

解析 MP3 音频格式

一、概述

MP3 文件是由帧(frame)构成的,帧是 MP3 文件最小的组成单位。MP3 的全称应为 MPEG1 Layer-3 音频文件, MPEG(Moving Picture Experts Group)在汉语中译为活动图像专家组,特指活动影音压缩标准, MPEG 音频文件是 MPEG1 标准中的声音部分,也叫 MPEG 音频层,它根据压缩质量和编码复杂程度划分为三层,即 Layer-1、Layer2、Layer3,且分别对应 MP1、MP2、MP3 这三种声音文件,并根据不同的用途,使用不同层次的编码。MPEG 音频编码的层次越高,编码器越复杂,压缩率也越高,MP1 和 MP2 的压缩率分别为 4: 1 和 6: 1-8: 1,而 MP3 的压缩率则高达 10: 1-12: 1,也就是说,一分钟 CD 音质的音乐,未经压缩需要 10MB 的存储空间,而经过 MP3 压缩编码后只有 1MB 左右。不过 MP3 对音频信号采用的是有损压缩方式,为了降低声音失真度,MP3 采取了“感官编码技术”,即编码时先对音频文件进行频谱分析,然后用过滤器滤掉噪音电平,接着通过量化的方式将剩下的每一位打散排列,最后形成具有较高压缩比的 MP3 文件,并使压缩后的文件在回放时能够达到比较接近原音源的声音效果。

名词解释:

比特率是指每秒传送的比特(bit)数。单位为 bps(Bit Per Second),比特率越高,传送的数据越大。比特率表示经过编码(压缩)后的音、视频数据每秒钟需要用多少个比特来表示,而比特就是二进制里面最小的单位,要么是 0,要么是 1。比特率与音、视频压缩的关系,简单的说就是比特率越高,音、视频的质量就越好,但编码后的文件就越大;如果比特率越少则情况刚好相反。

采样率(也称为采样速度或者采样频率)定义了每秒从连续信号中提取并组成离散信号的采样个数,单位用赫兹(Hz)来表示。采样频率的倒数是采样周期(也称为采样时间),它表示采样之间的时间间隔。这里要注意不要将采样率与位速相混淆。

二、MP3 文件整体结构

MP3 文件大体分为三部分: TAG_V2(ID3V2), Frame, TAG_V1(ID3V1)

ID3V2 标签帧	ID3v2 在文件头,以字符串“ID3”为标志,包含了演唱者,作曲,专辑等信息,长度不固定,扩展了 ID3V1 的信息量。
音频数据帧	一系列的帧,个数由文件大小和帧长决定,每个 FRAME 的长度可能不固定,也可能固定,由位率 bitrate 决定,每个 FRAME 又分为帧头和数据实体两部分,帧头记录了 mp3 的位率,采样率,版本等信息,每个帧之间相互独

	立
ID3V1 标签帧	ID3v1 在文件结尾，以字符串“TAG”为标记，其长度是固定的 128 个字节，包含了演唱者、歌名、专辑、年份等信息。

三、MP3 结构具体分析

1.ID3V2 标签

每个 ID3V2.3 的标签都一个标签头 and 若干个标签帧或一个扩展标签头组成。关于曲目的信息如标题、作者等都存放在不同的标签帧中，扩展标签头和标签帧并不是必要的，但每个标签至少要有有一个标签帧。标签头和标签帧一起顺序存放在 MP3 文件的首部。

标签头

在文件的首部顺序记录 10 个字节的 ID3V2.3 的头部。数据结构如下：

```
char Header[3]; /*必须为"ID3"否则认为标签不存在*/
char Ver; /*版本号 ID3V2.3 就记录 3*/
char Revision; /*副版本号此版本记录为 0*/
char Flag; /*存放标志的字节，这个版本只定义了三位，稍后详细解说*/
char Size[4]; /*标签大小，包括标签头的 10 个字节和所有的标签帧的大小*/
```

注:最后 4 个字节表示 ID3V2 标签的大小，在实际寻找首帧的过程中，我发现大部分的 mp3 文件的标签大小是包含标签头的，但有的又是不包含的，可能是某些 mp3 编码器写标签的 BUG，所以为了兼容只好认为其是包含的，如果按大小找不到，再向后搜索，直到找到首帧为止。

计算 ID3V2 标签帧的大小：

最后四个字节表示，但每个字节只用 7 位，最高位不使用恒为 0。所以格式如下：

0xxxxxxx 0xxxxxxx 0xxxxxxx 0xxxxxxx

计算大小时要将 0 去掉，得到一个 28 位的二进制数，就是标签大小(不懂为什么要这样做)，计算公式如下：

下：

```
int ID3V2_Size;
ID3V2_Size = (Size[0]&0x7F)*0x200000
+(Size[1]&0x7F)*0x4000
+(Size[2]&0x7F)*0x80
+(Size[3]&0x7F)
```

(2) 标签帧

每个标签帧都有一个 10 个字节的帧头和至少一个字节的固定长度的内容组成。它们也是顺序存放在文件中，和标签头和其他的标签帧也没有特殊的字符分隔。得到一个完整的帧的内容只有从帧头中的到内容大小后才能读出，读取时要注意大小，不要将其他帧的内容或帧头读入。

一般我们从标签头中获取信息计算出整个 ID3V2 标签的大小即可，对于 ID3V2 的标签帧就不再仔细分析。

2. 音频数据

数据帧往往有多个，至于有多少，由文件大小和帧大小来决定。每个帧都有一个四字节的帧头，接下来可能有两个字节的 CRC 校验，其存在由帧头中的具体信息决定。接着就是帧的实体数据，也就是 MAIN_DATA 了。

