STADIUM
TECH REPORT

Keeping IPTV Audio in the IP Stack

A technical case for native multicast audio ingestion in professional venue infrastructure

Robare Pruyn
Principal Engineer, AmpThink
March 2026

Executive Summary

For years, venue technology teams have treated IPTV and distributed audio as parallel but fundamentally separate disciplines. The practical consequence is a signal chain that introduces unnecessary hardware, unnecessary costs, and unnecessary failure points between the audio source and the speaker.

This paper describes an alternative approach that is now proven and in production across professional sports venues in North America ranging from 12,000 to 70,000 seats. That approach is: by receiving IPTV audio natively from multicast MPEG transport streams directly into the venue's IP-based audio DSP infrastructure, integrators can eliminate dedicated media player decoders, analog audio distribution equipment, and the signal conversion hardware that connects them. The audio signal remains digital and network-native from headend to speaker, with the only digital-to-analog conversion occurring at the amplified loudspeaker endpoint - where it was always going to occur regardless of signal chain architecture.

This paper makes three claims and supports each with field evidence. First, the native multicast audio ingestion approach has been demonstrated at full professional sports scale, including venues exceeding 150 premium hospitality spaces. Second, the primary technical concern this approach raises – that of latency and lip synchronization – has a specific, well-characterized solution that has been implemented successfully across every deployment in a portfolio of venues that served as the case studies for this white paper. Third, the cost and complexity reductions are real and directly measurable, and are the result of removing an entire hardware class from the signal chain.

One architectural prerequisite applies: a converged network infrastructure. Venues where IPTV and audio transport already share a common IP backbone can implement this approach today with no additional infrastructure investment. This paper also addresses a codec compatibility constraint that arises from proprietary audio encoding in some IPTV headend platforms, describes a validated middleware solution, and provides specification language suitable for inclusion in a basis-of-design document.

