

国 际 电 信 联 盟

**ITU-R**

国际电联无线电通信部门

**ITU-R BS.1196-3 建议书**

(08/2012)

**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卫星业务等所有无线电通信业务合理、平等、有效、经济地使用无线电频谱，不受频率范围限制地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世界或区域无线电通信大会以及无线电通信全会在研究组的支持下履行。

## 知识产权政策 (IPR)

ITU-R的IPR政策述于ITU-R第1号决议的附件1中所参引的《ITU-T/ITU-R/ISO/IEC的通用专利政策》。专利持有人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/en>获得，在此处也可获取《ITU-T/ITU-R/ISO/IEC的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/en>)

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

**说明：** 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  
2013年，日内瓦

© 国际电联 2013

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.1196-3 建议书\*\*\*

## 数字广播的音频编码

(ITU-R第19/6号研究课题)

(1995-2001-2010-2012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源编码系统。它还详细说明了一种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。

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) 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；
- d) 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；
- e) 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；
- f) 不同业务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；
- g) 若干种广播业务早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；
- h) 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；

---

\* 无线电通信第6研究组根据国际电联无线电通信部门（ITU-R）第44号决议于2003年对本建议书进行了编辑性修订。

\*\* 应提请国际标准化组织/国际电工委员会（ISO/IEC）注意本建议书。

- 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) 音频编码系统应最好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之一：

- 扩展版HE AAC，如ISO/IEC 23003-3（2012）所述；
- E-AC-3，如ETSI TS 102 366（2008-08）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。

**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半采样速率，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 AAC第二版，如ISO/IEC 14496-3（2009）所述；
- AC-3，如ETSI TS 102 366（2008-08）所述。

注3 – 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4 – 鼓励ITU-R成员以及接收器和芯片制造商支持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ISO/IEC 23003-1（2007）所述的多声道音频扩展；

注5 – 由于ISO/IEC 23003-1（2007）所述的MPEG环绕声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ISO/IEC 11172-3层II编码，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180 kbit/s（不包括辅助数据）；

**5** 对于解说链路，可使用ISO/IEC 11172-3层III编码，单信道信号的比特速率至少为60 kbit/s（不包括辅助数据），利用联合立体声编码，立体声信号的比特速率至少为120 kbit/s（不包括辅助数据）；

**6** 对于高质量应用，取样频率应为48 kHz；

7 低比特率音频编码器的输入信号不应加重，而编码器不应采用任何加重方式；

8 遵守本建议书是自愿的。然而，本建议书可能包含某些强制性规定（为保证互操作性或可应用性等），当符合所有这些强制性规定时，便遵守了本建议书。“应”或“必须”等一些强制性语言以及对应的否定性词语用来表示要求。不得将这些词汇的使用解释为必须部分或完全地遵守本建议书，

#### 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它用户要求的编码系统配置的信息，应参见ITU-R BS.1548建议书。

注1 – 本建议书中有关编解码器的信息见附录1-5。

## 附录 1

### MPEG-1和MPEG-2，层II和层III音频

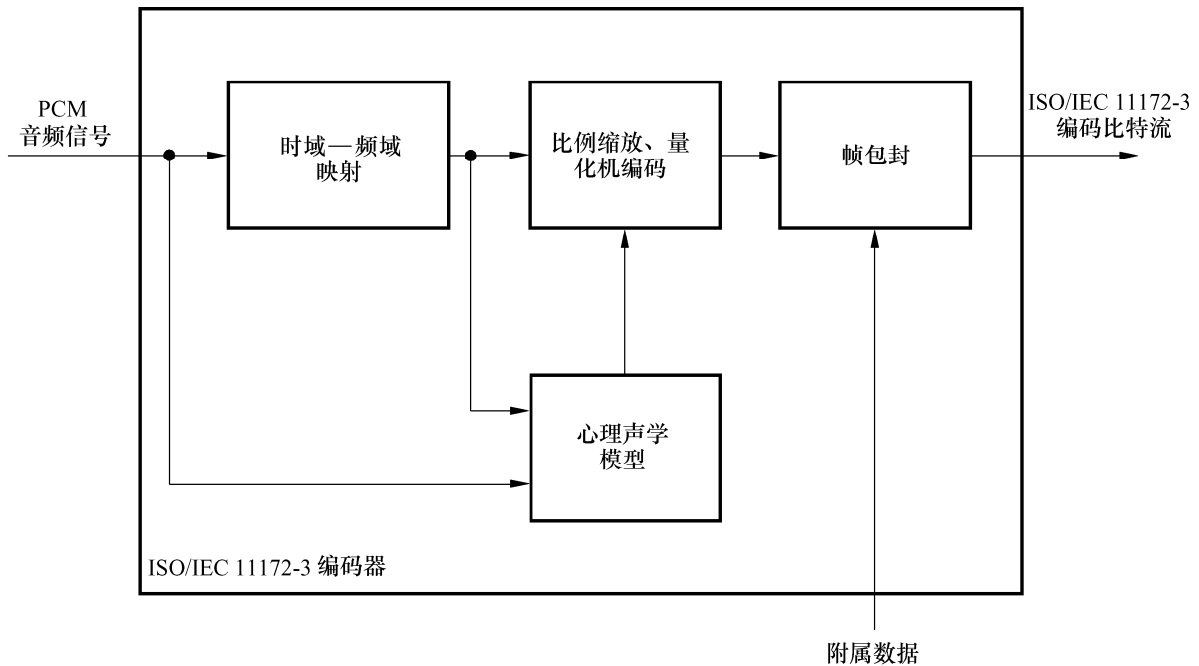
#### 1 编码

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，诸如听觉掩蔽门限估计、量化和比例缩放（注 1）。编码器输出必须使符合于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 中所示的典型编码器。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 III 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，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。“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一起，并当需要在辅助数据字段内加上其他信息（例如，误码纠错）。

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：**层 I 中包含：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例分配以及利用数据块的压扩和格式化进行量化。一个层 I 帧表征每信道 384 个样本。

