

国 际 电 信 联 盟

ITU-R

国际电联无线电通信部门

ITU-R BS. 1196-8 建议书
(10/2019)

数字广播的音频编码

BS 系列
广播业务 (声音)



国际电信联盟

前言

无线电通信部门的职责是确保所有无线电通信业务（包括卫星业务）合理、公平、有效和经济地使用无线电频谱，在不受频率范围限制的情况下开展研究，并在此基础上通过建议书。

无线电通信部门的规则和政策职能由研究组支持的世界和区域无线电通信大会以及无线电通信全会来执行。

知识产权政策（IPR）

ITU-R的知识产权政策在ITU-R第1号决议所参引的《ITU-T/ITU-R/ISO/IEC通用专利政策》中进行描述。专利持有者用于提交专利声明和许可声明的表格可从<http://www.itu.int/ITU-R/go/patents/zh>处获取，在此还可获取《ITU-T/ITU-R/ISO/IEC通用专利政策实施指南》和ITU-R专利信息数据库。

ITU-R系列建议书

（也可在线查询<http://www.itu.int/publ/R-REC/zh>）

系列	标题
BO	卫星传送
BR	用于制作、存档和播出的录制；电视电影
BS	广播业务（声音）
BT	广播业务（电视）
F	固定业务
M	移动、无线电测定、业余和相关卫星业务
P	无线电波传播
RA	射电天文
RS	遥感系统
S	卫星固定业务
SA	空间应用和气象
SF	卫星固定业务和固定业务系统间的频率共用和协调
SM	频谱管理
SNG	卫星新闻采集
TF	时间信号和频率标准发射
V	词汇和相关问题

注：该ITU-R建议书的英文版本根据ITU-R第1号决议详述的程序予以批准。

电子出版
2020年，日内瓦

© 国际电联 2020

版权所有。未经国际电联书面许可，不得以任何手段复制本出版物的任何部分。

ITU-R BS.1196-8*建议书

数字广播的音频编码

(ITU-R第19-1/6号课题)

(1995-2001-2010-2012-02/2015-10/2015-2017-01/2019-10/2019年)

范围

本建议书详细说明了适用于数字声音和电视广播的音频源编码系统。它还详细说明了适用于数字声音系统及电视广播系统后向兼容的多声道增强型系统。

关键词

音频、音频编码、广播、数字、广播、声音、电视、编解码器

国际电联无线电通信全会，

考虑到

- a) ITU-R BS.1548建议书详细说明了数字广播音频编码系统的用户需求；
- b) 有或没有附图的多声道音响系统是ITU-R BS.775建议书的主题，使用有效比特率压缩的高质量、多声道音响系统对于数字广播系统至关重要；
- c) ITU-R BS.2051建议书中规定的高级音响系统包含三维声道配置并采用静态或动态元数据控制基于对象的、基于场景的和基于声道的信号；
- d) 对带有微小损害的音频系统（包括多声道音响系统）的主观评价是ITU-R BS.1116建议书的主题；
- e) 对中等音频质量音频系统的主观评价是ITU-R BS.1534（MUSHRA）建议书的主题；
- f) 国际电联无线电通信部门已经测试了高质量音频的低比特率编码；
- g) 不同业务间音频源编码方法的共性可能会提供更高的系统灵活性并降低接收机成本；
- h) 许多广播业务已使用或规范了使用MPEG-1、MPEG-2、MPEG-4、AC-3和E-AC-3系列的音频编解码器；
- i) ITU-R BS.1548建议书列出了被证实可满足广播机构有关馈送、分配和发射要求的编解码器；
- j) 尚未开始提供业务的广播机构应可以选择最适合其应用的系统；
- k) 在选择系统时，广播机构可能有必要考虑与原有广播系统和设备的兼容问题；

* 本建议书应引起国际标准化组织（ISO）和国际电工委员会（IEC）的关注。

- l) 在引进一个多声道音响系统时，还应顾及现有的单声道和立体声接收机；
- m) 对现有音频编码系统的后向兼容多声道扩展能够比同时联播提供更好的比特率效率；
- n) 音频编码系统最好应能够以同样高的保真度对语音和音乐进行编码，

建议

1 如不需要与原有传输设备兼容，数字声音或电视广播发射的新应用应采用下述低比特率音频编码系统之一：

- 扩展版HE AAC，如ISO/IEC 23003-3:2012所述；
- E-AC-3，如ETSI TS 102 366 (2014-08)所述；
- AC-4，如ETSI TS 103 190-1 v1.3.1 (2018-02)和ETSI TS 103 190-2 v1.2.1 (2018-02)所述；
- MPEG-H 3D音频LC型，如ISO/IEC 23008-3:2019所述；
- DTS-UHD，如ETSI TS 103 491 v1.1.1 (2017-04)所述。

注1 – 扩展版HE AAC是比MPEG-4-HE AAC-V2、HE AAC和AAC-LC更加灵活的超集，并包括MPEG-D统一语音和音频编码（USAC）。

注2 – E-AC-3是比AC-3更加灵活的超集。

注3 – AC-4、MPEG-H 3D音频LC型和DTS-UHD规范包括支持高级音响系统（如ITU-R BS.2051建议书所述）的兼容性，用户应参考ITU-R BS.1548建议书的编解码器合规性。

2 如需要与原有传输设备兼容，数字声音或电视广播发射应用应采用下述低比特率编码系统之一：

- MPEG-1层II，如ISO/IEC 11172-3:1993所述；
- MPEG-2层II半取样速率，如ISO/IEC 13818-3:1998所述；
- MPEG-2 AAC-LC或MPEG-2 AAC-LC（带SBR），如ISO/IEC 13818-7:2006所述；
- MPEG-4 AAC-LC，如ISO/IEC 14496-3:2009所述；
- MPEG-4 HE AACv2，如ISO/IEC 14496-3:2009所述；
- AC-3，如ETSI TS 102 366 (2014-08)所述。

注4 – ISO/IEC 11172-3有时可被称为13818-3，因为该规范包括对11172-3的引证。

注5 – 鼓励支持如ISO/IEC 23003-3:2012所述的扩展版HE AAC。其中包括所有上述AAC版本，因此保证与世界各地使用相同单一解码器的未来及传统的广播系统相兼容；

3 数字电视和声音广播系统的后向兼容多声道扩展应采用如ISO/IEC 23003-1:2007所述的多声道音频扩展；

注6 – 由于ISO/IEC 23003-1:2007所述的MPEG环绕声技术独立于用于后向兼容信号传输的压缩技术（核心编码器），因此可结合建议1和建议2中建议的任何编码系统，使用所述的多声道增强工具；

4 对于分配和馈送链路，可使用以下编码之一，每个音频信号（即，每个单声道信号或每个独立编码的立体声信号的组成部分）的比特速率至少为以下比特率（不包括附属数据）；

- MPEG-1 层 II，如 ISO/IEC 11172-3 所规定，每个单声道信号的比特率至少为 180 kbit/s；
- MPEG-4 AAC，如 ISO/IEC 14496-3 所规定，在多达五级级联的情况下，每个单声道信号的比特率至少为 144 kbit/s；
- AC-4，如 ETSI TS 103 190-1 v1.3.1 (2018-02) 和 ETSI TS 103 190-2 v1.2.1 (2018-02) 所规定，在多达五级级联的情况下，每个单声道音频信号的比特率至少为 128 kbit/s；
- MPEG-H 3D 音频，如 ISO/IEC 23008-3 所规定，在多达五级级联的情况下，每个单声道音频信号的比特率至少为 144 kbit/s；

5 对于解说链路，可使用 ISO/IEC 11172-3 层 III 编码，单声道信号的比特率至少为 60 kbit/s（不包括附属数据），利用联合立体声编码，立体声信号的比特率至少为 120 kbit/s（不包括附属数据）；

6 对于高质量应用，取样频率应为 48 千赫兹（kHz）；

7 低比特率音频编码器的输入信号不应加重，且编码器不应采用任何加重方式，

进一步建议

1 有关已被证实满足馈送、分配和发射质量及其他用户要求之编码系统配置的信息，应参见 ITU-R BS.1548 建议书最新版本；

2 需要对 ITU-R BS.2051 建议书最新版本中规定的高级音响系统的要求做进一步研究，且当完成这些研究后，应对本建议书进行更新。

注 – 本建议书中有关编解码器的信息见附件 1 至附件 8。

附件 1 (资料性)

MPEG-1 和 MPEG-2，层 II 和层 III 音频

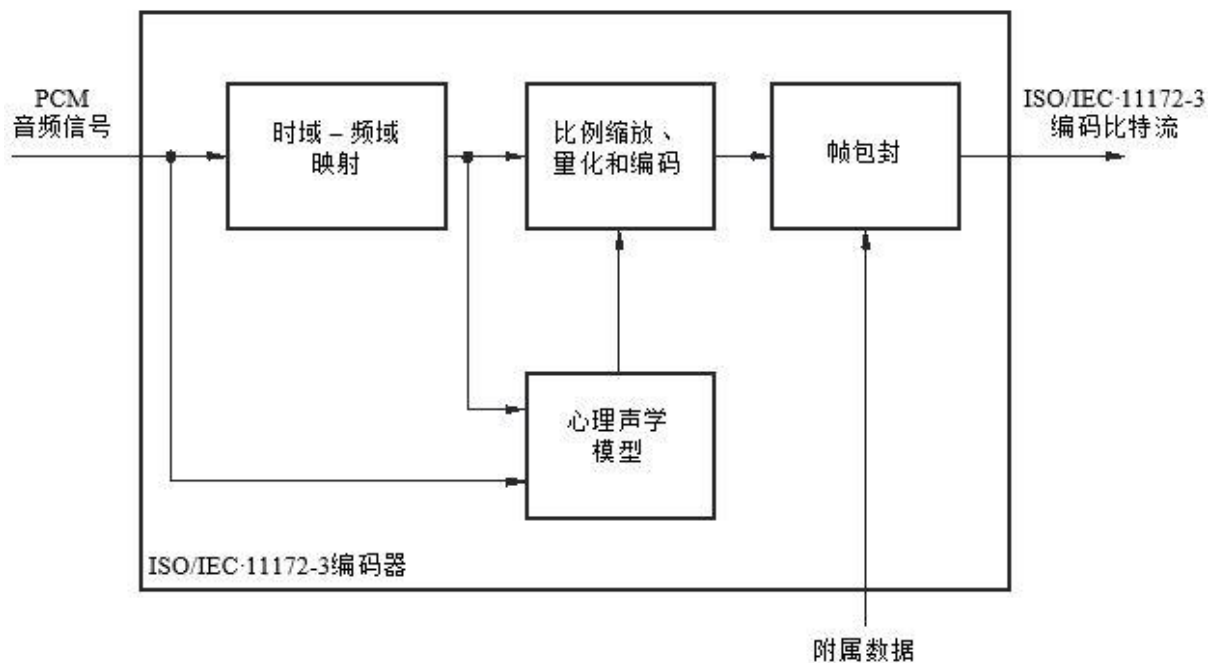
1 编码

编码器对数字音频信号进行处理，产生压缩比特流。编码器算法并无标准化的规定，可采用各种编码手段，诸如听觉掩模门限估计、量化和比例缩放（注 1）。不过，编码器输出必须使符合本建议书规定的解码器能解码给出适合于预定应用的音频信号。

注 1 – 遵循 ISO/IEC 11172-3 (1993) 附件 C 和附件 D 中技术说明的编码器，能满足最低的性能标准。

下面，说明如图 1 所示的典型编码器。输入的音频样本传送到编码器上，由时域到频域的映射形成经滤波和亚取样的输入音频流信号。映射后的样本可以是子带样本（如层 I 和层 II 中，参见下述），或者是变换后的子带样本（如层 III 中）。应用与音频信号的时域到频域映射并行运算的快速傅里叶变换，由心理声学模型产生一个数据集以控制量化和编码。取决于实际编码器的实施方案，这些数据是不相同的。一种可能的方法是应用掩模门限的估计来控制量化器。由“比例缩放、量化和编码”框从映射的输入样本中产生一个编码符号集合。另外，该框的传递函数取决于编码系统的实施方案。“帧包封”框使来自其他框的输出数据（例如，比特分配数据、比例因子和编码的子带样本）中用于所选定层的实际比特流组装在一起，并当需要在附属数据字段中加上其他信息（例如，误码保护）。

图1
典型编码器方框图



BS 1196-01

2 分层

依据应用场合，可以采用具有增大复杂性和增高性能的不同编码系统层级。

层I: 该层包含将数字音频输入映射入基本的32个子带；固定的分段使数据格式化成为数据块；用心理声学模型确定自适应的比特分配以及利用数据块的压扩和格式化进行量化。一个层I帧表征每声道384个样本。

