Áudio MPEG-H: principais características

Autores aprofundam conhecimentos e oferecem dados sobre o sistema de áudio da próxima geração para a TV 3.0 do Brasil.

Por Gabriel Thomazini e Uirá Moreno

Foto: Chauhan Moniz em Unsplash

Introdução

O Fórum do Sistema Brasileiro de Televisão Digital Terrestre (SBTVD) submeteu, em julho de 2020, uma Chamada de Propostas (CfP) em busca de contribuições para o sistema de TV Digital da próxima geração do Brasil, no âmbito do "Projeto TV 3.0". Sem as restrições de compatibilidade retroativa, o Projeto da TV 3.0 está preparando o caminho para um sistema de televisão de próxima geração avançado e moderno no Brasil. [1].

Após uma complexa fase de testes e avaliações, o sistema de áudio MPEG-H foi selecionado como o único sistema de áudio obrigatório para a próxima geração da TV no Brasil. Durante a avaliação, o sistema de Áudio MPEG-H cumpriu todos os requisitos obrigatórios para a TV 3.0.

Finalizado em 2015, o Áudio MPEG-H 3D é o mais recente padrão de compressão de áudio desenvolvido pelo MPEG (***Moving Pictures Experts Group***), seguindo um processo competitivo e colaborativo, com a participação dos principais especialistas mundiais no campo da tecnologia de codificação de áudio. Desde o início do desenvolvimento do Áudio MPEG-H, o objetivo definido foi proporcionar a melhor experiência possível com som imersivo, além de permitir recursos avançados de acessibilidade, interatividade e personalização em uma única solução, elevando o áudio a um novo patamar.

A tecnologia de áudio MPEG-H proporciona mais realismo com o som vindo de uma camada superior e em torno do ouvinte, além de um grau de liberdade sem precedentes aos consumidores para personalizar a experiência de áudio. Ao mesmo tempo, com seu rico conjunto de metadados, o Áudio MPEG-H torna o conteúdo mais acessível para o público com deficiência auditiva ou visual e proporciona às emissoras e criadores de conteúdo o controle total sobre as opções de interatividade.

Características do Áudio MPEG-H e experiência do usuário

Se comparado aos padrões antigos de compressão de áudio usados na transmissão de TV e *streaming*, como o AAC e HE-AAC, o sistema de Áudio MPEG-H se traz uma solução mais eficiente e com mais recursos.

Isto inclui som imersivo e capacidade de personalização, com opções avançadas de acessibilidade, funcionalidade de renderização e adaptações do conteúdo, em um único projeto de sistema para conectividade entre vários dispositivos. A combinação de tecnologias de codificação altamente eficientes, a capacidade de representar conteúdo de áudio usando formatos baseados em canais e em objetos, bem como gerenciamento avançado de ***loudness*** e de controle da faixa dinâmica (***Dynamic Range Control*** – DRC) formam a base do sistema de Áudio MPEG-H.

Som imersivo

O Áudio MPEG-H abre uma nova perspectiva de som que vai além do estéreo e do *surround*. Com os sons vindos de cima, além do entorno do ouvinte, uma terceira dimensão é adicionada à experiência de áudio. Com a adição do eixo vertical, os ouvintes experimentam uma representação sonora mais realista e natural.

Os telespectadores podem desfrutar de música e eventos esportivos imersos nos sons que vêm de todas as direções, como os avisos que vem de cima de um sistema de som em um estádio ou passos de uma pessoa caminhando, vindos de baixo. O caso de uso mais comum para aplicações de transmissão é a utilização da mistura de um arranjo imersivo fixo, a chamada "cama de canais" (por exemplo, no formato 5.1+4H) e vários objetos de áudio adicionais como: diferentes comentaristas para um jogo de futebol, uma áudio descrição para a transmissão de uma novela ou ainda efeitos sonoros para conteúdo cinematográfico, por exemplo.

O som 3D pode ser convenientemente reproduzido em ambientes residenciais usando um *soundbar* imersivo, como o modelo da Sennhesier da linha Ambeo (Figura 1, à esquerda) ou um sistema de entretenimento doméstico imersivo, como o Sony HT-A9, com suporte nativo para o *360 Reality Audio* (Figura 1, à direita).

