ADM-OSC: an industry initiative for communicating object-based audio data

Preprint · September 2023

Michael Zbyszyński
L-Acoustics

Guillaume Le Nost
L-Isa

ADM-OSC: an industry initiative for communicating object-based audio data

Michael Zbyszyński
Creative Software
L-Acoustics
London, UK
michael.zbyszynski@l-acoustics.com
0000-0003-3771-0869

Hervé Déjardin
Audio Innovation Department
Radio France
Paris, France
herve.dejardin@radiofrance.com

David Marston
Research & Development
BBC
London, UK
david.marston@bbc.co.uk

Guillaume Le Nost
Creative Software
L-Acoustics
London, UK
guillaume.lenost@l-acoustics.com

***Abstract*—Spatial and immersive audio has become increasingly mainstream, presented in concert halls and more recently (e.g. Apple: June 2021) through music streaming services. There is a diverse ecosystem of hardware and software controllers and renderers in both live and studio settings that would benefit from a standardized communication protocol. Since 2021 a growing group of industry stakeholders has been working to develop ADM-OSC to fill this need. ADM-OSC proposes a standard for transmitting position data for object-based audio by implementing a namespace in parallel with the Audio Definition Model (ADM), a metadata standard developed in the broadcast industry. Open Sound Control (OSC) is a well-established data transport protocol developed for flexible and accurate communication of real-time performance data. By leveraging these open standards, we have created a lightweight specification that can be easily implemented in audio software, plugins, game engines, consoles, and controllers.**

This paper will discuss the design principals of ADM-OSC, developed to meet the needs of specific stakeholders and use cases. The core address space for position data is described, including a solution for representing both physics-based and cinematic distance. Proposed extensions to the next version of the specification will add comprehensive support for additional broadcast use cases, including different types of audio content (e.g., different languages or assistive descriptions), user interactivity, and cues for reducing channel count. We conclude with an overview of future ADM-OSC development, including next steps in bringing together ideas and discussion from multiple industry partners.

***Keywords*—object-based audio, immersive audio, transmission protocol, Audio Definition Model, Standard, Audio Metadata**

I. Introduction

The Audio Definition Model (ADM) was first published by the European Broadcast Union (EBU) in 2015 as a standard representation of audio metadata. [1] The goal of ADM is to support a broad range of Next Generation Audio (NGA) use cases that include spatial and immersive audio, as well as interactive personalization and accessibility features. ADM can be used to represent channel-based, scene-based, and object-based audio. It is defined by the EBU in ITU-R BS.2076 [2]. One typical representation of ADM is as an XML chunk in a RIFF/WAV file. This archival format is able to represent multiple "programmes" with diverse audio content. For streaming purposes, there is a serialized representation of ADM (S-ADM), as described in ITU-R BS.2125 [3]. This can be synchronized with audio signals in SMPTE ST 2116. [4]

ADM offers a flexible, open standard for representing audio metadata and it can be streamed for broadcast use via S-ADM. However S-ADM is still a relatively complex format, both in terms of the amount of data and the technical challenge of implementation. Live production ecosystems which have more limited bandwidth and computing power require a simpler protocol and implementation to enable real-time communication of immersive audio metadata between software and hardware controllers.

Open Sound Control (OSC) [5], [6] is a data transport specification for real-time message communication among applications and hardware that is commonly used in musical (and non-musical) applications. There are multiple, open-source implementations that simplify adoption for developers. The OSC specification does not define specific messages; these vary from according to the implementation. There are currently many audio devices implementing immersive data in OSC, but there is not necessarily agreement on how to represent similar data. ADM-OSC proposes a standardized message address space for transmitting position data for object-based audio, enabling interoperability between devices from different manufacturers. It is not a complete re-implementation of ADM in OSC, but a subset of position and control messages that can be used on their own or translated into ADM messages for streaming or storage (see sections III-B and V-D below).