Table of Contents

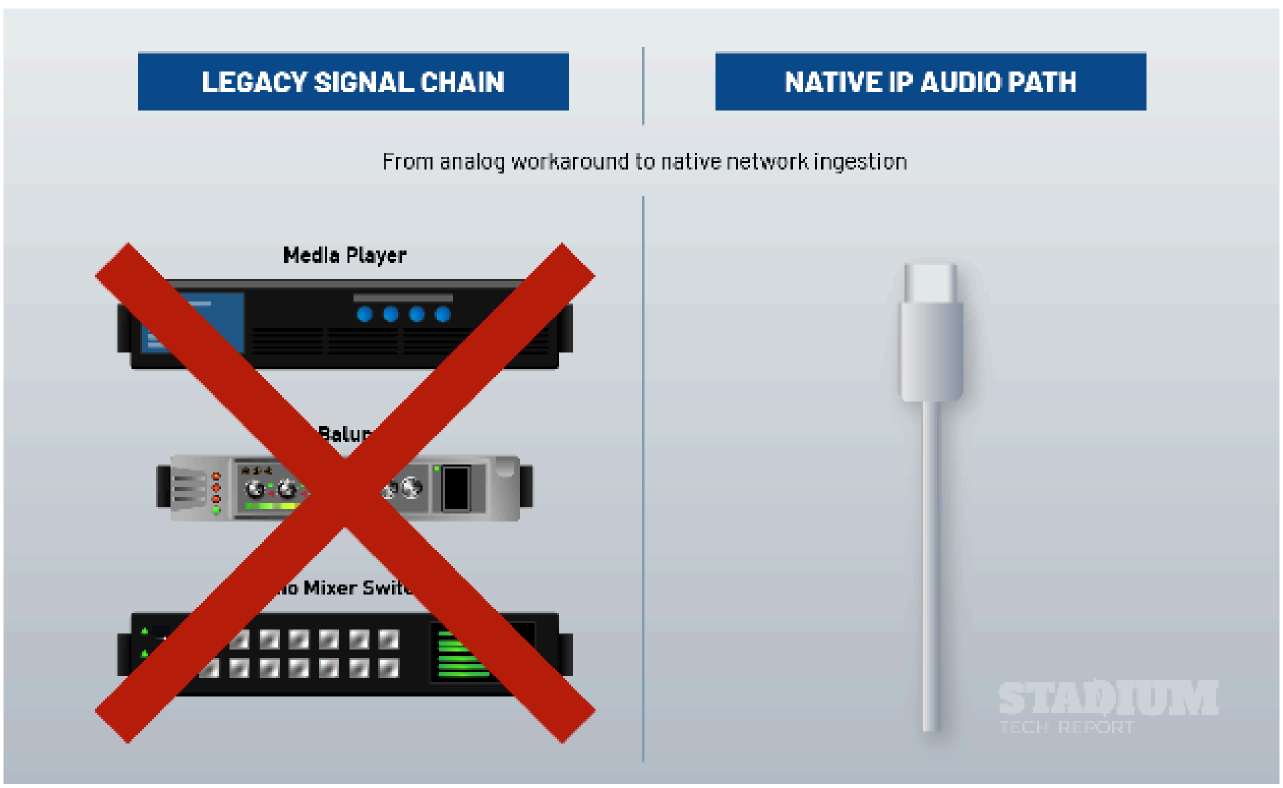
The Problem Hiding in the Signal Chain

IPTV distribution in large venues is, at this point, a solved problem. Encoders ingest source feeds, encapsulate video and audio into MPEG transport streams, and distribute them as IP multicast across the venue network. Endpoint decoders subscribe to the relevant multicast group, decode the stream, and drive displays. The protocols are mature, the components are well understood, and dozens of vendors compete in the space. Getting live television to hundreds of screens in a stadium is a known quantity.

What the industry was not asking, for a surprisingly long time, was why the audio component of that stream has to leave the IP network before it can reach a speaker.

In a conventional venue IPTV deployment, the signal chain for delivering television audio to an overhead speaker system works as follows: The IPTV headend encodes video and audio together into an MPEG transport stream, typically carrying the audio payload as Dolby AC-3, MPEG-2 Layer II, or AAC depending on the source and headend platform. That stream travels across the venue network as IP multicast. At each premium space (e.g. suite, club, restaurant, back-of-house area, etc.) a dedicated media player decodes the stream and produces two outputs: video to feed displays via HDMI, and audio from an analog line-level output, typically a 3.5mm or RCA jack.

That analog audio output then begins a second journey. It passes through a balun to convert the unbalanced signal for structured cabling transport. It arrives at an analog input on a matrix switcher or audio DSP. It is then digitized, processed, and re-encapsulated onto the IP network (often as Dante or AES67) for delivery to networked amplifiers and speakers.



Read that signal path again. Audio that originated as a digital stream on an IP network was decoded to analog at a media player, transported as analog through a balun, re-digitized at a matrix switcher or DSP input, and placed back on an IP network for final delivery. Three unnecessary signal conversions and at least three discrete hardware components. Per space. In every space in the building.

We are taking digitally network-distributed audio, converting it to analog, ingesting it as analog, reconverting it to digital IP, and sending it over IP back to a speaker. We are spending time and money to recreate an analog cable path inside a digital infrastructure.

At a venue with 20 premium suites and two club lounges, that means a rack of media players and a rack of analog audio distribution equipment that exist solely to bridge two IP-based systems that already share the same network. At a 70,000-seat venue with 150-plus suites and dozens of clubs and restaurants, the hardware count, the rack space, the cabling, the commissioning time, and the ongoing maintenance burden compound accordingly.

The question is why this pattern persisted as long as it did. The honest answer is that this was never an engineering decision. It was institutional momentum. The AV integration industry developed its practices during the analog era, and those practices were carried forward into IP infrastructure largely unexamined. Specifications called for dedicated media players because specifications had always called for dedicated media players. Nobody had documented the alternative at scale, and the specification process is, rightly, conservative about adopting approaches that lack field evidence. This paper exists to provide that evidence.

