

# マルチモーダルデジタルツインによるアクセシブル社会の実現

研究代表者 森島 繁生  
(先進理工学部 応用物理学科 教授)

## 1. 研究課題

平成 19 年度から 22 年度まで、JST 未来社会創造事業「分散型匿名化処理によるプライバシープリアード AI 基盤構築」のサポートにより研究を遂行した。現在は開発技術のコンテンツ・映像業界への技術展開のみならず、障害者支援など社会貢献をも視野に入れている点が特徴である。近未来において実現されるデジタルツイン社会のあるべき姿として、環境やユーザ自身の正確なコピーを仮想世界に展開することのみならず、実世界と仮想世界間の相互のシームレスなインタラクション、デジタルツイン同士の情報共有を経ることによって、最終的に安全・安心で豊かな生活を、個人の特性を越えて誰もが恩恵にあずかれる世界を想定している。現在、デジタルツイン社会で構想されるサービスの多くは健常者にとっての貢献が前提となっており、障害者や高齢者、幼児など社会的な弱者に必ずしも十分に配慮した内容とはなっていない。本研究課題では、特に視覚障害者等が健常者と分け隔てなくサービスを楽しめるデジタルツイン社会の実現を目的としている。視覚のみならず、聴覚、触覚などの五感のマルチモダリティを駆使することによって、現在の社会的な弱者が健常者と同等レベルのサービスを楽しむことが可能となり、真のダイバーシティ社会の実現に貢献するものである。

## 2. 主な研究成果

本年度は、視覚障害者を案内するためのマルチモーダル大規模言語モデル (MLLM)、視覚障害者の探索を支援する地図なしロボット、視覚障害者を案内するための Visual Language Navigation (VLN) モデルとそのベンチマークの研究開発を行なった。

また、視覚障害者の探索を支援する地図なしロボットの開発にも取り組んだ。視覚障害者を案内するためのロボットは活発に研究開発が行われているが、それらは全て使用する建物において地図を用意する必要があるため、さまざまな場所で使うことができない。また、これまで主に視覚障害者の移動を支援する手法が確立されてきたが、一人で周囲を歩いて楽しむ探索の支援は、そのタスクが非常に困難であるため行われてこなかった。そこで本研究では、地図なしで視覚障害者の探索を、事前準備した地図なしで行うシステムを提案する。本システムをデザインするために、私は日本科学未来館および虎ノ門ヒルズにおいて 10 名の視覚障害者を対象に実験を行い、システムの要件を明らかにした。例えば、ユーザ個人によって好む説明の粒度は異なるため、システムは 3 段階 (短い、普通、長い) の長さで周囲を説明し、それらを調節できるようにすべきことを明らかにした。また、要件に基づいて

開発したシステムを使用し、最終的に日本科学未来館で実験を行い、システムの有効性を確かめた。本研究は HCI 分野の最難難関国際会議である CHI2025 に採択された。

また、視覚障害者を案内するための VLN モデルとそのベンチマークの研究開発も行なった。VLN モデルは、言葉で行き先を伝えることによりロボットが自動で地図なしに目的地まで行けるようにする可能性がある。具体的には、視覚障害者周囲のせい感謝にいきたい場所までの行き方をきき、晴眼者に説明してもらい、それを聞いた目的地までロボットが自動で走行する。しかし、現在の VLN モデルは R2R データセットのような実世界からかけ離れた環境において研究が行われている。そこで本研究では、実世界の特性を反映したベンチマークを開発した。ベンチマークは、実世界の構造や見た目を再現した仮想環境と経路説明のデータからなる。経路説明のデータを実際にユーザスタディを通じて収集することにより、説明データも実際に人から得られるものを取得した。また、大規模言語モデル (LLM) とロボットの API を融合した手法も提案した。本ベンチマークにおいて既存の VLN モデルを評価したところ、既存の VLN モデルでは実世界のシナリオにおいては低い性能しか発揮できないことを示し、また提案手法の有効性を確認した。本研究はロボティクスの難関会議である ICRA2025 にてポスター論文として採択された。

また、昨年度実験を実施した、視覚障害者を案内する地図なしスマートフォンシステムおよび視覚障害者のウィンドウショッピングを支援するスマートフォンシステムの発表も、2024 年 10 月の Melbourne 開催の MobileHCI2024 において行なった [6][7]。