Although ADM supports channel-based and scene-based audio, ADM-OSC implements exclusively object-based [7] metadata. Object-based representation encodes audio tracks along with positional and other data about how that audio should be reproduced, or rendered, during playback. Positional data is speaker-agnostic, allowing object-based mixes to be highly portable. A musician might audition a mix on headphones using a binaural renderer then perform at a venue with dozens of loudspeakers using a spatial renderer. That mix might then be rendered for streaming with a third renderer. To realize this workflow, interoperability is crucial. Hardware controllers and consoles, DAWs and audio plugins, and audio renders need to follow a standard representation of position, even if diffusion to loudspeakers is realized differently depending on the technology. ADM-OSC extends the ADM standard for positional data further into live and studio settings.

This paper will describe the ADM-OSC initiative and detail real-world use cases from Radio France and the BBC. The current, 0.4 specification of the ADM-OSC address space develops organically out of requirements of these cases and is described in the next section. Possible future development is then outlined in the following areas:

- Extent
- Distance
- Touch messages
- Conversion to S-ADM
- Program configuration
- Grouping

Finally, the paper concludes with a discussion of the organization of ADM-OSC stakeholders and how we can continue to develop an open-source standard that is simple to understand and implement, but comprehensive enough for the widest range of use cases.

II. ADM-OSC INITIATIVE

The ADM-OSC initiative started as a collaboration between L-Acoustics, Radio France, and Flux:: as a way to generalize interoperability between object editors for live sound performances. Radio France needed to drive two different object renderers, in real time, from a single object editor. A speaker-based renderer would drive the loudspeaker system for immersive shows, while a binaural renderer would allow a real-time radio broadcast on the French Radio Network (see section III-A). All spatial mixing decisions were done by the live mixing engineer during the performance.

Once the first proof of concept was achieved, this initiative was presented to the EBU and to other industry partners to gather interest. The key arguments for implementation of ADM-OSC over S-ADM for live sound were the following:

- OSC is already implemented (or at least familiar) in many live performance devices: mixing consoles, live renderers, show control systems, lighting systems, video systems, game engines.
- OSC has many open-source implementations.
- ADM-OSC can be transmitted aside from audio streams. It allows for separate streams and more modular audio system designs.
- ADM-OSC provides a more lightweight solution than S-ADM if devices are only interested in object positions.
- ADM-OSC implementation can follow more closely the needs and capabilities of a specific device as opposed to a full S-ADM implementation.
- Once the streams of audio and metadata become more unidirectional (towards a broadcast encoder, for example), ADM-OSC and audio streams can be converted into S-ADM by a specific device (see section V-D).

Today, project contributors include BBC, Dolby, d&b Audiotechnik, DiGiCo, Lawo, Magix, Merging Technologies, Meyer Sound, Sound Particles, and Steinberg. Regular online meetings are organized to discuss proposals and next steps (see section V).

III. USE CASES

ADM-OSC has been designed to solve real problems for live and broadcast sound producers. The section describes situations where Radio France and BBC have used the standard to expedite and enable object-based mixes for live and broadcast consumption simultaneously, using hardware and software from multiple stakeholders.

A. Radio France – FIP 360

Fig. 1. Irène Drésel “Live 360°” at Le Petit Palais, Paris, December 2022

Radio France used ADM-OSC in the production of the FIP 360 podcast. [8] to control multiple renderers. In episode #9

[9], Irène Drésel performed live in Le Petit Palais, Paris with electronic audio and video projection. L'Acoustics L-ISA [10] was used to route audio to a circular speaker array for the live performance what ADM-OSC data from L-ISA was sent to a Flux:: [11] binaural renderer for streaming as a podcast. ADM-OSC was also used for video mapping during the performance. Broadcast images were generated in real time by a Modulo Kinetic [12] server and their movement in space was controlled by ADM-OSC data.

The use of L-ISA in this context is the result of earlier experiments on tour with Molécule [13]. ADM-OSC's interoperability and compatibility with multiple types of spatial renderers enabled producers to create multiple spatial mixes with one set of controls.

B. BBC – Eurovision

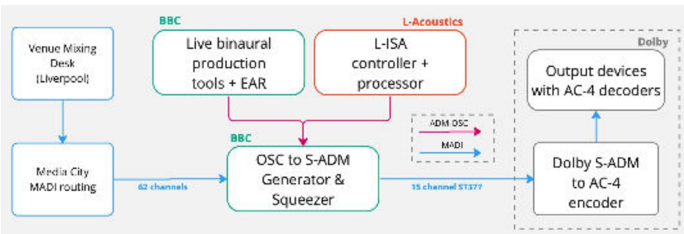

Fig. 2. Audio and metadata flow for Eurovision test.

