

MPEG-H 电视音频系统及国际标准化发展

The MPEG-H TV Audio System and Its Developments in International Standardization

■ Fraunhofer Institute for Integrated Circuits (IIS)

Stefan Meltzer Max Neuendorf Jan Plogsties Robert Bleidt

摘要 重点介绍了被 ATSC 3.0 采纳并成为其最终标准的下一代广播电视音频标准——MPEG-H 电视音频系统，阐述该标准的技术特征及其在下一代广播电视标准应用中的核心性能；阐述 MPEG-H 在 ATSC 3.0 标准发展过程中的大事记，以及对下一代电视音频标准商业推广的相关思考。

关键词 MPEG-H 电视音频 3D 音频 沉浸式声音 互动式声音

Abstract: The article introduces the major features of the MPEG-H TV Audio System and its core elements. The MPEG-H TV Audio system has been adopted and fully specified in the ATSC 3.0 Digital Television standard. The article also describes the major steps in the process of the standardization and, necessary considerations regarding the commercial roll out of the next generation TV standard.

Keywords: MPEG-H TV Audio; 3D Audio; immersive sound; interactive sound

1 引言

MPEG-H 电视音频系统（下一代音频编解码器）基于音频编码标准化组织 MPEG 最新开发的 MPEG-H 3D 音频标准，能够为家庭端带来沉浸式声音体验，增强电视节目的真实感和沉浸感，并通过音频对象为观众呈现个性化的音频内容。为了方便用户使用交互性功能，广播电视公司可以提供“预设”服务，即预定义对象增益和位置的混音效果。而沉浸式声音的制作播出则可通过将基于声道的信号或基于场景的声音元素与静态或动态音频对象相结合的方式来实现。改良后的响度和动态范围控制，能够根据不同类型的用户设备和收听环境来定制声音效果，从而实现最佳的声音回放。

目前，MPEG-H 电视音频系统已经被 DVB 标准采纳，并已成为 ATSC 3.0 数字电视标准中的正式音频标准。此外，韩国于 2017 年 5 月 31 日推出基于 ATSC 3.0 标准的超高清电视服务，这也是首个采用 MPEG-H 电视音频系统的常规广播电视服务。

2014 年，为了支持包括沉浸式声音和交互性能

等在内的全新音频功能并提升用户体验，美国高级电视系统委员会（ATSC）发布公告征集下一代广播电视音频系统标准方案，并将其作为 ATSC 3.0 标准的一部分。与此同时，ATSC 还要求解决方案能够实现部分高级功能，例如：将通过互联网发送的音频元素与通过地面广播路径传输的音频元素相结合；支持观众调整节目中的某些声音元素；使音频可以适用于各种终端用户设备，从而通过同一音频码流就可以服务于所有类型的终端用户设备；通过 binaural（双耳回放）技术实现耳机端的沉浸式声音回放等。而 MPEG-H 3D 音频标准可以满足以上所有需求。MPEG-H 电视音频系统是 MPEG-H 3D 音频标准的子集，专为满足 ATSC 3.0 和 DVB 所规定的广播电视需求而定制。

2 MPEG-H 电视音频系统的特征

2.1 沉浸式声音

MPEG-H 电视音频系统能够通过垂直维度上扩展声像实现“3D”声音回放，进而实现沉浸式声

音并区别于传统环绕声。该功能能够使声音听起来更加真实，使用户不再是一名旁观者，而是体验身临其境一般的感受。研究表明，与环绕声相比较，在沉浸式声音环境下，人们感知到的整体音质明显改善，两者之间的差异就像立体声与环绕声之间的差异一样大。图 1 为与 22.2 声道参考信号相比，环绕声、沉浸式 /3D 声音格式的声音还放系统扩展后的整体音质改善。

