

第14部

SoftwareDefinedMediaコンソーシアム

SDM WGメンバー

第1章 はじめに

Software Defined Media (SDM)コンソーシアムは、2014年に設立され、オブジェクト指向のデジタルメディアと、インターネット環境が前提の映像・音響空間を用いたビジネス創造を目指してきた。昨年までの報告書では、SDMのコンセプトやアーキテクチャ、試作プラットフォーム[63, 64]、データ収録[65]、収録データベース設計[66, 67, 68]、SDMプラットフォームを利用したアプリケーションであるSDM3602[69, 70]やLiVRation[71, 72]、Web3602[73]などを報告してきた。

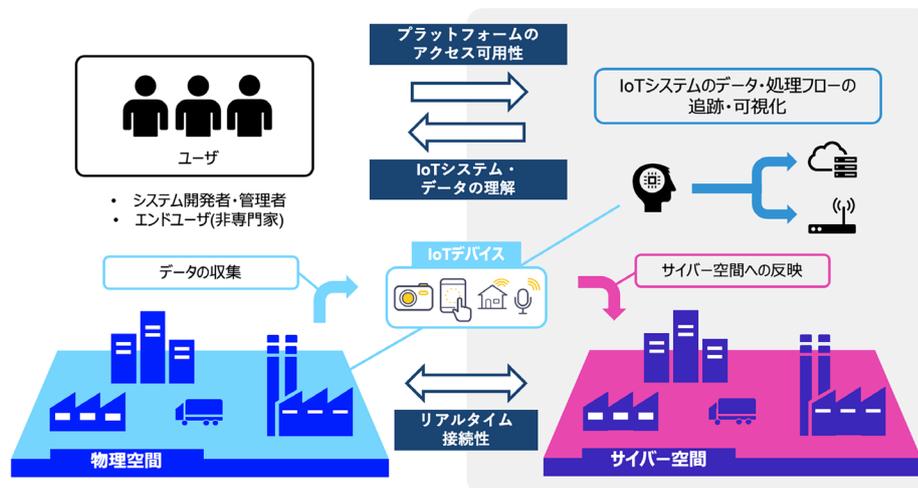
本報告書では、2024年度に実施した活動を報告する。

第2章 IoTデバイスの効率的運用に向けたイベント駆動プロトコルとオープン基盤

近年、Society 5.0の実現に向けて、IoT技術の活用が進んでいる。IoTシステムは、物理空間とサイバー空間を統合し、データ駆動型の効率的なシステム管理を可能にする。しかし、IoTデバイスのバッテリーや計算リソースの制約、システム間の相互運用性の欠如、一般ユーザーにとってのアクセスの難しさといった課題が、持続可能な発展を阻む要因となっている。本研究では、これらの課題に対応するために、イベント駆動プロトコルとオープン基盤を提案した。

2.1 イベント駆動プロトコル

IoTデバイスのエネルギー効率を向上させるため、本研究ではイベント駆動プロトコルを開発した。このプロトコルは、デバイスの状態やネットワーク環境に応じて適応



的に制御を行うことで、不要な電力消費を抑え、持続可能な運用を実現する。本プロトコルは、以下の二つの技術で構成される。

間欠的デバイス制御: イベントの発生時のみデバイスをアクティブにすることで、消費電力を削減する。

オフローディング最適化: スループットやデータサイズに基づき、クラウドまたはエッジ処理のどちらが適切かを動的に選択する。

2.2 オープン基盤

IoTシステムの相互運用性を向上させるため、Webベースのデジタルツイン技術を活用したオープン基盤を設計・実装した。本基盤の要件は以下の三つである。

Webベースのアクセス: 特定のデバイスや専用ソフトウェアを必要とせず、汎用的な環境での利用が可能である。

視覚的なデータ理解: IoTデータを直感的に可視化し、非専門家でもシステムの状態を把握できる設計である。

リアルタイム性: デバイスからのデータを即座に反映し、迅速なフィードバックを可能とする。

第3章 遠隔ライブ視聴のための音声遅延が与える影響の調査

デジタルツインの概念があらゆる業界で注目を浴びており、様々なシミュレーションのために現実の世界をそのまま仮想世界上に再現したモデルである。一方で、メタバースは、同様に仮想空間を構築するためデジタルツインと似た概念であるものの、アバターを介した人々の交流や経済活動が行われる場であり、メタバース空間上には現実空間で実在するもの、あるいはしないものの両方が混在する。また、仮想空間でアバターを使って交流するメタバースは、近年注目されているサービスであり、スポーツ観戦や音楽ライブなどのイベントも開催される。しかしメタバース空間における音楽ライブ(例:VR

