



A evolução do áudio de próxima geração: Transporte de metadados MPEG-H áudio via S-ADM e SMPTE ST 2110-41

Por Gabriel Thomazini

Autor explica que a combinação do MPEG-H Audio com o S-ADM e o SMPTE ST 2110-41 estabelece o caminho para a transmissão IP de metadados em tempo real, tornando o áudio de próxima geração uma realidade prática nas produções ao vivo.

Metadados modulados em um canal de áudio

A indústria de broadcast passa por uma transformação fundamental, afastando-se de infraestruturas de ponto a ponto para a flexibilidade e escalabilidade das redes de Protocolo de Internet (IP). Esta transição não é meramente uma atualização tecnológica, mas uma reavaliação completa de como os sinais de mídia são transportados, processados e geridos. No centro desta evolução está a mudança da **Interface Digital Serial** (SDI), um padrão que serviu bem a indústria durante décadas, para o conjunto de padrões SMPTE ST 2110, que introduz uma arquitetura desagregada e baseada em essências.

O SDI tem sido a espinha dorsal das instalações de broadcast, fornecendo um método maduro e robusto para o transporte de vídeo digital, áudio e dados auxiliares sobre um único cabo coaxial. A sua principal característica definidora é a natureza "agrupada" do fluxo de transporte. Dentro de um único fluxo de dados SDI, os sinais de vídeo, múltiplos canais de áudio e pacotes de dados auxiliares (ANC) estão intrinsecamente entrelaçados numa estrutura temporal rígida. Embora este design garanta uma sincronização inerente e simplicidade para ligações ponto-a-ponto, impõe restrições significativas aos fluxos de produção modernos.

Sistemas como o MPEG-H dependem de metadados dinâmicos para permitir funcionalidades como a personalização por

parte do usuário e recursos de acessibilidade avançados. No paradigma SDI, a integração destes metadados é brilhantemente endereçada pelo **MPEG-H Audio** através da modulação das informações em uma pista de áudio, usando o "**Control Track**" (CT). Essa solução é bastante robusta, permitindo edições (cortes) e passar por reamostragem (**resampling**) sem corromper os metadados do MPEG-H. Porém, para modificar apenas os metadados (para alterar uma opção de idioma ou uma audiodescrição) um dispositivo de autoria deve processar todo o fluxo SDI para atualizar a informação e então modular novamente o CT para a substituição da informação anterior. Esta necessidade exemplifica a ineficiência inerente do modelo agrupado.

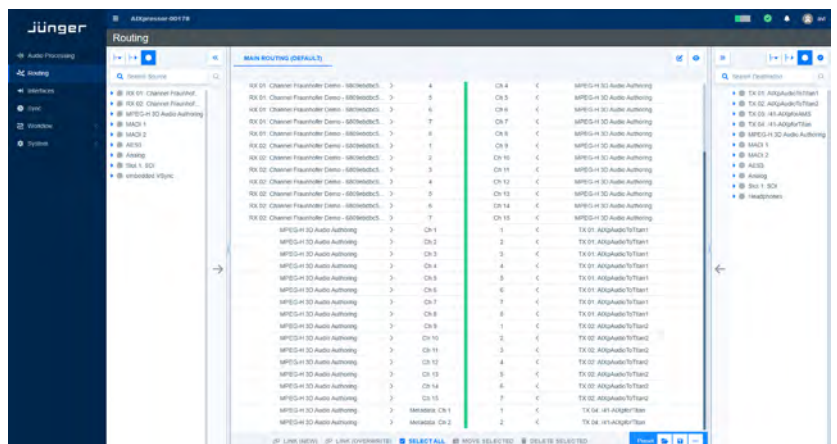


Gráfico 1: Reprodução

A revolução ST 2110: Essenciais para a máxima flexibilidade

O conjunto de padrões SMPTE ST 2110 representa uma ruptura radical com o paradigma SDI. Em vez de simplesmente encapsular o fluxo SDI numa rede IP — uma abordagem adotada pelo padrão de transição SMPTE 2022-6 — o ST 2110 desconstrói o sinal nos seus componentes fundamentais, conhecidos como “essências”. Cada essência é transportada como um fluxo IP separado e independente, permitindo uma flexibilidade sem precedentes no roteamento e processamento.

