

A Generalisable Tool for Spatial Audio Metadata Networking in 5G Virtual Production

Patrick Cairns
CoSTAR Screen Lab
Ulster University
Belfast, United Kingdom
p.cairns1@ulster.ac.uk

Nurudeen Oladehinbo Salau
CoSTAR Screen Lab
Ulster University
Belfast, United Kingdom
n.salau@ulster.ac.uk

Richard Campbell
CoSTAR Screen Lab
Ulster University
Belfast, United Kingdom
r.campbell1@ulster.ac.uk

Adnan Ahmad Cheema
School of Engineering
Ulster University
Belfast, United Kingdom
a.cheema@ulster.ac.uk

Frank Lyons
CoSTAR Screen Lab
Ulster University
Belfast, United Kingdom
fr.lyons@ulster.ac.uk

Declan Keeney
CoSTAR Screen Lab
Ulster University
Belfast, United Kingdom
d.keeney@ulster.ac.uk

Abstract—Modern 5G virtual production contexts include the deployment of networked spatial audio systems. Here object tracking systems provide positional metadata required by spatial audio rendering systems. Discrete custom metadata parameter formats used by discrete systems, however, mean that native interoperability does not typically exist between object tracking and spatial audio rendering systems. Furthermore, the performance of typical best-effort approaches to metadata networking has not been validated for 5G deployments. This paper presents development towards a generalised bridging tool for spatial audio metadata networking in virtual production, and provides preliminary benchmarking of the performance of best-effort metadata streaming over a private 5G network deployment.

Index Terms—Spatial audio, wireless networks, 5G, cinematography, object tracking, virtual production

I. INTRODUCTION

Several recent papers have highlighted the potential for 5G networks to support the deployment of networked spatial audio systems in Virtual Production (VP) contexts [1]–[6]. Potential use-cases in VP include networked asset capture, remote monitoring and pre-visualisation, media distribution and presentation to audience endpoints, and interactive immersive installations. Here audio and object tracking metadata capture is streamed over a 5G network and used in dynamic spatial audio rendering and playback. In these VP applications the 5G standard [7] is considered to provide suitable network support. Here Ultra-Reliable Low-Latency Communication (URLLC) and Software Defined Networking (SDN) features of the 5G standard allow optimisation of Quality of Service (QoS) for media streaming. Multi-Access Edge Computing (MEC) resources may also be deployed to local base stations to provide Digital Signal Processing (DSP) services for relevant spatial computing and spatial audio rendering processes.

In VP contexts object tracking systems and spatial audio systems typically communicate using Open Sound Control (OSC) messages [8]. Here a string-based addressing system is used to associate OSC messages with object tracking or

spatial audio rendering parameters. Both object tracking and spatial audio rendering systems, however, typically implement bespoke addressing definitions. Native interoperability is therefore typically not possible, and VP deployments require implementation of bridging tools which provide OSC translation utility. This paper contributes development towards a general-purpose bridging tool for networked object tracking and spatial audio systems in VP use-cases. The solution allows for selection of input and output OSC profiles, and performs translation between input and output profiles, providing interoperability between object tracking and spatial audio systems. Highlighted here is the potential for adoption of Serialised Audio Definition Model (S-ADM) [9] as an interoperable standard.

The OSC protocol was developed such that OSC messages may be transported as a payload of common TCP/IP transport protocols. For practical tools used in object tracking and immersive audio for VP, however, OSC is typically implemented under best-effort UDP transport. The performance of best-effort OSC transport, however, has not been validated for 5G deployments. This paper contributes preliminary testing of best-effort OSC transport for 5G networked spatial audio. The developed bridge application is deployed across the CoSTAR Screen Lab¹ private 5G network, considering a remote monitoring use-case in the context of VP cinematography. Tracking metadata loss ratio, Mean Squared Error (MSE) and windowed MSE between a reference signal and a signal sent over the 5G network is reported. Results indicate that though 5G networks may be optimised for media and metadata transport, lack of 5G native capabilities in VP hardware requires wireless connection to a 5G router, where significant packet loss may occur. Despite this, simple Packet Loss Concealment (PLC) methods such as linear interpolation are suggested as suitable for concealing tracking metadata loss in 5G VP deployments.

¹www.costarnetwork.co.uk/labs/screen-lab

TABLE I

A SUMMARY OF 5G AUDIO SYSTEM DEPLOYMENT FOR VP, HIGHLIGHTING RELEVANT USE-CASES, AND TYPE OF 5G DEPLOYMENT (PUBLIC, OR PRIVATE). SPATIAL AUDIO RENDERING AND METADATA STREAMING SOLUTIONS ARE HIGHLIGHTED WHERE THEY HAVE BEEN IMPLEMENTED.