The Native IP Approach

The foundational observation is straightforward. If audio is already present on the venue network as a component of an IPTV multicast transport stream, and the audio DSP system serving the venue's speakers is also on that network, then no boundary needs to exist between them.

Modern IP-based audio DSP platforms include processing blocks that can subscribe directly to an MPEG transport stream multicast group on the network and extract the audio payload. The DSP tunes to the multicast group address, demultiplexes the transport stream, decodes the audio elementary stream (AC-3, AAC, or MPEG-2 Layer II depending on the source encoding), and presents it as a routable input within the DSP's audio matrix. No media player. No analog output stage. No balun. No external analog-to-digital conversion. The audio enters the DSP's processing environment as a digital signal extracted directly from the network.

The most widely deployed example of this capability in the professional sports venue space is with QSC Q-SYS, though the principle applies to any DSP platform with native multicast stream reception. The technical description that follows is generalizable, not vendor-specific.

From that point forward, the IPTV audio source behaves identically to any other input in the audio system. It can be routed to any output zone, mixed with other sources, processed with equalization and dynamics, and delivered over Dante, AES67, or any other networked audio transport to amplified loudspeakers. The only digital-to-analog conversion in the entire signal chain occurs at the speaker's integral amplifier - which is where it was always going to occur regardless of the upstream architecture.

The Native IP approach is compatible with any IPTV headend platform that produces standards-compliant MPEG transport stream multicast output. Platforms in common use in professional venue deployments include Cisco (now Wipro) VisionEDGE, VITEC, TriplePlay and Exterity, among others. The DSP does not care which headend produced the stream. It subscribes to a multicast group and decodes a standard container format. That decoupling is an architectural feature, not a limitation; it means the IPTV platform and the audio platform can be specified, procured, and evolved independently.

Implementation at Scale

We've discussed that audio stream consumption is generally standards-driven and vendor agnostic. The next question becomes: "How does a centralized solution like this scale?" The answer comes in the form of two primary implementation patterns, differentiated by venue scale.

This has been deployed across two primary implementation patterns, differentiated by venue scale. The underlying principle is identical in both cases; only the stream management strategy differs.

Per-Space Stream Model (Small to Mid-Scale Venues)

In a venue with a manageable number of premium spaces - on the order of 20 to 40 suites and a small number of club environments - the implementation creates one dedicated audio stream per controllable space. Each suite or club receives its own multicast audio stream configured to align with the speaker zones in that room. Channel selection is handled via API calls from the venue's room control system into the IPTV headend's media streaming infrastructure, using the same control plane that manages IPTV channel selection on the displays and audio volume in the room. When a guest or operator changes the television channel in a suite, the audio stream feeding that suite's speakers changes in lockstep.

The first venue to validate this model was a 12,000-seat professional soccer stadium built as a public-private partnership under significant budget pressure. The original specification called for dedicated media players and analog audio distribution equipment in every premium space. The project team, recognizing that the venue was being built on a converged network infrastructure, asked whether the network already in place could do the work that dedicated hardware was being specified to do. It could. The result was a system that was not only less expensive but measurably more flexible and simpler to operate than the original specification would have delivered. What had been specified as multiple equipment racks per floor reduced to a single Category 6A cable handoff per space, with all audio routing handled in the DSP software.

Bulk Ingestion Model (Large-Scale Venues)

At venues with a large number of premium spaces - more than 100 suites, dozens of clubs, restaurants, and back-of-house areas - per-space stream provisioning becomes operationally cumbersome. The bulk ingestion model addresses this by pulling the entire multicast output of the IPTV headend into the audio DSP system simultaneously. Every channel the IPTV platform produces, the audio system receives. Routing from channel to speaker zone is then handled entirely within the DSP's audio matrix, synchronized to the IPTV channel assignments on the displays in each space.