Esta arquitetura “separada” é definida por um conjunto de padrões principais, cada uma abordando uma essência específica:

- **ST 2110-10: Sincronização e Temporização do Sistema:** Este é o padrão fundamental que torna todo o sistema funcional. Estipula o uso do Protocolo de Tempo de Precisão (PTP), conforme definido na IEEE 1588, para sincronizar todos os dispositivos na rede com um relógio de referência comum (Grandmaster) com precisão de microssegundos. Este tempo partilhado permite que os fluxos das essências, que podem percorrer caminhos de rede diferentes, sejam perfeitamente realinhados.
- **ST 2110-20: Essência de Vídeo Não Comprimido:** Define o transporte de vídeo ativo não comprimido como um fluxo RTP (Protocolo de Transporte em Tempo Real).
- **ST 2110-30: Essência de Áudio PCM (Sem Compressão):** Especifica o transporte de áudio PCM, baseando-se no padrão de interoperabilidade de áudio sobre IP AES67.
- **ST 2110-40: Essência de Dados Auxiliares:** Define como os pacotes de dados ANC tradicionais (do mundo SDI) são transportados em um fluxo RTP separado.

A separação das essências é mais do que uma conveniência técnica; é uma mudança filosófica fundamental. Transforma a infraestrutura de broadcast de uma coleção de dispositivos de função fixa, ligados por cabos ponto-a-ponto, para uma arquitetura definida por software que opera sobre uma infraestrutura de TI padrão.

Num ambiente ST 2110, um processador de áudio pode subscrever e processar dezenas de fluxos de áudio sem nunca precisar de tocar ou ser sobrecarregado pelos fluxos de vídeo associados. Esta capacidade de encaminhar, processar e escalar cada essência de forma independente é o pilar que suporta fluxos avançados e ágeis, abrindo caminho para a integração nativa de sistemas NGA (*Next Generation Audio* – Áudio de Próxima Geração) complexos.

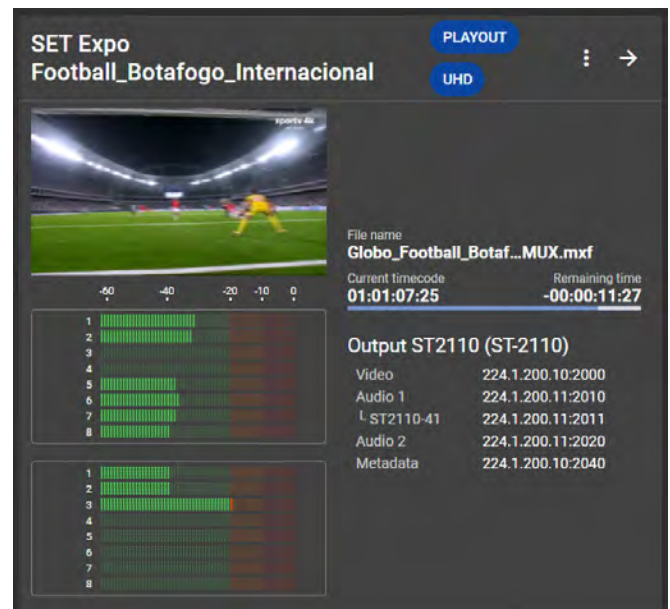


Imagem reprodução do SET EXPO 2025

A importância dos metadados: Por que é crítica uma nova camada de transporte?

À medida que o áudio evolui de um modelo baseado apenas em canais para um modelo baseado em canais, objetos e cenas, os metadados que o acompanham transformam-se. Deixam de ser meramente “auxiliares” — informações secundárias como rótulos ou simples descrições — e tornam-se um componente primário e essencial da experiência sonora. Os metadados em sistemas NGA como o MPEG-H definem a posição tridimensional dos objetos sonoros, as regras para a interação do usuário e as informações de renderização que adaptam a reprodução a diferentes sistemas de

reprodução. Sem estes metadados, a essência de áudio é incompleta e o seu potencial não pode ser completamente atingido.

A atual importância dos metadados expõe uma lacuna crítica no conjunto inicial de padrões ST 2110. Embora o ST 2110-40 forneça uma ponte para o passado ao transportar dados ANC do SDI, é inadequado para o futuro. A indústria precisava de um novo método de transporte: um que tratasse os metadados como uma essência de primeira classe, livre das restrições do SDI e totalmente sincronizado dentro do ecossistema ST 2110.