Reference	Use case	Audio streaming	Metadata streaming	Spatial audio	Public 5G	Private 5G
[10], [11]	NMP	SoundJack	-	Ambisonic microphone	✓	-
[12]	NMP	FPGA, UDP	-	-	✓	-
[1]	Museum display	-	-	Resonance Audio, MEC	-	-
[13]	NMP	JackTrip	Custom MaxMSP bridge	Spat, HISS, ICST	-	✓
[14], [15]	NMP	Elk Live	-	-	✓	✓
[16]	NMP	JackTrip, SonoBus	-	X-MCFX, DBAP	✓	-
[5]	NMP	-	-	-	Simulated	Simulated
[17]	-	-	-	-	-	Simulated
[18]	XR Cinematography	UE5 Pixel Streaming	-	-	-	✓
This paper	Remote monitoring	-	Generalisable bridge	-	-	✓

II. RELATED WORK

A. Object Tracking and Spatial Audio

With respect to object tracking in VP contexts, spatial audio rendering typically considers Object Based Audio (OBA) assets. Here a discrete audio object is described by a mono audio channel and associated object tracking metadata². This provides the required assets to support spatial audio rendering with dynamic sound sources and listener navigation with 3 or 6 Degrees of Freedom (3DoF/6DoF).

Object tracking metadata may be acquired using a range of VP solutions. Some tracking systems provide native OSC output, for example motion capture system (e.g., Vicon³) or Radio Frequency (RF) stage tracking systems (e.g., StageTrack II⁴ or Follow-Me⁵). Alternatively it is possible to develop systems which provide OSC output via game engines (for example using Unreal Engine 5, UE5, OSC plugins⁶, or third party Unity Tools such as extOSC⁷ or OSCJack⁸). This allows for OSC output of tracking metadata for character controllers and game objects, and from tracking systems which communicate via game engines (e.g., the Vive Mars CamTrack⁹ system, which may connect to UE5 via Live Link¹⁰).

In real-time VP contexts a range of spatial audio rendering solutions are available for binaural stereo headphone playback and playback over multichannel loudspeaker systems (for example consumer [20], specialist Ambisonic arrays [21], and venue systems). A wide range of audio plugin solutions¹¹ are available which support native OSC parameter control and may be hosted in Digital Audio Workstation (DAW) or other plugin hosting software.

²Native spatial audio microphone capture [19] and relevant formats are also utilised in the wider context of OBA. In the context of object tracking systems for VP, however, this is not relevant, and is not discussed further in this paper.

³www.vicon.com

⁴www.tta-sound.com/

⁵www.follow-me.com/

⁶<https://dev.epicgames.com/documentation/unreal-engine/osc-plugin-overview-for-unreal-engine>

⁷www.github.com/Iam1337/extOSC

⁸www.github.com/keijiro/OscJack

⁹www.mars.vive.com/

¹⁰<https://dev.epicgames.com/documentation/unreal-engine/live-link-in-unreal-engine>

¹¹A fair summary of relevant spatial audio plugins may be found in [22] and [23].

Pro-audio solutions are also commercially available which support native OSC parameter control (for example d&b Audiotechnik SoundScape¹², L-Acoustics L-ISA¹³, and FLUX Immersive Spat Revolution¹⁴).

As highlighted in section I, it is typical for both object tracking and spatial audio rendering systems to implement bespoke OSC message definitions. Achieving interoperability therefore typically requires the implementation of bridging software, which provides translation between object tracking and spatial audio rendering OSC profiles, and any relevant spatial computing. In recognition of this lack of interoperability, industry stakeholders have presented the definition ADM OSC¹⁵. This describes an OSC profile where object tracking metadata is defined following Serialised Audio Definition Model (S-ADM) definitions [9]. S-ADM is a serialised iteration of the Audio Definition Model (ADM) standard [24] for OBA. Considering the objective of providing a general-purpose bridging tool for object tracking and spatial audio systems in VP, ADM OSC profile support should be implemented, and ADM OSC definitions may serve as a common reference point for translation across bespoke OSC profiles.

B. 5G Spatial Audio Deployment

A range of recent studies have demonstrated the both private and public 5G networks are capable of supporting low-latency audio streaming (see table I). These studies largely focus on the deployment of Network Music Performance (NMP) systems [25], designed to support musical performance between network endpoints. Practical implementation and evaluation has shown that stable audio streaming with sub-30 ms latency (as required for unimpaired interaction and response between NMP endpoints) can be achieved across a single cell in private 5G deployments [13], [14]. Across larger 5G deployments and public 5G networks audio streaming with sub-30 ms latency has been demonstrated as resulting in significant signal degradation from packet loss [12], [14], [16], however stable audio streaming at sub-100 ms latency levels (as required for teleconferencing solutions) is still possible.