### 3. 共同研究者

斎藤 隼介, Meta Reality Labs. USA

谷田川達也, 准教授, 一橋大学

浅川 智恵子, 日本科学未来館館長 (2021 年 4 月～), IBM フェロー

高木 啓伸, 日本 IBM, 日本科学未来館副館長 (2021 年 4 月～)

Hubert Shum, Associate Professor, Dahrur University

斎藤 英雄, 教授, 慶応大学

杉本 麻紀, 教授, 慶応大学

五十川真理子, 准教授, 慶応大学

中村 哲, 教授, 奈良先端大学院大学

吉井 和佳, 教授, 京都大学

浜中 雅敏, 理研

### 4. 研究業績

#### 4.1 国際会議論文 (査読あり) 2023.04–2024.03

[1] SynchroDexterity: Rapid Non-Dominant Hand Skill Acquisition with Synchronized Guidance in Mixed Reality, Ryudai Inoue, Qi Feng, Shigeo Morishima, IEEE Conference on Virtual Reality and 3D User Interfaces 2025 年 3 月

[2] SyncViolinist: Music-Oriented Violin Motion Generation Based on Bowing and Fingering, Hiroki Nishizawa, Keitaro Tanaka, Asuka Hirata, Shugo Yamaguchi, Qi Feng,

Masatoshi Hamanaka, Shigeo Morishima, equal contribution, IEEE/CVF Winter Conference on Applications of Computer Vision, WACV2025 2025年2月

[3] Unsupervised Pitch-Timbre-Variation Disentanglement of Monophonic Music Signals Based on Random Perturbation and Re-entry Training, Keitaro Tanaka, Kazuyoshi Yoshii, Simon Dixon, Shigeo Morishima, APSIPA Transactions on Signal and Information Processing 14(1) 2025年2月

[4] Capturing Dynamic Identity Features for Speaker-Adaptive Visual Speech Recognition, Sara Kashiwagi, Keitaro Tanaka, Shigeo Morishima, APSIPA ASC 2024 2024年12月

[5] Onset-and-Offset-Aware Sound Event Detection via Differentiable Frame-to-Event Mapping, Tomoya Yoshinaga, Keitaro Tanaka, Yoshiaki Bando, Keisuke Imoto, Shigeo Morishima, IEEE Signal Processing Letters 32 186-190 2024年11月

[6] ChitChatGuide: Conversational Interaction Using Large Language Models for Assisting People with Visual Impairments to Explore a Shopping Mall, Yuka Kaniwa, Masaki Kuribayashi, Seita Kayukawa, Daisuke Sato, Hironobu Takagi, Chieko Asakawa, Shigeo Morishima, Proceedings of the ACM on Human-Computer Interaction 8(MHCI) 1-25, 2024年9月24日

[7] Snap&Nav: Smart Phone-based Indoor Navigation System For Blind People via Floor Map Analysis and Intersection Detection, Masaya Kubota, Masaki Kuribayashi, Seita Kayukawa, Hironobu Takagi, Chieko Asakawa, Shigeo Morishima, Proceedings of the ACM on Human-Computer Interaction 8(MHCI) 1-22 2024年9月24日

[8] 話者固有の発話特性に着目したマルチタスク学習に基づく読唇精度向上手法, 柏木 爽良, 田中 啓太郎, 森島 繁生, Visual Computing, VC2024 2024年9月

[9] Detect Fake with Fake: Leveraging Synthetic Data-driven Representation for Synthetic Image Detection, Hina Otake, Yoshihiro Fukuhara, Yoshiki Kubotani, Shigeo Morishima, Trust What You learn (TWYN) Workshop (Organized in conjunction with ECCV 2024) Women in Computer Vision workshop (Organized in conjunction with ECCV 2024) 2024年9月

[10] The Gap in the Strategy of Recovering Task Failure between GPT-4V and Humans in a Visual Dialogue, Ryosuke Oshima, Seitaro Shinagawa, Shigeo Morishima, Proceedings of the 25th Annual Meeting of the Special Interest Group on Discourse and Dialogue 728-745 2024年9月

[11] Adaptive Sampling for Monte-Carlo Event Imagery Rendering, Yuichiro Manabe, Tatsuya Yatagawa, Shigeo Morishima, Hiroyuki Kubo, ACM SIGGRAPH 2024 Posters, 2024年7月