层II: 该层对比特分配、比例因子和样本提供附加的编码。一个层II帧表征每声道 $3 \times 384 = 1152$ 个样本。

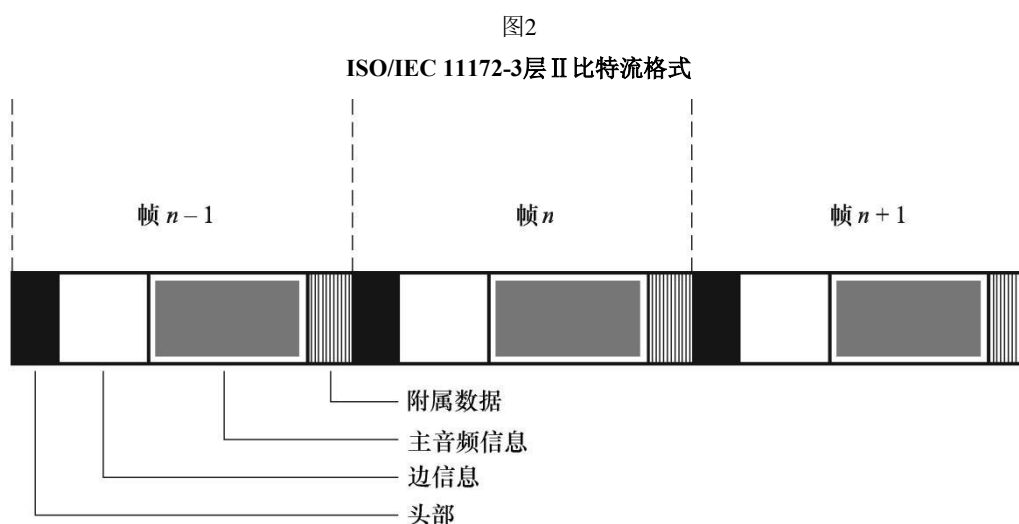
层III: 该层中基于混合滤波器组（带有可变长度修正离散余弦变换的32子带滤波器组）引入增大的频率分辨率。它增加了非均匀量化器、自适应分段和量化值熵编码等处理。一个层III帧表征每声道1152个样本。

任一层可以有四种不同的模式：

- 单声道；
- 双声道（两个独立的音频信号在一个比特流内编码，例如双语种场合）；
- 立体声（立体声对的左、右信号在一个比特流内编码）；
- 联合立体声（立体声对的左、右信号在一个比特流内编码，带有开发的立体声不相关性处理和冗余处理）。可利用联合立体声模式来在低比特率上提高音频质量和/或降低立体声信号的比特率。

3 编码比特流格式

ISO/IEC 11172-3 层 II 比特流的概况如图 2 所示，层 III 比特流的概况如图 3 所示。编码比特流由连续帧构成。依据层级，一个帧内包括如下字段：

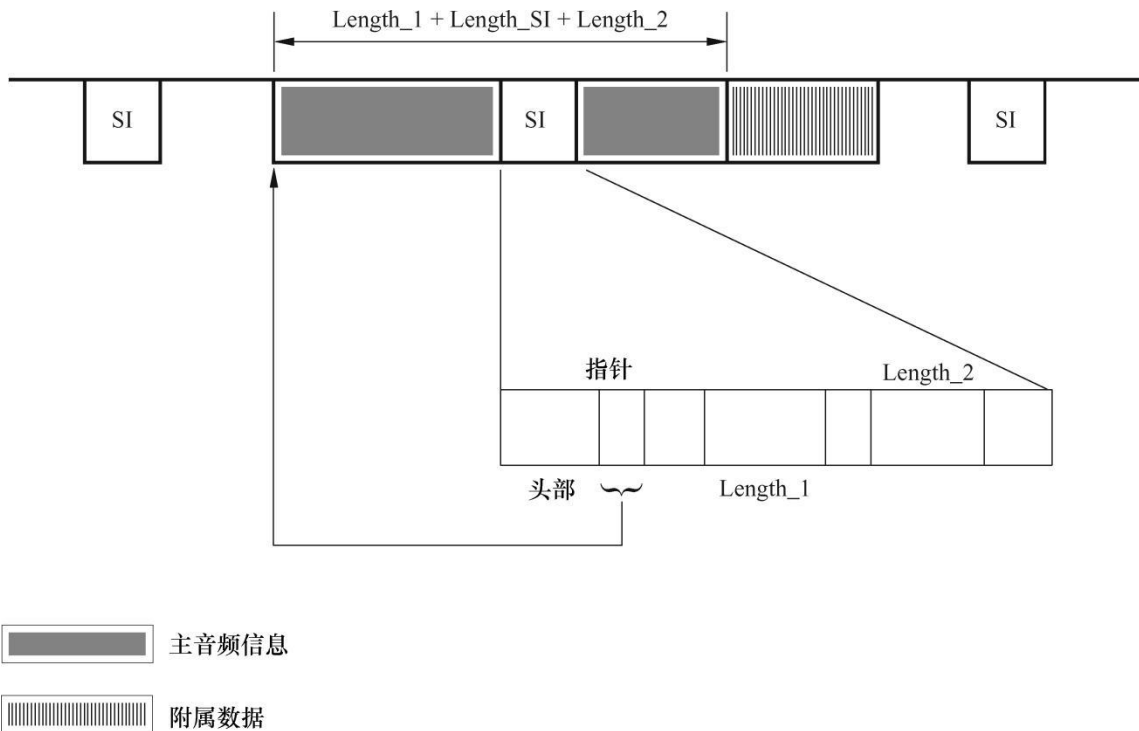


层 II:

- 头部： 比特流内包含同步和状态信息的部分
- 边信息： 比特流内包含比特分配和比例因子信息的部分
- 主音频信息： 比特流内包含编码的子带样本的部分
- 附属数据： 比特流内包含用户可定义数据的部分

图3

ISO/IEC 11172-3层III比特流格式



层 III:

边信息(SI):	比特流中包含头部、指针、length_1和length_2及比例因子信息等的部分;
头部:	比特流中包含同步和状态信息的部分;
指针:	指明主音频信息的开始;
Length_1:	主音频信息第一部分的长度;
Length_2:	主音频信息第二部分的长度;
主音频信息:	比特流中包含编码音频的部分;
附属数据:	比特流中包含用户可定义数据的部分。

BS.1196-03

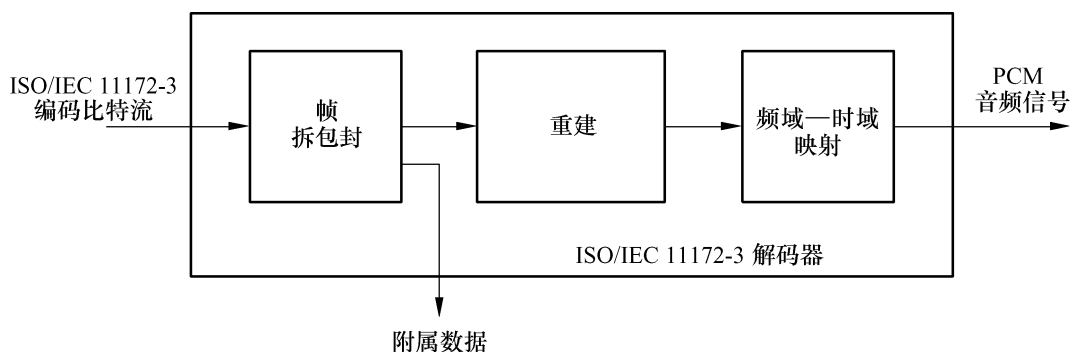
4 解码

解码器接收在 ISO/IEC 11172-3 中定义的句法内的编码音频比特流，解码出数据元素，并应用数据信息产生数字音频输出。

编码的音频比特流传送入解码器。如果在编码器内加入误码校验，则比特流拆包封和解码处理时视情实施误码检测。对比特流拆包封以恢复各个信息段，诸如音频帧头部、比特分配、比例因子、映射的样本和可选的附属数据等信息。重建处理中，重建映射的样本集中的量化值。频域到时域的映射使这些映射的样本变换回线性 PCM 音频样本。

图4

解码器方框图



BS.1196-04

附件2 (资料性)

MPEG-2和MPEG-4 AAC音频

1 引言

ISO/IEC 13818-7 说明了被称为 MPEG-2 高级音频编码 (AAC) 的 MPEG-2 音频非后向兼容标准，它是一种与要求 MPEG-1 后向兼容相比能达到更高质量的多声道标准。

为了能够在所需存储器和处理能力以及音频质量之间做到兼顾，AAC 系统由三种档型组成：

— 主型

在任一给定的数据率上，主型提供最高的音频质量。可应用除增益控制之外的所有工具以给出高的音频质量。与 LC 型相比，它需要的存储器和处理能力更高。主型解码器能解码 LC 型编码的比特流。

— 低复杂度 (LC) 型

LC 型所需的处理能力和存储器小于主型，但质量性能可保持较高。LC 型中没有预测器和增益控制工具，但具有阶次有限的时间噪声成形 (TNS)。

— 取样速率可缩放 (SSR) 型

SSR 型中借助增益控制工具能提供取样速率可缩放的信号。它能选择频段进行解码，因而解码器需要较少的硬件。例如，在 48 kHz 取样频率下为了只解码最低的频段，解码器能够以最小的解码复杂度重现 6 kHz 带宽的音频信号。

AAC 系统能支持如表 1 中所示的、从 8 到 96 kHz 取样频率范围内的 12 种类型，并可高达 48 路音频声道。表 2 显示了默认声道配置，它们包括单声道、双声道、5 声道 (3 个前方声道/2 个背后声道) 和 5 声道加低频效果 (LFE) 声道 (带宽 < 200 Hz) 等。除默认配置外，还可能在每个位置 (前方、侧面和背后) 处规定扬声器的数目，以做到灵活的多声道扬

声器布局。还支持缩混能力，用户能指定一个系数，将多声道音频信号缩混到双声道上。因此，能够控制仅使用两声道的重放装置的声音质量。

表1
支持的取样频率

取样频率 (Hz)
96 000
88 200
64 000
48 000
44 100
32 000
24 000
22 050
16 000
12 000
11 025
8 000

表2
默认的声道配置⁽¹⁾

数值 ⁽²⁾	扬声器数	接收指令中列出的音频句法元素	扬声器映射的默认元素 ⁽³⁾	ITU-R BS.775或ITU-R BS.2051建议书规定的声道名称 ⁽⁴⁾
1	1	single_channel_element	M+000	单声道
2	2	channel_pair_element	M+030, M-030	左、右
3	3	single_channel_element()	M+000	中央
		channel_pair_element()	M+030, M-030	左、右
4	4	single_channel_element()	M+000	中央
		channel_pair_element()	M+030, M-030	左、右
		single_channel_element()	M+180	单声道环绕
5	5	single_channel_element()	M+000	中央
		channel_pair_element()	M+030, M-030	左、右
		channel_pair_element()	M+110, M-110	左环绕、右环绕
6	5 + 1	single_channel_element()	M+000	中央
		channel_pair_element()	M+030, M-030	左、右
		channel_pair_element()	M+110, M-110	左环绕、右环绕
		lfe_element()	LFE1	低频效果

表2 (续)

数值 ⁽²⁾	扬声器数	接收指令中列出的音频句法元素	扬声器映射的默认元素 ⁽³⁾	ITU-R BS.775或ITU-R BS.2051建议书规定的声道名称 ⁽⁴⁾
7	7 + 1 前	single_channel_element()	M+000	N/A ⁽⁵⁾
		channel_pair_element()	M+030, M-030	
		channel_pair_element()	M+045, M-045	
		channel_pair_element()	M+110, M-110	
		lfe_element()	LFE1	
8-10	—	—	预留	—
11	6 + 1	single_channel_element()	M+000	N/A
		channel_pair_element()	M+030, M-030	
		channel_pair_element()	M+110, M-110	
		single_channel_element()	M+180	
		lfe_element()	LFE1	
12	7 + 1 后	single_channel_element()	M+000	中央
		channel_pair_element()	M+030, M-030	左、右
		channel_pair_element()	M+090, M-090	左侧环绕、右侧环绕
		channel_pair_element()	M+135, M-135	左后环绕、右后环绕
		lfe_element()	LFE1	低频效果
13	22 + 2	single_channel_element()	M+000	前中央
		channel_pair_element()	M+030, M-030	左前中央、右前中央
		channel_pair_element()	M+060, M-060	左前、右前
		channel_pair_element()	M+090, M-090	左侧方、右侧方
		channel_pair_element()	M+135, M-135	左后、右后
		single_channel_element()	M+180	后中央
		lfe_element()	LFE1	低频效果-1
		lfe_element()	LFE2	低频效果-2
		single_channel_element()	U+000	前中
		channel_pair_element()	U+045, U-045	左前上、右前上
		channel_pair_element()	U+090, U-090	顶部左侧、顶部右侧
		single_channel_element()	T+000	顶部中央
		channel_pair_element()	U+135, U-135	顶部左后、顶部右后
		single_channel_element()	U+180	中后上
		single_channel_element()	B+000	中前下
		channel_pair_element()	B+045, U-045	左前下、右前下

表2 (结束)

数值 ⁽²⁾	扬声器数	接收指令中列出的音频句法元素	扬声器映射的默认元素 ⁽³⁾	ITU-R BS.775或ITU-R BS.2051建议书规定的声道名称 ⁽⁴⁾
14	7 + 1顶部	single_channel_element()	M+000	中央
		channel_pair_element()	M+030, M-030	左、右
		channel_pair_element()	M+110, M-110	左环绕、右环绕
		lfe_element()	LFE1	低频效果
		channel_pair_element()	U+030, U-030	顶部左前、顶部右前
15	—	—	预留	—