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

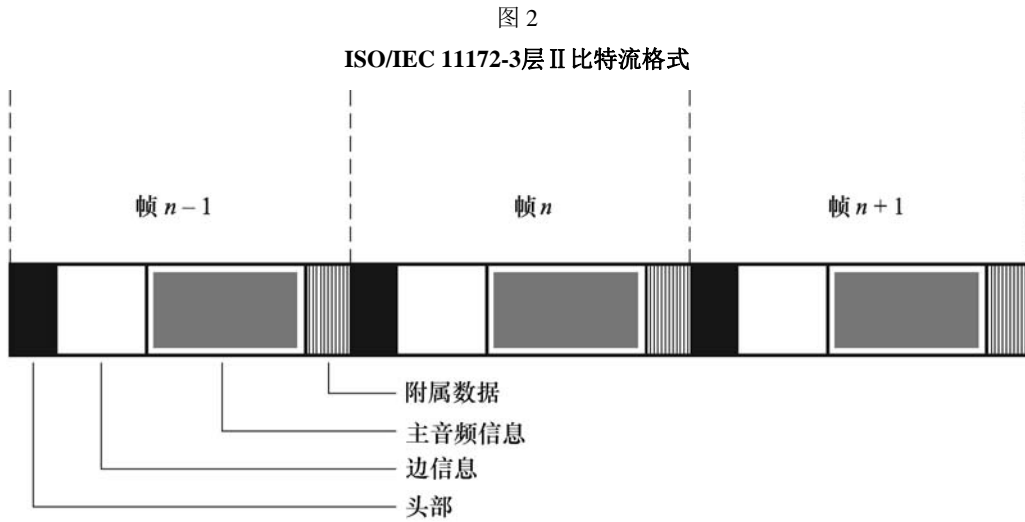
**层 III：**层 III 中基于混合滤波器组（带有可变长度修正离散余弦变换的 32 子带滤波器组）引入增大的频率分辨率。它增加了非均匀量化器、自适应分段和量化估熵编码等处理。一个层 III 帧表征每信道 1152 个样本。

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）；以及，
- 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以使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。

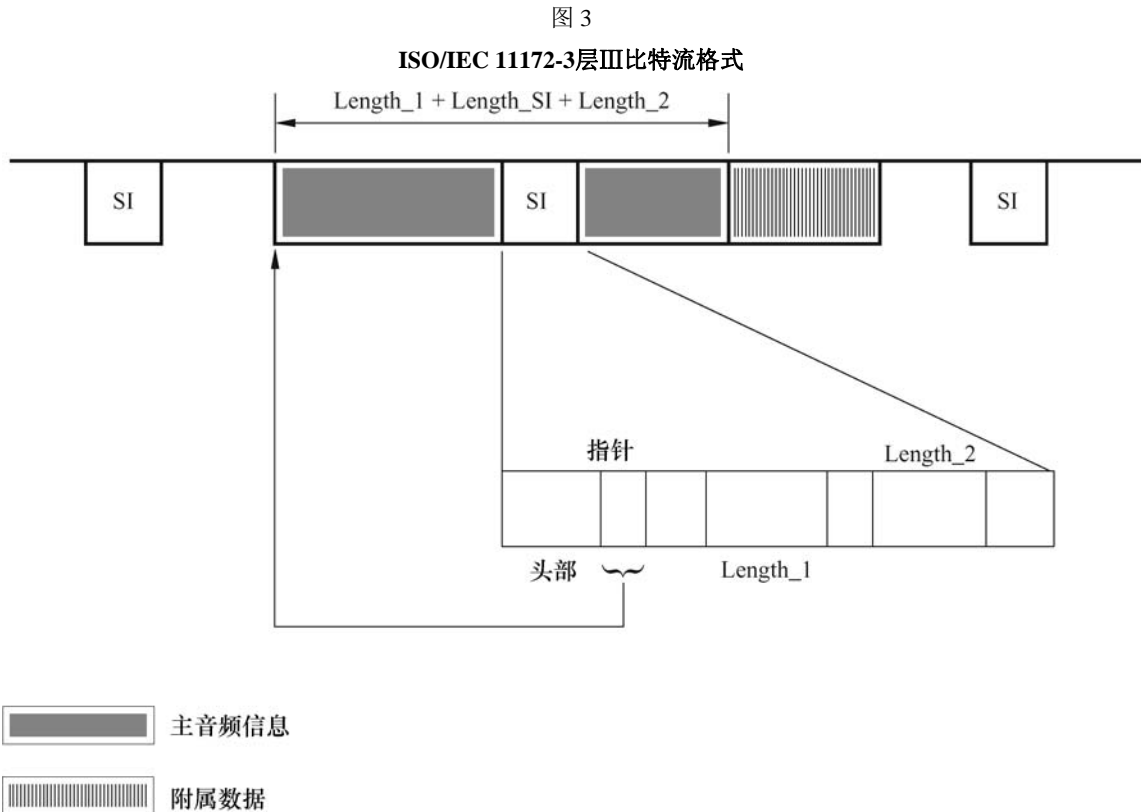
### 3 编码比特流格式

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。编码比特流内连续帧构成。依据层级，一个帧内包括有如下字段：



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



## 层 III:

边信息(SI):	比特流中包含首标、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。

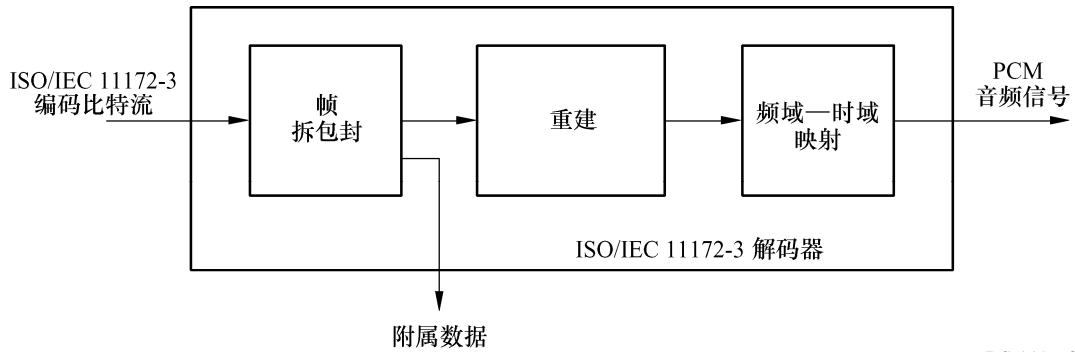
BS.1196-03

## 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。

如图 4 所示，编码的音频比特流传送入解码器。如果在编码器内加入有误码校验，则比特流拆包封和解码处理时相应实施误码检测。对比特流拆包封以恢复各个信息段，诸如音频帧音标、比特分配、比例因子、映射样本和可选的辅助数据等信息。重建处理中，重建映射样本集合中的量化值。频域到时域的映射使那些映射样本变换回线性 PCM 音频样本。