In May 2023, the BBC conducted an audio delivery experiment using content from the Eurovision Song Contest. [14] The purpose of this test was to create S-ADM data live to accompany audio and video content which could be streamed to an NGA encoder outputting AC-4. 62-channels of audio was fed from the performance venue in Liverpool to the BBC's immersive listening room in Salford. Audio was distributed to the 32 speakers in the listening room using L-Acoustics L-ISA Studio. L-ISA also sent ADM-OSC data to a custom OSC to S-ADM converter developed by the BBC. Data from L-ISA, BBC's Live Binaural Production Tool, or EBU's EAR Production Tools [15] were used to generate a 15-channel SMPTE-encoded stream that could then be converted to a broadcast format. An Ateme Titan Live encoder converted the ST337 input into a Dolby AC-4 bitstream, which was then delivered in a DVB-T transport stream for over-the-air TV reception. It was also sent via a DASH server (Dynamic Adaptive Streaming over HTTP) for internet-based delivery to both a smartphone and to a television supporting HbbTV (Hybrid broadcast broadband TV) in the lab.

The tests demonstrated multiple formats and points of interoperability, and showed that an ADM-OSC to S-ADM workflow can be implemented in live production.

IV. ADM-OSC v0.4

A. Current ADM-OSC implementations

ADM-OSC is currently implemented in the following products:

- SPAT Revolution (FLUX::SE)
- L-ISA Controller (L-Acoustics)
- Ovation (Merging Technologies)
- Nuendo (Steinberg)
- SpaceMap Go (Meyer Sound)
- QLAB 5 (Figure 53)
- Space Controller (Sound Particles)
- Modulo Kinetic (Modulo Pi)
- Iosono (Barco)

The definition of implementation is currently under discussion, since many products don't need to implement every message in the schema. One proposal is for a "core" group of addresses that any implementation should handle, along with one or more extension groups that could be implemented, depending on use case. (see section VI below)

The most recent version of ADM-OSC specifies the addresses listed in table I. In that table, *f* stands for floating point values, *i* signifies integers, and *k* is the object number. So, the message `/adm/obj/4/azm 34.5` would specify that object number four should be panned to 34.5 degrees to the left of center. Messages can be translated directly into sub-elements of an ADM audioBlockFormat tag.

TABLE I
ADM-OSC v0.4
AUDIOBLOCKFORMAT SUB-ELEMENTS FOR OBJECTS

subelem	attr	unit	min	max	OSC
POLAR					
position	azimuth	degrees[1]	-180	180	/adm/obj/*k*/azim *f*
position	elevation	degrees[1]	-90	90	/adm/obj/*k*/elev *f*
position	distance	norm	0	1	/adm/obj/*k*/dist *f*
position	aed	see above			/adm/obj/*k*/aed *f f f*
CARTESIAN					
position	X	norm	-1	1	/adm/obj/*k*/x *f*
position	Y	norm	-1	1	/adm/obj/*k*/y *f*
position	Z	norm	-1	1	/adm/obj/*k*/z *f*
position	XYZ	norm	-1	1	/adm/obj/*k*/xyz *f f f*
COORDINATE SYSTEM					
cartesian		boolean	0	1	/adm/config/ obj/*k*/cartesian *i*
GAIN					
gain		linear	0	1	/adm/obj/*k*/gain *f*

[1]-90.0 degrees is on the right, 0.0 is in front.
[2]90.0 degrees above the listener.

For the coordinate system, the `/config` part of the OSC address is used to help optimize parsing on the server side. Any `/adm/config` messages would be "low frequency" compared to `/adm/obj` that could be parsed with higher time accuracy.

V. NEXT STEPS

A. Extent

There is currently a proposal for messages describing the size of an object, outlined in table II.