ライブ)などは、ネットワーク遅延による問題で演者は一箇所に集まらなければならない。また、観客は録音されたステレオの音源を聞くため、生のライブよりも臨場感に欠けることが多い。そこで、演者、観客ともにどこにいても実現できる臨場感の高いVRライブシステムの需要があるのではないかと考えられる。本システムの実現のためには、ネットワーク遅延の影響による遠隔演奏の精度を確かめる必要がある。そこで本稿では、まず、SYNCROOMを用いた事前検証において、遠隔セッションにおける音声遅延を測定した。実環境での検証により、今回の実験環境であれば演奏者間の往復遅延は最大で32.5ms程度に収まることが確認できた。ただし、特定の条件下では遅延が大きくばらつく場合もあり、演奏の質に影響を与える可能性が示唆された。次に、音声遅延シミュレーション実験を通じて、遅延が演奏者および聴講者の知覚に与える影響を評価した。演奏者においては、片道遅延が20msを超えた時点で遅延を感じ始め、60msを超えると演奏が困難になることが明らかとなった。さらに、楽器の種類や組み合わせによる遅延の影響の違い

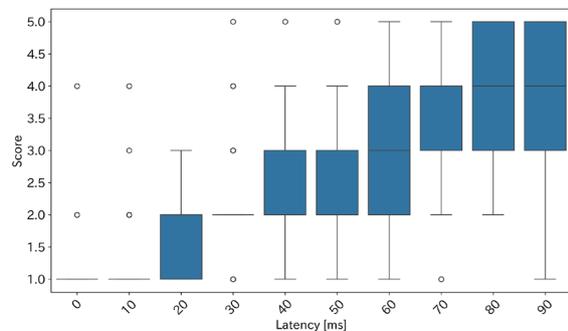


図1 遅延の感じ度合い(pops全体)

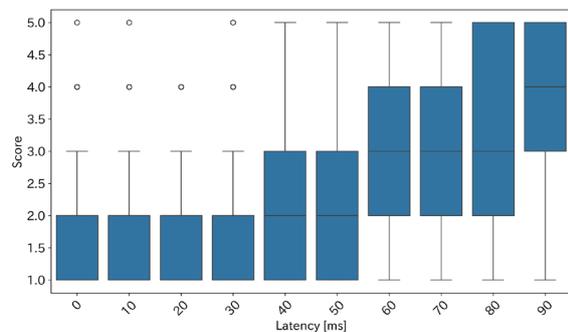


図2 第三者が感じた演奏のずれ(pops全体)

もみられた。例えば、ドラムは遅延に対して特に敏感である一方、ボーカルは比較的遅延を感じにくいという結果となった。さらに第三者評価実験では、録音した演奏を聴講者に提示し、遅延が音楽全体の印象や成立性に与える影響を検討した。その結果、遅延が40msを超えると、演奏のずれを感じる人が多くなることが判明した。逆に遅延が40ms以下であれば、演奏がほぼ自然に聞こえることが示された。これらの結果から遠隔演奏をエンターテインメントとして第三者に配信する際、聴講者に対して違和感のない演奏をすることは十分可能であると考えられる。今後の課題として3人以上で演奏した時や、音声遅延が行きと帰りで異なっている状況での調査、長時間の演奏や複雑な楽曲における遅延の影響の検討、さらに演奏のずれの知覚に関して音源再生機器による差を明確にし原因を特定することが挙げられる。

第 4 章 Edge Vision AI Co-Processing for Dynamic Context Awareness in Mixed Reality

4.1 Introduction

The integration of AI with MR technologies enhances spatial interactions but faces computational limitations on standalone headsets. This project introduces an edge AI co-processing system utilizing a Hailo-8 AI accelerator, a Rockchip RK3588-based SBC, and the Magic Leap 2 AR headset to offload real-time AI processing.

4.2 System Overview

The system streams video via WebRTC, processes AI inferences at the edge, and reprojects results 1:1 into 3D space. Key components include:

- **Hardware:** Rockchip RK3588-based SBC with Hailo-8 AI accelerator and Magic Leap 2.
- **Software:** AIORTC for streaming, GStreamer for preprocessing, and AI models (YOLOv8s, CenterPose, FastDepth, SCRFD+ArcFace) for vision analysis.

