1, pp. 1-11, 2020.
- [35] 武田昌一, 長谷川優, 津久井勤, 桐生昭吾. 小倉百人一首競技かるた選手の出札認識タイミングの聴取実験による計測と音響的手がかりの推定, 電子情報通信学会技術研究報告, vol.114, pp. 141-146, 2015.
- [36] 徳島智春, 曽我真人. 競技かるたにおける決まり字変化シミュレーションシステムの構築, 人工知能学会全国大会論文集, 第 32 回, 2018.
- [37] D. Yu, H. Liang, X. Lu, K. Fan, and B. Ens. Modeling endpoint distribution of pointing selection tasks in virtual reality environments. *ACM Transactions on Graphics*, Vol.38, 6, Article 218, pp. 1-13, 2019.
- [38] P. Baudisch, E. Cutrell, M. Czerwinski, D. Robbins, P. Tandler, B. Bederson, A. Zierlinger. Drag-and-Pop and Drag-and-Pick: Techniques for Accessing Remote Screen Content on Touch-and Pen-Operated Systems. *Interact*, 3, pp.57-64, 2003.
- [39] M. McGuffin and R. Balakrishnan. Fitts' law and expanding targets: Experimental studies and designs for user interfaces. *ACM Transactions on Computer-Human Interaction*, pp. 388-422, 2005.
- [40] P. Baudisch, E. Cutrell, K. Hinckley, and A. Eversole. Snap-and-go: helping users align objects without the modality of traditional snapping. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '05)*. Association for Computing Machinery, pp.301–310, 2005.
- [41] D. Ahlström. Modeling and improving selection in cascading pull-down menus using Fitts' law, the

- steering law and force fields. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '05)*. Association for Computing Machinery, pp. 61–70, 2005.
- [42] I. Poupyrev, M. Billinghurst, S. Weghorst, and T. Ichikawa. The go-go interaction technique: non-linear mapping for direct manipulation in VR. *Proceedings of the 9th annual ACM symposium on User interface software and technology (UIST '96)*. Association for Computing Machinery, pp. 79–80, 1996.
- [43] T. Grossman and R. Balakrishnan. The bubble cursor: enhancing target acquisition by dynamic resizing of the cursor's activation area. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '05)*. Association for Computing Machinery, pp. 281–290, 2005.
- [44] NASA, MAN-SYSTEMS INTEGRATION STANDARDS, <https://msis.jsc.nasa.gov/sections/section03.htm> (2023/8/29 確認)
- [45] I. Scott MacKenzie and W. Buxton. Extending Fitts' law to two-dimensional tasks. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '92)*. Association for Computing Machinery, pp. 219–226, 1992.
- [46] D.Saffro, S.D.Bartolomeo, C.Yildirim, C.Dunne. Remote and Collaborative Virtual Reality Experiments via Social VR Platforms. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for

未来ビジョン

他の対人ゲームへの知見の適用

百人一首かるたはプレイヤ同士のインタラクションが少なく、札をとる時間を最小化するというゲームの目標がシンプルであるなどの特徴があり、支援をプレイヤごとに非均等、非対称に設定しても、それらが独立に個々のプレイヤにのみ作用しやすく、VRを用いたスキル調整が比較的行いやすいゲームであるといえる。百人一首かるたと似たような特徴を内包する他の対人ゲームを探査すれば、本研究で得た知見を直接的に応用することができるだろう。そのような事例の収集を通じて、対人ゲームの持つ特徴に応じた適切なスキル調整の提案のノウハウを蓄積していきたい。

Computing Machinery, No.:523, pp. 1-15, 2021.

- [47] M. Nancel, E.Lank. Modeling User Performance on Curved Constrained Paths, *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. Association for Computing Machinery, pp. 244–254, 2017.
- [48] Y. Itaguchi. Size Perception Bias and Reach-to-Grasp Kinematics: An Exploratory Study on the Virtual Hand with a Consumer Immersive Virtual-Reality Device. *Frontiers in Virtual Reality*, 2, 2021.

インクルーシブな百人一首かるたの実現

より多様な人が百人一首かるたを楽しめるように、支援の幅を広げたい。障がいへの対応としては、例えば歌の読み上げなどの聴覚情報を視覚情報として提示することで、聴覚障がいへの対応が可能になると考える。また、提示する色情報をプレイヤごとに調整することで、色覚多様性への対応も可能になるだろう。さらに、多言語対応機能を実装することで、非日本語話者も一緒に百人一首かるたを楽しむことがより容易になることが期待される。

なお、特に競争的な対人ゲームにおいては、多様な特徴を持った人々が参画できたとしても、参画意欲を維持するには適切なスキル調整が必要不可欠であろう。本論文で得られた知見はその実現に貢献するものである。

VR リダイレクションを用いたゴルフスイングトレーニングシステム

余 志皓 * 廖 振傑 * 小池 英樹 *

概要. 本論文は、ゴルフのトレーニングにおいて、スイング軌道を修正するために、仮想現実（VR）技術を用いたボール位置のリダイレクションを提案する。一連のスイング動作の中で、インパクト前にボールを前方に移動することで、ゴルファーが無意識にボールの移動に反応し、スイング軌道を修正する。このリダイレクションを達成するために、仮想環境（VE）を開発し、高速光学トラッキングデバイスを使用してゴルフクラブのモーションデータを収集し、リアルなシミュレーションを行った。本研究ではリダイレクションのためのボール移動パターンを3つ提案し、12人の被験者を対象としてユーザ実験を行った。この実験により、異なるパターンのトレーニングをしたゴルファーの学習率を評価し、それぞれのパターンの効果を検証した。その結果、ボールの「間欠的な（Interspersedly）」移動が、ゴルファーに無意識にスイング軌道を修正することができ、持続的な学習効果を示した。この手法をトレーニングに取り入れることにより、ユーザの偏ったフォームを改善し、技能レベルを向上させる可能性がある。

1 Introduction

For golf beginners, there is often a high degree of instability in the ball's flying direction when they strike it. This problem also happens to golfers with years of experience playing the round. They often perform a strong bias in their swing path, particularly tend to deflect from outward to inward, causing the ball to veer left of the target. This outward-to-inward path tendency often leads to a series of errors, one of which is commonly known as a “slicing shot”[1]. The slicing shot causes the ball to veer sharply in the opposite direction (right side) and is a challenge for many golfers [8]. There are various reasons for this outward-to-inward bias, most of which can be attributed to an erroneous swing motion called the “over-the-top swing”. Seeking guidance from a professional coach is arguably the most efficacious method to rectify these swing inaccuracies. However, the associated costs might deter beginners, particularly those pursuing golf as a mere pastime. A common alternative, self-practice, whether indoors or outdoors, may risk ingraining poor techniques in the absence of proper guidance.

This work proposes a virtual reality (VR) golf training system to help golfers correct inappropriate habits during a swing. We use a spatial distortion technique of VR to change the user's behavior, also called redirection, while the users

do not perceive the change of position. In the VR system, we dynamically move the ball position when the users are performing a downswing. When the ball shifts a short distance ahead, the user promptly shifts his downswing path forward, thus correcting his swing path to some extent.

Based on previous research exploring the prolonged aftereffect of visuomotor adaptation to gradually distorted reality in VR[7], we introduce three modes to shift the ball during the downswing. To evaluate the effectiveness of the proposed system, we develop a VR golf simulation system incorporating the three modes above and conduct a user study.

2 RELATED WORK

2.1 Visual/Object redirection

Visual dominance[2][3] can be used to influence the user's experience in the virtual environment and thus achieve specific desired effects, with redirection techniques being one of the most popular. For example, in redirected walking [13], users may perceive themselves as walking in a straight line, although they are walking on a curved path. A non-strict mapping between head rotation and orientation in the real and virtual world achieves this work.

Kasuga et al.[7] had participants wear VR headsets and had them perform the task of repetitive 3D arm reaching for a target on a touchpad in front of them. The target had multiple movement modes. In the abrupt condition, the im-

Copyright is held by the author(s).

* 東京工業大学

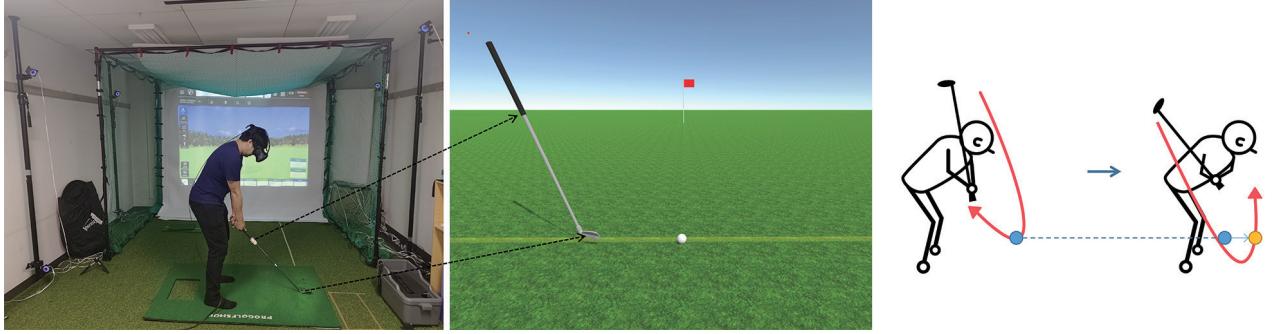


Fig.1. Make the ball shift forward by tracking the golf club while downswing, avoiding “over-the-top”-like motion that can cause outside-to-inside swing path.

ages were rotated by 20° , while in the gradual condition, the rotation was increased stepwise from 0° to 20° . Their results showed that the decay of the aftereffect was slower in the case of slow movement. This research motivated us to design modes in which the user is unaware of the ball’s movement. We hypothesize that this type of mode could potentially help users to maintain their training results effectively.

2.2 Golf training in HCI

The application of technology in golf training mainly includes the use of AR [4], VR [9], projection systems [14][5][6], and monitors to aid users to train through various visualizations. Other approaches involve incorporating forms of feedback, such as auditory cues [14][6], haptic feedback achieved by modifying golf clubs [10][11], or external haptic feedback devices [15]. Moreover, research that helps users train by deep learning the swings of skilled players has been actively researched in recent years [12].

In a series of research [14][5][6] focusing on the use of the so-called “Virtual Golf Shadow” for golf training, a skilled golfer’s silhouette captured through motion capture technology is projected onto an indoor golf training course. Users can learn and improve their golf swing in real-time by consciously aligning their silhouette with the skilled golfer’s. However, this study is only applicable to beginners who need to start learning from posture. Moreover, habits such as “over-the-top swing”, caused by many personal preferences and details, are challenging to correct simply by differentiating from experts. Our approach can be used not only for beginners to practice but also for experienced users to correct the swing path.

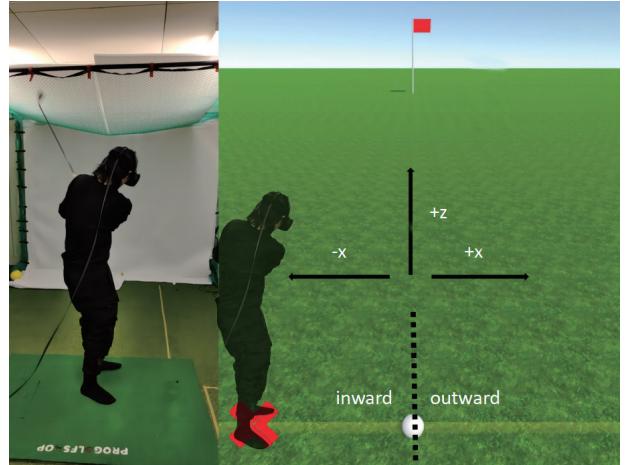


Fig.2. Left: The user takes the HMD for a real swing. Right: the user interface in VR.

3 Design and implementation

The system design is shown in Fig.1. To get accurate motion tracking, we use a high-precision motion capture system (OptiTrack) to track the golf clubs and simulate the outcome of the ball being hit in VR in real time. Finally, we developed three different golf ball shift modes for training using the VR golf setting.

3.1 User interface in VR

The user interface in VR is shown in Fig.2. The user must stand on the yellow line and swing, with the red flag as a target on the user’s left side.

According to golf rules, a golf ball’s diameter should not be less than 42.7mm. Hence, the diameter of our model ball is set to 43mm. To make it easier for beginners to hit the ball, we let the ball model’s collision body diameter be

VR リダイレクションを用いたゴルフスイングトレーニングシステム

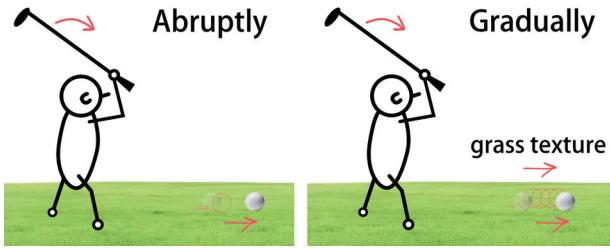


Fig.3. Ball shift sketch for Abruptly and Gradually modes.

1.6 times larger than itself, which is 68mm.

In the VR system, when the user hits the ball with a certain force, the ball flies slightly higher than in reality, while the speed in the x and z directions remains the same as in reality. Our research focuses on the swing path rather than the angle of the golf club face, which allows the user to observe the ball's flight trajectory more comprehensively. The ball's flight trajectory disappears after the ball hits the grass, while a new ball subsequently appears on the ball spawner for the user to continue training.

3.2 Shift of the ball

We developed three shifting modes in which the ball shifts the same distance. The distance of the ball movement is two times the diameter of the ball model, 86mm. The previous study by Kasuga et al.[7] mentioned four modes: abrupt condition, gradual condition, a mixture of abrupt condition and baseline, and a mixture of gradual condition and baseline. Among them, gradual condition and the mixture of gradual condition and baseline are the most effective, with a slower decay of after-effects. In contrast, the mixture of abrupt conditions and baseline mode has the worst effect. To verify whether abrupt and gradual conditions produced similar results in our study, we implemented “Abruptly”, “Gradually”, and the “Interspersedly” modes for the user study.

Abruptly (A) In this mode, when the user is preparing to hit the ball, the ball suddenly shifts towards the front for a distance (Fig.3 Left).

Gradually (G) This mode makes the ball shift slowly towards the front for a distance and tries not to be noticed by the user. To achieve this, we take advantage of the feature of the tile texture of grass by having the texture move forward with

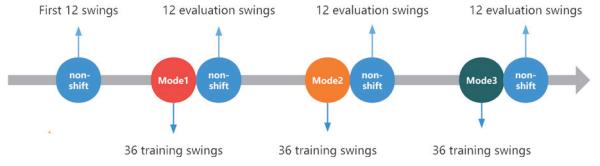


Fig.4. The procedure of user study: Mode 1, Mode 2, and Mode 3 represent the 3 modes in a randomized order.

the ball such that the user would lose their frame of reference and be as unaware of the ball's shift as possible (Fig.3 Right).

Interspersedly (I) This mode alternates between the **G** mode and the non-shift mode. We aim to correct the user's habit with the **G** mode swing while preventing over-correction with the “non-shift” mode, allowing the user to learn and verify between the two modes.

4 User Study

To test and verify the hypothesis that shifting the ball during the swing can improve the user's swing path, we conducted a user study involving 12 participants.

4.1 Design and procedure

We first give instructions to the participants on how to swing the golf club properly by having them practice on a real-world indoor golf course. As done in actual training and golfing, we asked the participants to aim at the ball and keep their eyes on it throughout the swing. After we ensure they are familiar with the basic swing by completing a full swing 5 times, we let them train with VR headsets and try to hit the ball in VR. We have all participants use a #7 Iron club, stay in a line throughout the test, and the participant can see the ball's flight path. When the participants hit the ball four times continuously in VR, and it lands in the range of -30° to 30° , we consider that they are ready to train in VR. Then, we give instructions to the user about the experiment, including matters such as the purpose of the study and the three different modes.

During the experiment, we had each participant test all three modes randomly. Before testing the three modes, the user performed 12 swings in the non-shift mode as the baseline. As shown in Fig.4, for the three different modes, such as **A** mode, we let the user perform 36 swings in the

A mode, followed by 12 swings again in the non-shift mode. We mainly use the last 12 swings to evaluate the effective learning rate. Particularly, in the **I** mode, participants first performed 6 swings in non-shift mode, followed by 6 swings in the **G** mode. The participants repeat these two modes until they complete 36 swings and then perform the final 12 swings in the non-shift mode for evaluation. The training is approximately 10 minutes for 48 swings in each condition. We asked the participants to take a 10-minute break between conditions while filling out a questionnaire about this part of the experiment. The whole experiment lasts approximately one hour.

4.2 Participants

We invited a total of 12 participants ($\min_{age} = 20$, $\max_{age} = 31$, $\text{mean}_{age} = 24.25$, $sd_{age} = 3.16$). Eight of them were men, and four were women. Three had no knowledge of golf, six had some knowledge of golf, they may have watched golf videos or practiced swings, and the other three had some golf experience, they frequently engage in indoor golf training or have practical experience playing golf games. Four of them used VR regularly, five of them had VR experience, and three of them had no VR experience at all.

4.3 Questionnaire

We used a customized questionnaire to evaluate the participant's subjective experiences in each mode, which served to support the discussion of the result. We presented four questions for each mode using a 5-point Likert scale questionnaire:

- Q1: How about your experience of this condition?
- Q2: Can you feel the ball shifting when you swing?
- Q3: Do you think the shift of the ball disturbed you?
- Q4: Do you think your swing form changed?

5 Results

5.1 Velocity bias

In our VR environment, the $-x$ direction represents the inward, the $+x$ direction represents the outward, and the $+z$ direction represents the correct direction, i.e., the center(Fig.2). If the x

component of the club velocity is negative, the hit will cause the ball to fly to the left and vice versa to the right.

v_x/v_z represents the offset value of the velocity acting on the ball concerning v_x , which we call the **velocity bias**. $\arctan(v_x/v_z)$ represents the offset of the velocity in degrees. When $v_x/v_z = 0.1$, it represents an offset of 5.71° , and 0.2 represents an offset of 11.3° . If v_x/v_z is negative, the swing path swings outward to inward and vice versa. If all other factors are accurate, then v_x/v_z will represent the final deflection of the ball. We want to see a smaller absolute value of velocity bias during training than during the initial measurement. More importantly, we want the absolute value of velocity bias to be as small as possible after training, representing a significant training effect in the normal mode even after training.

We first conducted a two-way repeated-measure ANOVA with turn and mode as variables for the three modes(normally distributed and satisfies the spherical assumption). The results showed that for the three modes, both turn($df = 47, f = 1.93, p < 0.001$) and v_x/v_z ($df = 2, f = 39.10, p < 0.001$) or the interaction between turn and v_x/v_z ($df = 94, f = 1.40, p < 0.01$) have significantly differences. We then performed a Tukey-Kramer posthoc test on the three models and found significant differences between **(A, I)** and **(G, I)** in all three models (both $p < 0.05$). No significant difference occurred between **A** mode and **G** mode($p = 0.09 > 0.05$).

5.2 Average velocity bias

Using the data from the 12 times non-shift mode swing at the beginning, we found that 9 of the 12 participants(P1, P2, P3, P4, P5, P6, P7, P9, P12) were habitually biased from outward to inward, two participants(P10, P11) were between inward and outward, one participant(P8) was habitually biased from inward to outward. Because our ideal target is who is used to playing from the outward bias toward the in-ward, we first averaged the data of these nine participants individually, as shown in Fig.5. From the average data, the velocity bias (v_x/v_z) of these nine participants in the first 12 shots ranged between -0.2 and -0.1 , which means that on average, each swing would have a bias of approximately 9° from outward to inward.

We analyze the distribution of velocity bias

VR リダイレクションを用いたゴルフスイングトレーニングシステム

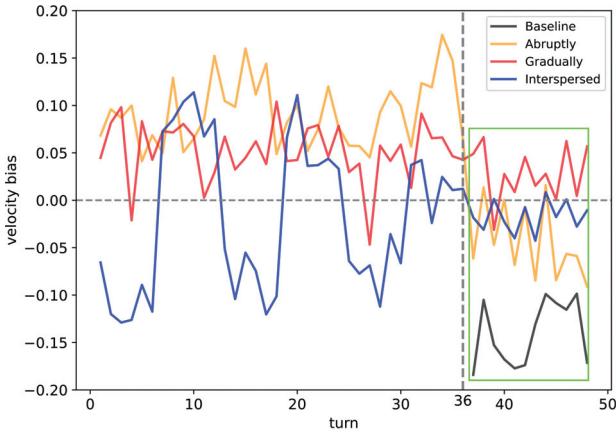


Fig.5. The average Velocity Bias of the 9 out of 12 participants who have a habit of swinging from the outside to the inside.

for the first 36 times in all three modes. In the **A** mode, the values were the largest among the three modes most of the time, and the velocity bias was between 0.05 and 0.15 mostly, meaning that the swing path was from inward to outward. In the **G** mode, the velocity bias fluctuated around 0.05, meaning that the average bias angle was close to 3 degrees, and the bias towards either side was not very strong. Intriguing results exist in the **I** mode because it is performed in a cycle of six times in the non-shift mode and six times in the **G** mode. The results of the first 36 times are distributed in positive and negative sides with six as the number of cycles. However, we can observe an overall tendency to converge to 0, and the velocity bias is already close to 0 in the 30th to 36th times in the **G** mode.

5.3 Questionnaire evaluation

11 out of 12 participants thought it was easier to hit a golf ball in VR than in reality, and one thought it was about the same difficulty. Nine people thought golf training using VR was reliable and preferred practicing in a VR environment. The questionnaire results for each mode are depicted in Fig. 6. In terms of the overall experience, **G** ($M = 4.25$, $sd = 0.72$) and **I** ($M = 4$, $sd = 0.91$) were better than **A** ($M = 3.42$, $sd = 1.03$). The majority of participants noticed the movement of the ball in the **A** mode ($M = 1.5$, $sd = 0.76$), whereas in the **G** ($M = 3.92$, $sd = 1.26$) and **I** ($M = 3.58$, $sd = 1.1$) modes, users did not perceive the ball's movement frequently. Additionally, users felt

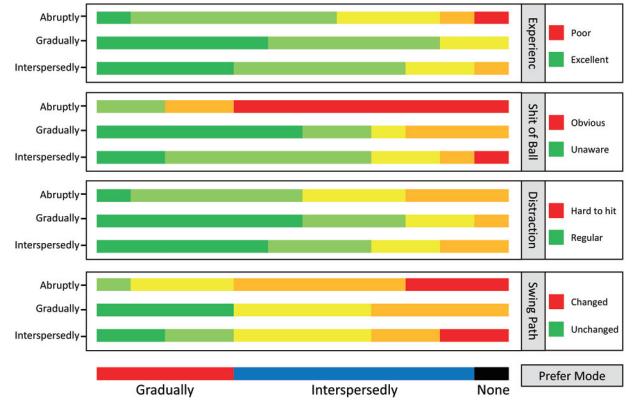


Fig.6. Questionnaire on a 5-point Likert scale for participants' experiences, perception of ball movement, discomfort, and conscious alteration of the downswing path for each mode.

more disrupted and found it more challenging to hit the ball in the **A** mode, requiring a conscious effort to modify their downswing path. After conducting all three modes and understanding the purpose of our user study, the majority of participants said they preferred the **I** mode for training.

6 Discussion

6.1 Main findings

From the velocity bias, it can be found that all three modes have shown positive effects during the training. **A** and **G** modes showed the most significant change. Due to its cyclical nature, the **I** mode experiences fluctuations over time, but gradually approaches zero. This result is further supported by the questionnaire responses. The questionnaire results indicate users were not aware of the shift of the ball in the **G** and **I** modes. Therefore, users reported slight discomfort only in the **A** mode due to the instantaneous shift of the ball, which made it challenging to make accurate hits.

After 36 swings of training, the velocity bias in the **A** mode regressed to the baseline level. The **G** mode and **I** mode maintained the training effects in the short term and showed similar levels. (*RQ 4* is true) The questionnaire also supported the superior overall performance of the **I** mode. We believe this is because users can make non-deliberate swing path corrections during training in the **I** mode, which allows the user to validate

and learn.

6.2 Subjective Feedback

Some participants thought that for golf beginners, training directly with a real ball is challenging to hit and not safe enough. “I like that there was no danger of having to pick up the ball or risk a flying ball.”(P6) Hence, golf training using VR is appropriate for beginners. VR golf can help beginners get up to speed rapidly and is very motivating. “It’s exciting that you could hit the ball constantly without catching the ball.”(P4) Furthermore, visualizing the ball’s trajectory, where it lands, and other information in VR can be very beneficial for regular indoor golf training. “it is really fun that I can hit the ball and can see the trajectory of the ball.”(P7) “I firstly touch the golf club, but I really enjoyed the golf training in all section, because I can hit the ball. It is useful to beginners to the training to motivate them.”(P12)

For perceptions of ball shift in different modes. P6 mentioned: “In the **A** mode, the ball position changed so significantly that I felt I needed to consciously adjust my posture. There wasn’t much difference between the **G** and **I** modes because I didn’t even notice the change in ball position.” P12 commented: “I occasionally feel a sense of space being stretched in the **G** and **I** modes, but despite this, I don’t consciously wonder if this is the ball shifting because the downswing is really fast.”

Other comments were concerned with correction, i.e., what was originally an outside-in trend but was corrected to an inside-out trend. Under the distance when we set the ball to shift, the learning result of the **A** mode was not effectively retained, and the over-correction phenomenon occurred under the **G** mode. At the same time, the **I** mode can be contracted to a more reasonable interval by the user’s verification during training. However, for user trends that are not outside-in or more exaggerated, it is unsuitable to train directly at this direction and distance of ball shift.

6.3 Limitations and future work

Several participants also indicated that even in the **G** and **I** modes, they were still distracted by the ball’s movement. In our user study, 1 of the 12 participants had no significant swing bias, and 2 had an inward-to-outward bias, which

might cause poor bias guidance when using our method.

Omnidirectional Redirection For different deflections, including inward-to-outward or outward-to-inward, other users may have different deflections when hitting the ball. Each user has a different bias; therefore, fixed ball shift distance and direction do not suit everyone. We plan to develop an adaptive training system that constantly adjusts to each user’s bias. This will not only help users with a significant bias to standardize their swing path but also help users with a smaller bias each time to reduce their bias trend.

Spatial Distortion of Angles Using object position redirection to correct the downswing path may somewhat change the overall swing plane. Our evaluation is mainly based on velocity bias, i.e., angles, so we plan to use spatial distortion of angles as well to allow users to learn unconsciously. We plan to use virtual arrows on the ground to indicate the direction of the user’s swing while gradually changing the direction of the arrows during the swing process. This may result in better correction with less discomfort.

7 Conclusion

In this work, we developed a VR method to help users correct their swing path by shifting and redirecting golf balls. We proposed three ball-shifting modes and conducted a user study to compare the advantages and disadvantages of the three modes. Subsequently, we analyzed the three modes’ learning rate by visualizing the golf club’s speed and swing path during and after the training sessions. Our findings revealed that training with the **I** mode, involving golf ball redirection, was particularly effective in helping individuals with an outward-to-inward swing bias standardize their swing path. In the future, we plan to enhance the system by incorporating other feedback mechanisms and omnidirectional adaptive redirection and conduct further studies on other motor skill training. These future improvements aim to provide many users with further assistance and immersive training experiences.

謝辞

本研究は JST CREST JPMJCR17A3 の支援を受けています。

参考文献

- [1] B. Baic. LEARNING GOLF DRIVE: NATURAL SWING PATH TENDENCY TO SLICE, FADE OR PULL. 2014.
- [2] E. Burns, S. Razzaque, A. Panter, M. Whitton, M. McCallus, and F. Brooks. The hand is slower than the eye: a quantitative exploration of visual dominance over proprioception. In *IEEE Proceedings. VR 2005. Virtual Reality, 2005.*, pp. 3–10, 2005.
- [3] J. J. Gibson. Adaptation, after-effect and contrast in the perception of curved lines. *Journal of experimental psychology*, 16(1):1, 1933.
- [4] A. Govil, S. You, and U. Neumann. A Video-Based Augmented Reality Golf Simulator. In *Proceedings of the Eighth ACM International Conference on Multimedia, MULTIMEDIA '00*, p. 489–490, New York, NY, USA, 2000. Association for Computing Machinery.
- [5] A. Ikeda, D.-H. Hwang, and H. Koike. Real-Time Visual Feedback for Golf Training Using Virtual Shadow. In *Proceedings of the 2018 ACM International Conference on Interactive Surfaces and Spaces, ISS '18*, p. 445–448, New York, NY, USA, 2018. Association for Computing Machinery.
- [6] A. Ikeda, Y. Tanaka, D.-H. Hwang, H. Kon, and H. Koike. Golf Training System Using Sonification and Virtual Shadow. In *ACM SIGGRAPH 2019 Emerging Technologies, SIGGRAPH '19*, New York, NY, USA, 2019. Association for Computing Machinery.
- [7] S. Kasuga, R. Mori, S. Kasahara, J. Rekimoto, and J. Ushiba. Prolonged Aftereffect of Visuomotor Adaptation to Gradually Distorted Reality Displayed on a See-Through Head-Mounted Device. *International Journal of Human-Computer Interaction*, 35(15):1345–1351, 2019.
- [8] J. J. Kawada. Kinematic analysis of an outside-to-inside golf swing path. p. 74. California State University, Sacramento, 2022.
- [9] P. Kelly, A. Healy, K. Moran, and N. E. O'Connor. A Virtual Coaching Environment for Improving Golf Swing Technique. In *Proceedings of the 2010 ACM Workshop on Surreal Media and Virtual Cloning, SMVC '10*, p. 51–56, New York, NY, USA, 2010. Association for Computing Machinery.
- [10] C.-C. Liao, H. Koike, and T. Nakamura. Realtime Center of Mass Adjustment via Weight Switching Device inside a Golf Putter. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems, CHI EA '20*, p. 1–8, New York, NY, USA, 2020. Association for Computing Machinery.
- [11] T. Nakamura, D. Saito, E. Wu, and H. Koike. Actuated Club: Modification of Golf-Club Posture with Force Feedback and Motion Prediction in VR Environment. In *ACM SIGGRAPH 2020 Emerging Technologies, SIGGRAPH '20*, New York, NY, USA, 2020. Association for Computing Machinery.
- [12] T. Nakamura, D. Saito, E. Wu, and H. Koike. Actuated Club: Modification of Golf-Club Posture with Force Feedback and Motion Prediction in VR Environment. In *ACM SIGGRAPH 2020 Emerging Technologies, SIGGRAPH '20*, New York, NY, USA, 2020. Association for Computing Machinery.
- [13] S. Razzaque, Z. Kohn, and M. C. Whitton. Redirected Walking. In *Eurographics 2001 - Short Presentations*. Eurographics Association, 2001.
- [14] Y. Tanaka, H. Kon, and H. Koike. A Real-Time Golf-Swing Training System Using Sonification and Sound Image Localization. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology, VRST '18*, New York, NY, USA, 2018. Association for Computing Machinery.
- [15] M. P. Woźniak, J. Dominiak, M. Pieprzowski, P. Ladoński, K. Grudzień, L. Lischke, A. Romanowski, and P. W. Woźniak. Subtletee: Augmenting Posture Awareness for Beginner Golfers. *Proc. ACM Hum.-Comput. Interact.*, 4(ISS), nov 2020.

拡張現実を用いた姿勢可視化によるスキートレーニングシステム

松本 高^{*} Erwin Wu^{*†} Chen-Chieh Liao^{*} 小池 英樹^{*}

概要. スキーは人気の高いウィンタースポーツであり、トレーニング効率の向上を目的としたシステムが数多く提案されている。しかし、既存システムの多くは特定のシミュレータマシン上でしか使用できず、実際のトレーニングとのギャップが大きい。また、スキーの重要な要素の一つである上半身の姿勢の改善に焦点を置いたシステムは少ない。本論文では、シースルー型ヘッドマウントディスプレイ (HMD) を用いた拡張現実 (AR) スキートレーニングシステムを提案する。提案システムでは、滑走姿勢を改善するための3種類の視覚的フィードバックをリアルタイムに提示する。評価実験によれば、ユーザの背骨を可視化する手法は従来手法よりも約7%姿勢が改善され、最も高いパフォーマンスが得られることが示された。最後に、実際のスロープ上でのパイロットテストにより、実環境においても提案システムが適用できることを確認した。

1 はじめに

仮想現実 (VR) や拡張現実 (AR) などのXR技術の発展に伴い、仮想空間は技術習得の場面にも多く用いられるようになってきている。

アルペンスキーのトレーニングでは、場所や季節に対する制限や、リアルタイムに姿勢評価ができない問題がある。これらの課題に対する様々なシステムが提案されている[1, 2, 8, 14, 16, 18, 19, 21, 23, 28, 29]。その中には、ウェアラブルセンサ[23]や聴覚フィードバック[14]を活用して、実際のスロープ上での指導を行うものがある。しかし、スキーの効果的なトレーニングのためには、視覚的フィードバックを用いることが最適であると示唆されている[16]。

既存の視覚ベースのスキートレーニングシステムの多くはVRを用いてユーザに情報を提示している[16, 21, 29]。しかし、ユーザが周囲を見ることができないVRシステムは、安全上の理由からユーザの位置が変化しない小規模なシミュレータ上のみでの使用に制限されるものがほとんどである。

本論文では、これらの問題に対処するため、ビデオシースルー型のヘッドマウントディスプレイ (HMD) を利用して、視覚的フィードバックを提示するARスキートレーニングシステムを提案する(図1)。提案システムの主な対象は、スラロームスキーにおける旗を通過することができる一方、滑走中に正しい姿勢を意識する余裕がない中級スキーヤーである。

提案システムを評価し、ARにおける最適な視覚的フィードバックを探索するための評価実験を行う。季節の制限や安全上の理由から、実験には横幅約6mの大規模スキーシミュレータを代替として用いる。

評価実験では、3つの視覚化手法(図2)を比較する:Follow The Proはコーチに追従する既存の可視

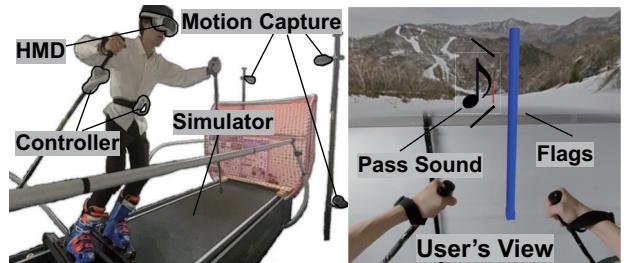


図 1. 提案システムの概要。

化手法[29], Spine Angle Indicatorはユーザの背骨を可視化するもの、Keep Horizontal Poleは仮想のポールを水平に保つことで間接的に姿勢を矯正するものである。条件間の比較として、作業負荷[12]、ユーザ経験[25]、ユーザビリティ[4]に関する複数のアンケートによる定性的評価と、滑走のパフォーマンスによる定量的評価を行う。結果として、ユーザの背骨を可視化する手法は、複数の難易度条件・評価指標において既存の可視化手法を上回り、今後のARトレーニング開発の方向性を示唆している。

最後に、提案システムを用いて、HMDを装着した状態でも実際の雪上でターンが行えることを確認する(図4)。

2 関連研究

2.1 ARを用いた技能習得支援

様々な分野の技術習得において、ARが活用されている。Hanら[11]は、太極拳の自己トレーニングを支援するシステムを提案した。ユーザは、シースルー型のHMDを用いて、コーチや自身の動きを観察することができる。Dorzhievaら[6]は、ARとドローン、および触覚インターフェースを組み合わせたアーチェリートレーニングシステムを提案した。

Copyright is held by the author(s).

* 東京工業大学

† 華為技術日本株式会社

彼らの実験結果は、技能習得において仮想的な視覚フィードバックと物理的な小道具を組み合わせることの有効性を示唆している。Suzuki ら [30] は、合気道の指導において用いられる比喩表現を AR グラス上に視覚化することで、身体の動かし方に関する感覚の理解を支援するシステムを提案した。他にも、メンテナンス [15, 27]、回路基板のデバッグ [5]、ヨガ [17]、歩行リハビリ [10]、バスケットボールのフリースロー [20]、ポールジャグリング [22]、脳外科手術のカテーテル [7] などの AR システムが存在している。

2.2 スキーシミュレーション

スキーシミュレーションを用いて、スキーのトレーニング効果を高める研究が数多く行われている。Aleshin ら [1, 2] は、屋内用スキーシミュレータと仮想スロープの投影を組み合わせたシステムを提案し、初心者とアスリートの両方に対して提案システムが有効であることを示した。Panizzplo ら [23] は、スラロームスキーにおける筋活動パターンを、自然雪上、本格的なトレッドミル、家庭用トレーニングマシンの 3 つの異なる条件下で比較した。Hasegawa ら [14] は、聴覚フィードバックを用いて滑走中の重心位置をリアルタイムに提示した。

VR とスキーシミュレータを組み合わせたシステムも多く提案されており、VR コンテンツによりユーザーの集中力を高める [18]、様々な機能を用いてユーザーとコーチの差を可視化する [28, 29]、触覚フィードバックを用いて理想的な膝の角度を提示する [16]、時間の経過速度を変化させる [21] システムなどが存在している。また、AR を用いてスロープ上に仮想の障害物を動的に生成するシステム [8] や、ユーザー間で通過ルートを共有するシステム [9] が提案されている。

しかし、既存のスキーシステムの多くは、ユーザーの位置がほとんど動かない簡易的なスキミュレータ上での使用のみに制限されており、実際のトレーニングとのギャップが大きい。また、足首の角度などユーザーの下半身の姿勢を評価した研究 [16, 29] はあるが、上半身の姿勢が改善されたかどうかを定量的に評価したものは少ない。

3 システムデザイン

3.1 ハードウェア

提案システムは、ビデオシースルー型 HMD(PICO4¹)、及び 2 つのコントローラから構成される(図 1(左))。今回は、実際のスロープの代替として、U.S. Ski Team²がトレーニングに使用している大規模な屋内スキーシミュレータ(SkyTechSport³)を用い

る。ユーザはスキーブーツを履いてスキミュレータに乗り、ロール軸に足を傾けることで実際のパラレルターンに近い軌道で左右方向に移動できる。また、スキミュレータの周囲には 8 台のモーションキャプチャカメラ⁴を設置し、ユーザの 3 次元姿勢を取得する。

3.2 仮想スロープ

没入感の高いスキートレーニングを実現するために、Unity を用いて仮想スキースロープを作成する(図 1(右))。スロープには仮想の旗が設置されており、一定の速度で手前に移動する。ユーザはスキーシミュレータを操作して横方向に移動し、旗の外側を通過する。ユーザが装着した HMD の横方向の座標を参照し、旗の外側を通過した際にターン成功の効果音が鳴る。

没入感を高めるため、仮想空間の背景には雪山の球体画像を投影する。ただし、球体画像の下半分は透明になっており、ユーザは HMD を装着していても自身の位置や姿勢を把握することができる。

3.3 視覚的フィードバック

スキーにおいて、姿勢は重要な要素の一つである。提案システムでは、以下の 3 つの視覚的フィードバックを用いてユーザの滑走時の姿勢改善を支援する。

3.3.1 Follow The Pro (Cond.F)

スキーの一般的なトレーニング方法として、コーチの動きを模倣しながら追従するものがある。Follow The Pro(Cond.F) では、Wu ら [29] の方法を参考にし、事前に記録したエキスパート 1 名(ワールドカップレーサー)の全身姿勢をユーザの 2.5m 前方に表示する(図 2(左))。

3.3.2 Spine Angle Indicator (Cond.S)

スキーの技術を向上させるためには、上半身を地面に対して垂直に保つことが重要である⁵。しかし、滑走中の自身の姿勢を握りし、改善することは困難である。Spine Angle Indicator(Cond.S)では、PICO4 のコントローラを腰の中心に固定し、ユーザの背骨の向き(コントローラから HMD へのベクトル)をインジケータとして可視化する(図 2(中))。インジケータは、ユーザが上半身を地面と垂直に保てている間は白色であり、一定以上の角度に傾くと赤色に変化する。事前の検証実験により、インジケータが赤色になるしきい値を 10 度とした。

3.3.3 Keep Horizontal Pole (Cond.K)

初中級者の場合、自分の姿勢を握りしいても、それを改善する余裕がないことが考えられる。これに対し、フィードバックを間接的なものに置き換え

¹ <https://www.picoxr.com>

² <https://www.simsportsarena.com/us-ski-team>

³ <https://www.skytechsport.com>

⁴ <https://qualisys.archivetips.com>

⁵ <https://ltad.alpinecanada.org/coaches>

拡張現実を用いた姿勢可視化によるスキートレーニングシステム

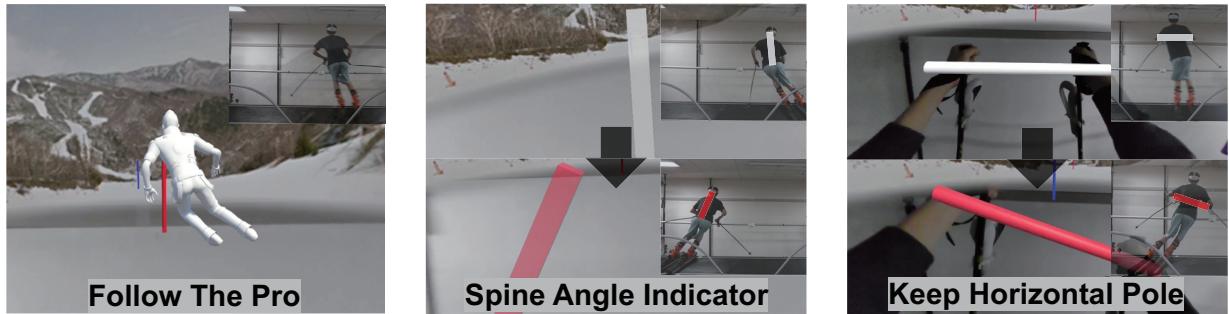


図 2. 提案システムの 3 つの視覚的フィードバック.

ることより、姿勢改善を支援するアプローチが考えられる [30]. Keep Horizontal Pole(Cond.K) では、Alpine Canada⁵が提唱するトレーニング法を参考にし、ユーザの手元に仮想のポールを表示する(図 2(右)). このポールを水平に保つことで、間接的な上半身の姿勢の改善を促す。ユーザの手の位置は 2 本のスキーストックに取り付けられたコントローラを用いて追跡する。(1) 手の位置が頭から 50cm 以上低い、(2) 左右の手の高さが水平から 20 度以上傾いている、(3) 手が前方に 35cm 以上伸びていない、のいずれかの場合に、ポールが白色から赤色に変化する。各しきい値は Cond.S と同様に決定された。

4 評価実験

提案システムの評価および、3つの視覚的フィードバックの有効性を比較するための評価実験を行った。

本実験は、東京工業大学倫理審査委員会の承認を受け実施された。

4.1 仮説

本実験では、以下の仮説を立て、後述する評価指標を用いて検証を行った:

H1: 抽象的な姿勢可視化によるトレーニング効果

コーチの全身姿勢をそのまま提示する Cond.F に比べ、ユーザの姿勢を抽象化して提示する Cond.S と Cond.K はより効果的である。

H2: 間接的な姿勢可視化によるトレーニング効果

直接的に姿勢を提示する Cond.S に対し、間接的に姿勢を提示する Cond.K はより効果的である。

すなわち、3つの視覚的フィードバックは Cond.F < Cond.S < Cond.K の順に効果が高い。

4.2 実験参加者

18名のスキー経験者(女性 5名、男性 13名、年齢 19~28 歳、平均 21.7、標準偏差 2.05)が本実験に参加した。参加者のうち、2名が 2 年以上 5 年未満のスキー経験があり、残りの 16 名は 5 年以上の経験があった。参加者全員が実際のスロープ上でパラルターンができると回答しており、スキー中級者と

みなせる。XR 機器の使用経験は、4名が年に 1 回程度、1名が週に 1 回程度の経験があり、残りの 13 名は全くの経験がなかった。参加者には謝礼として 2,000 円が支払われた。

4.3 評価指標

各条件を、滑走時のパフォーマンスを用いて定量的に、かつ質問票により定性的に評価した。

4.3.1 定量的指標

Number of Passed Flags (N.P.F): 30 本の旗を 1 セットとして、正しく通過することができた旗の本数。値が大きいほど良い。

Mean Spine Angle (M.S.A): 参加者の背骨(モーションキャプチャによって計算される骨盤から胸部へのベクトル)のロール方向の傾きと、地面に垂直な方向がなす角の平均値。1 セット(旗 30 本)を通過するまでの全フレームの背骨の角度が使用される。値が小さいほど良い。

Error (E.R.R): M.S.A を旗の通過率で割った値。背骨の角度に対し旗の通過率が低い分のペナルティを与えることになり、値が小さいほど良い。例えば、M.S.A が 10、N.P.F が 15 の場合、E.R.R は $10 \div (15/30) = 20$ となる。

4.3.2 定性的評価

定性的評価には System Usability Scale(SUS)[4]、User Experience Questionnaire(UEQ)[24]、NASA-TLX[13] 質問票を用いた。ただし、実験手順の簡略化のため、Short version UEQ(UEQ-S)[25] と RAW NASA-TLX[12] を使用した。

4.3.3 難易度設定

本実験では、1 セット 30 本の旗を奥行き 15m 間隔、横幅 1m で左右交互に設置した。複数の難易度におけるパフォーマンスを測定するため、Easy Level(時速 60km、中級者がある程度余裕を持って旗を通過できる速度)と Hard Level(時速 80km、上級者でないと全ての旗を通過するのが難しい速度)の 2 つの難易度を設定した。

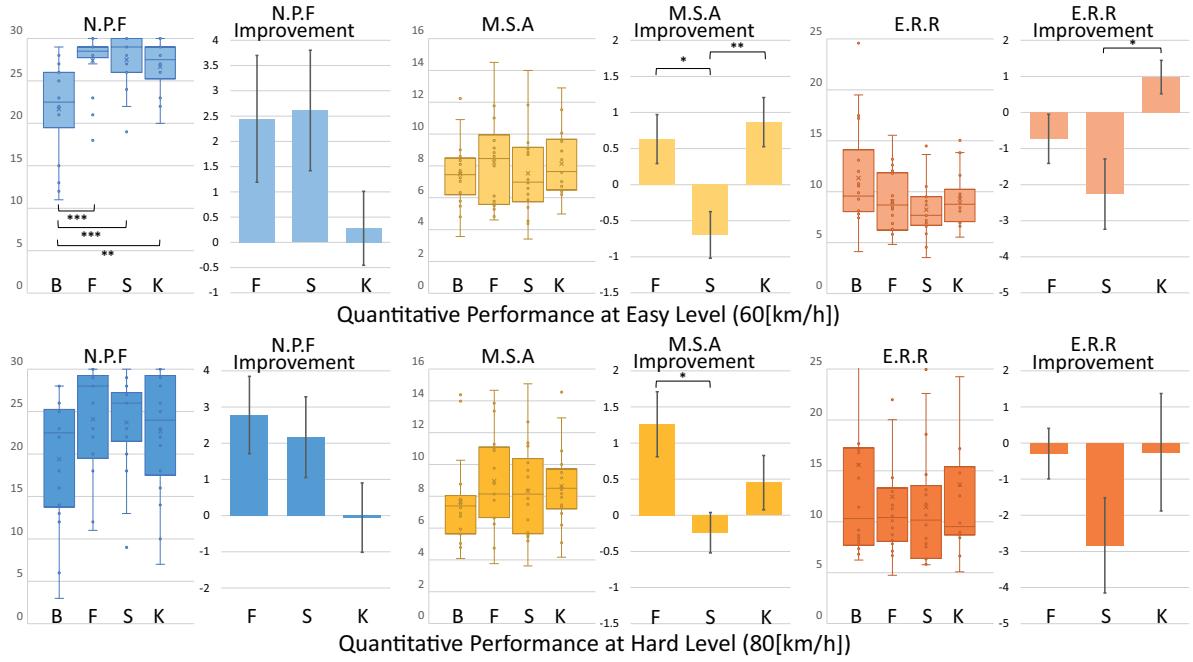


図 3. 定量的結果 (* : $p < 0.05$, ** : $p < 0.01$, *** : $p < 0.001$). N.P.F は値が大きいほど良く, M.S.A 及び E.R.R は値が小さいほど良い.

4.4 実験手順

参加者への説明の後, 以下の手順で実験を行った.

- (1) シミュレータに慣れるための約 15 分間のウォーミングアップを行う.
- (2) HMD を装着し, 参加者が提案システムに慣れるまで視覚的フィードバックのない基本条件 (Cond.B) で初期トレーニングを行う.
- (3) 3 つの視覚的フィードバックのうち 1 つの条件で Easy Level の練習を 3 セット行う.
- (4) Cond.B の 1 セットを (3) と同じレベルで行う.
- (5) Hard Level で (3) と (4) を繰り返す.
- (6) その条件についての質問票に回答する.
- (7) 最低 10 分間の休憩を取り, 直前に行った条件の学習効果を軽減する.
- (8) (3)～(7) を 3 つの条件で繰り返す. 条件の順番は参加者ごとにカウンターバランスする.
- (9) システム全体と各条件の有効性や改善点についての口頭インタビューを行う.

手順 (2) の最後の 1 セットと (4) で記録された Cond.B(視覚的フィードバック無し) のパフォーマンスを用いて, 各条件のトレーニング効果を比較する.

4.5 結果

4.5.1 定量的結果

図 3 の箱ひげ図は定量的指標の結果を示し, 棒グラフは直前に行った条件と比較した改善度合い (Im

provement) を示す. ただし, 3 つの条件のうち, その参加者の最初に行った条件の場合, 実験手順 (2) の最後の 1 セットの結果 (Cond.B) と比較する.

Easy Level に関して, One-way ANOVA により, N.P.F ($F_{3,68} = 9.220, p < 0.001$), M.S.A Improvement ($F_{2,51} = 6.004, p < 0.01$), E.R.R Improvement ($F_{2,51} = 4.569, p < 0.05$) に条件間の有意差が見られた. 他の指標では有意差は見られなかった (N.P.F Improvement: $F_{2,51} = 1.539, p = 0.266$, M.S.A: $F_{3,68} = 0.455, p = 0.715$, E.R.R: $F_{3,68} = 6.004, p = 0.0682$).

Hard Level に関して, M.S.A Improvement において条件間の有意差が見られた ($F_{2,51} = 3.801, p < 0.05$). 他の指標では有意差は見られなかった (N.P.F: $F_{3,68} = 1.795, p = 0.156$, N.P.F Improvement: $F_{2,51} = 1.910, p = 0.159$, M.S.A: $F_{3,68} = 0.692, p = 0.560$, E.R.R: $F_{3,68} = 0.685, p = 0.564$, E.R.R Improvement: $F_{2,51} = 1.268, p = 0.290$).

有意差が見られた指標については, Tukey HSD [26] を用いてどの条件間に有意差があるかを調べた. その結果を図 3 に角括弧で示す.

4.5.2 定性的結果

各質問票の結果を表 1 に示す. One-way ANOVA の結果, UEQ-S の Pragmatic Quality ($F_{2,51} = 3.815, p < 0.05$) と Overall ($F_{2,51} = 3.663, p < 0.05$) に有意差が見られた. Raw NASA-TLX ($F_{2,51} = 1.236, p = 0.299$), SUS ($F_{2,51} = 1.680, p = 0.197$),

表 1. 定性的結果 (# : $p < 0.1$, * : $p < 0.05$).