Currently, different renderers implement some or all size dimensions, and these dimensions might vary with distance. There is also less agreement about how size could be realize perceptually. Because of interpretation of extent parameters

TABLE II
ADM-OSC EXTENT

subelem	unit	min	max OSC	
POLAR				
width	degrees	0	180	/adm/obj/*k*/width *f*
height	degrees	0	180	undefined
depth	ratio	0	1	undefined
CARTESIAN				
width	norm	0	1	/adm/obj/*k*/w *f*
height	norm	0	1	/adm/obj/*k*/h *f*
depth	norm	0	1	/adm/obj/*k*/d *f*

varies substantially between different renderers, it is not clear whether they are a good candidate for standardized messages.

B. Distance

1) Context: The 3D paradigm chosen by the ADM standard is a normalised (dimensionless) reference volume, defined in Cartesian or spherical coordinates:

- dimensionless ADM cube:
  $-1 \leq x \leq 1; -1 \leq y \leq 1; -1 \leq z \leq 1$
- dimensionless ADM sphere:
  $0 \leq d \leq 1$ (d=distance or radius)

This paradigm is used by studio/broadcast mixing tools such as Dolby Atmos or MPEG-H.

On the other hand, some audio renderers represent a physics-based world, and the notion of source distance relates to a physical unit, such as meters. Aside from direct sound gain, the source physics-based distance d_m relates to advanced audio object parameters such as propagation delay, air attenuation, and energy levels of early/cluster reflections and late reverberation, or sound field behaviors (plane vs spherical waves).

These audio renderers include L-Acoustics L-ISA, Flux:Spat, d&B Soundscape, but also, more generally in the AR/VR domain, game audio engines such as Unreal, Unity, Wwise, or XR audio engines such as Magic Leap Soundfield Audio. A common challenge for all these renderers based on physical distance is that if the gain follows physical attenuation laws (such as "-6dB per doubling of distance"), there are some singularities when d_m gets close to 0. Hence, most of these renderers include a "volume of reference" or "unit volume" where the rendering (and in particular the gain) do not follow physically-informed laws anymore. This is true for Unreal and Spat, for example.

2) Proposal: An object position in a physics-based world can be described as:

- Cartesian:
  $xMin \leq x_m \leq xMax$ (meters)
  $yMin \leq y_m \leq yMax$
  $zMin \leq z_m \leq zMax$
- spherical:
  $0 \leq d_m \leq dMax$ (meters)

$dMax$ corresponds to the `absoluteDistance` parameter in an ADM audioPackFormat element.

$dRef$ is a new parameter can be defined as the radius in meters of a volume of reference, which would serve the two purposes. It coincides with the dimensionless volume used in the ADM standard and it is used by physically informed renderers as the "volume of reference" where the laws of physics do not apply, and the gain(dB) is constant regardless of distance.

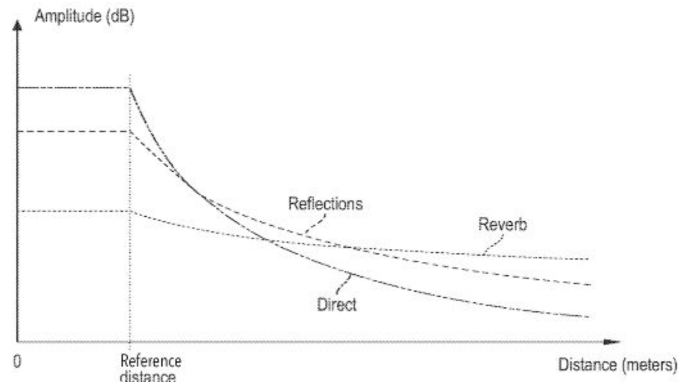

Fig. 3. (from Jot et. al.) [16] The volume of reference is defined by $d_m < dRef$ and shows an amplitude behavior (dB) independent from the distance.

By definition: $0 \leq \frac{dRef}{dMax} \leq 1$
and the following cases arise:

- $\frac{dRef}{dMax} = 1$: the world is a dimensionless reference volume, matching the ADM standard
- $\frac{dRef}{dMax} = 0$: no reference volume within the physical world

These two values could be represented in ADM-OSC as follows. The messages are optional, with a default value of $dMax = 1$. We propose that referenceDistance could be added to a future version or the ADM specification.