¹²www.dbaudio.com/global/en/soundscape/

¹³www.l-acoustics.com/products/l-isa-immersive/

¹⁴www.flux.audio/project/spat-revolution/

¹⁵<https://github.com/immersive-audio-live/ADM-OSC>

A small number of studies describe practical deployment of spatial audio systems over 5G networks (see table I). Prior work in an NMP and virtual concert use-case [10], [11] includes streaming and playback of Ambisonic microphone capture in an NMP and virtual concert use-case. Here the NMP software SoundJack¹⁶ is used for low-latency audio streaming. In other work [13] detail is provided for an OBA system deployment across a single cell, where object tracking metadata is derived from a Supperware head tracker¹⁷, and bridged using Max MSP¹⁸ to an MEC spatial audio rendering service using Spat alongside HISS¹⁹ and ICST Ambisonic plugins²⁰. Here the NMP software JackTrip²¹ is used for low-latency audio streaming. Server-based spatial audio rendering is deployed in other work [16] using X-MCFX convolver²² instances with a Distance-Based Amplitude Panning (DBAP) algorithm to interpolate between Spatial Impulse Response (SIR) outputs. Here JackTrip and SonoBus²³ are used for low-latency audio streaming, and object tracking metadata is output from a Unity application (implemented using extOSC). Notably no prior work implements a generalisable solution for routing object tracking metadata to spatial audio rendering. Additionally, where object tracking is integrated, OSC parameter control is implemented under best-effort UDP transport, and no validation of the performance of this method over 5G deployment is provided.

III. DEFINING A GENERAL PURPOSE BRIDGING TOOL

In order to enable interoperability between a range of object tracking and spatial audio systems used in VP, this paper explores the implementation of a general purpose bridging tool. The requirements of this tool may be defined:

- To support multiple audio objects (as required for OBA systems), bridging requires support for multiple OBA channels.
- To support spatial audio rendering which allows listener navigation, support for definition of channels as listener or sound source is required.
- To support input from different object tracking systems, OSC input profile and port configuration is required.
- To support output to different spatial audio systems, OSC output profile and port configuration is required. To support networked deployment for remote monitoring, output OSC IP address configuration is also required.
- To support bridging, translation between input and output profiles is required. ADM OSC may be used as a suitable intermediate format for translation. Here translation from any object tracking OSC format may be translated to ADM OSC via an input translation function. Translation

from ADM OSC to any spatial audio OSC format may be translated via an output translation function. In this manner support for any input or output OSC format may be provided via a single input translation and single output translation function, rather than implementing a discrete translation function for each discrete combination of input and output OSC formats.

- To support dynamic rendering, where a listener channel is defined, DSP which computes relative positional metadata for source channels is required.
- To support deployment over 5G networks, it is required that the bridging tool is 5G-capable (i.e., should provide a stable metadata stream for remote monitoring contexts). As highlighted in section II-B, it is unknown whether typical best-effort UDP is sufficient for reliable transport of OBA metadata over 5G networks. As such, the technical requirements of being 5G-capable are not straightforward to define.

IV. PROTOTYPE IMPLEMENTATION AND 5G DEPLOYMENT

A prototype implementation of a general purpose bridging tool was developed as a desktop application and VST plugin using the C++ JUCE²⁴ framework. As a proof-of-concept implementation, the bridge currently only supports a single OBA metadata stream. As only a single metadata channel is supported, specification of a channels as source or listener was not implemented. Specification of tracking metadata input port and OSC profile is supported, where ADM OSC and a bespoke camera tracking OSC profile are selectable. Specification of spatial audio metadata OSC output profile, port and IPv4 address are supported, where ADM OSC and IEM Room Encoder OSC profiles are selectable. OSC input and output are implemented as best-effort UDP. OSC translation overhead is expected to be negligible (i.e., in the order of microseconds), however was not profiled.

The prototype bridge application was tested in an example VP deployment over a 5G network (see figure 1), where a remote monitoring and pre-visualisation context is considered. Here an asset capture endpoint and a monitoring endpoint are connected via the CoSTAR Screen Lab private 5G network. This network operates on the N77 band, and consists of a base station processor and two radio dots providing radio access to the network. At the asset capture endpoint object tracking for a single source is captured relative to a static listener position. This may represent a sound source, such as a performer or presenter, relative to a static viewpoint, such as a camera or virtual camera. The object tracking metadata is bridged to the monitoring endpoint for input to a spatial audio rendering system, such that a producer at the monitoring endpoint is provided with a preview of spatial audio delivery during asset capture, or live monitoring of spatial audio delivery in live broadcast production.

¹⁶www.soundjack.eu

¹⁷www.supperware.co.uk/headtracker-overview

¹⁸www.cycling74.com

¹⁹[www.github.com/HISSTools/](https://github.com/HISSTools/)

²⁰www.ambisonics.ch

²¹[www.github.com/jacktrip/jacktrip](https://github.com/jacktrip/jacktrip)

²²[www.github.com/JB-Luke/X-MCFX](https://github.com/JB-Luke/X-MCFX)