(1) 该列表引用自ISO/IEC 14496-3:2009/Amd.4:2013的表1.19。

(2) 音频输出声道配置用ISO/IEC 23001-8:2013“独立于编码的代码点”规定的、承载声道配置值的四比特字段来表示。MPEG-2适用于不超过7的声道配置值。MPEG-4 AAC适用于不超过15的声道配置值。

(3) 通过符合ITU-R BS.2051建议书的标签来识别扬声器。

(4) 声道标签和名称取决于实际的声道配置。

(5) N/A: 不适用；在ITU-R BS.2051和ITU-R BS.775建议书中未规定声道配置。

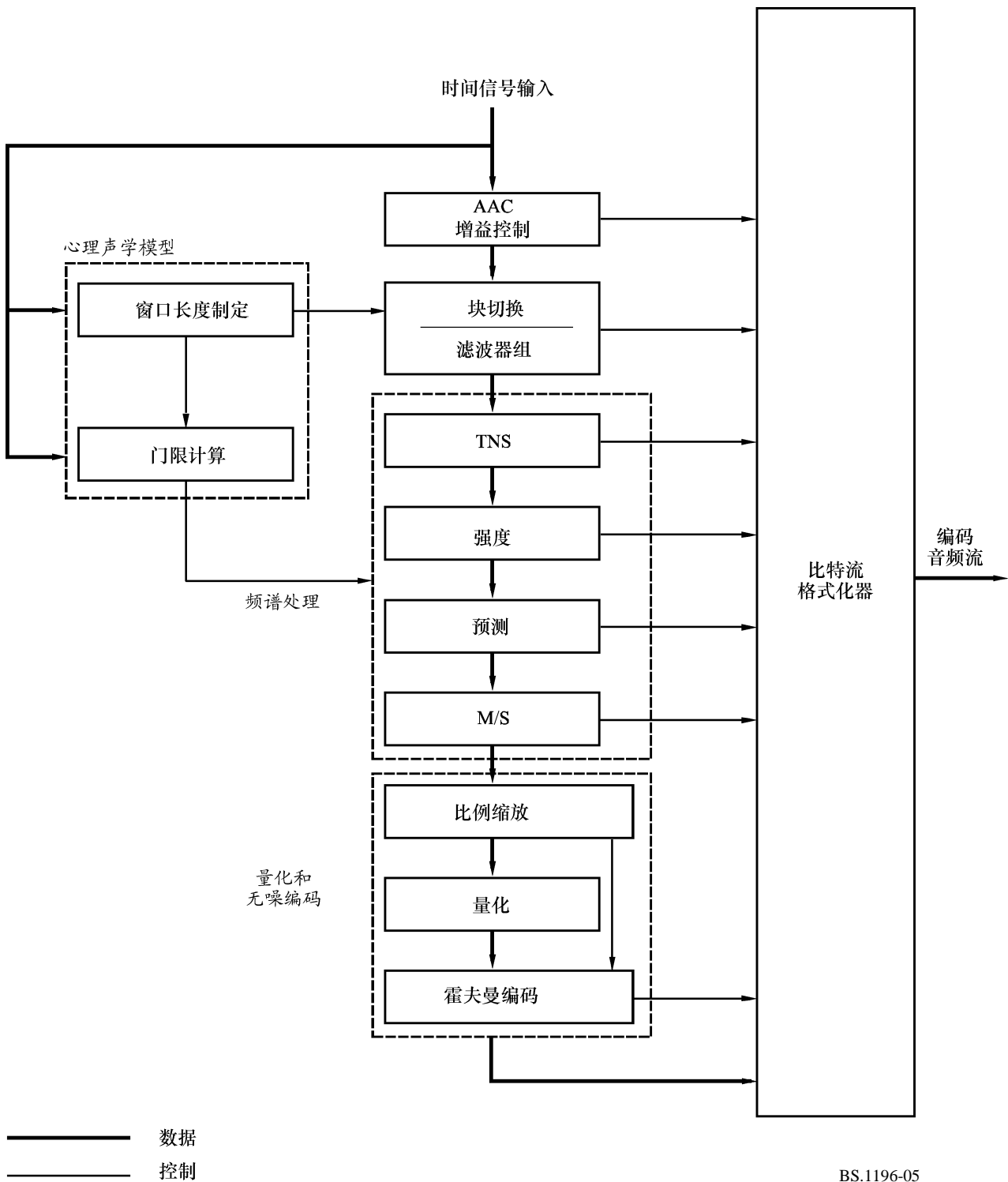
2 编码

MPEG-2 AAC 编码器的基本结构如图 5 所示。AAC 系统由下面的编码工具构成：

- 增益控制：增益控制中，将输入信号分割成4个等间隔的频段。增益控制用于SSR型。
- 滤波器组：修正的离散余弦变换（MDCT）滤波器组将输入信号分解成亚取样的频谱分量，以48 kHz取样时，频率分辨率为23 Hz，时间分辨率为21.3毫秒（ms）（128个频谱分量），或者频率分辨率为187 Hz、时间分辨率为2.6 ms（1024个频谱分量）。窗口形状在两个可选的窗口形状之间选择。
- 时间噪声成形（TNS）：经滤波器组解析之后，实施TNS运算。TNS技术能使编码器对量化噪声在时间上做出精细结构控制。
- 中间/边侧（M/S）立体声编码和强度立体声编码：对于多声道音频信号，可应用强度立体声编码和M/S立体声编码。在强度立体声编码中，只传输声能包络，以减少传输的方向信息。在M/S立体声编码中，传输归一化的和信号M（M代表中间）及差信号S（S代表边侧）而非传输原来的左与右信号。
- 预测：为降低静止信号的冗余，在后续帧的亚取样频谱分量之间实施时域预测。
- 量化和无噪编码：在量化工具中，应用步进大小为1.5分贝（dB）的非均匀量化器。对量化的频谱、不同的比例因子和方向信息，采用霍夫曼编码。

- 比特流格式化器：最后，使用比特流格式化器来复用比特流，其中包括量化和编码的频谱系数以及来自每一工具的某些附加信息。
- 心理声学模型：使用心理声学模型自输入信号计算当前的掩模门限。可采用类似于ISO/IEC 11172-3中心理声学模型2的心理声学模型。在量化处理过程中，使用自掩模门限和输入信号电平导出的信号-掩模比，以便最小化可闻的量化噪声，并附加地借以选择适当的编码工具。

图5
MPEG-2 AAC编码器方框图



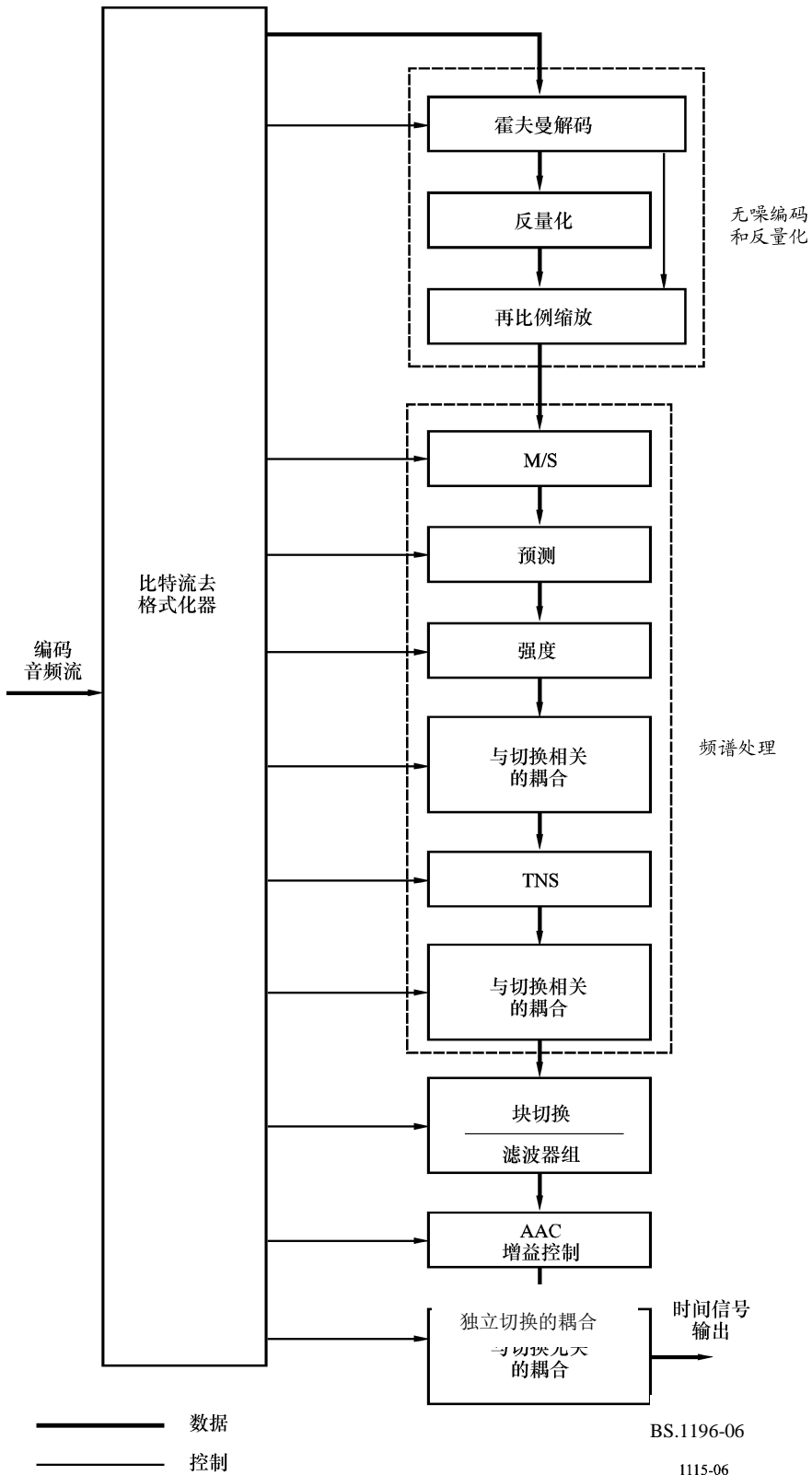
BS.1196-05

3 解码

MPEG-2 AAC 解码器的基本结构如图 6 所示。解码处理基本上是编码处理的逆过程。

图6

MPEG-2 AAC解码器方框图

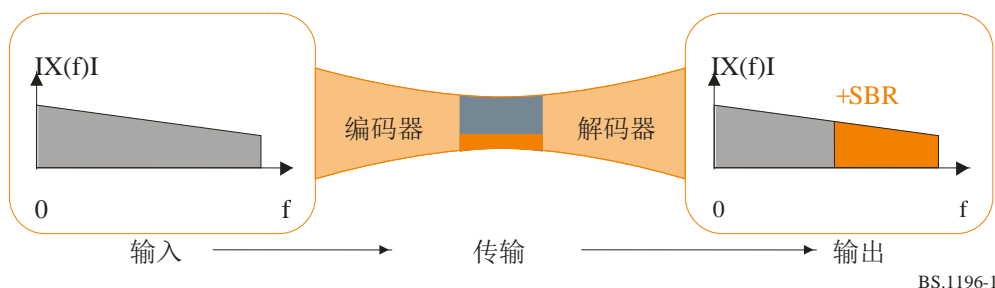


解码器的功能是寻找出比特流中对量化音频频谱的描述，解码其量化值和其他重建信息，重建量化的频谱，借助比特流中凡是起作用的工具对重建的频谱进行处理，以便得到输入比特流中描述的实际信号频谱，并最终将频域频谱变换为时域频谱，处理中可以用或不用可选的增益控制工具。在初始的频谱重建和比例缩放频谱重建后，有许多可选的工具可用来修正一个或多个频谱，以便提供更高效率的编码。对工作于频域中的每一个可选工具，保留其“旁通”选项，在省略频谱运算的所有场合下，其输入端上的频谱不做修正地直接通过工具。

4 高效AAC和频段复制

高效 AAC (HE AAC) 引入了频段复制 (SBR)。SBR 是音频压缩算法中一种高效的高频编码方法。通过以既定的比特率增加音频段宽，或者以既定的质量水平提高编码效率，它使低比特率音频和语音编解码器的性能得以改进。

被编码和传输的部分仅是频谱中较低的部分，这是人耳最为敏感的那部分频谱。SBR 不发送频谱中较高的部分，而是基于对所发送较低频率的分析，重建较高频率的一个解码后的过程。通过在所编码比特流中以很低的数据速率发送 SBR 相关的参数，确保准确的重建。

