

# MPEG-2 先进音频编码 (AAC) 的研究和软件仿真

包益平 陈 健

(上海交通大学 上海 200030)

1998 年 9 月 28 日收到

**摘要** MPEG-2 AAC 是 ISO/IEC MPEG 组织继 MPEG-1、MPEG-2 音频编解码标准后, 制定的最新国际标准 (ISO/IEC 13818-7), 在 64kbps/ 每声道低码率时能提供全透明音质。本文对其编解码算法进行了简要介绍。论文作者对 MPEG-2 AAC 进行了软件仿真, 文中介绍了仿真结果, 仿真时对 AAC 标准所附编码范例中的一些不妥之处, 作了改进, 文中列举了 6 方面的修正。

**关键词** MPEG, AAC, 心理声学模型, 数字音频处理

## Study and simulation of the MPEG-2 advanced audio coding(AAC)

Bao Yiping Chen Jian

(Shanghai Jiao Tong Univ, Shanghai 200030)

**Abstract** The MPEG-2 Advanced audio coding (AAC) was the newest standard(ISO/IEC. 13818-7) designed by ISO/IEC MPEG in audio coding. The AAC can deliver "indistinguishable" audio quality at 64kbps/channel data rate. The main features of the AAC system are described briefly. The simulation result of the AAC is presented and an improved scheme on six aspects is proposed.

**Key words** MPEG, AAC, Psychoacoustic model, Digital audio signal processing

### 1 引言

MPEG-2 BC(13818-3) 是与 MPEG-1 音频编解码技术向下兼容的, 除了对多声道与 16kHz、22.5kHz、24 kHz 低采样率的扩展外, 与 MPEG-1 基本相同。MPEG-2 BC 在码率为 640-896 kbit/s/5 声道时, 能提供高质量的音频效果, 但达不到国际电信联盟 (ITU-R)1991 年提出的 384 kbit/s/5 声道传送音频时 “不可区

分的质量” 的规定。MPEG 组织在 1994 年开始制定 AAC, 集成了世界各大公司和组织最新的音频编解码技术, 不再与 MPEG-1 保持兼容, 突破了兼容带来的限制, 1996 年经英国 BBC 和日本 NHK 的测定, 表明 AAC 已经达到 ITU-R 的上述规定。1997 年 4 月, AAC 正式批准为国际标准 (13818-7)。AAC 是数字音频领域新的核心技术。

## 2 AAC 编解码技术概述 [1]

AAC 是利用人耳听觉的掩蔽特性对变换域中的谱线进行编码。编码过程如图 1 所示, 框图中的有些模块是可选的。根据应用的不同, 在编码质量与复杂性之间进行折衷, AAC 分为主要档次、低复杂性档次、可分级档次。除了增益控制模块外, 其余模块主要档次都可能用到。我们主要讨论主要档次。

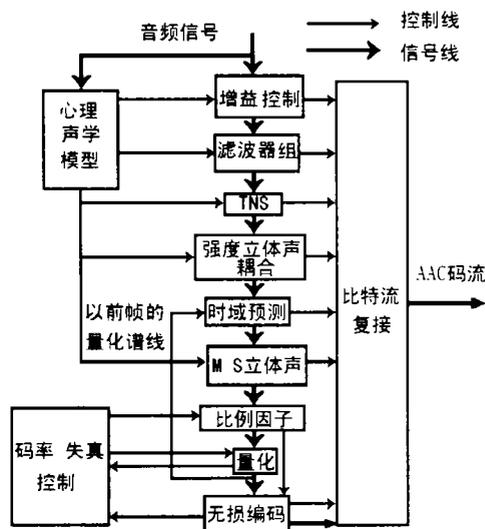


图 1 MPEG-2 AAC 编码框图 [1]

编码时, 时域信号首先经滤波器组分解成为频域谱线, 同时根据输入的时域信号进行心理声学模型分析, 计算当前的掩蔽域值。另外, 心理声学模型还输出中间/边立体声(M/S)和强度立体声处理所需的控制信息。时域噪声成形(TNS)模块对滤波器组输出的部分谱线进行线性预测(LPC), 用得到的残差信号代替原谱线, TNS 控制噪声在时域窗内的分布形状, 获得编码增益。强度立体声模块利用人耳对高频信号的相位不敏感的特点, 只传输高频信号的包络, 大大降低了所需比特数。时域预测与 TNS 不同, 它是对每一条谱线在前后帧之间进行后向预测, 利用帧间信号的相关性获得编码增益。M/S 模块以和信号 M 与差信号 S 代