图4  
解码器方框图

BS.1196-04

## 附录 2

### MPEG-2和MPEG-4 AAC音频

#### 1 引言

ISO/IEC 13818-7 说明了被称为 MPEG-2 先进音频编码 (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)。

##### — 取样频率可缩放 (SSR) 型

SSR 型中借助增益控制工具能提供取样频率可缩放的信号。它能选择频带进行解码，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带，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 2 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 路声道。表 3 示明默认声道配置，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 声道加低音效果 (LFE) 声道 (带宽 < 200 Hz) 的 5.1 声道等。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

扬声器安排。还支持缩混能力，用户能将多声道音频信号缩混到指定缩混系数的两声道上。所以，能够控制仅应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

扬声器数	接收指令中列出的 音频句法元素	映射的扬声器默认元素
1	single_channel_element	前置中央扬声器
2	channel_pair_element	前置左、右扬声器
3	single_channel_element()	前置中央扬声器
	channel_pair_element()	前置左、右扬声器
4	single_channel_element()	前置中央扬声器
	channel_pair_element(),	前置左、右扬声器
	single_channel_element()	后置环绕扬声器
5	single_channel_element()	前置中央扬声器
	channel_pair_element()	前置左、右扬声器
	channel_pair_element()	后置左环绕、右环绕扬声器
5+1	single_channel_element()	前置中央扬声器
	channel_pair_element()	前置左、右扬声器
	channel_pair_element()	后置左环绕、右环绕扬声器
	Lfe_element()	低音效果扬声器

表 2 (完)

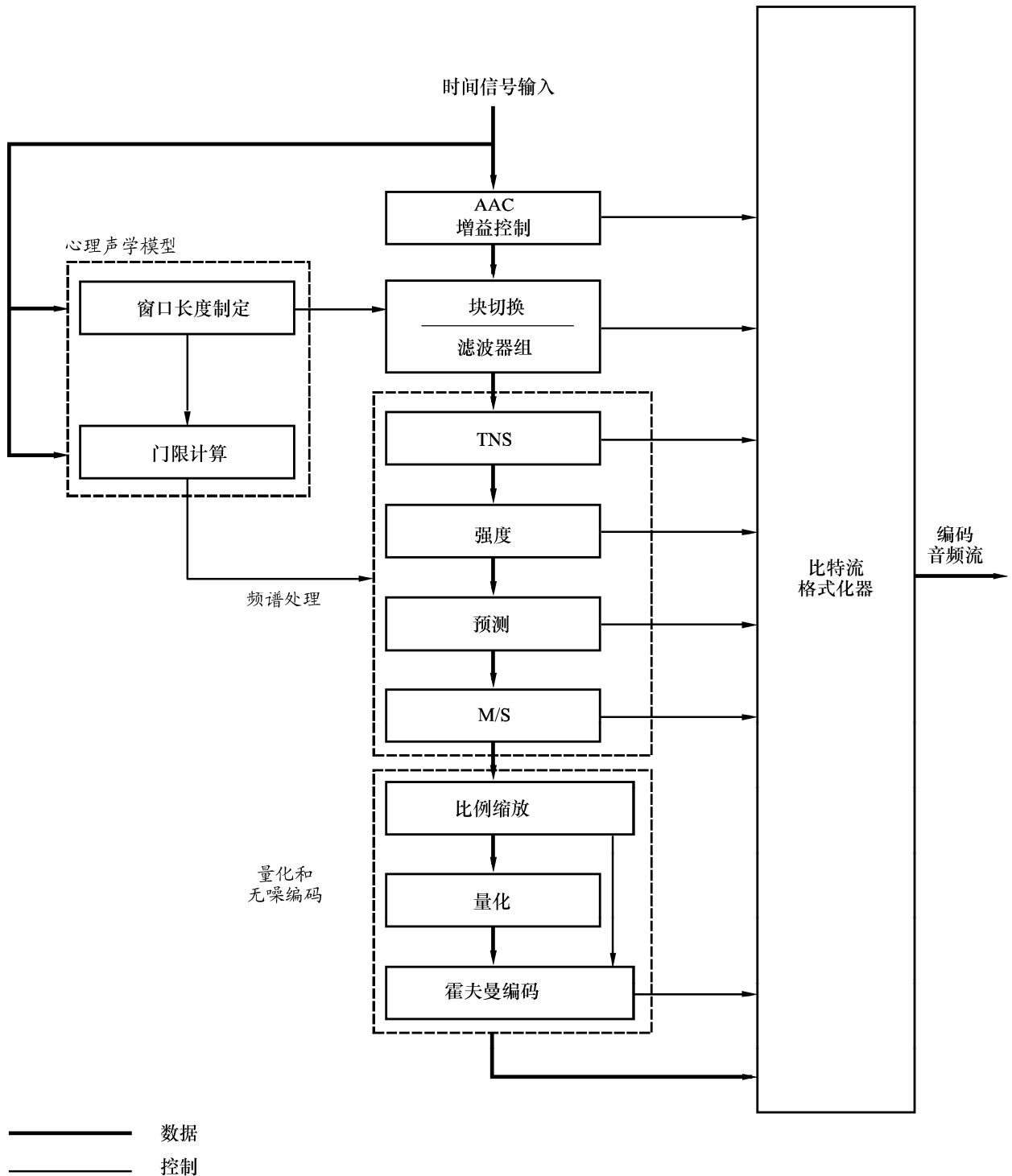
扬声器数	接收指令中列出的音频句法元素	映射的扬声器默认元素
7+1	single_channel_element()	前置中央扬声器
	channel_pair_element(),	前置左、右扬声器
	channel_pair_element()	前置外侧左、右扬声器
	channel_pair_element()	后置左环绕、右环绕扬声器
	lfe_element()	低音效果扬声器