Figura 1 - Som imersivo em casa (A esq: Sennheiser AMBEO. A dir. Sony HT-A9) / Fotos: Divulgação

Personalização e interatividade

A adição de objetos de áudio introduz a opção de interatividade e personalização para o espectador, aprimorando a experiência do usuário. Isto implica mudar o paradigma de produção para uma abordagem de áudio baseada em objetos. Os metadados do Áudio MPEG-H trazem um rico conjunto de informações que oferece aos espectadores grande flexibilidade para se envolver ativamente com o conteúdo e adaptá-lo às suas preferências. A maneira mais fácil de interagir com o conteúdo é selecionar um dos vários ***presets*** de áudio definidos para a produção. Esses ***presets*** são mixagens de áudio predefinidas, com uma identificação descritiva atribuída a eles, por exemplo, "Mixagem de TV", "***Dialog+***" ou "Local". Além disso, são possíveis ajustes individuais, tais como aumentar o volume do diálogo em relação à música de fundo e efeitos. Por fim, os ouvintes podem personalizar cenários avançados nos quais certos elementos sonoros da mixagem de áudio podem ser selecionados e ter níveis e posições ajustados.

Uma interface de usuário é disponibilizada com todas as opções de personalização nos dispositivos ou aplicativos habilitados para áudio MPEG-H, para que os telespectadores possam personalizar seu conteúdo usando, por exemplo, o controle remoto da TV ou a tela sensível ao toque em um dispositivo móvel. A interface de usuário do áudio MPEG-H se adapta automaticamente às intenções do criador do conteúdo e exibe apenas as opções de interatividade disponíveis no momento, como mostrado na Figura 2.

Por exemplo, durante um jogo de futebol os

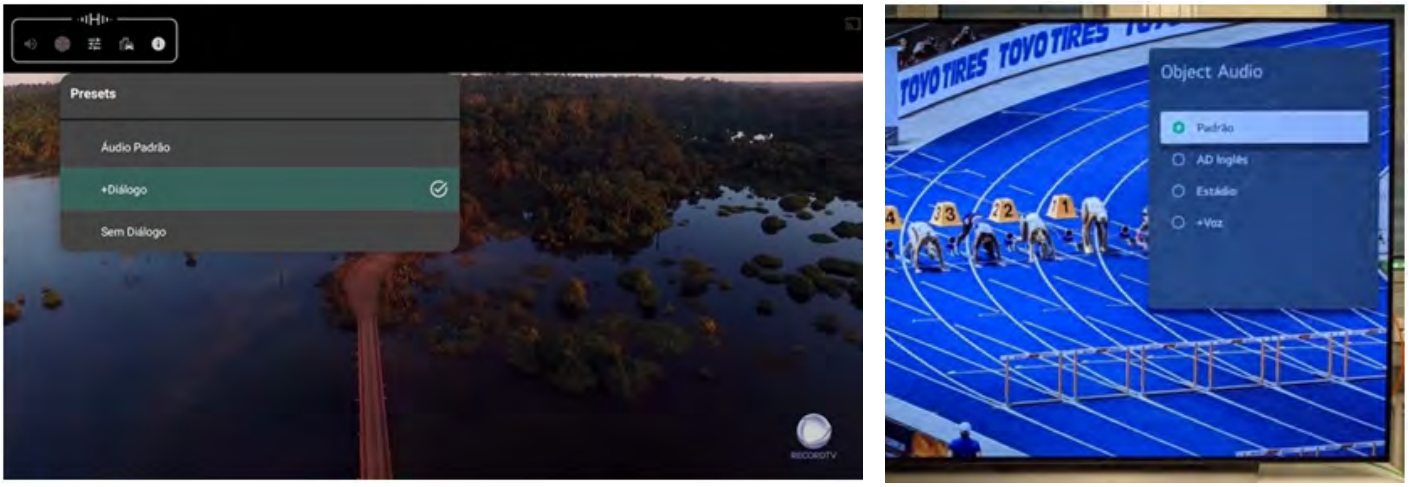

Figura 2 - Exemplo de interface do usuário na TV recebendo sinal de áudio MPEG-H (esquerda) e em dispositivo móvel recebendo o Áudio MPEG-H por *streaming* (direita) / Fotos: Reprodução