MPEG-H Audio: A anatomia de uma cena interativa e imersiva

Para compreender como transportar os metadados do MPEG-H é, primeiro, essencial dissecar a natureza desses metadados. O MPEG-H não é simplesmente um codec de áudio; é um sistema de áudio abrangente concebido para oferecer experiências sonoras imersivas e personalizadas. Desenvolvido pelo grupo de trabalho ISO/IEC *Moving Picture Experts Group* (MPEG) e padronizado como

ISO/IEC 23008-3, o sistema foi projetado desde o início para ir além do som surround tradicional, introduzindo uma nova dimensão de controle, tanto para o criador de conteúdo como para o consumidor final. A sua arquitetura é construída sobre uma base de metadados ricos que descrevem a cena de áudio, permitindo uma flexibilidade sem precedentes na entrega e reprodução.

Para além do som surround: Os conceitos centrais do MPEG-H

A força do MPEG-H reside na sua abordagem híbrida, que permite a representação simultânea de conteúdo de áudio em formatos distintos. Esta flexibilidade permite que os produtores de conteúdo utilizem a melhor ferramenta para cada elemento sonoro dentro de uma única produção:

- **Canais:** Este é o método tradicional de representação de áudio, onde cada sinal corresponde a uma posição de uma caixa de som específica numa configuração predefinida. O MPEG-H expande este conceito para suportar arranjos 3D imersivos, como 7.1 + 4H (sete canais surround, um subwoofer e quatro canais para representar o eixo vertical), fornecendo uma base sólida para camadas de som ambientes e paisagens sonoras.
- **Objetos:** Um objeto de áudio é um elemento sonoro individual, como uma linha de diálogo, um comentário específico ou um efeito sonoro distinto. Ao contrário dos canais, os objetos não estão ligados a uma posição fixa em uma caixa de som específica. Em vez disso, são acompanhados por metadados que definem a sua posição no espaço tridimensional, o seu ganho e outras propriedades. Esta posição pode

variar ao longo do tempo, permitindo movimentos dinâmicos. O decodificador no dispositivo de reprodução utiliza estes metadados para renderizar o objeto de forma inteligente na configuração de caixas de som disponíveis (ou fones de ouvido), garantindo uma localização espacial precisa.

A capacidade de combinar estes elementos dentro de um único *bitstream* confere ao MPEG-H um poder e flexibilidade extraordinários, permitindo a criação de cenas de áudio ricas e complexas.

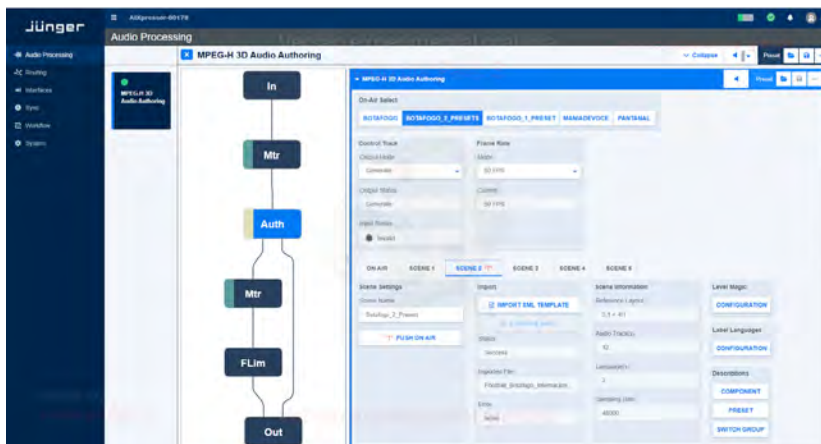


Gráfico 3: Reprodução

O poder dos metadados: Habilitando a personalização e a imersão

Os metadados são muito importantes no sistema MPEG-H; são eles que transformam uma coleção de sinais de áudio numa experiência interativa e adaptável. Estes metadados, criados durante o processo de autoria, contêm toda a informação necessária para que o decodificador reconstrua a intenção do criador e responda às escolhas do usuário. Os principais tipos de metadados incluem:

- **Metadados Posicionais:** Para cada objeto de áudio, estes metadados especificam as suas coordenadas tridimensionais (azimute, elevação, distância) ao

longo do tempo. Isto permite que os objetos se movam dinamicamente através do espaço sonoro.