²³www.sonobus.net

²⁴www.juce.com

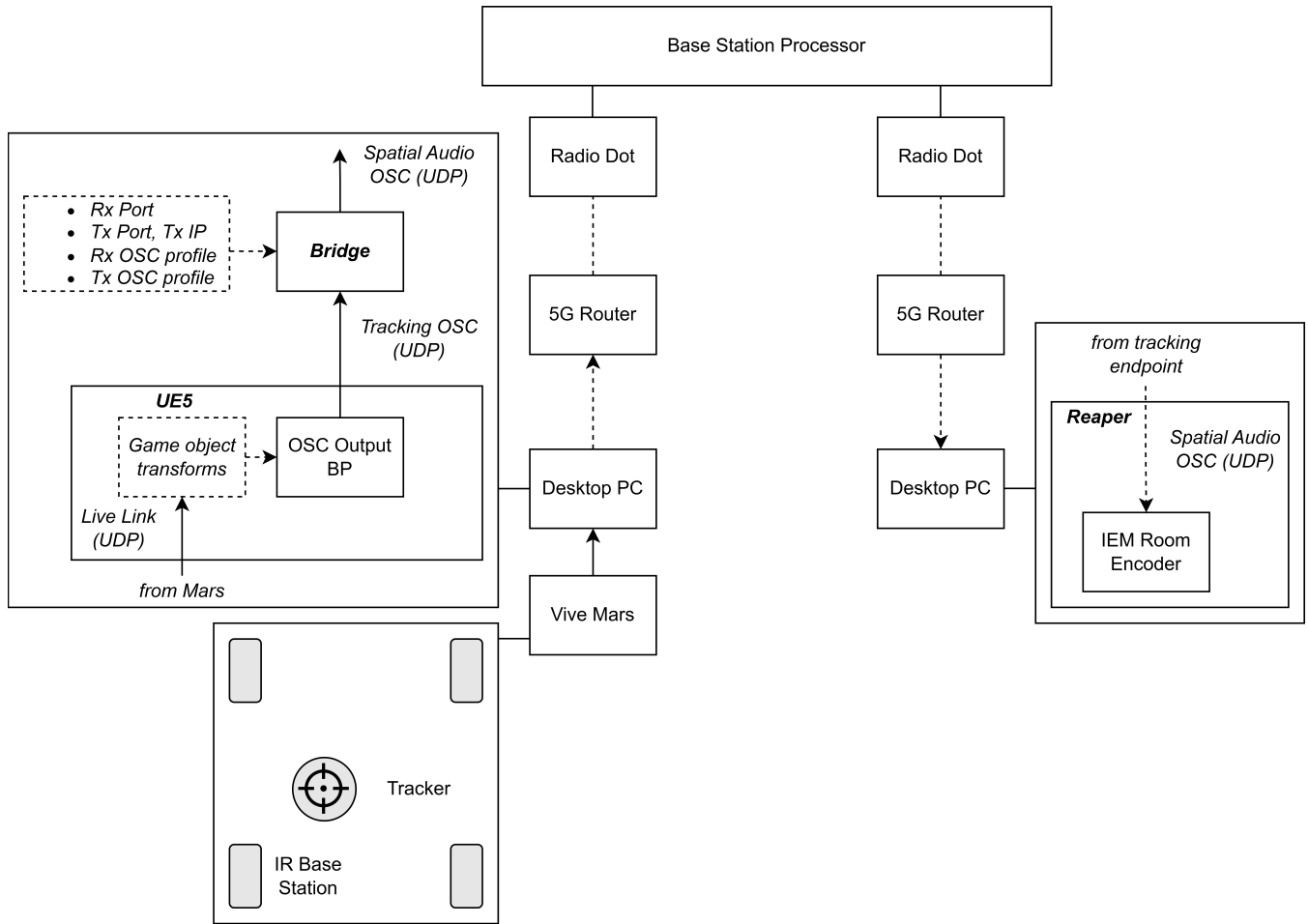


Fig. 1. Overview of a 5G deployment of a networked object tracking and spatial audio system. The system is deployed over the CoSTAR Screen Lab private 5G network at Studio Ulster. Here tracking data is obtained from a Vive Mars CamTrack system via UE5. Tracking OSC output from UE5 is received by the bridge application. The bridge translates Rx OSC to provide sound source position metadata required by IEM Room Encoder. Spatial audio OSC is output from the bridge application and transported as best effort UDP to the spatial audio rendering endpoint, where it is received by an instance of IEM Room Encoder for spatial audio rendering.

At the asset capture endpoint a Vive Mars CamTrack system is used to acquire object tracking metadata. This system consists of four IR base stations and a tracker module, which may be attached to mobile sound sources.

Object tracking metadata is output by the Vive Mars module, which is connected via ethernet to a desktop PC. Object tracking metadata is passed to UE5 via Live Link as UDP, where the tracking positional metadata is applied to a game object (here a cineCam). A custom BP developed for virtual camera tracking is used to obtain game object transforms and output positional metadata for the tracked source following a bespoke OSC message definition. This OSC output is passed in UDP payloads to the bridge application. The bridge application provides translation between the custom camera tracking OSC profile and the relevant OBA profile (here IEM Room Encoder²⁵). The bridge application then outputs OSC control data to the monitoring endpoint via the 5G network.

Here the desktop is connected to a 5G router via wireless 5G connection, and the 5G router is connected to the 5G base station via a radio access installed in the CoSTAR Screen Lab research offices. Access via the 5G router is required for devices which are not 5G-native, and require wireless connection to a 5G-capable router to achieve radio access to the private 5G network.

At the monitoring endpoint a desktop computer hosts spatial audio rendering. Here Reaper²⁶ is used to host spatial audio rendering based on the IEM Plugin Suite. With respect to OBA rendering metadata, OSC parameter control received from the asset capture endpoint is input to an instance of IEM Room Encoder to control the position of a sound source. Received OSC control messages are sampled to plugin parameters with a 20 ms refresh rate. Similarly to the asset capture endpoint, at the monitoring endpoint the desktop has a wireless connection to a 5G router connected to the 5G base station via the radio access.