Because audio is not a high-bandwidth application relative to video, even a full channel lineup of decoded audio streams represents modest network throughput. Hundreds of audio channels can be extracted from their MPEG transport stream sources and made available in the DSP routing matrix over a standard gigabit Ethernet link without congestion.

The definitive proof point for this model is a 70,000-seat professional football stadium, one of the most technically sophisticated venues in North America, where the bulk ingestion architecture is in production across more than 150 premium suites and dozens of additional club and hospitality spaces. This deployment resolved the scale question definitively: the bulk ingestion approach holds at maximum professional sports venue scale without requiring architectural modification.

Latency, Lip Synchronization, and Clock Domains

The first technical objection the native multicast audio ingestion approach encounters, reliably and in every design review where it is proposed, is latency and lip synchronization. If video and audio are taking different paths through the system, how do they stay in sync at the point of delivery?

It is a legitimate concern, and it deserves a precise answer.

Why Audio Arrives First

The IPTV video path includes an inherent buffering stage that the audio path does not. An IPTV endpoint decoder must receive, buffer, and reassemble video frames before it can begin playback. Depending on the encoder's Group of Pictures (GOP) structure, the codec profile, and the decoder's implementation, this introduces a decode latency that typically ranges from one to several video frames - roughly 30 to 150 milliseconds in practice, varying by platform and configuration.

Audio extracted directly from the MPEG transport stream by the DSP has no equivalent buffering requirement. The DSP's multicast receiver demultiplexes and decodes the audio elementary stream with minimal latency - on the order of single-digit milliseconds for the decode operation itself. The result is that audio arrives at the speaker path with less total latency than video arrives at the display, not more.

The Practical Solution: Fixed Delay Offset

The synchronization solution is to add a small, fixed delay to the audio path within the DSP to align it with the video decode latency of the IPTV endpoints in the room. This is a standard DSP operation: a configurable delay block inserted in the audio signal chain, requiring no additional hardware.

In a well-designed venue network, where the multicast distribution topology is consistent and IPTV endpoints are typically one or two network hops from the distribution source, the video decode latency is predictable and consistent across the venue. A single delay offset value, determined during commissioning by measuring the actual video decode latency of the deployed IPTV endpoints, is generally sufficient to achieve acceptable lip synchronization across all spaces. For venues or individual spaces requiring finer adjustment, per-source delay offsets are easily configured within the DSP.

Perceptual Threshold and Engineering Margin

The International Telecommunication Union (ITU-R BT.1359-1) establishes that audio-to-video synchronization errors below approximately 45 milliseconds for audio leading video, and below approximately 125 milliseconds for audio lagging video, are generally imperceptible to viewers. The more conservative broadcast production guideline, often cited as ±20 milliseconds for professional monitoring environments, applies to production control rooms rather than hospitality playback. In a venue premium space, the practical perceptual threshold provides a meaningful engineering margin.

Across every deployment in the deployments cited in this paper, lip synchronization performance using the native IP approach has been equivalent to or better than what the legacy analog handoff approach delivers. This is not a theoretical claim. It is a measured outcome across venues of varying scale.

A Note on Clock Domains

A question that arises in technically rigorous design reviews concerns the relationship between the MPEG transport stream's timing domain and the audio network's PTP (IEEE 1588) clock domain. In a Dante-based audio infrastructure, for example, all audio devices synchronize to a PTP grandmaster clock to maintain sample-accurate synchronization across the network. The IPTV multicast stream originates outside that clock domain; it carries its own Program Clock Reference (PCR) embedded in the transport stream.