torcedores nas suas casas podem decidir ouvir apenas a torcida de seu próprio time, sem nenhum comentarista, e experimentar o jogo como se estivessem presentes no estádio. Alternativamente, eles podem trocar entre diferentes comentaristas ou habilitar um locutor em seu idioma preferido. Os deficientes visuais podem desfrutar do conteúdo com seus amigos e familiares, ajustando de forma independente o nível da Audiodescrição e posicionando-a no espaço 3D, proporcionando uma separação espacial do diálogo principal e aumentando a inteligibilidade.

Os provedores de conteúdo e as emissoras desejam controle sobre as opções dos telespectadores para alterar a forma como o conteúdo é exibido. Portanto, os metadados do áudio MPEG-H possibilitam aos radiodifusores total controle sobre as opções de interatividade oferecidas, permitindo definir estritamente os limites nos quais o usuário pode interagir com o conteúdo. A Figura 3 mostra um exemplo das configurações dos ajustes de interatividade na etapa de criação (***authoring***), na qual o criador do conteúdo é capaz de decidir a extensão dos ajustes oferecidos ao espectador, como o nível ou a posição do diálogo.

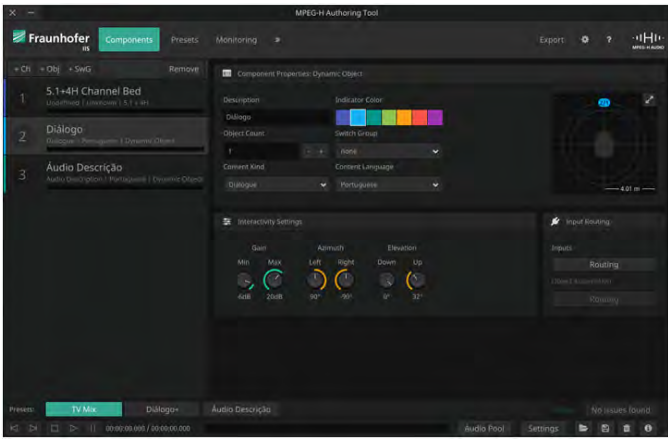

Figura 3 - Ferramenta de Autoração MPEG-H. Exemplo de configurações de interatividade / Foto: Reprodução

Controle avançado de Loudness e DRC (Controle de Faixa Dinâmica)

Os metadados de áudio MPEG-H contêm todas as informações necessárias para reprodução e renderização em diferentes esquemas de reprodução para garantir a melhor experiência de áudio em qualquer plataforma. O sistema inclui a funcionalidade de renderização, redução de canais (***downmix***) e também gerenciamento avançado de ***loudness*** e controle de faixa dinâmica (***Dynamic Range Control*** - DRC). O módulo de normalização de ***loudness*** garante um nível consistente entre programas e canais, para diferentes configurações predefinidas de reprodução, baseado em informações de ***loudness*** já incorporadas ao fluxo de áudio MPEG-H. A presença dessas informações para cada ***preset*** permite a normalização instantânea e automatizada dos níveis quando o usuário alterna entre diferentes ***presets***. Além disso, podem ser inseridas informações de ***loudness*** para situações específicas, com diferentes opções para redução de canais (***downmix***).

O padrão MPEG-H áudio 3D suporta informações de alto nível que são obrigatoriamente incluídas nos metadados do fluxo de áudio MPEG-H. Vários padrões para a aferição de ***loudness***, como ITU-R BS.1770, EBU R-128 e ATSC A/85 são suportados para cumprir com os regulamentos e recomendações de transmissão aplicáveis. O sistema permite especificar se as informações estão relacionadas ao ***loudness*** de um programa completo ou se são referenciadas a um elemento de âncora específico, tal como o diálogo ou comentário.

A Figura 4 ilustra o conceito geral de normalização de

loudness em três exemplos de conteúdos transmitidos com valores aferidos diferentes da referência definida no decodificador.