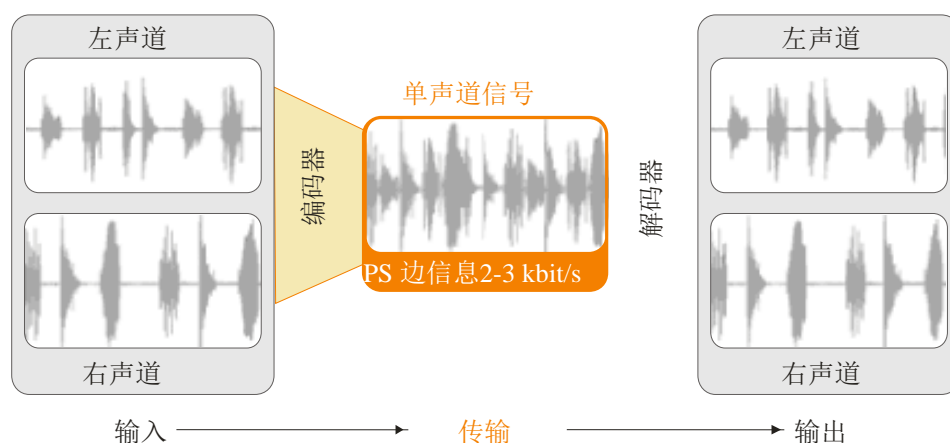


HE AAC 比特流是 AAC 音频比特流的增强型。额外的 SBR 数据嵌入 AAC 填充元素中，从而保证与 AAC 标准的兼容性。HE AAC 技术是一个双重速率系统。后向兼容的普通 AAC 音频比特流以半 SBR 增强型取样速率运行，因而，无法对 SBR 增强数据进行解码的 AAC 解码器将以 HE AAC 解码器生成之取样速率的半倍速率生成一个输出时间信号。

5 高效AAC第2版和参数立体声

HE AAC 第 2 版是 HE AAC 的扩展版，并引入了参数立体声 (PS)，以增强低比特率立体声信号的音频压缩效率。

编码器对立体声音频信号进行分析，然后构建立体声图像的参数表达式。现在，不需要发送两个信道，仅对原立体声信号的单声道表达式进行编码。该信号与重建立体声图像所需的参数一起发送。



BS.1196-2

结果是，与没有参数立体声的类似比特流相比，含有参数立体声的低比特率音频比特流（例如，24 kbit/s）的可感知音频质量得到大幅提升。

HE AAC 第 2 版比特流以 HE AAC 比特流为基础。增加的参数立体声数据嵌入单声道 HE AAC 流的 SBR 扩展元素中，从而保证与 HE AAC 及 AAC 的兼容性。

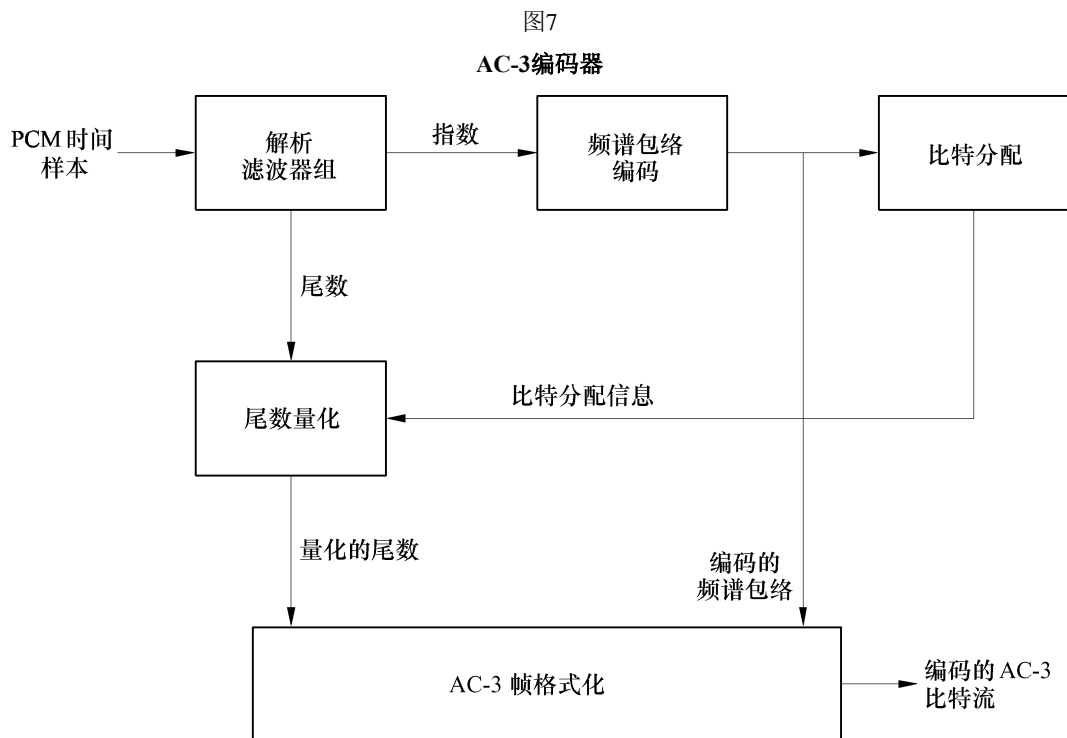
无法对参数立体声增强解码的 HE AAC 解码器以全带宽产生一个单声道输出信号。无法对 SBR 增强数据解码的普通 AAC 解码器以半倍取样速率生成一个单声道输出时间信号。

附件3 (资料性)

AC-3和E-AC-3音频

1 编码

AC-3 数字压缩算法能够将来自 PCM 表示的 1 至 5.1 声道的音频源编码成数据率范围为 32 kbit/s 至 640 kbit/s 的串行比特流。通过对音频信号的频域表示粗略量化，AC-3 算法能获得高的编码增益（输入比特率与输出比特率之比）。AC-3 算法处理过程方框图如图 7 所示。编码处理中的第一步是将来自 PCM 时间样本序列的音频表示变换成频率系数的数据块序列，由解析滤波器组完成这一变换。对 512 个时间样本的重叠数据块乘以一个时间窗口，并变换到频域上。由于数据块重叠，每个 PCM 输入样本在两个相继的变换块中得到表示。于是，频域表示将大幅下降 2 个级数，因此，每个数据块包含 256 个频率系数。各个频率系数均以二进制指数标记法表示成二进制的指数和一个尾数。指数集合编码成信号频谱的粗略表示，它被称为频谱包络。该频谱包络供核心比特分配程序使用，它决定以多少比特位来对每个单个的尾数进行编码。将 6 个音频数据块（1 536 个音频样本）的频谱包络和粗略量化的尾数格式化成 AC-3 帧。AC-3 比特流是由 AC-3 帧构成的一个序列。



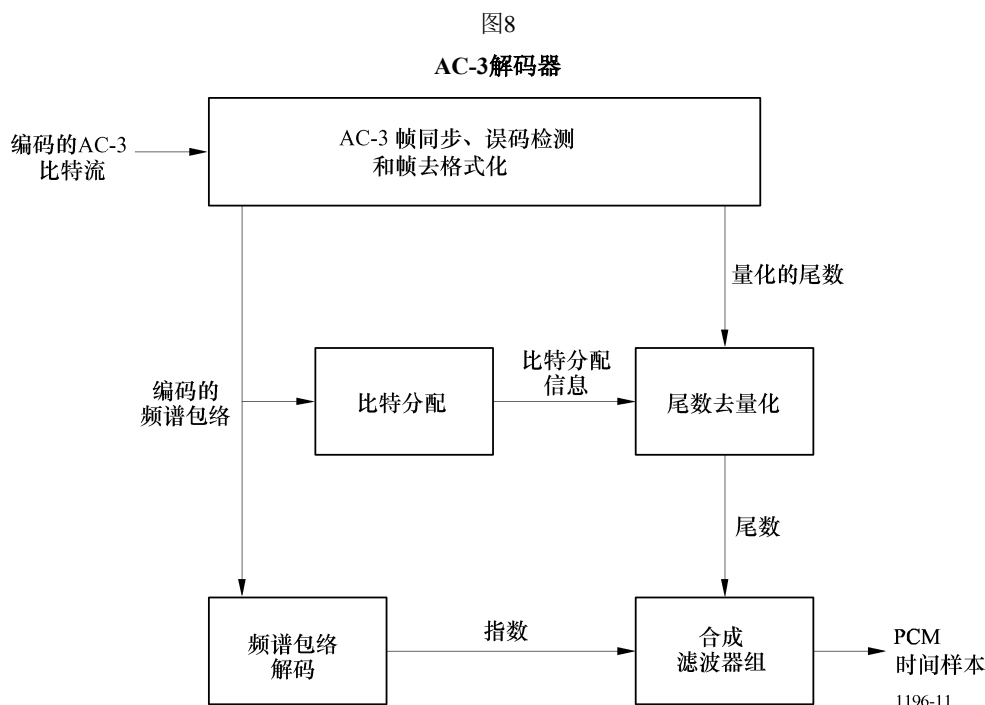
BS.1196-07

实际的 AC-3 编码器要比图 7 中示出的复杂，它还包括上面未示出的以下功能：

- 附上一个帧头，当中包含同步和解码已编码比特流所需的信息（比特率、取样速率和编码声道数等）；
- 插入误码检测码，以使解码器能验证接收到的数据帧无误码；
- 解析滤波器组的频谱分辨率可动态地改变，以便能更好地适配每个音频数据块的时间/频率特性；
- 频谱包络可以用可变的时间/频率分辨率进行编码；
- 可以执行更复杂的比特分配，并修正核心比特分配程序的参数，以便生成更优化的比特分配；
- 在高频上各声道可以耦合在一起，以便在较低比特率上运行时能达到较高的编码增益；
- 在双声道模式中，可以选择性地执行再矩阵处理，以便提供附加的编码增益，并在用矩阵环绕声解码器解码双声道信号的场合下能得到经改善的结果。

2 解码

解码处理基本上是编码处理的逆过程。图 8 中所示的解码器必须与经编码的比特流同步，校验误码，以及对各种类型的数据（如经编码的频谱包络和经量化的尾数）实施去格式化。运行比特分配程序，并使用其结果进行拆包和尾数去量化。解码频谱包络以得到指数。将指数和尾数变换回时域内，以产生经解码的 PCM 时间样本。



BS.1196-08

实际的 AC-3 解码器要比图 8 中示出的复杂，还包括上面未示出的以下功能：

- 在检测到数据出错的场合下，可采用误码隐匿或误码抹去处理；
- 对其高频内容耦合在一起的各声道，必须进行去耦合；
- 无论何时当对声道实施了再矩阵处理时，都必须进行解矩阵（在双声道模式中）；
- 必须以编码处理中动态改变编码解析滤波器组的同样方式，动态地改变合成滤波器组的分辨率。

3 E-AC-3

增强型 AC-3（E-AC-3）为上述基本型 AC-3 编解码器增加了若干额外的编码工具和特征。其额外的编码工具提供了更好的编码效率，允许以更低的比特率运行，同时额外的特征提供了更多的应用灵活性。

额外的编码工具：

- 自适应混合转换 – 用于分析/合成滤波器组的额外层，用于提供更精细的（AC-3的 1/6）频谱分辨率。
- 瞬时预噪声处理 – 减少瞬时预噪声的额外工具。
- 频谱扩展 – 基于编码器创建之边信息的、最高频率组件的解码器合成。
- 耦合增强 – 处理声道耦合中的相位和振幅。

额外的特征：

- 更精细的数据率粒度。
- 更高的最大数据率（3 Mbit/s）。
- 子数据流可承载更多的音频声道，如 7.1 声道或解说声道。

附件4 (资料性)

MPEG环绕

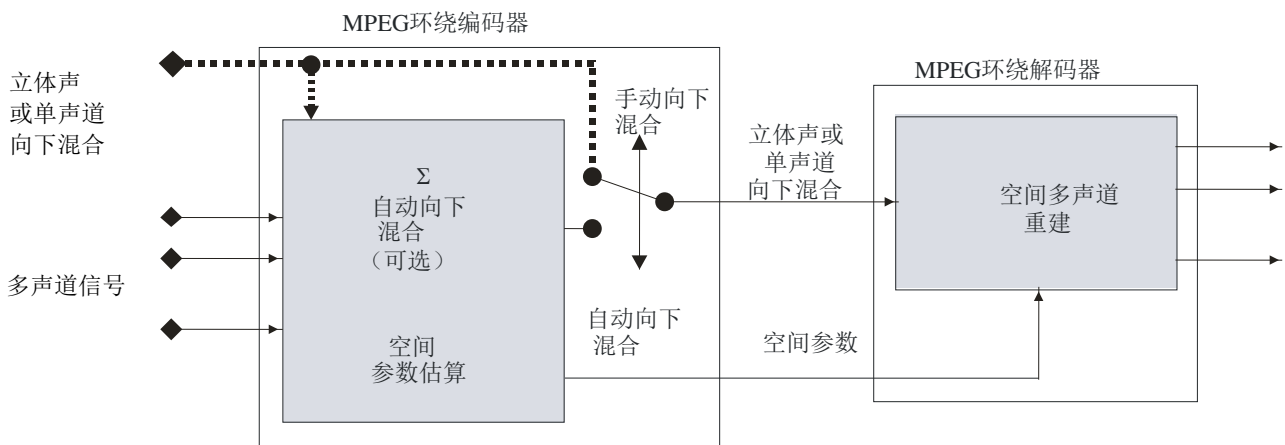
1 引言

ISO/IEC 23003-1 或 MPEG 环绕技术提供一个极为有效的多声道声音编码方法，并允许以通常用于单声道或立体声声音编码的比特率传输环绕声音。它能够基于 $M < N$ 声道向下混合和额外的控制数据表示一个 N 声道多声道音频信号。在优选运行模式中，MPEG 环绕编码器从多声道音频输入信号中产生一个单声道或立体声向下混合。该向下混合利用一个标准的核心音频编解码器进行编码，例如，建议 1 和建议 2 中建议的一种编码系统。除向下混合外，MPEG 环绕还可产生一种对多声道音频的空间图像参数描述，以后向兼容的形式添加为核心音频编解码器的一个附属数据流。原有的单声道或立体声解码器将忽略附属数据并回放立体声或单声道向下混合音频信号。带有 MPEG 环绕功能的解码器首先对单声道或立体声向下混合进行解码，然后使用从附属数据流采集的空间图像参数来生成高质量的多声道音频信号。