Condition	Raw NASA-TLX		SUS		UEQ-S Pragmatic Quality		UEQ-S Hedonic Quality		UEQ-S Overall	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Cond.F	57.13	13.70	63.61	17.46	0.50	1.41	1.04	1.49	0.77	1.12
Cond.S	56.48	12.01	70.00	14.88	* 1.65	1.16	1.81	1.10	* 1.73	1.01
Cond.K	62.92	13.65	60.00	15.88	# 0.64	1.40	1.26	1.28	0.95	1.16

UEQ-S の HedonicQuality($F_{2,51} = 1,555, p = 0.221$) には有意差が見られなかった。Tukey HSD の結果を表 1 に角括弧で示す。

4.6 参加者からのコメント

口頭インタビューにより各条件に対する参加者のコメントが得られた:

Follow The Pro (Cond.F): Cond.F の大きな利点は、コーチの動きを観察することで、ターンのタイミング (P1, P7, P12, P17) や旗を通過する際の適切な位置 (P2, P3) を参考にできることであった。特に、コーチの膝の傾け具合が参考になったというコメントがあった (P4, P8)。

一方、コーチが少し前方を走っているため、完全に動きをトレースすると旗を通過できないのが難しい (P4, P6, P14)、コーチと旗が重なってしまい見難い (P4, P5, P11)、特に Hard Level ではコーチの動きを観察する余裕がない (P3, P5, P10) といった意見があった。さらに、自分の姿勢に対するフィードバックがなく、姿勢改善には繋がらない (P12, P18) というコメントもあった。

Spine Angle Indicator (Cond.S): Cond.S の利点として最も多く挙げられたのは、実際の雪上トレーニングでは不可能な自身の姿勢をリアルタイムに確認できることであった (P1, P3, P6, P7, P11, P15, P16, P18)。その結果、上半身のぶれを軽減する (P1, P4, P5, P12) だけでなく、重心移動を最小限に抑えたコンパクトなターンが可能になった (P5)。また、ターンが左右非対称になっていることを認識できた (P13, P15, P18) というコメントもあった。

ただし、上半身の姿勢を意識する必要があるのは主に初心者であり、意図的に反対方向に姿勢を傾けることがある上級者には効果がないという指摘があった (P2, P8)。また、Hard Level では、旗を通過することだけに意識が向いてしまい、姿勢を気にすることが難しい参加者もいた (P13, P17)。

Keep Horizontal Pole (Cond.K): Cond.K の主な利点は、両腕を水平に伸ばした状態を維持することにより、上半身を垂直に保つことを意識できることであった (P1, P2, P8, P11, P12)。また、肩を地面と水平に保つことを意識するようになつた

(P5, P8, P13)。さらに、普段無意識に動かしてしまう腕の動きを視覚化することが効果的である (P10) というコメントもあった。

一方、HMD では視野が制限されたため、ポールを視界に入れ続けるのが難しいという指摘があった (P2, P3, P5)。さらに、コントローラを装着したことによりスキーストックが重くなったり (P6, P7, P13)、競技シーンでは旗を叩くように腕を振るため、そのためのトレーニングにはならない (P3, P5, P14) という問題が挙げられた。

4.7 考察

全体として、3 つの可視化フィードバック条件下でのトレーニング後の N.P.F はいずれもトレーニング開始時 (Cond.B) よりも増加しており (図 3)，一定のトレーニング効果が示されている。本節では、各条件の長所と改善点について述べる。

4.7.1 Follow The Pro

Cond.F はコーチに追従する一般的なトレーニング方法であり、新規性に乏しいため、参加者からの主観評価は高くなかった。一方で、Hard Level の N.P.F については、Cond.F が最も高く、旗を通過するという点で優れている。しかし、旗の通過を優先するあまり姿勢が崩れてしまい、M.S.A が悪化している。長時間トレーニングを行う場合、姿勢の悪さによる悪影響を発生する可能性が考えられる。

Cond.F では、上半身の姿勢のみに注目している他の 2 つの条件とは対照的に、コーチの全身姿勢を観察することができる。実際、Cond.F では下半身 (膝や脚) を意識していたとコメントした参加者が多く見られた。したがって、コーチの下半身の動きを参考し、ターンのタイミングを把握することをトレーニングの目的とする場合には、Cond.F が良いと考えられる。

4.7.2 Spine Angle Indicator

Cond.S は、3 条件の中で唯一 SUS スコアが平均である 68[3] を上回っており (表 1)，参加者の主観的評価が最も高い。また、実際に姿勢改善につながることが確認された (図 3 の M.S.A)。特に、普段はリ

アルタイムに確認することができない背骨を可視化することが高く評価されており、3つの条件の中でも最もARを活用できていると考えられる。

一方、より上級者向けには、別の部位を可視化すべきであるという意見があった。例えば、肩や太ももなども可視化することで、より高度なトレーニングが可能になる。ただし、画面上の情報量が多くなると、ユーザが処理しきれない可能性がある点に注意が必要である。例えば、姿勢の悪い部位だけを表示するなどの工夫を考えられる。

4.7.3 Keep Horizontal Pole

Cond.KもCond.Fと同様に、実際に用いられているトレーニング方法であるが、ポールの色が変化することにより問題点が認識しやすく、斬新であったと評価されている。また、腕だけでなく肩の姿勢の意識にもつながったというコメントがあり、一定の間接的な姿勢改善効果が見られた。ポールの表示位置を肩の位置に固定するなどの工夫により、ポールが見えにくいという課題を解決することができ、トレーニング効果の向上につながると考えられる。

Cond.Kでは、腕を高く水平に保ちながら旗を通過する必要があるため、N.P.Fが低く、Raw NASA-TLXのスコア(作業負荷)が高くなつたと考えられる。Cond.Kは、競技シーンで用いるのではなく、トレーニングの初期段階における正しい姿勢の意識づけとして使用するのが望ましいと考えられる。

4.8 仮説の検証

上記の結果より、本実験の仮説を検証する：

H1: 抽象的な姿勢可視化によるトレーニング効果

Cond.FとCond.Sを比較すると、Cond.Sの方が定量的かつ定性的に優れている(図3、表1)。一方で、Cond.FとCond.Kを比較すると、UEQ-SはCond.Kの方が高いものの、他の指標においてCond.KがCond.Fよりも優れているとは言えない。

ユーザの姿勢を抽象化して可視化することは有効であるが、その可視化方法には注意が必要である。したがって、H1は部分的に支持される。

H2: 間接的な姿勢可視化によるトレーニング効果

Cond.SとCond.Kを比較すると、Cond.Sの方が滑走時の姿勢が改善しており、参加者からの定性的評価も高くなっている(図3、表1)。

したがって、本実験の結果はH2を支持しない。しかし、4.7.3節で述べたように、より旗の速度が低く負荷が小さいタスクの場合や、姿勢の提示方法を変更した場合には間接的な姿勢可視化が効果的に働く可能性が考えられるため、詳細な検証が必要である。

5 今後の課題

第一に、本論文では既存の可視化手法との包括的な比較は行われていない。また、今回の評価実験で



図4. 実際の雪上にて行われたパイロットテストの様子。

はCond.SとCond.Kを別々に表示していたが、これら2つの可視化を同時に行うことが考えられる。

第二に、難易度の設定が適切であったかを検討する必要がある。評価実験では、Easy Level(時速60km)とHard Level(時速80km)の2つの難易度を設定しているが、中間のレベル(時速70km)を追加したり、より細かく難易度を分割することが考えられる。さらに、旗の通過率に応じて速度を動的に調整することで、最適な難易度でのトレーニングが可能になる。

第三に、今回の実験は短期的なトレーニング効果を検証するものであり、長期的な効果は検証されていない。特に、Cond.SとCond.Kはユーザに正しい姿勢を意識させることを目的としており、短期的には意識の改善による効果が表れない可能性がある。

6 実環境でのパイロットテスト

小規模なパイロットテストにより、提案システムが実環境にも適用できることを確認した(図4)。2人のスキーヤーが実際の雪上で提案システムを用いて仮想の旗を通過するターンを行うことができた。

しかし、HMDのトラッキングが屋内での使用に比べて不正確になる場合があった。現在のHMDの多くは屋内でのみの使用が想定されており、太陽光の影響を強く受けてしまうことが原因であると考えられる。ただし、LiDARセンサーなどを搭載した屋外でも使用可能なHMDの開発が進むことにより、これらの問題は解決できると考えられる。

7 まとめ

本論文では、滑走時の姿勢改善を目的とした、実環境のトレーニングとのギャップが小さいARスキートレーニングシステムを提案した。評価研究では、3つの視覚化フィードバック条件を比較した。その結果、ユーザの背骨を可視化する条件が定量的かつ定性的に優れていることが示された。最後に、提案システムが実環境でも適用できることを確認した。

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] V. Aleshin, V. Afanasiev, A. Bobkov, S. Klimenko, V. Kuliev, and D. Novgorodtsev. Visual 3D Perception of the Ski Course and Visibility Factors at Virtual Space. In *2011 International Conference on Cyberworlds*, pp. 222–226, 2011.
- [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. Bangor, P. Kortum, and J. Miller. Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. *J. Usability Studies*, 4(3):114–123, may 2009.
- [4] J. Brooke. SUS: A quick and dirty usability scale. *Usability Eval. Ind.*, 189, 11 1995.
- [5] I. Chatterjee, T. Pforde, A. Tng, F. Salemi Parizi, C. Chen, and S. Patel. ARDW: An Augmented Reality Workbench for Printed Circuit Board Debugging. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, UIST '22, New York, NY, USA, 2022. Association for Computing Machinery.
- [6] E. Dorzhieva, A. Baza, A. Gupta, A. Fedoseev, M. Cabrera, E. Karmanova, and D. Tsetserukou. DroneARchery: Human-Drone Interaction through Augmented Reality with Haptic Feedback and Multi-UAV Collision Avoidance Driven by Deep Reinforcement Learning. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 270–277, Los Alamitos, CA, USA, oct 2022. IEEE Computer Society.
- [7] S. Eom, D. Sykes, S. Rahimpour, and M. Gorlatova. NeuroLens: Augmented Reality-based Contextual Guidance through Surgical Tool Tracking in Neurosurgery. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 355–364, 2022.
- [8] K. Fan, J.-M. Seigneur, J. Guislain, S. Nanayakkara, and M. Inami. Augmented Winter Ski with AR HMD. In *Proceedings of the 7th Augmented Human International Conference 2016*, AH '16, New York, NY, USA, 2016. Association for Computing Machinery.
- [9] A. Fedosov, I. Elhart, E. Niforatos, A. North, and M. Langheinrich. SkiAR: Wearable Augmented Reality System for Sharing Personalized Content on Ski Resort Maps. In *Proceedings of the 7th Augmented Human International Conference 2016*, AH '16, New York, NY, USA, 2016. Association for Computing Machinery.
- [10] A.-L. Guinet, G. Bouyer, S. Otmane, and E. Desailly. Towards an AR game for walking rehabilitation: Preliminary study of the impact of augmented feedback modalities on walking speed. In *2020 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 264–268, 2020.
- [11] P.-H. Han, Y.-S. Chen, Y. Zhong, H.-L. Wang, and Y.-P. Hung. My Tai-Chi Coaches: An Augmented-Learning Tool for Practicing Tai-Chi Chuan. In *Proceedings of the 8th Augmented Human International Conference*, AH '17, New York, NY, USA, 2017. Association for Computing Machinery.
- [12] S. G. Hart. NASA-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 50(9):904–908, 2006.
- [13] 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.
- [14] S. Hasegawa, S. Ishijima, F. Kato, H. Mitake, and M. Sato. Realtime Sonification of the Center of Gravity for Skiing. In *Proceedings of the 3rd Augmented Human International Conference*, AH '12, New York, NY, USA, 2012. Association for Computing Machinery.
- [15] S. Henderson and S. Feiner. Exploring the Benefits of Augmented Reality Documentation for Maintenance and Repair. *IEEE Transactions on Visualization and Computer Graphics*, 17(10):1355–1368, 2011.
- [16] J. Hoffard, T. Nakamura, E. Wu, and H. Koike. PushToSki - An Indoor Ski Training System Using Haptic Feedback. In *ACM SIGGRAPH 2021 Posters*, SIGGRAPH '21, New York, NY, USA, 2021. Association for Computing Machinery.
- [17] H.-Y. Jo, L. Seidel, M. Pahud, M. Sinclair, and A. Bianchi. FlowAR: How Different Augmented Reality Visualizations of Online Fitness Videos Support Flow for At-Home Yoga Exercises. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, New York, NY, USA, 2023. Association for Computing Machinery.
- [18] 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.
- [19] 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.
- [20] T. Lin, R. Singh, Y. Yang, C. Nobre, J. Beyer, M. A. Smith, and H. Pfister. Towards an Under-

- standing of Situated AR Visualization for Basketball Free-Throw Training. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA, 2021. Association for Computing Machinery.
- [21] T. Matsumoto, E. Wu, and H. Koike. Skiing, Fast and Slow: Evaluation of Time Distortion for VR Ski Training. In *Augmented Humans 2022*, AHs 2022, p. 142–151, New York, NY, USA, 2022. Association for Computing Machinery.
- [22] B. Meyer, P. Gruppe, B. Cornelsen, T. C. Stratmann, U. Gruenefeld, and S. Boll. Juggling 4.0: Learning Complex Motor Skills with Augmented Reality Through the Example of Juggling. In *Adjunct Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*, UIST '18 Adjunct, p. 54–56, New York, NY, USA, 2018. Association for Computing Machinery.
- [23] F. Panizzolo, G. Marcolin, and N. Petrone. Comparative Evaluation of Two Skiing Simulators as Functional Training Devices for Recreational Skiers. *Journal of sports science & medicine*, 12:151–8, 03 2013.
- [24] M. Schrepp, A. Hinderks, and J. Thomaschewski. Applying the User Experience Questionnaire (UEQ) in Different Evaluation Scenarios. In A. Marcus ed., *Design, User Experience, and Usability. Theories, Methods, and Tools for Designing the User Experience*, pp. 383–392, Cham, 2014. Springer International Publishing.
- [25] M. Schrepp, A. Hinderks, and J. Thomaschewski. Design and Evaluation of a Short Version of the User Experience Questionnaire (UEQ-S). *International Journal of Interactive Multimedia and Artificial Intelligence*, 4:103, 01 2017.
- [26] J. Tukey. Comparing individual means in the analysis of variance. *Biometrics*, 5 2:99–114, 1949.
- [27] S. Webel, U. Bockholt, T. Engelke, N. Gavish, M. Olbrich, and C. Preusche. An augmented reality training platform for assembly and maintenance skills. *Robotics and Autonomous Systems*, 61(4):398–403, 2013. Models and Technologies for Multi-modal Skill Training.
- [28] 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.
- [29] 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.
- [30] 鈴木 淳登, 坂本 大介, 小野 哲雄. Gino .Aiki: 合氣道の身体の使い方の習得を支援する MR ソフトウェア. WISS 第 30 回 インタラクティブシステムとソフトウェアに関するワークショップ, pp. 46–55, 2022.

Smart Replay: e ラーニング動画における視覚的・時間的アクセシビリティの向上

SECHAYK Yotam * SHAMIR Ariel † 五十嵐 健夫 *

概要. e ラーニングは、教材への幅広いアクセスを可能にすることを目的としている。しかし、動画コンテンツを多用することは、アクセシビリティに大きな課題をもたらす。多様な参加者を対象とした予備的調査に基づき、e ラーニングのビデオコンテンツに存在するアクセシビリティの障壁を明らかにする。これらには、ユーザーの理解速度と動画速度の不一致、どこに注意を向けてよいかわからない視覚的複雑さ、ナビゲーションのしにくさなどが含まれる。また、我々の調査結果は、アクセシビリティの問題が、障害のある利用者とない利用者の両方に影響を及ぼすことを示している。さらに、既存のアクセシビリティツールには限界があり、さらなる対応が必要であることを示している。そこで我々は、e ラーニングアクセシビリティツールである「Smart Replay」を提案する。私たちのツールは、学習ビデオのビジュアルとコンテンツに基づいた分析を行い、アクセシブルな再生オプションを生成する。視覚的分野と時間的分野の両方を強化したビデオセクションの復習を可能にする。

1 はじめに

e ラーニングはインターネットを利用した教育で、世界中の学習者に利便性を提供している。近年、e ラーニング、特にビデオコンテンツによる教育が人気を博している。しかし、e ラーニングは、障がいのある利用者にとって、学習能力に影響を与えるという課題がある。静的メディアとは異なり、動画は情報を時間的に分散させ [31]、検索を困難にし [17]、双方向性を十分に活用できていない [14][3]。これは、学習可能性を妨げ、アクセス障壁を提示し、多様な学習者にとっての有効性を低下させる。

アクセシビリティ（バリアフリー）とは、どのような能力状態にある人々にとっても、障壁を減らし、困難を軽減することである。アクセシビリティ [29] および e ラーニング [13] 研究への関心が高まる中、ビデオベースの e ラーニングの領域は、依然として重大なアクセス障壁を抱えたままである。本研究の目的は、このようなアクセシビリティの課題を解明し、改善策として新しいアクセシビリティツールを紹介し、この領域における今後の試みを促すことである。まず、 $N = 53$ の参加者を対象としたウェブ調査から始める。年齢、興味、能力状態（感覚、身体、認知など）の多様性を含む、一般的なアクセス障壁とアクセシビリティ・ツールの利用パターンを精査する。この包括的な分析により、あらゆるタイプの学習者が遭遇する一般的な課題について、貴重な洞察を得ることができる。

その結果に基づき、動画学習のためのアクセシビ

リティツールである Smart Replay (SR) を提案する。SR は、ビデオ再生のアクセシビリティを向上させる要素を統合するために、視覚的領域と文脈的領域の両方でビデオ分析を採用している。このツールは、(1) どこに注意を向ければよいかわからないという問題、(2) 複雑なコンテンツナビゲーション、(3) 動作再生速度と利用者の理解速度の不一致をターゲットとしている。ウェブ・インターフェースを介して、ユーザーはアクセシビリティを向上させた e ラーニング・ビデオを視聴、ナビゲート、再生することができる。

2 関連研究

e ラーニングは、教育のためにインターネットを使用し、世界中の学習者にとって大きな利便性を提供している [30]。特に Covid-19 パンデミック以降、e ラーニングの研究は成長している。以前の研究では、異なるプラットフォームの利用とその影響についての調査が行われている [35][23][43]、また、e ラーニングの効果を評価する方法やアプローチについても研究が行われている [16][2][25]。さらに、異なるデザインのパラダイムに関する研究も行われており [28][45][24][12][1][10]、e ラーニングアクセシビリティについての研究も進行中である [7][47]。

e ラーニングアクセシビリティの研究では、ユニバーサルデザインの活用や特定の環境におけるアクセシビリティの分析が行われている [39][40][42]。また、新たなアクセシブルガイド生成技術に関する研究も行われている [22]。既存の研究とは異なり、私たちの研究では、さまざまな能力を持つ人々と共有されるアクセシビリティニーズを特定し、取り組ん

Copyright is held by the author(s).

* 東京大学

† Reichman University

でいる。

アクセシビリティ研究は進展しているが[29]、まだ多くの課題が残されている。過去の研究では、ビデオのアクセシビリティ評価や、よりアクセス可能なビデオを生成するツールの提供に焦点が当てられている[27][26][15][32][48]。静的な視覚フィルターや視線ベースのズーミングといった手法も探求されている[41][6]。また、最近の機械学習の進歩により、ビデオの説明やコンテンツの簡素化が改善されている[49][44]。

一般のビデオと異なり、e ラーニングビデオは固有の課題を提供している[18]。e ラーニングの中核的なニーズは、学習可能性（理解）の促進である。これまでの研究では、ゲーミフィケーションの利点やコンテンツの説明[50][21][51][19][44]、スライドのアクセシビリティ分析[38]、黒板のコンテンツ理解[11]、スクリーンリーダーに対応したスライド[36][37]、そして e ラーニングビデオのナビゲーションに関する研究が行われている[31]。

しかし、時間的な柔軟性と認知負荷の考慮は未だに解決されていない課題である。これまでの研究は、e ラーニングビデオをスクリーンリーダーなどの特定のアクセシビリティツールに統合することに焦点を当ててきたが、広範なアクセシビリティのニーズを見逃している可能性がある。私たちの研究では、既存の障壁と核心的な問題を特定し、それらに取り組むための基盤を築くことを目指している。

3 予備的調査

現在のアクセス障壁を理解するために、私たちは予備的調査を行った。障害のある利用者もない利用者も募集した。テルアビブ大学のバリアフリー支援室による参加者募集と、ソーシャルネットワークグループへの投稿を通じて募集を行った。10 個の質問からなるウェブアンケートを実施し、すべての質問は任意である。そのため、回答数が少ない質問もある(< 53)。主な対象は、(1)e ラーニングの利用状況、(2) 学習ビデオのユーザー・エクスペリエンス、(3) 既存のツールと未対応の問題である。私たちの知る限り、これは混在する参加者を対象にアクセスの障壁を調査した最初の研究である。また、アクセシビリティ・ツールが、障害の有無にかかわらず、個人にどのようなプラスの影響を与えることができるかを、より詳細に明らかにした。

参加者の年齢はさまざまで、26~35 歳が 27 人(51 %)、18~25 歳が 15 人(28 %)、46~60 歳が 7 人(13 %)、36~45 歳と 61 歳以上がそれぞれ 2 人(4 %) だった。性別や学歴に関する情報は収集していない。能力の多様性（すなわち、障害の状態）も、本研究の重要なポイントである（表 1）。先行研究では、特定のユーザー・グループ[39][8][4]に焦点を当てていたが、我々は、アクセシビリティ・

ニーズの共通性を理解するために、多様性を重視している。

表 1. 障害状態の概要、複数選択可。合計：いくつ挙げたか。排他的：排他的に言及した数。52 の回答がある選択問題。

状態	合計	排他的
正常な視力への修正	12	6
視覚障害または低視力	11	8
色覚異常	1	1
聴覚障害または難聴	11	8
認知障害	3	2
ディスレクシア	2	0
ADHD (注意欠陥多動性障害)	11	5
身体的障害	7	5
なし	7	7

Part 1: e ラーニングの利用状況 (Q1-Q6)。参加者はオンライン学習リソースを利用する傾向が高い（図 1）。「読書」は 49 名、「ビデオ（オンデマンド）」は 45 名、「オンライン授業（ライブ）」は 37 名が利用している。上記のすべてを利用しているのは合計 30 人である。さらに、43 人がインストラクションビデオ（ハウツー、ガイドなど）を、34 人が講義録を、28 人がビデオエッセイを視聴し、17 人がすべてを利用している。

よく視聴されるコンテンツの一般的なタイプには、専門的なトピック (38)、料理 (24)、数学 (18)、プログラミング (14) などが含まれる。人気のあるプラットフォームには、YouTube¹ (47)、Khan Academy² (11)、Coursera³ (11) などがある。

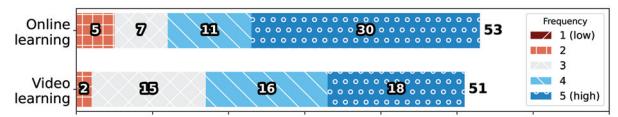


図 1. オンライン学習コンテンツ、特にビデオを利用する可能性。

Part 2: 学習ビデオのユーザー・エクスペリエンス (Q7-Q8)。参加者はスマートフォン (9 人) よりも PC (42 人) を非常に好む。また、能力の有無に関わらず、多くの困難を経験している。我々は、これらを時間的、視覚的、音声的な困難と区別している（表 2）。非障害者 (7 名) 全員が、学習ビデオを視聴する際に多くの困難に直面すると述べている。

¹ <https://www.youtube.com/>

² <https://www.khanacademy.org/>

³ <https://www.coursera.org/>

Smart Replay

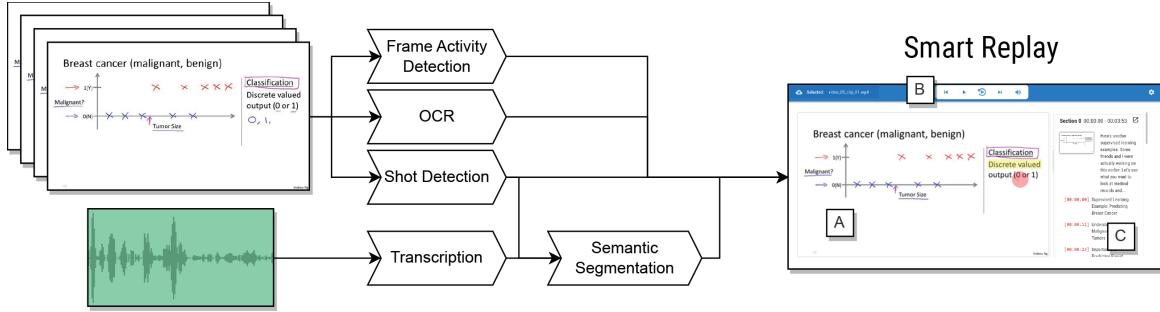


図 2. Smart Replay のシステム概要図. 映像の分析, 特徴の抽出, アクセス可能な再生のオンライン計算. スマート・リプレイ : (A) ビデオ表示, (B) 再生コントロール, (C) チャプター選択と概要.

表 2. 困難. (T.) 時間的, (V.) 視覚的, (A.) 音声的.
(* = 自由記述).

困難	合計
T. Speech can be too fast	26
Reading on-screen text in time	14
The pace is too fast (or slow)	10
V. Pause frequently to see better	21
On-screen text is too small	17
Noticing small on-screen details	11
Bright colors are overwhelming	9
Distinguishing between colors	7
Unsure about attention placement	7
Unnoticeable non-central elements	5
Unreliable audio descriptions*	2
Visual overload*	1
A. Hard to hear the video clearly	24
Unreliable or missing captions	20

Part 3: 既存のツールと未対応の問題である (Q9-Q10). 参加者が使用しているツールはごく少数である. 36人がビデオ速度の変更を, 27人がキャプションを, 20人がノイズ低減ヘッドフォンを, 8人がスクリーン・リーダー・ソフトウェアを使用している. その他, 画面拡大鏡, カラーフィルター, 視線追跡カーソルなどが挙げられている. さらに,多くのアクセシビリティ・ニーズが, 参加者から依然として提起されている. 音声やオーディオの強化, カスタマイズによる信頼性の高い書き起こし, スクリーン・リーダーのサポート不足, より簡単なナビゲーション, 双方向性などである. 現在, 研究は進んでいるものの, 利用可能なツールは限られており, 存在するツールに対する認識も不足している.

4 調査結果

学習者の能力に関係なく, アクセスの障壁は存在する. 現在のツールの状況には多くの不満が残る. また, ユーザによるカスタマイズの有用性は[20][9]で議論されているが, アクセシビリティに関するカスタマイズの可能性については, さらなる調査が必要である. さらに, 既存の研究にもかかわらず, 多くのアクセシビリティの側面が未解明のままである.

我々の発見に基づき, ユーザーが直面する3つの重要な問題を区別し, これらに対処するための新しいアクセシビリティ・ツールを提案する.

P1 (ついていくのが難しい) ビデオの時間的な再生速度がユーザーの理解速度と一致していない.

P2 (見づらい) 多すぎる視覚的要素が注意を誤導する可能性がある.

P3 (ナビゲーションが難しい) ビデオ内での検索は複雑で, 多くのユーザーアクションに依存している.

5 Smart Replay

このような問題を解決するために, 我々はeラーニングビデオのアクセシビリティツールである *Smart Replay* (SR) を提案する. システム概要図を図に示す(図2). スライド, ナレーション, ライブドローイングを組み込んだ一般的なビデオのスタイル[46][5]に焦点を当てている(図3).

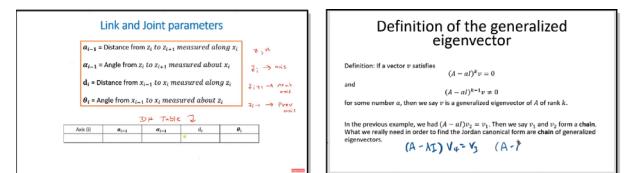


図 3. 対応コンテンツ例のフレーム.

映像解析により、複雑な映像部分でも利用しやすい再生機能を合成する。音声と視覚情報（図2）を利用し、既存の方法を用いて、光学的文書認識（OCR）、文字起こし、ショットセグメント情報を抽出する。

さらに、セマンティックセグメンテーション、OCR-音声相互参照、フレーム活性検出を導入する。後者は、eラーニングビデオのための新しいアプローチである。抽出された情報は再生システムに供給され、学習者は文脈上のセグメントを簡単にナビゲートしたり、アクセシビリティを強化して再生したり、学習体験をカスタマイズしたりするツールを得ることができる。

5.1 システムの特徴

ここでは、各機能の詳細とその問題点（P1～P3）について説明する。すべての機能にはカスタマイズが含まれる（図4-B）。ユーザーが体験をパーソナライズできるようにする。

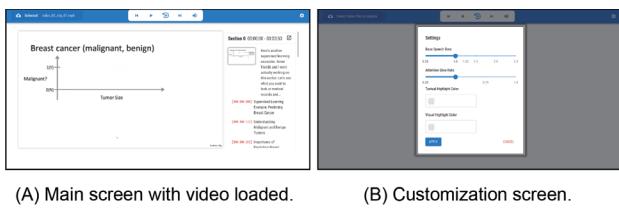


図 4. Smart Replay のウェブアプリケーション画面。

アクセシブル・リプレイ（P1）。 手動ナビゲーションを省略するリプレイボタン。アクセシビリティを強化したコンテンツセグメントを再生する。

時間調整（P1）。 認知負荷（発話速度、ハイライトなど）に応じて自動的に変化する、柔軟なビデオ速度。

OCR-音声クロスハイライト（P2）。 音声に基づいて画面上のテキストをハイライトする。学習者の注意を誘導し、フォローを助ける。

フレーム・アクティビティ・ハイライト（P2）。 フレームアクティビティとは、ポインタの動きやスライド上の描画などのアクションを指す。これらはハイライトされ、ユーザーがどこに注意を向ければよいかを簡単に示すことができる。

セマンティックセグメンテーション（P3）。 ビデオコンテンツからの意味情報に基づいた、きめ細かなセグメンテーション。ユーザーは簡単にコンテンツを選択し、ナビゲートし、概観することができる。

5.2 インタラクションの概要

図4-Aはメイン・ウェブ・インターフェースである。左側がビデオセクション、右側がセグメント化されたチャプターである。各メイン・セグメント（ショット）には、それぞれタイトルで定義されたセマンティック・セグメントのリストがある。ユーザーは興味のあるセグメントを見つけ、ビデオを簡単にナビゲートすることができる。

上部には、コントロールセクションがある。ユーザーは再生/一時停止▶、前の◀または次の▶セグメント、音量➡、SRボタン➡にアクセスできる。リプレイボタンを押すと、現在のセグメントが最初から再生され、ビデオにオーバーレイされた様々なアクセシビリティ機能が有効になる。

5.3 インタラクションの例

視覚障害を持つコンピュータ科学専攻のアリスの視点から、このシステムを動作を説明する。アリスは録画されたC++の講義を見ている。セマンティック・セグメンテーションによって、アリスは各章の概要を知り、簡単にナビゲートできる。

再生中、講師はスライドに絵を描いている。聴きながら講師が描いている場所を特定することは、アリスにとって難しい作業である。アリスの視覚障害は、このセクションを理解する能力を妨げ、認知的負荷を増加させる。アリスがSRボタンを押すと、そのセクションが再生される。大きな視覚的ハイライトがどこを見るべきかを示し、スピーチはゆっくりになり、アリスは要求される認知的負荷を減らすことができる。

後の章では、多くの文字情報を含むスライドが提示される。これはアリスに要求される認知的負荷を増加させる。アリスはもう一度SRボタンを押す。今度は、講師が話すとテキストの一部が強調表示され、アリスは楽についていける。

6 実装

概要に続いて、システム機能がどのように実装されているかについて説明する。フレームアクティビティ、セマンティックセグメンテーション、OCR-音声クロスハイライトに重点を置く。YouTubeから取得したビデオコンテンツを使用する[33]。

6.1 アクセシブル・リプレイと時間調整

SR（再生速度）機能は、ユーザーがアクティブなビデオセグメントを再生し直す。現在の再生時刻とセグメントの開始時刻との間隔がある閾値 ($th_{SR} = 2sec$) より小さい場合、リプレイは前のセグメントの開始時刻に移動する。

これにより、ユーザーの認知負荷を軽減するため、時間の柔軟性を組み込んでいる。再生中、ビデオの

時間経過速度が変わる。システムによって導入された視覚的ハイライトに関しては、ユーザー設定に基づいてビデオをさらに一時的に遅く再生する（デフォルトは $0.75 \cdot V_{speed}$ ）。

さらに、すべてのセンテンスに対して、所望の1秒あたりの単語数（WPS）値 (U^{wps})に基づいて再生速度を調整する。したがって、文字起こし行 T_n に対して次のようになる：

$$T_n^{wps} = \frac{|T_n^{words}|}{T_n^{end} - T_n^{start}} \quad (1)$$

ここで、 T_n^{end} および T_n^{start} はそれぞれ終了時刻と開始時刻である。したがって、ビデオ再生速度 V_{speed} は次のようになる：

$$V_{speed} = \frac{U^{wps}}{T_n^{wps}} \quad (2)$$

さらに、非音声の長い箇所を高速化することができる。時間の調整を行うことで、学習者が追従し、集中力を保つ能力が向上すると期待できる。

6.2 OCR-音声クロスハイライト

スライドの関連部分にユーザーの注意を誘導する。まず、OCR のショット候補を収集するために低感度で `scenedetect` を実行する。次に、各候補ショットからテキストを抽出する (*EasyOCR*⁴ を使用)。同一のテキストを持つ候補を統合する (図 5)。

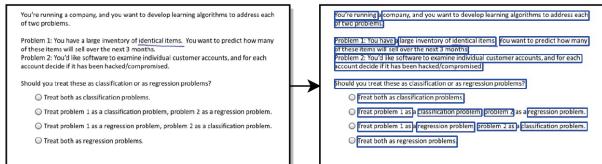


図 5. OCR 検出結果の例。

ビデオ再生中、トランスクリプションと適切なOCR要素をクロスリファレンスする(図 6-A)。各OCRフラグメントについて、フラグメント fr_n とトランスクリプション行 T_n の間で共通の最も長い連続単語 ($W_n^{T,fr}$)を見つける。つまり、 $|W_n^{T,fr}| \geq 2$ 。たとえば、トランスクリプションが "[...] try to predict the breast cancer [...]" であり、OCR フラグメントが "Breast cancer (malignant, benign)" の場合:

$$W_n^{T,fr} = ["breast", "cancer"] \quad (3)$$

その後、以下の手順に従う:

1. $|W_n^{T,fr}|$ が最大化される fr を検索する。
2. OCR フラグメントがフレーム $n-1$ に選択された場合、無視する。

⁴ <https://github.com/jaidedai/easyocr>

残りのOCRフラグメントは、その境界ボックスに基づいてハイライト表示される。

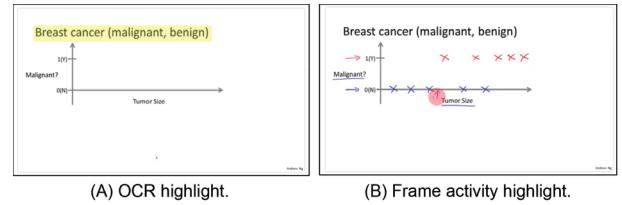


図 6. 異なる画面上のレンダリングハイライト、アクティビティ、および音声-OCR。

6.3 フレーム・アクティビティ・ハイライト

ビデオフレームで発生しているアクティビティを、2つのフレーム間の比較によって検出する。アクティビティとは、(1) マウス/ポインタの動き、(2) スライド上での描画、(3) 同じスライド上で新しいコンテンツの表示を指す (図 6-B)。アルゴリズムの概要は図 7 に示されている。

ビデオフレームのウィンドウ F_w 上で作業する。ここで $|F_w| = V_{fps}$ (frames-per-second) である。最初と最後のフレームを比較する。OpenCV⁵の `absdiff` を使用してフレームの差分をチェックし、輪郭計算を適用する。各フレーム n の輪郭 i (つまり、 $Ct_{n,i}$ は、次の簡略化された定義に変換される:

$$Ct_{n,i} = (A_{n,i}, P_{n,i}) \quad (4)$$

ここで、 $A_{n,i}$ はそれがカバーする面積であり、 $P_{n,i}$ は中心点である。これを用いて、中心から中心への距離に基づいて、見つかった輪郭をクラスタにグループ化するために連結コンポーネント (DFS) を使用する。すべての輪郭をノードとする完全に連結したグラフから始める。そして、中心点間の距離が与えられた閾値 $th_{RoI} = 0.1 \cdot \min(F_w, F_h)$ より大きい場合、エッジを反復的に削除する。ここで、 F_w 、 F_h はそれぞれフレームの幅と高さである。残った連結コンポーネントが注目領域 (RoI) である。このような RoI は次のように定義される:

$$R_{n,j} = (A_{n,j}, P_{n,j}, N_{n,j}) \quad (5)$$

ここで、 $N_{n,j}$ はそれに含まれる輪郭の数、 $P_{n,j}$ は輪郭の中心点、 $A_{n,j}$ はその領域の面積で、 $A_{n,j} = \sum_p A_{n,p}$ である、 p はその領域内の輪郭である。

ビデオ再生中、現在の再生時刻を基にすべての関連するフレームアクティビティ RoI を収集する。これらの RoI に対して、いくつかのアクションを実行する:

1. $\#(R_n) > th_R$ の場合、ハイライトをスキップする。ここで $th_R = 8$ である。

⁵ <https://opencv.org/>

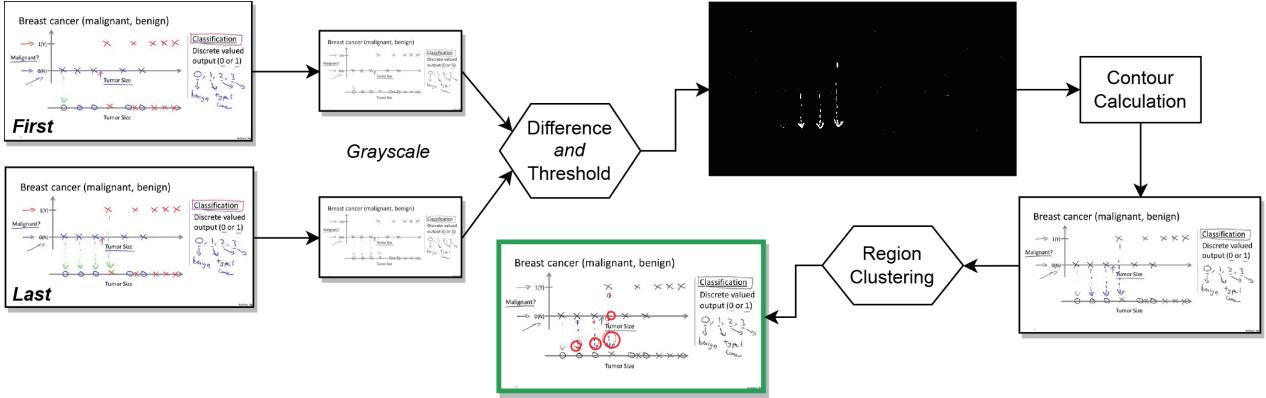


図 7. フレームアクティビティ検出アルゴリズムの概要。フレームのウィンドウで動作し、そのウィンドウ内の最初のフレームと最後のフレームを比較する。

2. 面積 ($A_{n,j}$) に基づいて RoI をソートする。
3. $A_{n,j} > 0.25 \cdot \max(F_w, F_h)$ または $A_{n,j} < 0.01 \cdot \min(F_w, F_h)$ の場合、RoI を無視する。
4. 以下のいずれかが前回のハイライトに対して正当化されているか確認する：
 - (a) 十分な時間が経過している $\Delta t \geq 10sec$.
 - (b) 中心点間の距離が大きい $P_{n,j} - P_{m,k} > 0.1 \cdot \min(F_w, F_h)$.

残りの RoI は、ハイライト表示され、その直径は $\sqrt{A_{n,j}} + A_b$ である。ここで A_b はユーザーによって設定された基本面積値である（デフォルトは 20px）。

6.4 セマンティックセグメンテーション

まず、スライドのナビゲーションであるショット変更を *scenedetect*⁶ Python パッケージを使用して検出する。それから、OpenAI Whisper⁷を使って文字起こしを行っている。意味的セグメンテーションアルゴリズムは、各ショット S_i と文字起こし行 T_j に対して、 $T_j \in S_i$ であるかどうかを計算する。次に、5つの異なるケースがある：

- C1 $T_j^{start} \geq S_i^{start}$ かつ $T_j^{end} \leq S_i^{end}$
- C2 $T_j^{end} \leq S_i^{start}$
- C3 $T_j^{start} \geq S_i^{end}$
- C4 $T_j^{start} < S_i^{start}$ かつ $S_i^{start} < T_j^{end} \leq S_i^{end}$
- C5 $T_j^{end} > S_i^{end}$ かつ $S_i^{start} \leq T_j^{start} < S_i^{end}$

C1 はショット内に完全に含まれているため、それに属する。C2 と C3 はショットの境界の外に完全にある文字起こしに対応する。C4 と C5 は、複数の

ショットにオーバーラップする中間の文字起こしに対応する。C4 と C5 に対して、比率の閾値として $th_{ratio} = 0.5$ を導入する。C4 の検証：

$$abs\left(\frac{T_j^{start} - S_i^{start}}{T_j^{end} - T_j^{start}}\right) < th_{ratio} \quad (6)$$

C5 の検証：

$$abs\left(\frac{T_j^{end} - S_i^{end}}{T_j^{end} - T_j^{start}}\right) < th_{ratio} \quad (7)$$

各ショットセグメントごとに収集された文字起こし行を使用し、GPT-3.5[34]を活用して生成されたショットの文字起こしを文脈のある段落に分割する。それぞれにユーザーに提示される表題が付いている。

7 結論

私たちは、さまざまな能力を持つ多様なグループを対象に、e ラーニングに関する広範な調査を実施した。その結果、3つの重要な課題が明らかになった。動画は(1)フォローしにくい、(2)見づらい、(3)ナビゲートしにくい。残念なことに、既存のツールはこれらの課題に十分に対処できていない。過去の研究では、これらの課題の側面について研究されているが、認知的負荷や柔軟な時間の使い方といった本質的な考慮は、顕著に欠けていた。

「Smart Replay」という革新的なソリューションを提案した。これは、e ラーニングビデオを詳細に分析し、フレームアクティビティの拡張や OCR スピーチの統合などの機能を追加したものである。セマンティックセグメンテーションとリプレイ機能によって、コンテンツ内での移動が容易になり、改善されたナビゲーションが実現される。時間的な柔軟性とユーザーのカスタマイズを組み込むことで、認知負荷のダイナミクスにより適合したインターフェイスで個人化された学習環境を作成することを目指している。

⁶ <https://www.scenedetect.com/>

⁷ <https://openai.com/research/whisper>

謝辞

本研究は *Google* の一部支援を受けて行われた。
その経済的なサポートに感謝の意を表す。

参考文献

- [1] G. E. Acuna, L. A. Alvarez, J. Miraflores, and M. J. C. Samonte. Towards the Development of an Adaptive E-learning System with Chatbot Using Personalized E-learning Model. In *ICFET 2021: The 7th International Conference on Frontiers of Educational Technologies, Bangkok, Thailand, June 4 - 7, 2021*, pp. 120–125. ACM, 2021.
- [2] M. Agatha Priska, D. Aulia, E. Muslim, and L. Marcelina. Developing a framework to evaluate e-learning system at higher education in Indonesia. In *2020 The 4th International Conference on Education and E-Learning*, pp. 27–32, 2020.
- [3] M. Akram, M. W. Iqbal, M. U. Ashraf, E. Arif, K. Alsubhi, and H. M. Aljahdali. Optimization of Interactive Videos Empowered the Experience of Learning Management System. *Comput. Syst. Eng.*, 46(1):1021–1038, 2023.
- [4] W. Aljedaani, R. Krasniqi, S. Aljedaani, M. W. Mkaouer, S. Ludi, and K. Al-Raddah. If online learning works for you, what about deaf students? Emerging challenges of online learning for deaf and hearing-impaired students during COVID-19: a literature review. *Universal access in the information society*, pp. 1–20, 2022.
- [5] D. An. Introduction to Jordan Canonical Form, 2021. YouTube video.
- [6] A. S. Aydin, S. Feiz, V. Ashok, and I. V. Ramakrishnan. Towards making videos accessible for low vision screen magnifier users. In F. Paternò, N. Oliver, C. Conati, L. D. Spano, and N. Tintarev eds., *IUI '20: 25th International Conference on Intelligent User Interfaces, Cagliari, Italy, March 17-20, 2020*, pp. 10–21. ACM, 2020.
- [7] J. D. Basham, J. Blackorby, and M. T. Marino. Opportunity in crisis: The role of universal design for learning in educational redesign. *Learning Disabilities: A Contemporary Journal*, 18(1):71–91, 2020.
- [8] C. Batanero-Ochaíta, L. de Marcos, L. F. Rivera, J. Holvikivi, J. R. Hilera, and S. O. Tortosa. Improving Accessibility in Online Education: Comparative Analysis of Attitudes of Blind and Deaf Students Toward an Adapted Learning Platform. *IEEE Access*, 9:99968–99982, 2021.
- [9] X. A. Chen, C. Wu, T. Niu, W. Liu, and C. Xiong. Marvista: A Human-AI Collaborative Reading Tool. *CoRR*, abs/2207.08401, 2022.
- [10] N. A. Dahan, M. Al-Razgan, A. Al-Laith, M. A. Alsoufi, M. S. Al-Asaly, and T. Alfakih. Metaverse framework: A case study on E-learning environment (ELEM). *Electronics*, 11(10):1616, 2022.
- [11] K. Davila, F. Xu, S. Setlur, and V. Govindaraju. FCN-LectureNet: Extractive Summarization of Whiteboard and Chalkboard Lecture Videos. *IEEE Access*, 9:104469–104484, 2021.
- [12] H. Deng. Strategies to Improve the Online-Teaching Effect of Colleges and Universities in the Post-Epidemic Era. In *CIPAE 2021: 2nd International Conference on Computers, Information Processing and Advanced Education, Ottawa, ON, Canada, 25–27 May, 2021*, pp. 786–789. ACM, 2021.
- [13] E. Djekic, J. Dégila, C. Bondiombouy, and M. H. Alhassan. E-learning bibliometric analysis from 2015 to 2020. *Journal of Computers in Education*, 9:727–754, 12 2022.
- [14] H. Haga. Combining video and bulletin board systems in distance education systems. *The Internet and Higher Education*, 5(2):119–129, 2002.
- [15] T. Han, M. Bain, A. Nagrani, G. Varol, W. Xie, and A. Zisserman. AutoAD: Movie Description in Context. *CoRR*, abs/2303.16899, 2023.
- [16] N. Harrati, I. Bouchrika, A. Tari, and A. Ladjalia. Exploring user satisfaction for e-learning systems via usage-based metrics and system usability scale analysis. *Comput. Hum. Behav.*, 61:463–471, 2016.
- [17] L. He, E. Sanocki, A. Gupta, and J. Grudin. Comparing presentation summaries: slides vs. reading vs. listening. In T. Turner and G. Szwillus eds., *Proceedings of the CHI 2000 Conference on Human factors in computing systems, The Hague, The Netherlands, April 1–6, 2000*, pp. 177–184. ACM, 2000.
- [18] K. Hornbæk, D. T. Engberg, and J. Gomme. Video lectures: HCI and e-learning challenges. In *Workshop on Human-computer interaction and E-learning*, 2002.
- [19] W. Huang, T. Xiao, Q. Liu, Z. Huang, J. Ma, and E. Chen. HMNet: a hierarchical multimodal network for educational video concept prediction. *International Journal of Machine Learning and Cybernetics*, pp. 1–12, 2023.
- [20] V. Ilin. The role of user preferences in engagement with online learning. *E-Learning and Digital Media*, 19(2):189–208, 2022.
- [21] A. Khaldi, R. Bouzidi, and F. Nader. Gamification of e-learning in higher education: a systematic literature review. *Smart Learn. Environ.*, 10(1):10, 2023.
- [22] J. Kong, D. Sabha, J. P. Bigham, A. Pavel, and A. Guo. TutorialLens: Authoring Interactive Augmented Reality Tutorials Through Narration and Demonstration. In F. R. Ortega, R. J. Teather, G. Bruder, T. Piomsomboon, B. Weyers, A. U. Batmaz, K. Johnsen, and C. W. Borst eds., *SUI '21: Symposium on Spatial User Interaction, Virtual Event, USA, November 9–10, 2021*, pp. 16:1–16:11. ACM, 2021.

- [23] A. Krouská, C. Troussas, and C. Sgouropoulou. Usability and Educational Affordance of Web 2.0 tools from Teachers' Perspectives. In N. N. Karanikolas, A. Voulodimos, C. Sgouropoulou, M. Nikolaïdou, and S. Gritzalis eds., *PCI 2020: 24th Pan-Hellenic Conference on Informatics, Athens, Greece, 20-22 November, 2020*, pp. 107–110. ACM, 2020.
- [24] S. N. Kurbakova, Z. N. Volkova, and A. V. Kurbakov. Developing Students' Cognitive Abilities in E-Learning Environment. In *IC4E 2021: 12th International Conference on E-Education, E-Business, E-Management, and E-Learning, Tokyo Japan, 10 - 13 January, 2021*, pp. 124–130. ACM, 2021.
- [25] Z. Li and Z. Zhan. Integrated infrared imaging techniques and multi-model information via convolution neural network for learning engagement evaluation. *Infrared Physics & Technology*, 109:103430, 2020.
- [26] X. B. Liu, R. Wang, D. Li, X. A. Chen, and A. Pavel. CrossA11y: Identifying Video Accessibility Issues via Cross-modal Grounding. In M. Agrawala, J. O. Wobbrock, E. Adar, and V. Setlur eds., *The 35th Annual ACM Symposium on User Interface Software and Technology, UIST 2022, Bend, OR, USA, 29 October 2022 - 2 November 2022*, pp. 43:1–43:14. ACM, 2022.
- [27] X. Liu, P. Carrington, X. A. Chen, and A. Pavel. What Makes Videos Accessible to Blind and Visually Impaired People? In Y. Kitamura, A. Quigley, K. Isbister, T. Igarashi, P. Bjørn, and S. M. Drucker eds., *CHI '21: CHI Conference on Human Factors in Computing Systems, Virtual Event / Yokohama, Japan, May 8-13, 2021*, pp. 272:1–272:14. ACM, 2021.
- [28] Y. Liu. Design and Application of Hybrid Teaching Platform Based on Internet +. In *CIPAE 2021: 2nd International Conference on Computers, Information Processing and Advanced Education, Ottawa, ON, Canada, 25-27 May, 2021*, pp. 1490–1493. ACM, 2021.
- [29] K. Mack, E. McDonnell, D. Jain, L. L. Wang, J. E. Froehlich, and L. Findlater. What Do We Mean by "Accessibility Research"?: A Literature Survey of Accessibility Papers in CHI and ASSETS from 1994 to 2019. In Y. Kitamura, A. Quigley, K. Isbister, T. Igarashi, P. Bjørn, and S. M. Drucker eds., *CHI '21: CHI Conference on Human Factors in Computing Systems, Virtual Event / Yokohama, Japan, May 8-13, 2021*, pp. 371:1–371:18. ACM, 2021.
- [30] I. A. Mastan, D. I. Sensuse, R. R. Suryono, and K. Kautsarina. Evaluation of distance learning system (e-learning): a systematic literature review. *Jurnal Teknoinfo*, 16(1):132–137, 2022.
- [31] G. N. V. P, and P. S. Interactive Audio Indexing and Speech Recognition based Navigation Assist Tool for Tutoring Videos. In *2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS)*, pp. 1678–1682, 2022.
- [32] R. Natalie, E. Jarjue, H. Kacorri, and K. Hara. ViScene: A Collaborative Authoring Tool for Scene Descriptions in Videos. In T. J. Guerreiro, H. Nicolau, and K. Moffatt eds., *ASSETS '20: The 22nd International ACM SIGACCESS Conference on Computers and Accessibility, Virtual Event, Greece, October 26-28, 2020*, pp. 87:1–87:4. ACM, 2020.
- [33] A. Ng. Introduction to Machine Learning — ML-005 Lecture 1 — Stanford University — Andrew Ng, 2017. YouTube video.
- [34] OpenAI. OpenAI: Introducing ChatGPT, 2022. (accessed on July 3, 2023).
- [35] D. Pal and V. Vanijja. Perceived usability evaluation of Microsoft Teams as an online learning platform during COVID-19 using system usability scale and technology acceptance model in India. *Children and youth services review*, 119:105535, 2020.
- [36] Y. Peng, J. P. Bigham, and A. Pavel. Slidecho: Flexible Non-Visual Exploration of Presentation Videos. In J. Lazar, J. H. Feng, and F. Hwang eds., *ASSETS '21: The 23rd International ACM SIGACCESS Conference on Computers and Accessibility, Virtual Event, USA, October 18-22, 2021*, pp. 24:1–24:12. ACM, 2021.
- [37] Y. Peng, P. Chi, A. Kannan, M. R. Morris, and I. Essa. Slide Gestalt: Automatic Structure Extraction in Slide Decks for Non-Visual Access. In A. Schmidt, K. Väänänen, T. Goyal, P. O. Kristensson, A. Peters, S. Mueller, J. R. Williamson, and M. L. Wilson eds., *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, CHI 2023, Hamburg, Germany, April 23-28, 2023*, pp. 829:1–829:14. ACM, 2023.
- [38] Y. Peng, J. Jang, J. P. Bigham, and A. Pavel. Say It All: Feedback for Improving Non-Visual Presentation Accessibility. In Y. Kitamura, A. Quigley, K. Isbister, T. Igarashi, P. Bjørn, and S. M. Drucker eds., *CHI '21: CHI Conference on Human Factors in Computing Systems, Virtual Event / Yokohama, Japan, May 8-13, 2021*, pp. 276:1–276:12. ACM, 2021.
- [39] G. Ravichandran, D. J. Sujathamalini, and D. K. Gunasekaran. E-Learning-Accessibility of Students with Visual Impairment in Higher Education. *International Journal of Research and Review*, 9(5):27–31, 2022.
- [40] S. Russ and F. Hamidi. Online learning accessibility during the COVID-19 pandemic. In S. R. Vázquez, T. Drake, D. Ahmetovic, and V. Yaneva eds., *W4A '21: 18th Web for All Conference, Virtual Event / Ljubljana, Slovenia, April 19-20, 2021*, pp. 8:1–8:7. ACM, 2021.
- [41] A. Sackl, F. Graf, R. Schatz, and M. Tschelegi. Ensuring Accessibility: Individual Video Playback Enhancements for Low Vision Users. In T. J. Guerreiro, H. Nicolau, and K. Moffatt eds., *ASSETS '20: The 22nd International ACM SIGACCESS Conference on Computers and Accessibility, Virtual Event, Greece, October 26-28, 2020*, pp. 67:1–67:4. ACM, 2020.

- [42] C. Smith. Challenges and opportunities for teaching students with disabilities during the COVID-19 pandemic. *International Journal of Multidisciplinary Perspectives in Higher Education*, 5(1):167–173, 2020.
- [43] B. Song, Y. Sun, J. Guo, D. Zhao, and J. Tan. Application of case teaching method and Microteaching Method in experimental teaching based on wechat platform. In *BIC 2021: International Conference on Bioinformatics and Intelligent Computing, Harbin, China, January, 2021*, pp. 407–411. ACM, 2021.
- [44] T. Sukardiyono, M. I. Luthfi, and N. D. Septiyanti. Breaking Down Computer Networking Instructional Videos: Automatic Summarization with Video Attributes and Language Models. *Elinvo (Electronics, Informatics, and Vocational Education)*, 8(1):26–37, 2023.
- [45] J. C. Sun and K. Chang. Design and development of a location-based mobile learning system to facilitate English learning. *Univers. Access Inf. Soc.*, 15(3):345–357, 2016.
- [46] ThatsEngineering. Denavit - Hartenberg (DH) Tables For Robotic Systems - Direct Kinematics II, 2020. YouTube video.
- [47] F. Vindigni, T. Lasala, C. Giraudo, C. Furia, M. Caramagna, M. Bosco, and G. Baratto. Implementing Universal Design for Learning: An Online Training Course for Peer Tutors to Students with Disabilities and Specific Learning Disorders: Implementing UDL: An Online Training Course for Peer Tutors to Disabled and SLD Students. In *Proceedings of the 2022 6th International Conference on Education and E-Learning*, pp. 157–161, 2022.
- [48] Y. Wang, W. Liang, H. Huang, Y. Zhang, D. Li, and L. Yu. Toward Automatic Audio Description Generation for Accessible Videos. In Y. Kitamura, A. Quigley, K. Isbister, T. Igarashi, P. Bjørn, and S. M. Drucker eds., *CHI '21: CHI Conference on Human Factors in Computing Systems, Virtual Event / Yokohama, Japan, May 8–13, 2021*, pp. 277:1–277:12. ACM, 2021.
- [49] B. F. Yuksel, P. Fazli, U. Mathur, V. Bisht, S. J. Kim, J. J. Lee, S. J. Jin, Y. Siu, J. A. Miele, and I. Yoon. Human-in-the-Loop Machine Learning to Increase Video Accessibility for Visually Impaired and Blind Users. In R. Wakary, K. Andersen, W. Odom, A. Desjardins, and M. G. Petersen eds., *DIS '20: Designing Interactive Systems Conference 2020, Eindhoven, The Netherlands, July 6–10, 2020*, pp. 47–60. ACM, 2020.
- [50] D. Zhang, L. Zhou, R. O. Briggs, and J. F. N. Jr. Instructional video in e-learning: Assessing the impact of interactive video on learning effectiveness. *Inf. Manag.*, 43(1):15–27, 2006.
- [51] H. Zhou, R. Martín-Martín, M. Kapadia, S. Savarese, and J. C. Niebles. Procedure-Aware Pretraining for Instructional Video Understanding. *CoRR*, abs/2303.18230, 2023.

未来ビジョン

我々のロードマップには、効果、インタラクションパターン、および使いやすさに関する標準的なユーザースタディと、実世界のe ラーニングのパフォーマンスに関する探索的なスタディが含まれる。このアプローチにより、ユーザー・エクスペリエンスを包括的に理解することができる。

さらに、今後、アクセシビリティのさらなる向上を探求したいと考えている。認知的負荷の軽減という観点に焦点を当て、理解を深めるために、ビデオの特定の領域を拡大（ズーム）することを検討する。

拡大表示は以前の研究 [6] でも検討されたツールであるが、提案された視線ベースのソリューションは、e ラーニングコンテンツに適用すると失敗する。加えて、提示されたインタラクション手法は認知負荷の軽減には逆効果であり、改善できると考えている。

また、クリエイターのためのオーサリングツールの開発を模索している。「ヒューマン・イン・ザ・ループ」アプローチを採用し、e ラーニング・ビデオのアクセシビリティを高めるために活用することができるツール。SR システムを活用することで、カスタマイズされたバリアフリーな学習体験を生み出すことを目指している。

プログラミング初学者の課題解決方針の立案能力を高める選択肢タップ式学習支援システム

岡 大貴 * 大西鮎美 * 西田健志 * 寺田 努 * 塚本昌彦 *

概要. プログラミングが教養となりつつある近年では、さまざまな教育機関でプログラミングに関する授業が実施されている。しかし、初学者にとってプログラミング学習はつまずきの連続であり、特にプログラムを使った課題解決の方針を考えることや、その方針をプログラムコードを書いて実装するのが難しいことが学習障壁となっている。そこで、本研究ではプログラミングを「課題の解決方針を立案すること」と「解決方針をコーディングすること」に分離し、前者のみに専念させることで高速な試行錯誤を可能にし、プログラム課題の解決方針を立てる能力を高めるための学習システムを提案する。提案システムのプロトタイプを、スマートフォンからの利用を想定したWebアプリケーションとして実装し、ユーザが利用することでプログラム課題に取り組む際にどのような効果が得られるか実験を行った。

1 はじめに

プログラミング的思考 [1] や計算論的思考 [2] を育むことを目的とし、小学校においてもプログラミング学習が必修化された昨今では、プログラミングが誰もが身につけるべき教養となりつつある。プログラミングに関する技能を伸ばすことは重要である一方で、初学者にとってプログラミング学習はつまずきの連続であり、多くの学習障壁が伴う。

プログラミング初学者が学習時に直面する困難については多くの調査がなされており [3, 4, 5]、関数やポインタなど「プログラミングに関する概念の理解に関するもの」と、プログラムを設計することやその設計方法の理解、実装する機能を複数の手順に分割することなどの「プログラム設計の過程に関するもの」とがある。プログラミングに関する概念やプログラミング言語の理解を支援する学習教材やツールは多数存在するが [6, 7]、プログラムの設計の過程を支援するものは少ない。プログラムを設計する力を身につけるには、実際に課題に取り組んで試行錯誤し、課題を解決する経験が必要である。しかし、プログラミングにおける試行錯誤は難しい。

プログラム課題に対して試行錯誤することが難しい原因として、「プログラム課題を解くためにまず何から考えればよいのかがわからない」ことがあげられる。初学者はプログラミングでの課題解決の経験が少ないので、課題解決のために解決方針を立案することが難しい。そして「考えた解決方針を実践することが難しい」ことも原因である。初学者がある解決方針を考えついたとしても、それをプログラムにし実践するためには、プログラミングに関する概念や文法を理解したうえでプログラムを記述し、エ

ラーを取り除いてプログラムを完成させなければならない。

そこで本研究では、プログラミングを「課題の解決方針を立案すること」と「解決方針をコーディングすること」に分離して前者のみに専念させ、課題解決のための解決方針を選択式で表示し、有限な解決方針を選択するのみで高速に試行錯誤を経験させることで、プログラム課題解決のためのアプローチを学ばせ、解決方針の立案能力を高めるシステムを提案する。

提案システムは視覚的な出力を用いたプログラミング入門講義の課題で用いられることを想定しており、スマートフォンからの利用を想定したWebアプリケーションとして実装した。ユーザは取り組む課題を選択し、課題の解決方針の選択肢をタップするのみで対応したコードが挿入され、手軽に課題解決のための方針を試すことができる。例えば「日本国旗をプログラムで描画する」という課題の場合、まずユーザは表示される3つの選択肢「円を描く」「図形を塗る色を赤に指定する」「四角形を描く」の中から、「円を描く」という解決方針を選択する。次に「円を描く関数を使い、描画領域の中心に円を描く」「円を描く関数を使い、原点に円を描く」といった方針の実装手順に関する選択肢が表示される。日本国旗のような円を描くためには描画領域の中心に円を配置する必要があるので、「円を描く関数を使い、描画領域の中心に円を描く」という選択肢を選ぶ。すると、この実装手順をプログラムで表現したもののが表示され、ユーザはそれを確認し選択すると、エディタにプログラムが挿入される。ユーザはこのプロセスを繰り返すことで課題を解いていく。本論文では、プロトタイプシステムを実装し、これを用いて初学者に3問のプログラム課題を初学者に解かせ、プログラム課題の解決方針の立案にどのような

Copyright is held by the author(s).

* 神戸大学

効果があるかを調査した.

2 関連研究

2.1 プログラミング学習の学習障壁に関する研究

プログラミング学習時の学習障壁については広く調査がなされている. Derus らはプログラミングの基礎講義を受講している学生に、プログラミング学習の難しさに関するアンケート調査を行った. その結果、難しい学習内容は多次元配列や繰り返し、関数などの抽象概念や、プログラムの構造を理解したり、問題を解決するためのプログラムを設計したりすることであった[3]. Lahtinen らも 500 人以上の学生と教師を対象に、プログラミング学習の難しさに関するアンケート調査を行った. その結果、最も難しい概念は再帰やポインタと参照、抽象データ型のような抽象的な概念であり、最も難しいタスクは「ある課題を解決するためのプログラムをどのように設計するかを理解すること」「機能をステップに分けること」そして「自分のプログラムのバグを見つけること」であった[4]. また、Ismail らはプログラミング教育の問題点について、5人のコンピュータサイエンスの教師にインタビュー調査を行った. その結果、学生がプログラミング学習で直面する問題の4つの主な原因は、(1) 問題を分析するスキルの不足、(2) 効果的でない問題解決のための表現、(3) 問題解決とプログラミングの効果的でない指導方法、(4) プログラムの文法と構造を理解し習得する能力の不足、であると述べた[5].

これらの結果をみると、プログラミングの学習障壁には、関数やポインタなど「プログラミングに関する概念の理解に関するもの」と、プログラムを設計することやその設計方法の理解、実装する機能を複数の手順に分割することなどの「プログラミングの設計過程に関するもの」とがある. 本研究では、プログラムの設計過程に関する学習障壁に焦点をあて、なかでもプログラム課題の解決方針の立案能力を高めることを目指す.