- **Metadados de Interatividade:** Este é talvez o aspecto mais revolucionário do MPEG-H. Estes metadados definem as regras e os limites para a personalização do usuário, dando ao produtor de conteúdo total controle sobre as opções oferecidas. As principais funcionalidades interativas incluem:
 - **Melhoria de Diálogo:** Ao designar o diálogo como um objeto separado, os metadados permitem que os usuários ajustem o seu

volume e posição independentemente dos sons de fundo (cama), melhorando a inteligibilidade.

- **Seleção de Idioma/Comentarista:** Vários objetos de diálogo (por exemplo, comentários em diferentes idiomas) podem ser agrupados num “grupo de comutação” (*switch group*). Os metadados instruem o decodificador a permitir que o usuário selecione apenas um objeto deste grupo de cada vez, garantindo que apenas um idioma seja ouvido.
- **Predefinições (Presets):** As emissoras e produtores de conteúdo podem criar e nomear múltiplas cenas predefinidas. Uma predefinição é um conjunto de metadados que define os

níveis de ganho, posicionamento e o estado de ativação/desativação de vários objetos. Por exemplo, numa transmissão de futebol, podem ser oferecidas predefinições como “+Diálogo” (que aumenta o som do comentarista em relação a torcida) ou “Ambiente do Estádio” (que remove todos os comentários).

- **Metadados de Renderização:** Estes metadados fornecem ao decodificador a informação necessária para adaptar de forma inteligente um único bitstream a qualquer sistema de reprodução. Isto inclui informações sobre a configuração de original, permitindo readaptações inteligentes para sistemas estéreo ou 5.1, e dados para uma renderização binaural que recria a experiência imersiva em fones de ouvido comuns.



Fraunhofer no IBC 2025/Foto: Fernando Moura

O perfil ADM do MPEG-H: Uma ponte para padrões abertos

Embora o sistema MPEG-H Audio tenha uma estrutura de metadados interna rica e completa, a indústria de broadcast busca convergir para formatos de metadados abertos e padronizados para garantir a interoperabilidade. O principal padrão para este fim é o **Audio Definition Model (ADM)**, especificado na recomendação ITU-R BS.2076. O ADM fornece uma sintaxe baseada em XML para descrever cenas de áudio complexas, incluindo conteúdo baseado em canais, objetos e cenas.

Para preencher a lacuna entre o seu ecossistema nativo e o mundo mais limitado dos padrões abertos, os desenvolvedores do

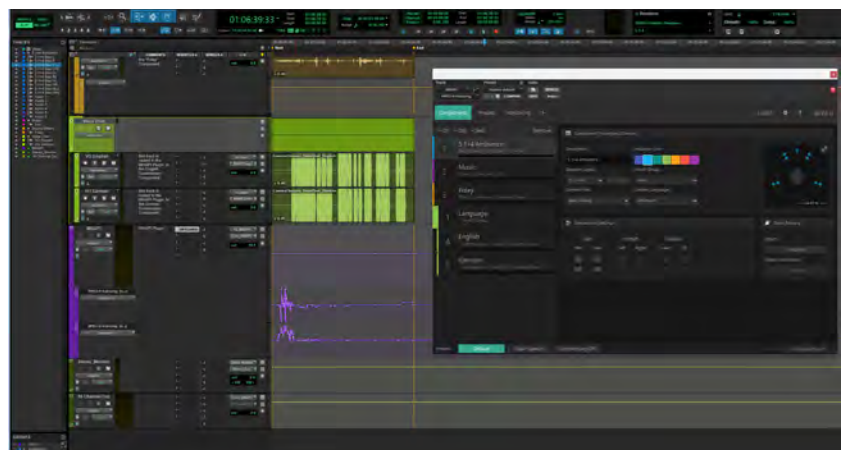


Gráfico 4: Reprodução

MPEG-H definiram o “Perfil ADM do MPEG-H”. Um perfil, neste contexto, é um conjunto de regras, restrições e mapeamentos específicos que ditam como a estrutura de metadados do MPEG-H deve ser representada dentro da sintaxe ADM padrão. Este perfil garante que arquivo ADM criado em conformidade pode ser utilizado por um codificador MPEG-H e interpretado sem ambiguidades, preservando toda a intenção artística e interativa original.