## 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个等间隔的频带。增益控制应用于SSR型中。
- 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，48 kHz取样时，频率分辨率为23 Hz，时间分辨率为21.3 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。强度立体声编码中，只传输声能包络以减少传输方向信息。M/S立体声编码中，采用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）以取代传输原来的左与右信号。
- 预测：为降低静止信号的冗余，在相继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，应用步级大小为1.5 dB的非均匀量化器。对于量化的频谱，不同的比例因子和方向信息，实施霍夫曼编码。
- 比特流格式化器：最后，应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：应用心理声学模型计算输入信号内当前的掩模门限。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。在量化处理中，应用从掩模门限和输入信号电平内导出的信号—掩模比，以使得可闻的量化噪声最小，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



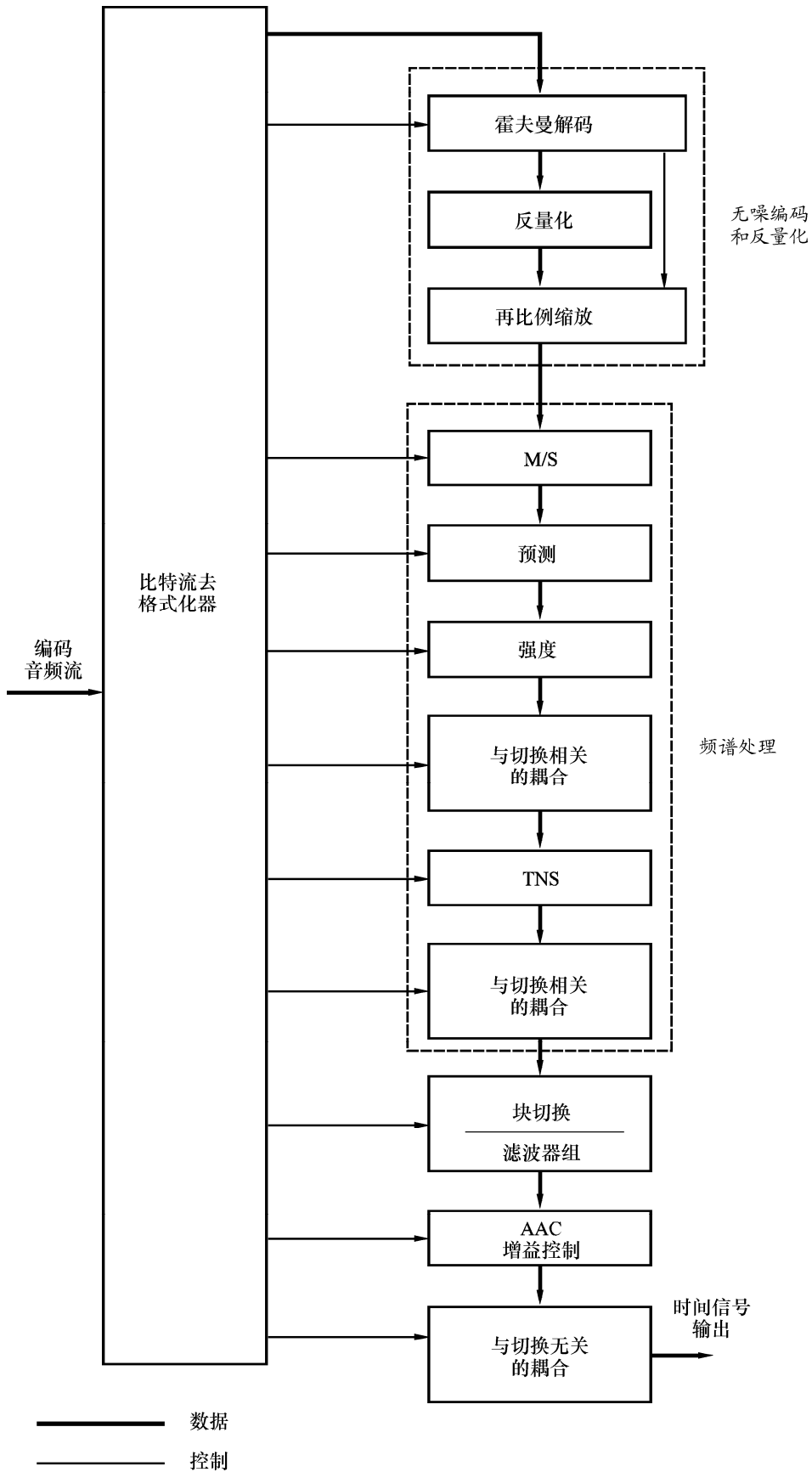
BS.1196-05

### 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

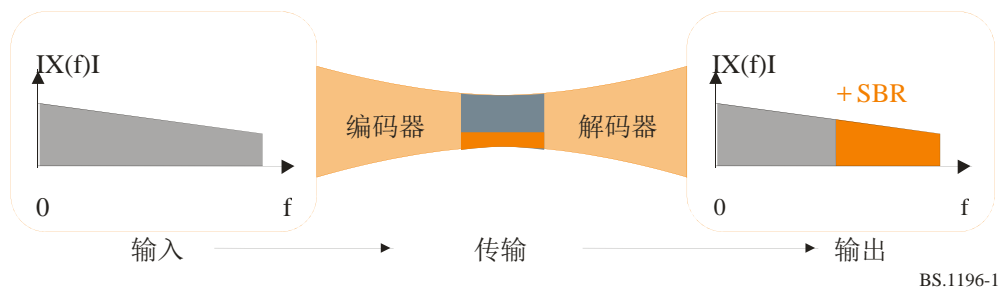


解码器的功能是寻找出比特流中对量化音频频谱的描述，解码其量化值和其他重建信息，重建量化的频谱，借助于比特流中凡是起作用的工具对重建的频谱进行处理，以便得到输入比特流中描述的实际信号频谱，并在最后将频域内频谱变换到时域信号上，处理中可以用或不用可选的增益控制工具。在初始频谱重建和比例缩放的频谱重建之后，有许多供选用的工具可修正一个或多个频谱，以便提供更高效率的编码结果。对于工作于频域内的每一个可选工具，保留其“旁通”选项，在省略频谱运算的所有场合下，可使输入端上的频谱不做修正地直接通过可选工具。