In practice, this clock domain boundary is handled transparently by the DSP's multicast receiver. The receiver decodes the audio payload, buffers it into the DSP's internal processing environment, and presents it to the audio matrix synchronized to the DSP's own clock. The audio enters the Dante (or AES67) clock domain at the point of ingestion into the DSP, and from that point forward it is transported with the same sample-accurate synchronization as any other source in the system. The clock domain transition is a non-issue in practice, but it is the kind of question a qualified reviewer will ask, and the designer should be prepared to answer it.

The Converged Network Prerequisite

Native IP ingestion has one non-negotiable infrastructure requirement: a converged network. The IPTV system and the audio system must share a common IP transport layer with multicast routing capability between them. In a venue where IPTV and audio operate on physically separate networks, this approach cannot be implemented without first bridging or consolidating those networks.

A converged network infrastructure provides this capability inherently. In practical terms, that means a single managed IP backbone supporting all venue subsystems: Wi-Fi, wired access, IPTV, audio transport, building management, security, and broadcast production. When all IP-based systems share a common transport layer, a multicast stream produced by the IPTV headend is a network resource available to any system with a connection and an IGMP subscription. The silos that traditionally separated IPTV from audio simply do not exist.

This is not a minor architectural point. It is the enabling condition. The legacy approach exists specifically to bridge two systems that were deployed on separate infrastructure. The entire hardware class this paper proposes to eliminate is compensating for a network architecture problem. In a converged infrastructure, the problem does not exist, and the compensating hardware serves no purpose.

Venues being designed and built today are increasingly adopting converged network infrastructure as the baseline architecture. For these venues, native IPTV audio ingestion is available from day one with zero incremental infrastructure cost. For existing venues operating on legacy parallel-network architectures, the converged network investment must come first - but that investment delivers value across every venue subsystem, not just audio. The native IPTV audio capability is one of many returns on that foundational investment.

Economic and Operational Impact

The cost reduction native IP ingestion delivers is not a matter of optimization or marginal savings. It is the consequence of removing an entire hardware class from the signal chain. Every component in the legacy approach that exists solely to extract audio from the IPTV stream and deliver it to the DSP as an analog input is eliminated: the media player, the balun, the analog matrix switcher input, the associated cabling, and the rack space.

The following table illustrates the per-space component comparison between the legacy approach and native IP audio ingestion.

Per-Suite Component	Legacy Approach	Native IP Approach
Dedicated IPTV media player / decoder	1 per source	Eliminated
Analog audio balun (Cat-to-XLR or RCA)	1 per source	Eliminated
Analog matrix switcher input	1 per space	Eliminated
Additional rack unit(s) per space	Yes	None
Analog cabling from player to DSP input	Per source run	Eliminated
DSP multicast stream receiver (config only)	N/A	1 per space (software)
Digital-to-analog conversions before speaker	3 (min)	1 (at speaker)
Commissioning: per-space hardware setup	Required	Eliminated
Failure points in audio signal chain	Multiple per space	Network + speaker

Table 1. Per-space component comparison: legacy analog handoff versus native IP audio ingestion.

Multiply the legacy column by every premium space in the building. At a venue with 150 suites, the hardware elimination is not incremental - it is a category removal from the bill of materials, the rack elevation drawings, the commissioning schedule, and the ongoing maintenance inventory.

Beyond installed cost, the operational benefits compound over the life of the system. Fewer hardware components means fewer failure points, fewer firmware update cycles, fewer devices drawing power and generating heat in equipment rooms, and less time spent troubleshooting a signal chain with multiple analog conversion stages. Commissioning time decreases because per-space hardware setup and calibration is replaced by DSP software configuration that can be templated and replicated across spaces.

Downstream and Unintended Benefits

Control System Simplification

The most significant downstream benefit of native IP ingestion, and one that was not fully anticipated in the initial deployments, was the effect on the room control system. In a legacy architecture, managing IPTV audio from a room control interface requires the control system to bridge two separate subsystems: the IPTV platform (for channel selection) and the analog audio distribution path (for source routing and volume). That bridge introduces additional control integration complexity, additional API endpoints, and additional failure modes.