图 9 说明了 MPEG 环绕技术的原则。

图9

MPEG环绕技术的原则，使用核心音频编解码器对向下混合进行编码



BS.1196-09

通过使用 MPEG 环绕技术，可对现有业务进行轻松升级，从而以后向兼容的方式实现环绕。虽然现有的老式消费设备中的立体声解码器可忽略 MPEG 环绕数据，并可回放立体声信号而不造成质量下降，然而，带有 MPEG 环绕功能的解码器将提供高质量的多声道音频。

2 编码

MPEG 环绕编码器的目的是将多声道输入信号表示为一个后向兼容的单声道或立体声信号，同时与有利于多声道输出重建的空间参数结合，从感知角度看，这类似于原始多声道输

入信号。除自动生成的向下混合之外，还可使用外部创建的向下混合（“人工向下混合”）。该向下混合将保存输入声音的空间特征。

MPEG 环绕技术与 HE-AAC 结合的参数立体声技术的基础上，产生 HE AAC 第 2 版标准规范。通过结合多个参数立体声模块和其他新开发的模块，定义了各种支持不同输出和向下混合声道数目组合的结构。举例来说，对于一个 5.1 多声道输入信号，可有三种不同的配置：一种基于立体声向下混合的系统的配置（525 配置），以及两种不同的、基于单声道向下混合的系统的配置（采用不同框串联的 515₁ 和 515₂ 配置）。

MPEG 环绕技术含有若干富有特色的工具，从而能够实现对标准的广泛应用。MPEG 环绕技术的一个关键特色是逐渐将空间图像质量从非常低的附属数据空间扩展至透明。另一个关键特色是解码器输入可与现有的矩阵式环绕技术兼容。

这些及其他特色是通过以下非同寻常的编码工具实现的：

- 残余编码：除空间参数外，通过混合编码技术还可传递残余信号。这些信号代替了部分去关联化信号（属于参数立体声框的一部分）。通过将 QMF 域信号转换为 MDCT 域来对残余信号进行编码，然后利用 AAC 对 MDCT 系数进行编码。
- 矩阵兼容：作为一种选项，可对立体声向下混合进行预处理，以便与原有的矩阵式环绕技术兼容，确保与解码器的后向兼容，解码器仅能对立体声比特流进行解码，但配备有矩阵式环绕解码器。
- 任意向下混合信号：MPEG 环绕系统不仅能处理编码器产生的向下混合，而且还能处理在多声道原始信号之外的、提供给编码器的人工向下混合。
- PCM 上的 MPEG 环绕：一般情况下，MPEG 环绕空间参数承载于基本音频压缩方案的附属数据部分。对于向下混合被作为 PCM 传输的应用，MPEG 环绕技术还支持一种允许空间参数通过未压缩音频声道承载的方法。这种基本技术被称为掩埋数据。

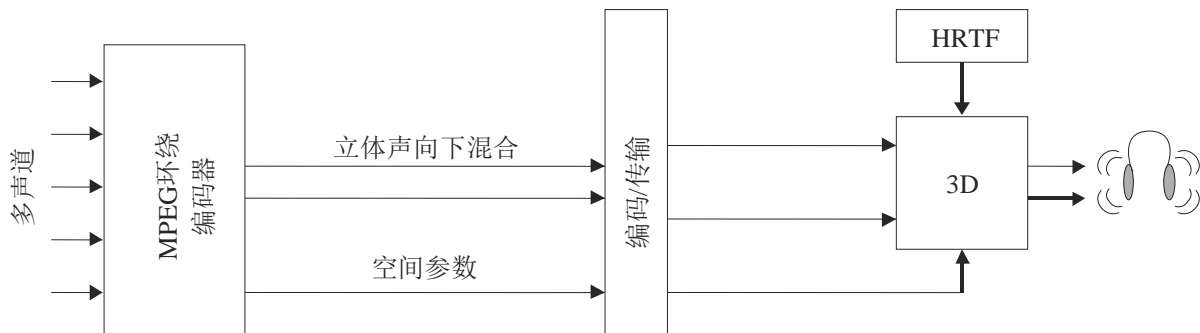
3 解码

除提供多声道输出外，MPEG 环绕解码器还支持其他输出配置：

- 虚拟环绕：MPEG环绕系统可利用空间参数将向下混合提供给立体声虚拟环绕输出，以便在老式耳机上进行回放。该标准未详细说明头部相关传输函数（HRTF），而仅说明了这些HRTF的接口，因而可根据使用情况自行采用。虚拟环绕处理既可用于解码器，也可用于编码器，后者提供了有关向下混合的虚拟环绕经验的可能性，而不必配备MPEG环绕解码器。不过，MPEG环绕解码器可在向下混合中取消虚拟环绕处理，并重新应用另外一种虚拟环绕。图10概括了其基本原则。

图10

MPEG环绕的虚拟环绕解码过程



BS.1196-10

- 增强型矩阵模式：如果是原有的立体声内容，其中未显示空间边信息，那么MPEG环绕技术能够从向下混合中估算出空间边信息，从而创建多声道输出，但可提供超出传统矩阵环绕系统的质量。
- 剪裁：作为基本结构的一个结果，MPEG环绕解码器可将其输出提供给声道数量低于编码器多声道输入声道数量的声道配置。

4 档次和级别

MPEG环绕解码器的应用分为高质量版本和低功率版本。尽管会产生不同的输出信号，但两种版本都在相同的数据流上运行。

MPEG环绕的基本档次定义了六种不同的级别，分别具有不同数量的输入和输出声道，采用不同范围的取样速率，并适用于不同的残余信号解码带宽。解码器的级别必须等于或大于比特流的级别，以确保正常解码。此外，第1、2和3级解码器能够对第2、3和4级的所有比特流进行解码，但因解码器的局限，质量可能会稍有下降。MPEG环绕解码器的输出质量和格式还取决于特定解码器的配置。不过，解码器的配置方面问题与该档次的不同级别完全有关。

5 与音频编解码器之间的关联

MPEG环绕技术在老式音频编码方案的基础上作为预处理和后处理扩展而运行。因此，它具备适合几乎任何核心音频编码器的方法。MPEG环绕技术的成帧极为灵活，以确保与各种类型的编码器保持同步，并意味着对与已使用参数工具之编码器（例如，频段复制）的连接进行优化。

附件5 (资料性)

扩展版高效AAC (扩展版HE AAC)

1 引言

ISO/IEC 23003-3 MPEG-D统一语音和音频编码 (USAC) 规定了扩展版HE AAC档次。USAC是一项音频编码标准, 允许对语音、音频或任何语音和音频的混合体进行编码, 并且确保在范围广泛的比特率上全部声音素材音频质量保持一致。该标准支持高比特率的单声道和多声道编码, 提供感知上透明的质量。同时, 该标准亦允许以非常低的比特率进行非常高效的编码, 并且保留全部的音频带宽。

之前的音频编解码器在对语音或音频内容进行编码时具有特定的优势和弱点, 而USAC能够在编码时确保所有内容具有同等的高保真度, 无论内容类型如何。

为实现同样良好的编码音频和语音质量, USAC采用了基于经证实的改进的离散余弦变换 (MDCT) 的MPEG-4音频 (MPEG-4 AAC、HE AAC、HE AAC v2) 编码技术, 并将其与代数编码激励线性预测 (ACELP) 等专用语音编码器元素相结合。MPEG-4频段复制 (SBR) 和MPEG-D MPEG环绕等参数编码工具得以增强, 并紧密地集成到编解码器中。其结果实现了高效编码, 并能够以最低的比特率工作。

目前, USAC标准规定了两种档次:

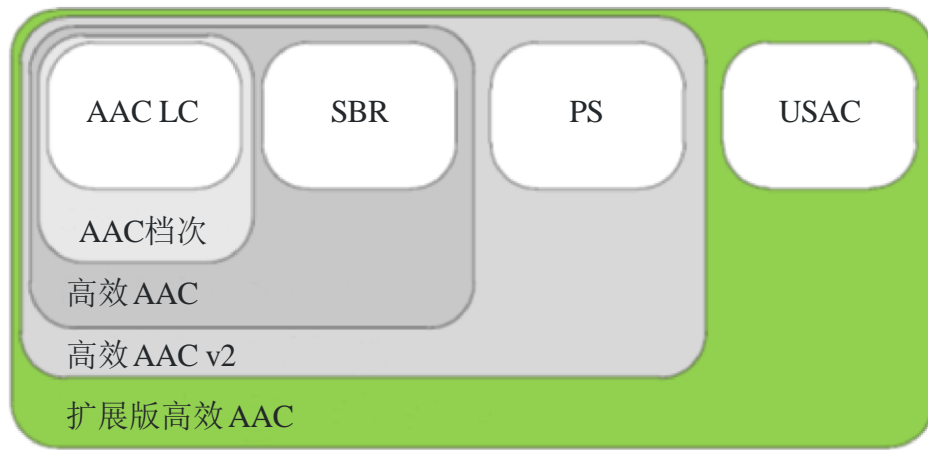
– 基准USAC档次

基准USAC档次提供USAC标准的完整功能, 同时保持较低的整体计算复杂性。对存储或处理能力需求过大的工具被排除在外。

– 扩展HE ACC档次

专门针对需要与现有AAC系列档次 (AAC、HE AAC和HE AAC v2) 保持兼容的应用, 该档次通过增加USAC功能, 对现有的HE AAC v2档次进行扩展。该档次包括基准USAC档次的第2级。因此, 扩展版HE AAC档次解码器可以对所有的HE AAC v2比特流以及USAC比特流 (多达两个声道) 进行解码。

图11
扩展版高效AAC的结构



BS.1196-11

USAC支持7.35 kHz至96 kHz范围内的取样频率，并显示能够在8 kbit/s至实现感知透明的比特率范围内提供良好的音频质量。这已在ISO/IEC JTC 1/SC 29/WG 11的验证测试中得到证实（MPEG2011/N12232号文件，附于6B/286（Rev.2）号文件）。

声道配置可自由选择。可以有效地为最常见的应用场景发出13种不同的缺省声道配置信号。这些缺省配置包括所有的MPEG-4声道配置，例如单声道、立体声、5.0和5.1环绕立体声或者甚至7.1或22.2扬声器的设置。

2 编码

作为MPEG标准化中的常用标准，ISO/IEC 23003-3标准仅规定了MPEG-D USAC文件和数据流的解码过程。该标准未明确规定编码过程的标准。

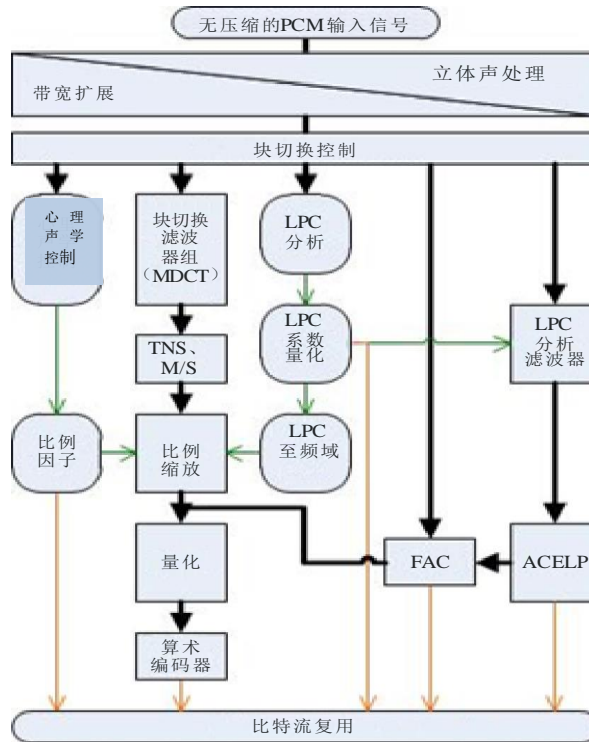
图12给出了一个典型的、可能的编码器结构。

编码器由以下编码工具组成：

- 立体声处理：在中/低比特率上，USAC采用参数立体声编码技术。这些技术原则上与ISO/IEC 23003-3附录2.5所述的PS工具类似，但基于附件4所述的MPEG环绕立体声，因此被称作MPEG环绕2-1-2（MPS 2-1-2）。编码器从输入音频信号中提取立体声图像的高效参数表示。这些参数与单声道向下混合信号一起，在比特流中进行发送。编码器可以选择发送一个残余信号，以修正解码器端的立体声信号重建过程。残余编码机制允许从全参数到全离散声道立体声编码的平滑缩放。MPS 2-1-2工具是USAC编解码器的固有组成部分。在比特率更高时，其中参数化编码和ACELP通常并不活跃，可借助复数值立体声预测，在MDCT域内完全独立地执行立体声编码。因此，该方法被称作复数预测立体声编码。这可以看作是传统M/S立体声编码的泛化形式。