Esta abordagem representa uma decisão estratégica crucial. Permite que as emissoras e produtores de conteúdo construam os seus fluxos em torno do padrão

aberto ADM, promovendo a interoperabilidade dos seus “masters” de áudio num formato padronizado, com a confiança de que estes podem ser utilizados para gerar um bitstream MPEG-H sem perda de funcionalidade. Ferramentas de autoria como o **MPEG-H Authoring Tool (MHAT)**, ou **Digital Audio Workstations** compatíveis (Pro Tools, Nuendo, Davinci Resolve, etc) podem exportar diretamente um “Master MPEG-H” como um arquivo **Broadcast Wave Format (BWF)** que contém tanto o áudio não comprimido (PCM) como os metadados ADM em conformidade com este perfil.

S-ADM (ITU-R BS.2125): Serializando a cena de áudio para produção ao vivo

Com uma compreensão clara da estrutura de metadados do MPEG-H e da sua representação padronizada através do Perfil ADM, o próximo desafio é adaptar este modelo para os rigores da produção e transmissão ao vivo. O padrão ADM (ITU-R BS.2076) foi concebido principalmente para workflows baseados em arquivos, onde o documento XML que descreve a cena de áudio está disponível de uma só vez. Este formato é ideal para arquivamento, pós-produção e troca de conteúdo, mas é inerentemente inadequado para ambientes de tempo real, como a transmissão ao vivo, onde os metadados devem ser gerados e transmitidos de forma contínua e sincronizada com o áudio e o vídeo.

num fluxo contínuo adequado para aplicações lineares.

A principal inovação do S-ADM é a transição do conceito de um “arquivo” para o de um “fluxo”. Um fluxo S-ADM é uma sequência de frames de metadados com uma referência temporal. Cada quadro contém os metadados ADM relevantes para um bloco de áudio correspondente, permitindo que a descrição da cena de áudio seja transmitida de forma incremental e em tempo real. Este modelo de quadros é explicitamente concebido para ser agnóstico em relação ao transporte, o que significa que um fluxo S-ADM pode ser transportado através de uma variedade de interfaces físicas e de rede, incluindo AES3, SDI e, mais importante para esta discussão, redes IP.

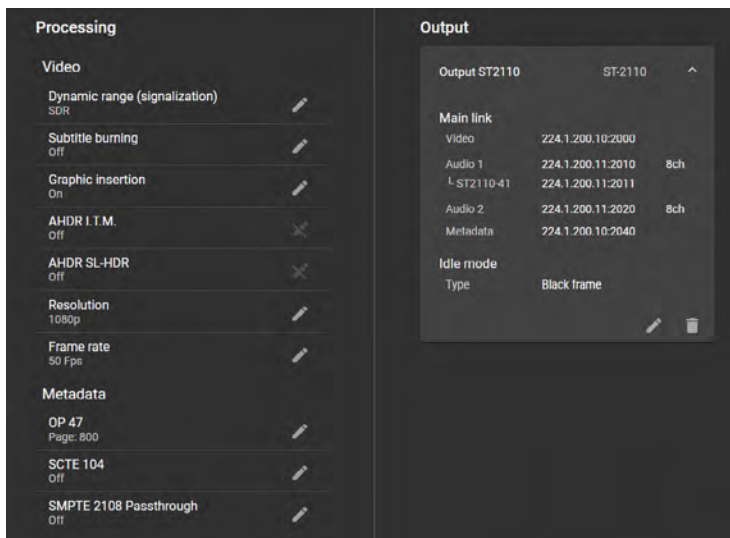


Gráfico 5: Reprodução

Para resolver esta incompatibilidade, a **International Telecommunication Union (ITU)** desenvolveu a recomendação ITU-R BS.2125, que define uma representação serial do **Audio Definition Model**, comumente conhecida como S-ADM. O S-ADM não é um novo formato de metadados, mas sim um método para segmentar e estruturar os metadados ADM existentes