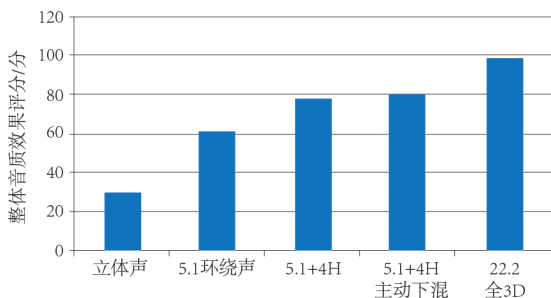


图 1 声音还放系统扩展后的整体音质改善

2.2 交互功能 / 个性化定制

MPEG-H 电视音频系统中包含“预设”功能。内容创作者可以定义“预设”内容，可以涵盖不同声音元素的组合以及调整不同元素在位置和音量上的关系。如此一来，广播电视公司可以让用户轻而易举地选择不同的声音体验。另外，通过 MPEG-H 电视音频系统，用户能够通过与其他混音元素的比较，直接选择和调整不同混音元素在位置和音量上的关系，从而对声音体验进行微调。其中一个典型的用例是调高或调低解说员的声音（也称为“对白增强”）或者选择不同的解说评论的语言。而实现传输交互内容其实也并不复杂，只需对现有制作流程稍作调整便能够为内容创作者提供全新选择，也能够为观众带来更好的体验。

2.3 通用传输

在多平台环境下，相同的 MPEG-H 内容通过不同的分发网络进行传输（例如：广播电视网络、移动网络和宽带网络），并且在不同的环境（如嘈杂的公交车和安静的客厅）和设备上（如电视机、移动设备和 AVR）进行播放。为了满足以上需求，MPEG-H 电视音频系统针对每个完整混音甚至是单一音频元素提供响度和动态范围控制，以及对输出的峰值和

削波预防进行控制。此外，全新的主动下混算法能够带来更高质量的下混音频信号。为了能通过耳机聆听沉浸声和环绕声音频内容，MPEG-H 电视音频系统还包含 binaural 渲染器。

3 MPEG-H 电视音频系统的核心性能

3.1 码流效率

下一代广播电视节目需要能够通过各种方式向用户提供内容，其中包括最高质量的有线电视和卫星电视传输方式以及移动设备上的流媒体传输方式等。为此，电视音频系统需要传输各种音频内容，从立体声道到 5.1 声道、7.1 声道和用于沉浸式音频内容的更多声道。MPEG-H 音频规范的目标码流范围旨在满足质量和效率的双重目标。对于当今的 5.1 环绕声，96~256 kbit/s 的码流便能够提供良好甚至优异的音频质量。对于未来具有更多声道（5.1 +4H 或者 7.1+4H）的音频内容，256 kbit/s~1.2 Mbit/s 的码流即可实现相同的音频质量。

MPEG-H 音频编解码器能够利用相同的码流承载更高质量和 / 或更多声道的音频内容。例如，在常用的广播音频数据传输码率（384 kbit/s）下，MPEG-H 能够传输最多包括四个附加对象的 7.1+4H 音频声道的信号，从而实现高度扬声器播放效果。

3.2 动态范围控制

动态范围控制（DRC）技术能够使产生的音频信号适应于不同的还放设备和聆听环境。在 MPEG-H 电视音频系统的 DRC 框架内，通过信号传输的不同 DRC 增益序列能够在播放设备中实现由编码器控制的动态范围处理。多个单独的 DRC 增益序列可以通过高分辨率信号传输，以适用于各种播放设备和收听环境，如：移动环境和家庭环境。通过 MPEG-H DRC 技术，还可以实现更好的削波预防和峰值限制效果。

3.3 响度标准化

为了符合不同的响度规范并带来一致的用户体验，响度信号传送和标准化对于任何下一代音频系统都扮演着至关重要的角色。在 MPEG-H 电视音频系统中，基于 EBU R128 或 ITU-R BS.1770-3 的

所有响度相关测量数据都被嵌入到用于响度标准化的数据流之中，并通过解码器对音频信号进行标准化，从而将节目响度映射至还放所需的目标响度。例如，在移动设备上，目标响度范围为 -12~-15 dB LKFS，而在家庭 AVR 上，目标响度通常设置为 -31 dB LKFS。然而，通过下混和动态范围控制，可能会改变信号的响度。因此，可以将专用的节目响度元数据植入到 MPEG-H 比特流中，以确保在还放端实现正确的响度标准化。