#### 4 高效AAC和频段复制

高效 AAC (HE AAC) 引进了频段复制 (SBR)。SBR 是音频压缩算法中一种高效的高频率编码方法。通过以既定的比特率增加音频带宽，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。

被编码和发射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 不是用来发送频谱中较高的部分，而是基于对发送的较低端频率的分析，重建较高端频率的一个解码后的过程。通过在编码的比特流中以很低的数据速率发送 SBR 相关参数，确保了准确的重建。

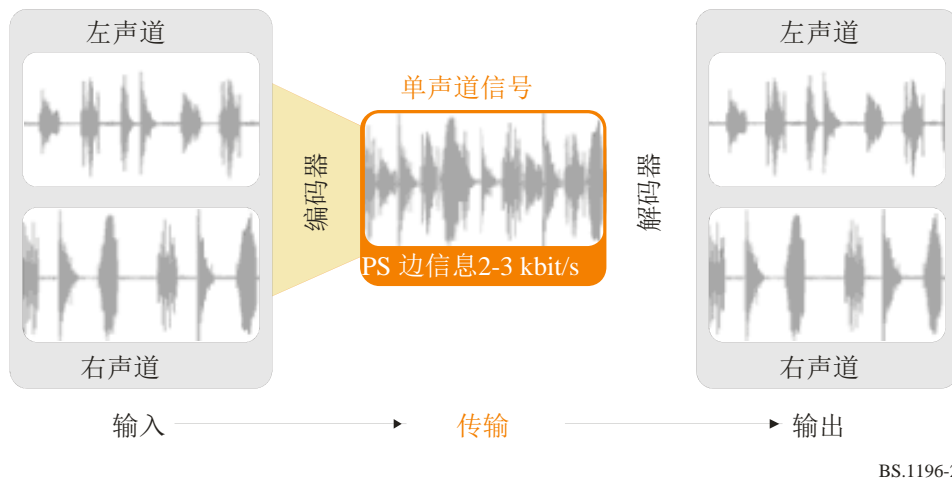


HE AAC 比特流是 AAC 音频比特流的增强型。额外的 SBR 数据嵌入在 AAC 填充元素中，从而保证了与 AAC 标准的兼容性。HE AAC 技术是一个双重速率系统。后向兼容的普通 AAC 音频比特流以半 SBR 增强型采样速率运行，因而，无法对 SBR 增强数据进行解码的 AAC 解码器将以 HE AAC 解码器生成的采样速率的半倍速率生成一个输出时间信号。

#### 5 高效AAC第2版和参量立体声

HE AAC 第 2 版是 HE AAC 的扩展版，并引进了参量立体声 (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，然后构建立体声图形的参数表达式。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。这个信号与重建立体声图像所需的参数一起发送。



其结果是，与没有参量立体声的类似比特流相比，含有参量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。增加的参量立体声数据嵌入单声道 HE AAC 流的 SBR 扩展成分中，从而保证了与 HE AAC 及 AAC 的兼容性。

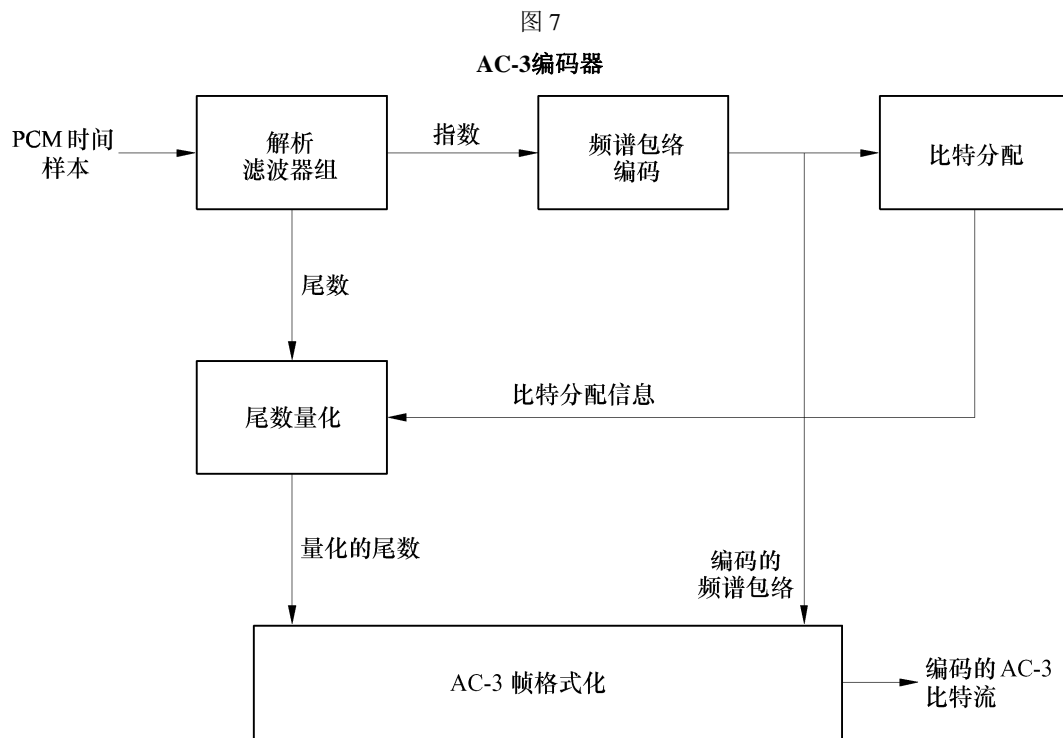
无法对参量立体声增强解码的 HE AAC 解码器以全带宽产生了一个单声道输出信号。无法对 SBR 增强数据解码的普通 AAC 解码器以半倍采样速率生成一个单声道输出时间信号。