- 带宽扩展：参数化带宽扩展是MPEG-4频段复制（SBR）的多重增强版本，这在ISO/IEC 23003-3附录2.4中予以描述。编码器对更高音频频段的频谱包络和音调进行估算，并向解码器发送对应的参数。编码器可以从两个不同的变调器类型（校音或复制）以及三个换位因子（1:2、3:8、1:4）中进行选择。增强型SBR工具是USAC编解码器的固有组成部分。
- 滤波器组、块切换：基于MDCT的滤波器组构成核心编码器的基础。根据所采用的量化噪声整形机制，可以选择从1024、512、256或128频谱线中选择一个变换分辨率。通过与3:8 SBR换位因子相结合，分辨率可更改成上述备选值的 $\frac{3}{4}$ ，即使在较低的取样速率上也能提供更好的时间粒度。
- 时域噪声整形（TNS）、M/S立体声编码、量化：已从AAC中采用这些工具，使用方式类似于ISO/IEC 23003-3附录2.2中所述的方式。
- 上下文自适应算术编码器：MDCT频谱系数的无噪声（即熵）编码由算术编码器进行处理，该编码器根据先前已编码的频谱线来选择概率表。
- 心理声学控制、比例因子缩放：基于心理声学模型的比例因子与AAC中使用的比例因子类似，请参阅ISO/IEC 23003-3附录2.2。
- 基于线性预测编码（LPC）参数的缩放：该频谱噪声整形工具可用于替代上述比例因子缩放。LPC滤波器系数集的频率表示的加权版本在进行量化和编码之前应用于MDCT频谱系数。
- ACELP：代数编码激励线性预测（ACELP）编码器工具采用了经过验证的自适应/创新性码书激励表示，即来自最先进的语音编解码器。
- 比特流多路复用：最终的比特流由编码器工具产生的各种元素组成。
- FAC：前向走样校正（FAC）工具提供了一种机制，用于实现从基于受走样影响的MDCT编码到基于时域的ACELP编码的无缝转换。

图12
MPEG-D USAC编码器的方框图



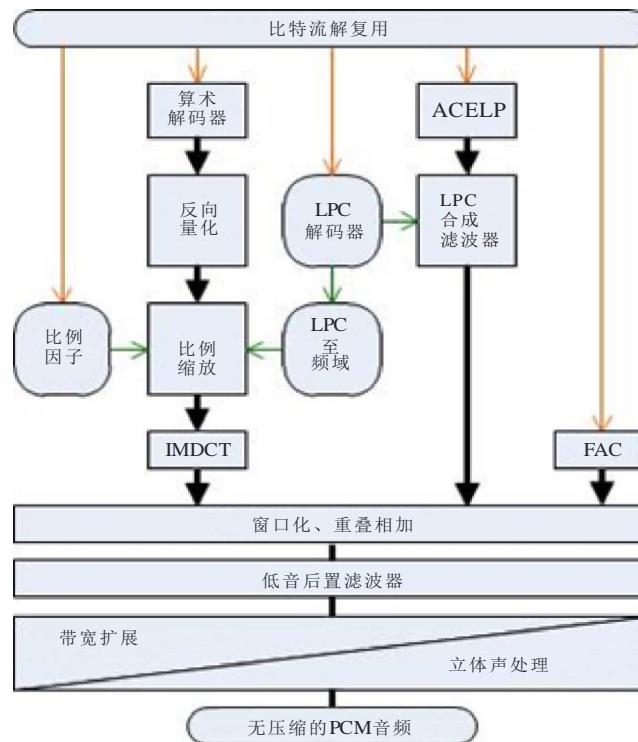
BS.1196-12

3 解码

图13给出了MPEG-D USAC解码器的基本结构。解码处理基本上是编码处理的逆过程。

图13

MPEG-D USAC解码器的方框图



BS.1196-13

解码过程可粗略地概述如下：

- 比特流解复用：解码器在比特流中找到与工具相关的全部信息，并将其转发至相应的解码器模块。
- 核心解码：根据比特流的内容，解码器：
 - 对MDCT频谱系数进行解码和反向量化，应用基于比例因子信息或LPC系数信息的缩放，并进一步应用（可选的）基于MDCT的工具（如果存在此类工具并适用）。最后，应用逆向MDCT，以获得相应的时域信号。
 - 或者，对ACELP相关信息进行解码，产生一个激励信号，并借助LPC滤波器合成一个输出信号。
- 窗口化、重叠相加：核心编码器的后续帧在AAC通常采用的重叠相加过程中被串联或合并。ACELP和基于MDCT的编码之间的转换是通过合并解码后的FAC数据来实现的。
- 低音后置滤波器：可应用一个可选的音调增强滤波器来提高语音质量。
- 带宽扩展、立体声处理：最后，运用适用于带宽扩展和立体声编码工具的参数编码工具来重建全带宽、离散立体声信号。

对于每个可选工具，保留“直通”的选项，并且在省略操作的所有情况下，在其输入端的数据将直接通过该工具，而无须任何修改。

4 档次和级别

MPEG目前定义了两个采用USAC编解码器的档次。

– 基准USAC档次

基准USAC档次包含完整的USAC编解码器，除了少数表现出极度最坏情况计算复杂性的工具。上文没有对这些工具做出说明。该档次为应用和使用案例提供一个明确的独立档次，其中支持AAC档次系列（AAC、HE AAC、HE AAC v2等档次）的功能并不相关。

– 扩展HE AAC档次

扩展高效AAC档次包含高效AAC v2档次的全部工具，因此能够对所有AAC档次系列比特流进行解码。另外，该档次集成了基准USAC档次的单声道/立体声功能。因此，该档次是HE AAC v2档次的自然演进，因为USAC（当以低比特率操作时）的单声道/立体声部分在低比特率的内容类型之间提供性能一致的附加价值。

附件6 (资料性)

MPEG编码独立于编码的代码点 (CICP)

1 引言

ISO/IEC23001-8:2013描述了音频节目与编码表达无关的编码问题，其中包括扬声器系统的位置和布局。默认声道配置包括ITU-R BS.775或ITU-R BS.2051建议书中规定的声道配置。所有声道配置如表3所示。

表3

声道配置和扬声器布局 (注1)

声道配置值*1 (注1)	扬声器数 (注2)	扬声器映射的 默认元素 (注3)	ITU-R BS.775或ITU-R BS.2051建议书中 规定的声道名称 (注4)
0	任意配置		
1	1/0.0 (0+1+0)	M+000	单声道
2	2/0.0 (0+2+0)	M+030	左
		M-030	右
3	3/0.0 (0+3+0)	M+000	中央
		M+030	左
		M-030	右

表3 (续)

声道配置值*1 (注1)	扬声器数 (注2)	扬声器映射的 默认元素 (注3)	ITU-R BS.775或ITU-R BS.2051建 议书中规定的声道名称 (注4)
4	3/1.0 (0+4+0)	M+000	中央
		M+030	左
		M-030	右
		M+180	单声道环绕
5	3/2.0 (0+5+0)	M+000	中央
		M+030	左
		M-030	右
		M+110	左环绕
		M-110	右环绕
6	3/2.1 (0+5+0)	M+000	中央
		M+030	左
		M-030	右
		M+110	左环绕
		M-110	右环绕
		LFE1	低频效果
7	5/2.1 (0+7+0)	M+000	N/A*2
		M+030	
		M-030	
		M+045	
		M-045	
		M+110	
		M-110	
		LFE1	
8	1+1	声道1	N/A
		声道2	
9	2/1.0 (0+3+0)	M+030	左
		M-030	右
		M+180	单声道环绕
10	2/2.0 (0+4+0)	M+030	左
		M-030	右
		M+110	左环绕
		M-110	右环绕

表3 (续)

声道配置值*1 (注1)	扬声器数 (注2)	扬声器映射的 默认元素 (注3)	ITU-R BS.775或ITU-R BS.2051建 议书中规定的声道名称 (注4)
11	3/3.1 (0+6+0)	M+000	N/A
		M+030	
		M-030	
		M+110	
		M-110	
		M+180	
		LFE1	
12	3/4.1 (0+7+0)	M+000	中央
		M+030	左
		M-030	右
		M+090	左侧方环绕
		M-090	右侧方环绕
		M+135	左后环绕
		M-135	右后环绕
		LFE1	低频效果
13	11/11.2 (9+10+3)	M+000	前中央
		M+030	左前中央
		M-030	右前中央
		M+060	左前
		M-060	右前
		M+090	左侧方
		M-090	右侧方
		M+135	左后
		M-135	右后
		M+180	后中央
		LFE1	低频效果-1
		LFE2	低频效果-2
		U+000	中前上
		U+045	左前上
		U-045	右前上
		U+090	顶部左侧
		U-090	顶部右侧
		T+000	前中
		U+135	顶部左后
		U-135	顶部右后
U+180	中后上		
B+000	中前下		
B+045	左前下		
U-045	右前下		

表3 (续)

声道配置值*1 (注1)	扬声器数 (注2)	扬声器映射的 默认元素 (注3)	ITU-R BS.775或ITU-R BS.2051建 议书中规定的声道名称(注4)
14	5/2.1 (2+5+0)	M+000	中央
		M+030	左
		M-030	右
		M+110	左环绕
		M-110	右环绕
		LFE1	低频效果
		U+030	左前上
		U-030	右前上
15	5/5.2 (3+7+0)	M+000	中央
		M+030	左
		M-030	右
		M+090	左侧
		M-090	右侧
		M+135	左后
		M-135	右后
		U+045	左高
		U-045	右高
		UH+180	中央高
		LFE1	左低频效果
		LFE2	右低频效果
16	5/4.1 (4+5+0)	M+000	中央
		M+030	左
		M-030	右
		M+110	左环绕
		M-110	右环绕
		LFE1	低频效果
		U+030	左前上
		U-030	右前上
		U+110	左后高
		U-110	右后高
17	6/5.1 (6+5+0)	M+000	N/A
		M+030	
		M-030	
		M+110	
		M-110	
		LFE1	
		U+000	
		U+030	
		U-030	
		U+110	
		U-110	
		T+000	

表3 (续)

声道配置值*1 (注1)	扬声器数 (注2)	扬声器映射的 默认元素 (注3)	ITU-R BS.775或ITU-R BS.2051建 议书中规定的声道名称 (注4)		
18	6/7.1 (6+7+0)	M+000	N/A		
		M+030			
		M-030			
		M+110			
		M-110			
		M+150			
		M-150			
		LFE1			
		U+000			
		U+030			
		U-030			
		U+110			
		U-110			
		T+000			
19	5/6.1 (4+7+0)	M+000	中央		
		M+030	左		
		M-030	右		
		M+090	左侧环绕		
		M-090	右侧环绕		
		M+135	左后环绕		
		M-135	右后环绕		
		LFE	低频效果		
		U+045	左前上		
		U-045	右前上		
		U+135	顶部左后方		
		U-135	顶部右后方		
		20	7/6.1 (4+9+0)	M+000	中央
				M+SC	左屏
M-SC	右屏				
M+030	左				
M-030	右				
M+090	左侧环绕				
M-090	右侧环绕				
M+135	左后环绕				
M-135	右后环绕				
LFE	低频效果				
U+045	左前上				
U-045	右前上				
U+135	顶部左后方				
U-135	顶部右后方				

表3（结束）

声道配置值*1 (注1)	扬声器数 (注2)	扬声器映射的 默认元素 (注3)	ITU-R BS.775或ITU-R BS.2051建 议书中规定的声道名称(注4)
21-63	预留		

*1 音频输出声道配置用ISO/IEC 23001-8:2013“独立于编码的代码点”中规定的、用于承载声道配置值的一个六比特字段来表示。

*2 N/A: 不适用；声道配置未在ITU-R BS.2051或ITU-R BS.775建议书中规定。

注1 – 列表摘自ISO/IEC 23001-8:2013/Amd.1:2015的表8。

注2 – 扬声器编号的概念规定在“前置扬声器/环绕扬声器、LFE扬声器”以及括号中的“上层扬声器 + 中层扬声器 + 底层扬声器”的约定中，但不包括LFE扬声器。

注3 – 通过符合ITU-R BS.2051建议书的标签来识别扬声器。

注4 – 声道标签和名称取决于实际的声道配置。

附件7 (资料性)

AC-4

1 引言

AC-4是一种用于数字广播的高级编码系统，它使用数字压缩算法和不同的参数编码工具来改善效率和功能。AC-4支持基于声道的和基于对象的输入和输出格式。

表4提供了AC-4本身支持的声道格式清单，它涵盖ITU-R BS.1548建议书中定义的、发射所需的声道配置。AC-4还支持高声道计数格式的编码，它可用于支持ITU-R BS.2051建议书的高级音频系统。

表4

本身支持的声道格式概述

格式	声道数	注
单声道 (1/0格式*)	1	
立体声 (2/0格式*)	2	

表4 (结束)

格式	声道数	注
3.0 (3/0格式) ⁽¹⁾	3	
5.0/5.1 (3/2格式) ⁽¹⁾	5/6	
7.0/7.1 (系统I) ⁽²⁾	7/8	三种不同扬声器配置的信令
7.0.4/7.1+4 (系统J) ⁽²⁾	11/12	声道较少的子集的信令
9.0.4/9.1.4	13/14	声道较少的子集的信令
22.2 (系统H) ⁽²⁾	24	