3.4 音频对象

将音频对象作为附加音轨嵌入到音频节目内容中，可以实现一系列的全新应用。其中，最为突出的应用就是允许用户通过改变混音效果实现交互性功能。用户可以选择不同的语言音轨、附加音轨，例如导演对电影的评论或针对视障者的场景叙述等。同时，用户能够调节特定音轨声音，以便获取个性化收听体验，例如调高解说评论的声音使其高于背景音。

诸如对白的音频对象可以根据其动态范围进行单独控制，这样可确保在任何压缩模式下都能清晰地听到对白内容。利用音频对象的概念，还可以在不同播放场景下，实现准确的空间声音还放。为实现以上需求，可将描述几何位置的对象元数据嵌入到比特流中。MPEG-H 电视音频系统的解码器包含一个对象渲染器，它能够根据元数据和用户家中扬声器的位置将对象信号映射并分配至扬声器中。

3.5 Ambisonics

Ambisonics 能够通过特定的数学方程式呈现声场压力。随着 Ambisonics 阶数的提升，其还放准确度也随之升高。此外，Ambisonics 的呈现不受终端扬声器配置的影响，因此，在声音还放之前可以轻而易举地对其进行调整。得益于以上特性，Ambisonics 还放技术非常适用于当今的虚拟现实应用。目前，结合音频对象的 FOA(一阶高保真)技术和 HOA(高阶高保真)技术在虚拟现实制作中的应用越来越广泛。

3.6 流媒体

对于传输系统和其所涉及的媒体来说，将互联网流媒体音频内容传输到移动设备的过程极具挑战性。甚至当今常用的 3GPP 或 LTE 网络的移动数

据连接也无法保证提供恒定的带宽。为了确保内容还放的连续性，最新的广播电视设备采用了诸如 MPEG-DASH 等动态自适应流媒体技术，从而确保内容码流能够完全适应于当前的信号连接质量。

其实，自适应流媒体的概念并不新鲜，早在 MPEG-4 HE-AAC 流媒体应用中已经采用了这一方式。MPEG-H 电视音频系统在设计阶段就已考虑到自适应流媒体的需求，进而可以更容易地实现基于 DASH 服务。音频帧的内置概念缩短了调准时间，并且显著降低了实施复杂性，从而可以在制作过程中轻松进行码流拼接和广告插播，即使是在本地接收机端也可实现个性化广告内容。

3.7 灵活的渲染和还放

针对立体声和 5.1 声道的音频制作和监听扬声器的安装已经很成熟。然而，消费者家中扬声器的配置、数量和位置都有所不同，为了应对这些问题，在 MPEG-H 电视音频系统中，可以通过格式转换器实现针对不同扬声器布局的灵活渲染，从而使内容格式适应于播放端的扬声器设置。

过去几年间，媒体消费已更多转向移动设备，而用户也主要通过耳机收听音频内容。因此，MPEG-H 电视音频系统解码器中集成了 binaural 渲染组件，专门应用于将沉浸声和环绕声内容还放到耳机端。

4 MPEG-H 电视音频系统成功纳入 ATSC 3.0 标准

2010 年，在 ATSC 举办的下一代广播电视最新技术研讨会上，Fraunhofer IIS 提出了全新方案，即通过对象与声道相结合的方式作为下一代音频技术标准。2015 年 1 月，由几家公司组成的联盟向工作组提交了基于 MPEG-H 3D 音频标准的提案。除了 MPEG-H 电视音频系统以外，ATSC 还收到了另外两份提案。图 2 为 MPEG-H 电视音频系统纳入 ATSC 3.0 标准的发展过程。