4.3 Deployment and Results

The system was demonstrated at **Interop Tokyo 2024**, showcasing five AI pipelines running at **30 FPS**. Evaluation metrics included:

- **Power Consumption:** Efficient scaling with concurrent models.
- **Latency:** Photon-to-photon latency of 167-205ms (3 out of 5 models), ensuring real-time interaction.
- **Pipeline Delay:** Single model delays ranged 9-18ms, increasing with multiple models.

4.4 Future Work

This system enables low-latency, AI-enhanced MR interactions while reducing headset processing load. Future work includes possibly integrating split learning techniques to optimize bandwidth, evaluating multi-user scenarios, and deploying vision AI coprocessing at an infrastructure level.



図3 Live demonstration at Interop 2024



図4 Photo of user testing PLATONE system and remote-desktop view of audio spatialization

4.5 Publications and Demonstrations

- **Conference Papers:** AINA 2025
- **Short Abstract and Poster:** IEEE VR 2025
- **Live Demonstration:** Interop Tokyo 2024

第 5 章 PLATONE: An Immersive Geospatial Audio Spatialization Platform

5.1 Introduction

The convergence of physical and virtual environments highlights the importance of accurate spatial audio for immersive urban experiences. PLATONE is a platform designed to augment city-scale environments with environment-dependent audio spatialization. By leveraging georeferenced 3D building data, custom RTK-GNSS+IMU hardware, and cloudbased audio processing, PLATONE enables realtime, location-aware binaural audio rendering.

5.2 System Overview

PLATONE integrates geospatial data, real-time positioning, and spatialized audio streaming through the following components:

- **Georeferenced 3D Data:** Uses open-source city-scale datasets, such as Japan's MLIT Project Plateau and Switzerland's swisstopo data, to simulate urban sound propagation.
- **Position and Orientation Tracking:** Employs custom hardware, which combines RTK-GNSS and an IMU for precise geolocation and head tracking.
- **Cloud-Based Audio Rendering:** Utilizes Microsoft Project Acoustics and Unreal Engine to precompute sound propagation effects, including occlusion and reverberation, streamed in real time to mobile devices.

5.3 Experimental Deployment and Results PLATONE was evaluated through two key studies:

- **IEEE MetaCom 2024 (Hong Kong):** Foundation paper that demonstrated real-time geospatial audio in urban navigation and immersive media applications.

- **ACM ICEA 2024 (Tokyo):** Assessed the accuracy of simulated audio in Montreux, Switzerland, by comparing real-world recordings to digital twin-generated audio.

The evaluation found relative offsets in decay time between real-world and simulated audio ranging from **-4.03% to +12.13%**, suggesting the need for improved frequency-dependent material modeling and further testing.

5.4 Future Work

PLATONE enables city-scale, spatialized audio experiences with high positional accuracy. Future work will focus on developing an extensive testing methodology to objectively evaluate all acoustic properties of DT audio spatialization in comparison to real-world characteristics, therefore expanding large-scale validation in diverse urban settings.

5.4.1 CREST AIP

New hardware is being developed as a result of CREST AIP funding. The goal of this new hardware is to enable high-precision positioning and orientation for visual Mixed Reality applications through sensor fusion of RTK-GNSS, IMU, barometric pressure sensor, and MR headset self-localization.

5.5 Publications

- **Conference Papers:** IEEE MetaCom 2024, ACM ICEA 2024

第 6 章 その他の活動

- 塚田学, "デジタルツインが拓く新たな建築空間", In: 建築雑誌(特集: データが再構成する建築), vol. 139, no. 1794, pp. 16-17, 2024. [74]

第 7 章 まとめ

本報告書では、2014年より開始したSDMコンソーシアムで進めて来た、視聴空間サービスのソフトウェア制御に関する研究として、2024年度に行なった活動について報告した。

SDMコンソーシアムでは、ソフトウェア処理による視聴空間の制御、映像音声を制御するネットワーク機器、インタラクティブ・ユーザインターフェイス、SDMプラットフォームを利用したコンテンツ作成など、共同研究活動に参加するパートナーを募集しています。

ご協力をよろしくお願いいたします。