替左右声道信号, 量化模块进行比特分配, 尽量使量化噪声小于掩蔽域值。量化后的信号与比例因子进行哈夫曼编码, 最后组成 AAC 码流。

AAC 支持各种声道组合方式, 缺省声道配置有单声道、双声道和 5.1 声道, AAC 支持的采样率从 8k 到 96kHz。

与以前的 MPEG 编码技术 [2,3] 相比, AAC 新增加了 TNS、预测、增益控制等技术, AAC 对滤波器组、心理声学模型、立体声处理、多声道处理、哈夫曼编码等都有不同程度的改进, 效果更佳。以下具体阐述。

解码是编码的逆过程, 如图 2 所示。

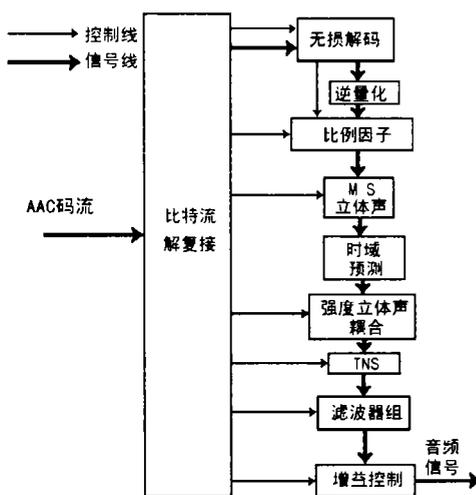


图 2 AAC 解码框图 [1]

## 3 软件仿真

基本上按照 AAC 标准实现了软件仿真。把 44.1kHz 采样的单声道和立体声音频信号压缩为 64kbit/s/ 声道的 AAC 比特流, 再进行解码, 重建的声音质量与原声听不出区别。

仿真时我们对 AAC 标准中的一些不妥之处进行了修正, 提出了改进的方法, 主要有下列几方面。

### 3.1 求扩展函数

按 AAC 标准上计算得到的扩展函数 sprd-

ngf(i,j) 曲线如图 3 示。由心理声学模型的第 6 步知, 扩展函数 sprdngf(i,j) 是 i 处谱线对 j 处谱线产生掩蔽时的曲线。而按照文献 [4], 扩展函数曲线应当如图 4 示, 因为低频信号容易掩蔽高频信号。仿真时我们交换了 i,j 的位置。

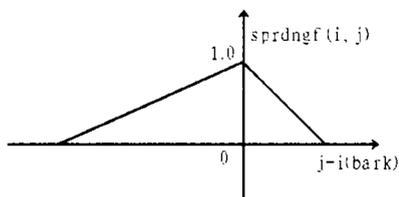


图 3 标准中的 sprdngf(i,j) 曲线

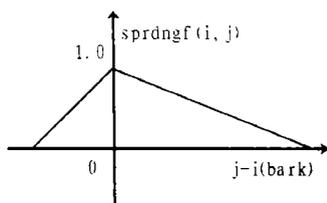


图 4 文献 [4] 中的 sprdngf(i,j) 曲线

### 3.2 短块的预回音控制

AAC 标准的心理声学模型中第 11 步为

$$temp = \min(nb(b), nb_l(b) * rpelev) \quad (1)$$

$$nb(b) = \max(qsthr(b), temp) \quad (2)$$

其中  $nb_l(b)$  是上一块的掩蔽域值,  $qsthr(b)$  是绝对域值, 对短块  $rpelev=1$ , 对长块  $rpelev=2$ 。按上两式若有一短块全是零, 则以后所有短块的掩蔽域值都为绝对域值, 这是不应该的。在 48kHz 采样率时, 因为短块只有 128 点 2.6ms 长, 不需要再做预回音控制。因此, 仿真时对长块的心理声学模型第 11 步不变, 对短块的第 11 步修改为

$$nb(b) = \max(qsthr(b), nb(b)) \quad (3)$$

对鼓声测试, 听不到预回音。

应用声学

### 3.3 M/S 立体声方式开关的决定

AAC 标准中 M/S 控制方式是通过先分别计算开和关的情况下所用的比特数, 然后再决定开还是关, 这种方法计算量太大。仿真时我们根据左右声道信号的相关性决定 M/S 方式开关。考虑某一比例因子频带中左右声道谱线  $lr[k], rr[k]$ , 相关矩阵  $C$  为