⁽¹⁾ 定义于ITU-R BS.775建议书。

⁽²⁾ 定义于ITU-R BS.2051建议书。

AC-4能够将1至22.2个声道的源PCM音频编码为串行比特流，数据速率从24 kbit/s到1 536 kbit/s。除了支持基于声道的表示外，AC-4也支持动态音频对象的编码。在ETSI TS 103 190-2文件中可以找到对AC-4比特流语法的完整描述。

表5提供了建议的比特流，用于不同的声道配置，以满足ITU-R BS.1548建议书中所述的要求。

表5

用于特定声道配置的比特流概述

模式	比特流
2.0立体声	96 kbit/s
5.1环绕声	192 kbit/s
22.2环绕声	1536 kbit/s

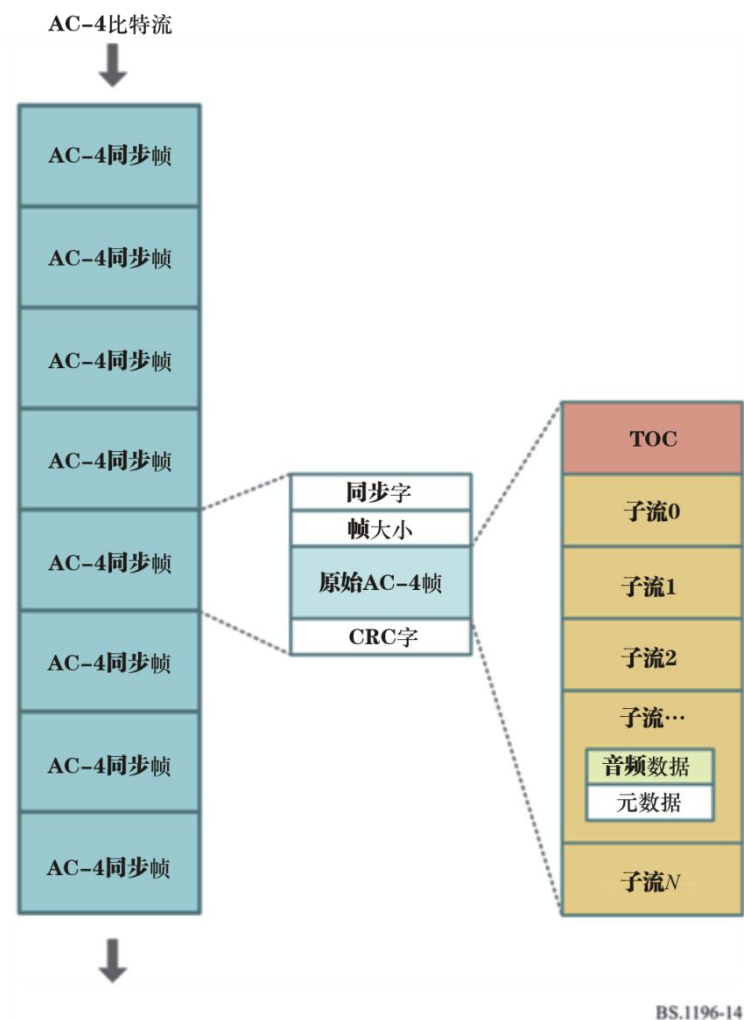
AC-4也本身支持下述系统特征：

- 符合ITU-R BS.1770和ITU-R BS.1771建议书的智能响度管理，包括表明符合目前有效之国际和若干区域响度规定的信令。
- 支持对基于声道的和基于对象的音频表示的编码和解码。
- 支持如ITU-R BS.1909建议书要求的收听环境，即“家庭环境”和“移动环境”。包括适用于“家庭”和“移动”环境的众多设备类型的高级动态范围控制。
- 对话增强。
- 视频帧同步编码，使音频帧与视频帧对齐。

- 本身支持增强型附属数据或元数据的传送和信令。

图14所示的AC-4比特流包括AC-4同步帧，它以同步字开始，并以循环冗余校验（CRC）字结束。同步字允许解码器简单地识别AC-4帧并开始解码，CRC字则允许解码器检测比特流错误的发生并执行任何所需的错误隐藏。实际的编解码器帧或“原始AC-4”帧由TOC（目录）和至少1个子流组成。

图14
高级比特流同步



每个子流包括经编码的数据和相关的元数据（附属数据）。TOC包括有关应如何解码子流或多个子流的必要信息。

2 编码

没有关于AC-4编码器的规范性说明，但它支持各种编码工具，因此可以生成符合要求的编码比特流。

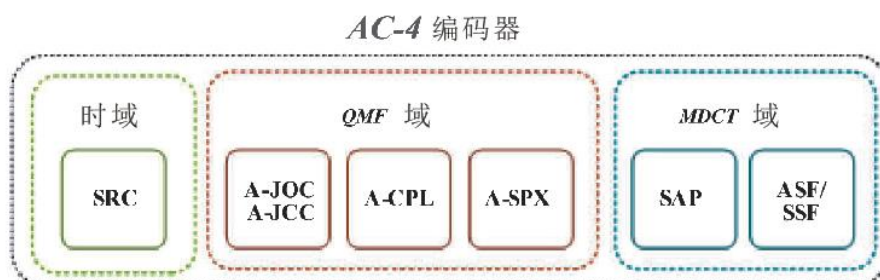
AC-4使用转换域量化以及使用伪正交镜像滤波器组中具有灵活窗口切换和参数编码的、经修正的离散余弦变换（MDCT）进行编码。AC-4支持将1至22.2个声道的源PCM音频

以24 kbit/s至1536 kbit/s的数据速率编码为串行比特流。对于立体声2.0和5.1声道的编码，数据速率96 kbit/s和192 kbit/s满足ITU-R BS.1548建议书中规定的性能要求。AC-4也支持三种比特分配模式：恒定比特率、平均比特率和可变比特率。

AC-4编码器可以通过图15所示的各种编码工具来工作，以便根据操作模式和/或应用来提高效率和灵活性/功能。工具工作的顺序遵循从左至右的顺序，这意味着PCM音频将先从左侧输入工具，然后AC-4经编码的比特率从右侧输出。对于编码工具的描述如下所示：

- **SRC**：AC-4编码器中需要取样速率转换器工具来启用与帧同步编码模式的视频帧匹配的帧持续时间。取决于帧速率，输入信号被转换为内部取样速率46,034 Hz、46,080 Hz、48,000 Hz或51,200 Hz之一，供后续基于QMF和MDCT的编码器工具使用。
- **A-JCC**：高级联合声道编码工具将沉浸声道（大于5声道）输入向下混合为较少数量的声道并对相关参数进行编码。通过这些参数，解码器可以完成对所有输入声道的重建。
- **A-JOC**：高级联合对象编码工具将音频对象作为输入，并对这些对象进行空间编码以产生更少数量的输出对象，因此而减少MDCT编码的信号数量。通过利用数量减少后的输出对象对参数进行编码，能够完成对解码器中对象的重建。
- **A-CPL**：高级耦合工具执行从2声道到1声道的向下混合，并对相关参数进行编码，以重建原始2声道输入。
- **A-SPX**：高级频谱扩展工具对与输入信号的高频内容相关的参数进行编码，然后在解码器中进行频谱扩展。这些参数包括包络、音调和噪声测量结果。参数的频谱和时间分辨率可以适应输入信号的特性。
- **SAP**：立体声音频处理工具在两个或多个输入声道之间的MDCT域中进行联合声道编码。
- **ASF**：音频频谱前端工具是一种基于MDCT的量化和编码工具，它利用变换窗口切换。窗口切换控制模块根据输入信号的类型来选择帧的最佳变换长度。在比特流的非线性量化和无噪编码之后，存储MDCT系数和附加控制信息。时间和频率上的比特分布均由比特缓冲和感知模型来控制。比特缓冲模型也考虑到了其他编码器工具和通用元数据所用的比特。
- **SSF**：语音频谱前端工具是一种替代的、基于MDCT的量化和编码工具，专用于在短变换上进行的语音编码。它通过子带预测器，在MDCT域中执行量化和编码。SSF和ASF工具是相互排斥的，因此只能使用ASF或SSF来对MDCT进行编码。

图15
AC-4系统中可用的编码工具



3 解码

解码处理基本上是编码处理的逆过程，如图16所示。

图16
解码工具



AC-4比特流和解码器设计本身支持实现较低复杂度的解码器，以支持功能有限的设备（例如移动电话/平板电脑）。

采用两种方法来支持此性能：

- 1 核心解码：AC-4支持两种解码模式：完全解码和核心解码。AC-4解码器可以在核心解码模式下操作，该模式对所编码节目的声道的核心子集（表示所有输入声道的内容）进行解码，从而使节目的兼容性再现具有较低的计算复杂度。
- 2 可扩展解码：AC-4也支持可扩展的取样速率，AC-4能够以可扩展的方式来支持较高的取样速率，特别是96 kHz和192 kHz。仅支持48 kHz输出的设备只需解码基本层。

除了支持原生支持的声道格式的解码外，如表4所示，AC-4还支持使用辅助数据或元数据。这允许按照设备所需的可预测的方式，将来自较高声道计数的经解码声道输出向下混合到较低输出声道计数。

附件8 (资料性)

低复杂度型MPEG-H 3D音频

1 引言

MPEG-H 3D音频是一种音频编码标准，用以支持将音频编码为声道、音频对象或高阶高保真度立体声响复制（HOA），并为响度归一化和动态范围控制提供解决方案。每种内容类型（声道、对象或HOA）可单独使用或彼此之间组合使用。音频声道组、对象或HOA的使用使得节目可以实现交互或个性化，例如，可在MPEG-H解码器的渲染期间选择不同的语言音轨或者调整对象的增益或位置。

MPEG-H 3D音频规范发布在ISO/IEC 23008-3:2015修正3中，规范明确了低复杂度型（LC型）MPEG-H 3D音频，其他技术于2017年年初发布。

LC型MPEG-H 3D音频可支持最多24个输出扬声器以及56个编解码器核心声道（其中28个声道可以一次解码）。

可能的目标扬声器布局示例如下：

- 2.0立体声（ITU-R BS.775建议书中定义的2/0格式）。
- 5.1多声道音频（ITU-R BS.775建议书中定义的3/2格式）。
- 10.2沉浸式音频（ITU-R BS.2051建议书中定义的系统F）。
- 22.2沉浸式音频（ITU-R BS.2051建议书中定义的系统H）。

该标准可用于众多应用，包括立体声和环绕声存储与传输。该标准支持交互式和沉浸式声音，以满足下一代媒体分发要求，尤其是新的电视广播系统和娱乐流媒体服务以及与虚拟现实有关的内容和服务。

例如，在电视广播中，评论或对话可以作为音频对象发送，并与MPEG-H 3D音频解码器中的沉浸式声道床相结合。这允许以多种语言进行的对话可以实现高效传输，并允许收听者根据自己的偏好调节对话与其他声音元素之间的平衡。该概念可被扩展至广播中通常不会出现的其他元素中，诸如视觉障碍者的音频描述、导演的评论或者体育赛事参与者的对话。

LC型MPEG-H 3D音频支持符合ITU-R BS.1770和ITU-R BS.1771建议书要求的响度管理，包括表明符合国际和一些区域性响度规定的合规信令。它还支持“家庭”和“移动”环境下众多设备类型的高级动态范围控制（DRC）。

2 编码

MPEG-H 3D音频编解码器体系结构构建于用于压缩不同输入信号类别的感知编解码器之上，基于MPEG统一语音和音频编码（USAC）。USAC允许以每声道8 kbit/s或更高的速率来压缩单声道到多声道的音频信号。

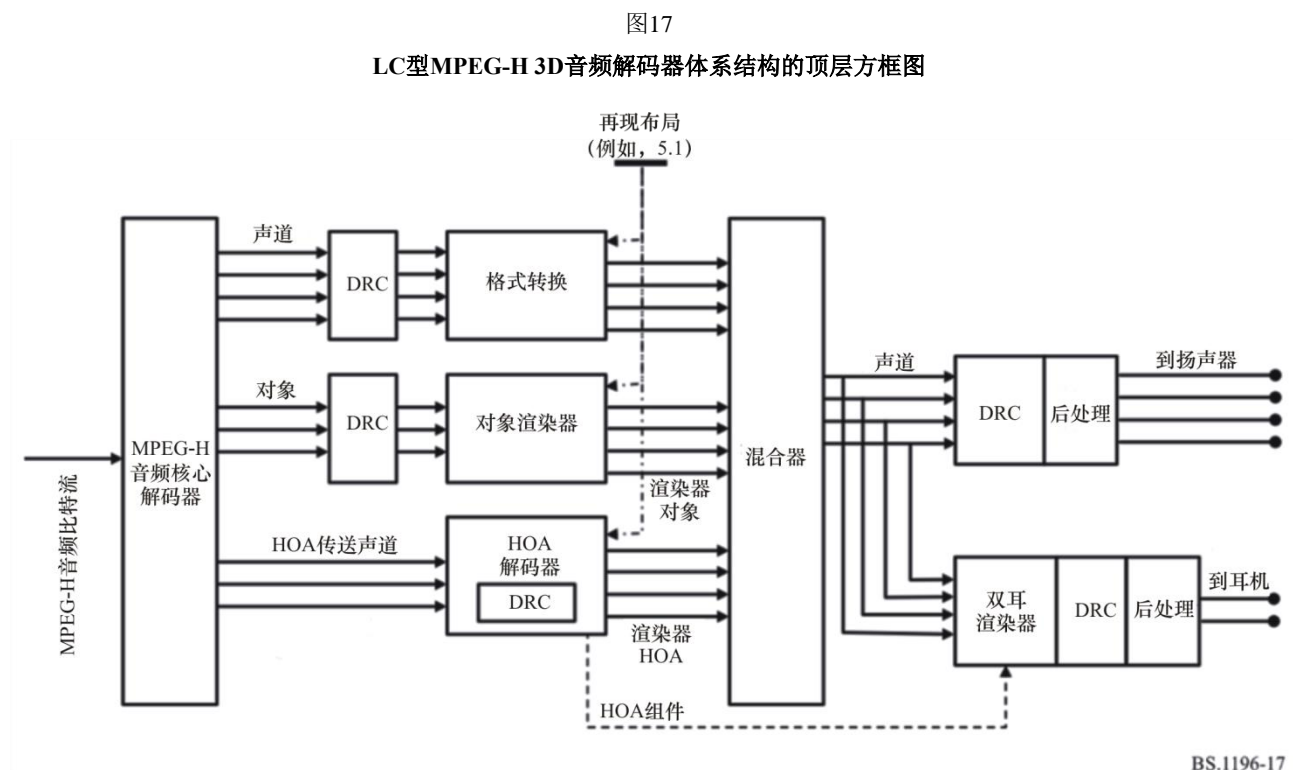
对3D音频场景中出现的新要求，工具对该技术进行了扩展，特别是利用3D再现的感知效果，并因此而进一步提高了编码效率，诸如：