2.2 初学者向けプログラミング学習のアプローチに関する研究

テキストベースのプログラミング言語についていきなり学ばせるのではなく、初学者のためにより学習しやすいアプローチを提案するシステムは数多く存在する. なかでも小中学校等のプログラミング学習には Scratch [8] を代表とするブロックベースのビジュアルプログラミング言語 (VPL: Visual Programming Language) が用いられることが多い. これらを用いた学習は構文エラーやタイプミスなどが発生しないため、考えた解決方針を実践することをある程度簡易化しているといえる.

また初学者向けプログラミング学習システムとし

て、テキストベースのプログラミング言語学習への足場かけを目的としたものがいくつか提案されている. 中西らは、ビジュアルプログラミングからテキストを用いたプログラミングへの橋渡しとして、初学者向けプログラミング学習環境である PEN のソースコードをフローチャートとして生成する PenFlowchart を開発し、高等学校の授業で運用した. その結果、PenFlowchart の導入により、フローチャートやプログラムの構造を把握する力が向上する可能性が確認された[9]. 末吉らは、VPL からテキストベースプログラミング言語への移行を支援するシステム Block Code を開発した[10]. このシステムは Web から利用でき、ブロックベースのブロックを使ってプログラムすると、それが C 言語のプログラムに変換されたものが表示される. 松澤らは、VPL からテキスト記述言語へのシームレスな移行を目的とした開発環境である BlockEditor を提案した. BlockEditor はブロックを組み合わせて作成したプログラムを Java プログラムに変換できる. これを文化系大学生向けのプログラミング入門教育で運用した結果、プログラミング学習が進行するにつれ、BlockEditor から Java へ徐々に移行していくことや、プログラミングに苦手意識をもつ学生ほど VPL の選択率が高いことがわかった[11].

又吉らは、プログラミング時のタイピングに着目し、用意された課題のプログラムをタイプすることで、逐次入力した行が実行され出力が閲覧できる学習システムを提案した[12]. このシステムを演習型プログラミング授業に導入した結果、ユーザのタイピング速度のばらつきが減り、タイピング速度が向上していることがわかった.

しかし、プログラム課題に対しどう解決方針を立てるかや解決方針をどのように実装していくのか、といった支援を行う学習システムは筆者らの知る限り存在しない. 本研究では、プログラム課題への取り組みを「課題の解決方針を立案すること」と「解決方針をコーディングすること」に分離し、前者のみに専念させることで高速な試行錯誤を可能にし、初学者のプログラム課題の解決方針を立てる能力を高める.

3 提案システムのプロトタイプ

本章では、プログラミングを「課題の解決方針を立案すること」と「解決方針をコーディングすること」に分離し、前者のみに専念させることで高速な試行錯誤を可能にし、プログラム課題の解決方針を立てる能力を高めるための学習システムのプロトタイプについて説明する. 本論文における解決方針の立案とは、プログラム課題を解決するためにまず何をするかといった大まかな方針と、それを表現する手順・アルゴリズムを考えることである.

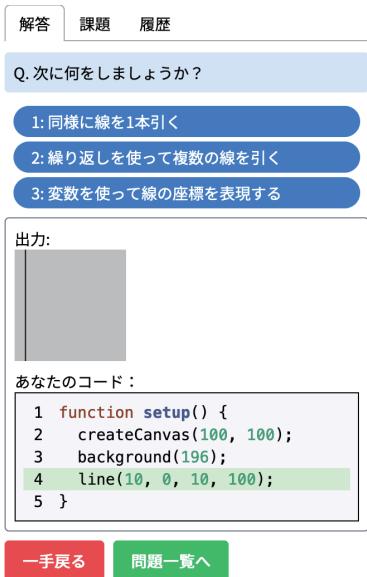


図 1: 提案システムの画面

3.1 設計指針

プログラム課題の解決方針の立案能力を高めるシステムの実装のため、4つの設計指針を設けた。

解決方針を気軽に試せる

プログラム課題は「解決方針を思いついでそれをコーディングできないので正誤がわからない」「実装プログラムのエラーが解決できずつまずいてしまう」などの障壁により、初学者が思いついで解決方法を試行錯誤することが難しい。よって選択式で解決方針を選択するのみでプログラムが自動で実装されるよう設計し、容易かつ高速に試行錯誤を繰り返すことができるよう設計した。

また選択肢が多すぎたり、学習システムの利用ハードルが高いと手軽に試行錯誤ができず、継続的な使用も見込めない。よって表示する選択肢は常に3個以下とし、スマートフォンでのタップ操作のみで課題に取り組めるよう設計した。

解決方針を段階的に実装することを意識させる

プログラムを書く際、例えばfor文で同じ処理を繰り返す場合でも、まずループ変数や条件式などを記述し、次にループ変数を利用して処理内容を記述する、といったように複数の小さなステップに分解できる。ある解決方針を考えた際に、それが複数のステップによって構成されており、その1つ1つを実装していくことで解決方針を実現できる、ということを体験させるために、表示する選択肢は複数の小さなステップに分かれるよう設計した。

過去の選択の振り返りを促す

プログラム課題を実際にプログラムを記述して解いた際、プログラムの流れと課題解決のための思考過程は必ずしも同じとは限らないため、どのような思考過程で解答に至ったかを振り返ることは初学者にとって難しいと考えられる。よって、課題に取り組む過程や課題解決時に思考過程を振り返ることを促すよう設計した。

実際のコーディングと段階的に接続する

提案システムの目的は解決方針を立てる能力を高めることであるが、これができるだけでは実際にコーディングして課題を解くことに繋がらない可能性がある。よって、ユーザが解決方針のみを選ぶだけでなく、どのようなプログラムの実装になるかを確認させるフェーズを設け、提案システムを使用せずにコーディングするための足場かけとなるように設計した。

3.2 対象とするプログラミング分野

提案システムは、初学者向けのプログラミング学習での利用を想定している。昨今では、初学者向けに Processing [13] や p5.js [14] といった視覚的な出力を用いたテキストベースのプログラミング環境を用いることが増えている [15]。提案システムは、プログラミング言語 JavaScript とその描画ライブラリである p5.js を利用したプログラミング講義での課題を解く際に、スマートフォンを利用して使用されることを想定している。例えば p5.js を使って日本国旗を描くプログラムは、プログラム 1 のようになる。

プログラム 1: p5.js のプログラム例

```

1 createCanvas(180, 120);
2 background(255);
3 fill(255, 0, 0);
4 noStroke();
5 ellipse(width/2, height/2, height*3/5,
height*3/5);

```

3.3 ユーザインターフェース

提案システムの利用時の画面を図 1 に示す。はじめに、ユーザは Web ブラウザから指定の URL にアクセスし、取り組む課題を選ぶ。そして表示される選択肢をタップ操作で選択することで課題を解いていく。図 1 の画面は、現状のユーザのコードとその出力画像、選択肢からなる「解答タブ」、課題が表示される「課題タブ」、ユーザがこれまでに選択した過去の選択肢を閲覧できる「履歴タブ」が表示されており、それぞれのタブをタップすることで対応した内容を閲覧できる。ユーザが取り組む課題を選ぶとその解決方針が選択式で表示され、タップで選



図 2: 解答タブのフェーズ

択肢を選んでいくことで、それを実現したコードがテキストエディタに追記される。これを繰り返すことで、ユーザはプログラムの実装をせずとも課題の解決方針のみを考え、選択するだけで手軽にさまざまな解決方針を試すことができる。また、誤りの選択肢を選んでしまったり、自分の解決方針に違和感を感じたりした場合は「一手戻る」ボタンをタップし、前の選択肢に戻り選択をやり直すことができる。

解答タブ

このタブに各設問の選択肢が表示され、ユーザはそれらをタップすることで課題を解いていく。各設問は3つのフェーズ「方針選択」「アルゴリズム選択」「コード確認」に分かれており、これらのフェーズを繰り返すことでユーザは最終的な解答を完成させる。それぞれのフェーズの画像を図2に示す。

1つ目の「方針選択フェーズ」(図2(a))では、ユーザは課題の解決のためにどのようなアプローチを取るかを選択する。例えば、「円を描く」、「繰り返しを使って複数の線を引く」などである。選択肢は青色で表示される。

2つ目の「アルゴリズム選択フェーズ」(図2(b))では、選択した課題解決のためのアプローチをプログラムするための手順やアルゴリズムを選択する。例えば、「円を描く関数を使い、canvasの中心に円を描く」「ループ変数を使って9本の線を引く」などである。選択肢は黄色で表示される。

3つ目の「コード確認フェーズ」(図2(c))では、アルゴリズム選択フェーズで選択したコードを確認し、それを選択するとテキストエディタにコードが挿入され、ハイライトされる。このフェーズにより、立案した解決方針を実際にコーディングするとどのようなコードになるのかをユーザに確認させる。選択肢は紫色で表示される。

方針選択フェーズとアルゴリズム選択フェーズは、選択肢をタップして高速に試行錯誤することで、解決方針を立てる能力を高めるためのフェーズであり、コード確認フェーズは、実際のコーディングへの接



図 3: 解決方針の例

続を意図したフェーズである。

課題タブ

このタブには取り組んでいる課題の説明文と再現する対象の画像が表示される。また、現状のユーザの方針選択により作成されたプログラムの出力も表示される。

履歴タブ

これまでにユーザが選択した選択肢の一覧が図3のように表示される。それぞれのフェーズの選択肢が色とインデントで階層的に表示されている。

3.4 実装

提案システムは、Web アプリケーションフレームワークである Next.js [16] を使い、Web アプリケーションとして実装した。選択肢は JSON 形式で表現されており、プロトタイプの実装では第一著者が手作業で用意した。

4 予備実験

提案システムのプロトタイプの利用により、学習者のプログラム課題の解決方針の立案能力にどのような効果があるかを調査するための実験を行った。

4.1 実験内容

実験の被験者は 10 から 20 代の学生 8 名 (女性 1 名、男性 7 名) である。被験者のプログラミング経験に関するアンケートの結果、「全くない」が 1 名、「独学や授業での学習経験あり」が 3 名、「変数や繰り返しなどの基礎的な概念を使ったプログラムが実装できる」が 3 名、「ソフトウェアを開発したことがある」が 1 名であった。

実験手順は、まずプログラミングに関する基礎概念の理解と確認のため、プログラミング初学者向け

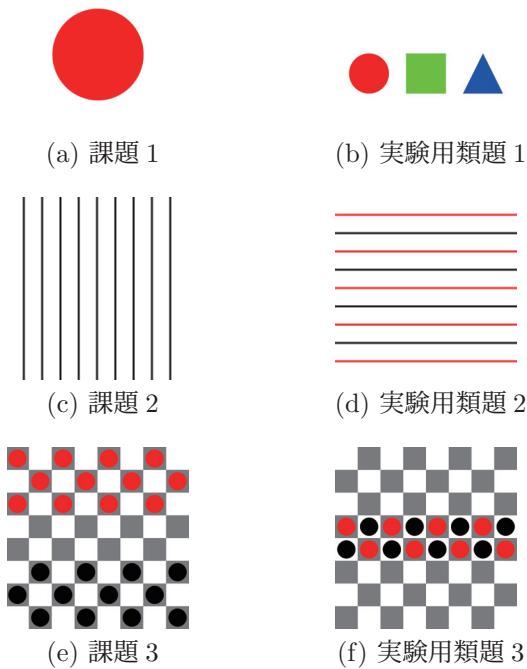


図 4: プロトタイプで用意した課題と実験用類題で用いた画像

講義のテキストを読ませ、プログラム課題を解くのに必要な知識を学習させる。そして、提案システムのプロトタイプを使用させ、用意しているプログラム課題 3 問に取り組ませる。取り組むプログラム課題の内容は、図 4(a), 4(c), 4(e) の画像をプログラムで再現するというものである。その後に提案システムの所感についてアンケート調査を行ったのち、プログラム課題の立案方針を問うテスト課題に取り組ませた。この課題は提案システムで取り組んだ課題の類題 3 問である。類題の画像を図 4(b), 4(d), 4(f) に示す。また、提案システムはコーディングそのものの支援は行っていないため、被験者には課題をコーディングして解かせるのではなく、日本語で解答方針のみを箇条書きで記述させた。なお本実験は、神戸大学大学院工学研究科人を直接の対象とする研究倫理審査委員会の承認を受けて行ったものである（審査番号 05-18）。

4.2 実験結果と議論

テストの解決方針

類題のテストで得られた解決方針の例を以下に示す。解決方針例 1, 2, 3 はそれぞれ別の被験者によるものである。このように粒度に違いはあるものの、多くの被験者が解決方針を日本語で段階的に記述できていることがわかった。解決方針例 1, 2 のように日本語のみで完結している解答もあれば、解決方針例 3 のように for 文などのプログラムの構文を用い

た解答もみられた。

解決方針例 1 (類題 1)

- ・塗る色を赤にする
- ・円を描く関数を座標と共に使って左側に描く
- ・塗る色を緑にする
- ・四角を描く関数を座標と共に使って中央に描く
- ・塗る色を青にする
- ・△を描く関数を座標と共に使って右側に描く

解決方針例 2 (類題 2)

1. 1 本の線を描画する
2. 繰り返し文を利用して線を 9 本に増やす
3. 条件分岐を利用して線の本目が奇数ならば赤色、偶数ならば黒色に線を塗るように指定する

解決方針例 3 (類題 3)

- ・for 文でループを作成 ($i=0$ とし、 i が 8 より小さい間、 i を 1 ずつ増やす)
- ・作成した for 文の中にもう一つ for 文を作成 ($j=0$ とし、 j が 8 より小さい間、 j を 1 ずつ増やす)
- ・ $i+j$ を 2 で割ったあまりが 1 のとき、 図形の色を指定 (グレー)、 正方形を作成 (x 軸が i , y 軸が j)、 正方形の枠線を削除
- ・ $i+j$ を 2 で割ったあまりが 1 かつ、 j が 2 より大きく 5 より小さい時、 図形の色を指定 (赤)、 円を作成 (x 軸が i , y 軸が j)、 図形の枠線を消す
- ・ $i+j$ を 2 で割ったあまりが 0 かつ、 j が 2 より大きく 5 より小さい時、 図形の色を指定 (黒)、 円を作成 (x 軸が i , y 軸が j)、 図形の枠線を消す

タップ操作のログ

被験者が提案システムの利用時に行った操作のログを分析した結果について述べる。被験者全体のタップ操作をみると「一手戻る」ボタンが平均 14.6 回使用されており、被験者が複数の選択肢を選んで試行錯誤してたことがわかる。履歴タブの参照回数は平均 3.6 回であり、どの被験者も解答の過程で何度も方針を振り返っていることが確認できた。また課題タブと解法タブの参照回数がそれぞれ平均 20.0 回、15.1 回であり、2 つのタブを往復して課題を確認しながら解いていく様子がみられた。

表 1: 実験のアンケート結果 (1: 全くそう思わない, 5: とてもそう思う)

コード	質問	AVE	SD
Q1	プログラム課題の解決方針を立てる力が向上したと感じるか	3.5	0.93
Q2	各問題の選択肢を気軽に試すことができたか	4.4	0.92
Q3	授業等のプログラム課題と比べて気軽に取り組めたか	4.6	0.52
Q4	各問題を解く方針を段階的に考えることができたか	4.5	0.53
Q5	各問題について、選択肢の数や粒度は適切だと感じたか	3.8	1.3
Q6	アプリは使いやすかったか	3.0	0.53
TQ1	テスト課題を解く際に段階的に考えることができたか	4.3	0.46
TQ2	テスト課題を解く際に提案システムの影響があったと感じるか	4.9	0.35

アンケート結果

実験で得られた提案システムのプロトタイプに関するアンケートの結果を表 1 に示す。Q1 の結果より、提案システムを利用することでプログラム課題の解決方針を立てる能力を高められる可能性がある。自由記述からは「日本語で次に何をすべきかを考えることが『解決方針を立てる』という面でみれば、プログラミングスキルの優劣にかかわらず役に立つのではないかと感じた」といったコメントが得られた。一方で「このレベルの問題は C 言語で何度もやってきたので、特に能力が向上したとは感じなかった」というコメントも得られたため、他のプログラミング言語の経験があり、解決方針の立案能力がすでに高い場合は効果が得られない可能性がある。Q2 と Q3 の結果より、高速な試行錯誤を促進できる可能性がある。さらに Q4 の結果より、スマートステップでの問題解決を促進できる可能性がある。

また Q5, Q6 の結果より、ある程度プロトタイプは使いやすく、手軽に利用できる可能性がある。しかし、UI に関しては「最初は問題がどこにあるのかわからず困ったので、課題と解答のタブが左右逆になっていた方がやりやすいと思う」などのコメントが得られたため、画面に表示する情報を整理し、使いやすさを向上させる必要がある。

なお、テスト課題に関するアンケートである TQ1 と TQ2 の結果から、提案システムを利用することで、プログラム課題を解く際に段階的に解決方針を立てられるようになる可能性がある。また「目標物が段階的に作り上げられている過程を見ることは良いと思った。コードを選択肢で選びつつ文字でも確認できるのは初学者には嬉しいのではないかと思った」「Web アプリで学んだプログラムの構成の手順に関する知識を活かして解答出来たと感じた」「プログラムを始める前に、いったん頭を整理しようという段階が明確に意識された気がした」「実際にあったら使ってみたい。普段無意識でやりがちな、完成に向けての各工程の言語化の大切さをわからせてくれるアプリだなと思った」などのコメントが得られ、課題の解決方針を設計・分割してか

ら取り組むことを促進できる可能性がある。しかし、今回の実験は 1 時間程度の短期間の実験であり、取り組んだ課題もプロトタイプで取り組んだ課題の類題であるため、提案システムの効果を検証するためにはより被験者数を増やし、長期的な実験を行う必要がある。

5 今後

実験結果を踏まえた提案システムの今後の方向性について述べる。まず、課題の作成方法についてである。今回の提案システムのプロトタイプでは、3 問の課題の選択肢を第一著者が手作業で用意したが、実際のプログラミング演習講義では多くの課題が提出され、すべての課題の選択肢を教授者が手作業で用意するのは現実的ではない。よって、今後は課題の選択肢を半自動的に作成できるエディタの実装や、ChatGPT [17] などの大規模言語モデルを利用した選択肢の自動生成を検討する。

そして、今回の実験は数名の学習者を対象に提案システムのプロトタイプを利用させることにとどまったため、プロトタイプを改善し、より多くの初学者を対象に長期的な実験を行う。具体的には、神戸大学国際人間科学部の講義「プログラミング基礎演習 1」の予習課題に提案システムを導入し、受講者たちの課題への取り組み方の変化を調査する。

6 まとめ

本研究ではプログラミングを「課題の解決方針を立案すること」と「解決方針をコーディングすること」に分離し、前者のみに専念させることで高速な試行錯誤を可能にし、初学者の課題解決方針を立てる能力を高める選択肢タップ式学習システムを提案した。提案システムのプロトタイプを、プログラム課題の解決方針を選択肢タップにより手軽に試すことができる Web アプリケーションとして実装した。プロトタイプの予備実験の結果、提案システムを利用することで課題解決方針を立てる能力を高められる可能性があるとわかった。今後はプロトタイプを改善し、より多くの初学者に対し長期的な実験を行う。

謝辞

本研究の一部は、JST CREST(JPMJCR18A3)およびJST科学技術イノベーション創出に向けた学フェローシップ創設事業JPMJFS2126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] 文部科学省: 小学校段階におけるプログラミング教育の在り方について(議論の取りまとめ), https://www.mext.go.jp/b_menu/shingi/chousa/shotou/122/attach/1372525.htm (accessed on 30 October 2023).
- [2] J. M. Wing: Computational Thinking, *Communications of the ACM*, Vol. 49, No. 1, pp. 33–35 (2008).
- [3] S. R. M. Derus and A. Z. M. Ali: Difficulties in Learning Programming: Views of Students, *Proc. of the 1st International Conference on Current Issues in Education (ICCIE 2012)*, pp. 74–78 (Sept. 2012).
- [4] E. Lahtinen, K. Ala-Mutka, and H. Jarvinen: A Study of the Difficulties of Novice Programmers, *Proc. of the 10th Annual SIGCSE Conference on Innovation and Technology in Computer Science Education (ITiCSE 2005)*, pp. 14–18 (June 2005).
- [5] M. N. Ismail, N. A. Ngah, and I. N. Umar: Instructional Strategy in the Teaching of Computer Programming: A Need Assessment Analyses, *Turkish Online Journal of Educational Technology (Turk. Online J. Educ. Technol.)*, Vol. 9, No. 1, pp. 125–131 (2010).
- [6] Progate: <https://prog-8.com/> (accessed on 30 October 2023).
- [7] ドットインストール: <https://dotinstall.com/> (accessed on 30 October 2023).
- [8] M. Resnick, J. Maloney, A. Monroy-Hernández, N. Rusk, E. Eastmond, K. Brennan, A. Millner, E. Rosenbaum, J. Silver, B. Silverman, and Y. Kafai: Scratch: Programming for all, *Communications of the ACM*, Vol. 52, No. 1, pp. 60–67 (2000).
- [9] 中西 涉, 辰己丈夫, 西田知博: PenFlowchartによるプログラミング導入教育の評価, 研究報告コンピュータと教育(CE), Vol. 2013, No. 9, pp. 1–7 (Oct. 2013).
- [10] 末吉春一, 佐藤 喬: ビジュアルプログラミングを用いたテキストベースプログラミング学習支援システム, 第78回全国大会講演論文集, pp. 897–898 (2016).
- [11] 松澤芳昭, 酒井三四郎: ビジュアル型言語とテキスト記述型言語の併用によるプログラミング入門教育の試みと成果, 情報処理学会研究報告コンピュータと教育(CE), Vol. 2013, No. 2, pp. 1–11 (Mar. 2013).
- [12] 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, 情報処理学会, Vol. 2020, No. 1, pp. 1–8 (Aug. 2020).
- [13] Processing, <https://processing.org/> (accessed on 30 October 2023).
- [14] p5.js, <https://p5js.org/> (accessed on 30 October 2023).
- [15] 三好きよみ: Processingによる初学者向けプログラミング教育の実践, 情報教育シンポジウム論文集, Vol. 2020, pp. 89–95 (Dec. 2020).
- [16] Vercel: Next.js, <https://nextjs.org/> (accessed on 30 October 2023).
- [17] OpenAI: ChatGPT, <https://chat.openai.com> (accessed on 30 October 2023).

AR ハンドと EMS による手指の同期を用いた遠隔でのピアノ演奏指導

守 新太郎 * 井原 圭一 † 川口 一画 ‡

概要. 遠隔同期型の楽器演奏指導において細かな演奏動作を確認するためには、学習者は複数の視点から指導者の動作を見る必要がある。そのため、複数カメラの設置と映像のスイッチングが必要となり、操作が負担となる。また、各カメラの視点は固定されているため、学習者はより細かな動作の確認ができない。加えて、対面で行う指導と比べて、指導者の存在感を感じづらい。そこで、我々は、拡張現実感 (AR) と電気的筋肉刺激 (EMS) を組み合わせて遠隔からの楽器演奏支援システムを提案する。指導者の手の動きを変換し学習者の HMD 越しに表示することによって、演奏時の手の形の効率的な伝達および遠隔者の存在感の向上を目指す。また、指導者が楽器の弦や鍵盤などを押さえたときに、学習者の手に対して EMS を与えることによって、指の動きや押す強さを提示する。本論文では、ピアノの演奏指導における提案システムのプロトタイプを実装し、ピアノ指導における具体的なインタラクション方法を示す。

1 はじめに

近年、楽器を始める人の増加に加えて、オンライン指導も増加している。オンライン指導には、動画を視聴して学習するオンデマンド型や、リアルタイムで指導者が遠隔から学習者に指導する遠隔同期型がある。ただ、遠隔同期型の楽器演奏指導において細かな演奏動作を確認するためには、指導者および学習者がともに複数の視点から相手の動作を見る必要がある。そのためには複数カメラの設置と映像のスイッチングが必要となり、操作が負担となる [6]。また、各カメラの視点は固定されているため、演奏時の細かい指の動きや角度を確認したい場合（弦の押さえ方の確認など）に問題が生じる。

このような視点に対する先行研究として、拡張現実感 (AR) を用いたアプリケーションが検討されている。AR を用いた方法 [13] では、ピアノ演奏時の指導者の手を AR アバタに変換し、学習者が HMD を通して手の形を確認できるようにすることによって、指導者は学習者に対して鍵盤に置く手の位置や動作を効率的に伝えられる。ただし、AR アバタでは、楽器の弦や鍵盤などを押す強さは提示できない。また、手のみの AR アバタでは、遠隔同期型での指導の場合、指導者の存在感を感じづらい。これに対して、EMS を用いた方法 [26] では、電気刺激により楽器を演奏する学習者の指を直接制御することによって、楽器の弦や鍵盤などを押す強さを提示することができる。また、初心者は、打鍵時の再現性が低い [28] ことから、指を直接制御することによって指の角度や動きの再現性を高められる。ただし、EMS を用いた手法は、指導者が学習者に対

して遠隔からリアルタイムに指示を行うケースでは検討されていない。

そこで本研究では、楽器演奏支援に対するアプローチとして、AR アバタと電気的筋肉刺激 (EMS) を組み合わせて、遠隔からの楽器演奏を支援するシステムを提案する（図 1）。これは、指導者を AR アバタに変換し学習者の HMD 越しに表示することによって、演奏時の手の形や位置を効率的に伝え、遠隔指導時の存在感を向上させる。加えて、指導者が楽器の弦や鍵盤などを押さえたときに、学習者の手に対して EMS を与えることによって、指の動きや押す強さを提示する。これにより、学習者は能動的に自分の手を動かしながら、受動的に指導者からの指の動きの指示を受けられる。

本論文では、そのプロトタイプとしてピアノ演奏時に AR ハンドと EMS (2 チャンネル出力) で遠隔指導を行うシステムを実装する。また、本プロトタイプを用いた指導方法を示す。

2 関連研究

本システムは、AR アバタと EMS を用いて、遠隔からの楽器演奏を支援する。本章では、楽器演奏支援に対する既存手法、および EMS に関する関連研究について述べる。

2.1 ピアノにおける楽器演奏支援

楽器演奏支援として、いくつかのアプローチがある。ピアノの指導に対する学習者ベースのアプローチでは、(1) さまざまな身体部位の動きと姿勢を同期させる、(2) 譜読力を向上させる (3) 学習者のモチベーションを高める (4) アドリブ演奏を奨励することの 4 つが重要である [4]。身体の同期を用いたアプローチとして、ユーザに対して次にどの鍵盤を押せばいいかを色分けする、タイミングのエラーを表

Copyright is held by the author(s).

* 筑波大学 情報メディア創成学類

† 筑波大学 情報理工学位プログラム

‡ 筑波大学 システム情報系



図 1. 提案システムを示す. (a) 指導者の演奏を単眼カメラから撮影し, ハンドトラッキング情報を学習者側へ送信する. (b) 学習者は EMS を指に装着し, 指導者の手を模した AR ハンドに手の動きを追従しながら, EMS による指の動きの支援を受ける. (c) 指導者の手の俯瞰映像から関節点を取得する. (d) 指導者の手を模した AR ハンド. 関節点とそれらを結ぶボーンから構成される.

示させるといったフィードバックを与えるものがある [10, 11, 25]. また, ハンドトラッキングを用いて指導者の手の形を AR アバタに変換し, それを学習者に対して, HMD 越しに表示させることによって, 手の形や動きを効率的に伝える手法も提案されている [13]. この手法では, 現実空間の鍵盤の上に表示された指導者の手の動きを理解することによって, 学習者は演奏時の手を正しく位置づけ, 手の形を再現する能力が高まることが示唆されている. 加えて Liu らは, 指導者と学習者の手の動きをモーションキャプチャによって記録し, ポーン表示を用いて手の動きの違いを視覚的に表示する手法を提案している [14]. 譜読力への支援として, ピアノロールを鍵盤の延長線上に表示させ, 次に押す鍵盤の位置に加えて, タイミングをユーザに教える手法がある [2, 7]. 学習者のモチベーションを高めるアプローチでは, AR とゲーミフィケーションに基づいたアプローチがある. ピアノの演奏にゲームプレイの要素を追加したもの [17] や, 定期的にピアノを演奏してポイントを獲得するといったインセンティブ [18] によって動機づける手法がある. アドリブ演奏への支援として, 手のジェスチャーをカメラ [8] や EMG[12] を用

いて検出し, 手の形や動きに応じて, 音を変化させ, 音の表現を拡張している. 先行研究とは異なり, 本研究では, AR アバタと EMS を組み合わせることで, 手指の動きの同期をリアルタイムで行い, 学習効率の向上を図る.

2.2 電気的筋肉刺激 (EMS)

EMS を用いた楽器演奏アプリケーションはこれまでにいくつか検討されている. Stimulated Percussions[5] は, 前腕もしくは, ふくらはぎに電気刺激を与えて打楽器でのリズムをユーザに直接提示する. Niijima らは, トレモロ演奏するピアノ熟練者と初心者の筋活動を比較し, ピアノ熟練者が演奏時によく活動する上腕筋や前胸筋へ EMS を与えることによって, 初心者がトレモロ演奏するときに不必要的筋活動を抑制している [19]. ただ, この研究ではピアノの演奏手法としてトレモロ演奏のみに限定している. PossessedHand[27] は, 前腕周囲の筋肉に電気刺激を与えることによってユーザの指を制御する. 各筋肉は 28 個の電極パッドを介して刺激され, 手の 16 の関節の動きを制御できる. このシステムを用いて, 琴の楽器演奏支援を行い, タイミング

AR ハンドと EMS による手指の同期を用いた遠隔でのピアノ演奏指導

のエラー、および弦の間違いを軽減している。ただし、このシステムは、親指を除く PIP 関節（近位指節骨間関節）を独立的に制御できるが、それ以外の関節は制御できない。これは、前腕部にある指を屈曲するための指屈筋が密接しているため、前腕部に電気刺激を与えると関節を独立的に制御することが難しくなるからである。そこで、Takahashi らは、指の MCP 関節（中手指節間関節）を動かす虫様筋に電気刺激を与える手法をとっている [26]。これは、従来の前腕部に電気刺激を与える手法と比べて、MCP 関節を独立的に制御することができる。また、複雑な指の形を制御する方法として、EMS と外骨格を組み合わせたものが検討されている [20]。これは EMS によって指を作動させ、外骨格を用いて指を正確な位置で停止させるアプローチを取っている。これによって、MCP、および PIP 関節を独立して制御できる。先行研究では、指導者が学習者に対して遠隔からリアルタイムに指示を行うケースでは検討されていない。そこで、我々は、Takahashi らの先行研究を元に、EMS を用いて遠隔にいる指導者の指の動きを学習者の指へ提示する。

3 提案システム

提案システムは、HMD を通した AR アバタの表示と、EMS の制御の組み合わせで構成されている。将来的には、指導者の全身を AR アバタとして提示し、さらに両手の全ての指（10 チャンネル）に EMS を提示する予定であるが、本論文中ではそのプロトタイプとして、AR ハンドの提示と 2 チャンネル分の EMS 提示が可能なシステムを実装した。本章では提案システムにおけるこれらのシステム実装構成について述べる。

3.1 システム構成

提案システムの実装構成を図 2 に示す。単眼カメラ（カメラ：Sony α6300、レンズ：SEL1650）を用いて指導者の手の映像を取得し、MediaPipe を用いてハンドトラッキングを行う。ハンドトラッキング情報と、ピアノ演奏における MIDI 情報を、UDP 通信を用いて指導者側の PC(CPU:Intel Core i7-12700H, GPU:Intel Arc A370M, メモリ：32GB) から、学習者側の PC(CPU:Intel Core i7-11800H, GPU: GeForce RTX 3070 Laptop, メモリ：64GB) に送信し、HMD 上にアバタとして表示する。そして、マイコンから EMS を制御し、学習者の指を制御する。指導者と学習者は、ビデオ会議ツールを用いて、音声を共有する。また、指導者は、学習者の学習状況を確認するため、学習者側の手を映したカメラ映像を、ディスプレイにて確認する。単眼カメラの位置は鍵盤の真上に設定した。HMD は Hololens2[15] を、HMD への表示ソフトウェアとして Unity（バージョン 2021.3.21f1）を用いた。また、ピアノは Casio

LK-520[1] を用いた。

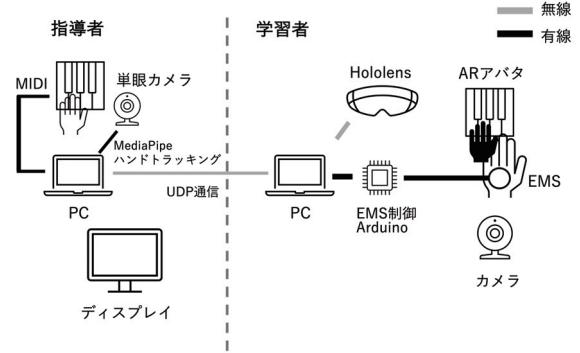


図 2. システム構成図。

3.2 指導者の演奏動作の取得

指導者の演奏時の手の動きを、ハンドトラッキングを用いて取得し、打鍵時の強さを EMS 入力として用いるため、MIDI 情報を取得する。

3.2.1 関節点の取得

OpenCV と Mediapipe のライブラリがパッケージ化されている CVzone[3] を用いて、指導者に対しハンドトラッキングを行い、UDP 通信にて学習側の PC へ送信する。ハンドトラッキングは、Mediapipe を用いて、手首と各指 4 点の計 21 点の関節点を取得した。リアルタイムでのハンドトラッキングにおけるフレームレートは 20FPS 程度であった。

3.2.2 MIDI 情報

Mediapipe を用いたハンドトラッキング情報では、真上からの映像であるため、指がどの程度曲がっているか取得することが難しく、EMS の強さを調節するための入力として用いることができない。そこで、指導者が演奏するピアノの MIDI 情報から、打鍵または離鍵した鍵盤の位置、および打鍵の強さ（音の大きさ）を取得し、打鍵の強さに応じて与える EMS の強さを調節する。このとき、打鍵の強さは 128 段階で取得できる。そして、打鍵または離鍵した鍵盤の位置、および AR 上に表示されたピアノと AR ハンドの衝突情報を用いて、弾いている鍵盤の指を推定し、EMS を与える指を選択する。具体的には、各指の先端の関節点と、仮想ピアノの各鍵盤にゲームオブジェクト同士の物理的な衝突を検出する機能を持つ Collider コンポーネントを持たせている。鍵盤の Collider コンポーネントを図 3 のように y 軸上に検出範囲を伸ばすことによって、x 軸と z 軸の 2 次元平面上において、関節点と鍵盤が重なったときに衝突を検出する（図 4）。そして、打鍵、または離鍵した鍵盤の位置と、その鍵盤と重なっている関節点から、弾いている鍵盤の指を推定する。

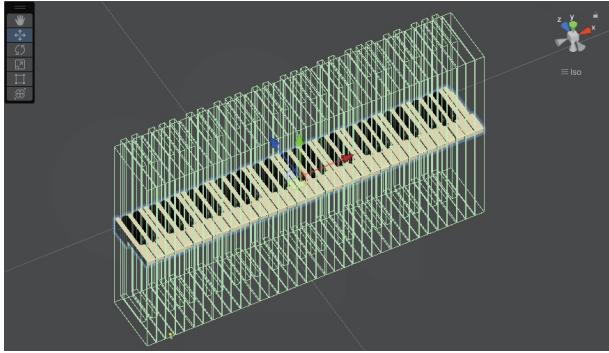


図 3. 鍵盤の Collider コンポーネント。

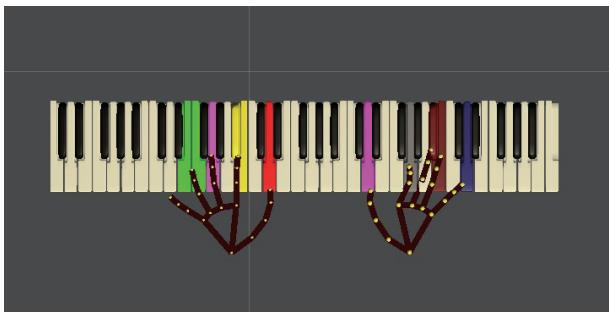


図 4. 関節点と鍵盤の衝突検出。衝突検出を視覚的に分かりやすくするために、鍵盤に対して衝突した指ごとに色付けを行っている (e.g. 左手親指: 赤色)。3 オクターブ目のレが押された場合、レの鍵盤と重なっている左手親指に対して EMS を与える。

MIDI 情報は、ピアノから USB 接続で指導者側の PC に送られ、Python ライブライアリの Pygame[22] を用いて情報を取得する。その後、UDP 通信によって学習者側の PC へ MIDI 情報が送信される。このとき、MIDI 情報を同時に送信すると、複数の鍵盤を同時に打鍵、または離鍵した情報が損失する場合があるため、即時送信ではなく、0.05 秒おきに送信する。

3.3 AR ハンド表示

指導者の手を模した AR ハンドは、関節点の球体と関節点をラインで結んだ構成をしている (図 1(d))。AR 空間上には現実のピアノとの位置を調節するために、現実のピアノと同等サイズの仮想ピアノを配置している。この仮想ピアノを現実のピアノに重畠表示して、位置合わせを行う。Mediapipe で取得した手のランドマーク座標は、Unity 上にて関節点を表示する球体に各関節の座標に対して入力する。AR ハンドのサイズは、仮想ピアノと比較しながら図 1(c) のような指導者の手と大きさが同じになるように、x,y,z 軸のすべてのスケールを手動で調整する。

3.4 EMS 制御

既製品として AR との組み合わせを想定している EMS デバイス Unlimitedhand がある [9]。このデバイスでは手指の動きに関する入出力を行えるが、指の動きが限定的であるため、ピアノ演奏に対して不適当である。そこで、電気刺激によって学習者の指を直接制御する EMS 制御装置として、Pfeiffer らの Let-Your-Body-Move toolkit [21] を使用した (図 5)。また、EMS 生成器として Sanitas SEM 43[23] を使用した。このツールキットにおける制御回路は、Arduino の出力に応じて、デジタルポテンショメータによって MOSFET を制御して抵抗値を調整し、EMS の出力電流を調節する。MOSFET の抵抗値は 256 段階で変更できるため、EMS の出力は 256 段階の調節ができる。

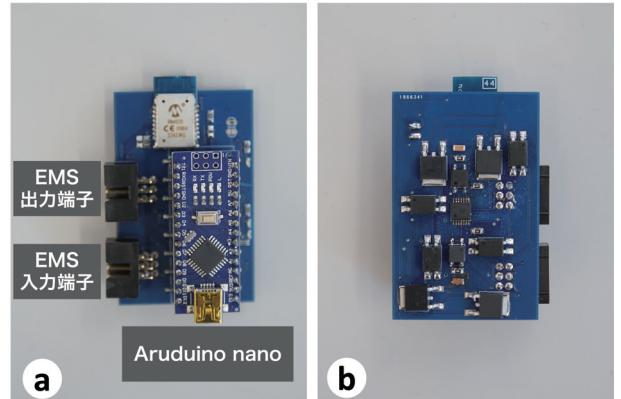


図 5. EMS 制御基板. (a) 表面. (b) 裏面.

ピアノの演奏では、音階を打鍵する場合、MCP 関節が屈曲し PIP 関節が伸展する。また、和音を打鍵する場合、手首、MCP 関節および、DIP 関節(遠位指節骨間関節)が屈曲する [24]。したがって、音階・和音の演奏動作から指の MCP 関節を屈曲・伸展させるため、電極パッドの装着位置として指の MCP 関節を独立して制御できる手の虫様筋に設定した (図 6)。電極パッドのサイズは、 $1 \times 3 \text{ cm}$ (虫様筋) と $3 \times 3 \text{ cm}$ (手首) とした。MCP 関節の手のひらとの最大屈曲角度は、全ての指で 90 度程度であった。ここで、親指の装着位置は、虫様筋が無いため、MCP 関節を屈曲させる長母趾屈筋に電極パッドを装着した。現段階での実装では、EMS の出力チャンネルが 2 チャンネルのみであるため、指を制御できる数は 2 本である。鍵盤の打鍵から指に EMS が与えられるまでの時間は約 0.2 秒であった。また、指が動き始めてから最終的な指の形になるまで、最大約 0.4 秒かった。研究室内の同一ネットワークを使用している環境では、キーボードの打鍵から光るタイミングまでの時間差から、ハンドトラッキングと MIDI イベントの受信との間には約 0.2 秒の遅延がある。

AR ハンドと EMS による手指の同期を用いた遠隔でのピアノ演奏指導



図 7. 実装したプロトタイプを用いた指導のながれを示す. (a) 学習フェーズ: 指導者は学習者へ手本を示す. (b)(c) 学習フェーズ: 学習者は AR ハンドに重なるよう自分の手を動かして学習. (d) テストフェーズ: 学習者は何も見ずにピアノ演奏のテストを行う. (e) フィードバックフェーズ: 指導者は学習者へテスト中に出た技術的課題に関してフィードバックを与える.

延が生じている. また, EMS が与えられてから最終的な指の形になるまで, 最大で約 0.2 秒かかることが観察されたため, この環境では, AR ハンドと EMS の出力がほぼ同時に実行されている. しかし, 遅延時間はネットワーク環境などに依存すると考えられるため, 今後の課題として適切な同期処理を検討する必要がある.

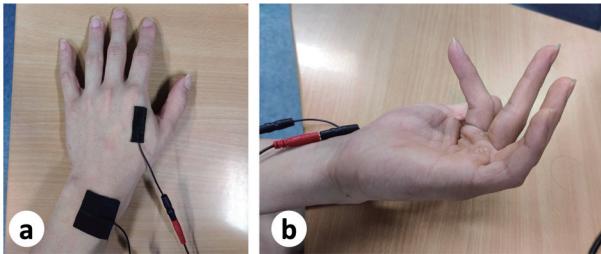


図 6. EMS 装着位置. (a)EMS を左手人差し指の虫様筋に装着した. 手首の電極パッドは GND として用いる. (b)EMS を与えたときの指の曲がり具合.

3.5 実装したプロトタイプを用いた指導のながれ

実装したプロトタイプを用いた指導では, 以下のように指導を行う.

1. 学習フェーズ: 指導者はピアノを演奏し, 学習者に対して課題曲またはフレーズの手本を示す(図 7(a)). 学習者は, 鍵盤上に表示された AR ハンドに重なるよう自分の手を動かすことで, 指導者の手の動きを学習する(図 7(b)(c)). このとき, 学習者の指には指導者の動きと連動して EMS が与えられ, EMS をトリガーとして学習者自身も指を動かしながら, 指の動きのタイミングや強さを学習する.
2. テストフェーズ: 学習後, 学習者は指導者の手の動きを思い出しながら, 何も見ずにピアノ演奏のテストを行う(図 7(d)).
3. フィードバックフェーズ: テスト後, 指導者は学習者の手の動きをディスプレイで確認しながら, テスト中に出た運指, 音の強弱, およびリズムの正確さなどの演奏時の技術的課

題について学習者へフィードバックを与える(図 7(e)).

そして最終的に学習者がテストの合格基準を満たすまで, 1 ~ 3 を繰り返し, 指導を行う.

4 今後の課題

実装したプロトタイプの評価実験計画, 現段階での提案システムの課題, および拡張したシステムの評価実験計画について述べる.

4.1 評価実験計画

今後, 提案システムの実用性と有効性を確認するために評価実験を行う. 評価実験の目的として, 提案システムが実際の楽器学習者に対してどれだけ効果的な支援を提供できるかを検証する. タスクとして, 3.5 節中の 1 ~ 3 で示した一連の指導を実施する. 調査項目として, 演奏技術の向上度や学習効果の定量的な評価, およびモチベーション向上度などを調査する. また, アンケートやインタビューを通じて, ハンドトラッキングのフレームレートや遅延がシステムに影響があるか, 学習者の意見や感想を収集し, システムの利点や課題を調査する.

また, 評価実験では指導者側からの評価を考慮する必要がある. 指導者の評価として, システムを通じて学習者に適切なフィードバックを提供できるか, 遠隔指導の実現がどの程度実用的かなどを評価項目として取り入れる. 指導者の視点からの評価を通じて, 指導者と学習者の双方にとって最適なシステムを調査する.

4.2 提案システムの拡張

今後の実装として, AR アバタ, および EMS の実装の拡張, ピアノ位置調整の簡素化, およびピアノ以外の楽器の拡張を行う.

4.2.1 AR アバタの実装

現段階での AR アバタの実装では, 指導者の手を AR 表示したのみであり, その他の身体部位の AR 表示は行っていない. 今後の実装として, 指導者の上半身, および顔の AR アバタ実装を行う必要がある. 具体的な実装方法として, 指導者の前にトラッキン

グ用のカメラを設置し、上半身と顔のトラッキング情報を取得し、学習者側へ情報を送る。トラッキングはハンドトラッキングと同様に Mediapipe を用いる。そして、そのトラッキング情報をもとに Unity にて AR アバタの実装を行う。

指導者の上半身および頭を AR アバタとして提示する場合、それを学習者に対してどのような配置で提示するかは検討が必要である。例えば、指導者を学習者の対面に表示することによって、手の動きや姿勢、体を使った表現などを、鍵盤との視線の遷移を小さくして確認できるようにする、指導者を学習者の隣に表示させることによって、指導者の存在感を高め、学習者の集中力を上げる、といった表示方法が考えられる。また、それぞれの配置に対して、EMS をどのように組み合わせるかについても検討が必要である。

4.2.2 EMS チャンネル数の拡張

現段階での提案システムの実装では、EMS の出力チャンネルが 2 チャンネルのみである。これは EMS 生成器から出力されるチャンネル数が 2 つであるためである。したがって、両手すべての指を動かすにはチャンネル数を 10 チャンネルに拡張する必要があるため、EMS 生成器を 1 台から 5 台に増やして制御する。

4.2.3 指の動きの提示

本論文では、指導者の指の動きを、MIDI 情報から間接的に取得している。また、指の動きの提示として MCP 関節の屈曲に限定している。音階の打鍵では MCP 関節が主として屈曲するが、和音の打鍵では MCP 関節に加えて手首や DIP 関節が屈曲する。したがって、より正確な指の動きの提示を行うため、筋電位を用いて指導者の指の動きの情報を取得し、MCP 関節に加えて手首や DIP 関節の動きの提示を行う。

4.2.4 ピアノ位置調整

AR 上のピアノと学習者側のピアノと位置調整を行う場合、学習者側のピアノを移動させることによって調整する。これは Unity から Hololens2 への出力として、Holographic Remoting を用いているが、この出力方法では、Hololens2 の初期位置を前回のアプリケーションが終了した位置としている。したがって、アプリケーション終了時にはあらかじめ決まった場所に置く必要があり、この置き場所が変わると都度調整する必要がある。このような調整方法では、学習者側のピアノが電子ピアノのような移動しやすいピアノに限定され、グランドピアノのような移動が難しいピアノでは適用できない。

この課題の対策として、QR コードを用いたトラッキングによる解決方法が考えられる。HoloLens 2 で

は、内蔵カメラから QR コードを検出し、現実空間での QR コードの位置を原点位置に設定することができる [16]。したがって、原点位置に AR、および現実空間のピアノの位置を設定することによって、自動的に AR 空間と現実空間の位置合わせができる。

5 まとめ

本研究では、提案システムとして指導者の AR アバタと EMS を組み合わせたピアノ演奏指導支援を提案した。そして、そのプロトタイプとして、指導者の手を模した AR ハンドと EMS を組み合わせたシステムを実装した。また、本プロトタイプを用いた、具体的な指導方法を示した。

今後の課題として、提案システムの実用性と有効性を確認するために、評価実験を行い、システムが学習者の演奏技術向上や学習効果促進にどの程度効果的かを調査する。また、学習者の評価に加えて指導者側にとって提案システムが実用的であるか調査し、AR アバタ、および EMS のチャンネル数の拡張を行う。

まとめとして、提案システムは遠隔楽器指導における新たな可能性を開拓するものであり、今後の改善によってより学習者に対する学習支援が実現できると考えられる。

参考文献

- [1] Casio. LK-520 — CASIO. <https://www.casio.com/jp/electronic-musical-instruments/product.LK-520>. (Accessed on 09/06/2023).
- [2] J. Chow, H. Feng, R. Amor, and B. C. Wünsche. Music Education Using Augmented Reality with a Head Mounted Display. In *Proceedings of the Fourteenth Australasian User Interface Conference - Volume 139*, AUIC '13, pp. 73–79, AUS, 2013. Australian Computer Society, Inc.
- [3] CVzone. CV Zone - Computer Vision Zone. <https://www.computervision.zone>. (Accessed on 09/06/2023).
- [4] J. A. Deja, S. Mayer, K. Čopíč Pucihař, and M. Kljun. A Survey of Augmented Piano Prototypes: Has Augmentation Improved Learning Experiences? *Proc. ACM Hum.-Comput. Interact.*, 6(ISS), nov 2022.
- [5] A. Ebisu, S. Hashizume, K. Suzuki, A. Ishii, M. Sakashita, and Y. Ochiai. Stimulated Percussions: Method to Control Human for Learning Music by Using Electrical Muscle Stimulation. In *Proceedings of the 8th Augmented Human International Conference*, AH '17, New York, NY, USA, 2017. Association for Computing Machinery.
- [6] C. A. Fraser, J. O. Kim, A. Thornsberry, S. Klemmer, and M. Dontcheva. Sharing the

- Studio: How Creative Livestreaming Can Inspire, Educate, and Engage. In *Proceedings of the 2019 Conference on Creativity and Cognition*, CC '19, pp. 144–155, New York, NY, USA, 2019. Association for Computing Machinery.
- [7] S. Glickman, B. Lee, F. Y. Hsiao, and S. Das. Music everywhere - augmented reality piano improvisation learning system. In *New Interfaces for Musical Expression*, 2017.
 - [8] N. Granieri and J. Dooley. Reach, a keyboard-based gesture recognition system for live piano sound modulation. 06 2019.
 - [9] I. H2L. UnlimitedHand について - Unlimited-Hand. <http://unlimitedhand.com/>. (Accessed on 09/06/2023).
 - [10] F. Huang, Y. Zhou, Y. Yu, Z. Wang, and S. Du. Piano AR: A Markerless Augmented Reality Based Piano Teaching System. In *2011 Third International Conference on Intelligent Human-Machine Systems and Cybernetics*, Vol. 2, pp. 47–52, 2011.
 - [11] B. István and S. Dieter. Augmented Reality Agents in the Development Pipeline of Computer Entertainment. *Entertainment Computing - ICEC 2005*, pp. 345–356, 2005.
 - [12] J. Karolus, A. Kilian, T. Kosch, A. Schmidt, and P. W. Wozniak. Hit the Thumb Jack! Using Electromyography to Augment the Piano Keyboard. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference*, DIS '20, pp. 429–440, New York, NY, USA, 2020. Association for Computing Machinery.
 - [13] K. Labrou, C. H. Zaman, A. Turkyasar, and R. Davis. Following the Master's Hands: Capturing Piano Performances for Mixed Reality Piano Learning Applications. 2023.
 - [14] R. Liu, E. Wu, C.-C. Liao, H. Nishioka, S. Furuya, and H. Koike. Synchronized Hand Difference Visualization for Piano Learning. In *ACM SIGGRAPH 2022 Posters*, SIGGRAPH '22, New York, NY, USA, 2022. Association for Computing Machinery.
 - [15] Microsoft. Microsoft HoloLens — Mixed Reality Technology for Business. <https://www.microsoft.com/en-us/hololens>. (Accessed on 09/06/2023).
 - [16] Microsoft. QR コード追跡の概要 - Mixed Reality — Microsoft Learn. <https://learn.microsoft.com/ja-jp/windows/mixed-reality/develop/advanced-concepts/qr-code-tracking-overviewdevice-support>. (Accessed on 09/06/2023).
 - [17] D. Molero, S. Schez-Sobrino, D. Vallejo-Fernandez, C. González-Morcillo, and J. Albusac. A novel approach to learning music and piano based on mixed reality and gamification. *Multimedia Tools and Applications*, 80:165–186, 2020.
 - [18] W. Molloy, E. Huang, and B. C. Wünsche. Mixed Reality Piano Tutor: A Gamified Piano Practice Environment. In *2019 International Conference on Electronics, Information, and Communication (ICEIC)*, pp. 1–7, 2019.
 - [19] A. Niijima, T. Takeda, K. Tanaka, R. Aoki, and Y. Koike. Reducing Muscle Activity When Playing Tremolo by Using Electrical Muscle Stimulation to Learn Efficient Motor Skills. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 5(3), sep 2021.
 - [20] R. Nith, S.-Y. Teng, P. Li, Y. Tao, and P. Lopes. DextrEMS: Increasing Dexterity in Electrical Muscle Stimulation by Combining It with Brakes. In *The 34th Annual ACM Symposium on User Interface Software and Technology*, UIST '21, pp. 414–430, New York, NY, USA, 2021. Association for Computing Machinery.
 - [21] M. Pfeiffer, T. Duente, and M. Rohs. Let Your Body Move: A Prototyping Toolkit for Wearable Force Feedback with Electrical Muscle Stimulation. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '16, pp. 418–427, New York, NY, USA, 2016. Association for Computing Machinery.
 - [22] Pygame. [pygame.org/news](https://www.pygame.org/news). <https://www.pygame.org/news>. (Accessed on 09/06/2023).
 - [23] Sainitas. For regeneration and pain relief SEM 43 - Digital EMS/TENS — Sanitas Onlineshop. <https://sanitas-online.de/en/p/sem-43-digital-ems-tens-1>. (Accessed on 09/06/2023).
 - [24] N. Sakai, M. C. Liu, F.-C. Su, A. T. Bishop, and K.-N. An. Motion Analysis of the Fingers and Wrist of the Pianist. *Medical Problems of Performing Artists*, 11(1):24–29, 1996.
 - [25] C.-H. Sun and P.-Y. Chiang. Mr. Piano: A Portable Piano Tutoring System. In *2018 IEEE XXV International Conference on Electronics, Electrical Engineering and Computing (INTERCON)*, pp. 1–4, 2018.
 - [26] A. Takahashi, J. Brooks, H. Kajimoto, and P. Lopes. Increasing Electrical Muscle Stimulation's Dexterity by Means of Back of the Hand Actuation. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA, 2021. Association for Computing Machinery.
 - [27] E. Tamaki, T. Miyaki, and J. Rekimoto. PossessedHand: Techniques for Controlling Human Hands Using Electrical Muscles Stimuli. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pp. 543–552, New York, NY, USA, 2011. Association for Computing Machinery.
 - [28] 雅展 三浦. ピアノ演奏の初期教育における指の構えに関する考察. 音楽音響研究会資料, 39(8):37–42, 02 2021.

EmoCast: 顔への表情変化情報投影による話し相手の印象向上システム

浅野 恭志 *大西 鮎美 †寺田 努 †塚本 昌彦 †

概要. 会話時の表情は話し相手の印象に大きな影響を与えており、表情が豊かでないと話し相手に悪い印象を与える。本研究では、静止対象物にグレースケールの動き情報を投影して対象物が動いているように錯覚させる技術を応用し、顔のパーツが動く会話中などでも運動パターンを投影できるウェアラブル表情拡張システム EmoCast を提案する。提案システムでは、カメラで取得した顔映像をスマートフォンに送信して顔の 3D モデルを作成し、この 3D モデルにあらかじめ作成した運動パターンをマッピングして、プロジェクタで本人の顔に投影する。これにより、無表情からの表情の生成、表情の増強、表情の打消しといった機能が実現できる。評価実験により、動きのある顔への運動情報の投影で表情がより変化したと錯覚させることができ、会話時に提案システムを使用することで印象が良くなることが示唆された。

1 はじめに

会話時の表情は、話し相手の印象に大きな影響を与える。Mehrabian は、会話時に印象を与える要素として会話内容が 7%，音声トーンや話し方が 38%，そして表情や身振りが 55% であると示した [4]。このように、表情や身振りは話し相手の印象に大きな影響を与えており、表情が豊かでないと話し相手に悪い印象を与えててしまう。

会話時の表情を変化させる先行研究では、顔の一部をディスプレイで覆ったり [8][10]、顔の部位にプロジェクションマッピングを行って塗りつぶしたりすることで [1]、理想的な表情を提示するものが多い。しかし、顔の一部を覆ったり、部分的にでも口や鼻などの部位をプロジェクションによって塗り替えたりすると、細かい表情の変化を塗りつぶしてしまうため、投影対象者の顔自体が話し相手から見えづらくなる。

静止対象物にグレースケールの動き情報を投影して対象物が動いているように錯覚させる変幻灯がある [9]。この手法は、ユーザの顔の部位を塗りつぶさずに表情を変えるのにも有効であると考えられるが、静止物体への投影のみを対象としており、会話中などの動きのある顔への投影を対象としていない。

そこで本研究では、動きのある顔のパーツに追従して運動パターンを投影し、会話時の表情を豊かに見せることで話し相手への印象を良くするウェアラブル表情拡張システム EmoCast を提案する。提案システムでは、カメラで取得した顔映像から現在の顔の 3D モデルを作成し、この 3D モデルにあらかじめ生成した運動パターンをマッピングして、プロジェクタで投影する。これにより、図 1 の例のような無表情からの表情の生成、表情の増強、表情の打消しといった機能が実現できる。運動パターンを投影している性質上、図のような静止画ではシステムの機能が把握しづらいため、動画¹を確認いただきたい。



図 1: 提案システムの機能の一例

ジェクタで投影する。これにより、図 1 の例のような無表情からの表情の生成、表情の増強、表情の打消しといった機能が実現できる。運動パターンを投影している性質上、図のような静止画ではシステムの機能が把握しづらいため、動画¹を確認いただきたい。

本稿では提案システムのプロトタイプを実装し、評価実験として、顔に運動情報を付与して動きのある顔のパーツがより動いて見えたか、その結果会話時の印象が良くなったかを確認する質問紙調査を実施した。結果より、動きのある顔に運動情報を投影することでより表情が変化したと錯覚させることができ、会話時に提案システムを使用することで印象が良くなることが示唆された。

Copyright is held by the author(s).