$$C = \begin{pmatrix} c_{ll} & c_{lr} \\ c_{lr} & c_{rr} \end{pmatrix} \quad (4)$$

$$\begin{aligned} \text{其中 } c_{ll} &= \frac{1}{N} \sum_{k=0}^{N-1} lr[k]^* lr[k], \\ c_{lr} &= \frac{1}{N} \sum_{k=0}^{N-1} lr[k]^* rr[k], \\ c_{ll} &= \frac{1}{N} \sum_{k=0}^{N-1} rr[k]^* rr[k], \end{aligned}$$

$N$  是该频带谱线的数目。

对 (4) 做 Karhunen-loeve 变换,

$$RCR^T = \Lambda = \begin{pmatrix} \lambda_{ii} & 0 \\ 0 & \lambda_{ee} \end{pmatrix} \quad (5)$$

其中

$$\begin{aligned} R &= \begin{pmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \cos \alpha \end{pmatrix}, \\ \alpha &= \left[ -\frac{\pi}{2}, \frac{\pi}{2} \right] \end{aligned} \quad (6)$$

故旋转角度  $\alpha$  满足

$$\text{tg}(2\alpha) = \frac{2c_{lr}}{c_{ll} - c_{rr}} \quad (7)$$

当  $\alpha=\pi/4$  时, 就是 M/S 方式。因此当  $\pi/8 < |\alpha| < 3\pi/8$  时, 我们置 M/S 方式为开, 否则置为关, 我们称此法为 XY 判据。我们比较了另外三种方法, 即 Dolby 公司在 AC-3 中用的方法、MPEG-1 中用的方法和 J.D.Johnston 在文献 [5] 中用的方法, 结果如图 5 示。其中 NMR- 表示量化噪声小于掩蔽域值的比例因子频带占总频带的百分比, NMR- 所占的百分比越高越好, 以上量化结果的统计表明, 我们所采用的方法 (即 XY 判据) 是最好的。

### 3.4 长短块切换时判决门限 switch\_pe 的确定

于 switch\_pe 时, 由长块切换到短块。AAC 标准没有给出 switch\_pe 的具体值。我们对各种信号的 PE 值进行了统计, 数据如下。

当心理声学模型中计算出的感知熵 PE 大

表 1 单声道 PE 的分布百分比

PE 值	0-500	500-1000	1000-1500	1500-2000	2000-2500	2500-3000	3000-3500	3500-4000	4000-4500	4500-5000
男歌声	0	43.5	56.5	0	0	0	0	0	0	0
月亮河	0	9.65	90.35	0	0	0	0	0	0	0
女歌声	0	52.81	47.19	0	0	0	0	0	0	0
卡门	0	10.17	89.83	0	0	0	0	0	0	0
铃声	2.2	46.98	37.09	7.97	3.02	1.37	1.37	0	0	0
鼓声	0	65.98	24.35	8.12	0.86	0.17	0.17	0	0.35	0

表 2 双声道 PE 的分布百分比

PE 值	0-500	500-1000	1000-1500	1500-2000	2000-2500	2500-3000	3000-3500	3500-4000	4000-4500	4500-5000
男歌声	0	18.5	68	8.5	3.5	0	0	0	0	0
月亮河	0	67.74	31.64	0.62	0	0	0	0	0	0
女歌声	0	85.66	12.69	1.26	0.39	0	0	0	0	0
卡门	0	64.19	33.51	1.72	0.3	0.14	0.14	0	0	0
铃声	2.2	46.98	37.64	7.69	2.75	1.65	1.1	0	0	0
鼓声	0	65.98	24.35	7.77	1.04	0.17	0.17	0	0.35	0