Enquanto o comercial e o programa esportivo mostram níveis mais altos, o *loudness* do filme fica muito baixo se comparado com o "alvo" pretendido. O lado esquerdo da figura mostra os itens não processados e o lado direito ilustra seus níveis e faixa de *loudness* depois que a normalização do MPEG-H foi aplicada.

O *loudness* de reprodução de todos os três itens do programa é o mesmo e corresponde ao nível definido no decodificador.

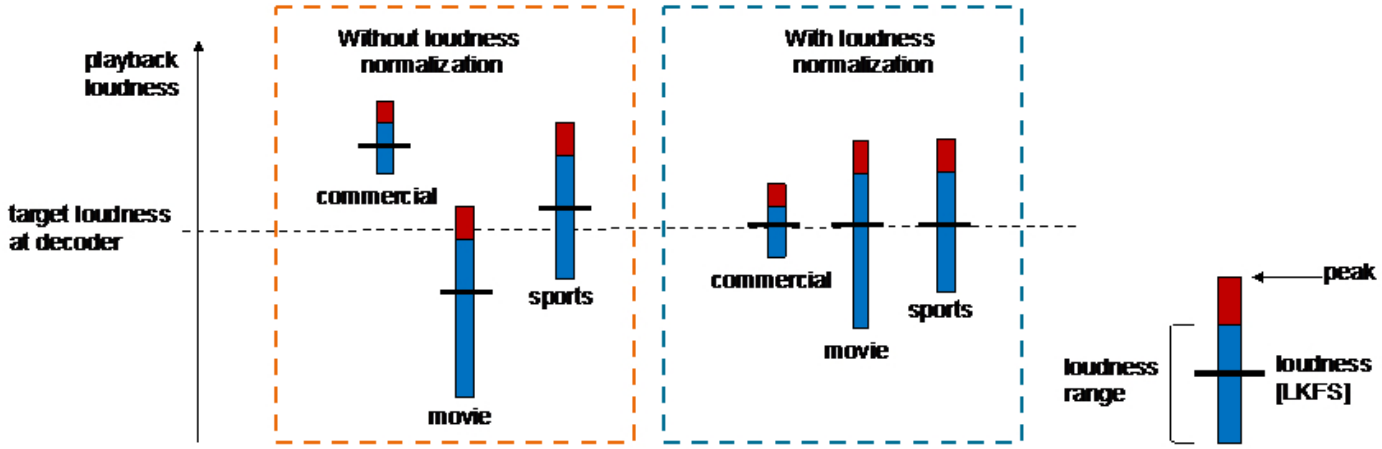

Figura 4: Exemplo de normalização de **loudness** para três diferentes conteúdos com extensão dinâmica e níveis variados / Foto: Reprodução

No caso de um fluxo de áudio MPEG-H conter vários *presets* do mesmo programa, as informações de *loudness* serão inseridas separadamente para cada *preset*. Isto permite o controle imediato e automático do *loudness*, mesmo para o áudio interativo e personalizado. Por exemplo, quando o usuário alterna entre diferentes *presets*, a normalização de *loudness* é instantaneamente ajustada para assegurar uma reprodução com níveis consistentes em todos as opções oferecidas.

Além do módulo de normalização de *loudness*, o áudio MPEG-H ainda inclui um componente adicional para a compensação do *loudness*, responsável por ajustar os níveis após as interações do usuário. Por exemplo, se o usuário aumenta o nível do diálogo isso faria o nível geral de *loudness* aumentar também. Nesse caso, o decodificador do áudio MPEG-H vai diminuir automaticamente o nível da mixagem completa, após a interação do usuário, de tal forma que o *loudness* geral seja compensado e permaneça constante, como mostrado na Figura 5.