* 神戸大学工学部

Faculty of Engineering, Kobe University

† 神戸大学大学院工学研究科

Graduate School of Engineering, Kobe University

¹ <https://youtu.be/h2qkCgmjvMA>

2 関連研究

2.1 顔に表情を提示する研究

表情を豊かにするために、顔の上に表情を提示する研究は多数行われている。石井らの HappyMouth は、ユーザの選択した表情や会話内容を提示するマスク型ディスプレイで、会話の促進を目的としている [10]。また Kumasaki らは、ユーザの表情の印象を変化させるマスク型ディスプレイ TransEmotion を提案している [8]。このデバイスはユーザの口の形や動きに同期してマスク型のディスプレイに CG で生成した口唇を表示し、対話相手がもつユーザの印象を変化させることを目的としている。しかし、これらのデバイスは顔の一部をディスプレイで覆うため、ユーザの表情を直接見せず、あらかじめ用意した映像を表示するため、投影対象者の顔自体が話し相手から見えづらくなる。

そこで本研究ではプロジェクタを用いて表情変化の運動情報のみを顔に投影することで、ユーザの顔の一部を覆うことなくユーザの表情を変化させ、相手に与える印象を良くするデバイスを提案する。

2.2 顔へのプロジェクションマッピング

プロジェクタを用い、表情変化を支援する研究はいくつか存在する。Bermano らが提案する Make-UpLamp は、舞台で演じる俳優や女優の外見を変化させるシステムであり、顔の特徴やしわを強調するようなエフェクトを使用できる [1]。また渡辺らは、高速カラープロジェクタと高速カメラを組み合わせて高速に顔にプロジェクションマッピングを行い、バーチャルメイクなどに応用可能なシステムを提案している [11]。

これらの研究は顔のパーツの動きに追従してプロジェクションマッピングを行う点で本研究と共通するが、しわや顔の部位といった、新たな顔のパーツの映像をユーザの顔に重ねて投影し、ユーザの見た目を別のものに塗り替えるため、細かな表情の変化が塗りつぶされてしまう。また、鼻や口といった顔の部位を新たに塗り替えるために、適切な色をもつ解像度の高い映像を投影できる高性能な大型プロジェクタが必要となってしまい、システムを利用できる場所が限られる。よって本研究で提案するデバイスは、運動情報のみ投影することで、ユーザの細かな表情の変化を塗りつぶすことなく表情を豊かにすることを目指し、さらに会話で使用することを想定しているため、構成デバイスを持ち運び可能な大きさにする。

2.3 運動情報を投影して表情を変化させる手法

運動情報をユーザに投影し、ユーザの顔の部位を塗り替えずに投影する技術として、吹上らの変幻灯が有効であると考えられる [9]。変幻灯は、静止

対象物にグレースケールの運動パターンを投影し、静止対象物が動いているかのように錯覚させる技術である。この技術の仕組みは、色や形、動きの情報を独立に分析した後に脳内で統合する人間の視覚特性に基づいており、運動情報のみをプロジェクタから投影し、静止対象物の色や形を塗り替えることなく動いているように知覚させる。この技術を用いることで、顔の部位の色や形を塗りつぶすことなく、表情が変化しているように見せることができる。また、運動情報は空間解像度が低くても認知でき、グレースケールでも違和感が生じないため、高性能でない小型プロジェクタでも投影する映像に違和感が生じにくい。

変幻灯は静止物体への投影のみを対象としているが、本研究では会話中など動きのある顔への投影を対象とする。そのため、顔のパーツの動きに追従して運動パターンを投影できるよう、リアルタイムに顔映像を処理し、顔の変化に応じて映像をマッピングして投影する必要がある。本研究では、動きのある会話時の顔にリアルタイムで表情の運動パターンを投影して表情が変化しているように錯覚させ、会話時の印象を良くすることを目指す。

3 提案システム

提案システムでは、表情が豊かに見える映像を会話時の顔に投影して、話し相手への印象を良くすることを目指す。提案システムは、以下のような機能に対応できるように設計する。

表情の生成

無表情の顔に特定の表情を生成する

表情の増強

表情が少ししか変化しない場合に表情を増強する

表情の打消し

出すべき表情が間違っている場合に表情をキャンセルする

提案システムの使用例を以下に示す。

- ・ 顔の表情がでない人が、周りの人に面白く思っていると思わせたい状況で、特定の表情を生成する機能を活用することで、その人が笑顔で話を聞いているように周りの人に感じさせる
- ・ 顔の表情がでにくい人が、話し相手に感じている印象をより深く伝えたい状況で、表情を増強する映像を活用することで、その人が話をよく聞いているように周りの人に感じさせる
- ・ 適切な顔の表情を作ることができない人が、笑ってはいけない状況で笑ってしまった時に周りに笑っていることを隠したい状況で、表情を打消す機能を活用することで、その人が

EmoCast: 顔への表情変化情報投影による話し相手の印象向上システム



図 2: 作成したデバイス

あまり笑っていないように周りの人に感じさせる

3.1 システム構成

本研究では、会話における話し相手への印象を良くすることを目的としている。会話で用いるためには、デバイスは持ち運びできるよう小型化する必要がある。また、会話時は特に口元が大きく動くため、顔のパーツの動きに追従して、表情を強めたいタイミングでリアルタイムに表情を増強する必要がある。図 2 に提案システムのプロトタイプの構成を示す。プロトタイプがヘルメット型であるのは、顔全体に投影を行うことで様々な表情を拡張するため、部分的な顔のパーツの拡張に限定すればシステムは小型化できる。システムはヘルメット型デバイスとスマートフォンで構成されており、ヘルメットに小型プロジェクタと Web カメラを固定している。ユーザは小型プロジェクタや Web カメラをスマートフォンに接続し、システムを利用する。

提案システムの処理の流れを図 3 に示す。スマートフォンのストレージに、あらかじめ作成した表情の運動パターンと、映像内の顔の特徴点を表す UV 座標を格納する。システムが起動すると、デバイスに搭載したカメラが顔の映像を撮影し始める。スマートフォン内では、撮影した顔の映像を機械学習モデルに渡して顔の特徴点を取得し、3D モデルを作成する。3D モデルの作成と並列して、カメラやマイクなどから音声や表情などの情報を受け取る。これらの情報を分析し、トリガとなる情報を捉えた際に、あらかじめ作成した運動パターンを、顔の特徴点の UV 座標を用いて顔の 3D モデルにマッピングする。

表情の運動パターンの作成方法

提案システムで用いる運動パターンは、無表情からある表情に変化する様子をカメラで撮影し、無表情画像との差分を取って作成する。この方法は、グレースケールの静止画像と静止画像を動的に変形した画像列との差分を計算して作成する変幻灯の運動誘導パターンを参考にしている。

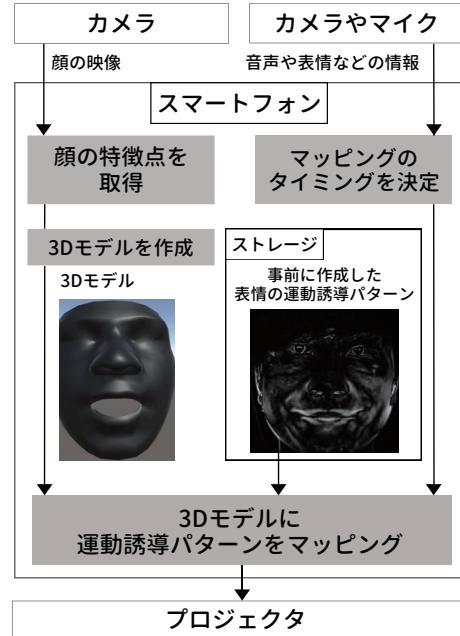


図 3: 提案システムの処理の流れ

各感情に対応する表情の作成には、Sato らの日本人の感情と表情の関係に関する研究にて提示されている、基本 6 感情(怒り・嫌悪・恐怖・喜び・悲しみ・驚き)に対応するシナリオをもとにした表情生成方法を採用した [5]。図 5 のように、基本 6 感情に対応する表情に変化する様子をカメラで撮影し、無表情の画像との差分を取り、図 4 に示すような基本 6 感情に対応する表情の運動パターンとした。

さらにより効果的な映像の作成を目指し、表情の変化によって大きく動き、運動情報の適切な差分が取れない輪郭や顔の部位といった投影時に白くなりすぎる箇所の映像を削除した。また、3D モデルの輪郭部分をユーザの顔のサイズよりも大きくしてから投影することで顔全体にプロジェクタの光が当たるようにし、プロジェクタの投影光による違和感を軽減した。

投影モデルの作成方法

会話時には様々な顔のパーツが動くため、顔のパーツの動きに追従して映像を投影する必要がある。装着者の顔と投影する映像の間に大きなズレがあると、顔の動きと投影する映像が別々で知覚されてしまう [9]。そのため、投影モデルを作成する処理は高速に行う必要がある。

提案システムでは、機械学習モデルで顔の特徴点を取得し、Unity 内で顔の 3D モデルを作成する。スマートフォンを用いた投影モデル作成を高速化するため、機械学習モデルに MediaPipe の FaceMesh



図 4: 笑顔の運動パターン

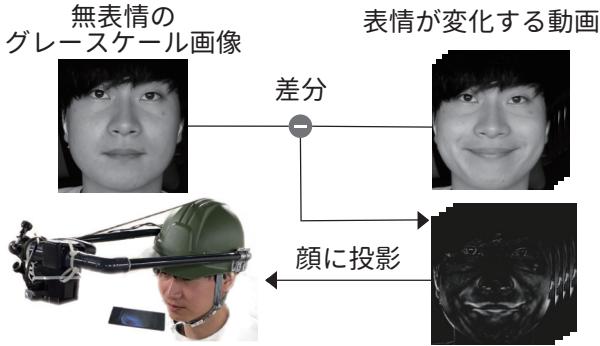


図 5: 表情の運動パターンの作成方法

を用い [3], TensorFlowLite の GPUDelegate という高速化手法を取り入れた。FaceMesh はオープンソースの機械学習ライブラリである MediaPipe のモデルの 1 つであり、顔の 478 個の特徴点を高速に取得できる。このモデルはエッジデバイスに最適化されているため、今回使用した Android 端末上でも高速に顔の特徴点を取得できる。また、GPUDelegate はエッジデバイス内の GPU を活用して推論時間を短縮する高速化手法であり、XNNPACK や GPUDelegate などの高速化手法を比較し、最も高速だった GPUDelegate を取り入れることで、顔の特徴点の取得にかかる時間を短縮した。

提案システムでは、3D モデルの作成後、運動パターンのフレーム毎に 3D モデルの顔の特徴点と運動パターンの顔の特徴点をリアルタイムでマッピングして投影モデルを作成する。フレーム毎に高速に処理する必要があるため、運動パターンの顔の特徴点はあらかじめ保存した。FaceMesh を用いて取り出した運動パターンの顔の特徴点を運動パターンの映像とともにバイナリファイルとして Android のストレージに保存し、リアルタイムなマッピング処理を軽量化した。

表情を増強する機能

提案システムを使用することで様々な機能を実現できるが、本研究では以下の機能を実装した。

- ユーザが希望するタイミングでシステムを動かし、特定の表情を生成する
- ユーザの顔映像を分析し、表情を増強する

この機能のうち、表情を増強する機能について説明する。ユーザの表情を増強する際に投影する表情とユーザの表情が一致しない場合、ユーザの意図しない印象を話し相手に与えてしまい、誤解を生じさせてしまう。そこで、ユーザの表情が変化したタイミングで素早くユーザの表情を判別して適切な運動パターンを投影する必要がある。

表情の判別方法に FACS というものがある [2]。FACS(Facial Action Coding System) は Action Unit という表情筋動作の単位の組合せを用いて感情を推定する方法であり、高精度で表情に基づく感情推定ができる。一方で、日本人の顔面運動が西洋人に比べて全体的に減少する文化的な差異が報告されているように [6]、文化圏の違いや顔の筋肉の個人差によって、表情と感情の対応関係が変化してしまう。

そこで提案システムでは、FACS に Sato らが提示した日本人の感情と ActionUnit との関係に関するデータ [5] を組み合わせることで、日本人の表情と感情の対応関係に即した手法を実装した。特定の感情に対応する Action Unit が動作した際に Sato らによって示された表情の強度を合算する。表情の強度を各強度の総和で除算し、基本 6 感情のうちのどれに該当するかを判別する。

4 評価実験

提案システムで顔に運動情報を付与し、以下の 2 項目を評価した。

項目 A 動きのある顔がより動いて見えたか

項目 B 会話時の印象がよくなったか

4.1 実験方法

装着者は表情を作るのが苦手な 20 代男性で、170 lx の照明環境下で提案デバイスを装着させた。装着者には実験中、常に感情を出さないでいるように伝えられた。今回の評価実験では、装着者とは異なる人物(本論文の第一著者)の顔から作成した笑顔の運動パターンを投影し、笑顔の表情を増強する効果を被験者に見せた。各被験者が見る装着者の状態は、会話あり・なし、システムによる投影あり・なしを組み合わせた計 4 状態で、見せる状態の順番は被験者ごとにランダム化した。実験で提示した 4 状態については、

EmoCast: 顔への表情変化情報投影による話し相手の印象向上システム



図 6: 実験時の顔の様子



図 7: 実験の様子

動画²を確認いただきたい。装着者の会話なしと会話ありの顔の例を図 6 に示す。顔のパーツの動きのない会話なしの状態と、動きのある会話ありの状態を比較し、錯視の効果を定量的に調べる。実験の様子を図 7 に示す。被験者は装着者と机を挟んだ 170 cm の距離で装着者の表情を見せ、表情や印象の変化について表 1 に示す質問紙調査を行った。被験者は 20 代の男女 10 名である。

項目 A の顔に運動情報を付与したことで動きのある顔がより動いて見えたかは、表 1 の Q1 の笑顔度を用いて評価する。今回の評価実験では、笑顔の運動パターンを投影し、笑顔の表情を増強する効果を評価するため、装着者の笑顔の表情がシステムによってどれだけ増強したかを比較する。

項目 B の印象がどのように変化したかは、表 1 の Q2~Q6 を用いて評価する。Q2~Q5 を用いて印象が良くなったかを調査し、Q6 でどのような印象を抱いたかを調べる。Q2 と Q4 は会話に関する内容のため、会話ありの状態でのみ質問した。

システムのリアルタイム性や問題点は、表 1 の Q7 を用いて評価する。システムのリアルタイム性が十分であるかを評価するために、違和感を覚えた箇所を調査した。

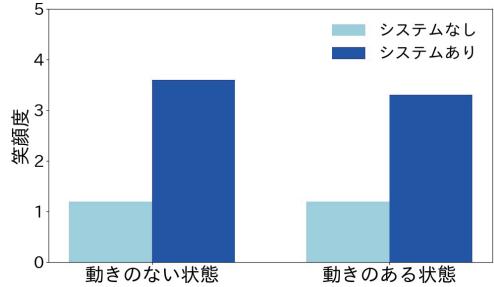


図 8: 笑顔度の評価結果

4.2 実験結果

項目 A の動きのある会話時の顔が笑顔に見えたかについて、Q1 の回答で得られた笑顔度の平均値を図 8 に示す。会話なしにおいて、システムなしは 1.2、システムありは 3.6 であった。また、会話ありにおいて、システムなしは 1.2、システムありは 3.3 であった。よって、動きのある顔でも運動情報を投影することにより動いて見え、動きのある顔でも、動きのないときと同様な錯視の効果が得られることが示唆された。

項目 B の印象の変化については、Q2~Q5 の回答で得られた回答の平均値を 2 に示す。全ての質問で動きの有無に関わらず、システムなしよりもの方が点数が高かった。したがって、システムを利用した装着者の印象が良くなったことが示唆された。

装着者に投影されている映像の違和感について、違和感を覚えたと答えた被験者は、会話なしでは 10 名中 6 名、会話ありでは 4 名だった。そのように回答した理由について、適切なタイミングで投影することが違和感を減らしたと答えた被験者が複数いた。このことから、会話時の動きのある顔に投影した方がより自然であることが示唆された。また、作成した映像が装着者の顔と異なっていると指摘した人はおらず、第一著者の顔で作成した映像でも装着者の顔に適切にマッピングされていると考えられる。

5 議論

今後、より違和感のない映像を投影するためにどのようなシステムを作成すべきかについて、被験者の回答から違和感を覚えた内容とその原因について議論する。

システムの速度について、口元の追従や瞬きのある目元などの映像のズレに違和感を覚えたと答えた人は 10 名中 1 名であり、適切な速度で装着者の動きに追従できていたと考えられる。一方で、プロジェクタの光量について、3 名が映像が白すぎることによる違和感を指摘した。これは運動情報以外の不要な情報もプロジェクタから投影されていたことが原因であると考えられる。より効果的な錯視を生む

² https://youtu.be/qG_dGsIvrNY

表 1: 質問紙調査の内容

項目	内容
A	Q1. 笑顔度はどうか (1. 全く笑顔に見えない, 5. 強い笑顔に見える)
	Q2. 話を聞いているように感じたか (1. 全く聞いていない, 5. とても聞いている)
	Q3. 愛嬌よく感じたか (1. 全くそう思わない, 5. とてもそう思う)
B	Q4. リアクションはどうだったか (1. 全くない, 5. とても大きい)
	Q5. よく笑う人と感じたか (1. 全く感じない, 5. 強く感じた)
	Q6. どのような印象を感じたか (自由記述)
Q7. 違和感を覚えたか (自由記述)	

表 2: 印象に関する評価結果

会話 システム	なし		あり	
	なし	あり	なし	あり
話を聞いているか	-	-	1.6	2.8
リアクション	-	-	1.5	2.7
話を聞いているか	1.3	3.4	1.1	3.1
よく笑う人か	1.1	3.2	1.4	3.0

ために、投影する映像の明るさや投影する部位を調整する必要があると考えられる。

装着者の装着感について、装着者にシステムを使用した感想を答えさせた。装着感については、デバイスの重量が大きく、1時間を超える長時間の装着が難しいとの指摘があった。提案手法は、様々な表情を表現するために顔全体に映像を投影することを選択しており、プロジェクタの位置が顔から離れてしまうために、装着者に重さを感じさせていた。より長時間の装着が可能であるデバイスを作成するため、今後は特定の表情に特化し、部分的な顔のパーティにのみ投影を行うことで小型にする。

装着者が感じる眩しさについて、提案デバイスでは目の部分を黒い映像にして、装着者が眩しくないように投影を行っていた。しかし、長時間の装着によって、プロジェクタと顔の位置関係が変わってしまい、目に白い映像が入るようにならざるを得ないという指摘があった。今後、プロジェクタから投影する映像の位置や角度を自動で機能する機能を実装し、プロジェクタと顔の位置関係が変わってしても、装着者にとって眩しくないデバイスの作成をする。

6 まとめ

本論文では会話時の印象を良くするために、表情が豊かに見える映像を顔に投影するウェアラブル表情拡張システム EmoCast を提案した。提案システムでは、カメラで取得した顔映像から現在の顔の3D モデルを作成し、この3D モデルにあらかじめ作成した運動パターンをマッピングして、プロジェクタ

で投影する。これにより、無表情からの表情の生成、増強、打消しといった機能が実現できる。提案システムの有効性を評価した結果、動きのある顔に対しても運動情報を投影することでより動いているように錯覚させることができ、会話時にシステムを使用することで印象が良くなったことが示唆された。しかし、装着者を観察した被験者が感じる違和感や装着者の長時間装着には課題があった。今後は、より効果的な錯視を生む投影方法や装着者の装着感の向上についてより詳しく調査し、日常利用できるようにシステムを改良していく。

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] A. H. Bermano, M. Billeter, D. Iwai, and A. Grundhöfer: Makeup Lamps: Live Augmentation of Human Faces via Projection, *Computer Graphics Forum*, Vol. 36, No. 2, pp. 311–323 (May 2017).
- [2] P. Ekman and W. V. Friesen: Facial Action Coding System, *Environmental Psychology & Nonverbal Behavior* (1978).
- [3] Google: MediaPipe | Google for Developers. <https://developers.google.com/mediapipe>. (Accessed 2023-08-04).
- [4] A. Mehrabian: Communication without Words, *Communication theory*, pp.193 – 200 (2017).
- [5] W. Sato, S. Hyniewska, K. Minemoto, and S. Yoshikawa: Facial Expressions of Basic Emotions in Japanese Laypeople, *Frontiers in psychology*, Vol. 10, No. 259, pp. 1–11 (Feb. 2019).
- [6] C. H. J. Tzou, P. Giovanoli, M. Ploner, and M. Frey: Are There Ethnic Differences of Facial Movements between Europeans and Asians?, *British journal of plastic surgery*, Vol. 58, pp. 183–195 (Mar. 2005).
- [7] T. Kawabe, T. Fukuiage, M. Sawayama, and S. Nishida: Deformation Lamps: A Projection Technique to Make Static Objects Perceptually

EmoCast: 顔への表情変化情報投影による話し相手の印象向上システム

- Dynamic, *ACM Trans. Appl. Percept*, Vol. 13, No. 2, pp. 1–17 (Mar. 2016).
- [8] R. Kumazaki and A. Inoue: Development and Evaluation of a Mask-type Display Transforming the Wearer's Impression, *In Proceedings of the 31st Australian Conference on Human-Computer-Interaction (OzCHI '19)*, pp. 568—571 (Dec. 2019).
- [9] 吹上大樹, 河邊隆寛, 西田眞也: 変幻灯一錯覚を利用した光投影による実物体のインタラクティブな動き編集—, 情報処理学会研究報告, Vol. 2017-CVIM-206, No. 8, pp. 1–6 (Mar. 2017).
- [10] 石井綾郁, 小松孝徳, 橋本直: HappyMouth: マスク型デバイスによる対面コミュニケーション能力の拡張, 情報処理学会研究報告, Vol. 2018-HCI-177, No. 7, pp. 1–7 (Mar. 2018).
- [11] 渡辺義浩: ダイナミックプロジェクトマッピングによる虚実融合, 日本画像学会誌, Vol. 61, No. 1, pp. 14–21 (Feb. 2022).

Taste Time Machine：飲食物を過去や未来の味に変える装置の実現に向けて

宮下 芳明*

概要. 本稿では熟成や酸化など、数日や数年にわたる「飲食物における味の時間変化」に着目した。実現可能性の検証として、熟成過程に焦点を当て、トマトとカレースープの味を日ごとに味覚センサで測定してモデル化した。そして、味物質や中和剤を添加する手法で、時間の順行・逆行両方向での味変化を、実食・味覚センサ測定の両面から検証して確かめた。未熟なトマトや作りたてのカレーの味を数日後の味に変えられるだけでなく、熟れたトマトや一晩おいたカレーの味を、以前の味に戻すこともある程度できるようになった。

1 はじめに

お酒やチーズなど、熟成・発酵によって、時間が経つと美味しくなる飲食物がある。逆に、腐敗・酸化によって、時間が経つほど風味が落ちる飲食物もある。「時間」とはいわば、飲食物を美味しくしたり不味くしたりする調味料なのかもしれない。

そしてもし、我々人間に「時間を自在に操る能力」があったとしたら、できたてのカレーの時間を1日進めたり、熟成年数の浅いウィスキーの時間を10年進めたり、熟れすぎて黒くなってしまったバナナの時間を1週間前に戻して食していることだろう。

本稿では、飲食物における味の時間変化に着目する。味覚センサによる実測と理論によって各味と時間の関係をモデル化し数式で表す。その式をもとに、現時点での食品の味と、設定した日時に推定される味との差を求め、味の逆算も行えるTTTV3で味を変える[3]。タッチパネルで時間軸をタップするだけで、通常ならば何日・何年も待たなければ得られない味を瞬時に得たり、風味が落ちた食品を新鮮なときの味に戻したりできることを目指す。

本稿では特に熟成に焦点を当て、トマトおよびカレースープの味の経時変化を味覚センサで測定し数式としてモデル化した。これにより、未測定の値や未知の時刻の味推定を行えるようになる。そのうえで、順行・逆行それぞれの時間方向での味変化を試み、実食と味覚センサ測定の両面から検証した。本稿の提案手法は、発酵や熟成などの体験的理理解、テイスティングの訓練、腐敗物の味の安全な学習に役立つ。さらに、時間を操ることによって賞味期限の概念を打破し、フードロス削減につなげることもできると考えている。

Copyright is held by the author(s).

* 明治大学

2 TTV3と味の減算

TTTVは、塩味や甘味などを呈する溶液をフィルム等に噴霧混合して、味を再現する味覚ディスプレイである[1]。TTTV2は、味溶液を飲食物に噴霧してその味を加え、飲食物の味を変える調味家電である[2]。TTTV3は、高性能チューブポンプ（RP-QIII A1.5S-3Z-DC3V）を20機搭載し、0.02ml単位で1000段階の体積の味溶液を混合して飲食物に滴下する機器である[3]。一味に対して複数の物質を用いることが可能で、たとえばカカオの味についての事例[4]では、酸味の度合いを揃えるだけでなく3種類の酢酸・乳酸・プロピオン酸を用いる配合で風味を近づけた。その結果、安価なコートジボワール産カカオを稀少なペルー産カカオの味に近づけることに成功している。味覚センサだけでなくプロのパティシエによる監修と調整も行い、EC2023でデモを実施し、約100名が体験し、デモ優秀賞・デモ推薦認定を受けた[5]。

TTTV2では「味の加算」のみを行うので、牛乳をカニクリームコロッケの味に変化させるなど、味が濃くなる方向での変化に限られていた。唯一の味の減算手段は、水を入れた無味タンクからの噴霧による希釈であった。しかしTTTV3では希釈以外にも多くの「味の減算」手法を開拓しており、アルカリ性の中和剤によって酸味を弱めたり、味の相互作用を利用したり、味覚修飾物質を使用したりする方法を提案している[3]。前述のカカオの事例では、変化前のコートジボワール産の苦味が強いが、これをうま味によるマスキングで抑制して味変化を実現している。また、白ワインを赤ワインの味に変える事例[6]では、味覚センサだけでなく酒類のプロによる評価および監修によるレシピを用いてEC2023でデモを実施し、約60名の方に比較・体験してもらい、デモ最優秀賞を受賞した[5]。赤ワインと比べると白ワインの方が酸っぱいが、これを中和剤による抑制、

甘味添加にともなうマスキングで打ち消し、タンニン酸の渋味を加えることで赤ワインの味に変化させることができている。

3 バナナジュースの熟成（予備実験）

本稿の構想の単純な予備実験事例として「バナナジュースの味を変化させる」コンテンツを TTTV3 用に試作した（図 1）。バナナジュースの入ったコップ（量は事前に計量）を置き、何日経過した味にしたいか、何日前の状態に戻したいかをタップすると、その味に変えられるソフトウェアである。



図 1. TTTV3 でバナナジュースの味を変化させる例。

この例を単純と称したのは、バナナの熟成においては甘味以外の味変化が少なく、甘味も単調増加であること、また、熟成に伴う甘味増加も純粋に糖量測定（の文献を読むこと）で解決できるからである。後述のトマトやカレーに比べれば遙かに容易なプロセスで味の推定と味の変化を実現できる。以下、このバナナの事例について味の推定と変化方法についてやや詳細に説明する。以降はこの基本説明を圧縮し、より記述が必要な部分に紙面を割く。

バナナの熟成は、バナナ自体から放出されるエチレンガスによって起こり、バナナの中の澱粉が单糖に分解されていく過程である。澱粉自体は、唾液のアミラーゼと反応させなければ無味であり、今回はバナナジュースなので唾液の影響は考慮せずに済み、甘味は単調増加すると考えた。野呂らは、バナナの熟成における糖量増加および澱粉量の減少について、糖分画操作のうえで定量調査している[7]。このデータでは、可食部 100gあたりの糖量が追熟開始 0 日目 7.5g → 3 日目 11.7g → 5 日目 12.1g と増加している。ここから、たとえば追熟開始 0 日目のバナナ 100g をジュースにしたとき、それを 5 日後の味に変化させたい場合は、 $12.1 - 7.5 = 4.6\text{g}$ のグルコースを添加するとよいと推測できる。ただ、1 日目の味にしたい場合そのデータはない。原理上 7.5g から 11.7g の間だと想像はできるが、近似式でモデル化する必要が出てくる。論文[7]を読むと澱粉量は 0 日目 12.05g → 3 日目 0.24g → 5 日目 0.2g となってお

り、糖として分解される澱粉の残量がほとんどなくなっている。このことから、線形近似でなく指数関数による飽和を示す近似式 $C = 7.5 + 4.719(1 - e^{-0.736t})$ を作成した（図 2）。この式に基づくと、 $t = 1$ のときに C は 9.96 となり、 $9.96 - 7.5 = 2.46\text{g}$ のグルコースを添加すればよいと推測される。時間を逆行させた場合の味変化についても実現可能である。甘味の減算を実現する方法として、キサンタンガム系とろみ剤（味やにおいがほとんどなく、唾液などによる影響も受けにくい）を添加した水による希釈を用いる。5 日目のバナナジュースの味を 3 日目の甘味にしたい場合、 $12.1 \times 100 \div 11.7 - 100 = 3.4\text{g}$ のとろみ剤添加水を入れることで同程度の甘味にできる（が他の味も希釈される）。

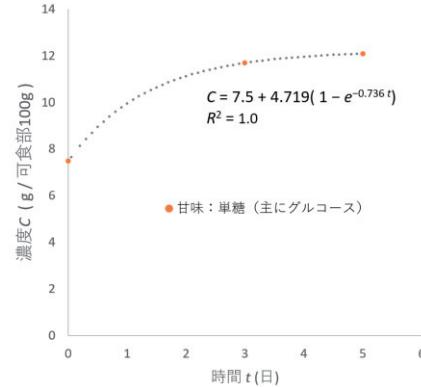


図 2. バナナの追熟による甘味変化と日数の関係。

[7]のデータに基づいて著者が作成。

著者が実際にシステムを試すと、時間順行方向は、熟れていくほど甘ったるくなるさまがよく表現できていると感じた。バナナジュースの色はそのままなので、どんどん黒ずんでいく視覚変化があればなお良いと感じた。時間逆行方向も概ねうまくいっているが、大きく味を減算する場合には他の味や香り、そして色も薄まるため、極端な設定ではバナナジュースの風味から離れていってしまう印象を得た。以上より、完全に自由自在・完全に味再現ができたわけではないが、「甘味を加減している」というより「時間を利用して味を操作している」という感覚のインターラクション体験を実際に得ることができた。

4 ウィスキーの熟成（予備実験）

本章では、ウィスキーの熟成の表現を試み、失敗した予備実験事例を述べる。ウィスキーについては成分分析のもので経年での成分変化を調べた研究が 70 年代からある[8]。以下の図 3 は、そのデータをプロットし著者が作成したものである。同論文ではモデル化も試みられており、総酸が線形近似、エステルとタンニンは累乗近似に沿うと主張されているため、それに沿って近似式を作成した。グラフを見

Taste Time Machine：飲食物の味を過去や未来に変える装置の実現に向けて

ても測定値に変動が少なく、妥当性の高いモデルであるようにみえる。

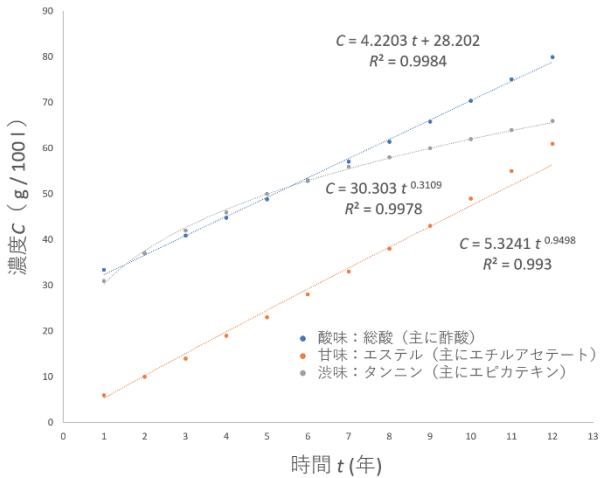


図 3. 成分分析に基づくウィスキーの経年変化。

[8]のデータに基づいて著者が作成。

これらの式に基づくと、たとえば4年熟成させたウィスキーを9年熟成させたように感じさせるためには、1リットルあたり0.21gの酢酸、0.24gのエチルアセテート、0.14gのエピカテキンを加えるとよいことになるが、本当にそうなるのか実際に作成して味見したところ、全くうまくいかないことがわかった。まず、ウィスキーの熟成と共に増える成分であるエピカテキンはそもそもそれほど苦味・渋味が強くない。同様に、エチルアセテートの甘味も弱く、ウィスキー全体への「味への貢献」が小さい物質であった。成分分析では重量変化の大きなものが優先して検出されるが、それは味覚変化の大きさを示すわけではない。つまり物質重量としての変化は大きくても、感覚変化が小さいものが存在するということである。重量変化は大きいのに無味なものもあるかもしれない。逆に、物質質量としての変化は小さくても、感覚変化として大きなものが存在する可能性も考えられる。こうした意味で、成分分析を行った論文は、参考にすることはできても、それだけで味を再現できるとは限らないと思われる。

こう考えると、前章のバナナについても「バナナは熟するほど甘くなる」という主観的・選択的な経験と、糖分増加量だけを調べた論文からの再現は極めて危ういといえる。甘味以外、たとえばうま味が本当に変化していないのかは調べてみないとわからない。このことから、味覚センサやパネラーによって味の変化を調べることの重要性が明らかとなった。

なお、ウィスキーについてはそもそも検証が難しいことがわかった。まず、市販のウィスキーにおける年数表記は熟成年数ではなく、様々な熟成年数の原酒を混ぜて造られたうちの最も若い原酒の熟成年

数を指していることがわかった。加えて、これらの原酒のブレンドにおいて人為的なデザインが介入していることが、酒造メーカーからのヒアリングでわかった。同じメーカー、同じブランドの、異なる年数のウィスキーを購入したところで、それらの違いは年数だけではないため、市販ウィスキーを用いた検証は全く無意味なことがわかった。

5 トマトの熟成（本実験）

バナナジュースの例は、時間順行が甘味の添加、時間逆行が希釈という単純な解釈での試作であった。しかし、一般的な熟成プロセスというのは、熟成に従って増えていく味と減っていく味が両方あり、そう簡単にはいかない。また、熟成過程によって「味として」増加するのか、減少するのか、それとも変化しないのかは実際に味覚センサや人間で判断する必要性が前章からわかった。本章では、複雑な熟成が行われる事例としてトマトを取り上げる。トマトの熟成は、エチレン合成から起こる様々な化学反応によると言われている。熟度が進むと糖分が増え、グルタミン酸は大きく増加する一方で、酸の成分が減っていくと言われている。ただし、物質としての変化はそうであっても、感覚（味覚）としての変化がどうなっているかは、あらためて測定・検証する必要がある。

そこでこの実態を測定すべく、同程度の熟成度合いのトマト「桃太郎」を12個選定し、1日経過ごとに測定していく(図4)。熟成条件は室温27~30°C、暗所、湿度67~69%である。1日目は、トマトの肩がまだ黄色く硬い状態であった。ザラザラしており、食してみると酸味がありフレッシュであった。2日目は、まだやや固い感じであった。食してみると酸味がやや減り、うま味が感じやすくなっているように思えた。3日目は、硬さもほどよく、食べ頃に思えた。食してみるとうま味も感じやすくなっていた。酸味は少しだけ残っていた。4日目は全体的に柔らかくなっていた。食してみると熟成感が強く、酸味をほとんど感じられなくなっていた。



図 4. 測定に用いたトマト「桃太郎」。

味覚センサ TS-5000Z による測定ではトマトの主要呈味であるうま味 (AAE, 先味) とそのうま味の余韻 (AAE, 後味), 酸味 (CA0, 先味) に影響があった。また, 甘味は屈折糖度計で測定し, それにも影響があった, よって, それらの軸でグラフを描画した結果を図 5 に示す。熟成段階ごとに酸味減少・うま味上昇という結果に表れている。甘味はトマト 1 日目を 100% として換算したものであるが, どの時点での測定結果もすべて 100% であったため, 甘味に変化がないことがうかがえた。これは一般的な見解と異なる。しかし, トマト「桃太郎」を室温で追熟し有機酸・糖を測定した論文[9]を調べたところ, 同様に糖度が変わらない結果が記されていた。さらに, 実際に食して甘味に変化がないことも確認したので, 測定に問題はなかったと思われる。

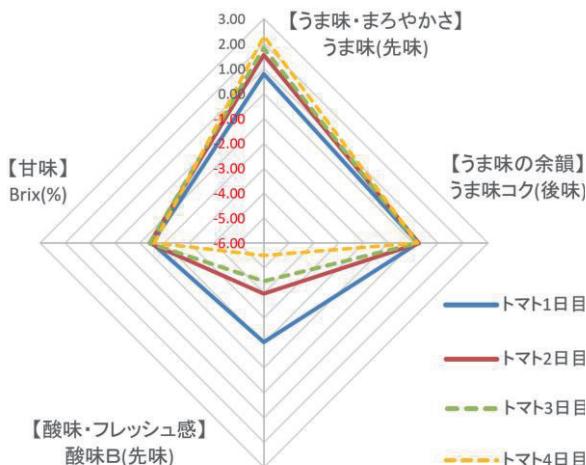


図 5. トマトの熟成過程の味覚センサ測定結果。

センサ値を添加量に変換するには, まずセンサ値の差を計算し, 各呈味濃度差を弁別できる濃度倍したうえで, トマト 1 日目が 100 となるように換算する (味推定値濃度差)。うま味としてのグルタミン酸添加量は, 論文 [10] から熟度ステージライトレッドのトマト「桃太郎」のグルタミン酸含有量 230.4(mg%) がわかるのでこれを乗じることで算出できる。ここから計算すると, 1 日目のトマト 100g の味を 2 日目, 3 日目, 4 日目のトマトの味にするには, それぞれ 34.6mg, 48.4mg, 73.7mg のグルタミン酸を添加するとよいことがわかる。次に, 酸味を抑制する中和剤 (炭酸水素ナトリウム) の添加量については, 合成トマト液に中和剤を入れたときの酸味変化量から実際のトマトの酸味変化分を換算し予測した。

これに基づくと, 添加すべき中和剤の量は, 55mg, 67mg, 87mg となる。これを図 6 のようにプロットすると, 対数の近似式が得られる。

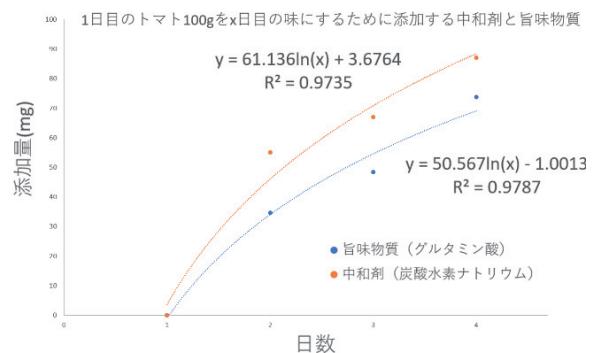


図 6. 1 日目のトマト 100g を x 日目の味にするために添加する中和剤とうま味物質.

味を戻す場合は, うま味物質を減らし酸味を増やすというアプローチになる。うま味の減算手段は水による希釀とし, 酸味はクエン酸を加えるとして計算を行う。たとえば 4 日目のトマトの味を, 1 日目のトマトの味にするには, 100gあたり水を 24g 加えてうま味を下げるができる。また, 味推定濃度差計算と論文[9]にあるトマト「桃太郎」のクエン酸含有量 (0.61%) から, 添加すべきクエン酸が 0.65g であると算出できる。これを図 7 のようにプロットすると, 対数近似の式を得ることができる。

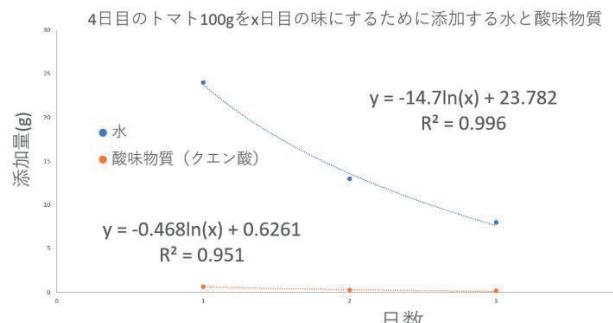


図 7. 4 日目のトマト 100g を x 日目の味にするために添加する水と酸味物質.

この配合で実際に時間を制御して体験してみると, 順方向はまさにトマトが熟成している感じが得られ, 実用に足る再現性だと思われた。逆行についても, 1日前, 2日前の味再現は満足がいくものの, 3日前の再現は甘味まで希釀されていることが気になった。今回, 甘味の測定値が変わらなかつたため甘味の添加を考えていなかったが, 希釀によって他の味 (甘味・塩味・苦味) も薄まるため, これらの味をむしろ加えれば, さらに良い再現性が得られると考えられる。またこれらの加算も現状のデータと論文等からの情報で補えるはずだと考えられる。

Taste Time Machine：飲食物の味を過去や未来に変える装置の実現に向けて

この味の類似性を味覚センサでも測定すべく、トマト1日目の味を中和剤とうま味添加によって4日目の味に近づけた再現サンプルと、トマト1日目、4日目の味を味覚センサで測定したものが図8である。中和剤の効果が大きく出ており、酸味が抑制されすぎているようにみえるが、著者が食したときにも酸味がほとんど感じられない内観を得ているので、問題はないと考えている。うま味については目標よりも高い値となっており、熟成度が本来よりもやや高くなっているように見えるが、食してみた感じとしては違いがわからなかった。

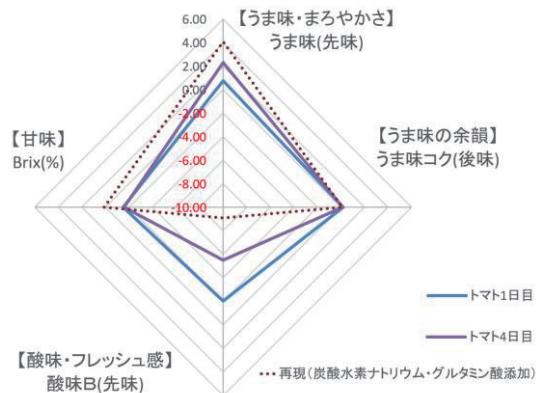


図8. トマトの熟成4日目の味再現結果。

6 カレースープの熟成（本実験）

本稿では最後に、おそらく最も複雑で難しいのはカレーなのではないかと考え、その味変化の測定と再現に挑戦した。カレーは一晩寝かせた方が美味しいとよく言われる。理由はジャガイモの澱粉によるろみであるとか、肉や野菜などから溶け出したうま味成分であるなど、具に起因する説明が多くなされている。しかし著者が試しにカレールーのみで作った具なしのカレーを、同濃度で12時間差で作り、同じ温度で味わったところ、それだけでも別物と思えるほどの味の違いがあった。そこで本稿ではまずカレーの味変化の基礎調査として、ルーのみによるカレーを対象にすることにした。

まず前章と同様に、実際に味覚センサでの測定を行う。バーモントカレーのカレールーを10倍希釀したカレースープをベースとした。腐敗を防止するため、以下のような手順を毎日行った。まず朝に蓋をして沸騰後5分加熱しサンプルを回収する。蓋をして常温静置し、4-5時間後(昼)、蓋をして沸騰後5分加熱・常温静置する。さらに8時間後(夜)、蓋をして沸騰後5分加熱・常温静置する。この手順を繰り返してサンプルを集め測定した。するとうま味(AAE、後味)、塩味(CT0、先味)とともに、意外に

も渋味(AE1、先味)にも変化があったのでその軸を使用することとし、図9に測定結果をまとめた。

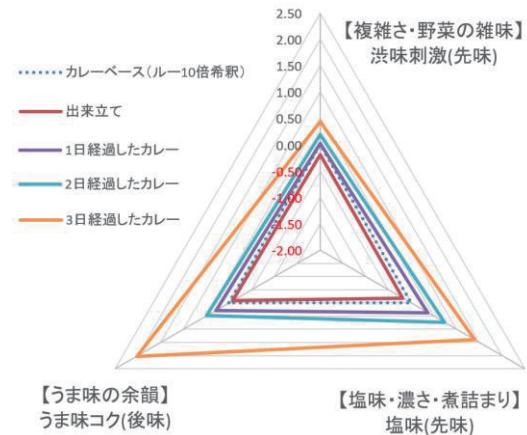


図9. カレースープの熟成過程の味覚センサ測定結果。

これを見ると、日が経過するごとに三角形が大きくなっていることがわかる。カレーの熟成の方向がもたらす変化は、煮詰まりによる渋味・塩味・うま味の余韻の味濃縮であると考えられる。この測定値を元に、添加する味物質の量を計算する。カレールー10倍希釀溶液にグルタミン酸ナトリウムまたはタンニン酸を入れ、そのうま味または渋味の変化量から実際のカレースープの変化分を換算し予測すると、できたてのカレースープの味を1日目、2日目、3日目の味にするには、100gあたりグルタミン酸ナトリウムを8mg, 14mg, 65mg、タンニン酸を3mg, 22mg, 50mgだけ入れればよい。また塩分分析値から塩化ナトリウムの添加量がわかり、各6mg, 379mg, 447mg 加えればよいことになる。これを図10のようにプロットして、各近似式を得た。

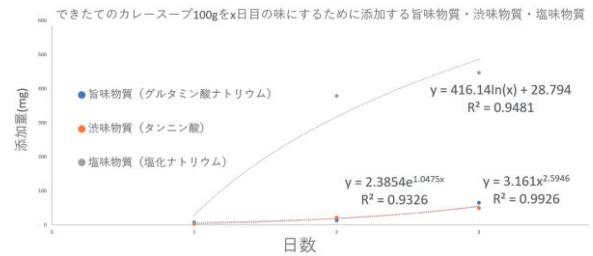


図10. できたてのカレースープ100gをx日目の味にするために添加するうま味物質・渋味物質・塩味物質。

逆に、3日目のカレースープの味を、1日前、2日前、3日前の味にするには、100gあたり水を5.46g, 52.6g, 51.6g 加える必要がある。これをプロットすると近似式が得られる（図11）。

この配合で実際に時間を制御して食してみると、順方向はかなり熟成している感じが得られた。「作りたてのカレーに、タンニン酸と食塩と味の素を入れ

れば熟成したカレーのようになる」という知見自体、家庭の料理で有用なノウハウになる気がするぐらいである。時間逆行についても味再現はある程度うまくいっているが、もともと味の薄いカレースープで行っているため、さらに薄くなても気にならない、ということのようにも思える。なお、再現品のほうが実際よりも香りが強い印象を得た。時間をかけた熟成を行う場合、香気成分が過熱で揮発していき、香りが弱まっていくが、再現品では希釈していくとまだその香りがやや残っている可能性がある。「希釈によって減退した味をそのまま加えることによって変わらないように見せる」という、前章での考察をここでも行うべきだと感じた。

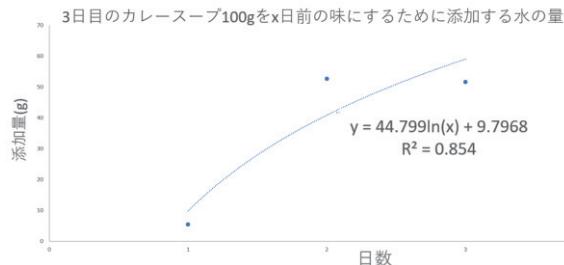


図 11. 3 日目のカレースープ 100g を x 日目の味にするために添加する水の量。

最後に、味の類似性を味覚センサでも確認した。ベース、3日経過したカレー、タンニン酸とグルタミン酸ナトリウムと塩化ナトリウムを加えて再現したものを作成し、味覚センサで測定したグラフを図 12 に示す。前述のように再現性が高いという内観を持っているので、データでも類似性が表れたと感じている。

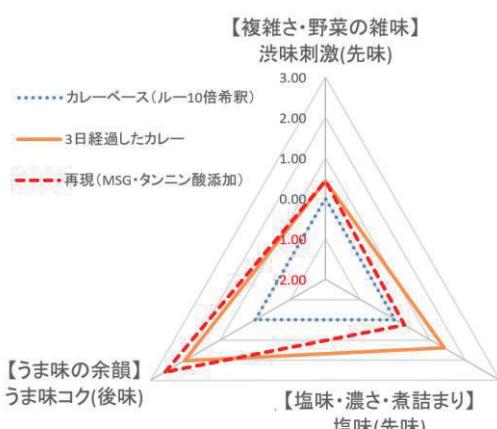


図 12. カレースープの熟成 3 日目の味再現結果。

7 まとめ・考察

本稿では日ごとに味覚センサで測定したデータをもとに、味の変化を連続関数の式としてモデル化した。これにより、測定点とは異なる時間の味推定

が行えるわけだが、モデル化のメリットは測定範囲を大きく超えたところの推定にあるはずなので、次のステップでは検証したい。

また本稿では、時間の順行・逆行両方向での味変化の可能性を、実食と味覚センサ測定値の両面から検証した。今後は、デジタル環境と物理環境がオーバーラップする AR 表現を導入することによって「時間を操る感覚」も増強していく[11]。

本稿の知見の応用範囲は広い。たとえば料理で用いる食材が未熟だったり熟れすぎていたりする場合に、味を調整して使用できる。今後この技術開発を推進した先には、消費期限の前に設定される「賞味期限」の概念を打破し、フードロス削減につなげることもできるかもしれない。他にも、発酵や熟成などを味覚体験的に理解させる教材が作成可能であると考えられる。たとえば、1日経つごとに味がどう変化するかを食べ比べる授業実習を行う場合、その何日も前から1日の時間差で同じものを用意し続ける必要がある。この、準備は面倒なはずなので、本稿での知見を活かして効率化できるはずである。今後、現場の教育者にヒアリングを行うなどしていきたい。他にも、味の熟成度合いを判別するテイスティング能力を訓練するためのアプリケーションや、腐敗物の味を学び食中毒を防ぐための学習コンテンツなどが作成可能である

料理や食に限らず、この物理世界では、変化の進行をただ待つしかない状況、取り消しがきかず諦めるしかない状況が多く起こる。だが、感覚世界ではこの限りではない。感覚をメディアとして自在に操作できれば、不可逆なことも可逆になるかもしれない。たとえば味覚メディア技術を駆使すれば、焼いた魚を刺身に戻すことだってできるかもしれない。

8 味覚メディア・味覚コンテンツの展望

味覚メディア技術の再現精度が向上してきたおかげで、「味覚コンテンツ」の研究も今後増えていくと著者は考えている。つまり、視覚メディア技術におけるコンピュータグラフィックス (CG) と同じように、今後はコンピュータティスト (CT) が振興するだろう。今日の CG 分野では、雪粒どうしの吸着や髪の毛のちぢれなど、細かな現象の再現がひとつひとつ研究テーマになっている。同じように CT 分野でも、一般的な調理行程はもちろんのこと、凍結や腐敗など様々な過程によって起こる細かな味変化現象を再現することが研究テーマになっていくだろう。最終的には、こうしたモデルが統合され、仮想空間上で多様な調理を行えるようになるとともに、その結果の「味のレンダリング」も正確に行えるようになるはずだと考えている。

Taste Time Machine：飲食物の味を過去や未来に変える装置の実現に向けて

参考文献

- [1] 宮下芳明. 液体噴霧混合式の味ディスプレイの試作, 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(WIIS2021)論文集, 2021.
- [2] 宮下芳明. TTVT2 (Transform The Taste and Visual appearance) : 飲食物の味と見た目を変える調味家電によるレイアウト, エンタテインメントコンピューティングシンポジウム 2022 論文集, Vol.2022, pp.143-150, 2022.
- [3] 宮下芳明, 村上崇斗, 大友千宙, 深池美玖. TTVT3 (Transform The Taste and reproduce Varieties): 産地や品種の違いも再現する調味機構と LLM による味覚表現. エンタテインメントコンピューティングシンポジウム 2023 論文集. Vol.2023, pp.236-243, 2023.
- [4] 彭雪儿, 深池美玖, 笠原暢仁, 村上崇斗, 吉本健義, 渡祥輝, 富張留斗, 宮下藏太, 川田健晴, 宮下芳明. 産地の異なるカカオの味の違いを定量化し純物質で再現する手法, 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.2023, pp.390-393, 2023.
- [5] エンタテインメントコンピューティング 2023
<https://ec2023.entcomp.org/award.html>
- [6] 金珉志, 村上崇斗, 宮下芳明. TTVT3 を用いたワインの味表現, エンタテインメントコンピューティング 2023 論文集, Vol.2023, pp.298-301, 2023.
- [7] 野呂哲, 葛西麻紀子, 山田綾子, 大中徹, 加藤陽治. バナナの澱粉. 弘前大学教育学部, Vol.105, pp.75-79, 2011.
- [8] Baldwin, S., & Andreasen, A. A. Congener Development in Bourbon Whisky Matured at Various Proofs for Twelve Years. *Journal of AOAC INTERNATIONAL*, Vol.57, No.4, pp.940–950, 1974.
- [9] 前澤重禮, 山田初男, 秋元浩一.トマト'桃太郎'追熟中の着色異常と収穫熟度および追熟温度との関係, 園芸学会雑誌, Vol.62, No3, pp.647-653, 1993-1994.
- [10] 高田式久, トマトのアミノ酸について, 日本家政学会誌, Vol.63, No.11, pp.745-749, 2012.
- [11] 藤澤秀彦, 宮下芳明. Taste-Time Traveller : 食品の時間を操る味覚 AR 装置. 第 31 回インタラクティブシステムとソフトウェアに関するワークショップ(WIIS2023)論文集, 2023.

未来ビジョン：透明トマトのカプレーゼ

右図に示しているのは「透明トマト」のレシピである。味香り戦略研究所が 10 種以上のトマトを測定した平均値（を公開用に改変してもらったもの）に基づいている。無色透明でありながら、完全にトマトの味であり、たとえこの世からトマトが絶滅してもこのレシピがあれば大丈夫と思えるほどの再現性がある。モツツアレラチーズにかけば、トマトがなくても美味しいカプレーゼが作れる。アレルゲンを含まないのでトマトアレルギーの方にも食してもらえる。また、熟成しないので同じ味を保持する。

さて、この透明トマトの味はややフレッシュ寄りであるため、本稿の手法で「熟成」させてみたところ、見事に熟成したトマトの味になった。ただ、中和剤を入れて酸味を減算するより、元のレシピから酸の分量を減らした方が手取り早い。このように、溶液混合だけで味を作り出すことは、今となっては楽である。

TTV3 で挑戦している「味の減算」は、いわば既に絵が描かれた壁面に違う絵をプロジェクトションするような技術であり、純白のスクリーンにそのまま映像を投影するのとは段違いに難しいのだが、一般の人にはこれをわかつてもらえない。



スクロース 48g, 塩化カリウム 3g, クエン酸 3g, クエン酸ナトリウム 4g, L-グルタミン酸 1g, グルタミン酸ナトリウム 0.9g, L-アスパラギン酸 0.5g, アスコルビン酸 0.15g, 硫酸カルシウム 2 水和物 0.25g, 塩化マグネシウム 1.02 μg

シャボン膜凍結式全周囲スクリーンの提案

橋本 侑樹 * 西村 晶太郎 * 寺澤 真一郎 * 大塚 真柊 * 佐藤 俊樹 *

概要. シャボン玉は美しく、そして儂いものとして親しまれており、幅広い分野で注目されてきた素材である。近年はシャボン膜をプロジェクトのスクリーンとして用いる研究が行われている。本研究では、新たにシャボン膜を凍結させ不透明度を高めることでのスクリーン化を試みる。特に、膜が完全に凍結するまでの過程で膜上に氷の欠片が発生し対流する「スノーグローブ現象」を用いた、新しい映像表現の可能性にも着目した。また、シャボン膜は「平面形状」以外にも空気を入れて膨らませることで、「半球(ドーム)形状」のような超広角プロジェクションを用いた全周囲映像投影に適した形状に変形させることも可能である。本稿では実際にシャボン膜を生成、凍結させ、さらにスノーグローブ現象を発現させることができたプロトタイプの実装を行い、その性能を評価し、今後の展望等について述べた。

1 はじめに

シャボン玉は美しくも儂い存在として古くから親しまれてきた存在であり、芸術・文学・音楽などの様々な表現で用いられてきた。一方、科学の分野でもその性質が様々な角度から分析され、応用が行われてきた[1]。Human-Computer Interaction (HCI) 分野においてもシャボン膜の持つ独特の性質は注目され、ディスプレイとして応用する試みが行われてきた。シャボン膜は約 $1\mu\text{m}$ 程の薄い膜であるため、壊れやすく、また透明であるため映像投影に適した素材であるとは言い難いが、シャボン膜に振動を与えることで膜を不透明化させ、映像投影が可能な美しいスクリーンを実現する試み[2]やシャボン玉に煙を閉じ込めて不透明化し、映像投影を可能にする研究[3]等の、シャボン玉ならではの特性に着目したディスプレイ化の試みが行われてきた。

しかし、スクリーンとしてのシャボン膜の持つ面白さはこれだけではない。本研究が新たに着目するのは、シャボン膜の「凍結」である。シャボン膜は本来透明度が高く映像投影には不向きであるが、膜を低温で凍結させることで不透明度が上昇する。この凍結状態のシャボン膜が、映像を美しく拡散するスクリーンになり得るのではないかと考えた。

また、本研究ではシャボン膜の「凍結過程」で観測できる「スノーグローブ現象(図1)」にも着目する。この現象は、シャボン膜上に生じた氷の欠片(以下「氷片」と呼ぶ)が膜上を対流する凍結過程独特的美しい現象であり、スクリーン化により映像投影技術とこの美しい自然現象を組み合わせた新しい表現の可能性にも着目した。

本稿では、シャボン膜ならではの美しさや儂さ、また後述するスクリーンとしての面白い性質を活か

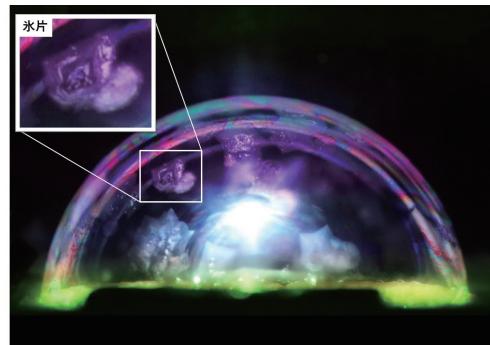


図 1. スノーグローブ現象の氷片

しながら、「凍結」という従来とは異なるアプローチでシャボン膜をスクリーン化する試みについて述べる。

2 提案と目的

本研究の目的は、シャボン膜を低温環境下において「凍結」させることで膜の不透明度を上昇させ、シャボン膜上への映像投影を可能にする新しい「シャボン膜凍結式スクリーン」を実現することである。また、その凍結過程にも着目した新しい映像表現の可能性を示すことである。

本研究の提案するシャボン膜凍結式スクリーンは、プロジェクトスクリーンとして次のようなユニークな性質を持っていると考える。まず、シャボン膜の凍結過程は膜の周囲の温度条件に大きく依存するため、膜の全体が完全に凍結するまでの過程は様々で、必ずしも膜の全体が均一に、もしくは一瞬で凍結するとは限らない。つまり多くの場合、その凍結過程は、同一膜上に「凍結している部分」と「そうでない部分」が混在する過程を経て、徐々に凍結していく過程をたどる。この性質は、改めてプロジェクタ

Copyright is held by the author(s).