[12] Idea Track: Improving Sample Efficiency in World Models through Semantic Exploration via Expert Demonstration, Kensuke Tatematsu, Hideki Tsunashima, Morishima Shigeo, Fortieth International Conference on Machine Learning, ICML2024, 2024年7月

[13] Keep Eyes on the Sentence: An Interactive Sentence Simplification System for English Learners Based on Eye Tracking and Large Language Models, Taichi Higasa, Keitaro Tanaka, Qi Feng, Shigeo Morishima, Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, 2024年5月

#### 4.2 国内会議論文（査読あり）

- [1] Enhancing Non-Dominant Hand Skills Through Inverted Visual Feedback in a Mixed Reality Environment, Ryudai Inoue, Qi Feng, Shigeo Morishima, 第32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, WISS2024, 2024年12月
- [2] AIによる質問を利用したメール返信支援システムの評価, 三浦 悠輔, 楊 期蘭, 栗林 雅希, 松本 啓吾, 葛岡 英明, 森島 繁生, 第32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, WISS2024, 2024年12月
- [3] 視覚障害者の探索を支援する事前準備した地図を必要としない案内ロボット, 栗林 雅希, 上原 康平, Allan Wang, 森島 繁生, 浅川 智恵子, 第32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, WISS2024, 2024年12月
- [4] 浅水方程式における剛体カップリングのための浮力計算法の検討, 平江 陽香, 森島 繁生, 安東 遼一, VisualComputing, VC2024, 2024年11月
- [5] 生成画像から獲得された表現を活用した生成画像検出, 大竹 ひな, 福原 吉博, 久保谷 善記, 森島 繁生, VisualComputing, VC2024, 2024年9月
- [6] 少数の参照画を用いたアニメ線画の自動彩色, 高野 悠, 前島 謙宣, 山口 周悟, 森島 繁生, Visual Computing, VC2024, ポスター発表, 2024年9月
- [7] Projection-Based Monocular Depth Prediction for 360 Images with Scale Awareness, Qi Feng, Shigeo Morishima, VisualComputing, VC2024, 2024年9月

#### 4.3 国内会議論文（査読なし）

- [1] 低リソース言語の自動音声認識における他言語データの効率的利用, 三森俊祐, 柏木 爽良, 田中啓太郎, 森島繁生, 情報処理学会 第87回全国大会, 2025年3月
- [2] プロソディ特徴を考慮した感情豊かな音声駆動 3D 発話顔生成, 坂本翔之進, 森島繁生, 情報処理学会 第87回全国大会, 2025年3月
- [3] Microfacet 理論に基づく複数回の反射を考慮した偏光レンダリング, 大羽英仁(早大), 谷田川達也(一橋大), 森島繁生(早大), 情報処理学会 第87回全国大会, 2025年3月
- [4] 少数の参照画を用いたアニメ線画の自動彩色, 高野 悠, 前島 謙宣, 山口 周悟, 森島 繁生, 情報処理学会 第87回全国大会, 2025年3月
- [5] ダンスの楽曲と振付に基づく対話型カメラワーク探索システムの提案, 鈴木 悠, 岩本尚也, 森島繁生, 情報処理学会 第87回全国大会, 2025年3月
- [6] SyncViolinist: Music-Oriented Violin Motion Generation Based on Bowing and Fingering, Hiroki Nishizawa, Keitaro Tanaka, Asuka Hirata, Shugo Yamaguchi, Qi Feng, Masatoshi Hamanaka, Shigeo Morishima, 第142回音楽情報科学研究発表会, 2025年3月
- [7] 変分オートエンコーダを用いた単旋律音楽信号の音高・音色・変動への分解, 田中 啓太郎, 吉井 和佳, Simon Dixon, 森島 繁生, 情報処理学会第141回音楽情報科学研究会, 2024.8
- [8] Investigation of Non-Dominant Hand Training through Virtual Reality and Inverted Visual Feedback, Inoue Ryudai (Waseda University), Feng Qi (Waseda Research Institute for Science, Engineering), Morishima Shigeo (Waseda Research Institute for Science, Engineering), ヒューマンコンピュータインタラクティブ研究会 (IPSJ-HCI) 33, 2024.6
- [9] フロアマップ解析と交差点検出を用いた不慣れた屋内空間における視覚障害者のための

案内システム, 久保田, 雅也, 栗林 雅希, 粥川 青汰, IBM, Research - Tokyo, 高木 啓伸, IBM, Research - Tokyo, 浅川 智恵子, IBM Research, 森島 繁生, ヒューマンコンピュータインタラクション研究会 (IPSI-HCI) 35 2024. 6