²⁵<https://plugins.iem.at/>

²⁶www.reaper.fm

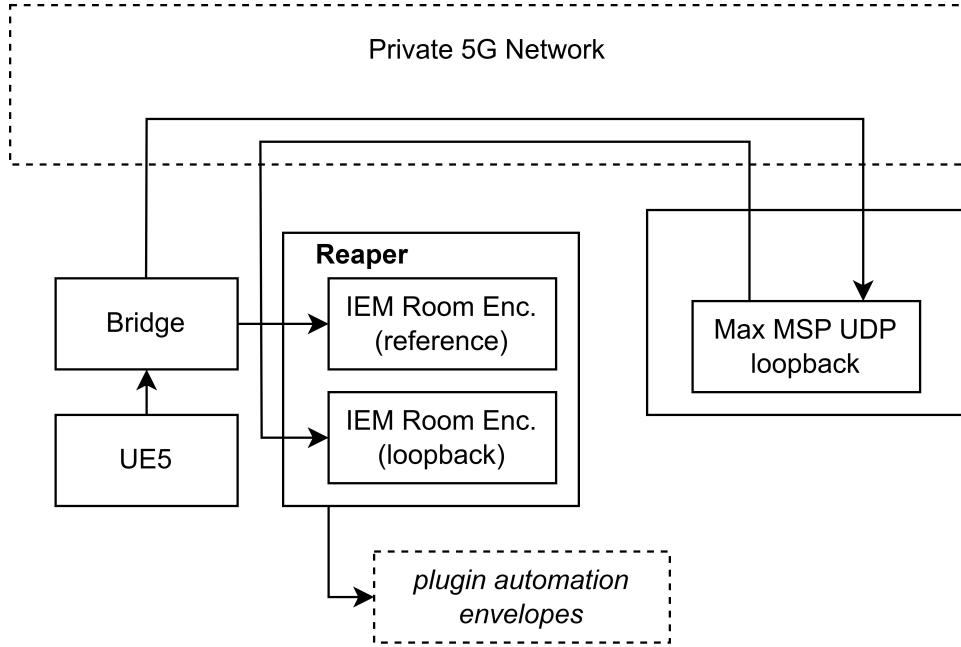


Fig. 2. Overview of preliminary test deployment .

V. PRELIMINARY TESTING OF BEST EFFORT UDP

As highlighted in section II-B the performance of best-effort UDP methods typically used for OSC in object tracking and spatial audio systems have not been validated over 5G networks. In order to contribute a preliminary evaluation, an iteration of the example system described in section IV was deployed over the private 5G network (figure 2).

Here object tracking data is output from UE5 to a modified version of the bridge application. To provide test signal a custom blueprint was implemented which may be attached to an arbitrary actor object and output positional tracking as OSC, following the OSC profile used by the Mars CamTrack system detailed in section IV. This blueprint was attached to a child actor of a player character to track player character movement. The modified bridge receives OSC from UE5, performs translation between the camera tracking OSC profile to the OSC profile for IEM Room Encoder, and outputs spatial audio OSC to two endpoints. The first is an instance of IEM Room Encoder hosted in Reaper. The second is a Max MSP UDP loopback patch hosted on a desktop at the monitoring endpoint. This loopback routes received OSC back to the asset capture endpoint, where the metadata is input to a second instance of IEM Room Encoder. The two IEM Room Encoder instances write parameter control envelopes to the DAW, refreshing OSC parameter control input at 20 ms intervals. This provides a reference signal (the automation for the first IEM Room Encoder instance, which receives OSC directly from the bridge application), and the test signal (the automation for the second IEM Room Encoder instance, which receives OSC via the loopback over the 5G network).

TABLE II
SUMMARY OF RESULTS FROM PRELIMINARY TESTING.

Nsamp. (reference)	Nsamp. (loopback)	Loss ratio	MSE
2,506	1,207	0.5184	0.000006

To analyse the reference and tests signals the source x position envelopes from the reference and loopback signal input to IEM Room Encoder instances is exported from Reaper using a custom Lua script which writes the envelopes to JSON files. This was developed using the ReaScript API²⁷ and the RXI JSON library²⁸. A custom python script was then used to analyse the reference and loopback envelopes. This script returns the tracking metadata loss ratio, Mean Squared Error (MSE) between the envelopes, and windowed MSE across the trace. tracking metadata loss ratio is calculated by comparing the number of breakpoints in the reference envelope to the number of breakpoints in the loopback signal envelope. As the reference and loopback envelopes may be of different lengths the two envelopes are re-sampled to an interpolated even length of 18,000 samples using numpy²⁹ (see figure 3 and figure 4). MSE is then calculated across the resampled envelopes. To provide an indication of variance in signal quality over time, windowed MSE is also calculated. Here the resampled envelopes are windowed across the trace using a 200 ms window size. MSE is then calculated for the windowed envelopes, and associated with the window centre time (see figure 5). Results of tracking metadata loss ratio and MSE measurement across the complete trace are reported in table II.