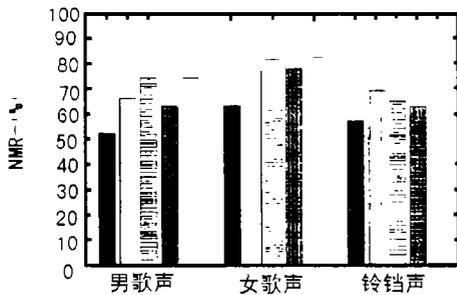


图 5 各种 M/S 判据的测试结果

■ 不用M/S方式    ▨ Johnston方式  
□ Dolby AC-3方式    □ XY方式  
▤ MPEG-1方式

法。要判断谱线对感知是否起重要作用, 涉及人耳的感知特性, 尚无可靠的判据。仿真时, 我们对信号固定地选用 sine 窗或 KBD 窗, 听不出有什么区别。

### 3.6 计算量的考虑

AAC 的计算量非常大, 我们采用了快速算法。N 点 MDCT 用 N/2 点 FFT 实现 [6], N 点实数傅里叶变换用 N/2 点 FFT 实现 [7]。量化模块中 AAC 标准提供的两层循环方法计算量很大, 仿真时还能忍受, 实时处理时可能不行, 选用高速处理芯片如 TI 的 C6x DSP, 同时考虑以牺牲部分性能简化量化过程, 减少计算量。

## 4 结论

在 128 kps/ 双声道时, MPEG 第二、三层、Dolby AC-3、Lucent PAC 的 5 级 MOS

结合信号时域波形的观察, 在一般情况下, PE 小于 1500, 有突发信号时, PE 大于 3000。我们选 switch\_pe 为 2500。

### 3.5 窗口形状的选择

AAC 标准虽提供了 sine 窗和 KBD 窗两种窗口形状, 但没给出自适应窗口形状选择的算

主观评分均不到 4 分, 只有 MPEG-2 AAC 得到了 4.53 的高分<sup>[8]</sup>。MPEG-2 AAC 为低码率高质量音频压缩领域提供了最好的方案, 将是未来 MPEG-4 音频压缩技术的核心。AAC 解码比编码简单得多, 这使 AAC 特别适合于数字音频广播、数字音频存储, 网络上 Hi-Fi 音频的传输等应用。美、日、欧等国计划近几年内推出基于 AAC 技术的数字卫星广播, 研究 AAC 技术, 尤其实时实现时的要求对我国有重大现实意义。

### 参 考 文 献

1 ISO/IEC 13818-7, "Information Technology - Generic Coding of Moving Pictures and Associated Audio, part 7: Advanced Audio Coding", 1997 April.

2 ISO/IEC 11172-3, "Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 Mbit/s, part 3: Audio", 1992.  
3 ISO/IEC 13818-3, "Information Technology-Generic Coding of Moving Pictures and Associated Audio, part 3: Audio", 1994.  
4 Columes c, Lever M, Rault J B, et al. *J.Audio Eng. Soc.*, 1995 **43**(4): 233-239.  
5 Johnston J D, Ferreira A J. *IEEE Proc. ICASSP*, 1992, **2**: 569-572.  
6 Masahiro Iwadare, Akihiko Sugiyama, Fumie Hazu, et al. *IEEE J. on Selected Areas in Communications*, 1992, **10**(1): 138-144.  
7 John Makhoul. *IEEE Trans. on ASSP*, 1980, **28**(1): 27-34.  
8 Soulodre G A, Grusec T, Lavoie M, et al. *J. Audio Eng. Soc.*, **43**(3): 164-177.

~~~~~  
(上接第 14 页)

8 Kanazaws K K, Gordon J G. *Anal. Chem.*, 1985, **57**, 1771.  
9 Schmitt N, Tessier L, Watier H, et al. *Sensors and Actuators*, 1997, B 43, 217-223.  
10 Thompson M, Dhaliwal G K, Arthur C L, et al. *IEEE Trans. on Ultra., Ferro. and Freq. control*, 1987.  
11 Ebato H, Gentry C A, Herron J N, et al. *Anal. Chem.*, 1994, **66**: 1683-1689.  
12 吕安德. 薄膜科学与技术, 1993, **6**(3): 241-246.

13 Chang D F, Yu D M, Hesketh P J, et al. *Sensors and Actuators*, 1996, B 35, 431-434.  
14 Welsch W, Klein C, Schickfus M V, et al. *Sensors and Actuators*, 1997 A 62, 562-564.  
15 Leidl A, Filser H, Labatzki A. *Sensors and Actuators*, 1995, A 46-47, 353-356.  
16 Dahint R, Seigel R R, Harder H, et al. *Sensors and Actuators*, 1996, B 35-36, 497-505.  
17 Du J, Harding G L. *Sensors and Actuators*, 1998, A 65, 152-159.  
18 Morgan C L. *Clin. Chem.* 1996 **42**(2): 193-209.  
19 Go Sakai, Takahiro saiki, Taizo Uda, et al. *Sensors and Actuators*, 1995, B 24-25, 134-137.