#### 4.4 受賞・表彰

[1] 少数の参照画を用いたアニメ線画の自動彩色, 高野 悠, 前島 謙宣, 山口 周悟, 森島 繁生, 情報処理学会 第 87 回全国大会, 学生奨励賞, 2025 年 3 月

[2] Understanding and Supporting Formal Email Exchange by Answering AI-Generated Questions, Yusuke Miura, Chi-Lan Yang, Masaki Kuribayashi, Keigo Matsumoto, Hideaki Kuzuoka, Shigeo Morishima, CHI2025 HONORABLE MENTION, 2025. 3

[3] Understanding and Supporting Formal Email Exchange by Answering AI-Generated Questions, Yusuke Miura, Chi-Lan Yang, Masaki Kuribayashi, Keigo Matsumoto, Hideaki Kuzuoka, Shigeo Morishima, 小野梓記念学術賞, 2025. 3

[4] Onset-and-Offset-Aware Sound Event Detection via Differentiable Frame-to-Event Mapping, Tomoya Yoshinaga, Keitaro Tanaka, Yoshiaki Bando, Keisuke Imoto, Shigeo Morishima, 小野梓記念学術賞, 2025. 3

[5] WanderGuide: Indoor Map-less Robotic Guide for Exploration by Blind People, Masaki Kuribayashi, Kohei Uehara, Allan Wang, Shigeo Morishima, Chieko Asakawa, 小野梓記念学術賞, 2025. 3

[6] 音響イベント検出のための隠れセミマルコフモデルに基づくイベント単位損失, 吉永朋矢, 坂東宜昭, 田中啓太郎, 井本桂右, 大西正輝, 森島繁生, 第 29 回日本音響学会, 学生優秀発表賞, 2024. 12

[7] 汎用事前学習済みモデルを用いた音響イベント検出のための HSMM に基づくイベント単位学習, 吉永朋矢, 田中啓太郎, 坂東宜昭, 井本桂右, 大西正輝, 森島繁生, 応用音響研究会 (EA), 学生研究奨励賞, 2024. 11

[8] 話者固有の発話特性に着目したマルチタスク学習に基づく読唇精度向上手法, 柏木爽良, 田中啓太郎, 森島繁生, Visual Computing, VC2024, 学生研究賞, 2024. 9

[9] Microfacet 理論に基づく複数回の散乱を考慮した偏光レンダリング, 大羽 英仁, 谷田川 達也, 森島 繁生, Visual Computing, VC2024, VC ポスター賞

[10] プロソディ特徴を考慮した表情豊かな音声駆動 3D 発話顔生成, 坂本 翔之進, 森島 繁生, Visual Computing, VC2024, VC ポスター賞, 2025. 9

[11] 変分オートエンコーダを用いた単旋律音楽信号の音高・音色・変動への分解, 田中 啓太郎, 吉井 和佳, Simon Dixon, 森島 繁生, 情報処理学会 第 141 回音楽情報科学研究会 夏のシンポジウム ベストプレゼンテーション賞 (Best Research 部門), 2024. 8

[12] 布の部分空間シミュレーションに向けたニューラルネットによる非線形項評価法の検討, 田中 瑞城, 谷田川 達也, 森島 繁生, 第 192 回 CGVI 研究会学生発表賞, 2025. 6

#### 5. 研究活動の課題と展望

2025 年度は、JST CRONOS 「AI 時代の知的通信サービス技術開発 (代表：延原 章平)」, JST CREST (2 件) 「分散連携型モバイル AI エージェントによる協働行動誘発ナビゲーション (代表：斎藤 英雄)」 「演奏における実環境音楽 AI 基盤の創出 (代表：浜中 雅俊)」に Co-PI

としてプロジェクト提案を行う。また JST 日英 ASPIRE に日本側の PI として「Trust-Embodied Human-Robot Interaction: A Human-Centric Framework for Dynamic Trust Building(代表: Hubert Shum)」の提案を行う。さらに現在進行している NEDO の調査研究プロジェクト「アニメクリエイター創出・制作支援のためのアニメ×生成 AI の利活用に関する調査(代表機関: オー・エル・エム・デジタル)」を、次の本格研究フェーズ立ち上げのための準備に取り掛かる。