- 通过智能间隙填充（IGF）提供了一种增强型噪声填充。IGF是一种工具，它使用来自频率和时间上相邻之频谱片段的适当信息，以参数方式恢复部分发射的频谱。对这些频谱片段的分配和处理通过基于输入信号分析的编码器来控制。因此，频谱间隙可以用频谱系数来填充。相比传统噪声填充的伪随机噪声序列，它在感知上具有更好的匹配性能。
- 除了编码效率的提高外，为3D内容/扬声器的布局和压缩流（音频声道、音频对象、HOA信号）中的信号类型，USAC-3D核心配备了新的信令机制。

压缩音频有效负载设计中的另一个新问题是瞬时速率切换或快速提示的行为做了改进，因为它出现在MPEG动态自适应流媒体（DASH）中。出于此目的，语法中添加了所谓的“即时播放帧”，使得从一个流到另一个流的无缝过渡变得可能。这对于IP网络上的自适应流媒体尤为有利。

3 解码

图17中描绘了整个LC型MPEG-H 3D音频解码器体系结构的顶层方框图。



主要的组件是一个所谓的USAC-3D核心解码器，它是一组用于不同信号类别和混合器的渲染器。在第一个阶段，借助于所谓的USAC-3D解码器，将其经数据压缩的表示转换为不同的基本信号。

然后将不同的信号类别（声道信号和目标信号或HOA系数信号的波形）馈送到其相关联的渲染器，这些渲染器将这些信号映射到扬声器馈送，用于在接收机侧可用的特定再现设置。一旦所有经渲染的信号都以再现格式提供，它们就会在混合阶段结合起来，以形成一个扬声器馈送。在请求双耳表示的情况下，信号被转换成用于耳机再现的虚拟3D场景。可以在单个MPEG-H流中传输不同信号类型的任意组合，例如，声道信号与对象信号或HOA场景与对象的组合。

渲染器为：

- 用于将声道信号从其制作扬声器格式转换为再现扬声器布局的格式转换器。
- 用于将静态或动态对象音轨放入再现布局的对象渲染器。
- 将基于场景的HOA表示转换为实际的再现布局的HOA渲染器。
- 将虚拟扬声器布局转换为耳机输出的双耳渲染。

此外，如果相应的元数据将这些信号标记为启用用于交互，则不同信号类别的播放和渲染可由用户接口来控制。

附件9 (资料性)

DTS-UHD

1 引言

DTS-UHD编码系统包括音频编码引擎（ACE），能够有效压缩音频波形。DTS-UHD支持基于频道的音频（CBA）编码、三维声场和基于对象的音频（OBA）。DTS-UHD在ETSI TS 104 491中实现标准化，并成为DVB MPEG 2传输流规范ETSI TS 101 154的一部分。该系统提供全沉浸式体验，通过运用对象、声道和声场及其综合效果，实现互动和个性化，提升使用体验。此外，该系统还为视力和听力受损用户提供无障碍获取服务。

DTS-UHD支持ITU-R BS.2051建议书规定的所有扬声器布局。

DTS-UHD不仅全面支持基于声道的音频，还能对多达224个离散音频对象进行编码，并在一个音频流中将其进一步编入32个对象组和32个方案中。

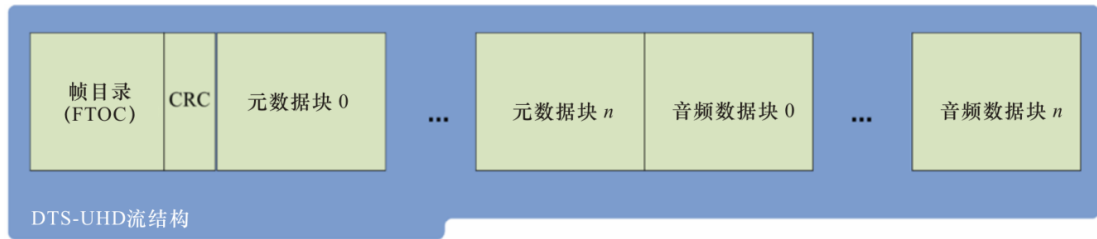
一个DTS-UHD流由一系列包含以下三个主要元素的DTS-UHD帧组成：

- 帧目录（FTOC）– 该元素允许解码器直接导航至帧的相关元素；
- 元数据块元素；
- 音频数据块元素。

结果如下所示：

图18

DTS-UHD流结构



BS.1196-18

FTOC存在于DTS-UHD的每一帧，表示某一帧是同步帧或非同步帧，同步帧为发起解码会话提供了全部必要参数。

音频数据块承载以最优方式优化压缩和承载的波形，可以为同步帧也可为非同步帧。音频数据块编组为多个子流，可表示单声道波形、立体声波形和LFE波形。

元数据块全面描述了一个音频组件，包括组件的类型，需要渲染组件的音频数据块，以及响度和动态范围压缩的信息。接入点（DTS-UHD同步帧）的元数据块引用的音频数据块为同步帧，包含启动解码的一整套元数据。

2 编码器

欧洲电信标准学会（ETSI）并未对DTS-UHD ACE编码器作规范说明。该编码器采用的一整套工具能够提供多个比特流，支持多种扬声器配置。

DTS-UHD编码技术能够保证音频的质量，并能以以下比特率产生满足ITU-R BS.1548建议中规定的高质量发射的比特流：

扬声器配置	比特率 (kbit/s)
0+2+0 (立体声)	128
0+5+0 (5.1)	192
4+7+0 (7.1.4)	288

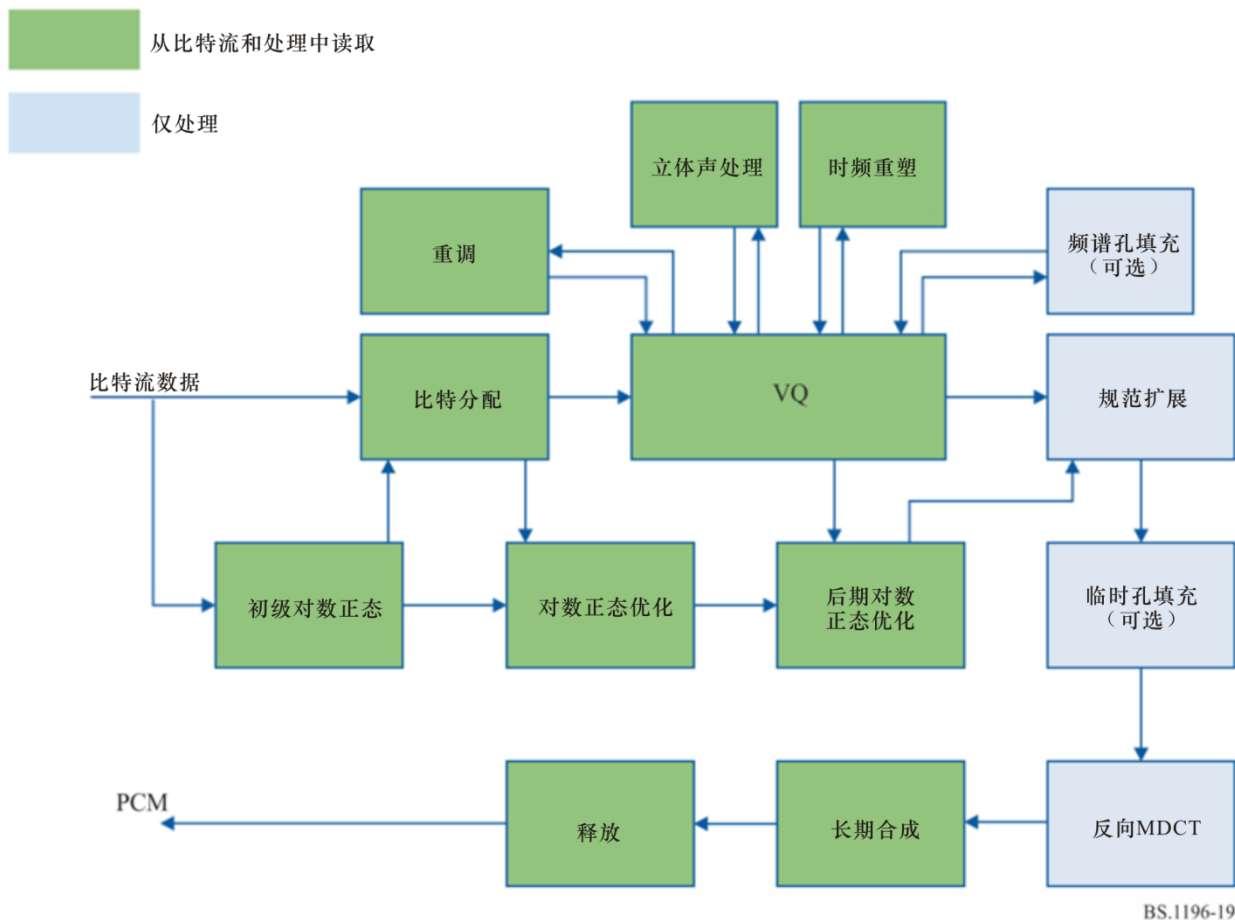
DTS-UHD编码器能够充分控制同步帧的放置，使某种业务可以控制进入音频业务的进入点。

DTS-UHD可变比特率（VBR）功能保证了各GoF（组帧 – 两个连续同步帧的间隔）的恒定比特率（CBR）。同步帧由编码器设定，同步间隔在编码过程中可按需动态调整。DTS-UHD编码器还能按要求设定最大瞬时VBR峰值速率。DTS-UHD比特率控制（BRC）模型在一个比特流中的对象、声道和声道对之间的各音频帧动态分配比特。BRC为每一帧中最相关的内容分配更多比特进行编码，同时确保同一帧内的所有对象和声道具有相同的音频质量。通过对象和声道之间相关内容的不同的分配，可以调用更多音频元素，挺高编码效率。编码器还能为音频帧动态分配比特，实现时间轴上恒定的音频质量（在GoF-wise CBR和规定的最大峰值比特率的限制内）。值得注意的是，BRC功能无须传输其他信息，解码器不知道BRC功能。在BRC模式下，各对象、声道或声道对保持相互独立的可解码性，并“可充分执行”。

3 解码器

ACE解码器对ACE立体声流、ACE单声道流和ACE LFE流进行操作。在ACE帧中，流被分组为上述三种流中的一种。立体流通过利用双声道共同的信息再次进行优化。

图19
ACE解码器



解码单声道或立体声ACE流包括以下几个阶段：

初级对数正态：从比特流中检索频段功率的第一近似值。在对数域中表示这些功率值，称为对数正态值。

比特分配：对于每个频段，计算标准化频段解码可用的比特数。该计算取决于比特流中的信息和（可能的）初级对数正态值。

对数正态优化：采用一小部分比特分配用于优化初级对数正态值。

VQ：（矢量量化），采用大部分比特分配用于重建标准化频谱频率值，每次一个频段。

重调：对于初始“已满”的频段以及较小的比特分配值，初始的重构稀疏频段经过重调，更接近全谱。

立体声处理：对立体声流而言，两个重构的声道转化为一个原始（非混合的）表示。

时频重塑：通过简单的哈尔变换，重新处理频段，将所有频段整合在同一时频表示。

频谱孔填充：初始能量重构为零频段填充“噪声”，噪声由早先重构的频段值产生。

后期对数正态优化：VQ后剩下的比特将用于最终的对数正态优化。

规范扩展：将频段功率值恢复至对应最终对数正态值的值。

临时孔填充：用合成噪声填充全零频段，取决于之前帧的频段功率。

反向MDCT：频谱表示转换为时域表示（PCM）。

长期合成：采用长期合成滤波器重构编码时移除的信号组件。