工作组根据预认证阶段所提交的文件，对三套提案系统进行了评估，其中两套系统获准进入下一阶段的遴选，包括正式的多站点双盲聆听测试和系

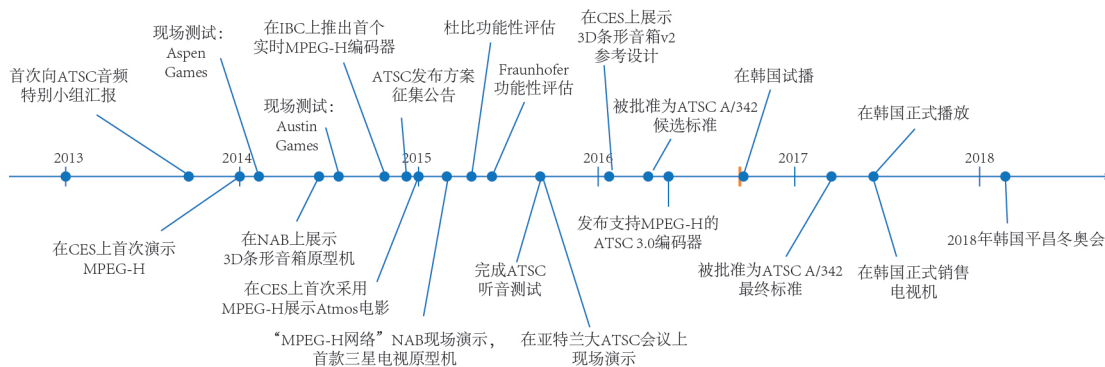


图 2 MPEG-H 电视音频系统发展大事记

统特征评估。听音测试结果表明：两套提案均实现了所需的编码效率，在某些项目中稍微倾向于选择 MPEG-H 电视音频系统，特别是具有低码流语音的项目 (MPEG-H 电视音频系统编解码器的语音编码工具可在其中使用)。2015 年 5、6 月，ATSC 评估员到访两个提案方的实验室，以评估提案特征的演示。最终，两套系统均符合 ATSC 规定的所有要求。2015 年 7 月，ATSC 组织了一次特别的演示活动并对 MPEG-H 测试床进行了展示。2015 年 11 月，ATSC 将两套系统升级为候选标准状态，并建议每个区域使用一套系统。2016 年底，两套系统均被升级为 ATSC 建议标准。2017 年 3 月，两套系统被确认为 ATSC 3.0 的最终标准。目前，两套系统信息都能够在 ATSC 官网 A/342 标准中下载。

5 结语

MPEG-H 电视音频系统的附加功能 (如音频对象和灵活渲染等)，能够为用户带来交互式 and 身临其境的沉浸式音频体验。经过改良响度和 DRC 控制功能，充分体现了以用户为中心的音频标准设计理念。

在 MPEG-H 电视音频系统内部以及 ATSC 3.0 标准开发过程中的测试均表明，MPEG-H 电视音频系统可以满足广播电视公司对于下一代音频系统的所有要求。MPEG-H 电视音频系统已被 ATSC 3.0 采纳为最终标准，并被韩国选定为基于 ATSC 3.0 标准的唯一音频系统，并于 2017 年 5 月 31 日起提供常

规服务。此外，支持 ATSC 3.0 标准 (包含 MPEG-H 电视音频系统) 的电视机已经在韩国正式销售，韩国的广播电视公司将在 2018 年平昌冬季奥运会上采用 MPEG-H 电视音频系统进行电视转播。此外，MPEG-H 电视音频系统也是 DVB 标准系列的一部分，并被考虑应用于其他电视标准中，如中国广播电视 3D 音频标准等。[5]

参考文献

- [1] ISO/IEC . ISO/IEC 23008-3:2015 Information technology High efficiency coding and media delivery in heterogeneous environments - Part 3: 3D audio[S/OL]. [2017-05-10].<https://www.iso.org/standard/63878.html>.
- [2] Herre J,Hilpert J, Kuntz A, et al. MPEG-H Audio-the new standard for universal spatial/3D audio coding[J]. Audio Eng. Soc, AES Journal, 2015, 62: 821-830.
- [3] ATSC 3.0. ATSC 3.0 A/342 Part[EB/OL]. [2017-03-03]. <http://atsc.org/atsc-30-standard/a342-part-32017-mpeg-h-system>.
- [4] Max Neuendorf, Jan Plogsties, Stefan Meltzer, et al. Immersive audio with MPEG 3D Audio - status and outlook[C]//Proceedings of 2014 NAB Broadcast Engineering Conference. 2014: 52-55.
- [5] Robert L, Bleidt, Deep Sen, et al. Development of the MPEG-H TV audio System for ATSC 3.0[J]. IEEE Transactions on Broadcasting, 2017, 63(1): 202-236.

(收稿日期: 2017-05-12)