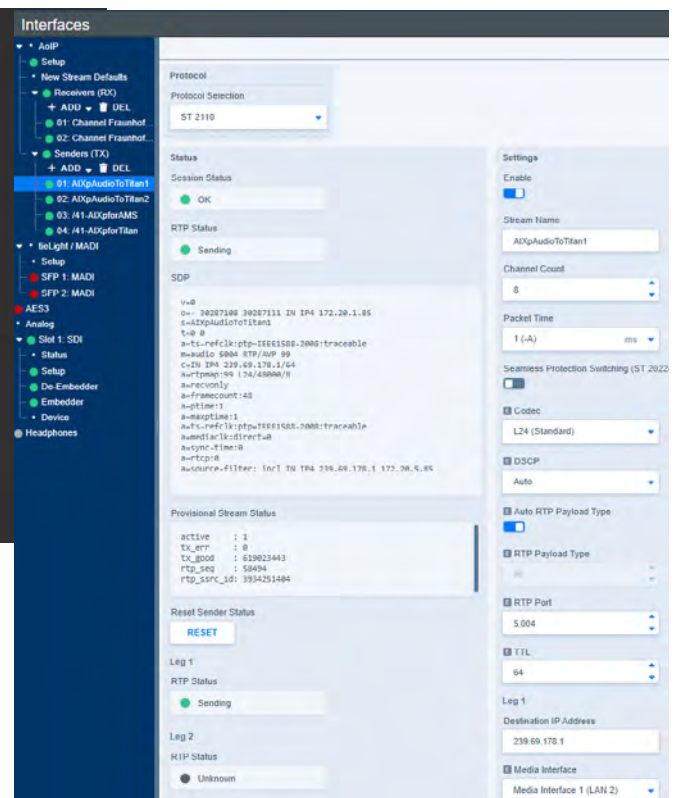


Gráfico 6: Reprodução

Ao serializar a estrutura ADM, o S-ADM permite que os fluxos de produção ao vivo aproveitem o poder do áudio baseado em objetos. Mudanças dinâmicas na mixagem – como um comentarista alternativo ou uma nova

predefinição ativada – podem ser refletidas em quadros S-ADM subsequentes e transmitidas instantaneamente, garantindo que os metadados permaneçam sempre perfeitamente sincronizados com a ação.

O padrão SMPTE ST 2110-41: A camada de transporte IP para metadados rápidos

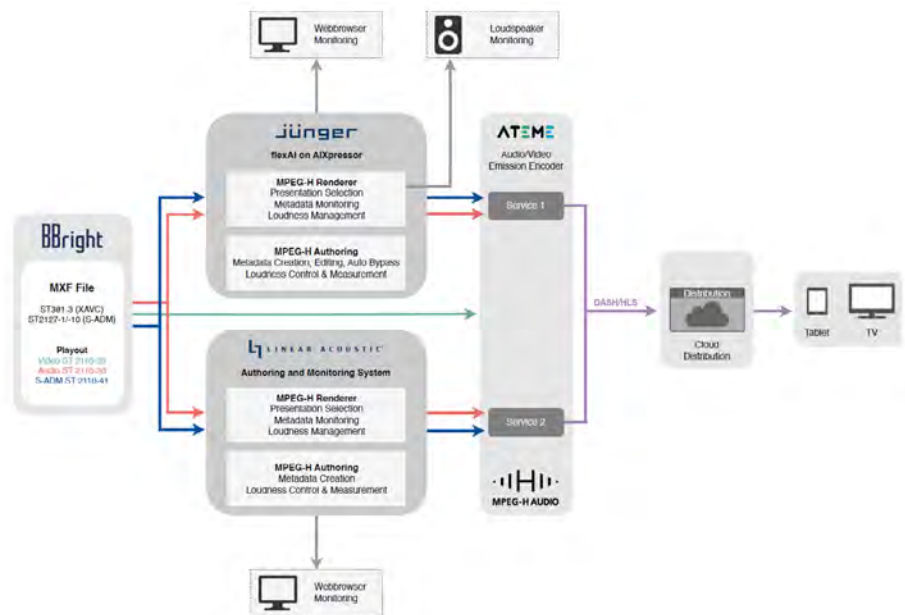
Com a cena MPEG-H descrita em ADM e serializada em um fluxo S-ADM, a peça final do quebra-cabeça é o mecanismo de transporte. O ecossistema SMPTE ST 2110 precisava de um “veículo” robusto e flexível para transportar este novo tipo de essência de metadados através da rede IP, de forma sincronizada com as essências de áudio e vídeo. A solução inicial, ST 2110-40, provou ser uma ponte útil para o passado, mas inadequada para o futuro. Isto levou ao desenvolvimento do ST 2110-41, um padrão projetado especificamente para as necessidades de metadados nativos de IP.