* 北陸先端科学技術大学院大学

スクリーンの性質として考えてみると、同一の膜上に「映像を映し出す不透明部分」と、「映像を透過させ、膜の外（例えば周囲の壁等）に映像を映し出す透明部分」を同時かつ動的に存在させる、面白い性質であるとも捉えられる。さらに、このシャボン膜の凍結過程においては、温度等が関係する特定の条件を満たすことで、「スノーグローブ現象[4]」と呼ばれる美しい現象が観測可能である。この現象は、凍結過程にあるシャボン膜上に小さな氷片が生じシャボン膜上を動き回る視覚的に美しい現象であり、第7章で詳しく述べるように、映像投影との組み合わせで新しい映像表現への応用の可能性も考えられる。

加えて、平面的なスクリーンとしての「膜形状」に限らず、空気を送り込んで膜を膨らませていくことで、「半球状」、さらには「球状」などの非平面的な形状への変化も可能な形状の可変性にも注目したい。これは超広角投影技術と組み合わせることで、形状・大きさ可変の全周囲投影用の半球・球体スクリーンとしての応用可能性も持つと考える。

本稿では以下に、次のようなことを述べる。

1つ目に、上記の平面および全周囲映像の投影が可能な2種類の非平面形状シャボン膜スクリーンの生成とスノーグローブ現象の発現を含むシャボン膜の凍結、および映像投影が可能なディスプレイシステムの要件についてまとめ、最初のプロトタイプ制作について述べる。

2つ目に、実装した試作機および凍結させたシャボン膜のスクリーンとしての最初の性能評価として、実装したシステムを用いたシャボン膜の凍結成功率、凍結するまでに要する時間、また凍結膜の不透明度について調べた実験について述べる。

3つ目に、実装したプロトタイプの改善点や今後の拡張案について述べ、凍結式シャボン膜スクリーンやスノーグローブ現象と映像投影との組み合わせの可能性について考察する。

3 関連研究

薄く透明であるシャボン玉やシャボン膜はそのままの状態ではディスプレイやプロジェクタのスクリーンとして用いることは困難であるが、これまでにも様々な解決手法が提案してきた。例えば、シャボン玉の中に不透明度を高めるための煙や霧を閉じ込めて、外からの映像投影を可能にする手法が提案されている[3][5]。この手法の場合、シャボン玉に対して外からの映像投影が可能になる点に加え、映像が投影された状態のシャボン玉が割れた瞬間に、空間中に解放された煙に映像が映る美しい映像表現が可能になる利点がある。

次に、複数のシャボン玉を2次元的に配置し、各シャボン玉の大きさを制御することでディスプレイの画素として用いる試み[6]や、液体中に発生させた泡を液体の表面[7]や底部[8]に2次元的に並

べ、文字や模様を形作る試みもある。

透明なシャボン膜のスクリーン化に着目した研究としては、シャボン膜に超音波を当て、膜を振動させることにより反射特性を変化させ、映像投影を可能にする試みもある[2]。またシャボン膜に超音波を当てて振動させる手法は立体的に形作った泡の集合体に対しても用いられており[9]、泡への映像投影の視認性を向上させている。さらにこの手法では、泡にヘリウムガスを含ませ、形作った形状を空中に浮かせる試みも行われている。

このようにシャボン玉（膜）や泡をディスプレイやスクリーンとして用いる試みが複数なされてきており、どの試みも様々な角度から注目したシャボン膜ならではの美しさや面白さをうまく活かしたものであるといえる。本研究では、「シャボン膜の凍結」というこれまでにない方向性で、凍結過程を含めたシャボン膜の新しい側面の美しさのプロジェクトスクリーンとしての応用可能性に着目したものである。

4 投影システムの設計

提案するシャボン膜凍結式スクリーンを有するディスプレイシステムに必要な機能について、シャボン膜の「生成」・「凍結」・「投影」機構の3つに分けてそれぞれの設計の指針を述べる。

4.1 シャボン膜の生成に関する機構の設計

まず、シャボン膜は乾燥等の影響で壊れやすく長時間の維持が困難であると考え、シャボン膜スクリーンは実際に映像投影を行う直前にユーザ（もしくはシステム側）が簡単な操作で生成することとした。

なお、凍結前のシャボン膜は透明に近いが、膜を透過する映像は若干であるが膜の模様が映り込む、また映像がゆがむ等の視覚的な影響を受ける。しかしシャボン膜が存在しない状態が初期状態であればこれらの影響はなくなり、プロジェクタ映像はシャボン膜スクリーンの位置よりさらに前方にある別のスクリーンへの投影も可能になる。

次に、スクリーンとして用いるシャボン膜は、ユーザ/システム側が「膜を膨らませる」ことで、平面から半球形状まで自由に形状を変えられることとした。また、この操作は凍結前に行うこととし、ユーザはボタン等の簡単な操作で無段階に膨らみ加減を調整できることが望ましいと考えた。このシャボン膜の生成が完了すれば、すぐ凍結のための冷却処理を開始する。なお、半球形状からさらに膨らませることで球に近い形状までの変形が可能であると考えられるが、凍結時に膜の耐久性が低下することが予想されたため、今回の実装では半球形状までの変形に留めることにした。

4.2 シャボン膜の凍結に関する機構の設計

Ahmadi らの報告 [4] によると、地面に接地した状態の球状のシャボン玉において、次の 2 種類の異なる凍結現象が報告されている。

1 つ目の凍結は、冷たい基部上にシャボン玉を置いたときに、接地部分から頂点部分の方向に向かって徐々に凍結していく凍結過程である。この凍結過程は、下部から徐々に凍結していくことからボトムアップ凍結と呼ばれている。

2 つ目の凍結は、シャボン玉の接地部分の温度がシャボン玉の融点よりも低く、かつ接地部分から頂点部分に向けた対流が膜上に起こっている場合に観測できる現象である。この現象では、接地部分で凍結した氷片が対流により削られ、対流に乗って膜上を上昇していく。このとき、氷片は次第に大きく成長しながら互いにくつきあい、全体が成長した氷片に覆いつぶされるように凍結していく過程となる。これがスノーグローブ現象である。

以上をまとめると、スノーグローブ現象が発現する後者の凍結パターンを再現するためには、シャボン膜スクリーンの接地部分が十分に冷却される必要があり、特に上部の方がより低温になるような凍結条件を再現する必要があると考える。

4.3 シャボン膜への投影に関する機構の設計

本研究で実現するシャボン膜スクリーンは空気で膨らませることで平面から半球形状まで変化する。特に半球形状や、将来的に球形状まで膨らませた場合の投影も可能にする拡張案を考えると、複数台のプロジェクタを用いざともシャボン膜スクリーン全体への投影が可能になるシャボン膜内部（下部）からの画角が 180 度に近い超広角投影が望ましいと考える。

5 実装

次に、以上の設計をもとに実際に制作したプロトタイプの実装手法について述べる。

5.1 シャボン膜の生成

システム全体のハードウェア構成は、「シャボン膜生成機構」、「シャボン膜冷却機構」および「全周囲投影システム」の大きく分けて 3 つのユニットから構成されている（図 2）。

まずシャボン膜生成機構（図 3）は、シャボン液に浸して張り詰めた状態の紐を（1）、銅パイプを曲げて制作した円形（直径 10cm）の「シャボン膜基部」全体に擦りつけることで（2）、平面的な初期状態のシャボン膜を生成する（3）。この紐の両端は 2 本のスライドレール上に固定されており（図 4），ユーザは紐を引っ張ることができ、いつでも容易にシャボン膜を張れるようになっている。

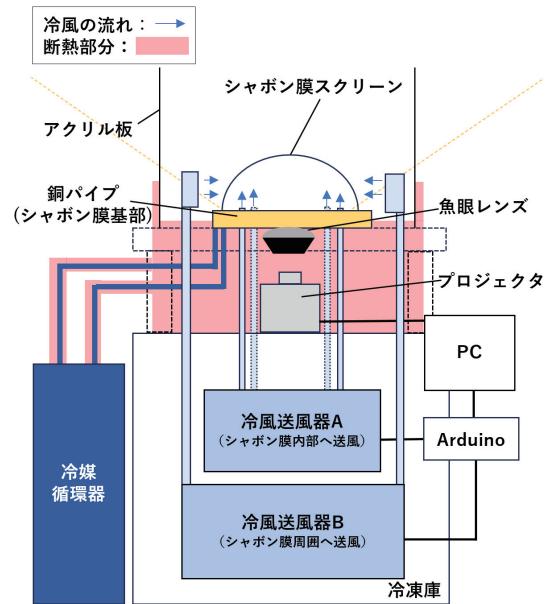


図 2. システム構成

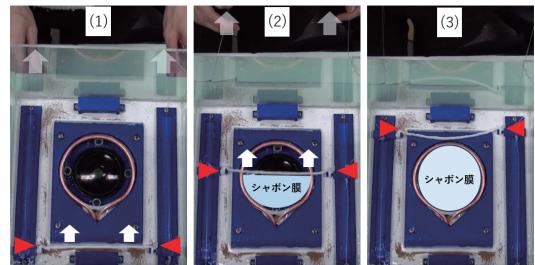


図 3. シャボン膜生成機構

次に、シャボン膜を膨らませる場合、シャボン膜内部へ開口する送風口（A）から市販の冷凍庫（MAXZEN 製 JF200ML01）を改造して制作した冷風送風機 A から断熱パイプを通して -10 度に冷却した空気を少量ずつ送り込む（この段階ではシャボン膜はまだ凍らない）。シャボン膜の形状は、この時送り込む冷却空気の量（電磁弁開放時間）に応じて、平面（初期状態）から半球形状（約 7 秒間開放）まで無段階に変化させることができる。今回の実装では、この空気量のコントロールはユーザが手動（ボタン操作）で行うようになっている。

なお、今回の実装で用いたシャボン液は、質量比で 80 % の水、10 % のポリビニルアルコール（PVA）を、そして 10 % の市販の食器用洗剤用いて調合した。これは一般的なシャボン液と同様の水および洗剤に加え、シャボン膜の強度を高める素材として PVA を加えて作成したものである。なお、シャボン玉の強度を上げるための様々な手法が存在するが [10] [11] [12]、加えすぎると逆に強度が低下したり、凝固点降下が起こり凍りにくくなったりする問題があり、

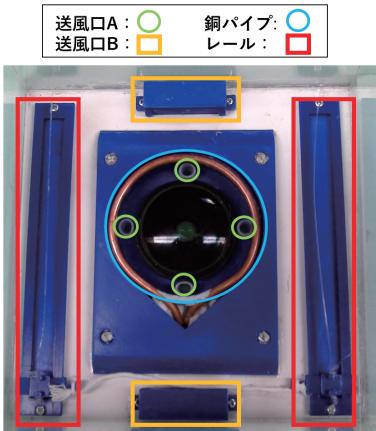


図 4. シャボン膜基部および送风口の位置関係

現時点では材料やその割合についてはまだ試行錯誤が必要な状態である。

5.2 シャボン膜の冷却

シャボン膜の生成が完了すると、即座にシャボン膜冷却機構を稼働させ、シャボン膜を凍結させる冷却プロセスを開始する。今回の実装では、入手および管理が困難で危険性が高い液化窒素や液化炭酸等を用いた冷却手法は用いない、次の2種類の異なる冷却手法を組み合わせて実装した。まずシャボン膜スクリーンの接地部を冷却するために、シャボン膜基部にある膜が接地する冷却パイプに冷媒循環装置(三商製 SCA-12)を用いて-2度に冷却した不凍液(エチレングリコール)を循環させた。同時に、シャボン膜上部の空間を冷却するために、冷風送風機Aから-10度の冷却空気の送風を行った。

なお、冷風送風機のダクトと冷媒循環器から冷却パイプに至る配管はグラスウール断熱材で覆い、またレンズおよびプロジェクタ部の周囲も発泡素材で覆うことで断熱処理を施した。また今回の実装においては、冷却空気をシャボン膜周囲に滞留させるために、スクリーンの周囲を高さ21cmの薄いアクリル板で囲むこととした。しかし、将来的にはシャボン膜に直接手で触れるなどの対話を可能にするために、開口部と外気の間に風の流れを作り外気の侵入を防ぐエアカーテン機構を取り入れる等の冷却効率の改善を行いたい。

5.3 スノーグローブ現象の再現

スノーグローブ現象の発現条件として重要なのは、シャボン膜の接地部分で氷の結晶が生成されていることである[4]。シャボン膜ではマランゴニ対流と呼ばれる液体中の温度勾配によって表面張力が変化し、それに伴い液体内部で対流が発生する現象が起きている。スノーグローブ現象ではシャボン液が凍り、結晶になる際に生じる熱によりシャボン膜内で

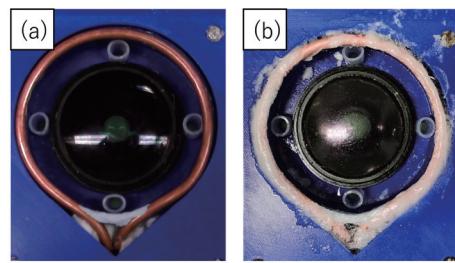


図 5. (a) シャボン膜基部と (b) 付着した霜

温度勾配が生まれ、マランゴニ対流が発生する。

シャボン膜の接地部分で生成された氷の結晶がマランゴニ対流によって液状部に流れ、大きな氷片となる。このようにして、スノーグローブ現象は発現する。このことから、シャボン膜の接地部分である銅パイプは、全く何も付着していない状態でシャボン膜を生成すると、ボトムアップ凍結が起こりやすくなると考える。実際に、事前に銅パイプにシャボン液を塗布し、銅パイプ表面に結晶(霜)を付着(図5)させてからシャボン膜を生成すると、スノーグローブ現象が発現する際に、より多くの細かな氷片が膜上を舞う様子が観察された。

5.4 超広角映像投影

全周囲映像投影システムは、約180度の画角を持つ小型の広角レンズ(Opteka製0.3X UltraFisheye)と広角レンズを通してフォーカスが合うように調整した小型プロジェクタ(UO製SmartBeam-Laser、輝度100ルーメン)から構成されている。広角レンズはスクリーンの膜面(平面時)から1cm下方に中心(光軸)を一致させた同心円状に配置し、プロジェクタはレンズ下面から13cmの距離に、広角レンズ後玉の直径と投影画像の短辺の長さが一致する角度で配置した。これにより、凍結状態のシャボン膜スクリーンの内側から、シャボン膜の全周囲に對しての映像投影を可能にした。

また投影画像は、事前に推定しておいた広角レンズを通したプロジェクタの内部・外部パラメータ、およびレンズの歪み係数を用いて、Fisheyeカメラモデルを実装したUnity内のシェーダプログラムにより、リアルタイムにレンダリングできるようにした。

なお、将来的に半球形状からさらに膜を膨らませた場合の球に近い形状のスクリーンに投影するためには、現在のレンズ位置を、膜を膨らませながら、膜面より上の部分にせり出させて配置する必要があると考える。

6 性能評価

今回作成したプロトタイプの性能を評価するため、2つの実験を行った。

6.1 凍結成功率と凍結に要する時間およびスノーグローブ現象の発現率

今回提案するシャボン膜の凍結によるスクリーンを生成する実験を行った。膨らませない平面形状、約3秒間送風を行うドーム形状（中間）、約7秒間送風を行う半球形状を形成したのちに凍結させた。凍結はそれぞれの形状ごとに30回行い、シャボン膜周辺の温度が5度を超えた場合、実験を一度停止し、0度まで冷却を行った後に実験を再開した。このとき、シャボン膜を生成したタイミングから膜全体が完全に凍結していることを視認するまでの時間（凍結時間）を測定した。なお、実験時の室温湿度は平面形状の時は21.3度、51%，ドーム形状の時は20.5度、49%，半球形状の時は20.3度、52%であった。また、ドーム形状、半球形状のシャボン膜についてはシャボン膜の接地位置からの高さを測定した。また、凍結終了時のシャボン膜周辺の温度をアクリル板の内側に設置した温度計から読み取った。

平面形状、ドーム形状、半球形状の凍結前および凍結後の様子を斜め上方向と水平方向から撮影した写真をまとめたものが図6である。また、各形状について成功率と凍結時間の平均を算出した。平面形状では、30回中23回凍結が成功し、成功率は76.7%であった。また、凍結時間の平均は45秒であった。ドーム形状では、30回中22回凍結が成功し、成功率は73.3%であった。また、凍結時間の平均は43秒であり、高さの平均は2.12cmであった。半球形状では、30回中11回凍結が成功し、成功率は36.7%であった。また、凍結時間の平均は63秒であり、高さの平均は4.99cmであった。なお、スノーグローブ現象についてだが、今回の実験ではどの形状においても氷片が少なからず接地面から伸びる氷の結晶から離れた場所に見られた。つまりスノーグローブ現象はシャボン膜を生成することができれば発現することが確認できた。

この実験から3点が判明した。1つ目は氷片の数や大小に差はあるが、シャボン膜が生成でき、凍結する際にはスノーグローブ現象が発現していたことである。2つ目は凍結の際に大部分が凍結しスノーグローブ現象が発現しなくなった後は冷風が届きづらいレール側が凍結しにくいくことである。3つ目は半球形状で失敗した際は、頂点付近では氷片がとどまらず、接地位置から成長した氷の結晶とつながり、頂点付近のみ凍結しない状態が見られたことである。2点目、3点目は冷風の発生場所に関する問題なので試作機の改良が必要となる。

6.2 凍結シャボン膜の不透明度の変化

シャボン膜が凍結した際の不透明度の変化を測定する実験を行った。5mm厚アクリル製の型を作り、シャボン液に浸することで四角いシャボン膜を張った。この型を-15度の冷凍庫に60秒間入れ、シャボン

表1. 計測した可視光透過率。

測定した物質	可視光透過率 (%)
凍結後のシャボン膜	80.9
凍結前のシャボン膜	99.8
アクリル板 (5mm厚)	97.0
曇りアクリル板 (5mm厚)	64.0

膜を凍結させた。凍結したシャボン膜に対し、可視光透過率測定器 (LS162 Transmission Meter) を用いることで可視光透過率を測定した。なお、比較のために凍結前のシャボン膜、アクリル板 (5mm厚)、曇りアクリル板 (5mm厚) の可視光透過率の測定も行った。また、凍結前後のシャボン膜に、黒背景に白文字で「Bubble」と表示した画像を投影し、その様子を水平方向から撮影した。

凍結前、凍結後のシャボン膜、アクリル板 (5mm厚)、曇りアクリル板 (5mm厚) の可視光透過率を表したもののが表1である。この結果からシャボン膜の凍結によって透過度の低下が示された。

また、図7(a)は非凍結時の、(b)は凍結時のシャボン膜に投影を行ったときの様子である。実際に文字の視認性が向上していることが見て取れる。

以上の可視光透過率の変化と投影した際の実際の視認性の変化から、スクリーンとして扱うことができるよう変化が起こっていると考える。

7 考察と展望

7.1 スノーグローブ現象の応用可能性について

スノーグローブ現象の氷片は、その位置や形状・サイズまでもが刻々と変化する、動的で視覚的に美しい自然現象である。この現象は、同一のシャボン膜スクリーン上に「透明部分」と「不透明部分」を局所的に作り出し、シャボン膜上の氷片への投影と、膜を透過した先にある別のスクリーン（壁やかざした紙や手等）への投影の、2種類の異なる美しい多層映像表現を可能にする。

また、将来的にプロジェクタ部にハーフミラーを介した同軸カメラを接続すれば、スノーグローブ現象の氷片をカメラで追跡し、氷片に対しリアルタイムに色や映像・文字等を提示する等の動的な氷片の美しさを拡張する映像提示も可能になると考える。

また、カメラを用いて凍結過程をシャボン膜内側から撮影した全周囲映像をHMDを通してユーザに提示することで、大きなシャボン膜の中に入り、中から凍結過程を眺めるような新しいVR体験にも発展可能であると考える。

また、スノーグローブ現象自体も、シャボン液の配合や温度環境等によって、その氷片の形状や成長過程、移動速度等も様々に変化する。今後は、美し

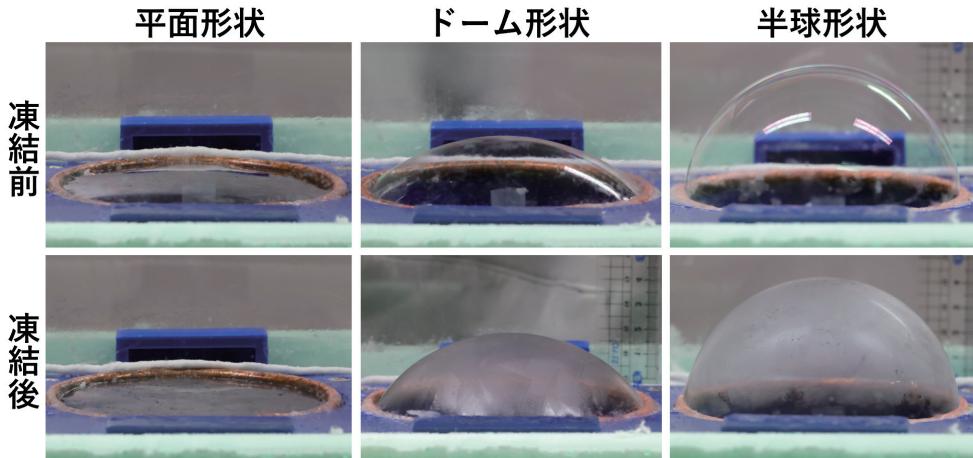


図 6. 水平方向から撮影した形状ごとのシャボン膜の凍結前後の様子



図 7. (a) 非凍結時, (b) 凍結時の文字画像の投影結果

い氷片を作り出すのに必要な様々な条件について検討し、これらをプログラマブルに作り出す冷却手法を探っていきたい。さらに、もしシャボン膜の周囲の温度変化の制御や、超音波・風等を用いて外部からシャボン膜上の対流や氷片の動きをコントロール可能になるのであれば、複数の氷片で模様や文字を作ることも将来的には可能になるかもしれない。

7.2 インタラクション要素の可能性

今回の試作においては、凍結の過程で起こるスノーグローブ現象の再現に重きを置き、その凍結過程をユーザにゆっくり見せることに着目した実装になっていた。

一方で、試作機の冷却・断熱性能の向上により、今後より低い温度での急速なシャボン膜の冷凍が可能になれば、シャボン膜スクリーンを生成と同時に瞬時に完全凍結させ、必要な時に、瞬時に手元に生成できる「インスタントスクリーン」として活用する可能性もあると考える。これは、例えばプロジェクトのメニュー画面等を、遠くの画面ではなく、レンズ直前に生成したインスタントスクリーン等に一時的に映し出し、手元での操作を可能にする等の利用方法もあるのではないかと考える。

また、シャボン膜スクリーンは、手で直接触れる

ことで破壊することも容易で、不要になった場合に完全に取り除くことも可能である。さらに、凍結したシャボン膜は膜の一部を溶かした場合でも部分的に穴が開くだけで、全体が割れて完全に消えてしまうことはない。この性質を応用し、手で直接触れる、部分的に風を当てる等でスクリーンに穴を開けて、映像を膜の外に選択的に映し出すようなインタラクションも可能になると考える。

7.3 シャボン膜の形状の制約

実装した試作機では、直径 10cm のシャボン膜を半球形状より大きく膨らませた状態で凍結させることは困難であった。シャボン膜に直接当たる冷風はおよそ -10 °C 程度に対し、シャボン膜周辺の温度は最低でも -1 °C 程度であったことからシャボン膜が大きくなるほど膜全体に冷風が届かず、十分に冷やされない部分が生じてしまうことが原因だと考えている。今後は送風口の数や位置を改良し、より高い場所まで冷風を送ることができれば、シャボン膜の直径や高さをさらに大きくすることも可能になるのではないかと考えている。

8 まとめ

本稿では、シャボン膜を凍結させることで不透明度を高め、凍結した様々な形状のシャボン膜の全周囲への映像投影を可能にする手法を提案した。また、凍結過程におけるスノーグローブ現象の再現にも着目し、これを可能にするシャボン膜の凍結および投影システムの最初のプロトタイプの設計と実装について述べた。

また初期のプロトタイプを用いた基本的な凍結性能の評価として、凍結時間、凍結の成功率、凍結前後の不透明度の変化を調べる実験を行い、最後にスノーグローブ現象を発現させるシャボン膜スクリーンの応用可能性や今後の拡張案について述べた。

謝辞

本研究は JSPS 科研費 21K11993 の助成を受けたものです。

参考文献

- [1] Xi Yang and Eiji Miyako. Soap bubble polination. *iScience*, Vol. 23, No. 6, p. 101188, 2020.
- [2] Yoichi Ochiai, Alexis Oyama, Takayuki Hoshi, and Jun Rekimoto. Poppable display: A display that enables popping, breaking, and tearing interactions with people. In *2013 IEEE 2nd Global Conference on Consumer Electronics (GCCE)*, pp. 124–128, 2013.
- [3] Masahiro Nakamura, Go Inaba, Jun Tamaoki, Kazuhito Shiratori, and Junichi Hoshino. Mounting and application of bubble display system: Bubble cosmos. In *Proceedings of the 2006 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology, ACE '06*, p. 47–es, New York, NY, USA, 2006. Association for Computing Machinery.
- [4] S. Farzad Ahmadi, Saurabh Nath, Christian M. Kingett, Pengtao Yue, and Jonathan B. Boreyko. How soap bubbles freeze. *Nature Communications*, Vol. 10, No. 1, p. 2531, Jun 2019.
- [5] Sue Ann Seah, Diego Martinez Plasencia, Peter D. Bennett, Abhijit Karnik, Vlad Stefan Otrocel, Jarrod Knibbe, Andy Cockburn, and Sriram Subramanian. Sensabubble: A chrono-sensory mid-air display of sight and smell. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '14*, p. 2863–2872, New York, NY, USA, 2014. Association for Computing Machinery.
- [6] Shiho Hirayama and Yasuaki Kakehi. Shaboned display: An interactive substantial display using soap bubbles. In *ACM SIGGRAPH 2010 Emerging Technologies, SIGGRAPH '10*, New York, NY, USA, 2010. Association for Computing Machinery.
- [7] Ayaka Ishii, Manaka Fukushima, Yasushi Matoba, and Itiro Siio. Utakata: Floating bubble display. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems, CHI EA '20*, p. 1–8, New York, NY, USA, 2020. Association for Computing Machinery.
- [8] Harpreet Sareen, Yibo Fu, Nour Boulahcen, and Yasuaki Kakehi. Bubbletex: Designing heterogeneous wettable areas for carbonation bubble patterns on surfaces. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, CHI '23*, New York, NY, USA, 2023. Association for Computing Machinery.
- [9] Yuki Kubo, Hirobumi Tomita, Shuta Nakamae, Takayuki Hoshi, and Yoichi Ochiai. Bubble cloud: Projection of an image onto a bubble cluster. In *Proceedings of the 13th International Conference on Advances in Computer Entertainment Technology, ACE '16*, New York, NY, USA, 2016. Association for Computing Machinery.
- [10] Marina Pasquet, Laura Wallon, Pierre-Yves Fusier, Frédéric Restagno, and Emmanuelle Rio. An optimized recipe for making giant bubbles. *The European Physical Journal E*, Vol. 45, No. 12, p. 101, Dec 2022.
- [11] 笹田和希, 山本七彩, 横山遙, 手嶋日菜子, 五十川奈穂, 柴田航志, 中原涼花, 乘次優希奈, 宮川光林, 田中聖子, 平野貴士, 上野桃子, 藤戸文子, 高嶋綾香, 菅野憲一. 〈論文・報告〉割れないシャボン玉の開発と割れなくなるメカニズムの解明. かやのもり:近畿大学産業理工学部研究報告, No. 24, pp. 15–20, 07 2016.
- [12] Stephen Frazier, Xinyi Jiang, and Justin C. Burton. How to make a giant bubble. *Phys. Rev. Fluids*, Vol. 5, p. 013304, Jan 2020.

導電性・強磁性・及び土台の可撓性を持つ毛構造の造形手法の提案

鎌田 航誠^{*} 高橋 治輝[†] 塚田 浩二^{*}

概要. 本研究では、安価な熱溶解積層方式の3Dプリンタで造形可能な細かい毛の集合体（毛構造）の拡張について提案する。先行研究で提案されていた毛構造の造形手法を活用し、導電性／磁性等を備えるフィラメントを造形に用いる方式や、柔軟に動かせるソフトレジンに毛を埋め込む方式を提案・試作する。試作した毛構造について、ユーザテストを通して、材料やパラメータ変化による触感の印象を調査する。さらに、制作例を通して、提案手法をインタラクション研究に活用する方向性を示す。

1 はじめに

近年、3Dプリンタの造形技術の進歩により、低価格の3Dプリンタで数ミリ単位の細かい形状や複雑な構造の立体物を造形可能となり、直径1mm以下の細長い毛のような構造の造形手法が提案されている[12]。毛状の造形物はVRコントローラ[4]やテクスチャ表現[1]のように、毛の長さによる触り心地の変化を活かして、実世界の触感の再現などに応用されている。ピンアレイや毛構造に上下運動や回転運動が可能なアクチュエータを組み合わせて、触覚ディスプレイとして扱う研究も行われている[13][8]。

一方、先行研究で提案された毛構造にはまだ拡張の余地が残されている。まず、毛を造形する際の材料に関して、これまでの手法では主にPLA樹脂が用いられており、フィラメントの種類について十分検証されていない。また、毛構造の毛の部分はプラスチック製の硬い土台に固定されており、この土台部分を柔軟に折り曲げて活用するような事例はない。さらに、こうした毛構造がもたらす触感の傾向を理解するためのユーザテストも行われていない。

本研究では、3Dプリントによって造形可能な細かい毛の集合体（毛構造）を用いた情報提示手法の構築を目指す。そのための手法として、導電性／磁性を持つフィラメントを造形に用いる方式や、ソフトレジンに毛構造を埋め込む方式を提案・試作する。これらの方針を用いることで、毛構造をセンサやアクチュエータとして扱い、静電容量／磁力／圧力等を活用したインタラクティブな触覚ディスプレイを構築できると考える。また、本論文はインタラクション2023でインタラクティブ発表[11]を行った内容に加筆したものであり、UIST2023でのデモ発表[2]を予定している。

Copyright is held by the author(s).

* 公立はこだて未来大学

† 立命館大学

2 関連研究

2.1 毛の造形手法

高橋らは熱溶解積層方式3Dプリンタを用いた毛構造の造形方法やその応用方法について提案している。具体的に、文献[12]では熱溶解積層方式3Dプリンタを使用し、樹脂を引き伸ばすG-codeを直接記述することで、毛を3Dプリントする方法を提案している。文献[8]では、G-codeを使わずにライサソフトウェアのみを用いて行えるようにした。さらに、毛構造とサーボモータを組み合わせることで、毛先を移動・回転可能な情報提示手法を提案している。Laputら[3]は、3Dプリンタの造形時に起こる糸引き現象を利用した毛の造形手法を提案している。文献[12][8]は二つの壁を橋渡しするようにして造形するが、この手法では一つの壁から水平方向にフィラメントを引き延ばしている。その影響として、毛の先端部分にかけて細くなる特徴がある。Ouら[6]は、光造形方式の3Dプリンタを用いて、細かい毛を表面テクスチャとして持ち、毛が生える傾きなどを調整できる毛構造を作製した。応用例として、毛の上に置いた物体の動きを制御できるアクチュエータを提案している。本研究では、先行研究[8]の造形方法を活用して毛構造の造形を行う。造形する毛構造に対し、導電性／磁性等の多様なフィラメントを利用することで、素材の性質を活用した情報提示手法を構築する。

2.2 毛構造を使った触覚提示・センシング

Leeら[4]は、毛を用いて指先に触覚フィードバックを与えるVRコントローラを作製した。デバイスを用いてVR上で仮想的な物体に接触することで、デバイスに内蔵されたモータが毛の長さと角度を制御し、柔らかい触感や粗い触感を指先に伝えることができる。Degraenら[1]は、VRで仮想的な物体の質感を知覚するためのテクスチャとして、3Dプリンタで作られた毛構造に着目した。毛の長さのみが異なる五種類の毛構造を配置し、HMDで布や金属などの視覚的な質感を毛構造の位置に重ね合わせて表

示する。この状態でユーザーに5種類の毛を仮想空間上で触れさせ、毛を使って再現したテクスチャと視覚的な質感を一致できるのか実験を行っている。梅津ら[9]は、形状記憶合金をフレキシブルチューブで覆い、光センサーを組み込むことで、光量によって曲がり具合が変わる直径3mmの毛を作製した。この毛を複数本合わせてLCD上に置くことで、複数の毛をLCDの画面映像の光によって制御することなどができる。中嶋ら[14]は、半透明のプラスチック光ファイバーと赤外線による接触検出技術を用いて、毛状のマルチタッチディスプレイを提案している。光ファイバーが配置されている隙間から赤外線ライトを照射し、接触による赤外線の反射によってタッチ部分を検出する。プロジェクターを使い、光ファイバーの束を発光させることで、視覚的なフィードバックを行う。遠藤ら[10]は、3Dプリンタを使用して毛と毛根を造形し、食用のアガーを固めた弾性体に毛を埋め込むことによって、生物の毛と皮膚を模したものを作製した。毛根の下部からカメラで撮影し、弾性体による光弹性の変化を観測することで、微細な毛のタッチ検出を行うことができる。本研究では、毛の土台部分などにセンサやアクチュエータを埋め込むことで、静電容量／磁力／圧力等を活用したインタラクティブな触覚ディスプレイを単純な構造で構築することを目指す。

3 先行研究の再現およびシステム実装

毛構造の造形に関して、本研究で使用する3Dプリンタと材料で再現可能か確かめるため、先行研究[8]で提案された手法の再実装を行った。造形環境として、熱溶解積層方式3Dプリンタ(Creality Ender3, Ender3 v2)とPLAフィラメント(Voxelab PLA+)を用いた。ノズルは0.4mm径のものを使用した。左右に壁を作り(図1①)，その間を往復して橋渡しするように樹脂を出力することで毛を造形していく(図1②)。これを繰り返して複数の毛が集まった毛構造を造形することができる(図1③)。

毛構造の3DモデルはRhinocerosやそのプラグインであるGrasshopper等のプログラマブルなCAD環境(図1下部)を用いて設計する。Grasshopperで毛の長さや毛同士の配置間隔、毛の本数等のパラメータをUI内のスライダを動かすことで調整し、Rhinoceros上で調整した毛構造の3Dモデルを反映させる。以降の章では、このシステムを用いて設計した造形データを用いて造形実験を実施する。

4 異なる材料を用いた毛構造の特性変化

3Dプリンタでは、さまざまな種類の材料(フィラメント)を使用して造形することができる。一般的にはPLA樹脂やABS樹脂を用いるが、造形する目的や用途に合わせて、耐衝撃性や耐熱性を持つ

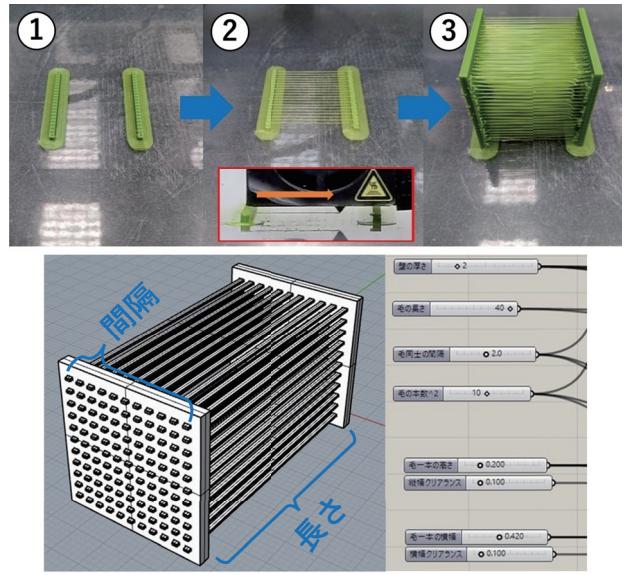


図1. 毛構造の造形の様子および設計システム

樹脂などを使い分ける。また、PLA樹脂に粉末素材(フィラー)を混ぜ合わせたフィラメントを使用することで、素材の特性を造形物に持たせることができる。

4.1 導電性フィラメント

導電性フィラメントとは、PLA樹脂に炭素系の粉末材料を混ぜた樹脂フィラメントである。導電性フィラメントを使うことで造形物に導電特性を付与することができる。そのため、このフィラメントはインタラクティブなオブジェクトの制作や電子回路用の部品の作成によく使用されている[5][7]。導電性フィラメントで造形した立体物は静電容量の変化によって接触を検知したり、曲げセンサのように立体物を曲げたときの状態を抵抗値から計測したりすることができる。

導電性を持つ毛構造を造形するため、Protopasta製のフィラメント(Proto-pasta CDP)¹とMSNJ製のフィラメント²の2種類の導電性PLAフィラメントを使用した。当初の造形実験では、入手性の観点からProtopastaの導電性フィラメントを使用して行っていた。このフィラメントを使用した造形には成功したが、造形後に土台から抜けてしまうことや折れやすいといった耐久性の問題が見られた。一方、MSNJ製の導電性フィラメントでは、こうした耐久面での問題が改善されたものの、材料に関する詳細な情報が公開されておらず入手性や材料の信頼性の面で課題がある。導電性フィラメントを使った毛構造の3Dプリントでも、先行研究と同様に、毛の長さ、密度、太さなどのパラメータを指定してモデルを造形可能である。

4.2 磁性フィラメント

磁性フィラメントとは、PLA樹脂に鉄粉末を混ぜた樹脂フィラメントである。ここでは磁鉄PLAフィラメント(Proto-pasta MIPLA)³を材料として毛構造を造形した。鉄粉末が混ぜこまれていることにより、造形物に金属の特性を付与することができる。このフィラメントは造形物に強磁性を持たせることを目的として作られているが、副次的な性質として、造形後の後処理により金属光沢や赤錆を造形物の表面に出すこともできる。

磁性フィラメントを使用して毛構造を3Dプリントするために、先行研究の造形方法に対して、本研究で用いた環境では、スライサの設定でノズル温度を推奨である200°Cよりも20°C高く設定し、フィラメントの押し出し量を110%にすることで安定して造形できた。磁性フィラメントを使った毛の3Dプリントでも、先行研究と同様に、毛の長さ、密度、太さなどのパラメータを調整してモデルを造形可能である。造形した毛構造は、図2のように、毛の部分が磁石に引き寄せられる特性の変化が見られた。さらに、これまでのPLA樹脂製の毛構造よりも丈夫であり、磁性の毛構造を手に持った状態で軽い磁石を持ち上げることが可能である。

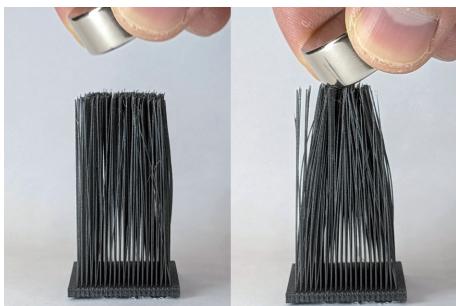


図2. 磁石に引き寄せられる毛構造

4.3 毛構造のソフトレジンへの埋め込み

3Dプリントした直後の毛構造の両端は、安定して造形するために二つの壁によって毛の両端が固定されている(図1③)。3Dプリント後は、カッターナイフを用いて片側を切り離すことにより毛構造を完成させる。そのため、毛構造の土台となっている部分は硬い板状になっており、曲げることは不可能である。土台部分をより柔軟にして曲げることが可能になれば、毛の動きの可能性をさらに広げたり、曲面に貼り付けたりすることができるだろう。ここ

では、毛構造の3Dプリント後に、土台部分を異なる素材に置き換える手法を紹介する。毛構造を柔軟に動作させることを可能にするため、土台にはレジンクラフトなどで使用されるソフトレジンを使用して柔軟な土台への植毛を行う。

4.3.1 ソフトレジンへの埋め込み手順

ソフトレジンに毛を埋め込む手順を図3の順番に沿って説明する。まず、3Dプリントした直後の毛構造の片方の壁と毛を切り離し(図3a), UVライトで硬化するソフトレジン液を型に流しておく(図3b)。その後、毛の先端をレジン液に埋める(図3c)。この状態でUV照射機を用いてUVライトを照射し、レジン液を硬化させる(図3d)。このようにして、硬化したレジンの壁と樹脂の壁で挟まれた毛構造を作ることができる(図3e)。その後、樹脂製の壁と毛を切り離すと柔軟な土台の毛構造を作ることができる(図3f)。

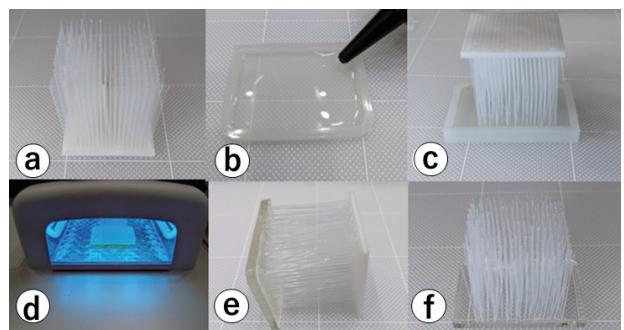


図3. レジンに毛を植毛する方法

4.3.2 シリコンモールドの作成

ソフトレジンへの毛の埋め込みには市販のシリコンモールドを使用できるが、ソフトレジンが硬化する前の段階(図3c)では毛の接地が不安定であり、UV照射機に入る際に、レジンに浸けた毛の先端が動いたり、毛構造自体が倒れたりしやすい。また、市販のシリコンモールドはサイズが限られているため、ソフトレジン土台の毛構造は目的に合わせたサイズで制作することが難しい。そこで、これらの問題の改善を目的としたシリコンモールドを作成する。図3の製作過程において、図3cの部分で使用されていた市販のシリコンモールドを新しく作成するシリコンモールドに置き換えることで、より安定した埋め込みが行えるようになる。

作成手順を図4a, b, cに示す。ソフトレジンを流し込むスペースとなる部分は、あらかじめ3Dプリンタを用いて造形する。ここでは、シリコンモールドに毛の先端が入り込む穴を設けるために、埋め込む毛の本数に合わせた窪みのある立体物を造形する(図4a)。このような形状で型をとることによって、

¹ <https://www.proto-pasta.com/collections/all/products/conductive-pla>

² <https://www.amazon.co.jp/dp/B08Z3NZCB7>

³ <https://www.proto-pasta.com/collections/all/products/magnetic-iron-pla>

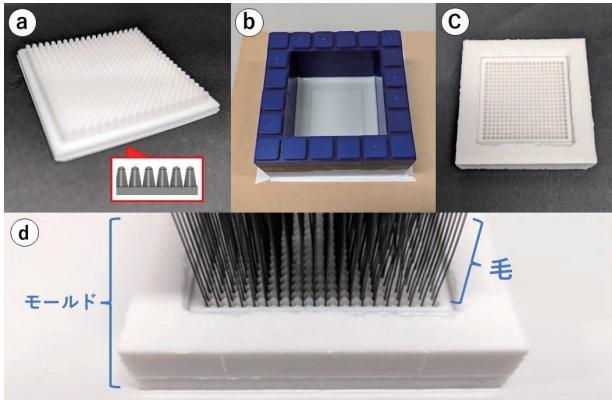


図 4. 植毛に適したシリコンモールドの作成手順および作成したシリコンモールドで植毛をする様子

埋め込む毛構造のサイズに合わせたシリコンモールドが作成可能である。この立体物(図4a)をシリコン作成用の枠に収め、シリコン液を枠に流し込み、硬化させることでシリコンモールドを作成する(図4b)。この状態で一日経過させ、固まったシリコンから枠と立体物を取り外すことで、図4cのようなシリコンモールドが作成できる。これに毛構造を設置すると、図4dのように、毛の先端がシリコンモールドに作られた細かな穴に入り込み、毛が均等に広がり、倒れにくい状態にできる。この改善により、図3の工程をより安定して行うことが可能になる。

4.3.3 柔らかい土台での毛の動き

ソフトレジンへ植毛した毛構造によって、土台部分を曲げることで毛が開く／閉じる動きや土台をねじることによる動きができるようになる(図5)。また、UVライトでレジンを硬化させる前の段階で、図6のように毛と一緒に電子部品を埋め込んで、毛構造の柔軟な土台部分を利用したセンシングやアクチュエーションが可能となる。試作として埋め込んだ曲げセンサ(SEN-10264)のような、薄く柔軟性のある電子部品が埋め込みに適していると考える。

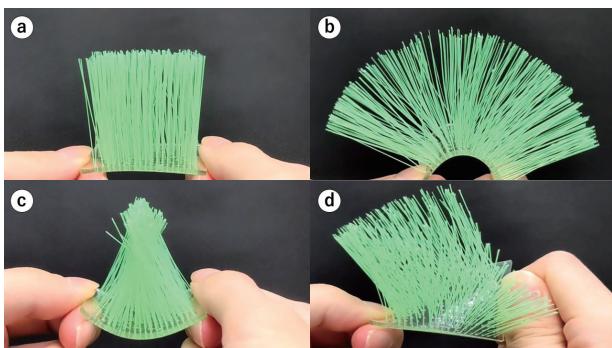


図 5. 土台を曲げたときの動作

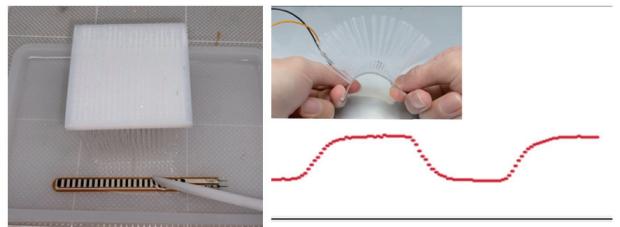


図 6. 電子部品の埋め込みと曲げセンサを用いた計測例

5 毛構造の触感の印象評価

5.1 目的

パラメータが異なる毛を被験者に実際に触らせて、それぞれの毛の触感について印象評価を行う。これにより、多様な毛構造の触感の違いや類似性について調査し、定量的な評価を測る。また、本稿で提案する導電性／磁性の毛構造を使った制作の参考にするため、パラメータは変更せずに造形材料だけを変えた場合の触感への影響も調査する。

5.2 手法

事前準備として、5種類の「パラメータを変更して造形した毛構造」と「導電性／磁性フィラメントを造形材料とした毛構造」を用意した。これらのパラメータと材料についてまとめたものを表1に示す。毛構造自体が独特の触感を持つことを考慮し、標準的なパラメータで造形した毛構造Aを触れた印象をベースラインとして、毛構造B-Gに触れた印象を回答させた。

被験者は、20～25歳の男性12名、女性2名の合計14名であった。まず、被験者には、実験で使用する毛構造が3Dプリンタで作製したものであることを伝え、表1の毛構造Aを提示し、形状と触感を確認させた。その後、実験者は毛構造B-Gの中から一つを順序効果を考慮して被験者に提示し、毛構造Aと触感を比較しながら評価を行わせた。なお、毛構造B-Gの形状は被験者からは見えないように配慮した。実験の様子と実験で使用された毛構造を図7に示す。実験では、被験者に片手で毛構造の毛先部分に指を押し付けるようにして触れさせながら、利き手で紙媒体のアンケートに回答してもらった。この手順を毛構造B-Gに対して繰り返し実施した。アンケートに使用した形容詞対はSD法で用いられる指標の中から触感に関するものを選出し、それぞれ7段階で回答させた(図8参照)。

5.3 結果

アンケートの結果の平均値と標準偏差について、毛の長さが異なる場合の結果を図8、毛の配置間隔が異なる場合の結果を図9、毛の造形材料が異なる場合の結果を図10にそれぞれ示す。

導電性・強磁性・及び土台の可撓性を持つ毛構造の造形手法の提案

表 1. 実験に使用した毛構造の詳細

	毛の長さ	毛の間隔	造形材料
毛構造 A	40 mm	2 mm	PLA
毛構造 B	20 mm	2 mm	PLA
毛構造 C	80 mm	2 mm	PLA
毛構造 D	40 mm	1 mm	PLA
毛構造 E	40 mm	5 mm	PLA
毛構造 F	40 mm	2 mm	導電性 PLA
毛構造 G	40 mm	2 mm	磁性 PLA



図 7. 実験の様子と使用された毛構造

図 8 より、「-3: やわらかい～+3: かたい」の質問については、長さが 20 mm の毛構造 (B) は平均 1.93 (標準偏差 0.92) となり、長さが 80 mm の毛構造 (C) は平均 -2.14 (標準偏差 0.66) という結果が得られた。また、「-3: なめらか～+3: 粗い」、「-3: 軽い～+3: 重い」、「-3: 弱弱しい～+3: 力強い」についても差がみられた。

図 9 より、「-3: 軽い～+3: 重い」の質問については、間隔が 1 mm の毛構造 (D) は平均 0.21 (標準偏差 1.31) となり、間隔が 5 mm の毛構造 (E) は平均 -2.21 (標準偏差 0.70) という結果が得られた。また、「-3: 弱弱しい～+3: 力強い」の質問については、間隔が 1 mm の毛構造 (D) は平均 1.14 (標準偏差 1.17) となり、毛の配置間隔が広い毛構造 (E) は平均 -2.29 (標準偏差 0.83) という結果が得られた。他にも、「-3: なめらか～+3: 粗い」、「-3: 親しみにくい～+3: 親しみやすい」、「-3: 悪い～+3: 良い」についても差が見られた。しかし、

「-3: やわらかい～+3: かたい」の質問についてのみ、差はほとんど見られなかった。

図 10 より、造形に使用する材料を変えた毛構造では、印象についてあまり大きな変化は見られなかつた。材料の特性の違いから、「なめらか～粗い」「やわらかい～かたい」等の項目で若干の分散は見られるが、毛の長さや間隔を変更した場合と比較すると、印象の差は小さい。

5.4 考察

毛の長さに関する実験結果から、毛が長いほどやわらかさを感じ、短いほど硬く感じる傾向を確認できた。また、やわらかさだけでなく、なめらかなどの触感の印象を変えたい場合も、毛の長さの変更が有用であると考えられる。毛の配置間隔を狭め、密度を高めることによって、力強い印象を受けることが確認できた。毛の配置間隔を広くすることで、毛の密度が低くなる。これによって、撫でた時に指に触れる毛の本数が減少し、軽くまたは弱弱しくを感じる傾向を確認できた。

毛の長さを変えた場合とは異なり、毛の配置間隔を変えた場合では、かたさの印象について差がでないことがわかる。さらに、材料を導電性／磁性に変えた場合でも、触感に大きな変化を与えないと考えられる。

評価実験の結果は、今後毛構造を用いたアプリケーションの実装に活用できると考える。例えば、毛を撫でてリラックスするためのデバイスを作りたいユーザには、やわらかい／なめらかな触感を重視して、「毛を長くする／毛の密度を高くする」等の設定を推薦できる。

6 制作例

6.1 導電性の毛構造を使った接触検知

導電性の毛に接触することで LED が点灯する作例を図 11 に示す。使用した毛構造は触れた際にやわらかい印象を得られるように、毛を 80 mm と長めに造形した。静電容量の測定には Arduino の CapacitiveSensor ライブライアリを用いており、手が接

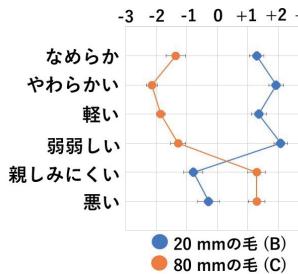


図 8. 毛の長さが異なる場合 (A-B, A-C) の平均と標準偏差

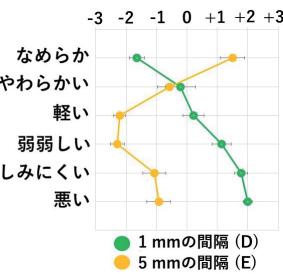


図 9. 毛の間隔が異なる場合 (A-D, A-E) の平均と標準偏差

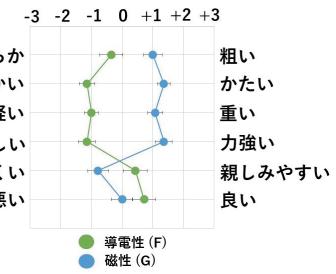


図 10. 毛の素材が異なる場合 (A-F, A-G) の平均と標準偏差

触することによる静電容量の変化を検出している。この制作例では、3つの状態（触れていない状態、指が触れている状態、指が多く触れている状態）を検出することができる。現状では、指で撫でる際の位置や方向の検出はできていないが、毛構造と回路が接触する部分を工夫することによって可能になると見える。毛構造等をトラックパッド等に固定して、撫でた位置や方向を検出するようなインターフェースに応用できると考える。

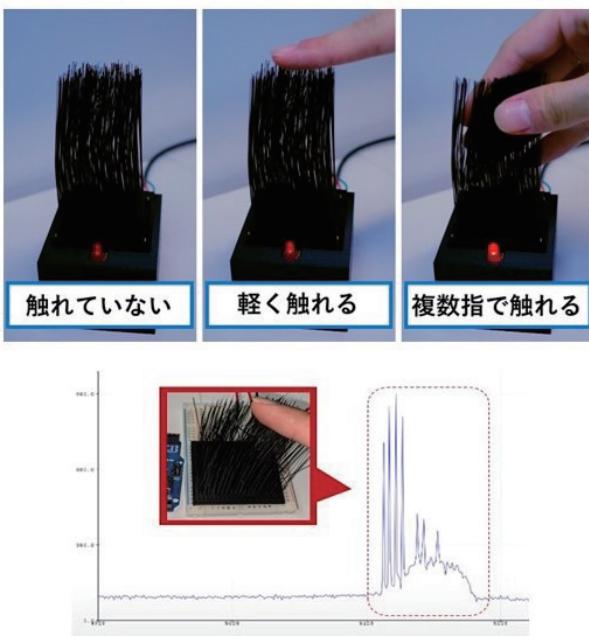


図 11. 毛への接触による LED の点灯および静電容量の変化

6.2 磁性を持つ毛を筆にした例

磁性の毛構造と永久磁石を使った作例について紹介する。まず、平筆のような形状の毛構造を棒状の立体物に取り付けた。次に磁石を一定間隔で配置できるボードを3Dプリントで作成した。ブラシを持ち、磁石が一定間隔で配置されたボードの上を筆でなぞることで、ボードの磁石に毛先が引き寄せられる（図12左）。そのため、ブラシでボード上の記号をなぞる際に部分的な抵抗感がある触覚を表現することができる。この抵抗感を感じる仕組みを利用して、埋め込む磁石の位置や磁力の大きさを制御することによって、特定の文字やパターンによる触覚をユーザに提示するアプリケーションの設計が可能であると考える。

6.3 ソフトレジンに埋め込んだ磁性フィラメント

毛構造をソフトレジンに埋込むことで、毛構造の土台部分を柔軟に扱うことができるようになった。

そのため、毛構造を手足などの湾曲した部分にも貼り付けることが可能になった。図12右では、毛の部分が磁性フィラメント、土台部分がソフトレジンの毛構造を手のひらに両面テープで貼り付けている。このようにすることで、磁石を近づけた際に、磁力による引き付けられる触覚を毛が集まる様子だけでなく手のひらの感覚で感じることができる。磁石を動かすことで、磁石の位置に伴う磁力の変化を感じることができた。

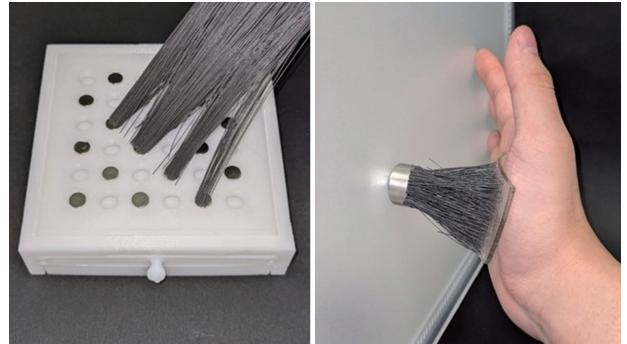


図 12. 磁性フィラメントを使用した制作例

7 まとめと今後の展望

本研究では、3Dプリントによって造形可能な毛構造を用いた情報提示手法の構築を目指し、そのために毛構造へ導電性／磁性を付与するための方式、ソフトレジンに植毛する方式を提案した。毛構造は3Dプリントによって、情報提示の目的に合わせて自由に触感を変更することができるため、制作物への没入感を向上させることにつながる。提案した方式を元に、毛構造を電子的に制御可能な新たなインターフェース・コンポーネントとして確立していくたい。

謝辞

本研究の一部は、科研費 20H04231 の支援を受けた。

参考文献

- [1] D. Degraen, A. Zenner, and A. Krüger. Enhancing texture perception in virtual reality using 3D-printed hair structures. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pp. 1–12, 2019.
- [2] K. Kamata, H. Takahashi, and K. Tsukada. Conductive, Ferromagnetic and Bendable 3D Printed Hair for Designing Interactive Objects. In *Adjunct Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, 2023.

