

文章编号: 1006 - 5342 (2002) 06 - 0077 - 03

MP3 文件格式剖析*

范建军

(咸宁师范高等专科学校 计算机系, 湖北 咸宁 437005)

摘 要: MP3 的全称是 MPEG Audio Layer - 3. 由于 MP3 具有压缩程度高、音质好的特点, 所以 MP3 是目前最为流行的一种音乐文件. 文章以国际标准化组织 (ISO) 和国际电工委员会 (IEC) 联合颁布的 ISO/IEC 11172 - 3 (MPEG-1 第三部分: 音频) 标准为依据, 结合 MP3 的特点, 详细剖析 MP3 文件的格式结构.

关键词: MPEG; MP3 文件; MP3 解码器

中图分类号: TP37 文献标识码: A

1 MP3 解码器的组成和工作原理^[1]

一个基本的 MP3 解码器组成如图 1 所示:

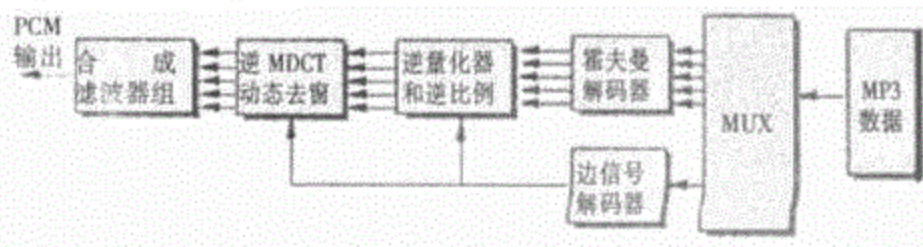


图 1 ISO/MPEG audio 层 3 (MP3) 解码器的结构

它一般包括分流器 (MUX)、霍夫曼解码器、逆量化器、逆 MDCT 器、合成滤波器组和边信号解码器. 其工作流程是从输入的 MP3 比特流中查找第一个帧头部进行同步、接着进行边信号解码、比例因子解码、霍夫曼数据解码、逆量化频谱、频谱重新排序 (动态去窗口)、逆 MDCT 处理、通过多相滤波器组合成, 最后产生 PCM 数据流并送往音箱播放声音.

2 MP3 文件的格式结构^[2]

MP3 音频文件的结构如图 2 所示:



图 2 MP3 音频文件的结构

一个 MP3 音频文件被分成若干个称为帧的小块, 每一个帧都是独立的. 每一个帧都有自己的头部和音频信息. MP3 音频文件没有头部. 因此, 可以切割 MP3 音频文件的任何部分并正确地播放它.

当希望读取有关 MP3 音频文件的信息时, 通常是查找第一个帧就足以获得, 读这第一个帧的头部, 同时假定其它帧也有同样的信息 (并不总是这样的).

Tag 常常被用于描述 MP3 音频文件.

MP3 帧结构如图 3 所示:

帧头
CRC 校验 (可选)
边信号信息
比例因子
音频数据

图 3 MP3 帧结构

在一个帧中, 其头部是由最前面的 4 个字节 (32 位) 组成. 头部最前面的 11 位总是被设置为 1 并称它为“帧同步”. 因此, 可以搜索整个文件以确定第一次出现连续 11 个被设置为 1 的地方, 然后读整个头部并检测其值是否正确. 在下面的表中你将了解到在头部的每一位的精确含义, 以对每个值检测其有效性. 被指定为保留, 无效, 坏和不允许的每一个值说明这是一个无效的头部.