TABLE III
ADM-OSC DISTANCE PROPOSAL

Sub-element	Units	Min	Max	OSC
distance	norm	0	1	/adm/obj/*k*/dRef *f*
distance	meters	0	n/a	/adm/obj/*k*/dMax *f*

C. Touch messages

Partners at Flux::, Merging Technologies, and Steinberg have implemented messages for touch control. These messages are proposed for the next version of ADM-OSC. These messages are notable because they do not directly translate to an existing ADM value. The message syntax involves the initiation of a `touch` message, updates to parameters, and then finally a `release` message:

D. Conversion for Broadcast

The workflow for an ADM broadcast is complex and mostly beyond the scope of this paper. One key area where

TABLE IV
ADM-OSC TOUCH

ADM-OSC address	message
/adm/obj/*k*/xyz	"touch" (string)
/adm/obj/*k*/xyz	*float float float*
/adm/obj/*k*/xyz	"release"

ADM-OSC becomes part of the broadcasting chain is in the generation of S-ADM metadata.

ADM-OSC is not intended to be directly converted to S-ADM, as ADM-OSC only covers a sub-set of the S-ADM features. However, ADM-OSC messages can be used to drive values in S-ADM parameters, in particular positional metadata values for object-based audio channels.

A S-ADM generator will generate a series of S-ADM frames containing ADM based on a predefined template. For example, the ADM program names, and content and object elements would have been defined in the template. The template will leave some values undefined (or set to default values), that will allow them to be populated by values from the ADM-OSC messages. As the ADM-OSC messages will be generated in real-time from the production tool, this will allow the S-ADM metadata to be updated in real-time from these messages.

E. Program configuration

Partners at Dolby have proposed an expansion of the namespace to help configure broadcast program. The whole proposal is online [17]. Some interesting challenges in the development of this namespace can be highlighted.

Position messages are typically sent at a high frequency over UDP. This is a typical OSC implementation. These messages are intended to be acted upon in real-time as soon as possible. The density of the stream means that packet loss is not a concern; it is assumed that late (or dropped) packets have been superseded and can be ignored. *Configuration* messages are the opposite. Setting the dialog content of a track, for example, might only happen once. That message can be handled relatively slowly, but it cannot be dropped. Part of the proposal is to use an `/adm/config/` naming convention, so real-time messages could be directed to a low-latency receiver while configuration messages could be handled differently. Handling `/config` messages over TCP/IP would also increase reliability.

The other challenge this proposal addresses is to reduce the amount of repeated data being sent. An ADM file with a 7.1.4 bed would contain data for the speaker positions of all of the speaker channels. Dolby's proposal is to implement common layout codes so standard bed layouts could be specified with one message. One step further is the proposal for templates. Many broadcast programs have the same basic layout. Sending a large amount of boilerplate metadata over OSC would be much less efficient than specifying a predefined template. Compared to the current ADM-OSC specification, templates would require much more shared data and coordination between the sender and the receiver. For such a critical setup, some kind of acknowledgment message would be appropriate.

F. Grouping

Another broadcast use case involves reducing the number of audio channels. A live object-based mix might have a very high number of channels, while broadcast streams are more limited. During the conversion process, channel count is reduced by combining or discarding objects. Messages have been proposed for grouping objects and specifying their importance, to help inform the reduction process.

VI. DISCUSSION

Organization and planning of the ADM-OSC initiative takes place on Github [18] and via online stakeholder meetings. Those wishing to be included in future meetings should contact the authors directly via email. The goal of the group is to develop the standard where it will be useful, while maintaining simplicity for new implementations. As can be seen from section V, there are multiple, non-overlapping use cases under consideration. Because a particular implementation might only be relevant to some of the possible uses for ADM-OSC, it is not expected that every possible message can be sent or received. This makes it difficult to define what it means to be "ADM-OSC compliant."