導電性・強磁性・及び土台の可撓性を持つ毛構造の造形手法の提案

- [3] G. Laput, X. Chen, and C. Harrison. 3D printed hair: Fused deposition modeling of soft strands, fibers, and bristles. In *Proceedings of the 28th annual ACM symposium on user interface software & technology*, pp. 593–597, 2015.
- [4] C.-J. Lee, H.-R. Tsai, and B.-Y. Chen. Hair-touch: Providing stiffness, roughness and surface height differences using reconfigurable brush hairs on a vr controller. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–13, 2021.
- [5] S. J. Leigh, R. J. Bradley, C. P. Purssell, D. R. Billson, and D. A. Hutchins. A Simple, Low-Cost Conductive Composite Material for 3D Printing of Electronic Sensors. *PLOS ONE*, 7(11):1–6, 11 2012.
- [6] J. Ou, G. Dublon, C.-Y. Cheng, F. Heibeck, K. Willis, and H. Ishii. Cilllia: 3D printed micro-pillar structures for surface texture, actuation and sensing. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 5753–5764, 2016.
- [7] M. Schmitz, M. Khalilbeigi, M. Balwierz, R. Lissermann, M. Mühlhäuser, and J. Steimle. Capricate: A Fabrication Pipeline to Design and 3D Print Capacitive Touch Sensors for Interactive Objects. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, UIST ’15, p. 253–258, New York, NY, USA, 2015. Association for Computing Machinery.
- [8] H. Takahashi and J. Kim. Designing a Hairy Haptic Display using 3D Printed Hairs and Perforated Plates. In *Adjunct Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, pp. 1–3, 2022.
- [9] S. Umezawa, M. Ohkubo, Y. Ooide, and T. Nogima. Hairlytop interface: A basic tool for active interfacing. In *Adjunct Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, pp. 95–96, 2014.
- [10] 遠藤暁友, 原田雅文, 錦織さくら, 佐藤俊樹. 微弱な接触を検知可能な毛状タッチセンサ. In *In Proceeding of WISS2022*. 日本ソフトウェア学会, 2022.
- [11] 鎌田航誠, 高橋治輝, 塚田浩二. 導電性樹脂・磁性樹脂を用いて3Dプリントした毛構造の拡張と応用. In Proceeding of インタラクション2023. 情報処理学会, 2023.
- [12] 高橋治輝, 宮下芳明. ブリッジ構造と樹脂の引き伸ばしを用いた高密度な毛構造の造形手法. In Proceeding of WISS2017. 日本ソフトウェア学会, 2017.
- [13] 高見太基, 斎藤大雅, 亀岡嵩幸, 梶本裕之. スライムにピンアレイを刺突することによる表面凹凸の変化が平滑感知覚に与える影響. In Proceeding of 第26回日本バーチャルリアリティ学会大会論文集. 特定非営利活動法人日本バーチャルリアリティ学会, 2021.
- [14] 中島康祐, 伊藤雄一, 築谷喬之, 藤田和之, 高嶋和毅, 岸野文郎. FuSA 2 Touch Display: 大画面毛状マルチタッチディスプレイ. 情報処理学会論文誌, 53(3):1069–1081, 2012.

MusicCommentVisualizer: 音楽動画へのコメントの可視化に基づく音楽推薦・鑑賞インタフェース

倉持 友哉^{*} 濱崎 雅弘[†] 中野 倫靖[†]

概要. 本論文では、音楽動画に対する YouTube コメントを可視化しながら、ユーザの状況や感情に合った音楽推薦（プレイリスト自動生成）と音楽鑑賞を実現するインタフェース MusicCommentVisualizer を提案する。従来、ユーザの状況や感情に合った曲を聴くための音楽推薦やインタラクティブシステムに関する研究はあったが、音楽動画へのコメントをもとに状況と感情の両方を考慮した研究はなかった。ユーザの状況と感情は、音楽鑑賞時におけるユーザの音楽の好みへ影響を与えるため、これら両方を同時に考慮できることは、多様な音楽推薦・鑑賞を実現するうえで重要である。MusicCommentVisualizer は音楽動画へのコメントから、状況と感情に関する単語を自動的に抽出して分類し、インタラクティブに操作可能な出現頻度分布として可視化する。さらに、その分布からユーザによって選択された状況と感情に基づいて、音楽プレイリストを生成したり、それらの楽曲が推薦された理由となるコメントを表示したりできる。

1 はじめに

オンライン音楽ストリーミングサービスの普及によって人々が聴くことのできる楽曲数は膨大となり、人々の音楽鑑賞を支える技術として音楽推薦システムが活用されている [18]。そのような音楽推薦システムは、主に音楽の情報（タグや音響信号等）に基づいて動作しているが、人の音楽の好みは楽曲を聴く時の状況や感情に影響されることから [2, 19]、ユーザの情報を反映できるシステム構築が望ましい。例えば、睡眠前に聴きたい音楽とジムで運動するときに聴きたい音楽は異なったり、嬉しい感情と悲しい感情でも音楽の好みが変化したりするが、そのような動的な変化を反映できるとより効果的である。

従来、音楽推薦システムとして、音楽に関する情報や音響信号等を用いる研究 [4, 12] や、ユーザの状況や感情等を用いる研究があった [9, 11, 19]。ここで、ユーザのコンテキストを扱う推薦は、コンテキストアウェア音楽推薦システムと呼ばれる [25]。ユーザのコンテキストとしては、ユーザの置かれた状況（場所、時間帯、曜日、気温、天候、活動、等）[19] を、ユーザの持つデバイスやセンサから取得して活用する [9]。このようなユーザのコンテキストに基づいた音楽推薦を実現する為には、音楽に関する情報が必要となり、ソーシャルタグや歌詞等のテキスト情報 [12]、音色やメロディ等の音響情報 [4]、が用いられた。また、自動的な推薦システムの他にも、Spotify や Amazon Music 等、人手によって楽曲選定されたプレイリスト（「朝」など、状況を表す名称が付与されている）をユーザが選択することもある。

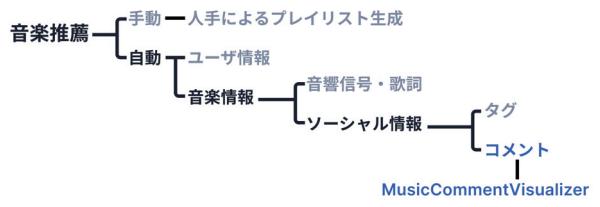


図 1. 音楽推薦における MusicCommentVisualizer の位置づけ。音楽に関するソーシャル情報として、自然言語による大量のコメントを扱う点で新しい。

図 1 に、音楽推薦の分類を示す¹。我々の知る限り、音楽に対するソーシャルメディア（YouTube 等）上のコメントを音楽推薦システムに活用した研究はない。そのようなコメントには、実際に音楽を聴いて得た感情や、音楽を聴いている状況を含む多様な発信が存在することを我々は確認しており [28]、新しいコンテキストアウェア音楽推薦システムの一つとしてコメントを用いることが有益だと考えた。

そこで本論文では、音楽動画のコメントをもとに、状況や感情に合った音楽を推薦・鑑賞するためのインタフェース MusicCommentVisualizer を提案する（図 2）。ここで、楽曲のミュージックビデオや「歌ってみた」と呼ばれる同曲異唱コンテンツ [23] などが発信される、音楽との相性が良いソーシャルメディアの一つである YouTube のコメント（以下 YouTube コメントと呼ぶ）を推薦に用いる。

MusicCommentVisualizer は、ユーザが望む状況や感情を「能動的」に選択して音楽が推薦される

Copyright is held by the author(s).

* 筑波大学大学院

† 産業技術総合研究所

¹ 「音楽情報」の下位分類を内容（歌詞と音響信号）とソーシャルな情報としたが、テキスト情報 [12] と音響情報 [4] に分類することも考えられる。

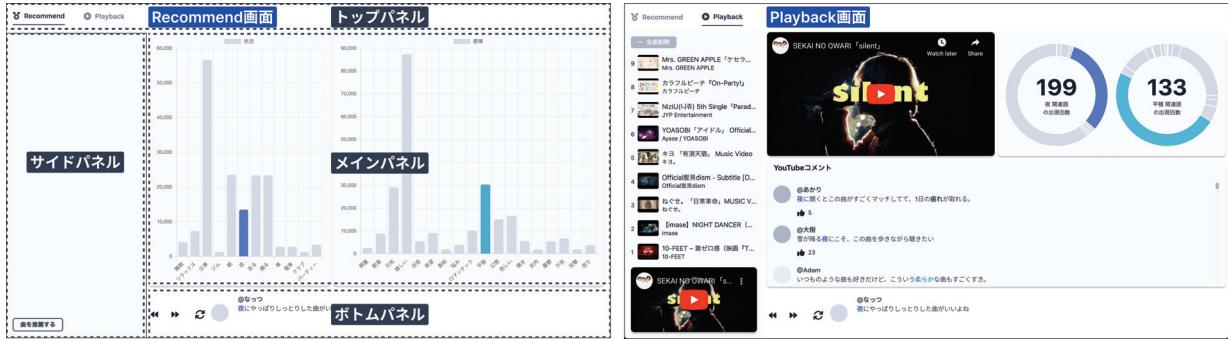


図 2. 音楽コンテキスト（感情と状況）を選択しプレイリストを生成する音楽推薦機能（左、Recommend 画面）と、選択された音楽コンテキストを含むコメントを可視化しながら音楽動画を鑑賞する機能（右、Playback 画面）

制御性と、推薦理由となったコメントが提示されてその理由が明確となる透明性とを持ち合わせた音楽推薦機能と音楽鑑賞機能をそれぞれ持つ。音楽推薦を全てシステムに任せる「受動的」な設計は、ユーザ自身がシステムをコントロールできる制御性や、ユーザに推薦の過程や理由を理解させる透明性を、いずれも低下させてしまう [10, 25]。ユーザの感情や状況に合わせて推薦すべき楽曲は常に同じではなく、変わりうるため、制御性は重要である。具体的には、ある感情状態で好む音楽は個人差が大きいこと、感情は変化しやすく、時と場合によって音楽の好みが変化すること、が知られている [19]。例えば、悲しい時に元気の出る音楽を聴きたい人がいたり、逆に、悲しい時には悲しい音楽を聴きたい人がいたりするし、状況が変わることで、好む音楽がその時々で変わりうることも容易に想像できる。さらに、システムが推薦した音楽がユーザの選択した状況や感情と合っている理由を説明できる透明性（説明可能な推薦システム）も、システムへの信頼を確立する上で重要である [20]。またそのような透明性は、ユーザの納得感や満足度を高めると考えられる。

図1に示すように、MusicCommentVisualizerでは、そのユーザが実際に置かれている状況やその真の感情（ユーザ情報）を用いるのではなく、音楽につけられたコメントに基づく状況と感情（音楽情報）を用いる。これらは、音楽や音楽動画から動画視聴者が感じた状況や感情であり、MusicCommentVisualizerのユーザ情報と明確に区別するためにこれらを「音楽状況」、「音楽感情」と呼ぶ。また、音楽状況と音楽感情を合わせて「音楽コンテキスト」と呼ぶ。つまり、ユーザが望む状況や感情を指定することで、それらをYouTubeの視聴者に対して誘発したであろう音楽をMusicCommentVisualizerは推薦する。

2 MusicCommentVisualizer

MusicCommentVisualizerのインターフェース画面は図2に示すように、画面上部のタブ（トップパネ

ル）で、音楽推薦機能（Recommend ボタンで選択）もしくは音楽鑑賞機能（Playback ボタンで選択）を切り替えることができる。いずれの場合も、トップ・サイド・メイン・ボトムの4つのパネルで構成され、メインパネルだけが機能に応じて切り替わる。

2.1 YouTubeコメントの分析結果を用いた インタラクティブな音楽推薦と音楽鑑賞

MusicCommentVisualizerでは、音楽推薦及び音楽鑑賞のいずれの場合においても、音楽コンテキスト（音楽状況と音楽感情）をクリッカブルな出現頻度分布で可視化し、これを音楽コンテキスト分布（音楽状況分布と音楽感情分布）と呼ぶ。この音楽コンテキスト分布は、本論文では、686曲 265,317件のYouTubeコメントから、状況や感情を説明する単語を自動的に抽出して生成したものである。Recommend画面においてはメインパネル（図2）に棒グラフとして表現され、Playback画面においてはドーナツグラフとして表現される。

ユーザはまず、Recommend画面の音楽コンテキスト分布を用いて、推薦に用いた状況と感情を決定し、音楽推薦機能により、音楽プレイリストを生成する。ここで、ユーザの決定を支援する目的で、ボトムパネルには選択した音楽コンテキストを含むYouTubeコメントがランダムに1つ表示され、更新して別のコメントを表示することもできる。つまりユーザは、共感できる意見や感想を含むYouTubeコメントをインタラクティブに探索でき、その探索を通して、求める音楽コンテキストを選択できる。

生成されたプレイリストはサイドパネルに表示され、最下段の楽曲が再生される。Playback画面に移動することで、その楽曲に対するコメントをメインパネル上で最大100件閲覧することができ、その楽曲に対するコメントの音楽コンテキスト分布も可視化される。このように推薦理由となるコメントが可視化されることで、再生中の楽曲が、ユーザの指定した音楽コンテキストに合っているかを確認できる。

このような音楽に付与されたコメントを用いるこ

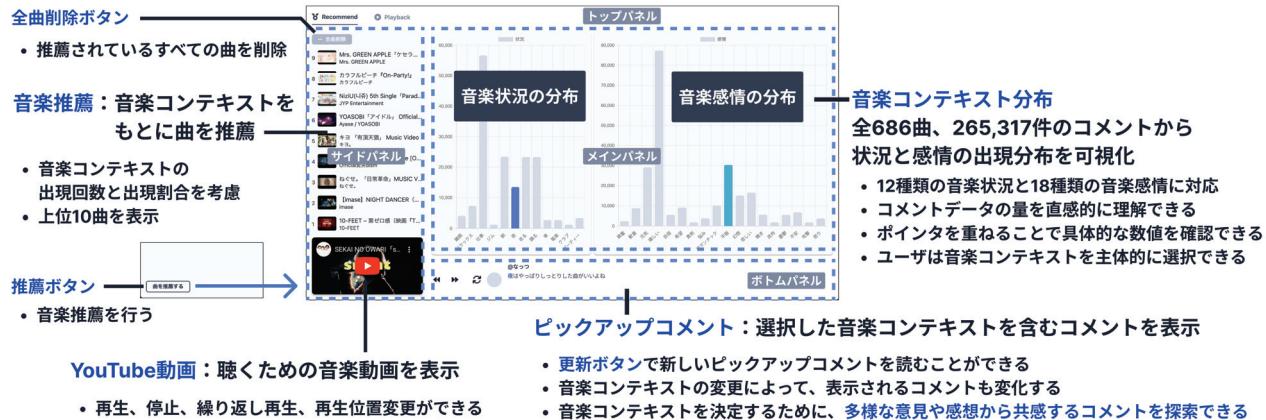


図 3. 音楽推薦機能 (Recommend 画面) の詳細

とで、音楽の多様な観点からの推薦を可能とする。まず、音楽状況と音楽感情を考慮できて有用である。具体的には、「夜」という状況と「不安」という感情を指定した場合、「夜に聴く」とコメントされた曲、不安な気持ちがコメントされた曲、不安な学生の心に響くとコメントされた曲などが実際に推薦された。また「電車」と「元気」の場合、電車で通勤や通学時に聴くとコメントされた曲、「元気が出る」とコメントされた曲などが推薦された。音楽の音響信号からのみでは、このような音楽推薦は難しいが、コメントを利用することで、音楽を聴いた状況や聴きたい状況と感情を扱える点が特長である。

次に、特徴的な歌詞を考慮できて有用である。コメントには視聴者が感銘を受けた歌詞が書かれることがあるため、「夜」や「不安」などの音楽コンテキスト分布を構成する単語が含まれている場合にその曲を推薦できる。しかもこれは、単に歌詞を用いた推薦とは異なり、動画視聴者が着目する歌詞を特に考慮できる推薦となって、特長的である。

最後に、音楽の社会的背景を考慮できる。コメントには音楽や音楽動画の映像に向けたものだけではなく、その外的要因として、アーティストの人間性や個性、音楽と関連した別のコンテンツ（アニメや映画など）に向けたものなどがある。音楽鑑賞においては、その音楽的な内容だけではなく、アーティストや映画のシーン等と関連した感情が誘発されることがあり、そのようなコメントも対象とできる。

以降、音楽推薦機能と音楽鑑賞機能におけるインタラクションを説明する。具体的には、音楽推薦に用いる音楽コンテキストをどのように決定するか、その結果、どうプレイリストが推薦されるか、そのプレイリストの楽曲をどう鑑賞するかを説明する。

2.2 音楽推薦インタラクション

音楽推薦機能としての Recommend 画面例を図 3 に示す。ユーザは、音楽プレイリストを生成する音

楽推薦のために、まず音楽コンテキストを決定する必要がある。メインパネルには YouTube 上の楽曲 686 曲に付与された 265,317 件の YouTube コメントに基づく音楽コンテキスト分布が可視化されており、マウスポインタをこの分布に重ねることで各音楽コンテキストの出現回数を見ながら、推薦に用いるコンテキストをクリックして決定できる。それぞれの音楽コンテキストとして、12 種類の音楽状況の分布（左）と、18 種類の音楽感情の分布（右）を棒グラフで可視化する。これらの音楽コンテキストは、それぞれ状況 [9] と感情 [6,7] に関する先行研究に基づいて決定した。分布の可視化により、ユーザは音楽コンテキストの出現傾向を把握できる。

ユーザは自身の状況や感情をもとに、この分布の棒の部分をクリックして選択することで、音楽コンテキストを選択できる（図 3 では、状況として「夜」、感情として「平穏」を選択）。この時、今の自分の状況や感情を選択するか、今の自分の状況や感情ではないが望むものを選択するか、はユーザの自由である。音楽コンテキストとしては、音楽状況と音楽感情のいずれか 1 つ、または両方を選択できる。

ユーザが分布の棒グラフから音楽コンテキストを選択すると、選択した音楽コンテキストが含まれるコメント 1 件がボトムパネルに表示される。これをピックアップコメントと呼び、ユーザの求める音楽コンテキストの決定を支援（状況と感情を決定するための参考情報として活用）する、本システムの特長的なインタラクションの一つである。ピックアップコメントでは音楽コンテキストを表す単語が太字で表示される。その中でも、選択中の音楽コンテキストを表す単語については、太字かつ着色されて強調表示される。これにより、コメントの中で音楽コンテキストに関する文や単語を認識しやすくなる。

ピックアップコメントは音楽コンテキストを選択する度に、異なるものが表示される。また、表示されているピックアップコメント左側の更新ボタンを



図 4. 音楽鑑賞機能 (Playback 画面) の詳細

押すことで、現在選択している音楽コンテキストのいずれかを含む新しいピックアップコメントに更新される。従来、状況や感情を入力するには、それを表す単語をクエリとして決定する必要があった。しかし、自身の求める状況や感情が確実に定まっていないような場合は、そのような方式が使いにくい。それに対して本インタラクションでは、自分が共感できたり、自分の状況や感情に気づきを与えてくれる実際のコメントをインタラクティブに表示できるため、音楽コンテキストを探索的に決定できる。

サイドパネルにある推薦ボタンを押すことで、現在選択中の音楽コンテキストに基づく 10 曲が同パネルに推薦される。先頭の音楽は YouTube 動画として再生や停止、繰り返し再生、再生時刻変更が可能である。また、ボトムパネルのアイコンで、プレイリスト中の前後の音楽を再生できる。プレイリストを再度推薦したい場合は、サイドパネルの全曲削除ボタンを用いる。推薦されているすべての曲が削除され、同パネルに推薦ボタンが再度表示される。

2.3 音楽鑑賞インタラクション

音楽鑑賞機能 (Playback 画面) を図 4 に示す。ここで、メインパネルは 3 つのフィールドを含む。1つ目は、動画フィールドであり、現在鑑賞している音楽動画を観覧用に大きく表示する目的を持つ。前述の通り、YouTube コメントでは楽曲の内容だけでなく、映像の内容について言及されることがあるため、大きな再生画面は重要である。

2 つ目は、コメントフィールドである。これは現在鑑賞している音楽に対して、選択した音楽コンテキスト (状況 and/or 感情) を含むコメントから、最大 100 件としてサンプリングして表示するフィールドである。各コメントにおける音楽コンテキストを表す単語は、ピックアップコメントと同様、太字や色付きで強調表示される。これらのコメントが推薦に利用されているため、ユーザは現在鑑賞している音楽が音楽コンテキストに合っているとして推薦された理由を理解できる。つまり透明性を高める役割があり、前述の通り、推薦システムへの信頼や推

薦結果に対する納得感につながると考えている。さらに、コメントに含まれる多様な感想を読むことで、自身の知らない新たな視点から見たアーティストや楽曲、また音楽動画に対する各個人の想いを楽しめる。よって、推薦結果の音楽を聴いて楽しむだけでなく、推薦理由を読んで楽しめる点も特長である。

3 つ目は、分析フィールドである。これは現在鑑賞している音楽について、音楽コンテキストを表す単語の出現回数とその割合を可視化するフィールドである。音楽状況分布が左、音楽感情分布が右にドーナツグラフで表示され、それぞれの中央に今選択している音楽状況・音楽感情を表す単語の出現回数が表示される。音楽推薦では各音楽コンテキストの出現回数だけでなく、出現割合も考慮しているため、現在鑑賞中の楽曲について、選択した音楽コンテキストの出現回数と割合が多いことをユーザは理解でき、透明性と推薦への納得感が向上する。

ここで、出現頻度を表すドーナツグラフはクリックブルであり、異なる音楽コンテキストを選択できる。これによりピックアップコメントとコメントフィールドを、新たに選択した音楽コンテキストを含むコメントに更新できる。つまり、今聞いている楽曲に対して異なる音楽コンテキストに関して、該当するコメントを鑑賞して楽しんだり、異なる観点から鑑賞したりすることにつながる。

3 データセット構築

システム実装のために収集したデータと構築したデータセットについて述べる。

3.1 YouTube データの収集

本研究で利用する YouTube コメントに関連するデータ (YouTube データと呼ぶ) を収集した。収集にあたって YouTube Data API² を用いた。また、週一回更新される「人気の曲トップ 100 - 日本」プ

² <https://developers.google.com/youtube/v3>

プレイリスト³を収集の対象として、2022年11月1日から2023年7月31日の9ヶ月間収集した。

収集した3種類のデータを以下に示す。

動画データ プレイリストの各動画データ。ただし、非公開に設定されている動画がプレイリスト内に含まれる場合、動画数は100曲以下となる。

コメントデータ 各動画のコメントに関するデータ。それぞれの動画に対して、人気順に最大2000件程度収集した。さらに、それらのコメントに対する返信を収集した。

メタデータ 順位、動画タイトル名、投稿者名、いいね数、視聴回数等のデータ。

コメント数や視聴数、いいね数の毎日の変動を追うため、プレイリスト内の100曲分を毎日収集した。毎日収集するため、またランキングが更新されても同じ動画が出現することもあるため、取得する動画とコメントが重複する。そのような重複を除き、686曲の音楽動画と3,828,473件のコメント（コメント作成者のチャンネルID数は696,946）を用いた。

3.2 単語セット

音楽コンテキストとして扱う音楽状況と音楽感情の種類（カテゴリ）は、我々の先行研究[28]で構築したものを用いた。音楽状況カテゴリは、Ibrahimらの先行研究[9]に基づく**睡眠**, **仕事**, **ジム**, **朝**, **夜**, **走る**, **踊る**, **車**, **電車**, **クラブ**, **パーティー**, **リラックス**の12種類である。音楽感情カテゴリは、音楽感情認識で用いられるラッセルの円環モデル[17]との整合性が示されたHuらの先行研究[6, 7]に基づく**平穏**, **悲しい**, **嬉しい**, **口マンチック**, **歓喜**, **憂鬱**, **怒り**, **嘆き**, **幻想**, **元気**, **悩み**, **攻撃**, **不安**, **自信**, **希望**, **真剣**, **皮肉**, **興奮**の18種類である。

これらのカテゴリに該当する単語の一式を単語セットと呼び、単語分散表現をk-means法により自動的にクラスタリングした後、人手で選別して構築した（手法の詳細は[28]参照）。ここで、上記カテゴリごとに単語（以下、音楽コンテキスト単語と呼ぶ）が1つ以上含まれるよう構築した。

音楽コンテキスト単語としては、例えば「睡眠」という状況カテゴリにおいては、朦朧、疲労、寝言、寝坊、遅刻等があった。また、「リラックス」という状況カテゴリにおいては、リラックス、くつろぎ、ゆったり、のんびり、まったり等があった。

3.3 データセット

YouTubeデータと単語セットから、コメントの特徴ベクトルと楽曲の特徴ベクトルを作成し、またコメント中の音楽コンテキストをマークアップした

データセットを構築した。まず、YouTubeコメントへの返信は、楽曲に対する直接的なコメントではないため、取り除いた。コメントは日本語を対象とし、ひらがな、カタカナ、漢字のいずれかを含むコメントのみを用いた。さらに、各コメントに対してMeCabを用いて単語分割と原形変換を行い、音楽コンテキスト単語を含まないコメントを除いた。

その結果、コメント数は265,317件になった。各音楽動画に付けられたコメント数の平均は437.82回、標準偏差は2262.64回であった。また、音楽動画ごとの各ユーザのコメント数の平均は1.64回、標準偏差は5.65回であった。YouTubeのコメントは一人で何回でも発言可能だが、1~2回コメントするユーザが多く、コメント数全体の60.33%を占める。

そのようにして得られたコメントごとに、音楽コンテキスト単語の出現回数を算出し、それをカテゴリごとに合計して、コメントの特徴ベクトルとする。各楽曲の特徴ベクトルは、その楽曲の全コメントの特徴ベクトルを加算して得る。すなわち、各コメントと各楽曲は、いずれも12次元の音楽状況ベクトルと18次元の音楽感情ベクトルで表現される。

次に、それぞれのコメントに対して、音楽コンテキスト単語をHTMLタグで囲み、各音楽コンテキストを表すクラスを設定したテキストを作成する。例えば「私は嬉しい」は「私は<strong class="glad">嬉しい」のようになる。

現在の実装では、否定表現もカウントされてしまうが、定性的な印象としては、その出現回数は肯定表現に比べて少なく、推薦への影響は小さいと考える。今後、否定表現の調査や対応が課題である。また、コメントの質を考慮して、不適切なコメントを除いたり、「高評価」が多いコメントを考慮できたりすると、より効果的な推薦につながる可能性がある。

4 システム実装

MusicCommentVisualizerはNext.jsとDjangoを用いて実装した。各機能の実装内容を述べる。

4.1 可視化の実装

サイドパネルと動画フィールドのYouTube動画の表示にはYouTube IFrame Player API⁴を用いた。サイドパネルに表示されるYouTube動画は、映像も表示するが、音楽の再生が目的であり、音楽動画は操作に支障が出ない最低限のサイズとした。

音楽コンテキスト分布を表すRecommend画面の棒グラフとPlayback画面のドーナツグラフはreact-chartjs-2⁵ライブラリを用いて実装した。Recommend画面における音楽コンテキスト分布について

⁴ https://developers.google.com/youtube/iframe_api_reference

⁵ <https://react-chartjs-2.js.org>

は、全楽曲における全音楽コンテキストの出現回数を、棒グラフの長さとして可視化する。ただし、出現回数が相対的に少ないことが原因で、棒グラフが短くなりすぎてクリックできなくなることを防ぐため、最低長さとして 10px を設定した。

ピックアップコメントには、選択中の音楽コンテキストに該当する単語が 1 回以上出現するコメントからランダムに 1 つ選択し、表示する。コメントフィールドには、現在鑑賞している曲に対して、現在選択されている音楽コンテキストを含むコメントをランダムに最大 100 件表示する。100 件のコメントは、1 曲を鑑賞中に読める十分な量であった。

選択中の音楽コンテキストを表す単語については、選択された音楽カテゴリに応じて CSS を適用する対象（クラス）を変更することで、色変更を行う。また、コメントに対するいいね数やユーザプロフィール画像は、YouTube データに含まれる値と URL を利用して、コメントと同時に表示する。

4.2 音楽推薦アルゴリズム

前章で構築したデータセットを利用して音楽推薦を行う。まず、各楽曲 s の音楽状況ベクトルを \mathbf{x}_s 、音楽感情ベクトルを \mathbf{y}_s 、選択された音楽状況を i 、選択された音楽感情を j として、全 686 曲を、4 つの基準で順位付けし、それぞれ $R_1 \sim R_4$ と呼ぶ。

- $R_1(s)$: $\mathbf{x}_s(i)$ に基づく順位
- $R_2(s)$: $\mathbf{x}_s(i) / \sum_i \mathbf{x}_s(i)$ に基づく順位
- $R_3(s)$: $\mathbf{y}_s(j)$ に基づく順位
- $R_4(s)$: $\mathbf{y}_s(j) / \sum_j \mathbf{y}_s(j)$ に基づく順位

ここで、 $\mathbf{x}_s(i)$ は楽曲 s における音楽状況 i の出現回数、 $\mathbf{y}_s(j)$ は楽曲 s における音楽感情 j の出現回数を意味する。次に、楽曲 s ごとに $R(s) = \sum_k R_k(s)$ を求め、 $R(s)$ が小さい（順位が高い）楽曲から上位 10 件を音楽推薦の結果として表示する。ここで、音楽状況のみが選択されている場合には $R(s) = R_1(s) + R_2(s)$ 、音楽感情のみが選択されている場合には $R(s) = R_3(s) + R_4(s)$ とする。

この推薦アルゴリズムにした理由は 2 つある。まず、出現回数 ($R_1(s)$ と $R_3(s)$) を考慮することで、選択された音楽コンテキストのコメント量が多い曲を推薦することができる（理由 1）。コメントフィールドに少量のコメントしか表示されないと、透明性を保つことができず、ユーザのコメントを読む楽しさも損なわれるため、コメント量が多くなるよう出現回数を評価した。しかしそれだけでは、コメントの総数が多い曲ほど評価されてしまって不適切であるため、割合として $R_2(s)$ と $R_4(s)$ も用いた（理由 2）。出現回数と割合の両方を考慮することで、コメントフィールドに十分な量のコメントを表示しつつ、推薦の多様性を高めるアルゴリズムとした。

今後はさらに、同じ状況と感情が選択されても異なる推薦結果となるように、音楽推薦結果のバリ

エーションを増加させる実装を予定している。

5 関連研究

ソーシャルメディア上の動画に対するコメントを対象とした研究として、ユーザのコメントを動画の検索やランキングに活用する研究 [22, 27] や、音楽動画の印象推定へ活用する研究 [26] がある。

また、インタラクティブな音楽推薦・探索システムのインターフェースについて研究が行われてきた [13]。例えば、音楽のソーシャルタグをもとに音楽探索・推薦を行う TagFlip [10]、アーティストを繋げたグラフによるジャンルの探索支援ツール TastePaths [16]、同曲異唱コンテンツを連続再生する Songrium RelayPlay [23]、歌詞のトピックに基づく楽曲探索インターフェース Lyric Jumper [21] などが提案された。可視化も研究されており、音楽ジャンルの等高線プロットによる可視化 [15] や、歌声から歌い方の傾向を可視化した SingDistVis が提案された [24]。

ユーザの状況に合った音楽推薦に関する研究では、旅程を提案し、ユーザが訪れた名所に合った曲を推薦する PlayingGuide [2] や、Ibrahim らのユーザや楽曲、デバイス情報をもとにした、ユーザ状況に合うプレイリスト生成システム [9] が提案された。

楽曲感情に着目した研究もある [5]。Hu らは楽曲の歌詞と音響特徴を用いて 18 カテゴリへのムード分類をし [6–8]、Laurier らはソーシャルタグを用いて 4 つのムードを表すクラスタを作成した [14]。感情を扱う音楽推薦システムについては、感情タグを利用した MoodPlay [1] や、ユーザが閲覧中の Web ページを利用した MusicSense が提案された [3]。

これらに対し、MusicCommentVisualizer では音楽動画に対するコメントをもとに音楽状況と音楽感情の両方を考慮する点、音楽コンテキストをインタラクティブに決定できる点、そして、推薦理由として強調表示したコメントを観覧できる特長がある。

6 まとめ

本論文では、音楽動画のコメントを用いて状況や感情に合った音楽推薦を実現するインターフェース MusicCommentVisualizer を提案した。MusicCommentVisualizer は音楽推薦機能に加えて、音楽鑑賞機能を持ち、共感できる音楽コンテキストをインタラクティブに探索しながら決定し、それを推薦に用いることができる制御性、推薦理由をユーザが理解しやすい透明性を持つ。また、音楽鑑賞時に音楽コンテキストを含むコメントを鑑賞したり、音楽コンテキスト分布を確認できる。

今後は、否定表現等のコメントに関する調査、音楽推薦アルゴリズムの改良、コメントに対するフィルタリング機能の実装、履歴機能等の音楽鑑賞機能の拡充、ユーザスタディを行う予定である。

参考文献

- [1] I. Andjelkovic, D. Parra, and J. O'Donovan. Moodplay: Interactive Mood-based Music Discovery and Recommendation. In *Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization (ACM UMAP 2016)*, Vol. 121, pp. 142–159, 2016.
- [2] M. Braunhofer, M. Kaminskas, and F. Ricci. Location-aware music recommendation. *International Journal of Multimedia Information Retrieval (IJMIR)*, 2(1):31–44, 2013.
- [3] R. Cai, C. Zhang, C. Wang, L. Zhang, and W.-Y. Ma. MusicSense: Contextual Music Recommendation Using Emotional Allocation Modeling. In *Proceedings of the 15th ACM International Conference on Multimedia (ACM MM 2007)*, p. 553–556, 2007.
- [4] Y. Deldjoo, M. Schedl, P. Cremonesi, and G. Pasi. Recommender Systems Leveraging Multimedia Content. *ACM Computing Surveys (ACM CSUR)*, 53(5):1–38, 2020.
- [5] J. S. Gómez-Cañón, E. Cano, T. Eerola, P. Herrera, X. Hu, Y.-H. Yang, and E. Gómez. Music Emotion Recognition: Toward new, robust standards in personalized and context-sensitive applications. *IEEE Signal Processing Magazine (IEEE SPM)*, 38(6):106–114, 2021.
- [6] X. Hu and J. S. Downie. When Lyrics Outperform Audio for Music Mood Classification: A Feature Analysis. In *Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR 2010)*, pp. 619–624, 2010.
- [7] X. Hu, J. Downie, and A. Ehmann. Lyric Text Mining in Music Mood Classification. In *Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR 2009)*, pp. 411–416, 2009.
- [8] X. Hu and Y.-H. Yang. Cross-cultural music mood classification: A comparison on English and Chinese songs. *Proceedings of the 13th International Society for Music Information Retrieval Conference (ISMIR 2012)*, pp. 19–24, 2012.
- [9] K. M. Ibrahim, E. V. Epure, G. Peeters, and G. Richard. Exploiting Device and Audio Data to Tag Music with User-Aware Listening Contexts. In *Proceedings of the 23rd International Society for Music Information Retrieval Conference (ISMIR 2022)*, pp. 186–192, 2022.
- [10] M. Kamalzadeh, C. Kralj, T. Möller, and M. Sedlmair. TagFlip: Active Mobile Music Discovery with Social Tags. In *Proceedings of the 21st International Conference on Intelligent User Interfaces (ACM IUI 2016)*, p. 19–30, 2016.
- [11] M. Kaminskas and F. Ricci. Contextual music information retrieval and recommendation: State of the art and challenges. *Computer Science Review*, 6(2):89–119, 2012.
- [12] P. Knees and M. Schedl. A Survey of Music Similarity and Recommendation from Music Context Data. *ACM Transactions on Multimedia Computing, Communications, and Applications (ACM TOMM)*, 10(1):1551–6857, 2013.
- [13] P. Knees, M. Schedl, and M. Goto. Intelligent User Interfaces for Music Discovery. *Proceedings of the 20th International Society for Music Information Retrieval Conference (ISMIR 2020)*, 3(1):165–179, 2020.
- [14] C. Laurier, M. Sordo, J. Serrà, and P. Herrera. Music Mood Representations from Social Tags. In *Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR 2009)*, pp. 381–386, 2009.
- [15] Y. Liang and M. C. Willemsen. Interactive Music Genre Exploration with Visualization and Mood Control. In *26th International Conference on Intelligent User Interfaces (ACM IUI 2021)*, pp. 175–185, 2021.
- [16] S. Petridis, N. Daskalova, S. Mennicken, S. F. Way, P. Lamere, and J. Thom. TastePaths: Enabling Deeper Exploration and Understanding of Personal Preferences in Recommender Systems. In *27th International Conference on Intelligent User Interfaces (ACM IUI 2022)*, p. 120–133, 2022.
- [17] J. Russell. A Circumplex Model of Affect. *Journal of Personality and Social Psychology*, 39:1161–1178, 1980.
- [18] M. Schedl, P. Knees, B. McFee, D. Bogdanov, and M. Kaminskas. Music Recommender Systems. In *Recommender Systems Handbook*, pp. 453–492, 2015.
- [19] M. Schedl, H. Zamani, C.-W. Chen, Y. Deldjoo, and M. Elahi. Current Challenges and Visions in Music Recommender Systems Research. *International Journal of Multimedia Information Retrieval (IJMIR)*, 7(2):95–116, 2018.
- [20] R. Sinha and K. Swearingen. The Role of Transparency in Recommender Systems. In *Extended Abstracts on Human Factors in Computing Systems (ACM CHI 2002)*, p. 830–831, 2002.
- [21] K. Tsukuda, K. Ishida, and M. Goto. Lyric Jumper: A Lyrics-Based Music Exploratory Web Service by Modeling Lyrics Generative Process. In *Proceedings of the 18th International Society for Music Information Retrieval Conference (ISMIR 2017)*, pp. 544–551, 2017.
- [22] K. Tsukuda, H. Masahiro, and M. Goto. SmartVideoRanking: Video Search by Mining Emotions from Time-Synchronized Comments. In *2016 IEEE 16th International Conference on Data Mining Workshops (IEEE CDMW 2016)*, pp. 960–969, 2016.
- [23] 濱崎 雅弘, 中野 倫靖, 後藤 真孝. Songrium RelayPlay: リレー再生による同曲異唱コンテンツ視聴支援インターフェース. 第23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ピング (WISS 2015) , pp. 1–6, 2015.

- [24] 伊藤 貴之, 中野 倫靖, 深山 覚, 濱崎 雅弘, 後藤 真孝. 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. 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) , pp. 1–8, 2021.
- [25] 奥 健太. 楽曲推薦システム —プレイリスト, コンテキスト, インタラクション—. 人工知能学会誌, 34(3):300–308, 2019.
- [26] 土屋 駿貴, 大野 直紀, 中村 聰史, 山本 岳洋. ソーシャルコメントからの音楽動画印象推定手法の提案. 第 8 回データ工学と情報マネジメントに関するフォーラム (DEIM 2016) , E3–3:1–7, 2016.
- [27] 中村 聰史, 田中 克己. ソーシャルアノテーションに基づく動画検索手法. 第 1 回データ工学と情報マネジメントに関するフォーラム (DEIM 2009) , D6–1:1–8, 2009.
- [28] 倉持 友哉, 濱崎 雅弘, 中野 倫靖. 気分に合った音楽推薦に向けた動画共有サイト上のコメント分析. 情報処理学会 研究報告 音楽情報科学 (MUS) , 2023-MUS-137(47):1–8, 2023.

SwipeGANSpace: 潜在空間の関心次元探索によるスワイプ操作に基づく嗜好画像生成

中島 柚斗 * 楊 明哲 * 馬場 雪乃 *

概要. GAN は従来の生成モデルより高品質な画像を生成できる。GAN の潜在変数を操作することで、好みの画像（嗜好画像）を生成できる。GAN を利用して嗜好画像を生成するには、適切な潜在変数を GAN に与える必要がある。しかし、GAN の潜在空間は高次元であり、探索が困難である。既存研究では、ユーザに複数のスライダを調整してもらい、画像を編集することで、嗜好画像を生成している。本研究では、より簡単なスワイプ操作によって嗜好画像を生成する手法を提案する。探索を効率化するために、潜在空間を主成分分析で縮小し、縮小した空間で探索を行う。ここで、潜在空間上でどの次元がユーザにとって関心があるかを把握するのは難しい。この課題を解決するために多腕バンディットアルゴリズムを利用して、探索すべき次元を動的に決定し、ユーザの関心次元を重点的に探索することで、さらなる探索の効率化を行う。実験の結果、提案手法がベースラインと比べて、効率的に嗜好画像を生成できることを明らかにした。

1 はじめに

高品質な画像生成が可能な Generative Adversarial Networks (GAN) [5] は、その潜在変数を操作することで生成画像を編集できる [6]。GAN を利用してユーザが好みの画像（嗜好画像）を生成するには、適切な潜在変数を GAN に与える必要がある。しかし、GAN の潜在空間は高次元であり、嗜好画像を生成する適切な潜在変数を探索するには、膨大な時間と労力が必要となる。探索を効率化するため、ユーザに複数のスライダを調整させる手法や画像を編集させる手法が提案されている [2]。これらは特にスマートフォンのような小さな画面では操作が難しい。また、嗜好画像を生成する方法として、自然言語から画像を生成する text-to-image の手法 [14] があるが、ユーザにとって嗜好画像を生成する適切なプロンプトを考えることは難しい場合がある。

本研究では、ユーザの負荷を軽減する画像生成手法を提案する。具体的には、ユーザにスワイプ操作のみを求める。図 1 に提案手法のユーザインターフェースを示す。ユーザは画面に提示された画像を左右にスワイプすることで嗜好を表現する。スワイプ操作のみを利用して嗜好画像を生成するには効率的に潜在変数を探索する必要がある。そこで本研究では、StyleGAN の潜在空間に主成分分析を適用し、画像の見た目を大きく変える主成分で構成された部分空間を作成する。この部分空間の中で、ベイズ最適化を用いて潜在変数探索の効率を高める。ここで、GAN の表現能力を維持するためには、部分空間の次元を大きく保つことが必要である。しかし、部分空間の次元を大きくすると、ベイズ最適化は探索効



図 1: スワイプ操作による問い合わせ例

率が悪くなるため、ベイズ最適化のみを利用して嗜好画像を生成することは困難になる。この問題に対処するために、我々は多腕バンディットアルゴリズムを採用する。このアルゴリズムによって、どの次元を探索すべきかを動的に決定し、特にユーザに関連する次元に焦点を当てることで、探索効率を高める。

提案手法がベースラインと比べて効率的に嗜好画像を生成できるかを検証するために、シミュレーション実験とユーザ実験を行った。その結果、提案手法がベースラインと比べて、効率的に嗜好画像を生成できることを確認した。

2 関連研究

GAN を用いた嗜好画像生成を行うために、ユーザに複数のスライダ調整や画像編集を繰り返し問い合わせる手法がある [2]。この手法では、人間の嗜好を目的関数、GAN の潜在変数を探索変数とした最適化問題を考え、human-in-the-loop 最適化を行

Copyright is held by the author(s).

* 東京大学

うことで嗜好画像を生成した。human-in-the-loop最適化とは、人間を計算資源とみなして最適化の反復計算に組み込む手法である。特に人間の嗜好を対象とした最適化問題では、人間が最適化の反復に入ることが有用である。近年、human-in-the-loop用のベイズ最適化の手法の研究が進められている[10, 11, 2, 4, 8, 16]。ベイズ最適化とはブラックボックス最適化手法の1つである。ベイズ最適化は他のブラックボックス最適化手法と比べて、目的関数の評価回数を抑えることができるため、人間の評価を必要とする目的関数の最適化に適している。このアプローチでは、特に人間にとて評価が容易な問い合わせを行うことや、問い合わせ回数を少なくすることが重要である。一般的な次元数のパラメータ空間を扱う場合のこのアプローチでは、スライダ調整[11]やN対比較[10, 16]など、様々な手法が提案されている。特に、既存研究[2]はGANの潜在空間が高次元であるため、複数のスライダ調整や画像編集など複雑な問い合わせをしているが、これは問い合わせの難易度が高く、ユーザの負担が大きい。

3 提案手法

3.1 画像生成のためのスワイプ操作

本研究では、嗜好画像を簡単なスワイプ操作のみで生成できるように設計した新しいユーザインタフェースを提案する。このインタフェースでは、スワイプ操作を繰り返し行うことで、ユーザの好みを捉えて嗜好画像を生成する。インタフェースは、Tinderのようなマッチングアプリを参考に、図1のように一度に1枚の画像を表示する。ユーザは、現在表示されている画像と以前のものを比較する一対比較を求められる。右にスワイプすることで現在の画像に対する好意を示し、左にスワイプすることで以前の画像に対する好意を表現する。このフィードバックに応じて、次の画像が生成され、ユーザに提示される。このプロセスは、嗜好画像が生成されるまで繰り返される。

3.2 効率的な潜在空間探索のための提案手法

3.2.1 画像生成モデル

本研究ではGANの一種、StyleGAN[7]を用いて画像を生成する。StyleGANでは、ランダムベクトル z をMapping Networkに入力し、中間潜在変数 w を得る。この w は年齢や性別といったスタイル情報を反映し、 w をSynthesis network入力することで画像が生成される。これにより、生成画像のスタイルの制御が可能であり、本研究ではこの利点を活用する。

3.2.2 目的関数

嗜好画像を生成するためには、最適な潜在変数を探索する必要がある。そこで本研究では、人間の嗜好を目的関数、StyleGANの中間潜在変数を探索変数とした最適化問題を考え、これを最適化する中間潜在変数 $w \in W$ を探査する。StyleGANのSynthesis networkを $s: W \rightarrow X$ とする。ここで、 $W \subset \mathbb{R}^d$ は $d = 512$ の高次元空間であり、 X は画像データ空間である。ユーザはこの生成モデル s を利用して画像データ $x \in X$ を生成しようとしているが、制御できるのは中間潜在変数 w のみである。画像データ x への嗜好度は関数 $g: X \rightarrow \mathbb{R}$ によって決定されるが、これはシステムにとって未知である。最終的な目標は、以下の最適化問題を解き、嗜好度の高い画像を生成することである。

$$w^* = \arg \max_{w \in W} g(s(w)) \quad (1)$$

ここで、 $g(s(w))$ は任意の点 w に対して計算可能であるが、 $g(\cdot)$ の計算には人間への問い合わせが必要であるため $g(\cdot)$ は評価コストが高く、評価回数をできるだけ少なくすることが重要である。

3.2.3 一対比較データを利用するベイズ最適化

本研究では、式(1)を解くために、目的関数の評価に人間の入力を用いたベイズ最適化を用いる。 g と s の合成写像 $f = g \circ s$ を定義し、 f を目的関数、 w を探索変数としたベイズ最適化を考える。提案手法は、ユーザから直接 f の値を受け取る代わりに、ユーザの一対比較結果を受け取る。Chuらの手法[3]を利用して、ベイズ推定の枠組みをもとに、一対比較の結果を実数値に変換し、 f を計算する。以降は既存のベイズ最適化と同様、ガウス過程回帰と獲得関数の最適化を行い、評価候補点を決定し、ユーザに問い合わせて目的関数の評価を行う。ユーザに問い合わせる際には、StyleGANのSynthesis networkを利用して、 w を画像 x に変換する。

3.2.4 次元削減

一対比較では、各反復で得られるユーザの嗜好情報が少なく、高次元空間で適切な潜在変数を探索するには、ユーザへの問い合わせ回数が増加するという課題がある。これを解決するために、探索空間を小さくすることで、問い合わせ回数を削減する。Elenaら[13]は、高次元空間でベイズ最適化を行う場合、探索空間に主成分分析(PCA)を適用し、実探索空間を縮小することで、効率的な探索が可能なことを示した。GANSpace[6]では、StyleGANの潜在空間にPCAを適用し、画像の見た目を大きく変化させる主成分を特定している。これらから着想を得て、StyleGANの潜在空間にPCAを適用し、画像の見た目を大きく変える主成分のみで構成される部分空

間での探索を行うことで、潜在変数の探索を効率化する。

具体的な手続きは以下の通りである。まず、標準正規分布から N 個のベクトル $z \in \mathbb{R}^d$ をランダムサンプリングする。次に、サンプリングした z を StyleGAN の Mapping network に入力することで、中間潜在変数 $w \in W \subset \mathbb{R}^d$ を獲得する。獲得した中間潜在変数 w を利用して、主成分分析を行い、寄与率上位 d' 個の主成分で部分空間 $W' \subset \mathbb{R}^{d'}$ を作成する。この部分空間 W' 内でベイズ最適化を行う。 W' 内でのベイズ最適化で提案される評価候補点 $w' \in W'$ の次元数は、 $d' < d$ であるため、 w' を直接 StyleGAN の Synthesis network に入力することはできない。そこで、縮小された次元に戻すために主成分分析の逆変換を行うことで、 w' を StyleGAN の潜在空間 W 上に写像し、Synthesis network に入力する。Synthesis network へ入力して得られた生成画像 x をユーザに提示し、直前の生成画像との一対比較結果を獲得する。

3.2.5 重要な次元の探索

GAN の表現能力を維持するためには、部分空間の大きさを大きく保つ必要がある。しかし、部分空間の次元数を大きくすると、高次元空間ではベイズ最適化の性能が低くなり、ベイズ最適化だけでは嗜好画像の生成が難しくなる。そこで本研究では、探索効率を高めるために、多腕バンディットアルゴリズムを利用して、探索すべき次元を動的に決定する方法を提案する。具体的には、あらかじめ各次元に対してベイズ最適化モデルを用意し、多腕バンディットアルゴリズムを用いて探索する次元を選択し、その次元に対してベイズ最適化モデルを用いて候補点を決定する。そして、候補点を画像化し、ユーザのフィードバックを得る。多腕バンディットモデルは、ユーザからの一対比較結果を用いて更新される。このアプローチを用いることで、部分空間に存在する多様な特性を持つ次元から、最も価値のある次元を重点的に選択し、効率的に探索することができる。

本研究では、多腕バンディットアルゴリズムの中で基本的なアルゴリズムである、UCB アルゴリズムを利用する [1]。UCB アルゴリズムでは、イテレーションごとに各次元の UCB スコアを計算し、最も UCB スコアが高い次元を探索する。 i^* を探索する次元とし、次元 i に関する UCB スコア U_i とすると、

$$i^* = \arg \max_{i=1}^{d'} U_i \quad (2)$$

となる。次元 i に関する探索回数を N_i 、報酬の推定値を \hat{r}_i とすると、多腕バンディットモデルを以下のように更新する。

$$\hat{r}_i = \frac{\sum_{j=1}^{N_i} r_{i,j}}{N_i} \quad (3)$$

ここで、 $r_{i,j} \in \{0, 1\}$ は i 次元目が j 回目に選ばれた時の報酬である。直前の生成画像が選ばれたら $r_{i,j} = 0$ 、現在の生成画像が選ばれたら $r_{i,j} = 1$ とする。選ばれた次元 i^* に関して、ベイズ最適化を利用して候補点の i^* 成分を決定する。獲得関数を A 、候補点の i^* 成分を w_{i^*} とすると、

$$w_{i^*} = \arg \max_{w \in R} A(w) \quad (4)$$

i^* 以外の次元については、過去の観測結果の各次元でのベイズ最適化モデルの最良値を割り当てる。各次元での最良値を $w_0, w_1, \dots, w_{d'}$ とすると、

$$w'_t = [w_0, w_1, \dots, w_{i^*}, \dots, w_{d'}] \quad (5)$$

w'_t を主成分分析の逆変換と Synthesis network を利用して画像に変換し、ユーザに問い合わせる。

4 シミュレーション実験

4.1 実験設定

ユーザが安定した回答を提供できると仮定した場合に、提案手法は効率的に嗜好画像を生成できるかを検証するために、シミュレーション実験を行った。我々は、提案手法からバンディットアルゴリズムを除いた対抗手法 (SimpleBO) と提案手法 (BanditBO) を比較した。目標画像を事前に設定し、生成画像と目標画像の類似度を計算し、1つ前のイテレーション時と比較することで、一対比較の結果を作成した。この一対比較結果をシステムに入力し、ベイズ最適化を行う。部分空間 W' の次元数は、 $d' \in \{4, 8, 16\}$ と設定した。10枚の目標画像を、探索空間からランダムサンプリングして用意した。本実験では、Flickr-Faces-HQ というデータセットで学習された、StyleGAN の事前学習済みモデルを利用した [7]。また、2つの画像の類似度を計算するために FaceNet [15] を使用した。FaceNet は、顔画像を埋め込みベクトルに変換する顔認識用の CNN モデルである。FaceNet を用いて、生成画像と目標画像の埋め込みベクトルを求め、埋め込みベクトル間のコサイン類似度を計算する。

4.2 実験結果

図2に、 d' ごとの、類似度の移動平均を示す。 $d' = 4$ の場合は、BanditBO と SimpleBO の差は大きく見られなかった。 $d' = 8$ の場合は、 $d' = 4$ の時と比べて、BanditBO が SimpleBO より効率的に目標画像に近くようになった。 $d' = 16$ の場合は、さらに BanditBO が SimpleBO より効率的に目標画像に近づいた。以上から、探索空間が高次元の場合に、提案手法は対抗手法と比べて効率的に目標画像に近づくことがわかった。

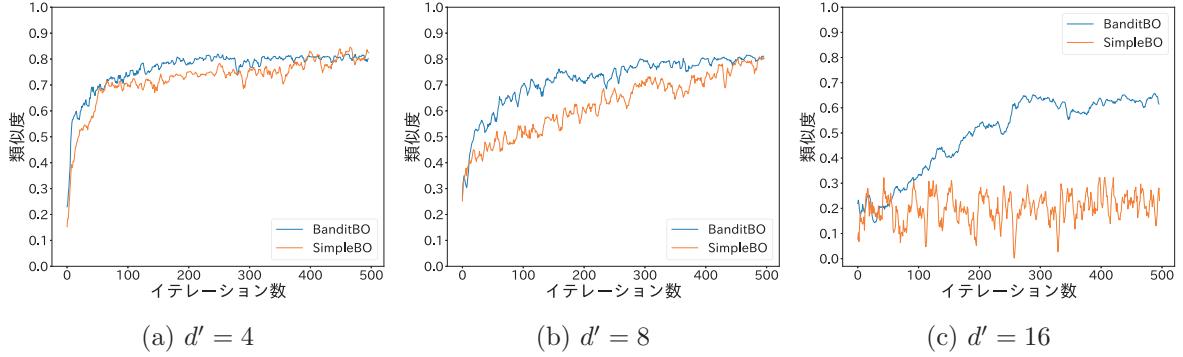


図 2: 各 d' に関する、生成画像と目標画像の類似度の推移（10枚の目標画像での移動平均の推移）

5 ユーザ実験

5.1 実験設定

ユーザの実評価を用いたときに、提案手法がベースラインと比較して効率的に嗜好画像を生成できるかどうかを検証するために、ユーザ実験を行った。被験者には、指定したシナリオに沿った好みのアバターを生成する実験を行った。指定するシナリオは6個(弁護士、教師、スポーツインストラクター、受付、図書館のスタッフ、保育士)用意した。被験者は14人(男性13人、女性1人、年齢:18歳から27歳)であった。被験者ごとに実施されるシナリオの順番はランダムである。各シナリオに対して、BanditBO, SimpleBO, 探索空間から一様分布で中間潜在変数をランダムサンプリングする手法(Random)の3つの手法と、2つの部分空間の次元数 $d' = 4, 16$ の組み合わせをランダムに割り当てた。実験中のユーザの嗜好を確認するために、各シナリオ開始前に、事前の嗜好画像のイメージを尋ねた。被験者には、事前アンケートの内容に固執しないように、与えられたシナリオのみに焦点を当てて画像の比較を行うように指示した。各シナリオでの回答終了後に、画像の提示・比較プロセスおよび最終アバターの満足度に関するアンケートを実施した。全てのシナリオ終了後、実験全体を通して感じたことをアンケートで尋ねた。

5.2 実験結果

図3に、「目的の画像まで効率的に辿り着くことができましたか?」というアンケートの結果を示す。これを見ると、提案手法がベースラインと比べて、目的の画像まで効率的に辿り着くことができたと回答した人が多かったことがわかる。提案手法についての自由記述では、「微調整しながら、途中で大幅に違うものと比較しては元の路線に戻ってくることを繰り返す、良い生成をしていた」、「適当に画像を出してきて、特に調整をしないシナリオや、ずっと微調整だけを繰り返すシナリオなどさまざまなシナリオがあったが、それらを混ぜ合わせたような、微調

整をしつつ、途中で一回大幅にずれてみて必要に応じてそっちに移るというシナリオが最も良かった」といった自由記述が複数みられた。これらから、提案手法は細かい変化と大きな変化を適切に組み合わせることで、より効率的に嗜好画像を生成できることがわかった。SimpleBOの4次元のアンケート結果も良い評価を得ている部分があるが、「画像は大抵の場合ほとんど変化せず、変化した場合も目的と無関係の変化が多かったので、画像の選択が難しかった」などの自由記述があり、ユーザの関心領域以外が変化してしまっていた可能性がある。

図4に、「提示される画像は徐々に好む画像に変化しましたか?」に関するアンケート結果を示す。これを見ると、提案手法がベースラインよりも好意的に評価されており、嗜好画像に変化した事例が多いことがわかる。図3、図4をみると、SimpleBOの4次元の結果は分散が大きいことがわかる。自由記述を見ると、「ほとんど画像の変化がなく、比較が難しかった」、「想定していた顔に非常に近い顔になつたため」といった自由記述がみられ、人によって受け取る印象に差があることがわかった。

図5に、各シナリオ後の「毎回の画像の比較は簡単でしたか?」というアンケートに対する結果を示す。これをみると、BanditBOとSimpleBOの4次元の場合の画像の比較が難しかったことがわかる。アンケートには、「同じ画像が生成され続け、選びにくい」という自由記述が多く見受けられた。ユーザから有用なフィードバックを得るために、類似した画像同士の比較は避けた方が良いことがわかった。

6 議論

6.1 生成画像の多様性とユーザの満足度の関係

本節では、ユーザ実験を通じて得られた知見について議論する。最終アンケートから、「効率が悪かったシナリオでは、互いにほぼ関係がない画像が次々と出力され、比較しても手応えがないように感じた。効率が良かったシナリオでは、少ない種類の画像を少しづつ変えていくようになり、選んだものが反映