帧可以有一个 CRC 校验, 但这较为罕见. 如果存在 CRC 校验, 这个 CRC 是 16 位长, 它跟在帧的头部之后; 如果你需要读其它帧的头部或仅仅希望计算这个帧的 CRC

* 收稿日期: 2002 - 09 - 20

以比较从文件中读出的帧, 你可以计算帧的长度并使用它。这实际上是一个非常好的检测 MP3 头部有效性的方法。在 CRC 的后面就是 Side Info (边信号信息)。边信号信息后面是 Scale factor (比例因子) 信息。再之后就是音频数据。

2.1 帧头部详细剖析

每个帧头部的长度为 4 字节, 其数据结构如下:

```
typedef struct tagHeader {
    unsigned int sync: 11; // 同步信息
    unsigned int version: 2; // 版本
    unsigned int layer: 2; // 层
    unsigned int error-protection: 1; // CRC 校正
    unsigned int bit-rate-index: 4; // 位率索引
    unsigned int sample-rate-index: 2; // 采样率索引
    unsigned int padding: 1; // 空白字
    unsigned int extension: 1; // 私有标志
    unsigned int channel-mode: 2; // 立体声模式
    unsigned int mode-extension: 2; // 保留
    unsigned int copyright: 1; // 版权标志
    unsigned int original: 1; // 原始媒体
    unsigned int emphasis: 2; // 强调方式
} HEADER, *LPHEADER;
```

其中:

sync (同步信息): 11 位, 全部被设置为 1。

Version (版本): 2 位, 设置如下:

- 00 - MPEG 版本 2.5.
- 01 - 保留.
- 10 - MPEG 版本 2.
- 11 - MPEG 版本 1.

layer (层): 2 位, 设置如下:

- 00 - 保留.
- 01 - 层 III.
- 10 - 层 II.
- 11 - 层 I.

error-protection (出错保护): 1 位, 设置如下:

0 - CRC 保护 (在帧头后面允许 16 位 CRC 校验码)。

1 - 无保护。

bit-rate-index (位率索引): 4 位, 设置如下:

设置值	V1, L1	V1, L2	V1, L3	V2, L1	V2, L2	V2, L3
0000	free	free	free	free	free	free
0001	32	32	32	32	32	8 (8)
0010	64	48	40	64	48	16 (16)
0011	96	56	48	96	56	24 (24)
0100	128	64	56	128	64	32 (32)
0101	160	80	64	160	80	64 (40)
0110	192	96	80	192	96	80 (48)
0111	224	112	96	224	112	56 (56)
1000	256	128	112	256	128	64 (64)
1001	288	160	128	288	160	128 (80)
1010	320	192	160	320	192	160 (96)

设置值	V1, L1	V1, L2	V1, L3	V2, L1	V2, L2	V2, L3
1011	352	224	192	352	224	112 (112)
1100	384	256	224	384	256	128 (128)
1101	416	320	256	416	320	256 (144)
1110	448	384	320	448	384	320 (160)
1111	bad	bad	bad	bad	bad	bad

注释: 所有的值单位是 kbps

V1 - MPEG 版本 1.

V2 - MPEG 版本 2 和 版本 2.5.

L1 - 层 I.

L2 - 层 II.

L3 - 层 III.

“free”意味着可变位率。

“bad”意味着这不是一个允许的值。

sample-rate-index (采样率索引): 2 位, 设置如下:

设置值	MPEG1	MPEG2	MPEG2.5
00	44100	22050	11025
01	48000	24000	12000
10	32000	16000	8000
11	保留	保留	保留

注释: 所有的值单位是 Hz.

Padding (空白字): 1 位, 设置如下:

0 - 帧没有填充位。

1 - 帧带有一个额外的填充位。

Extension (私有标志): 1 位, 目前此位的作用未知。

channel-mode (立体声通道模式): 2 位, 设置如下:

00 - 立体声。

01 - Joint 立体声 (Stereo)。

10 - Dual 立体声 (Stereo)。

11 - Single 立体声 (Mono)。

mode-extension (模式扩展): 2 位, 仅对 Joint 立体声通道模式有效, 设置如下:

设置值	Intensity stereo	MS stereo
00	off	off
01	on	off
10	off	on
11	on	on

Copyright (版权标志): 1 位, 设置如下:

0 - 音频数据没有版权保护。

1 - 音频数据有版权保护。

Original (原始媒体): 1 位, 设置如下:

0 - 原始媒体的拷贝。

1 - 原始媒体。

Emphasis (强调方式): 2 位, 设置如下:

00 - 无

01 - 50/15 ms

10 - 保留

11 - CCITT J.17

2.2 MPEG 音频 Tag 的 MP3v1 详细剖析

它包含的信息有艺术家 (Artist), 标题 (Title), 影集

(Album), 出版年份 (Year) 和音乐风格 (Genre). 对于注释 (Comments) 还有额外的空间. TAG 的大小始终是 128 字节长, 并且位于所有音频数据之后. 可以通过读 MP3 音频文件的最后 128 个字节而获取 Tag.