²⁷www.reaper.fm/sdk/reascript/reascripthelp.html

²⁸[www.github.com/rxi/json.lua](https://github.com/rxi/json.lua)

²⁹<https://numpy.org/doc/stable/reference/generated/numpy.interp.html>

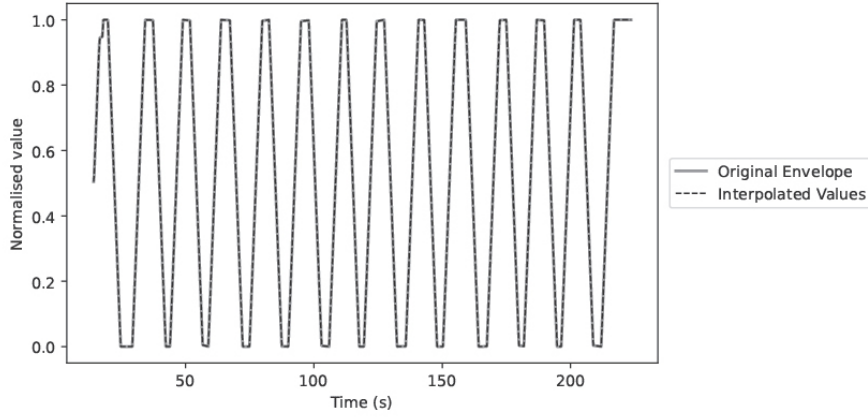


Fig. 3. Plot of the reference envelope recorded for preliminary testing. The original envelope is shown in grey as a solid line, and the resampled envelope used to calculate MSE metrics is shown in black as a dashed line. The value of x position samples is shown as a normalised value.

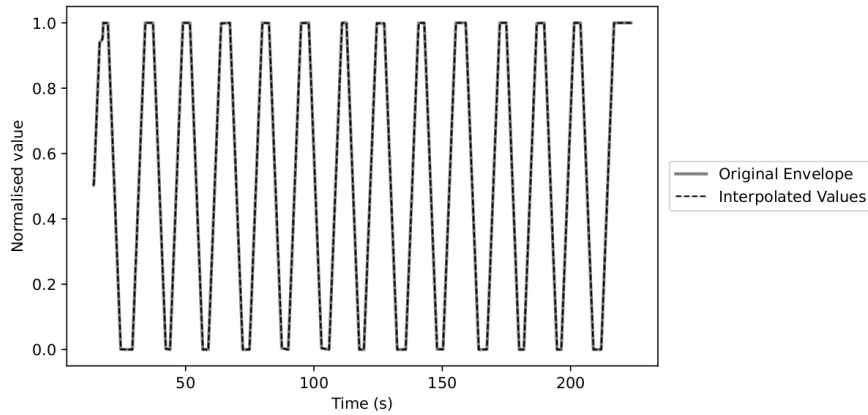


Fig. 4. Plot of the loopback envelope recorded for preliminary testing. The original envelope is shown in grey as a solid line, and the resampled envelope used to calculate MSE metrics is shown in black as a dashed line. The value of x position samples is shown as a normalised value.

VI. DISCUSSION

Most notably a tracking metadata loss ratio of 0.5184 was measured between the reference and loopback envelopes. This indicates that over half of tracking metadata samples sent via the loopback are being lost. Packet loss may be a result of interference in the wireless connection to the 5G router, rather than transport across the 5G network itself. Such interference can cause packet drops and significant jitter (in the order of tens of milliseconds). Latency and jitter across the wireless connection to the router, and across the 5G network itself, may also result in tracking metadata loss due to the refresh rate of IEM Room Encoder instances. Here packets routed via the loopback endpoint may not arrive within the 20 ms refresh cycles due to network jitter. This highlights a significant issue in 5G deployment of metadata transport systems in VP. Whereas 5G networks provide optimal QoS for wireless transport of media and metadata, the lack of native 5G capability with VP technologies (such as object tracking and spatial audio systems) requires that wireless connection to a 5G router is used to connect to 5G networks.

This connection may suffer significant packet loss and jitter depending on signal strength and potential interference. The potential for latency and jitter to cause late packet arrival with respect to spatial audio system refresh rate highlights the need for accommodation of some jitter and latency across 5G networks in tracking metadata streaming. Implementation of jitter buffering may be required to avoid this form of tracking metadata loss.

Notably, however, the MSE measured between the reference and loopback envelopes is miniscule. Observing the windowed MSE (see figure 5) it can be noted that MSE increases during periods of movement, and is not constant across the signal, however remains so small across the trace. These MSE measurements suggest an unnoticeable difference between the reference and loopback envelopes despite significant tracking metadata loss. This is reasonable to attribute to OSC parameter control in the spatial audio system. Here, the spatial audio system was configured to ramp between parameter values input from sequential OSC messages.