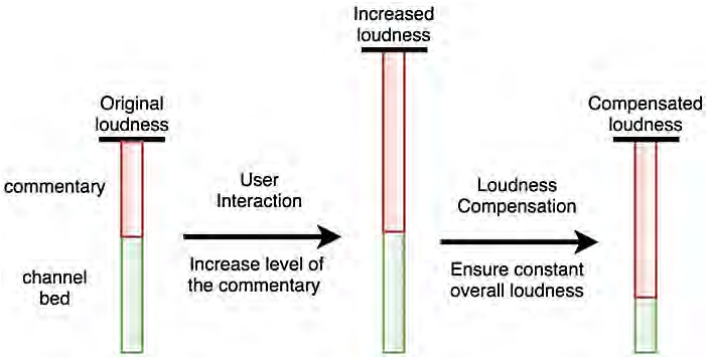

Figura 5: Ilustração do conceito de compensação de loudness para adequar o resultado da interação do usuário. Após aumentar o diálogo, o sistema diminui o nível da cama de canais para manter o loudness de acordo com a referência utilizada / Foto: Reprodução

Autoria de metadados e os fluxos de trabalho para produção

A produção e transmissão do áudio MPEG-H introduz novos conceitos em comparação com uma produção comum[2]. O sistema de áudio MPEG-H foi projetado especialmente para explorar estas novas opções criativas. Além do áudio 3D imersivo, os criadores de conteúdo podem preparar diferentes mixagens (além da mixagem padrão ou principal de um programa) usando as ferramentas de criação. Através de ajustes de ganho e posicionamento dos objetos é possível criar diferentes predefinições (*presets*) de mixagem que poderão ser apresentadas para o usuário em um simples menu.

Todos os recursos de interatividade oferecidos aos usuários são estritamente definidos pelo produtor durante o processo de criação dos metadados. Este processo de geração de metadados é chamado de "autoração". Comparativamente, essa etapa é a maior diferença entre uma produção de conteúdo de áudio MPEG-H e uma produção comum. Durante o processo de autoração, os

objetos de áudio devem ser mantidos separados dos outros componentes, tais como Música e Efeitos (M&E) – a chamada cama de canais. Os metadados gerados irão conter, por exemplo:

- Informações sobre a posição de reprodução dos objetos no espaço tridimensional;
- Limites de interatividade nos objetos de áudio e nos diferentes ***presets;***
- Informações de ***loudness*** sobre cada componente e cada ***preset;***
- Informações com os textos descritivos dos ***presets*** e, também, dos objetos de áudio (em diferentes idiomas);
- Esquema do arranjo de caixas acústicas usado como referência.

O sistema de áudio MPEG-H é projetado para trabalhar com os atuais equipamentos de *streaming* e transmissão usando fluxos de trabalho baseados em SDI ou IP. Em cenários de produção ao vivo, a autoração do áudio MPEG-H é realizada por um dispositivo chamado "Unidade de Autoração e Monitoramento" (AMAU), como ilustrado na Figura 6. Este dispositivo exporta os metadados em tempo real, modulados em um sinal de áudio que é sincronizado com o sinal de vídeo, e usa qualquer um dos formatos comumente utilizados em produções lineares, tais como SDI, MADI ou AoIP.

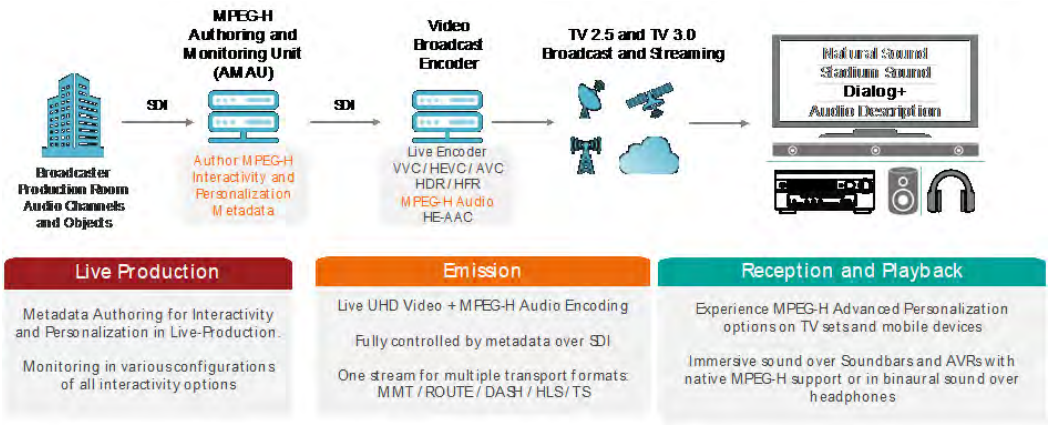

Figura 6: Fluxo de Produção ao vivo com Áudio MPEG- (simplificado) / Foto: Reprodução