Tag 各相关信息分配如下:

长度 (字节)	位置 (字节)	描 述
3	(0 - 2)	Tag 标识. 必须包含 'TAG', 如果 Tag 存在并且正确.
30	(3 - 32)	Title
30	(33 - 62)	Artist
30	(63 - 92)	Album

长度 (字节)	位置 (字节)	描 述
4	(93 - 96)	Year
30	(97 - 126)	Comment
1	(127)	Genre

规范要求所有的填充域都用 NULL 字符 (ASCII 0) 填充; 然而, 并不是所有的应用程序都遵守这个规定 (一个例子是 WinAmp 用空格 <space> ASCII 32, 填充填充域).

在 MP3v1.1 结构中, 有一小部分被建议改变. 注释域的最后 1 个字节可以被用于指定在一个影集中歌曲的音轨数. 如果信息是未知, 则它包含一个 NULL 字符 (ASCII 0).

MP3 文件对音乐的风格进行了限制, 共 254 种. Genre 返回的只是 MP3 风格的序列号, 具体还需要定位, 在这里把所有类型以常数形式列出:

0	'Blues'	20	'Alternative'	40	'AlternRock'	60	'Top 40'
1	'Classic Rock'	21	'Ska'	41	'Bass'	61	'Christian Rap'
2	'Country'	22	'Death Metal'	42	'Soul'	62	'Pop/ Funk'
3	'Dance'	23	'Pranks'	43	'Punk'	63	'Jungle'
4	'Disco'	24	'Soundtrack'	44	'Space'	64	'Native American'
5	'Funk'	25	'Euro - Techno'	45	'Meditative'	65	'Cabaret'
6	'Grunge'	26	'Ambient'	46	'Instrumental Pop'	66	'New Wave'
7	'Hip - Hop'	27	'Trip - Hop'	47	'Instrumental Rock'	67	'Psychadelic'
8	'Jazz'	28	'Vocal'	48	'Ethnic'	68	'Rave'
9	'Metal'	29	'Jazz + Funk'	49	'Gothic'	69	'Showtunes'
10	'New Age'	30	'Fusion'	50	'Darkwave'	70	'Trailer'
11	'Oldies'	31	'Trance'	51	'Techno - Industrial'	71	'Lo - Fi'
12	'Other'	32	'Classical'	52	'Electronic'	72	'Tribal'
13	'Pop'	33	'Instrumental'	53	'Pop - Folk'	73	'Acid Punk'
14	'R&B'	34	'Acid'	54	'Eurodance'	74	'Acid Jazz'
15	'Rap'	35	'House'	55	'Dream'	75	'Polka'
16	'Reggae'	36	'Game'	56	'Southern Rock'	76	'Retro'
17	'Rock'	37	'Sound Clip'	57	'Comedy'	77	'Musical'
18	'Techno'	38	'Gospel'	58	'Cult'	78	'Rock & Roll'
19	'Industrial'	39	'Noise'	59	'Gangsta'	79	'Hard Rock'

任何其它的值都被认为是 "Unknown".

MPEG 音频 Tag MP3v2

这是新建议的 TAG 格式, 它与 MP3v1 和 MP3v1.1 不同. 有关它的信息请查看网站 <http://www.id3.com/>.

参考文献:

- [1] 林福宗. 多媒体技术基础[M]. 北京: 清华大学出版社, 2001. 136~143.
- [2] 余崇智, 李存珠, 张志力等. 运动图像及其伴音的压缩与编码[M]. 南京: 南京大学出版社, 1995. 235~250.