O SMPTE ST 2110-41 fornece uma estrutura de **payload** RTP genérica e flexível, projetada para transportar itens de dados que não têm origem no mundo SDI. As características chave que tornam o ST 2110-41 ideal para transportar S-ADM e outros metadados NGA são:

- **Payload Arbitrário:** Ao contrário do ST 2110-40, o -41 não impõe uma estrutura de pacote de legado. Ele pode transportar de forma eficiente vários formatos de dados, incluindo XML, JSON ou BLOBs binários, sem a necessidade de um invólucro desnecessário.
- **Sincronização PTP:** Um fluxo ST 2110-41 usa os mesmos “carimbos” de data/hora RTP derivados do PTP que as essências de áudio e vídeo. Isso garante que os metadados possam ser rigorosamente sincronizados no tempo com os fluxos de mídia aos quais se referem, o que é absolutamente crítico para o áudio baseado em objetos.

- **Extensibilidade:** O padrão foi projetado para se adequar a novas necessidades no futuro, sendo capaz de acomodar novos tipos de metadados à medida que surgem. O ST 2110-41 fornece a estrutura unificadora para transportar a crescente família de fluxos de dados demandada pela indústria.

Após dissecar os componentes individuais – a cena de áudio MPEG-H, a serialização S-ADM e a camada de transporte ST 2110-41 – é hora de sintetizá-los num fluxo coeso de ponta a ponta. Este processo demonstra como os metadados de áudio viajam desde a sua criação numa estação de trabalho de áudio digital (DAW), ou em uma produção ao vivo, até à sua reconstrução num dispositivo receptor, atravessando toda a cadeia de transporte e transmissão.



Serialização S-ADM e a camada de transporte ST 2110-41/ Reprodução

Implementação prática: Adoção pela indústria

A transição da teoria para a prática expõe as realidades da construção de sistemas de broadcast robustos e interoperáveis baseados neste fluxo avançado. Embora os padrões forneçam os pilares, o sucesso da implementação depende da maturidade

do ecossistema circundante, do rigor nos testes de interoperabilidade e de uma abordagem holística ao design da rede. A indústria já está se adaptando para adotar esta tecnologia comercialmente, demonstrando a sua viabilidade para a produção do mundo real.

Validação no mundo real: Demonstração do Fraunhofer IIS no SET EXPO 2025

A confiança da indústria em novos padrões é construída através de demonstrações de interoperabilidade bem-sucedidas. A prova de que diferentes produtos de fornecedores podem trocar fluxos em conformidade com os padrões e funcionar em conjunto é fundamental para a adoção. Neste domínio, o workflow S-ADM sobre ST 2110-41 já alcançou marcos importantes.

Um exemplo notável é a colaboração entre a BBright, um fornecedor de soluções de playout, e a Jünger Audio, um especialista em processamento de áudio. As duas empresas demonstraram com sucesso a interoperabilidade entre a solução de playout da BBright e a plataforma *flexAI* da Jünger Audio, transmitindo metadados S-ADM através do padrão SMPTE ST 2110-41 durante a SET EXPO 2025, no estande do Fraunhofer IIS. Este teste de ponta a ponta validou que um sistema pode gerar o fluxo de metadados do MPEG-H Audio e outro pode recebê-lo e processá-lo corretamente, provando que o fluxo é prático e implementável.

O Fraunhofer IIS e os seus parceiros têm apresentado consistentemente fluxos IP de ponta a ponta que mostram a criação, transporte e reprodução de áudio MPEG-H utilizando S-ADM, ST 2110-30 e o ST 2110-41. Estas demonstrações servem como prova de conceito para a comunidade de broadcast em geral, aumentando a confiança e acelerando a curva de adoção.