With native IP ingestion, IPTV audio sources appear as standard inputs in the audio DSP's routing matrix. The control system already has established communication pathways into the DSP for managing speaker zones, volume levels, and source selection. Adding IPTV channel control requires only pointing the audio matrix to a different input using the same API calls already in use for every other audio function in the room. The control integration that would have required a separate bridge in the legacy architecture collapses into a configuration change within an existing control pathway.

Architectural Scalability

The native ingestion model has proven more scalable than the initial deployments suggested. As the portfolio has grown from the first implementation at a 12,000-seat venue to deployments exceeding 70,000 seats, the core architecture has held without fundamental modification. The two implementation patterns - per-space streams for smaller venues, bulk ingestion for larger ones - cover the full range of professional sports venue scale using the same underlying approach.

Future-Proofing and Upgrade Path

A system built on native IP audio transport is positioned to benefit from continued evolution in IP audio standards, DSP processing capability, and network infrastructure performance. As audio-over-IP protocols mature and DSP platforms gain processing headroom, the native ingestion model inherits those improvements automatically.

A system built on analog handoffs has no equivalent upgrade path. Its performance ceiling is fixed at the capabilities of the analog hardware in the rack. Replacing that hardware does not improve the architecture; it replaces components within an architecture that was suboptimal to begin with.

Field Evidence

This approach is deployed and operating across a growing portfolio of professional sports venues in North America. The following deployments represent the range of validated implementation contexts.

12,000-seat professional soccer stadium, new construction. Built as a public-private partnership under budget constraints, this venue is where the native IP audio approach was first developed and validated in production. The project team eliminated multiple racks of dedicated media players and analog audio distribution equipment from the original specification, replacing them with DSP-native multicast stream reception over the venue's converged network. This deployment established the viability of the per-space stream model and demonstrated that the approach produces a system that is simpler, less expensive, and more operationally flexible than the conventional alternative.

70,000-seat professional football stadium, new construction. One of the most technically sophisticated venues in North America, this stadium is the definitive answer to the scale objection. The bulk ingestion model is in production across more than 150 premium suites and dozens of club, restaurant, and hospitality spaces. The deployment confirmed that the approach holds at maximum professional sports venue scale with no architectural modification required.

Additional deployments. Beyond these two anchor implementations, the native IP audio ingestion approach is in active production at multiple additional professional sports venues ranging from approximately 20,000 to 60,000 seats. These deployments span arena and stadium configurations, new construction and retrofit projects, and a variety of ownership and operational models. Each has validated the core approach in a different context, and collectively they establish a body of field evidence that supports specification with confidence.

Codec Compatibility and Licensing Considerations

The native IP audio ingestion approach described in this paper depends on the audio DSP platform's ability to decode the audio codec carried within the IPTV headend's MPEG transport stream output. This is not a theoretical concern. It is a concrete specification dependency that has produced real integration challenges in the field, and a designer who does not account for it will discover the problem at commissioning time.

The Codec Landscape in Venue IPTV

IPTV headend platforms used in professional venue deployments encode audio using one or more of three primary codecs: Dolby AC-3, AAC, and MPEG-2 Layer II (MP2). The choice of codec is determined by the headend platform, the source feed characteristics, and in some cases the encoder configuration options available to the integrator.

Dolby AC-3 is the dominant audio codec in North American broadcast television and is consequently the default audio format on many IPTV headend platforms that ingest off-air or cable/satellite source feeds. AC-3 is a proprietary codec. Its use, including decode, requires a license from Dolby Laboratories. That license carries per-unit or per-channel costs that vary by platform and deployment context.

AAC and MPEG-2 Layer II are open-standard codecs that do not carry equivalent proprietary licensing requirements for decode. Some headend platforms offer these as configurable output options; others do not.