## 附录 3

### 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 音频信号的 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。借助于对音频信号的频域表示做出粗略量化，AC-3 算法能给出高的编码效率（输入比特率与输出比特率之比）。AC-3 编码处理方框图如图 7 所示。编码处理的第一步是将来自 PCM 时间样本序列的音频信号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，频域内的表示将除以 2，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。频谱包络供核心比特分配程序应用，它决定以多少比特位来对每一个别的尾数进行编码。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

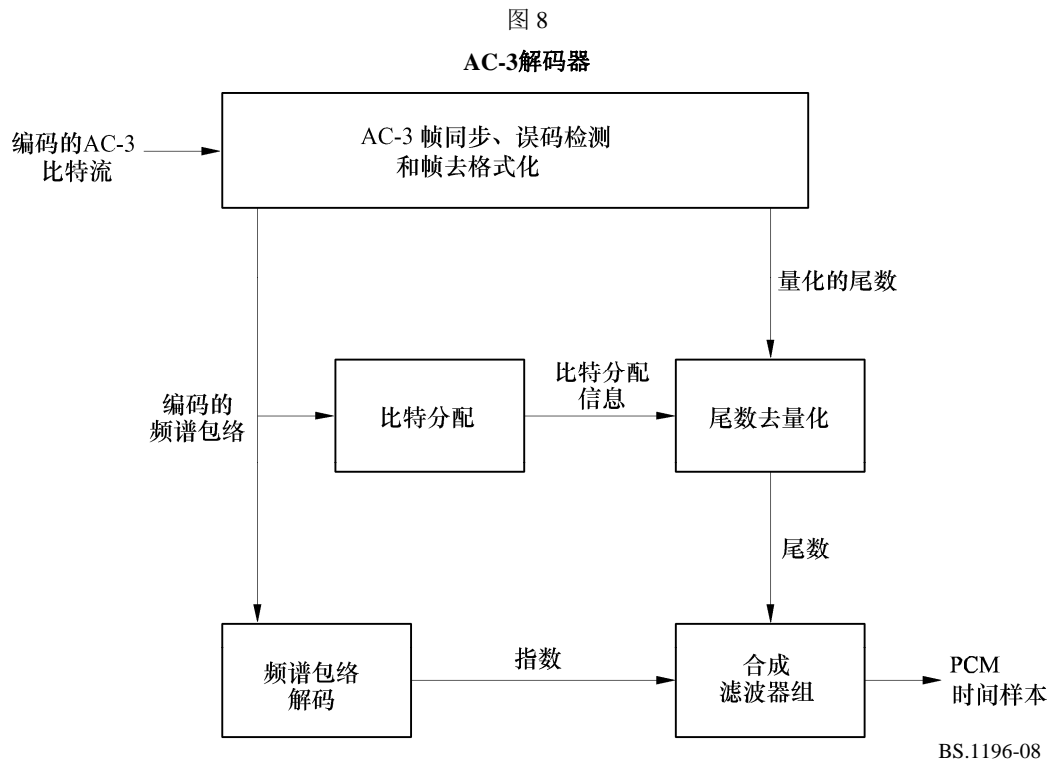
实际的 AC-3 编码器要比图 7 中示出的复杂，实际编码器中还包括上面未示出的下列功能：

- 附加上其中包含当同步和解码已编码 AC-3 比特流时所需信息（比特率、取样频率和编码声道数等）的帧头；
- 插入误码检测码，以使解码器能验明接收的数据帧无误码；
- 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；
- 用于实现较复杂的比特分配和修正核心比特分配程序的参数，借以得到更优化的比特分配；
- 在高音频范围内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；
- 双声道模式中，可以选择性地实施矩阵处理，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情况下能得到改善的结果。



## 2 解码

解码处理基本上是编码处理的逆过程。图 8 中示出的 AC-3 解码器必须与编码比特流同步，校验误码以及对各种类型的数据诸如编码的频谱包络和量化的尾数实施去格式化。运行比特分配程序，并应用其结果进行拆包封和尾数去量化。解码频谱包络以得到指数。将指数和尾数变换回时域内，产生解码的 PCM 时间信号。



实际的 AC-3 解码器要比图 8 中示出的复杂，实际解码器中还包括上面未示出的下列功能：

- 在检测到数据差错的场合下，可采用误码隐匿或误码抹去处理；
- 对于它们的高音频内容耦合在一起的各声道，必须进行去耦合；
- 对于已实施再矩阵处理的声道，必须进行解矩阵（在双声道模式中）；
- 编码处理中已使得解析滤波器组的频谱分辨率动态改变时，解码时合成滤波器组的频谱分辨率必须以同样的方式动态地改变。

### 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编码工具和特色。其额外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, 同时额外特色提供了更多的应用灵活性。

额外编码工具:

- 自适应混合转换 – 应用于分析/合成滤波器组的额外层提供了更精细的 (AC-3的1/6) 频谱分辨率。
- 瞬时预噪音处理 – 减少瞬时预噪音的额外工具。
- 频谱扩展 – 基于编码器创建的边信息对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。

额外特色:

- 更精细的数据率粒度。
- 更高的最大数据率 (3 Mbit/s)。
- 子数据流可携带更多的声道, 例如, 7.1声道, 或解说声道。

## 附录 4

### MPEG环绕

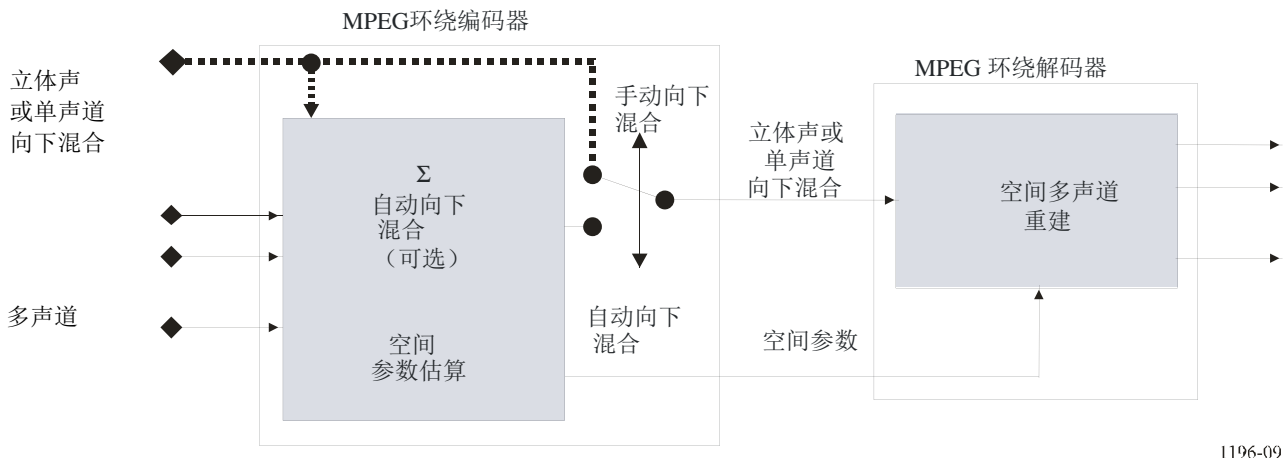
#### 1 引言

ISO/IEC 23003-1 或 MPEG 环绕技术提供一个极为有效的多声道音频编码方法, 并允许以通常用于单声道或立体声音频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 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, 以后向兼容的形式成为核心音频编解码器的一个辅助数据流。原有的单声道或立体声解码器将忽略辅助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, 然后使用从辅助数据流采集的空间图形参数产生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



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，具备 MPEG 能力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 结合的参量立体声技术的基础上，产生了 HE AAC 第 2 版标准规范。通过结合多个参量立体声模块和其它新开发的模块，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，可使用三种不同的配置：一种基于立体声向下混合系统的配置（525 配置），以及两种不同的基于单声道向下混合系统的配置（采用不同框串连的 515<sub>1</sub> 和 515<sub>2</sub> 配置）。

MPEG 环绕含有若干有助于发挥工具作用的特色，从而能够实现对标准的广泛应用。MPEG 环绕的一个重要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。另一个重要特色是解码器输入可与现有的矩阵式环绕技术兼容。

这些及其它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）。残余信号通过将 QMF 域转换为 MDCT 域被编码，然后利用 AAC 对 MDCT 系数进行编码。

- 矩阵式兼容：作为一种选项，立体声向下混合可被预处理，以便与原有的矩阵式环绕技术兼容，确保与配备有矩阵式环绕解码器而仅能实现立体声比特流解码的解码器的后向兼容。
- 任意向下混合信号：MPEG环绕系统不仅能够处理编码器产生的向下混合，而且还可处理在多声道原始信号之外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环绕的空间参数携带于基本音频压缩方案的辅助数据部分。对于向下混合被作为PCM发送的应用，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携带的方法。这种基本技术被称为掩埋数据。

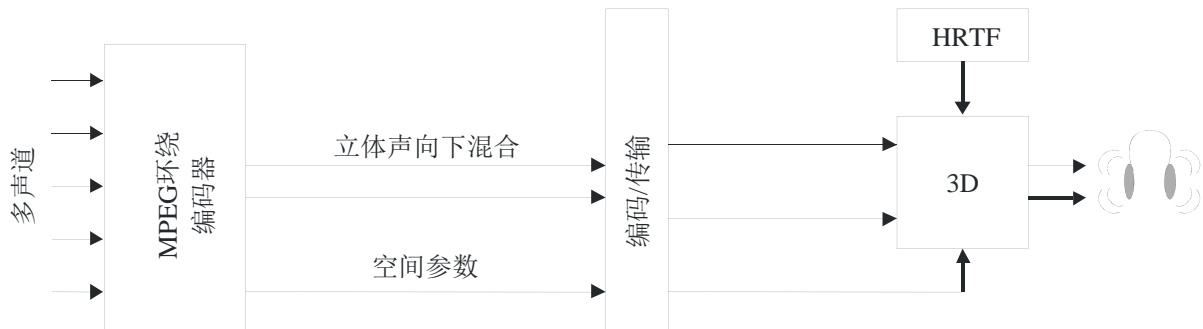
### 3 解码

除提供多声道输出外，MPEG环绕解码器还支持其它输出配置：

- 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。虚拟环绕处理可同时应用于解码器和编码器，后者提供了有关向下混合的虚拟环绕经验，而不必配备MPEG环绕解码器。但是，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



- 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，MPEG环绕技术能够从向下混合中估算出空间边信息，从而创建多声道输入，但可提供超出传统矩阵环绕系统的质量。
- 剪裁：由于基本的结构，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，采用不同范围的采样速率，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，以便确保正常解码。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，可能质量会稍微有所降低。MPEG 环绕解码器的输出的质量和格式还取决于特定解码器的配置。然而，解码器的配置完全与该档次的不同级别有关。

#### 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，同时确保对与已经使用参数工具编码器（例如，频段复制）的连接进行优化。

## 附录 5

### 扩展版高效AAC (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混合体进行编码，并且确保范围广泛的比特率上全部声音素材保持一致的音频质量。该标准支持高比特率的单信道和多信道编码，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则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）等专用语音编码器要素相结合。MPEG-4频带复制（SBR）和MPEG-D MPEG环绕等参数化编码工具得到了增强，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