Para garantir a integridade dos metadados no fluxo SDI, em todas as etapas de produção, é utilizada uma solução onde as informações são moduladas e entregues através de um canal de áudio chamado "Pista de Controle" (*Control Track*). Isto assegura a sincronia dos metadados com os sinais de áudio e vídeo correspondentes. A Pista de Controle do áudio MPEG-H é robusta o suficiente para suportar conversões A/D e D/A, mudanças de nível, conversões de taxa de amostragem ou edição de vídeo. Os metadados são coletados em pacotes sincronizados com o sinal de vídeo, e ordenados com modulação PCM, em um sinal que se encaixa na largura da banda do canal de áudio (veja Figura 7). Este sinal não é afetado por operações de filtragem, reamostragem ou escalonamento realizados pelos equipamentos de transmissão

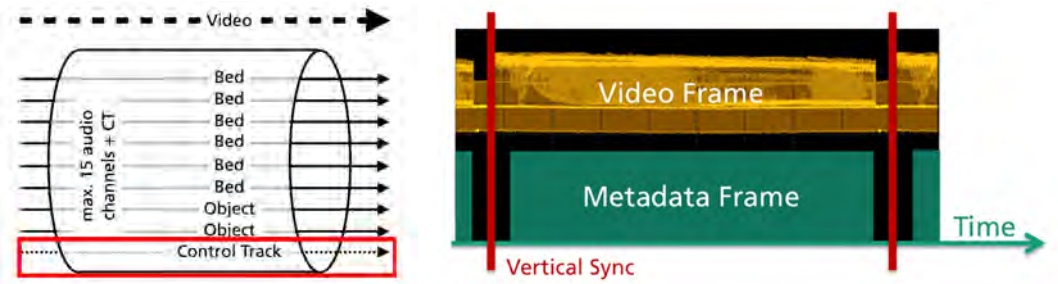

Figura 7: Ilustração da ControlTrack do Áudio MPEG-H / Foto: Reprodução

Como o sinal de sincronização vertical já está incluído, a precisão em nível de frame pode ser assegurada, possibilitando a comutação dos sinais de áudio sem nenhuma interrupção ou falha. Como a Pista de Controle é apenas um sinal de áudio semelhante a um código de tempo linear (***Linear Time Code*** -LTC), ele pode ser tratado como qualquer outra trilha de áudio em sistemas de edição de áudio ou vídeo.

Referências

[1] Fórum SBTVD - TV 3.0 - CfP Fase 2 / Teste e Avaliação ,

[2] Fluxos de Trabalho e Ferramentas de Produção MPEG-H

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. Realizou projetos em unidades móveis baseadas em IP, estúdios e salas de controle, bem como o desenvolvimento de fluxos de trabalho remotos e mixagem para formatos de áudio 3D. Colaborou em iniciativas de realidade estendida, desenvolvendo soluções de áudio para aplicações 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 de áudio MPEG-H no Brasil.

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

Uirá Moreno Rosário e Barros é mestre em Engenharia Elétrica com ênfase em Telecomunicações pela Universidade Mackenzie, atualmente é PO no Lab de Inovação Telecom, com foco em Futuro das Redes de Conectividade e de Distribuição de Conteúdo. Como destaques de cases de inovação, foi responsável pela Primeira Transmissão Comercial Imersiva de Áudio ao Vivo em ISDB-T e o primeiro uso de câmeras 5G no Carnaval, Rock in Rio e Prêmio Multishow de Música. Também é Professor de Rádio, TV e Internet, na Faculdade Cásper Líbero, lecionando as disciplinas de “Captação e Edição de Áudio”, “Tópicos Avançados em RTVI” e “Estéticas Sonoras e Musicais”. É membro do Grupo de Trabalho da TV 3.0 na SET, coordenador do tema Áudio Imersivo.

Contato: uira.moreno@g.globo