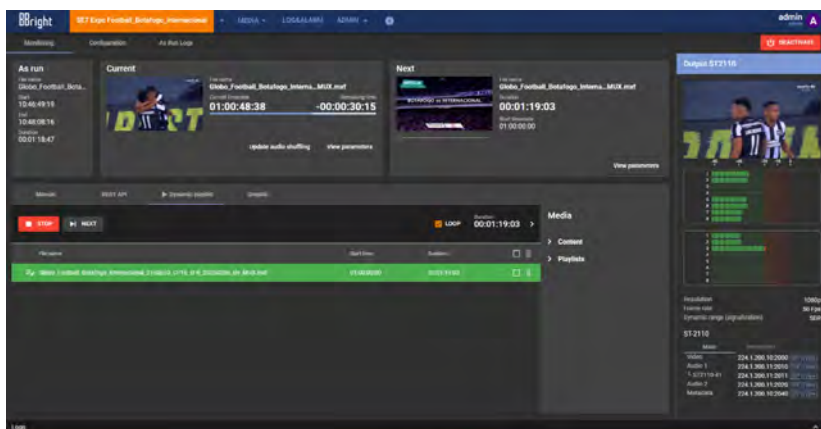


Gráfico 8: Reprodução

Conclusão: Perspectivas futuras e o impacto na TV 3.0

No início de 2025 foi publicada a recomendação ITU-R BS.2168, intitulada *“Audio definition model and serial representation of audio definition model profile for advanced sound systems emission”* (“Modelo de Definição de Áudio e representação serial do perfil do modelo de definição de áudio para emissão de sistemas de som avançados”), que estabelece um marco crucial para a próxima geração de experiências de áudio na radiodifusão. Publicada pelo Setor de Radiocomunicações da União Internacional de Telecomunicações (ITU-R), esta norma define perfis específicos para garantir a interoperabilidade na transmissão de áudio imersivo e interativo, como o som baseado em objetos e em cenas.

Em essência, a recomendação visa padronizar como os metadados de áudio avançado são transmitidos, assegurando que os receptores (como televisores e sistemas de *home theater*) possam interpretar e reproduzir corretamente a intenção original do criador do conteúdo. Isso abre portas para experiências de áudio mais ricas e personalizadas para o consumidor.

A convergência do sistema MPEG-H, da serialização S-ADM e do transporte ST 2110-41 não é apenas um avanço técnico; é um catalisador

que irá remodelar a produção de áudio e abrir novas possibilidades para a criação e consumo de conteúdo. Este fluxo padronizado e nativo de IP representa a peça final necessária para desbloquear todo o potencial do Áudio de Próxima Geração MPEG-H em escala, especialmente em ambientes de produção ao vivo.

Com a chegada da TV 3.0 (DTV+) o seu impacto será sentido em múltiplos domínios, desde o esporte e entretenimento até aos serviços de acessibilidade, anunciando uma era de experiências sonoras mais personalizadas, inclusivas, imersivas e interativas.



Gráfico 9: Reprodução

Referências

1. ISO/IEC 23008-3 Information technology — High efficiency coding and media delivery in heterogeneous environments — Part 3: 3D audio
2. ITU-R BS.2076 Audio Definition Model (ADM)
3. ITU-R BS.2125 Serial ADM (S-ADM)
4. ITU-R BS.2168 (02/2025) Audio definition model and serial representation of audio definition model profile for advanced sound systems emission
5. SMPTE ST 2110-10 Professional Media Over Managed IP Networks: System Timing and Definitions
6. SMPTE ST 2110-30 Professional Media Over Managed IP Networks: PCM Digital Audio
7. SMPTE ST 2110-31 Professional Media Over Managed IP Networks: AES3 Transparent Transport
8. SMPTE ST 2110-41 Professional Media Over Managed IP Networks: Fast Metadata
9. IETF RFC 8331 RTP Payload Format for SMPTE ST 291-1 Ancillary Data



Gabriel Thomazini

é graduado em Ciência da Computação, com Pós-graduação em Engenharia de Redes de Telecomunicações, Gabriel Thomazini iniciou sua carreira como engenheiro de gravação e passou por vários estúdios, trabalhando com artistas de diversos gêneros. Trabalhou como engenheiro de áudio em emissoras de TV e se envolveu cada vez mais com aplicações AV. Liderou projetos notáveis, como a implementação de unidades móveis baseadas em IP, integrações de estúdios e salas de controle, e o desenvolvimento de fluxos de trabalho para áudio remoto e 3D. Sua expertise também inclui a realidade estendida, onde contribuiu com o desenvolvimento de soluções de áudio para aplicações de VR e AR. Após mais de 20 anos na área de broadcast, ingressou na Fraunhofer IIS em 2021, onde atua como consultor de áudio para o desenvolvimento do ecossistema MPEG-H Audio no Brasil.

Contato: gabriel.thomazini@iis-extern.fraunhofer.de