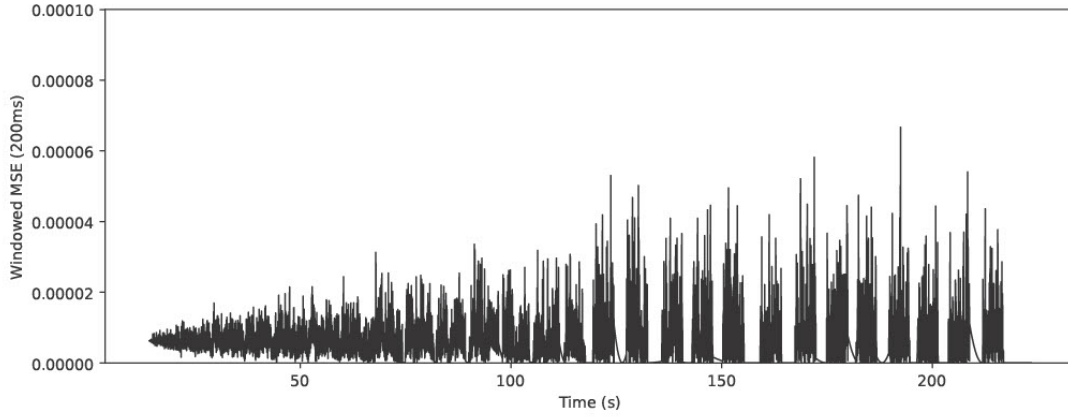


Fig. 5. Overview of preliminary test deployment .

Though a tracking metadata sample may be missing in the test signal, the same signal ramp may be followed as the reference signal, simply with less breakpoints in the ramp.

This would indicate that linear interpolation alone is likely sufficient with respect to Packet Loss Concealment (PLC) for tracking metadata transport to spatial audio systems over lossy network connections. These results suggest that a key issue in the deployment of VP systems over 5G networks is the lack of 5G native hardware.

This lack of native 5G capability for tracking and spatial audio OSC systems which operate on best-effort UDP is identified as requiring wireless connection to a router, where significant packet loss may occur. As such it can be suggested that some form of PLC is required when bridging object tracking and spatial audio systems over 5G networks. Measurement of MSE suggests that linear interpolation to replace missing tracking metadata samples will be sufficient to minimise signal degradation to negligible levels.

VII. CONCLUSION

This paper has presented the case of VP deployments over 5G networks, where object tracking endpoints communicate with spatial audio rendering endpoints via best-effort OSC. The need for OSC bridging to provide interoperability between object tracking and spatial audio systems is highlighted. Here, ADM OSC is highlighted as a suitable definition for interoperability, and the requirements of a general-purpose metadata bridging tool are contributed. A proof-of-concept implementation of a general-purpose bridging tool is contributed, and usage in an example 5G deployment is detailed. Preliminary testing of best-effort UDP for OSC transport over 5G networks is presented. Results indicate that the key obstacle in 5G deployment may be a lack of native 5G capability in hardware used in object tracking and spatial audio rendering, where wireless connection to a 5G router is required, which may result in significant packet loss.

Measurement of MSE, however, contributes suggestion that simple PLC methods such as linear interpolation may provide effective concealment of tracking metadata sample loss. Additionally, the potential for network latency and jitter to result in late packets is highlighted as possible to address by implementing jitter buffering.

The preliminary testing of best-effort OSC and the implemented bridging tool is, however, limited. Further work will require further development of the general-purpose bridging tool, including the implementation of the full set of features described in section III. Further work will also require the implementation of features required to record a comprehensive set of Key Performance Indicators (KPIs) for formal testing of the bridging tool and best-effort OSC methods. This includes timestamping and use of sequence indexing, required for latency measurement, packet ordering measurements, and packet loss measurements. Different PLC methods may be implemented directly in the bridging tool and evaluated for the case of 5G deployments. Comparison of the performance of different PLC methods for tracking metadata transport may also be a topic of further work. Further work may also consider evaluation for different audio system update rates, motion profiles, and network loads.

ACKNOWLEDGMENTS

This research is supported by the CoSTAR Screen Lab (AH/Y001044/1) and Studio Ulster.

REFERENCES

- [1] C. Rinaldi, F. Franchi, A. Marotta, F. Graziosi, and C. Centofanti, "On the Exploitation of 5G Multi-Access Edge Computing for Spatial Audio in Cultural Heritage Applications," *IEEE Access*, vol. 9, pp. 155 197–155 206, 2021.
- [2] L. Turchet, "Musical Metaverse: vision, opportunities, and challenges," *Personal and Ubiquitous Computing*, pp. 1–17, 2023.
- [3] L. Turchet, M. Lagrange, C. Rottondi, G. Fazekas, N. Peters, J. Østergaard, F. Font, T. Bäckström, and C. Fischione, "The Internet of Sounds: Convergent Trends, Insights and Future Directions," *IEEE Internet of Things Journal*, 2023.
- [4] L. Turchet and S. Krstulović, "DSP as a Service: Foundations and Directions," *IEEE Open Journal of the Communications Society*, vol. 5, pp. 6212–6226, 2024.