The current proposal is to organize a growing namespace into subsections, each with a working group relevant stakeholders in charge of that section. Possible subsections include:

- **position**—the current, 0.4 spec
- **control**—touch messages (see V-C) and triggering of scenes/snapshots/program changes. Since these messages do not translate literally into ADM tags, it is under consideration whether they should be in the `/adm` namespace
- **broadcast**—messages, such as interactivity, that are primarily or exclusively used in a broadcast context. For example, in the Eurovision trials (section III-B) there were multiple channels of commentary. These were marked in the S-ADM stream to allow viewers at home to select which one they were hearing.
- **tracking**—both head tracking and performer tracking are relevant to object-based audio. This is another area that might not directly map to ADM tags.
- **two-way communication**—there is a possible need for acknowledgment of important messages, or discovery of implementation details.

This list is a reasonably comprehensive summary of areas that have been discussed at recent meetings. But there is also a strong push to limit the scope of the standard, so it is unlikely that everything on the list will be implemented. Only messages that reach a minimum standard of utility would be recommended for future ADM-OSC specifications.

ADM-OSC is currently a set of recommendations; there is no certification body. Interoperability with existing implementations would be a primary motivation for any new implementation, so existing implementations define the standard. In addition, there are testing and validation tools available online.

The ADM-OSC group is currently working to develop and streamline documentation so that new ADM-OSC developers deploy the standard more easily.

ACKNOWLEDGMENT

The authors thank current and future ADM-OSC contributors, in particular James Cowdery and Tim McNamara at Dolby for their proposal summarized in section V-E and Gael Martinet from Flux:: for initial development and the test tools.

REFERENCES

[1] S. Füg, D. Marston, and S. Norcross, "The audio definition model—a flexible standardized representation for next generation audio content in broadcasting and beyond," in *Audio Engineering Society Convention 141*. Geneva, Switzerland: Audio Engineering Society, 2016.

[2] Recommendation ITU-R BS.2076, "Audio definition model," International Telecommunications Union, Geneva, Switzerland, Standard, 2019. [Online]. Available: https://www.itu.int/rec/R-REC-BS.2076

[3] Recommendation ITU-R BS.2125, "A serial representation of the Audio Definition Model," International Telecommunications Union, Geneva, Switzerland, Standard, 2019. [Online]. Available: https://www.itu.int/rec/R-REC-BS.2125

[4] M. Bolt, J. Smith, J. Roden, B. Leteneur, T. Habann, J. Hanschke, T. McNamara, S. Norcross, and K. Terry, "Practical implementation of new open standards for Next Generation Audio production and interchange," European Broadcasting Union, Geneva, Switzerland, Tech. Rep., 2021.

[5] M. Wright, "Open Sound Control: an enabling technology for musical networking," *Organised Sound*, vol. 10, no. 3, pp. 193–200, 2005.

[6] A. Schmeder, A. Freed, and D. Wessel, "Best practices for Open Sound Control," in *Linux Audio Conference*, vol. 10, 2010.

[7] N. Tsingos, "Object-based audio," in *Immersive Sound*. Routledge, 2017, pp. 244–275.

[8] [Online]. Available: https://www.radiofrance.fr/fip/podcasts/fip-360

[9] [Online]. Available: https://www.radiofrance.fr/fip/podcasts/fip-360/fip-360-9-irene-dresel-en-son-3d-3030805

[10] [Online]. Available: https://l-isa.l-acoustics.com/

[11] [Online]. Available: https://www.flux.audio/

[12] [Online]. Available: https://www.modulo-pi.com/media-servers/modulo-kinetic/

[13] [Online]. Available: https://l-isa.l-acoustics.com/projects/molecule/

[14] M. Firth and D. Marston. (2023) Live Next Generation Audio trial at Eurovision 2023. [Online]. Available: https://rdweb1.lh.bbc.co.uk/rd/blog/2023-06-eurovision-next-generation-audio

[15] [Online]. Available: https://ear-production-suite.ebu.io

[16] J.-M. Jot, R. Audfray, M. Hertensteiner, and B. Schmidt, "Rendering spatial sound for interoperable experiences in the audio metaverse," in *2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*. IEEE, 2021, pp. 1–15.

[17] [Online]. Available: https://immersive-audio-live.github.io/ADM-OSC/html/adm_spec_0.5_draft.html

[18] [Online]. Available: https://immersive-audio-live.github.io/ADM-OSC/