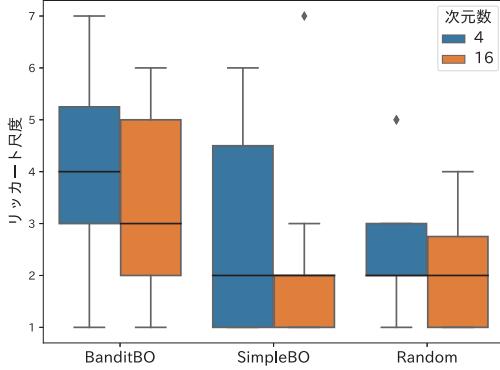


図 3: 「目的の画像まで効率的に辿り着くことができましたか？」に関するアンケート結果

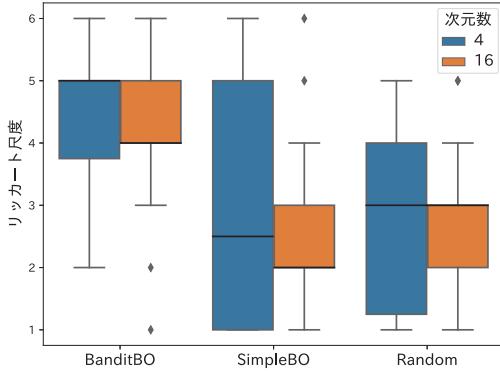


図 4: 「提示される画像は徐々に好み画像に変化しましたか？」に関するアンケート結果

されていくように感じたが、最初に選んだ画像でその後の出力がほとんど決まってしまい、局所的に最適な画像に落ちてしまつた気がする」といった自由記述が見られた。提案手法は Random と比べると生成画像の多様性に欠けてしまっている。

異なる次元数 d' における生成画像の多様性を確認するために、 d' ごとに、 w' をランダムサンプリングして複数枚の画像を生成した。図 6 は、その結果である。 $d' = 8, 16$ の場合は、多様な画像が生成できていることがわかる。しかし、 $d' = 4$ の場合は、類似した画像が生成されており、多様な画像の生成ができていない。StyleGAN の潜在空間を主成分分析によって次元削減することで、StyleGAN の画像生成の表現力が小さくなってしまうことが定性的に確認できた。また、4 次元空間の Random の自由記述で、「今回は過去に見た画像がまた出てくる事が多かった」といった自由記述が見られた。これらから、4 次元空間では 16 次元空間と比べて生成画像の多様性に欠けていることがわかる。しかし、両方の次元に関するアンケートから、「事前のアンケートで答えたイメージに近いアバターが生成できた」や「最

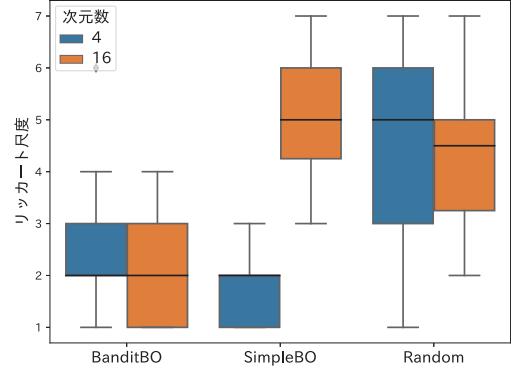


図 5: 「毎回の画像の比較は簡単でしたか？」に関するアンケート結果

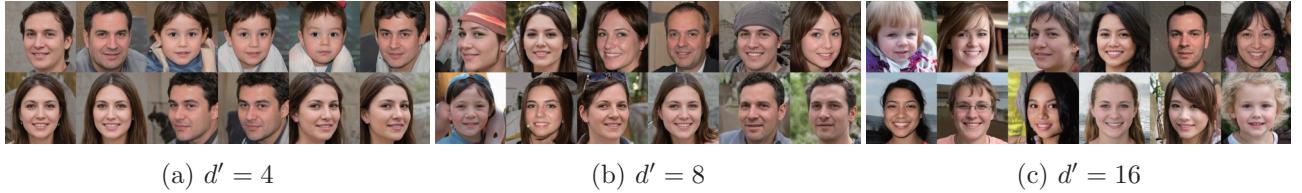
初のイメージとは異なるけど自分が納得する画像を生成できた」というコメントが見受けられた。このことから、次元の違いは生成される画像の多様性と表現力に影響を与えるが、ユーザが満足する画像を生成する能力には直接的な影響を与えない可能性がある。

6.2 画像の比較による嗜好の変化

アンケートの中に、「欲しいアバターの職業によっては性別のイメージが画像生成途中で変わることがよくあった。また、アバターを生成する中で明らかにイメージと異なる画像が生成されることも多かったが、従来と全く異なる系統の画像が表示されることで一種のインスピレーションを受けてアバターの好みが変わったこともある」といった自由記述が多く見受けられた。これにより、人間の嗜好は画像の比較を行うことで変化することがわかった。このような嗜好の変動を捉えるため、提案手法は多腕バンディットアルゴリズムを用いて各次元が現在のユーザの嗜好にどれだけ寄与しているかを動的に評価しており、ユーザの嗜好の変動に柔軟に対応し、適切な画像を生成することができる。

6.3 提案手法による多様な画像の探索支援

既存研究では、創造的なアイデアが独創性と有用性と関連しているとの議論が行われてきた [12]。ベイズ最適化の中の探索と活用は、それぞれ独創性と有用性という考え方に関連しており、これにより創造性とのつながりについて考察できる可能性があることが既存研究で提案されている [9]。BanditBO に関する自由記述で、「微調整しながら、途中で大幅に違うものと比較しては元の路線に戻ってくることを繰り返す、良い生成をしていた」といった自由記述がいくつか見られた。このことから、ベイズ最適化の探索と活用は、独創性と有用性の考え方に関連しており、提案手法はユーザの多様な画像の探索を

図 6: 部分空間の次元数 d' による表現力の違いの検証

支援できることがわかった。特に、BanditBO は有用性を追求しながらも独創性を失わないように作られている。

7 ユーザビリティ向上のための将来展望

7.1 比較画像の検討

本節では、嗜好画像生成のための人間との適切なインタラクションに関して議論する。アンケートの中で、「自分の選ばなかつた方の画像が次の比較対象となるとき、自分の好む画像から遠ざかっているような感じがして、少し苛立ちを覚えた」や「前回の比較で優位だった方を次の比較に引き継ぐ方が、より比較しやすく良い方法だと感じた」といった自由記述が複数みられた。これらから、ユーザは一対比較を行う際に前回選択した良い方と今回の生成画像の比較をシステムに求めていることがわかった。ただし、本研究ではスマートフォン操作を想定しており、画面の小ささから、Tinder のようなアプリを参考に 1 画面に 1 画像のユーザインターフェースを採用した。この場合、1 つ前の生成画像と比較することが自然だと考えている。スマートフォンで一対比較を行うための最適なユーザインターフェースを考える必要がある。

7.2 画像の変更箇所の可視化

アンケートの中で、「微調整のところでは両者がどのように違うのかがわかりづらいものも多く、もし、特微量ベースでやっているのなら、2 つがどのようない点で異なるかを文で教えてくれると助かると思った」という自由記述が見られた。ここから、新しい生成画像と前回の画像の変更箇所の可視化が求められていることがわかった。

7.3 探索と活用のバランスを表示

アンケートの中で、「画像の選択を繰り返すにあたって、大まかに絞り込んでいるのか、詳細な特徴を詰めているのかが提示されると選択中の不安感が減るように感じた」という自由記述が見られた。ユーザの不安感を軽減するために、現在のベイズ最適化の探索と活用のバランスをユーザに示してあげることが有用であることがわかった。ユーザがベイズ最適化の探索と活用のバランスを調整できるようにす

ることも今後の展望になる。

7.4 ユーザ主体のシステム

アンケートの中に、「眼鏡の部分を変えたいと思いつながら生成を繰り返していたが、眼鏡以外の部分がどんどん変化しており、なかなか思い通りの画像に近付かなかった」、「選びたくないのに 2 枚だから仕方なく選んだ画像によってだいぶ進路がブレたので、選ばないという選択肢があればいいなと思った」など、ユーザが思い通りに操作したい内容の自由記述が多く見られた。これらから、ユーザは自分が主体性を持つシステムを好むことがわかった。既存研究で、ベイズ最適化によるデザイン提案システムの中で、人間のデザイナが完全な主導権を持つ新しい概念が提案されている [9]。しかし、このコンセプトでの実証実験は行われていなく、ユーザから実際にユーザ主体のシステムが求められているかはわからなかった。これに対して、本研究のアンケートの結果から、ユーザ主体のシステム設計が実際に求められていることが明らかとなった。本研究はスマートフォン操作を想定しており、複数のスライダ調整などは操作性が低く難しいという制約がある。実際のシステム設計では、この制約を考慮してユーザ主体のシステム設計を検討する必要がある。

8 結論

本研究では、一対比較結果のみを利用する状況下で、提案手法がベースラインより効率的に嗜好画像を生成できるかを検証した。StyleGAN の潜在空間に対して主成分分析を適用して次元を削減し、ユーザの関心次元を重点的に探索することで、ベースラインより効率的に嗜好画像を生成できることを明らかにした。本研究により、スマートフォン上でスワイプ操作を用いて嗜好画像を生成する新しいアプローチを提案することができた。今後は、探索次元が大きい場合においても、少ないイテレーション数で嗜好画像を生成できるように、効率的な探索方法の開発に取り組む。具体的には、画像提示からスワイプまでの経過時間など、ユーザの行動から得られる情報を利用することで、嗜好画像生成までに必要なイテレーション数を下げるを目指す。また、7 節で述べたような、ユーザビリティの高いシステムにすることに取り組む。

謝辞

本研究は、株式会社メルカリ R4D とインクルーシブ工学連携研究機構との共同研究である価値交換工学の成果の一部である。

参考文献

- [1] P. Auer. Using Confidence Bounds for Exploitation-Exploration Trade-Offs. *J. Mach. Learn. Res.*, 3:397–422, 2003.
- [2] T. Chong, I.-C. Shen, I. Sato, and T. Igarashi. Interactive Optimization of Generative Image Modelling using Sequential Subspace Search and Content-based Guidance. *Computer Graphics Forum*, 40:279–292, 2021.
- [3] W. Chu and Z. Ghahramani. Preference Learning with Gaussian Processes. In *Proceedings of the International Conference on Machine Learning*, Vol. 119, pp. 137–144, 2005.
- [4] J. Dudley, J. Jacques, and P. O. Kristensson. Crowdsourcing Interface Feature Design with Bayesian Optimization. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2019.
- [5] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative Adversarial Nets. In *Proceedings of the Advances in Neural Information Processing Systems*, Vol. 27, 2014.
- [6] E. Härkönen, A. Hertzmann, J. Lehtinen, and S. Paris. GANSpace: Discovering Interpretable GAN Controls. In *Proceedings of the Advances in Neural Information Processing Systems*, Vol. 33, pp. 9841–9850, 2020.
- [7] T. Karras, S. Laine, and T. Aila. A Style-Based Generator Architecture for Generative Adversarial Networks. *IEEE transactions on pattern analysis and machine intelligence*, 43(12):4217–4228, 2021.
- [8] M. Khajah, B. Roads, R. Lindsey, Y.-E. Liu, and M. Mozer. Designing Engaging Games Using Bayesian Optimization. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 5571–5582, 2016.
- [9] Y. Koyama and M. Goto. BO as Assistant: Using Bayesian Optimization for Asynchronously Generating Design Suggestions. In *Proceedings of the ACM Symposium on User Interface Software and Technology*, pp. 1–14, 2022.
- [10] Y. Koyama, I. Sato, and M. Goto. Sequential Gallery for Interactive Visual Design Optimization. *ACM Trans. Graph.*, 39:1–12, 2020.
- [11] Y. Koyama, I. Sato, D. Sakamoto, and T. Igarashi. Sequential Line Search for Efficient Visual Design Optimization by Crowds. *ACM Trans. Graph.*, 36(4):1–11, 2017.
- [12] A. Kozielt, R. A. Beghetto, and M. A. Runco. *Theories of Creativity*, pp. 20–47. Cambridge University Press, 2010.
- [13] E. Raponi, H. Wang, M. Bujny, S. Boria, and C. Doerr. High Dimensional Bayesian Optimization Assisted by Principal Component Analysis. In *Proceedings of the International Conference on Parallel Problem Solving from Nature*, pp. 169–183, 2020.
- [14] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer. High-Resolution Image Synthesis With Latent Diffusion Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10684–10695, 2022.
- [15] F. Schroff, D. Kalenichenko, and J. Philbin. FaceNet: A Unified Embedding for Face Recognition and Clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 815–823, 2015.
- [16] Y. Zhou, Y. Koyama, M. Goto, and T. Igarashi. Generative Melody Composition with Human-in-the-Loop Bayesian Optimization. In *Proceedings of the Joint Conference on AI Music Creativity*, 2020.

未来ビジョン

本研究は、スマートフォンのスワイプ操作によって直感的に嗜好画像を生成する新しいフレームワークを提案し、そのようなインタラクションが可能な未来を開拓します。近年、テキストから画像を生成する「text-to-image」の技術が急速に進化し、質の高い画像生成が可能になっています。しかし、この手法では適切なプロンプトを考え出す作業が煩雑であり、ユーザにとっては一定の負荷が伴います。またその性質上、偶然的な発見が限られています。

一方で、近年マッチングアプリの急速な普及によって、スワイプ操作は日常的なインタラクションとなっています。この現状を踏まえ、本研究ではスワイプ操作を通じて気軽に効率的に嗜好に合った画像を生成できる方法を開発します。また、予想外の美しい画像に出会え

るような手法も探求します。これにより、複雑なプロンプトを考える必要がなく、手軽に高品質な画像を生成する未来が期待されます。

また、本研究の発展として産業連携を期待します。ゲーム産業では、ユーザが直感的にゲーム内のキャラクターや背景をカスタマイズできるようになれば、より深いエンゲージメントとユーザ体験が実現します。ファッション・デザイン業界では、ユーザの嗜好に合わせたアパレルやアクセサリーのデザイン案を即座に生成することができ、既存の商品開発プロセスを高速化し、多様化することができます。広告業界では、ユーザがスワイプ操作で気に入る画像やスタイルを生成できるという特性を利用して、将来的にはパーソナライズされた広告コンテンツの生成が可能になります。

動いている人の視線だけを惹きつけるモザイク投影手法

幸谷 有紗 * 夏山英高 * 浦垣 啓志郎 * † 宮藤 詩緒 * 小池 英樹 *

概要.

本研究では、運動中の観測者に対してのみ働き、静止中の観測者には影響を及ぼさない視線誘導手法を提案する。本手法は、画像中の視線を誘導したい箇所についてモザイク状に分解し、高速に重ね合わせることで実現される。これにより、画像内の顕著性を操作し、静止時と動作時の視線誘導を効果的に制御することが期待される。実験結果から、静止時と動作時では画像の観測結果が異なり、動作時において意図した箇所にモザイクが観測されることが示された。また、モザイクが観測される箇所において画像の顕著性が上がるすることが示された。本手法は歩行者への広告における視線誘導だけでなく、乗り物の広告や、実物体への視線誘導への応用が期待される。

1 はじめに

歩行者に向けた公共広告を提示する場合、広告前を通り過ぎるわずかな瞬間で歩行者の視線を誘導したい。また、歩行者が足を止めて広告を凝視した場合には、視線誘導効果が消え、内容が認識されるようにならう。そのため、歩きながら広告を見た場合は意図する箇所に視線が誘導される一方、静止すると誘導効果が消えるシステムが望まれる。

そこで本研究では、動きがある観測者に対してのみ働き、静止している観測者には働かない視線誘導方法を提案する。本手法は、画像を顕著性を操作する特殊なモザイクを用いて分解し、高速に投影することで実現される(図1)。顕著性は観測画像内の注目領域を示す指標であり、顕著性と観測者の視線には密接な関係がある[16]。本手法では、顕著性を変化させたい画像領域のみを複数枚のモザイクに分解し、ごく短い時間で連続投影する。これにより視覚系の時間積分効果が働くため[3]、静止している観測者には顕著性に変化がない通常の画像に知覚される。一方で観測者が動いている場合、一部のモザイクがズレて知覚されるため顕著性に変化が生まれ、視線誘導を促すことができる。本手法では観測者の運動状態に応じて人間の視覚特性を用いた視線誘導を行うため、特別なトラッキング装置を必要としない利点がある。

本論文では、以上の手法に関する評価を行い、その応用例と有用性について議論を行う。本論文の貢献は以下である。

- 運動中の観測者に対して顕著性を向上させる投影手法の提案
- 実験による顕著性操作の検証

Copyright is held by the author(s).

* 東京工業大学 情報理工学院

† 青山学院大学 理工学部

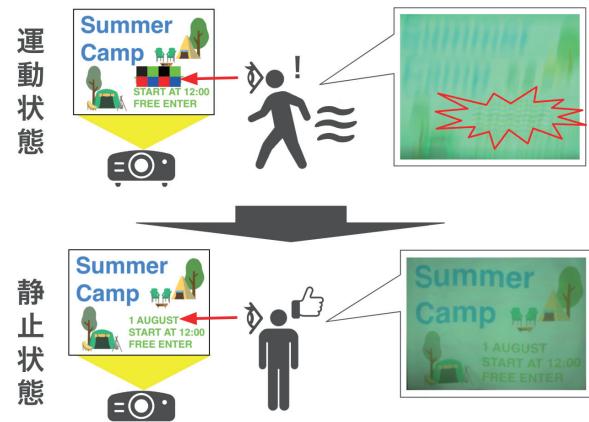


図 1. 運動状態の観測者の視線を惹きつけ、静止状態では通常の画像が認識される

- 投影手法を用いた応用事例の提示による有用性の議論

2 関連研究

2.1 顕著性の変化による視線誘導

顕著性(Saliency)とは、画像内の注目領域を特定する特徴量である。顕著性には一般にトップダウンの顕著性とボトムアップの顕著性がある。トップダウンの顕著性は、画像内の意味的な特徴量によって注意の集中する箇所を示す[5]。対してボトムアップの顕著性は、意識的な影響を考慮せず、画像の定量的な特徴量に基づいて注意を惹きつける箇所を示す手法である。ボトムアップの顕著性の算出手法についてさまざまな手法が提案されてきた[10, 8]。また、画像の顕著性を算出することにより、人間の視線を予測することができることが示されている[16, 11]。これに対し、顕著性を利用した逆のアプローチとして、顕



図 2. 顕著性を高めたい領域に対してモザイク手法を適応し、運動状態でモザイクが観測されるようにする

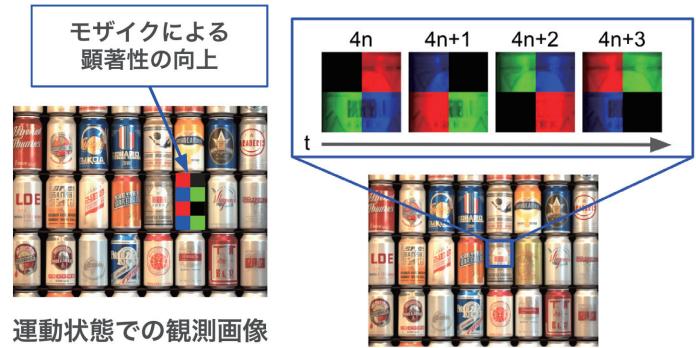


図 3. 位相の違うモザイクを順に高速表示する

著性を操作することで視線を誘導する手法が考えられている。例として、画像の色や特徴量を調整し、顕著性を操作することで指定した領域に人間の視覚的な注意を惹きつけることができる。[6, 12, 2, 17, 7]。また、画像内の特徴量を変更することで、画像中の特定の箇所から視覚的な注意を逸らすことも可能である[14]。Sutton らは AR を用いて顕著性の操作による視線誘導を行った[15]。仮想空間内において特定の場所のコントラストや彩度を変更することにより、画像内の顕著性を操作している。実験から、仮想空間内の顕著性を操作することで視線を誘導できることを実験的に示した。

これらの関連研究は、画像内の顕著性と視線誘導の関連性を示している。そこで本論文では、顕著性制御と視線誘導に関する新たなアプローチを提案する。

2.2 光の積分

人間の目は、残像効果により、一定の時間の光を重ね合わせて観測する[3]。この現象を用いて、動きに応じて観測される光を制御する研究が行われてきた[1, 18]。また、Ikeda らは観察者の動く方向に応じて異なる絵が認識される手法を提案している[9]。この研究では、観測者が動いた場合における光の時間的な重なり合わせによって絵を表現している。しかし、絵は線で表現されたバイナリ画像の重ね合わせで表現されるため、映像表現に限界がある。また、Sakaue らは観測者の移動速度に応じて異なる映像を見せる投影手法を提案している[13]。この研究では、観測者の移動速度に応じて複数枚のカラー画像を認識させることができる。ただし、観測者の移動は、投影画像に対して定められたベクトルで移動する必要がある。また、合成されて観測される画像は、表示の最適化の過程でノイズが発生し、元の画像を完全に再現することはできない。

そこで本研究では、観測者が静止状態である場合では投影するカラー画像を完全な状態で復元し、ユーザによる観測を阻害しない投影手法を提案する。

本手法は、運動状態の観測者に対する動作方向や速さへの制約が少なく、日常生活における歩行や乗り物等多様な場面での利用が見込まれる。

3 提案手法

3.1 手法の概要

本論文では、観測者の運動状態に応じて知覚される顕著性を高めることができる画像投影手法を提案する(図2)。本手法では、投影する画像中における顕著性を高めたい領域を複数枚のモザイクに分解する。モザイクは周期的に並んだ4色(赤、緑、青、黒)の正方形要素で構成され、図3のように時刻 $t = 4n, 4n+1, 4n+2, 4n+3$ ($n \in \mathbb{N}$) の経過とともに時間的・空間的に位相をずらして4フレーム分生成する。正方形要素は互いに色の変化が激しいため、指定領域の顕著性が高くなる。顕著性を向上する必要がない指定領域外では、投影画像の元の顕著性を変化させないように画素値を4等分する。最後にこれらの操作で生成した画像4フレーム分を高速に繰り返し投影する。

なお、モザイク中の正方形要素の大きさは、投影する画像中の顕著性を高めたい領域の大きさに応じてヒューリスティックに決定した。重ね合わせる正方形要素の色数が増えると、元の画像を知覚するためには高い投影周波数が必要となるため今回は4色(赤、緑、青、黒)の正方形要素を用いた。

3.2 モザイク内の色の割り当て

提案手法の適応前の画像を元画像 $I_g = (R_g, G_g, B_g)$ と呼ぶ。ここで R_g, G_g, B_g は I_g の各RGB成分である。元画像 I_g は R_g, G_g, B_g の加法混色で表現されており、 R_g, G_g, B_g の要素が重なることで I_g を再現することができる(図4)。

時間積分によって元画像 I_g を再現するために、投影するモザイク内の正方形要素には、時間方向に重ね合わせると R_g, G_g, B_g が全て重なり合うように色が割り当てられる。例えば、顕著性を高めたい指定

動いている人の視線だけを惹きつけるモザイク投影手法

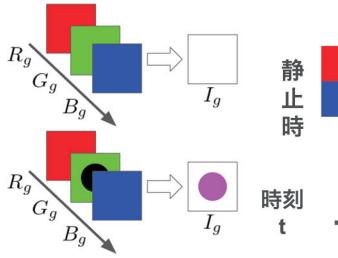


図 4. 画像を RGB の各成分に分解する例

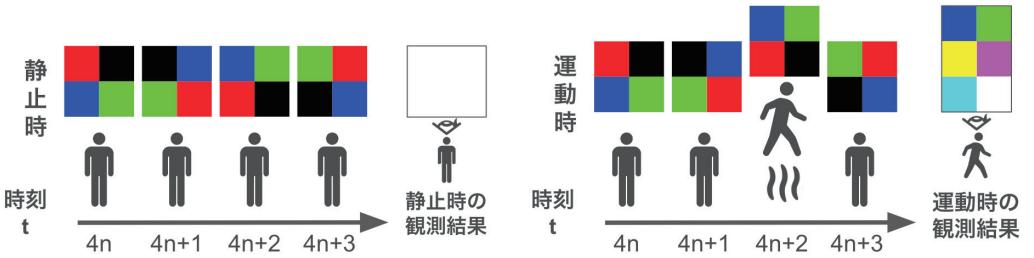


図 5. (左図) 静止時にはモザイクが重ねあわさり,(右図) 運動時にはモザイク中の正方形要素がずれて知覚される

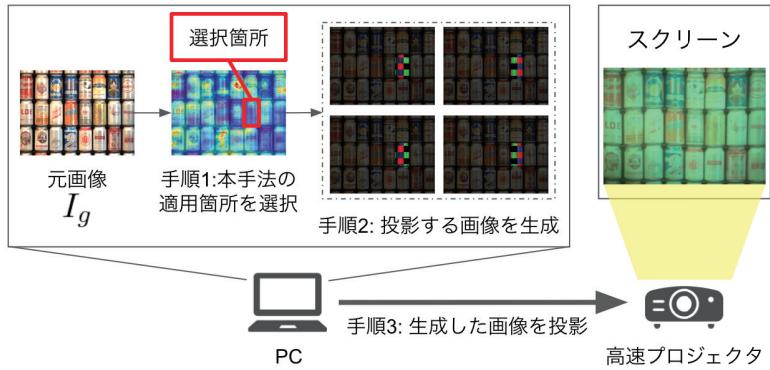
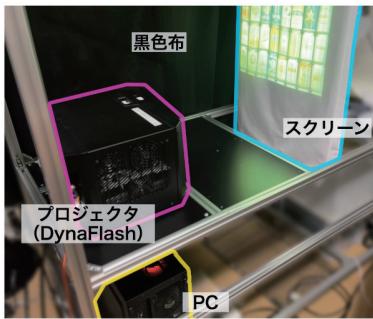


図 6. システム概要

領域において、元画像を $I_g = (R_g, G_g, B_g)$ としたとき、投影する画像 $I_p = (R_p, G_p, B_p)$ は以下のように変化する。

時刻 t	$4n$	$4n + 1$	$4n + 2$	$4n + 3$...
色	赤	緑	青	黒	...
R_p	R_g	0	0	0	...
G_p	0	G_g	0	0	...
B_p	0	0	B_g	0	...

一方で、顕著性を操作する対象の領域外において、投影する画像 I_p は以下のように変化する。

時刻 t	$4n$	$4n + 1$	$4n + 2$	$4n + 3$...
R_p	$R_g/4$	$R_g/4$	$R_g/4$	$R_g/4$...
G_p	$G_g/4$	$G_g/4$	$G_g/4$	$G_g/4$...
B_p	$B_g/4$	$B_g/4$	$B_g/4$	$B_g/4$...

結果として、静止している観測者にはモザイク指定領域の内外に依らず、投影画像 I_p が視覚系で時間方向に積分され、知覚される画像が I_g と概ね一致する。一方で、動いている観測者にはモザイク中の正方形要素がずれて知覚される(図 5)。

画像中のある領域における画素値の勾配が大きいと顕著性が大きくなりやすいため [10]、モザイクのズレにより I_p 中の指定領域内の顕著性が高まる。本

論文では、提案するモザイク生成により顕著性が高まる効果とともに、実際に観測者の運動状態に応じて顕著性が変化することを実験的に示す。

4 実装

本手法は、PCによる投影画像生成と高速プロジェクタによる投影によって実装した。図 6(左)にシステムで用いる装置を示す。また、高速投影にはDynaFlash と呼ばれる 925fps の高速カラー投影が可能なプロジェクタを使用する [19]。

本手法のシステム概要について図 6(右)に示す。まず、静止者に見せたい画像 I_g の顕著性マップを計算し、 I_g の顕著性マップを重ね合わせて表示する。この際、顕著性マップの計算には画像処理ライブラリである OpenCV を用い、Hou と Zhang ら [8] によって提案された、スペクトル残差を用いた手法を利用した。続いて、 I_g と I_g の顕著性マップを比較し、目を惹きたい箇所の顕著性を確認する。目を惹きたい箇所の顕著性が低い場合、その箇所を本手法の適用箇所として選択する(図 6 手順 1)。図 6 では、顕著性が低い缶を四角く囲んで選択している。つぎに、 I_g に対して提案手法を適用し、目を惹きたい箇所がモザイク状に分解されるように 4 枚の投影画像を生成する(図 6 手順 2)。最後に、手順 2 で生成した 4 枚の投影画像を高速プロジェクタを用いて

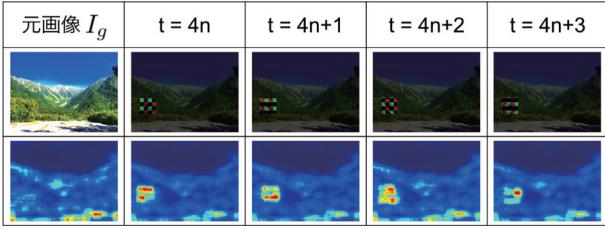


図 7. 風景画像の投影画像とその顕著性マップ

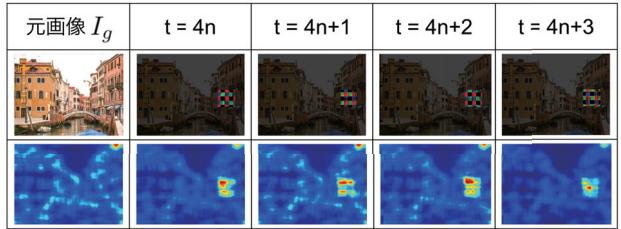


図 8. 街並み画像の投影画像とその顕著性マップ

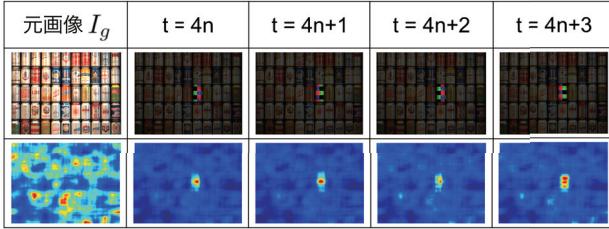


図 9. 商品画像の投影画像とその顕著性マップ

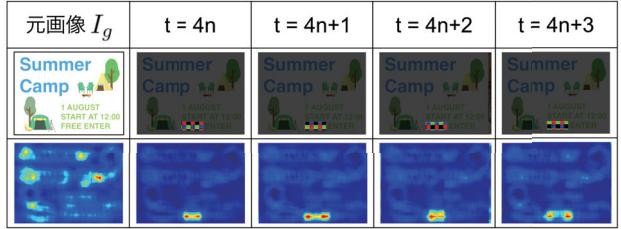


図 10. 文字を含むポスター画像の投影画像とその顕著性マップ

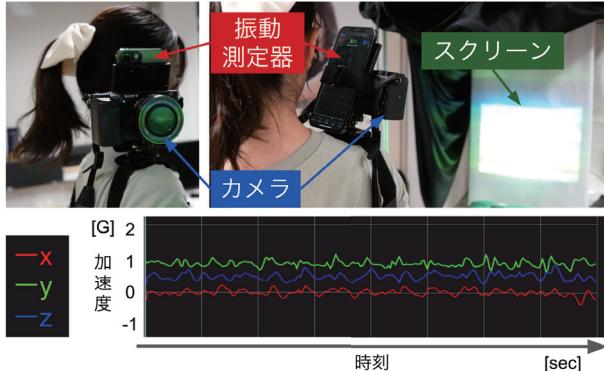


図 11. 歩行再現のための装置 (上) と実験撮影時のカメラ振動 (下)

投影する（図 6 手順 3）。フレームレートを変化させて観測したところ、フレームレートが高くなるにつれてモザイクは薄くなり、400fps ではモザイクは観測されなくなった。そのため、本システムの投影におけるフレームレートは 400fps とした。本システムの実装において、本論文では 400fps のフレームレートの実現のために高速投影が可能なプロジェクタを用いたが、リフレッシュレートの高いディスプレイを用いた実装も可能である。

5 実験

本実験は、提案手法によって顕著性を操作された画像が静止時と運動時でどのように観測され、顕著性がどのように変化するのかを検証するために行う。本実験で立てる仮説は以下の 4 つである。

- H1: 投影する画像において、モザイクの適応箇所の顕著性は向上する
- H2: 投影結果は静止時において画像 I_g が観測される
- H3: 投影結果は運動時には一部がモザイクとして観測される
- H4: 運動時にモザイクが観測される箇所は静止時に比べ顕著性が向上する

以上の仮説を検証するため、1. 投影画像の顕著性の評価、2. 投影結果の顕著性の評価、3. ユーザによる観測評価を行う。本実験では、風景、街並み、商品、文字を含むポスターの 4 種類の画像を用いた。風景の画像はエッジが少なく、顕著性がまばらである。対して街並みの画像はエッジが多く、顕著性に偏りがある。商品の画像では、類似した物体が敷き詰められており、激しい顕著性の変化がある。また、文字を含むポスターの画像では、画像内の要素に対応してはっきりとした顕著性が見られる。これらの画像を用いることにより、提案手法が多様な画像に対してどのような効果をもたらすか検証する。

5.1 投影画像の顕著性の評価

本実験では、4 種類の画像に対して提案手法の適応前の画像 I_g と、提案手法により生成された 4 枚の投影画像の顕著性を比較する。顕著性マップの計算には Hou と Zhang ら [8] によって提案された手法を利用し評価する。図 7, 8, 9, 10 に示される通り、元画像 I_g に比べ、4 枚の投影画像はモザイク付近の顕著性が向上していることが示されている。この結果から、仮説 H1 は支持される。

動いている人の視線だけを惹きつけるモザイク投影手法

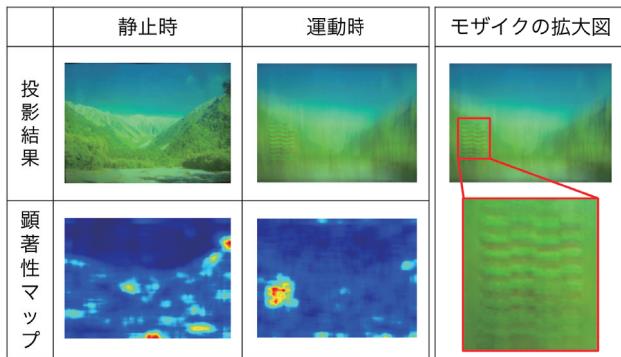


図 12. 風景画像の投影観測結果とその顕著性マップ

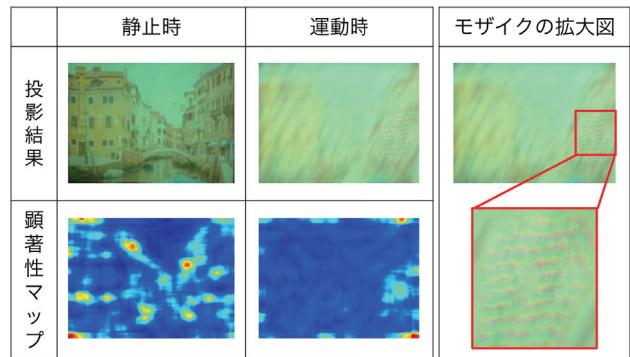


図 13. 街並み画像の投影観測結果とその顕著性マップ

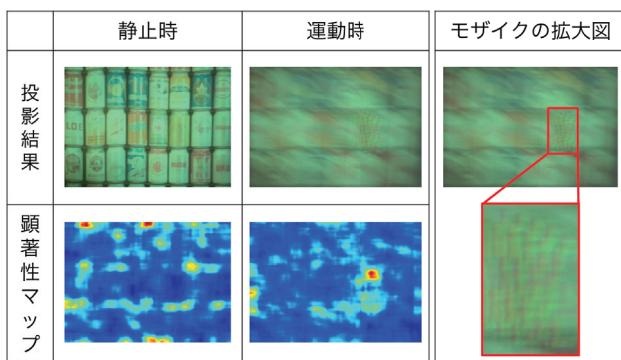


図 14. 商品画像の投影観測結果とその顕著性マップ

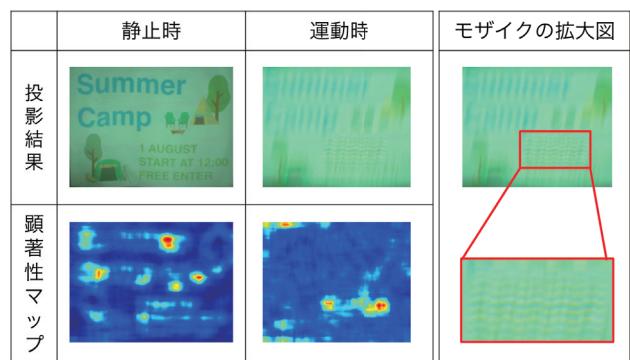


図 15. 文字を含むポスター画像の投影観測結果とその顕著性マップ

5.2 投影結果の顕著性の評価

本実験では、4種類の画像に対して提案手法を適応し、高速に投影することにより静止時と運動時の観測結果を比較する。歩行状態での観測結果をカメラで再現するため、カメラを肩に固定し、歩行して撮影を行なった(図11上)。カメラ(SONY α6400)に携帯(iPhone 13 ProMax)の振動測定器を固定し、歩行時のカメラ振動を計測した結果、図11下のような、振幅1G以内の波形となった。また、カメラのシャッタースピードは1/20secにして撮影した。このシャッタースピードは人間の目で考えると比較的遅い。しかし、カメラがデジタルで時間積分を区切る特性上、静止時の画像のチラつきを抑えられないため、限界値の1/20secを採用した。

提案手法を用いて画像を投影した際の、静止状態の時と運動状態における観測結果(上段)とそれぞれの顕著性マップ(下段)を図12、図13、図14、図15に示す。各実験結果から、静止時には画像が I_g がそのまま観測されるのに対し、運動時の観測結果では、モザイクが観測されている。この結果から、仮説H2、H3が支持される。また、下段の顕著性マップから、静止時に比べて運動時はモザイクが観測されている箇所において顕著性が向上していることが示

されている。このことから、仮説H4が支持される。

5.3 ユーザによる観測評価

研究室内の簡易的なユーザ評価として、4名の被験者に映像を観測してもらい、観測した結果について意見を得た。被験者には、モザイク投影手法の適応箇所について知らせずに、歩いている状態と静止した状態で映像を観測してもらった。その結果、4名全員から、歩いているときに変化を感じたという意見が得られた。歩いている時のそれぞれの被験者の意見として、「モザイクが観測された」「画像の一部が他の部分と違うように感じた」「画像がランダムで点滅しているように見えた」という意見も得られた。この結果から、仮説H3が支持された。

また、静止時の被験者の意見として、「画像に不快な点はなかった」「瞬きした時に目が疲れた時のようなちらつきを覚えた。しかしもう一度凝視すると綺麗な画像に見えたため、不思議に思った。」という意見が得られた。これらの意見から、肉眼で見た場合において静止時と運動時で映像の変化を感じられる一方、静止時であっても瞬きの瞬間にモザイクが観測されることが示唆された。この結果から、仮説H2は必ず支持されないことがわかった。

	(a)乗り物広告への応用	(b) トラック広告への応用	(c) 実物体への応用
運動時			
静止時			

図 16. アプリケーション例, (a) 乗り物広告への応用, (b) トラック広告への応用, (c) 実物体への応用

6 アプリケーション

本手法を用いたアプリケーションにより、日常生活の様々な場面で視線誘導を活用できると考えられる。以下に日常生活への応用として考えられるアプリケーション例を提案する。

広告に対する視線誘導

広告映像やポスターにおいて提案手法を活用すると、歩行者や電車の乗客などの運動している人に対して広告内の意図する箇所を目立たせることができるとある(図 16(a)). また、静止して広告を知覚する際に観測結果を損なわせることが無い。さらに、広告自体が移動している場合においても本手法を活用することができる。 トラック広告に対して本手法を応用することで、走行中には観測者の注意を惹きつけ、信号等で静止した場合には通常の広告が観測される応用が可能となると考えられる(図 16(b)).

白色光投影を用いた実物体への視線誘導

目標画像を白色画像とし、提案手法を適用することで、実物体に対して顕著性の操作が可能となる(図 16(c)). これにより、商品展示や美術館展示といった実物体の見た目を制御することが困難な場合においても視線誘導が可能となると考えられる。

7 議論と今後の展望

本論文では、静止状態での観測に影響することなく、運動状態の顕著性を操作する手法を提案した。実験では、静止状態と運動状態では画像の観測が異なり、運動状態でのみ対象箇所にモザイクが観測されることが示された。また、運動状態の観測結果について顕著性を調べたところ、対象箇所の顕著性を向上させることができると示唆された。これらの結果から、本手法は意図的に対象箇所の顕著性を向上させることができると示された。しかし、本論文において、実際にユーザーの視線が効果的に誘導されることについての検証はまだ行われていない。そのため、視線測定装置による視線誘導効果の検証が必要となる。

提案されたシステムにおいて、静止状態の観測者が瞬きをする際に一瞬モザイクが観測されるという問題が生じる。この現象は、瞬きの瞬間に光の時間積分が遮られることによって起こる。静止状態の観測者に対し瞬きの瞬間だけ顕著性を操作できる可能性が示唆される一方、ちらつきによる不快感を覚える可能性がある。そのため、静止者に対する映像体験についてさらなる応用や改善が期待される。

今後の展望として、元画像の顕著性が極度に偏っている場合の顕著性操作が挙げられる。特に、今回の手法はボトムアップの顕著性効果に注目したが、トップダウンの顕著性の効果が強い場合には、ボトムアップの顕著性効果が塗り替えられる場合がある[4]。そこで、本手法とは反対に、意図した箇所の顕著性を下げるような画像分解手法により、元の画像の顕著性の偏りが強い箇所に対して顕著性を低下させる操作が望まれる。例として、モザイクの要素を(赤、緑/2, 青、緑/2)とする手法が考えられる。この手法は提案手法に比べてモザイク内の輝度の変化が低く、顕著性の低下につながると予測される。

8 まとめ

広告の表示において、歩きながら広告を見た場合は視線が誘導される一方、静止すると誘導効果が消えるシステムが望まれる。そこで、本論文では観測者の運動状態に応じて知覚される顕著性を高めることができる画像投影手法を提案した。画像の視線を誘導したい箇所をモザイク状にして高速で投影することにより、顕著性を操作する。実験の結果、静止時には通常の画像が観測される一方、運動時にはモザイクが観測されることが示された。また、モザイクが観測される箇所の顕著性が向上することが確認された。しかし、現在の実験では視線誘導効果について定量的に評価されていないため、今後視線測定装置を用いた実験を行うことが望まれる。本手法は、広告に対する視線誘導や実物体の視線誘導への応用が期待される。

謝辞

本研究はJSPS科研費 20H04221の助成を受けたものです。

参考文献

- [1] H. Ando, J. Watanabe, T. Amemiya, and T. Maeda. Full-Scale Saccade-Based Display: Public / Private Image Presentation Based on Gaze-Contingent Visual Illusion. In *ACM SIGGRAPH 2007 Emerging Technologies*, SIGGRAPH '07, p. 5-es, New York, NY, USA, 2007. Association for Computing Machinery.
- [2] R. Bailey, A. McNamara, N. Sudarsanam, and C. Grimm. Subtle Gaze Direction. *ACM Trans. Graph.*, 28(4), sep 2009.
- [3] J. R. Brockmole, R. F. Wang, and D. E. Irwin. Temporal integration between visual images and visual percepts. *Journal of Experimental Psychology: Human Perception and Performance*, 28(2):315, 2002.
- [4] W. Einhäuser, U. Rutishauser, C. Koch, et al. Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. *Journal of vision*, 8(2):2–2, 2008.
- [5] S. Goferman, L. Zelnik-Manor, and A. Tal. Context-Aware Saliency Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(10):1915–1926, 2012.
- [6] A. Hagiwara, A. Sugimoto, and K. Kawamoto. Saliency-based image editing for guiding visual attention. *PETMEI'11 - Proceedings of the 1st International Workshop on Pervasive Eye Tracking and Mobile Eye-Based Interaction*, 09 2011.
- [7] H. Hata, H. Koike, and Y. Sato. Visual Guidance with Unnoticed Blur Effect. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, AVI '16, p. 28–35, New York, NY, USA, 2016. Association for Computing Machinery.
- [8] X. Hou and L. Zhang. Saliency Detection: A Spectral Residual Approach. Vol. 2007, 06 2007.
- [9] H. Ikeda, T. Hayakawa, and M. Ishikawa. Bilateral Motion Display: Strategy to Provide Multiple Visual Perception Using Afterimage Effects for Specific Motion. In *Proceedings of the 25th ACM Symposium on Virtual Reality Software and Technology*, VRST '19, New York, NY, USA, 2019. Association for Computing Machinery.
- [10] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1254–1259, 1998.
- [11] A. Kroner, M. Senden, K. Driessens, and R. Goebel. Contextual encoder-decoder network for visual saliency prediction. *Neural Networks*, 129:261–270, sep 2020.
- [12] R. Pal and D. Roy. Enhancing Saliency of an Object Using Genetic Algorithm. In *2017 14th Conference on Computer and Robot Vision (CRV)*, pp. 337–344, 2017.
- [13] F. Sakaue and J. Sato. Active 3D Motion Visualization Based on Spatiotemporal Light-Ray Integration. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1977–1985, 2020.
- [14] S. L. Su, F. Durand, and M. Agrawala. De-Emphasis of Distracting Image Regions Using Texture Power Maps. In *Proceedings of the 2nd Symposium on Applied Perception in Graphics and Visualization*, APGV '05, p. 164, New York, NY, USA, 2005. Association for Computing Machinery.
- [15] J. Sutton, T. Langlotz, A. Plopski, S. Zollmann, Y. Itoh, and H. Regenbrecht. Look over There! Investigating Saliency Modulation for Visual Guidance with Augmented Reality Glasses. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, UIST '22, New York, NY, USA, 2022. Association for Computing Machinery.
- [16] S. Treue. Visual attention: the where, what, how and why of saliency. *Current opinion in neurobiology*, 13(4):428–432, 2003.
- [17] E. E. Veas, E. Mendez, S. K. Feiner, and D. Schmalstieg. Directing Attention and Influencing Memory with Visual Saliency Modulation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, p. 1471–1480, New York, NY, USA, 2011. Association for Computing Machinery.
- [18] J. Watanabe, T. Maeda, and H. Ando. Gaze-Contingent Visual Presentation Technique with Electro-Ocular-Graph-Based Saccade Detection. *ACM Trans. Appl. Percept.*, 9(2), jun 2012.
- [19] Y. Watanabe, G. Narita, S. Tatsuno, T. Yuasa, K. Sumino, and M. Ishikawa. High-speed 8-bit Image Projector at 1,000 fps with 3 ms Delay. 12 2015.

LensTouch: スマートグラスのレンズ面を使った入力手法

川崎 竜也 * 真鍋 宏幸 *

概要. スマートグラスは、ユーザに拡張現実（AR）を体験させるだけでなく、大画面での映画や動画の観賞などを可能にさせるデバイスである。主な入力手法として、眼鏡のフレームでのタッチ入力や音声入力、ハンドジェスチャ入力などがあるが、いずれも入力の制限や入力速度の遅さなどの問題がある。この問題に対し本論文では、レンズ面でタッチ入力をを行うことでスマートグラスへの入力語彙を増やす LensTouch を提案する。抵抗膜方式のタッチパッドとスマートグラスを組み合わせて実装を行った。評価実験から、ユーザは設定に応じて素早く正確にターゲットを選択できることがわかった。次に、テキスト入力、音楽プレーヤ、ゲームの 3 つのアプリケーションの作成を行った。テキスト入力における入力速度とエラー率は 5.54WPM と 15.3% であり、残りの 2 つのアプリケーションにおいても、ユーザから肯定的な感想が得られたことから、提案手法の有用性を確認することができた。

1 はじめに

現在、Google Glass や Hololens、XREAL Air などのスマートグラスと呼ばれる光学シースルーなディスプレイを備えた眼鏡型デバイスが発売されている。スマートフォンのような携帯型のディスプレイやテレビのような固定型のディスプレイとは異なり、スマートグラスはユーザの労力を必要とせず常に利用可能な視聴覚体験の機会をユーザに提供する。スマートグラスを装着することで、大画面で映画を見たり、ゲームをプレイすることが可能である。さらに、Hololens のような周囲の物理環境の情報を取得できるスマートグラスでは、テーブルや棚などの実際の物体の上や周囲に“ホログラム”を表示させることもできる。現状のスマートグラスに使用されている標準的な入力手法は、デバイスによって異なっており、例えば、Google Glass では眼鏡のフレームでタッチ入力をを行うことができる。Hololens にはハンドトラッキングやアイトラッキングなどの機能が備わっているため、空中でのハンドジェスチャや、仮想オブジェクトへのタッチ、音声コマンドなどによる入力手法が可能である。スマートグラス自体に入力機能は備わっておらず、PC やスマートフォンなどの外部デバイスを入力デバイスとして使用する XREAL Air のようなスマートグラスもある。これらのスマートグラスに使用されている標準的な入力手法には、入力の制限や入力速度の遅さなどの課題がある。例えば Google Glass の眼鏡のフレームでのタッチ入力では、1 次元の入力しかできないため入力語彙がとても少ない。アイトラッキング機能を使用した視線入力では、ジェスチャ入力やマルチタッチのようなインタラクションに拡張することが困難

である。ハンドトラッキング機能を使用したハンドジェスチャ入力では、使用する上でユーザの周囲にある程度スペースが必要であり、使用環境に制限がある。

これらの問題を解決するため、スマートグラスのレンズ面を使った入力手法である LensTouch を提案する。なお、本研究の一部は UIST2023 においてデモ発表予定 [8] であり、本論文では、そこに含まれていない実装や実験の詳細について報告する。

2 関連研究

2.1 スマートグラスの標準機能を用いた入力機能の拡張

スマートグラスの入力機能を拡張するために、様々な研究が試みられてきた。Chun らはスマートグラスのフレームをタッチパッドとして使用することで、1 次元の一筆書きでのテキスト入力を実現した [14]。ほかにも、スマートグラスのブリッジに赤外線カメラを下向きに固定し、手と顔のジェスチャによって入力を行う研究 [13] がある。これらの研究は、ジェスチャ入力によってテキスト入力やスワイプ操作、電話をかけるなど様々な入力が可能である。しかし、ジェスチャ入力にはユーザが複数のジェスチャを覚えなければならない問題がある。[13] では 21 個ものジェスチャが用意されており、誰でもすぐに使うということは難しい。

スマートグラスでの標準的な入力手法として視線入力があり、視線を使ってテキスト入力を行う研究が存在する [2, 10]。視線入力は、ハンズフリーな入力が可能であるが、ミダスタッチ問題を回避する必要がある [3]。

Copyright is held by the author(s).

* 芝浦工業大学

2.2 外部デバイスを用いた入力機能の拡張

ジェスチャや視線入力以外にも、コントローラやスマートウォッチなどの外部デバイスを入力デバイスとして用いることで、スマートグラスやVRHMDの入力機能の拡張を行った研究が存在する。例えば、スマートグラスを装着しながら、手元のコントローラでの入力と音声入力の両方を用いてテキストを編集する研究[4]や、スマートウォッチや手首に装着したセンサでテキスト入力を行う研究などがある[1, 12]。ほかにも、VRHMDを装着しながら人差し指に取り付けた指輪の向きの情報を取得しテキスト入力を行う研究[6]などがある。このようにスマートグラスやVRHMDを装着しながら外部デバイスで入力を行うことで、現状の標準的な入力手法の問題点を解決することができる。例えばPalmType[12]では、Google Glassの眼鏡のフレームによるタッチ入力の問題点の1次元の入力しかできない点を、手首に装着したセンサを使って手のひらで入力をすることにより解決している。しかし、これらの研究のように外部デバイスを用いた入力手法は、入力をする前にセンサやデバイスを体に取り付ける必要があったり、コントローラを持つことで手が塞がってしまう。LensTouchでは、スマートグラスのレンズ面で入力を行うため、すぐに入力を始めることが可能であり、外部デバイスによって手が塞がることもない。

2.3 VRHMDへの入力デバイスの組み込み

入力手法に関する研究には、元々デバイスに備わっているセンサを使って行うものや、外部デバイスを使うもののほかに、スマートグラスやVRHMD本体に新たな入力デバイスを組み込む方法も取り組まれている。例えば、HMK[7]は、VRHMDの左右の側面に取り付けた分割型のキーボードでテキスト入力を行う研究である。ユーザは、VRHMDを装着した状態で実際にQWERTY配列のキーボードを使ってテキスト入力ができる。しかし、ユーザはテキスト入力を行うキーボードを直接見ることができないため、タッチタイピングが可能な熟練ユーザしか使用することができない。

FaceTouch[5]は、VRHMDの前面に配置されたタッチパッドを使用する。ユーザは、ディスプレイ上の視覚的な補助を受けながらタッチ入力やテキスト入力を行うことができる。このデバイスでタッチ入力を行う場合、ユーザは自身の指を見ることができない。そのため、表示されているオブジェクトを選択する際には、まずタッチパッド表面上のおよその位置に指を置き、次に選択したいオブジェクトの位置まで指を移動する必要がある。LensTouchは、FaceTouchをベースとし、VRHMDをスマートグラスに置き換えたものである。

2.4 目から近い位置でのタッチ入力手法

Blurry Touch Finger[9]は、クリップでスマートフォンにレンズを装着しただけの簡易型HMD（俗称はEasyVR）におけるタッチ入力の研究である。通常のHMDと違い、EasyVRではディスプレイの周囲を覆うカバーがないため、ユーザはディスプレイ（スマートフォン）に直接触れることができる。これをタッチ入力として用いることができるが、指が目とディスプレイの間に置かれるためユーザは対象物を視認できない、また指が目に近接しているため指がぼやけて見えてしまう、という課題がある。これらの課題に対し[9]では、ユーザは逆の目を使うことでタッチする対象のオブジェクトを見ることができること、そしてある程度の大きさのオブジェクトであれば正確な入力が可能であることを示した。LensTouchにおいても、目からの距離が近いスマートグラスのレンズ面を入力面としている。そのため指が目から近すぎることで指がぼやけて見えてしまう懸念点が存在する。しかし、この研究のように指がぼやけている状態であっても入力ができる可能性がある。

3 LensTouch

本論文においてスマートグラスのレンズ面とは、スマートグラスの外側の表面のことを指す。レンズ面の表面は、タッチ入力のための有力な候補であるにもかかわらず、光学シースルーハイブリッドディスプレイではあまり研究されていない。その主な理由として、指が目から近すぎるため正確なタッチ入力が難しい点や、タッチ入力によってレンズ面が汚れてしまう点などが挙げられる。前者については、[9]や[5]で示されているように、キャリブレーションや入力の確定方法などの実装上の工夫により正確なタッチ入力を実現できる可能性がある。また、スマートグラスの多くは映像のコントラストを向上させるために暗いレンズを採用しているため、指が触れることでレンズ面が汚れてしまう後者の課題については、あまり問題にはならないと考えられる。これらのアイディアに基づき、本論文ではスマートグラスのレンズ面をタッチして入力を行うLensTouchを提案する。FaceTouch[5]と似ているが、FaceTouchではVRHMDの前面にタッチパッドを設置しており、ユーザは入力する自身の指を直接見ることができない。一方でLensTouchは、光学シースルーハイブリッドスマートグラスのレンズ面にタッチ入力を行うため、ユーザは指がレンズ面に触れる前から自身の指を直接見ることができる。自身の指を直接見ながらターゲットや仮想オブジェクトをタッチすることができるため、FaceTouchと比較して正確で素早い入力が期待できる。また、スマートグラスに表示された仮想オブジェクトだけでなく、ユーザの周囲にある物理オブジェクト

オブジェクトを選択する際にも適用できると考えられる。さらに、LensTouchはタッチによる入力であるため、タッチジェスチャやマルチタッチなど、既存の様々なタッチインタラクションと組み合わせることもできる。左右両方のレンズ面を使うことで、入力語彙を増やすことも可能である。このように、LensTouchはスマートグラスへの効果的な入力手法として期待することができる。

3.1 実装

XREAL Air¹は市販されているスマートグラスの1つであり、PCやスマートフォンなどと接続することで、接続した画面を空中に表示することができる。そのレンズ面に、35mm × 42mmの1.8インチ透明抵抗膜方式タッチパッドを紫外線硬化樹脂で左右のレンズ面にそれぞれ1枚貼り付け、LensTouchの実装を行った(図1)。XREAL Airのレンズ面は比較的平らであり、レンズ面に指でタッチしてもタッチパッドが取り付けてあることに関する違和感はありません感じられない。ユーザは、ディスプレイを通して指の動きとディスプレイ上の映像の両方を見ることができる。実際にユーザがレンズ面にタッチした際のユーザの視点は図2のようであり、接近した指はぼやけて見える。タッチパッドの出力はArduinoでタッチ座標に変換され、シリアル通信でPCに送られる。スマートグラス自体はPCに接続され、そのPCの画面がスマートグラスから4mの距離に130インチサイズ相当で表示される。

ユーザごと、あるいは同一ユーザであってもデバイスの装着具合が変わり、目とディスプレイおよびレンズ面の相対的な位置関係に変化が生じると、スクリーン位置とタッチ位置の関係も変化する。そのため、使用前にタッチしたい座標と実際にタッチした座標の差を較正することが必要となる。これを実現するために、左上、左下、右上、右下、中央の5点をタッチすることによる簡単なキャリブレーションの実装を行った。

レンズ面での入力に関する研究があまりされていない理由のひとつに、指が触れることでレンズ面が汚れてしまう点が考えられた。しかし、実際にLensTouchを使用したユーザからは指紋の汚れにより画面が見えづらいなどの指摘は得られなかった。

4 実験

提案手法の懸念点として、指がぼやけて見える、タッチパッドが小さすぎて正確なポインティングができないことなどが挙げられる。そこで、20代の被験者5人を対象に、提案手法のポインティング性能を評価する実験を行った。

各被験者はまず、実験前に左右どちらのレンズ面



図1. 2枚のタッチパッドを取り付けたスマートグラス。



図2. タッチした際のユーザの視点。

を使用するか選択した。次にキャリブレーションを行い、その後15分程度ポインティングの練習を行った。実験では、各被験者が異なる数のターゲットを用いて3つのポインティングタスクを行う。ポインティングタスクのターゲットの数はそれぞれ9個(3x3)、21個(3x7)、45個(5x9)の3種類である(図3)。図3からわかるように、ターゲットの数を増やすと、ターゲットのサイズは小さくなる。例えば、9個のターゲットの場合、ターゲットの幅は4メートルの距離で770ミリ相当、45個の場合、260ミリ相当となる。ターゲットは大きく見えるように感じられるが、タッチパッド上の幅は、9個のターゲットの場合で10.8mm、45個のターゲットで3.6mmに相当する。上記のタッチパッド上のサイズは、タッチパッド全体をユーザの入力範囲として使用する場合の値であるが、キャリブレーションによって被験者それぞれの入力可能範囲はタッチパッドの一部に制限されるため、実際のタッチパッド上のサイズは上記よりもさらに小さくなる。タッチする対象のターゲットは赤く表示されており、ユーザがレンズ面にタッチ

¹ <https://www.xreal.com/jp/air/>

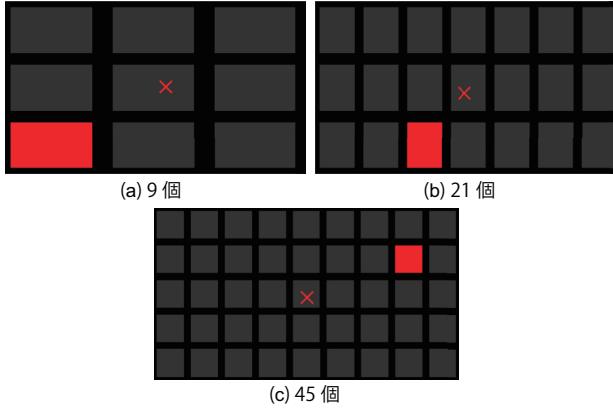


図 3. ターゲットの数.