The DSP Codec Support Gap

This is where specification meets field reality. QSC Q-SYS is the most widely deployed audio DSP platform in the professional sports venue market and the platform with the most mature native multicast MPEG-TS reception capability. It does not support AC-3 decode in its media stream receiver and has indicated no roadmap to add it. The Q-SYS MPEG-TS receiver supports AAC and MPEG-2 Layer II decode. It does not support Dolby AC-3.

This creates a direct specification dependency between the IPTV headend's audio output codec and the DSP's decode capability. If the headend produces AC-3 audio and the DSP cannot decode AC-3, the native ingestion approach as described in Section 2 of this paper does not work without an intermediary.

Some headend platforms provide configuration options that resolve this at the source. If the headend can be configured to encode or transcode its audio output to AAC or MP2, the DSP receives a codec it can decode natively and no intermediary is required. The designer should evaluate headend codec output options during the specification phase and select a configuration that aligns with the DSP's decode capability where possible.

However, not all headend platforms offer this flexibility. VITEC's current-generation IPTV encoders, for example, produce Dolby AC-3 audio on their MPEG transport stream outputs with no open-codec alternative at the transport stream level. VITEC's migration path beyond MPEG-TS goes directly to SMPTE ST 2110 with AES67 audio - which bypasses the MPEG transport stream model entirely and delivers uncompressed audio natively on the IP audio network. That is a valid architectural path, but it represents a different generation of headend equipment and a different system design, not a configuration change to existing infrastructure.

Bridging the Gap: Software Middleware

For deployments where the headend produces AC-3 audio and cannot be reconfigured, and the DSP does not decode AC-3, the solution is a software middleware layer that sits between the two. This is a lightweight service, deployable as a virtual machine or container on existing venue compute infrastructure, that subscribes to the IPTV multicast streams, extracts the audio elementary stream, transcodes from AC-3 to an open codec (MP2 or AAC) or outputs directly to Dante/AES67, and makes the result available to the audio DSP as a network audio source.

This middleware approach preserves the core value proposition of the native IP architecture. The dedicated per-space media player hardware, the analog output stage, the baluns, and the analog distribution infrastructure are still eliminated. The audio still never leaves the IP domain. The middleware adds a software processing stage, but it runs on shared compute infrastructure rather than dedicated per-space hardware, and it scales across all spaces in the venue from a single service instance.

Purpose-built software for this function exists and is in production use at venues across the deployments cited in this paper. The implementation details are beyond the scope of this paper, but the architectural point is critical: the codec compatibility gap is a known constraint with a known solution, and a designer should account for it in both the specification and the cost model.

What the Designer Must Verify

Before specifying native IPTV audio ingestion, the systems designer must confirm the following codec chain:

1. What audio codec(s) does the IPTV headend produce on its MPEG transport stream output? If AC-3, can the headend be configured to produce AAC or MP2 as an alternative?
2. What audio codecs does the specified DSP platform's multicast stream receiver support for decode? Does that list include the codec the headend will produce?
3. If there is a mismatch, is a software middleware transcoding layer required? If so, what compute infrastructure will host it, and what is the additional latency budget?

4. Are there proprietary codec licensing costs (particularly Dolby) that apply to any component in the chain (headend, middleware, or DSP) and are those costs reflected in the project budget?

Failing to ask these questions during the design phase produces a specification that looks complete on paper but encounters a hard integration wall during commissioning. The native IP audio approach works. The codec compatibility chain must be validated as part of making it work.

Specification Language

For systems designers developing a basis-of-design specification for a venue with converged network infrastructure, the following language is offered as a starting point for specifying native IPTV audio ingestion. It has been updated from earlier versions to reflect the codec compatibility considerations described in Section 9.