- [5] L. Turchet, C. Rinaldi, C. Centofanti, L. Vignati, and C. Rottondi, "5G-Enabled Internet of Musical Things Architectures for Remote Immersive Musical Practices," *IEEE Open Journal of the Communications Society*, pp. 1–1, 2024.
- [6] C. Rinaldi, C. Centofanti, L. Comanducci, M. Pezzoli, A. Bernardini, A. Marotta, and F. Antonacci, "Immersive Acoustics via Next-Generation Networks: Achieving High-Fidelity Audio in the Metaverse," *IEEE Open Journal of the Communications Society*, vol. 6, pp. 7899–7919, 2025.
- [7] ITU, "Recommendation ITU-R M.2150-3," Feb. 2026.
- [8] M. Wright, A. Freed, and A. Momeni, "2003: OpenSound Control: State of the Art 2003," in *A NIME Reader*, A. R. Jensenius and M. J. Lyons, Eds. Cham: Springer International Publishing, 2017, vol. 3, pp. 125–145, series Title: Current Research in Systematic Musicology.
- [9] ITU, "Recommendation ITU-R BS.2168-0 Audio definition model and serial representation of audio definition model profile for advanced sound systems emission," Feb. 2025.
- [10] A. Carot, C. Hoene, H. Busse, and C. Kuhr, "Results of the Fast-Music Project—Five Contributions to the Domain of Distributed Music," *IEEE Access*, vol. 8, pp. 47 925–47 951, 2020.
- [11] A. Carot, F. Sardis, M. Dohler, S. Saunders, N. Uniyal, and R. Cornock, "Creation of a Hyper-Realistic Remote Music Session with Professional Musicians and Public Audiences Using 5G Commodity Hardware," in *2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. London, UK: IEEE, Jul. 2020, pp. 1–6.
- [12] J. Dürre, N. Werner, S. Hämäläinen, O. Lindfors, J. Koistinen, M. Saarenmaa, and R. Hupke, "In-depth latency and reliability analysis of a networked music performance over public 5G infrastructure," in *Audio Engineering Society Convention 153*. Audio Engineering Society, 2022.
- [13] F. Martusciello, C. Centofanti, C. Rinaldi, and A. Marotta, "Edge-Enabled Spatial Audio Service: Implementation and Performance Analysis on a MEC 5G Infrastructure," in *2023 4th International Symposium on the Internet of Sounds*. Pisa, Italy: IEEE, Oct. 2023, pp. 1–8.
- [14] L. Turchet and P. Casari, "Assessing a Private 5G SA and a Public 5G NSA Architecture for Networked Music Performances," in *2023 4th International Symposium on the Internet of Sounds*. Pisa, Italy: IEEE, Oct. 2023, pp. 1–6.
- [15] —, "Performance Analysis of Slicing on a 10-node 5G Architecture for Networked Music Performances," in *2024 IEEE Symposium on Computers and Communications (ISCC)*. Paris, France: IEEE, Jun. 2024, pp. 1–6.
- [16] P. Cairns, H. Daffern, and G. Kearney, "Investigation of Server-Based Spatial Audio for Metaverse Concert Distribution," in *2024 IEEE 5th International Symposium on the Internet of Sounds (IS2)*. Erlangen, Germany: IEEE, Sep. 2024, pp. 1–8.
- [17] L. Vignati, G. Nardini, M. Centenaro, P. Casari, S. Lagén, B. Bojovic, S. Zambon, and L. Turchet, "Is Music in the Air? Evaluating 4G and 5G Support for the Internet of Musical Things," *IEEE Access*, vol. 12, pp. 38 081–38 101, 2024.
- [18] M. B. Khan, A. A. Cheema, A. Rahimian, G. Dunleavy, D. Charles, D. Finlay, F. Lyons, and D. Keeney, "5g-enabled next generation remote multi-site virtual production," *IEEE Network*, pp. 1–9, 2026.
- [19] H. Lee, "Multichannel 3D Microphone Arrays: A Review," *Journal of the Audio Engineering Society*, vol. 69, no. 1/2, pp. 5–26, Feb. 2021.
- [20] ITU, "ITU-R BS.2159-9 Multichannel sound technology in home and broadcasting applications," Mar. 2022.
- [21] M. Frank and A. Sontacchi, "Case Study on Ambisonics for Multi-Venue and Multi-Target Concerts and Broadcasts," *Journal of the Audio Engineering Society*, vol. 65, no. 9, pp. 749–756, Sep. 2017.
- [22] M. Tomasetti, A. Farina, and L. Turchet, "Latency of spatial audio plugins: a comparative study," in *2023 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*. Bologna, Italy: IEEE, Sep. 2023, pp. 1–10.
- [23] P. Müller, R. Haefeli, J. Schütt, and M. Ziegler, "Telemersive Audio Systems for Online Jamming," in *2024 IEEE 5th International Symposium on the Internet of Sounds (IS2)*. Erlangen, Germany: IEEE, Sep. 2024, pp. 1–15.
- [24] ITU, "Recommendation ITU-R BS.2076-3 Audio Definition Model," Feb. 2025.
- [25] C. Rottondi, C. Chafe, C. Allocchio, and A. Sarti, "An Overview on Networked Music Performance Technologies," *IEEE Access*, vol. 4, pp. 8823–8843, 2016.