すると、タッチ位置は図3のように×印で表示される。タッチ位置と赤く表示されているターゲットの位置が重なると、ターゲットの色は赤から青に変わる。ターゲットが表示される位置は図3のように固定であるが、次に選択するべきターゲットの順番はランダムである。被験者が指でタッチパッドにタッチした際の選択座標の確定方法として、「リリース」と「タッチ」の2つがある。両者の違いは座標を確定するタイミングである。リリース法では、タッチパッドから指を離したタイミングでの座標を取得し、それを選択座標とする。タッチ法では、指がレンズ面に触れた瞬間の座標を選択座標とする。なお、今回は確定のタイミングを意識させるためにタッチ法の実装・テストを行ったが、実際の場面ではリリース法でほとんどのケースをカバーできる。5人の被験者それぞれは9個、21個、45個のターゲットの数の順番で実験を行った。ターゲットの数が9個であれば9回タッチした時点で9個のターゲットの実験は終了となる。実験の際には、ターゲットへのタッチの成功の判定や、ターゲットが赤くなつてからタッチが終了するまでの時間を測定した。

4.1 実験結果

図4に各確定方法の平均タッチ成功率を示す。リリース法は最初のタッチでターゲットを外してしまったとしても指を動かし選択座標を修正することができるため、成功率が高くなる。実際、全ての場合でリリース法はタッチ法よりもタッチ成功率が高い。

特に45個のターゲットでは、タッチ法の成功率が約30%であるのに対し、リリース法では約90%であった。

図5に、タッチすべきターゲットが表示されてからタッチするまでの平均時間を示す。タッチ法はリリース法よりもタッチをするまでの時間が短く、素早い選択が可能である。また、リリース法はターゲットの数が増加しサイズが小さくなるにつれてタッチに時間がかかるようになっている。

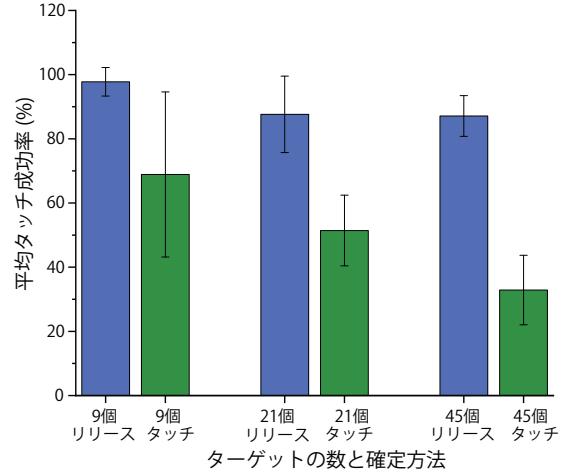


図 4. 平均タッチ成功率.

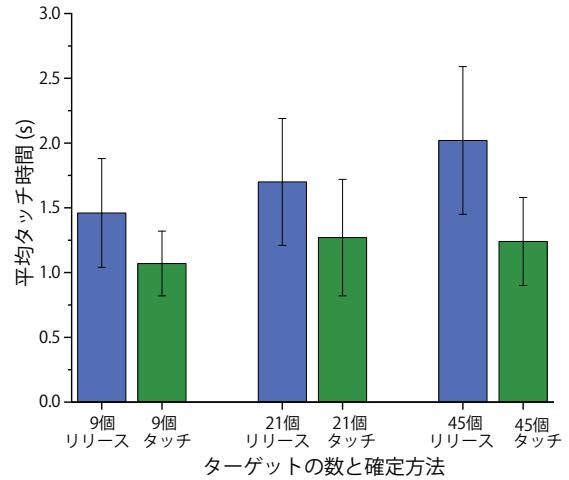


図 5. 平均タッチ時間.

5 アプリケーション

LensTouchは、タッチによるオブジェクトの選択やジェスチャなどを用いて様々なアプリケーションに適用することができる。アプリケーションによってLensTouchに要求される速度や精度が異なるため、アプリケーションに合わせて適切な設定を適用する必要がある。そこで、要求される速度や精度が異なる3つのアプリケーション（テキスト入力、音楽プレーヤー、ゲーム）の実装を行い、提案手法の有用性の検証を行った。

5.1 テキスト入力

テキスト入力のアプリケーションでは、スピードよりも正確さが重要である。ユーザは図6のように表示された仮想QWERTYキーボードで入力をを行う。

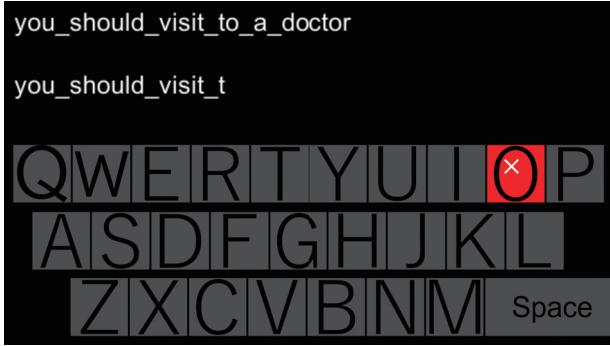


図 6. テキスト入力アプリケーション。

正確さが重要であるため入力の確定方法はリリースで行い、指をキーに重ねるとそのキーは赤くなり、指を離すとそのキーが入力される。入力した文字を削除する際には、入力に使用していないレンズ面にタッチすることで、テキストを削除する。その際にタッチする位置はどこでもよく、位置によって機能が変わることはない。今回実装したテキスト入力のアプリケーションの評価を行うため、20代の5人の被験者を対象にWPMとエラー率を測定した。測定をする際、被験者は15分程度練習を行った後に、1文につき16~43文字で構成される5文を入力した。入力する文章は、図6のように全て英文である。なおWPMは、例文の文字列を5文字（スペースを含む）で1ワードとみなして、次式で算出する[11]。

$$WPM = \frac{\text{例文の文字数}}{5 \text{ 文字}} \times \frac{60 \text{ 秒}}{\text{入力完了時間 (秒)}} \quad (1)$$

また、間違えて入力した文字を含めた入力文字数に対する、削除した入力数の割合をエラー率として次式で算出する。

$$\text{エラー率 (\%)} = \frac{\text{削除の入力数}}{\text{入力文字数}} \times 100 \quad (2)$$

測定の結果、5人の被験者の入力速度は平均で5.54WPM、エラー率は15.3%であった。

5.2 音楽プレーヤー

音楽プレーヤーはボタンの数が少ないため、素早いボタンの選択が可能である。音楽プレーヤーは図7(a)のように再生と巻き戻し、早送り、+/-の音量ボタンの5つのボタンで実装した。この音楽プレーヤーはボタンの数が少なく、1つ1つのボタンも大きいため、入力の確定方法はタッチ法で行った。ユーザは再生ボタンや音量ボタンに触れるだけで音楽プレーヤーを操作でき、スマートグラスを装着したまま音楽を楽しむことができる。実際に使用したユーザからは「じっくりと指とボタンを合わせる必要がないため

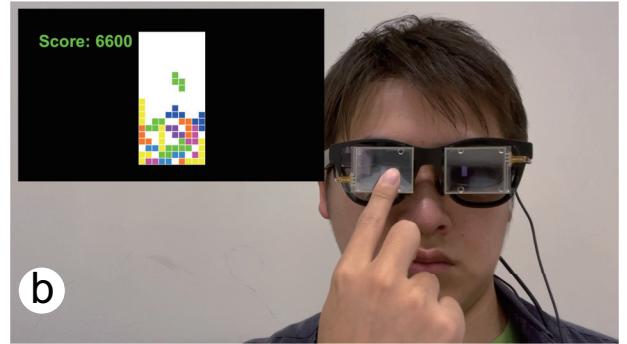
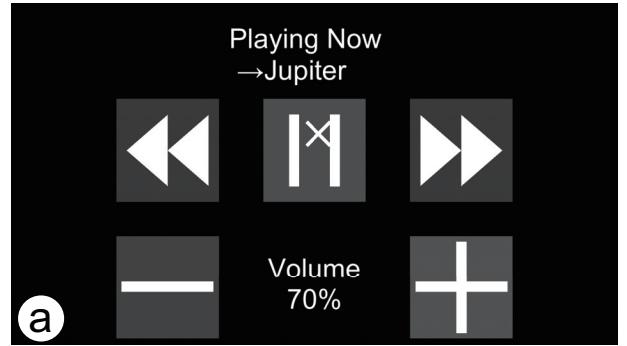


図 7. 2つのアプリケーション。(a) 音楽プレーヤー。(b) ゲーム。

操作しやすい」という感想が得られた。また、音量ボタンは連続で押すことが多く、それによりスマートグラスがずれてしまったり鼻パッドが押されることによる多少の痛みなども予想されるが、問題なく連続で押すことが可能であった。

5.3 ゲーム

左右両方のレンズを使い、素早い入力が必要な例として、「テトリス」を図7(b)のように実装した。図7(b)は実際にゲームをプレイしているユーザの図であり、左上の画像がユーザがスマートグラスを通して見ているゲームの画面である。テトリスの基本的な操作には、ブロックの左右の移動と回転、下へ落ちるスピードを上げるの4つがある。一方のレンズ面の左右をタッチすると、ブロックが左右に動く。もう一方のレンズ面の上半分をタッチするとブロックが右に90度回転し、下半分をタッチするとブロックが下に移動するスピードが上がる。実際に使用したユーザからは、「今までにない感覚と操作感で面白い」という肯定的な感想を得ることができた。一方で、「レンズ内での右と左の境界線が分からなくなる」というコメントがあった。ユーザからの視点では、レンズの(1つのレンズ内での)左右の境界は不明瞭であり、レンズの右半分をタッチしたつもりが実際にレンズの左半分をタッチしていたというミスがプレイ中に発生していた。また、「左右

のレンズで異なる操作をするため操作のミスが起こりやすい」という問題点も指摘された。

6 議論

LensTouch は、関連研究で挙げた FaceTouch[5] の VRHMD をスマートグラスに置き換えたものである。両者の違いとして、タッチ入力の際にユーザの指が直接見えるか見えないかという点がある。その影響の1つに、タッチパッドに指を置く位置の違いがある。指が見える LensTouch では、最初のタッチで選択したいオブジェクトの近くに指を置くことができる。一方、FaceTouch のように指が見えない場合では、およその位置に最初に指を置くことになる。そのため指が直接見える方が素早い入力が期待される。一方で、今回の実装で使用したタッチパッドのようにとても小さい入力面では、最初に指を置く位置は入力速度にあまり関係がない可能性もある。今後、指が見える状態と見えない状態でのタッチ入力の違いについて明らかにしていく必要がある。

静電容量方式ではなく抵抗膜方式のタッチパッドを使用しているため、現在の実装では1つのタッチパッドに対してシングルタッチしかサポートしていない。マルチタッチの実装が可能な静電容量方式のタッチパッドで実装ができれば、デザインスペースをさらに広げることができる。その具体例として、スマートグラスを装着したまま、ディスプレイに表示されている地図をピンチイン／アウトジェスチャにより拡大縮小したり、2本の指で上下にスライドすることで上下に画面をスクロールしたりできる。このように普段スマートフォンなどで行っている作業をスマートグラス上で同様に行うことが可能となる。今回スマートグラスのレンズ面のサイズに合った小さいサイズの静電容量方式のタッチパッドが見つけられなかったため、抵抗膜方式のタッチパッドを採用したが、今後はレンズ面のサイズに合った静電容量方式のタッチパッドの作成を行っていきたい。また、抵抗膜方式のタッチパッドでは、タッチが検出されるためにある程度の押下圧が必要となる。それに伴い、実験では誤入力や入力時間の遅延が見られた。静電容量方式のタッチパッドに切り替えることで、それらの課題の改善も見込まれる。

今回使用した XREAL Air は、PC やスマートフォンの画面をディスプレイに表示させることができるのが、アイトラッキングやハンドトラッキングの機能は備わっていない。一方 Hololens にはアイトラッキングやハンドトラッキングの機能が備わっているため、視線入力やハンドジェスチャ入力が可能である。この Hololens に LensTouch を実装できれば、タッチ入力と視線入力またはハンドジェスチャ入力を組み合わせたインタラクションが可能となる。しかし、Hololens はレンズ面が曲面になっているため、今回と同じ実装方法でタッチパッドを取り付け

ることはできない。そのため、Hololens に実装するためには、曲面に対応するタッチパッドの実装方法などを検討をする必要がある。

本論文の実装では、スマートグラスに表示されている PC の画面のオブジェクトを選択している。今後のインタラクションの拡張として、ユーザの選択できる範囲を表示されている画面だけではなく、ユーザ周囲の環境のオブジェクトにまで拡張することが考えられる。具体的な実装方法としてスマートグラスにカメラを取り付け、ユーザの周囲にある物理オブジェクトを認識する。そしてユーザがスマートグラスを装着した状態で物理オブジェクトをレンズ面でタッチすることで、物理オブジェクトの選択ができる。このように物理オブジェクトの選択が可能になることで、周辺環境を利用したゲームやインタラクションを実現することもできる。

7 まとめ

本論文では、スマートグラスのレンズ面を使ったタッチ入力技術である LensTouch を提案し、レンズ面に抵抗膜方式のタッチパッドを取り付けることで実装を行った。指がレンズ面から離れた座標を選択座標とするリリース法と、指がレンズ面に着地した座標を選択座標とするタッチ法の2つのタッチ確定方法を実装し、評価実験を行った。3種類のポインティングタスクを用いて評価実験を行った結果、ユーザはターゲットの数やタッチの確定方法に応じて、ターゲットを素早く、または正確に選択できた。次に、入力に対する要求条件の異なる3つのアプリケーション（テキスト入力、音楽プレーヤ、ゲーム）の実装を行った。テキスト入力アプリケーションを用いて、入力速度 5.54WPM、エラー率 15.3% でテキスト入力をを行うことができた。また、音楽プレーヤ、ゲームを実際に利用したユーザからは、肯定的な感想を得ることができた。

参考文献

- [1] S. Ahn, S. Heo, and G. Lee. Typing on a Smartwatch for Smart Glasses. In *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces*, ISS '17, p. 201–209, New York, NY, USA, 2017. Association for Computing Machinery.
- [2] S. Ahn and G. Lee. Gaze-Assisted Typing for Smart Glasses. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, UIST '19, p. 857–869, New York, NY, USA, 2019. Association for Computing Machinery.
- [3] C. Elmadjian and C. H. Morimoto. GazeBar: Exploiting the Midas Touch in Gaze Interaction. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*,

- CHI EA '21, New York, NY, USA, 2021. Association for Computing Machinery.
- [4] D. Ghosh, P. S. Foong, S. Zhao, C. Liu, N. Janaka, and V. Erusu. EYEEditor: Towards On-the-Go Heads-Up Text Editing Using Voice and Manual Input. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, p. 1–13, New York, NY, USA, 2020. Association for Computing Machinery.
- [5] J. Gugenheimer, D. Dobbelstein, C. Winkler, G. Haas, and E. Rukzio. FaceTouch: Enabling Touch Interaction in Display Fixed UIs for Mobile Virtual Reality. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, UIST '16, p. 49–60, New York, NY, USA, 2016. Association for Computing Machinery.
- [6] A. Gupta, C. Ji, H.-S. Yeo, A. Quigley, and D. Vogel. RotoSwype: Word-Gesture Typing Using a Ring. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, p. 1–12, New York, NY, USA, 2019. Association for Computing Machinery.
- [7] W. Hutama, H. Harashima, H. Ishikawa, and H. Manabe. HMK: Head-Mounted-Keyboard for Text Input in Virtual or Augmented Reality. In *Adjunct Proceedings of the 34th Annual ACM Symposium on User Interface Software and Technology*, UIST '21 Adjunct, p. 115–117, New York, NY, USA, 2021. Association for Computing Machinery.
- [8] T. Kawasaki and H. Manabe. LensTouch: Touch Input on Lens Surfaces of Smart Glasses. In *Adjunct Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, UIST '23 Adjunct, New York, NY, USA, 2023. Association for Computing Machinery.
- [9] Y. R. Kim, S. Park, and G. J. Kim. “Blurry Touch Finger”: Touch-Based Interaction for Mobile Virtual Reality with Clip-on Lenses. *Applied Sciences*, 10(21), 2020.
- [10] X. Lu, D. Yu, H.-N. Liang, and J. Goncalves. IText: Hands-Free Text Entry on an Imaginary Keyboard for Augmented Reality Systems. In *The 34th Annual ACM Symposium on User Interface Software and Technology*, UIST '21, p. 815–825, New York, NY, USA, 2021. Association for Computing Machinery.
- [11] R. W. Soukoreff and I. S. MacKenzie. Metrics for Text Entry Research: An Evaluation of MSD and KSPC, and a New Unified Error Metric. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, p. 113–120, New York, NY, USA, 2003. Association for Computing Machinery.
- [12] C.-Y. Wang, W.-C. Chu, P.-T. Chiu, M.-C. Hsiu, Y.-H. Chiang, and M. Y. Chen. PalmType: Using Palms as Keyboards for Smart Glasses. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '15, p. 153–160, New York, NY, USA, 2015. Association for Computing Machinery.
- [13] Y. Weng, C. Yu, Y. Shi, Y. Zhao, Y. Yan, and Y. Shi. FaceSight: Enabling Hand-to-Face Gesture Interaction on AR Glasses with a Downward-Facing Camera Vision. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA, 2021. Association for Computing Machinery.
- [14] C. Yu, K. Sun, M. Zhong, X. Li, P. Zhao, and Y. Shi. One-Dimensional Handwriting: Inputting Letters and Words on Smart Glasses. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, p. 71–82, New York, NY, USA, 2016. Association for Computing Machinery.

スマートウォッチの加速度センサを用いたThumb-to-finger ジェスチャ認識に基づく片手文字入力手法

角田 陸* 志築 文太郎†

概要. スマートウォッチの加速度センサを用いた thumb-to-finger ジェスチャ（親指から他の指へのジェスチャ）認識に基づく片手文字入力手法を示す。本手法では、スマートウォッチを装着した手の thumb-to-finger ジェスチャを用いて文字入力を実行する。本手法は画面のタッチを伴わないため、スクリーンオクルージョン問題、およびファットフィンガ問題の発生を回避し、スマートウォッチを装着した手のみを用いた片手入力を可能にする。また本手法は、スマートウォッチの加速度センサのデータに対する機械学習を用いてジェスチャを認識することによって、追加のセンサを用いることなく動作する。実験の結果、ユーザごとに認識性能の高いジェスチャを対象にすると、91.5%の正解率にて8個のジェスチャ認識が可能であることを確認した。加えて文字入力性能の予備調査の結果、平均3.69 WPMを示した。

1 はじめに

既存のスマートウォッチ（以降、超小型端末）における英語の文字入力手法の多くは、QWERTY キーボードを模したソフトウェアキーボードであるが、これらには以下の問題が存在する。(1) 超小型端末の画面は小さいため、文字入力の際にキーボードを表示すると、キーボード以外の表示領域が制限される（スクリーンオクルージョン問題 [11]）。(2) 画面に表示されるキーは小さい。そのため、誤入力が発生しやすく、加えて指に覆われたキーを視認することが困難である（ファットフィンガ問題 [25]）。

文字選択方法を工夫することによって、これらの問題を解決した手法 [22, 7, 12, 23, 19, 14, 21] がある。しかしこれらの手法では、超小型端末を装着していない手を用いて画面をタッチする必要があるため、両手の使用を要する。ゆえに、物の運搬、料理、傘の保持などの片手が塞がっているときの入力は不便になる [28]。超小型端末を装着した手のジェスチャを用いる片手文字入力手法 [29, 9, 30] はこの問題を解決するが、これらの手法では超小型端末以外の追加のセンサが必要である。また、音声入力は多くの超小型端末に搭載されている手法であるが、図書館などの声を出すことが難しい場所において使用できず、かつプライバシが漏洩し得る。

Thumb-to-finger ジェスチャ（親指から他の指へのジェスチャ）は、操作中に明確な触覚的手がかりを与えることができる入力方式 [13]、および指が本来持っている運動能力を活用する入力方式 [26] として注目を集めている。また、超小型端末に搭載された加速度センサを用いて、超小型端末を装着した手

のジェスチャを認識する研究 [17, 32, 24, 18] が行われている。しかし、加速度センサを用いた thumb-to-finger ジェスチャの認識およびその認識性能調査は行われていない。そこで、我々は加速度センサを用いて thumb-to-finger ジェスチャを認識し、加えてその認識性能を調査した。その上、超小型端末に搭載された加速度センサを用いたリアルタイムに動作する thumb-to-finger ジェスチャ認識に基づく英語の片手文字入力手法（図1）を開発した。本手法は、超小型端末を装着した手のジェスチャを用いた入力を採用することによって、スクリーンオクルージョン問題およびファットフィンガ問題を回避し、片手入力を可能にする。また、センサとして超小型端末に搭載された加速度センサのみを用いるため、本手法は市販の超小型端末において追加のセンサを用いることなく動作する。

本稿ではまず、超小型端末の加速度センサを用いて thumb-to-finger ジェスチャを認識する手法を述べた上で、ジェスチャの認識性能および主観評価の調査結果を述べる。その後、調査結果に基づいたキーボードの設計、および本手法を用いた文字入力性能の予備調査の結果を示す。

本研究の貢献を以下に示す。

- 超小型端末を装着した手の8個の thumb-to-finger ジェスチャを加速度センサを用いて91.5%の正解率にて認識できることを明らかにした。
- 超小型端末における追加のセンサを要しない thumb-to-finger ジェスチャを用いた片手文字入力手法を示した。

2 関連研究

本手法は、超小型端末の加速度センサを用いてジェスチャを認識し、それに基づいて動作する片手

Copyright is held by the author(s).

* 筑波大学 情報理工学位プログラム

† 筑波大学 システム情報系

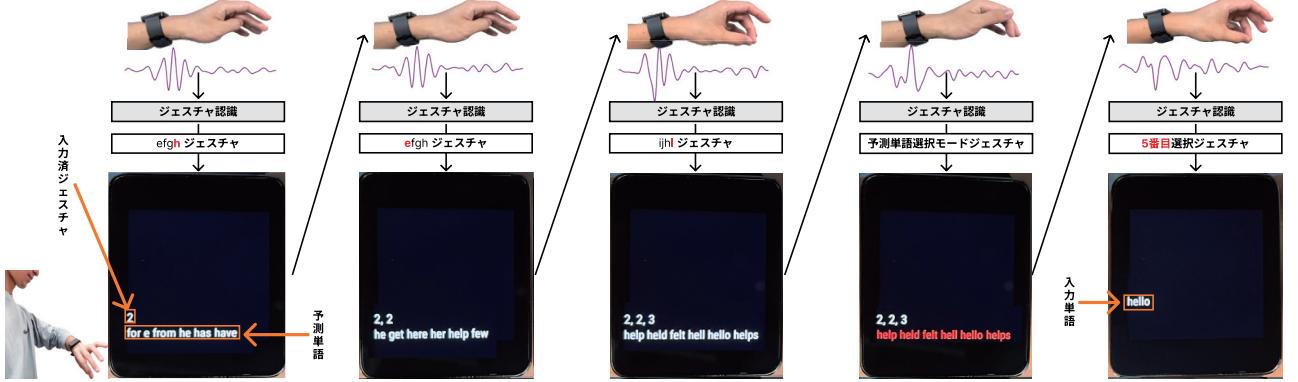


図 1. 本手法の概要. ユーザは超小型端末を装着した手の thumb-to-finger ジェスチャを用いて文字入力を行う. ジェスチャを用いてアルファベット群を繰り返し選択し, それに基づいて予測された単語をジェスチャを用いて選択する. この例では hello を入力している.

文字入力手法である. 本節では関連研究として, 超小型端末の加速度センサを用いてジェスチャを認識する手法, および超小型端末における文字入力手法を述べる.

2.1 超小型端末の加速度センサを用いたジェスチャ認識手法

超小型端末の加速度センサを用いてジェスチャを認識する研究が行われている. Serendipity [24] は, 市販の超小型端末に搭載された慣性計測ユニット(IMU)を用いて, 指をつまむ, 叩く, こするなどの5つの指のジェスチャを認識した. ViBand [18] は, 市販の超小型端末に搭載されている加速度センサのサンプリングレートを 4000 Hz に引き上げることによって, フリック, クラップ, スクラッチ, タップなどのハンドジェスチャ認識を可能にした. Taprint [5] は, 市販の超小型端末の IMU のサンプリングレートを引き上げることによって, 超小型端末を装着した手の甲における被タップ位置を認識した.

また Xu ら [32] は, 既存のジェスチャセットの性能を低下させることなくユーザからの最小限のデータ収集を用いたジェスチャカスタマイズを可能にした. Kimura [17] は, 自己教師あり学習を用いて, 少量のラベル付きデータを用いたハンドジェスチャ認識を行った.

我々は, 超小型端末の加速度センサを用いて thumb-to-finger という微細なジェスチャを認識し, 加えてその認識性能を調査する. その上, それに基づいた片手文字入力を実現する.

2.2 超小型端末における文字入力手法

超小型端末における文字入力手法は, ソフトウェアキーボードを活用する手法, ハンドジェスチャを用いる手法に大別される. 音声入力は多くの超小型端末に搭載されている入力手法であるが, ここでは, 超小型端末固有の文字入力手法を述べる.

ソフトウェアキーボードを活用する手法 [22, 7, 12, 23, 19, 14] の多くは, キーの2段階の選択を用いることによって, キーのサイズを大きくし, ファットフィンガ問題を解決する. Invisiboard [21] は, 透明なキーボードを用いることによってスクリーンオーバーレイ問題を解決する. しかし, ソフトウェアキーボードを用いる手法では, 超小型端末を装着していない手を用いて画面をタッチする必要があるため, 片手入力は不可能である.

ソフトウェアキーボードを活用した文字入力手法の多くでは, 画面専有面積と入力精度がトレードオフの関係にある. そのため, 画面のタッチを伴わないハンドジェスチャを用いる文字入力手法が研究されている [29, 33, 10, 34, 16, 9, 30]. ただし, これらの手法は両手を用いる手法, もしくは超小型端末に追加のセンサを必要とする手法である.

また, 超小型端末を装着した手を用いて描いた文字を, 追加のセンサを用いることなく認識する片手文字入力手法 [6, 31, 8, 20] が存在する. 我々の手法も追加のセンサを用いないジェスチャ認識に基づく片手文字入力手法であるが, 超小型端末を装着した手の thumb-to-finger ジェスチャを用いて文字入力を行う.

3 Thumb-to-finger ジェスチャの認識

我々は加速度センサのデータに対する機械学習を用いて, thumb-to-finger ジェスチャを認識する. 今回thumb-to-finger ジェスチャとして, タップ [26] を採用した. 我々は親指以外の指を分割して作った24箇所, およびタッチ面として使われることが多い掌 [4] の合計 25 箇所(図 2)をタップ対象とした.

3.1 実装

微細なジェスチャの認識のために, 超小型端末の Linux カーネルを変更することによって, 4000 Hz



図 2. 我々が採用した thumb-to-finger ジェスチャの 25 種類のタップ対象。タップ対象は、指前面、指側面、および掌の 3 種類から成る。指前面および指側面は、指の関節を境界として 3 つの領域に分割される。

のサンプリングレートにて加速度データを収集する [18]。実装には、IMU (InvenSense MPU6515) を搭載した超小型端末である LG G W100 を用いた。この IMU は他の多くの超小型端末に搭載されているものと同一のシリーズである。

3.1.1 セグメンテーション

文字入力のためには加速度データの分類をリアルタイムに行う必要があるため、ユーザが行ったジェスチャの発生の検知、および加速度データからジェスチャが行われた区間を切り出すこと（セグメンテーション）もリアルタイムに行う必要がある。今回我々は [27] にて用いられた変化率スコア R_x を検知に用いることにした。特に、微細な動きを捉えるために、40 点 (0.01 秒) 前との変化率を採用した。また、今回採用したジェスチャの所要時間は約 0.6 秒であるため、ウィンドウの大きさを 2400 (4000×0.6) とした。ジェスチャの開始点に対応する R_x の閾値 τ_R^1 を定め、最新の R_x のウィンドウにおいて、 τ_R^1 を初めて超える点が 200 点目から 600 点目に存在するとき、そのウィンドウをジェスチャが行われた区間とする。 τ_R^1 を初めて超える点はジェスチャの開始点より僅か後であることが多いため、これによってジェスチャの開始点からのデータをウィンドウに含めることができる。さらに、最新の R_x のウィンドウの最大値が閾値 τ_R^2 以上であるときのみ検知を行うことによって、誤検知を避ける。また、同一のジェスチャから生じる波形を複数回検知することを防ぐため、ジェスチャの検知後 0.6 秒間は検知処理を行わない。

3.1.2 機械学習

Thumb-to-finger ジェスチャは微細なジェスチャであるため同一の種類のジェスチャであっても、ユーザごとにその動きは異なる。したがって、ユーザご

との分類モデルを作成する。ただしこのためには、ユーザごとのデータ収集が必要となる。

我々は、ユーザごとのデータ収集を最小限にするため、Kimura の手法 [17] を採用することにした。この手法は自己教師あり学習を用いる手法であり、2 つのステップからなる。

最初のステップは、事前学習であり、多量のラベルなしデータに対する教師なし表現学習を行う。今回著者の 1 人が、超小型端末を装着した状態にて日常生活を約 2 時間送り、最新の R_x のウィンドウの最大値に対して閾値 τ_R^3 を用いて動きがあった際のデータのみを収集した。その結果、2400 (0.6 秒) の大きさであるウィンドウのデータが 5106 個得られた。このデータを Kimura の手法 [17] と同じく、3 つの周波数帯に分割し、時系列データに対する表現学習モデルの訓練を行った。

2 つ目のステップはファインチューニングであり、少量のラベル付きデータに対する教師あり学習をユーザごとに行う。まず、ユーザがジェスチャを行った際のデータを前節の方法にてセグメントする。次に、セグメントされたデータを事前学習時と同じく 3 つの周波数帯に分割する。その後、事前学習された表現学習モデルを用いて、そのデータを 320 次元の特徴表現に変換する。そして、この変換後のデータを用いてジェスチャ分類モデルの訓練を行った。機械学習アルゴリズムには LinearSVC を用いた。

3.2 ジェスチャの認識性能および主観評価の調査

本システムを用いたジェスチャの認識性能を調査するための実験を行った。また、文字入力に用いるジェスチャを決定するために、ジェスチャに対する主観評価を調査した。先行研究 [15] は、thumb-to-finger のタッチジェスチャを対象にジェスチャの主観評価を調査しているが、我々は掌を下に向かた掌が見えない状態においてジェスチャを行うため、この調査を行った。しかし、掌の向きおよび掌の見え方が主観評価に与える影響の違いについての調査はしていない。実験には、研究室の 6 名の大学生および大学院生 (21 歳–24 歳、平均 22.1 歳、全員男性、右利き) がボランティアとして参加した。

3.2.1 実験設計

立位の状態においてファインチューニングのためのデータ収集を行った。参加者は左手に超小型端末を装着した。各ジェスチャを 10 回ずつ行うことを 1 セッションとする。データ収集時間の短縮および参加者が誤ったジェスチャを行うことを防ぐために、それぞれのジェスチャは連続して行われた。ジェスチャの順番は全セッションにおいて同一である。また、指側面のジェスチャの前に最低 30 秒の休憩が設けられた。ジェスチャセットには、日常生活においてあまり行わない動きが含まれているため練習を行

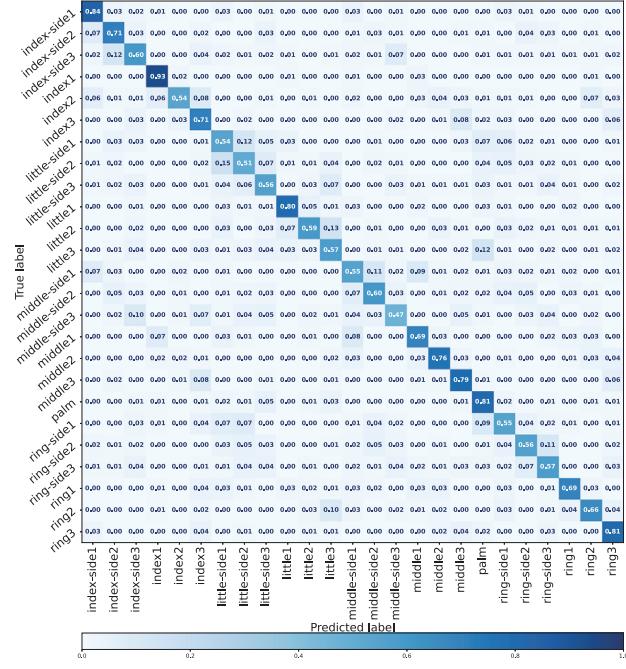


図 3. 全ジェスチャを対象にした混同行列。

2セッションを行い、その後本番を2セッション行った。セッションの間では最低2分の休憩、および超小型端末の再装着を行った。参加者は超小型端末の画面を見ながらジェスチャを行った。画面には行うべきジェスチャが表示された。また、参加者は同一の種類のジェスチャでは同様の指の動きをすること、および体と腕の姿勢を変化させないことを指示された。得られたデータは参加者1人あたり、25種類×10回×2セッション=500個である。

その後、参加者は5段階のリッカート尺度を用いて、ジェスチャの行いやすさおよびジェスチャの疲れにくさに関するアンケートに回答した。実験全体において参加者1人あたりが要した時間は50分程度であった。

3.2.2 結果および考察

参加者ごとの認識性能を調べるために、参加者ごとに、本番セッション1と本番セッション2のデータに対して、セッション間交差検証を行った。そのため、1度の検証における1つのジェスチャの訓練データ数は10、テストデータ数は10である。すべてのデータは事前学習された表現学習モデルを用いて特徴表現に変換された。結果として、全ジェスチャを対象にした正解率の平均は、65.7%（標準偏差：7.00%）であった。混同行列を図3に示す。

次に、ジェスチャごとの認識性能として、再現率が高い（他のジェスチャに分類されにくい）ジェスチャ、および適合率が高い（誤分類されにくい）ジェスチャを調査した。評価指標には、再現率と適合率

の調和平均であるF1スコアを用いた。ジェスチャごとのF1スコアの標準偏差の平均は15.4%であった。これは、特定のジェスチャに対する参加者間のF1スコアのばらつきが大きいことを示す。そのため、F1スコアが高いジェスチャセットは参加者ごとに異なっていた。

そこで、参加者ごとに、F1スコアが高いジェスチャのみを対象にしたセッション間交差検証を行った。その結果、F1スコアが高い上位8個のジェスチャのみを対象にすると、正解率の平均が9割を超えることがわかった（正解率平均：91.5%，正解率の標準偏差：2.54%）。

次にアンケートの結果を分析する。この分析ではリッカート尺度を便宜的に間隔尺度として用いる。ジェスチャの行いやすさに関して、「とても行いやすい」を5、「とても行いにくい」を1として平均値を求めた。ジェスチャの疲れにくさに関して、「全く疲れない」を5、「とても疲れる」を1として平均値を求めた。その結果、行いやすさの平均値が3以上かつ疲れにくさの平均値が3以上であったジェスチャは、index1, index2, index-side1, index-side2, index-side3, middle1, middle2, middle3, middle-side1, middle-side2, middle-side3, ring1, ring2, ring3, ring-side1, ring-side2, ring-side3, palm（以降、ジェスチャ群A）であった。これらのジェスチャを対象にした際の、参加者ごとのセッション間交差検証の正解率の平均は、71.6%（標準偏差：6.28%）であった。

一方、行いやすさおよび疲れにくさの両者に関して、回答が割れるジェスチャも存在した。これは、参加者によって指の可動域が異なることが原因の1つであると考えられる。

4 キーボードの設計

Thumb-to-fingerジェスチャを用いた英語の文字入力のためのキーボードの設計を述べる。まず、アルファベットの数に対してジェスチャの数が少ないため、我々は1つのジェスチャに対して複数のアルファベットを割り当てる（これらのジェスチャを入力ジェスチャと呼ぶ），これに単語予測を組み合わせる。単語予測には、予測単語列から単語を選択するための機能が必要である。今回、我々は予測単語選択モードを設け、このモードへの切り替えに専用のジェスチャ（選択ジェスチャ）を採用する。このモードでは、ユーザは入力ジェスチャを用いて単語を選択する。また、ジェスチャが誤認識されること、およびユーザが誤ったジェスチャを行うことが考えられる。そのため、我々は削除ジェスチャを採用する。

3.2節の結果よりF1スコアが高いジェスチャセットは参加者ごとに異なっていた。そのため、ユーザごとにジェスチャセットを作成すると、F1スコアが高いジェスチャセットとなる。しかし、このためには



図 4. 我々が設計したキーボード。文字はアルファベット順にて配置した。数字は、選択する予測単語の順番に対応している。

ユーザごとに全ジェスチャの認識性能を調査する必要およびユーザごとにキーボードをデザインする必要がある。したがって今回は、本手法の可能性を調査することを目的として、ユーザごとのジェスチャセットの作成は行わず、平均的に認識性能が高かったジェスチャを用いる。さらに、以下の方針をもとに文字入力に用いるジェスチャを決定する。(1) ジェスチャの主観評価が悪いジェスチャを除く(ジェスチャ群Aから採用する)。(2) 25クラス分類時の平均のF1スコアが特に低いジェスチャを除く。(3) 互いの誤分類率が高いグループからは1つのジェスチャのみを採用する。(4) 採用したジェスチャのみを用いた際の正解率が大きい。(5) 入力ジェスチャを最低5つ、選択ジェスチャ・削除ジェスチャ(ファンクションジェスチャ)のためのジェスチャを2つ採用する。(6) ユーザが文字入力に用いるジェスチャを覚えやすいように、指の面ごとに異なる組み合わせのジェスチャを採用しない(例として、中指の指前面ではmiddle1, middle2を採用し、薬指の指前面ではring2, ring3を採用することをしない)。(7) ファンクションジェスチャおよび入力ジェスチャは、ジェスチャの覚えやすさのためにジェスチャの種類(指前面のジェスチャまたは指側面のジェスチャ)または指が異なる。

この結果、我々は、ファンクションジェスチャとして、index-side1およびindex-side2を、入力ジェスチャとして、index1, middle1, middle3, ring1, ring3、およびpalmを採用した。これらのジェスチャを対象にした際の、参加者ごとのセッション間交差検証の正解率の平均は、87.6% (標準偏差: 5.35%)であった。以上のジェスチャを用いたキーボードを図4に示す。

このキーボードを用いた文字入力システム(図1)を説明する。ユーザは、目的の単語の各アルファベットに対応するジェスチャを行なう。システムは、入力されたジェスチャ列に基づき、予測単語を6つ提示する。次にユーザは選択ジェスチャを用い

て、モードを予測単語選択モードに切り替える。そして、予測単語の番号に対応するジェスチャを行なうことによって、単語を選択する。入力したい単語が予測単語列に存在しないときは、繰り返し選択ジェスチャを行なうことによって次の予測単語列の表示に切り替えることができる。以上の操作を繰り返すことによって文章を入力する。

単語予測では、入力されたジェスチャ列に対応するアルファベットの全ての組み合わせに対して、英単語の使用頻度およびauto-complete [9]を用いて予測単語を算出する。この手法は、使用頻度が比較的低いが短い単語の存在を考慮して、入力文字数以上の単語を予測単語として提示する。

また、文字入力中にジェスチャの誤認識および誤った操作が発生し得るため、認識されたジェスチャをユーザが確認できるように、システムは認識されたジェスチャを超小型端末に表示する。文字入力モードと予測単語選択モードの状態は、予測単語の色を変更することによって示す。なお、単語と単語の間にはスペースが自動的に挿入される。

5 文字入力性能の予備調査

本手法を用いた際の文字入力の速度(WPM: Words Per Minute [2])、およびエラー率(WER: Words Error Rate [3])を評価するための実験を行った。実験には、研究室内の3名の大学生および大学院生(21歳–24歳、平均22.3歳、全員男性、右利き)がボランティアとして参加した。

5.1 実験システムの実装

超小型端末とBluetooth接続されたスマートフォンをPCに接続する。PCにおいて超小型端末の加速度センサのデータを取得し、リアルタイムに信号処理、ジェスチャ認識、および単語予測を行う。また、課題文、入力中の文、入力中の単語のために行なったジェスチャ、および予測単語列を、PCからスマートフォンを経由して超小型端末に送信する。単語の使用頻度はコーパス[1]から得た。実験システムにおいて、ジェスチャを行なってからその認識結果が超小型端末に表示されるまでの所要時間は、主にデータ通信、ジェスチャ認識、および単語予測により約0.9秒であった。

5.2 実験設計

参加者は図4のキーボードを用いて、立位の状態において超小型端末に表示される課題文を入力する。初めに参加者は、ジェスチャ認識のファインチューニングのためのデータ収集のために、各ジェスチャを行なった。合計のジェスチャ回数が少なくデータ収集にあまり時間がかかるため、ジェスチャの種類が少なく誤ったジェスチャが発生しづらいため、および文字入力では連続して同一のジェスチャを行

うとは限らないため参加者はランダム順にてジェスチャを行った。また、アルファベットのヒントのラベルが、対応するジェスチャの領域の裏側に貼られた。参加者は、対応する文字がどの領域に割り当てられているか分からぬときにヒントを参照することができた [29]。参加者は実験の前に、本システムを用いた文字入力の練習を望むだけ行った。その後、最低3分間の休憩を行い、本番として5つの短文を入力すること（1セッション）を2回行った。セッション間には最低3分の休憩を取った。短文の文章および出題順は、全参加者において同一であり、同一の文章は練習および本番を通して繰り返し用いられない。短文は、フレーズセット [35] からランダムに選ばれた。また、参加者は高速かつ正確に入力することを求められた。

5.3 結果および考察

WPM の平均は 3.69（セッション 1: 3.60, セッション 2: 3.77), WER の平均は 0.0547 (セッション 1: 0.0694, セッション 2: 0.0400) であった。最も練習が長かった（約 40 分）参加者の WPM は、セッション 1 において 5.55. セッション 2 において 5.33 であり、WER は両セッションにおいて 0 であった。また、手法に習熟した著者が同一の条件において実験を行ったところ、セッション 1 において 7.68 WPM, セッション 2 において 8.38 WPM を達成した（両条件において WER は 0）。これは、手法に習熟することによってより高速に文字入力が可能になることを示唆する。キーの割り当ての確認に時間がかかったとコメントした参加者がいたことも、手法の習熟が高速な文字入力に繋がることを示唆する。

また、入力中にジェスチャの誤認識が多かったとコメントした参加者がいた。これは、入力中の体および手の姿勢の変化に伴って、データ収集時と異なる姿勢およびジェスチャになっていたためと考えられる。ジェスチャの誤認識に伴う直前の入力の取り消しおよび 2 度目の入力は、文字入力速度の低下につながるため、手の角度を変化させた際のデータ収集、または姿勢の変化に堅牢なジェスチャ認識が求められる。

6 議論および今後の課題

今回、微細なジェスチャの認識のために、4000 Hz のサンプリングレートにて加速度センサのデータを取得した。しかし、100 Hz のサンプリングレートの加速度データを用いたハンドジェスチャ認識の研究 [17, 32] がある。4000 Hz のサンプリングレートは消費電力の増大につながるため、サンプリングレートを小さくしていった際の認識性能を調査する必要がある。また、今回の実装ではジェスチャを行ってからその認識結果が超小型端末に表示されるまでに約 0.9 秒要した。したがって、入力は最も早く 0.9

秒に 1 回であった。この原因の 1 つは、ジェスチャの波形が完全に含まれる 0.6 秒のウィンドウを採用したことである。より小さいウィンドウを用いてジェスチャの途中までの波形にて認識を行うことができれば、ジェスチャを行ってからその認識結果を表示するまでに所要する時間が短縮されるため、文字入力速度の向上につながる。よって、今後はウィンドウに含めるジェスチャの波形の程度と認識性能の関係を調査する必要がある。加えて、高い性能にて認識可能なジェスチャの種類が増えれば、入力の語彙が増える。例として、人差し指、中指、および薬指の前面の 9 つの領域に対するタップの認識が可能であれば、T9 入力が可能になる。このように、入力の語彙が増えることによって、柔軟なキーボードの設計が可能になるため、高性能にて多種類のジェスチャ認識が可能な手法の開発が求められる。また、本手法では文字入力のための計算を PC において行った。そのため、屋外等の場所において用いるためには、超小型端末に接続されたスマートフォンにて計算をする必要、またはクラウドコンピューティングを活用する必要がある。

本論文では、thumb-to-finger ジェスチャとして指前面、指側面、および掌に対するタップジェスチャを採用した。しかし、指背面に対するタップ、素早く 2 回連続して行うタップ、およびスワイプ、サークル [4], フリック（指の弾き）などのタップ以外のジェスチャを採用することも考えられる。例として、フリックする指の認識に基づいた予測単語の選択ができれば、予測単語選択モードへの切り替えの必要がなくなる。これらのジェスチャは、今回採用したジェスチャと動きが大きく異なるため、今回採用したジェスチャの認識性能を低下させることなく認識できる可能性がある。加速度センサを用いた thumb-to-finger ジェスチャ認識に基づく片手文字入力のための最適なジェスチャセット、および主観評価が高いかつ性能が高いかつ学習コストが低いキーボード設計を調査する必要がある。

7 おわりに

本稿では、超小型端末の加速度センサを用いた thumb-to-finger ジェスチャ認識に基づく片手文字入力手法を示した。本手法では、スクリーンオクルージョン問題およびファットフィンガ問題が発生せず、市販の超小型端末において追加のセンサを用いることのない片手文字入力が可能である。ジェスチャ認識性能調査では、ユーザごとに F1 スコアの高いジェスチャを抽出することによって、平均 91.5% の正解率にて 8 個のジェスチャ認識が可能であることを示した。また、ジェスチャ認識性能および主観評価の調査を通じてキーボードの設計を行い、本手法を用いた文字入力性能の予備調査を行った。その結果、平均 3.69 WPM を示した。

参考文献

- [1] English Word Frequency. <https://www.kaggle.com/datasets/rtatman/english-word-frequency>. 最終参照日：2023年8月29日.
- [2] A. S. Arif and W. Stuerzlinger. Analysis of Text Entry Performance Metrics. In *2009 IEEE Toronto International Conference Science and Technology for Humanity (TIC-STH)*, pp. 100–105, 2009.
- [3] X. Bi and S. Zhai. IJQwerty: What Difference Does One Key Change Make? Gesture Typing Keyboard Optimization Bounded by One Key Position Change from Qwerty. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pp. 49–58, New York, NY, USA, 2016. Association for Computing Machinery.
- [4] E. Chan, T. Seyed, W. Stuerzlinger, X.-D. Yang, and F. Maurer. User Elicitation on Single-Hand Microgestures. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pp. 3403–3414, New York, NY, USA, 2016. Association for Computing Machinery.
- [5] W. Chen, L. Chen, Y. Huang, X. Zhang, L. Wang, R. Ruby, and K. Wu. Taprint: Secure Text Input for Commodity Smart Wristbands. In *The 25th Annual International Conference on Mobile Computing and Networking*, MobiCom '19, pp. 1–16, New York, NY, USA, 2019. Association for Computing Machinery.
- [6] W. Chen, L. Chen, M. Ma, F. S. Parizi, S. Patel, and J. Stankovic. ViFin: Harness Passive Vibration to Continuous Micro Finger Writing with a Commodity Smartwatch. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 5(1):1–25, 2021.
- [7] X. A. Chen, T. Grossman, and G. Fitzmaurice. Swipeboard: A Text Entry Technique for Ultra-Small Interfaces That Supports Novice to Expert Transitions. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, UIST '14, pp. 615–620, New York, NY, USA, 2014. Association for Computing Machinery.
- [8] Y. Gao, S. Zeng, J. Zhao, W. Liu, and W. Dong. AirText: One-Handed Text Entry in the Air for COTS Smartwatches. *IEEE Transactions on Mobile Computing*, pp. 1–14, 2021.
- [9] J. Gong, Z. Xu, Q. Guo, T. Seyed, X. A. Chen, X. Bi, and X.-D. Yang. WrisText: One-Handed Text Entry on Smartwatch Using Wrist Gestures. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pp. 1–14, New York, NY, USA, 2018. Association for Computing Machinery.
- [10] T. Götzelmann and P.-P. Vázquez. InclineType: An Accelerometer-Based Typing Approach for Smartwatches. In *Proceedings of the XVI International Conference on Human Computer Interaction*, Interacción '15, pp. 1–4, New York, NY, USA, 2015. Association for Computing Machinery.
- [11] C. Harrison and S. E. Hudson. Abracadabra: Wireless, High-Precision, and Unpowered Finger Input for Very Small Mobile Devices. In *Proceedings of the 22nd Annual ACM Symposium on User Interface Software and Technology*, UIST '09, pp. 121–124, New York, NY, USA, 2009. Association for Computing Machinery.
- [12] J. Hong, S. Heo, P. Isokoski, and G. Lee. SplitBoard: A Simple Split Soft Keyboard for Wristwatch-Sized Touch Screens. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pp. 1233–1236, New York, NY, USA, 2015. Association for Computing Machinery.
- [13] D.-Y. Huang, L. Chan, S. Yang, F. Wang, R.-H. Liang, D.-N. Yang, Y.-P. Hung, and B.-Y. Chen. DigitSpace: Designing Thumb-to-Fingers Touch Interfaces for One-Handed and Eyes-Free Interactions. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pp. 1526–1537, New York, NY, USA, 2016. Association for Computing Machinery.
- [14] A. Ishii, H. Hakoda, and B. Shizuki. Flicky: Flick-Based QWERTY Software Keyboard for Ultra-small Touch Screen Devices. In M. Kurosu ed., *Human-Computer Interaction. Interaction Technologies*, pp. 31–42. Springer International Publishing, 2018.
- [15] H. Jiang, D. Weng, Z. Zhang, and F. Chen. HiFinger: One-Handed Text Entry Technique for Virtual Environments Based on Touches between Fingers. *Sensors*, 19(3063), 2019.
- [16] J. Kim, W. Delamare, and P. Irani. ThumbText: Text Entry for Wearable Devices Using a Miniature Ring. In *Proceedings of the 44th Graphics Interface Conference*, GI '18, pp. 18–25, Waterloo, CAN, 2018. Canadian Human-Computer Communications Society.
- [17] N. Kimura. Self-Supervised Approach for Few-Shot Hand Gesture Recognition. In *Adjunct Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, UIST '22 Adjunct, pp. 1–25, New York, NY, USA, 2022. Association for Computing Machinery.
- [18] G. Laput, R. Xiao, and C. Harrison. ViBand: High-Fidelity Bio-Acoustic Sensing Using Commodity Smartwatch Accelerometers. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, UIST '16, pp. 321–333, New York, NY, USA, 2016. Association for Computing Machinery.
- [19] L. A. Leiva, A. Sahami, A. Catala, N. Henze, and A. Schmidt. Text Entry on Tiny QWERTY Soft Keyboards. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pp. 669–678, New York, NY, USA, 2015. Association for Computing Machinery.
- [20] X. Lin, Y. Chen, X.-W. Chang, X. Liu, and X. Wang. SHOW: Smart Handwriting on

- Watches. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(4):1–23, 2018.
- [21] A. Mottelson, C. Larsen, M. Lyderik, P. Strohmeier, and J. Knibbe. Invisi-board: Maximizing Display and Input Space with a Full Screen Text Entry Method for Smartwatches. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI ’16, pp. 53–59, New York, NY, USA, 2016. Association for Computing Machinery.
 - [22] S. Oney, C. Harrison, A. Ogan, and J. Wiese. ZoomBoard: A Diminutive QWERTY Soft Keyboard Using Iterative Zooming for Ultra-Small Devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’13, pp. 2799–2802, New York, NY, USA, 2013. Association for Computing Machinery.
 - [23] Y.-F. Shao, M. Chang-Ogimoto, R. Pointner, Y.-C. Lin, C.-T. Wu, and M. Chen. SwipeKey: A Swipe-Based Keyboard Design for Smartwatches. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI ’16, pp. 60–71, New York, NY, USA, 2016. Association for Computing Machinery.
 - [24] C. Shi, V. Lakafosis, M. H. Ammar, and E. W. Zegura. Serendipity: Enabling Remote Computing among Intermittently Connected Mobile Devices. In *Proceedings of the Thirteenth ACM International Symposium on Mobile Ad Hoc Networking and Computing*, MobiHoc ’12, pp. 145–154, New York, NY, USA, 2012. Association for Computing Machinery.
 - [25] K. A. Siek, Y. Rogers, and K. H. Connolly. Fat Finger Worries: How Older and Younger Users Physically Interact with PDAs. In *Proceedings of the 2005 IFIP TC13 International Conference on Human-Computer Interaction*, INTERACT ’05, pp. 267–280, Berlin, Heidelberg, 2005. Springer-Verlag.
 - [26] M. Soliman, F. Mueller, L. Hegemann, J. S. Roo, C. Theobalt, and J. Steimle. FingerInput: Capturing Expressive Single-Hand Thumb-to-Finger Microgestures. In *Proceedings of the 2018 ACM International Conference on Interactive Surfaces and Spaces*, ISS ’18, pp. 177–187, New York, NY, USA, 2018. Association for Computing Machinery.
 - [27] P. Streli, J. Jiang, A. R. Fender, M. Meier, H. Romat, and C. Holz. TapType: Ten-Finger Text Entry on Everyday Surfaces via Bayesian Inference. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI ’22, pp. 1–16, New York, NY, USA, 2022. Association for Computing Machinery.
 - [28] K. Sun, Y. Wang, C. Yu, Y. Yan, H. Wen, and Y. Shi. Float: One-Handed and Touch-Free Target Selection on Smartwatches. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI ’17, pp. 692–704, New York, NY, USA, 2017. Association for Computing Machinery.
 - [29] P. C. Wong, K. Zhu, and H. Fu. FingerT9: Leveraging Thumb-to-Finger Interaction for Same-Side-Hand Text Entry on Smartwatches. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI ’18, pp. 1–10, New York, NY, USA, 2018. Association for Computing Machinery.
 - [30] E. Wu, Y. Yuan, H.-S. Yeo, A. Quigley, H. Koike, and K. M. Kitani. Back-Hand-Pose: 3D Hand Pose Estimation for a Wrist-Worn Camera via Dorsum Deformation Network. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, UIST ’20, pp. 1147–1160, New York, NY, USA, 2020. Association for Computing Machinery.
 - [31] C. Xu, P. H. Pathak, and P. Mohapatra. Finger-Writing with Smartwatch: A Case for Finger and Hand Gesture Recognition Using Smartwatch. In *Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications*, HotMobile ’15, pp. 9–14, New York, NY, USA, 2015. Association for Computing Machinery.
 - [32] X. Xu, J. Gong, C. Brum, L. Liang, B. Suh, S. K. Gupta, Y. Agarwal, L. Lindsey, R. Kang, B. Shahsavari, T. Nguyen, H. Nieto, S. E. Hudson, C. Maalouf, J. S. Mousavi, and G. Laput. Enabling Hand Gesture Customization on Wrist-Worn Devices. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI ’22, pp. 496:1–496:19, New York, NY, USA, 2022. Association for Computing Machinery.
 - [33] Z. Xu, W. Chen, D. Zhao, J. Luo, T.-Y. Wu, J. Gong, S. Yin, J. Zhai, and X.-D. Yang. BiTipText: Bimanual Eyes-Free Text Entry on a Fingertip Keyboard. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI ’20, pp. 1–13, New York, NY, USA, 2020. Association for Computing Machinery.
 - [34] Z. Xu, P. C. Wong, J. Gong, T.-Y. Wu, A. S. Nittala, X. Bi, J. Steimle, H. Fu, K. Zhu, and X.-D. Yang. TipText: Eyes-Free Text Entry on a Fingertip Keyboard. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, UIST ’19, pp. 883–899, New York, NY, USA, 2019. Association for Computing Machinery.
 - [35] X. Yi, C. Yu, W. Shi, X. Bi, and Y. Shi. Word Clarity as a Metric in Sampling Keyboard Test Sets. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI ’17, pp. 4216–4228, New York, NY, USA, 2017. Association for Computing Machinery.

著者索引

- 浅野 恭志, 48
五十嵐 健夫, 25
井原 圭一, 41
Erwin Wu, 17
浦垣 啓志郎, 92
大塚 真柊, 62
大西 鮎美, 34, 48
岡 大貴, 34
夏山 英高, 92
鎌田 航誠, 69
川口 一画, 41
川崎 竜也, 99
倉持 友哉, 76
栗原 一貴, 1
小池 英樹, 10, 17, 92
幸谷 有紗, 92
佐藤 俊樹, 62
志築 文太郎, 106
SHAMIR Ariel, 25
SECHAYK Yotam, 25
高橋 治輝, 69
塚田 浩二, 69
塚本 昌彦, 34, 48
角田 陸, 106
寺澤 真一郎, 62
寺田 努, 34, 48
中島 柚斗, 84
中野 優靖, 76
西田 健志, 34
西村 晶太郎, 62
橋本 侑樹, 62
馬場 雪乃, 84
濱崎 雅弘, 76
松本 高, 17
真鍋 宏幸, 99
丸山 礼華, 1
宮下 芳明, 55
宮藤 詩緒, 92
守 新太郎, 41
山中 祥太, 1
余 志皓, 10
楊 明哲, 84
Chen-Chieh Liao, 17
廖 振傑, 10

日本ソフトウェア科学会 研究会資料シリーズ ISSN 1341-870X
No. 100 2023年11月29日発行

第31回 インタラクティブシステムと
ソフトウェアに関するワークショップ (WISS 2023) 論文集