The IPTV audio distribution system shall conform to the following requirements:

1. Signal Chain Architecture. The IPTV audio distribution system shall utilize native multicast MPEG transport stream reception to deliver audio from the IPTV headend to the venue's IP-based audio DSP infrastructure, eliminating dedicated media player decoders and analog audio distribution equipment from the IPTV-to-speaker signal chain.
2. Codec Decode Capability. The audio DSP platform shall be capable of subscribing to multicast groups carrying MPEG transport streams and decoding audio elementary streams encoded in AAC and/or MPEG-2 Layer II formats.
3. Codec Compatibility Middleware. Where the IPTV headend produces audio in a codec not natively supported by the DSP platform (e.g., Dolby AC-3), the system design shall include a software-based middleware transcoding service capable of receiving the IPTV multicast streams, extracting and transcoding the audio payload to a DSP-compatible codec or to Dante/AES67, and presenting the audio to the DSP as a network audio source. The middleware service shall be deployable on the venue's existing compute infrastructure.
4. Latency Alignment. Audio latency alignment with IPTV video endpoints shall be managed within the DSP environment using configurable per-source delay offsets calibrated to the measured video decode latency of the IPTV display endpoints deployed in each zone.
5. Network Infrastructure. The converged network infrastructure shall provide multicast routing capability between the IPTV distribution VLAN(s) and the audio network VLAN(s) with appropriate IGMP snooping and querier configuration.
6. Design-Phase Verification. The designer shall verify codec compatibility between the IPTV headend audio output and the DSP platform's decode capability during the design phase, and shall account for any proprietary codec licensing costs in the project budget.

This language is intentionally platform-neutral. It describes functional requirements rather than specifying particular vendor products. At the time of writing, QSC Q-SYS is the most widely deployed platform with native multicast MPEG-TS reception capability in the professional sports venue market, supporting AAC and MPEG-2 Layer II decode. The specification should be written to the functional requirement, allowing the market to compete on implementation while ensuring the codec compatibility chain is validated as a design-phase deliverable.

Conclusion

This paper opened with three claims. Each is supported by production deployments across professional sports venues of varying scale.

The approach works at scale. It has been implemented across venues ranging from 12,000 to 70,000 seats, including at maximum professional football scale with more than 150 premium spaces served by a single architectural model.

The latency and lip synchronization concern has a well-characterized, repeatable solution. Audio arrives at the DSP with less latency than video arrives at the display. A fixed delay offset in the DSP aligns the two. The solution has been implemented successfully in every deployment cited in this paper.

The cost and complexity reduction is real. Eliminating dedicated media players, analog audio distribution equipment, and baluns from the signal chain reduces installed cost, reduces commissioning time, reduces ongoing maintenance burden, and produces a system with fewer failure points and a cleaner upgrade path.

The prerequisite is a converged network, and the trend in new venue construction is toward convergence as the default architecture. For existing venues, the converged network investment delivers value across every building subsystem - audio is one of many beneficiaries. The codec compatibility constraint described in this paper is real and must be accounted for in the design phase, but it is a solved problem with documented solutions in production.

The equipment goes away. The audio stays on the network. The system is simpler, less expensive, more reliable, and better positioned for the future. The field evidence exists. The specification language is available. What remains is for the design community to examine the approach on its merits and decide whether the institutional momentum that sustained the legacy signal chain still serves the interests of the venues we are building.

About the Author

Robare Pruyn is a Principal Engineer specializing in converged network and media technology infrastructure for professional sports and entertainment venues. Over a career spanning more than sixteen years, he has led the design, deployment, and integration of large-scale IP network, IPTV, broadcast production, and AV control systems at venues ranging from major league stadiums to mid-scale amphitheaters and arenas across North America.

His work encompasses the full converged infrastructure stack: core network architecture, hyperconverged compute, IPTV distribution, real-time media transport, and control system automation. He is an active member of AVIXA, SMPTE, IEEE, and AES, and holds certifications from QSC and Audinate.

The approach described in this paper was developed through direct involvement in the design, deployment, and commissioning of the venue systems cited as field evidence.