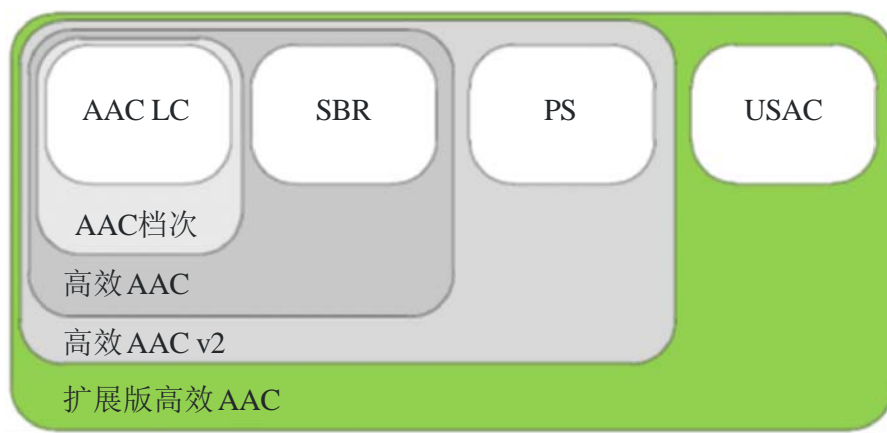
基准USAC档次提供USAC标准的完整功能，同时保持较低的整体计算复杂性。对存储或处理能力需求过大的工具被排除在外。

– 扩展版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）保持兼容性的应用，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HE AAC v2比特流以及USAC比特流（多达两个信道）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 范围内的采样频率，并经证明能够在8 kbit/s至实现感知透明度的比特率范围内提供良好的音频质量。这已在ISO/IEC 联合技术委员会1/29分委会/第11工作组的验证测试中得到证实（MPEG2011/N12232号文件，附于6B/286(Rev.2)号文件）。

信道配置可自由选择。可以有效地为最常见的应用场景发出13种不同的缺省信道配置信号。这些缺省配置包括全部MPEG-4信道配置，例如单声道、立体声、5.0和5.1环绕立体声或者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进程。该标准未明确规定编码进程的标准。

图12给出了一个典型的、可行的编码器结构。

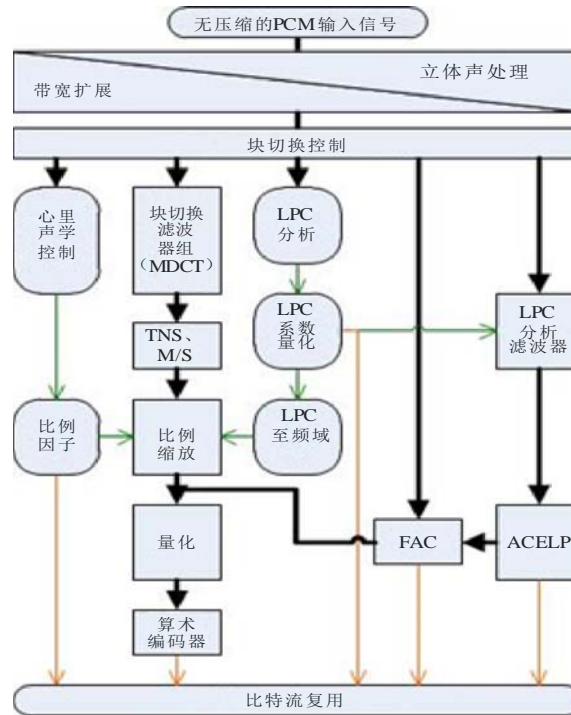
编码器由以下编码工具组成：

- 立体声处理：在中/低等比特率上，USAC采用参数化立体声编码技术。这些技术在原则上与附录2.5所述的PS工具类似，但基于附录4所述的MPEG环绕立体声，因此被称作MPEG环绕2-1-2（2-1-2 MPS）。编码器从输入音频信号中提取立体声像的高效参数表示。这些参数与单声道降混信号一起，在比特流中发射。编码器可以选择发射一个残差信号，以修正解码器端的立体声信号重建进程。残余编码机制允许从全参数化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频带复制（SBR）的多重增强版本，这在附录2.4中有所说明。编码器对更高音频频段的频谱包络和音调进行估算，并向解码器发射对应的参数。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。增强型的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替代数值的3/4，前提是在更低的采样率上实现更佳的时间粒度。
- 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，使用方式类似于附录2.2所述方式。
- 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：代数码激励线性预测（ACELP）编码工具采用了经过验证的适应/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要素组成。

- FAC: 前向走样校正 (FAC) 工具提供了一种机制, 可实现从受走样影响的MDCT编码无缝过渡到基于时域的ACELP编码。

图12

MPEG-D USAC编码器的方框图



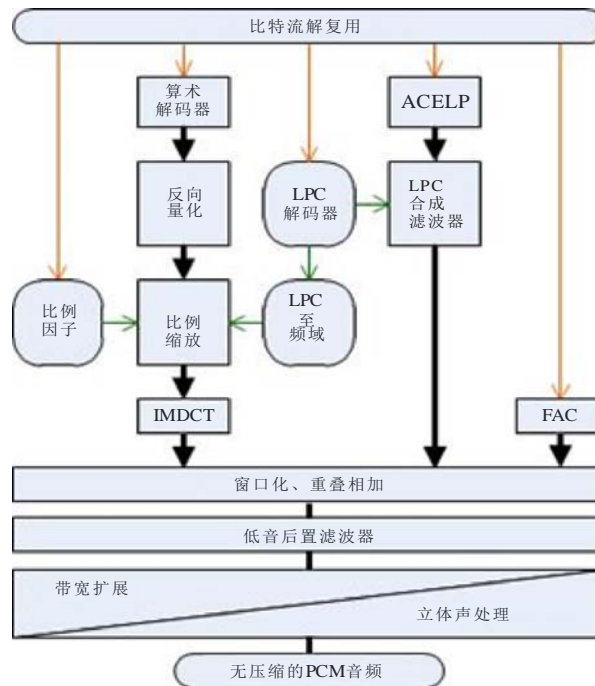
BS.1196-12

### 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

解码进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）基于MDCT的工具（如果存在此类工具并且适用）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连接或合并。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，适用于带宽扩展和立体声编码工具的参数化编码工具被用于重建全带宽、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，在其输入端的数据直接通过该工具，无需任何修改。

#### 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。

- 扩展版HE AAC档次

扩展版高效AAC档次包含高效AAC v2档次的全部工具，因此能够对所有AAC系列档次比特流进行解码。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档次的自然演进版本，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一致性能的附加价值。

---