(1) 音频数据帧的帧头

格式如下：

AAAAAAA AAABBCCD EEEFFGH IJJKLMM

符号	长度 (bit)	位置 (bit)	描述
A	11	(31~21)	帧同步 (所有位置 1)
B	2	(20~19)	MPEG 音频版本 ID 00 - MPEG 2.5 01 - 保留 10 - MPEG 2 (ISO/IEC 13818-3) 11 - MPEG 1 (ISO/IEC 11172-3)
C	2	(18~17)	Layer 描述 00 - 保留 01 - Layer III 10 - Layer II 11 - Layer I
D	1	(16)	校验位 (0 - 紧跟帧头后有 16 位即 2 个字节用作 CRC 校验 1 - 没有校验)
E	4	(15~12)	比特率 (位率) (见比特率索引表)
F	2	(11~10)	采样率 (见采样率索引表)
G	1	(9)	填充位 (填充用来达到正确的比特率。) 0 - 没有填充 1 - 填充了一个额外的空位
H	1	(8)	私有 bit，可以用来做特殊应用。例如可以用来触发应用程序的特殊事件。
I	2	(7~6)	声道 00 立体声 01 联合立体声 (立体声) 10 双声道 (立体声)

			11 单声道（单声）
J	2	(5~4)	扩展模式（仅在联合立体声时有效） 扩展模式用来连接对立体声效果无用的信息，来减少所需的资源。这两个位在联合立体声模式下有编码器动态指定。
K	1	(3)	版权 0 无版权 1 有版权
L	1	(2)	原创 0 原创拷贝 1 原创
M	2	(1)	强调 00 - 无 01 - 50/15 ms 10 - 保留 11 - CCIT J.17

【帧头信息表】

索引值	MPEG1			MPEG2&MPEG2.5	
	Layer1	Layer2	Layer3	Layer1	Layer2&3
0000	Free				
0001	32	32	32	32	8
0010	64	48	40	48	16
0011	96	56	48	56	24
0100	128	64	56	64	32
0101	160	80	64	80	40
0110	192	96	80	96	48
0111	224	112	96	112	56
1000	256	128	112	128	64
1001	288	160	128	144	80
1010	320	192	160	160	96
1011	352	224	192	176	112
1100	384	256	224	192	128
1101	416	320	256	224	144
1110	448	384	320	256	160
1111	Bad				

【比特率索引表】

Bits	MPEG1	MPEG2	MPEG3
00	44100	22050	11025
01	48000	24000	12000

10	32000	16000	8000
11	保留		

【采样率索引表】

(2) 如何计算音频数据帧长度

我们首先区分两个术语：帧大小和帧长度。帧大小即每帧采样数表示一帧中采样的个数，这是恒定值。其值入下表所示

	MPEG1	MPEG2	MPEG2.5
Layer1	384	384	384
Layer2	1152	1152	1152
Layer3	1152	576	576

【每帧采样数表】

帧长度是压缩时每一帧的长度，包括帧头。它将填充的空位也计算在内。LayerI 的一个空位长 4 字节，LayerII 和 LayerIII 的空位是 1 字节。当读取 MPEG 文件时必须计算该值以便找到相邻的帧。

注意：因为有填充和比特率变换，帧长度可能变化。

从头中读取比特率，采样频率和填充，

LayerI 使用公式：

帧长度（字节）= 每帧采样数 / 采样频率(HZ) * 比特率（bps）/ 8 + 填充 * 4

LayerII 和 LayerIII 使用公式：

帧长度（字节）= 每帧采样数 / 采样频率(HZ) * 比特率（bps）/ 8 + 填充
例：

LayerIII 比特率 128000，采样频率 44100，填充 0

=> 帧大小 417 字节

(3) 计算每帧的持续时间

之前看了一些文章都说 mp3 的一帧的持续时间是 26ms，结果在实际程序的编写中发现无法正确按时间定位到帧，然后又查了一些文章才知道，所谓 26ms 一帧只是针对 MPEG1 Layer III 而且采样率为 44.1KHz 来说是对的，但 mp3 文件并不都是如此，其实这个时间也是可以通过计算来获得，下面给出计算公式

每帧持续时间(秒) = 每帧采样数 / 采样频率（HZ）

可以这么理解：每帧采用数就是要采取的总数，采样率就是采取的速度，相除就得到时间。

这样通过计算可知 MPEG1 Layer III 采样率为 44.1KHz 的一帧持续时间为 26.12...不是整数，不过我们权且认为它就是 26 毫秒吧。

如果是 MPEG2 Layer III 采样率为 16KHz 的话那一帧要持续 36 毫秒，这个相差还是蛮大的，所以还是应该通过计算来获的，当然可以按 MPEG 版本，层数和采样率来建一个表，这样直接查表就可以知道时间了。

3.ID3V1 标签

ID3v1 标签用来描述 MPEG 音频文件。包含艺术家，标题，唱片集，发布年代和流派。另外还有额外的注释空间。位于音频文件的最后固定为 128 字节。可以读取该文件的最后这 128 字节获得标签。以最后 128 个字节的头三个字节为“TAG”作为有 ID3V1 的判断依据。

结构如下：
AAABBBBB BBBB BBBB BBBB
BCCCCCCC CCCCCCCC CCCCCCCC CCCCCCCC
DDDDDDDD DDDDDDDD DDDDDDDD DDDDEEEE
EEEEEEEE FFFFFFFF FFFFFFFF FFFFFFFF

符号	长度 (byte)	位置	描述
A	3	(0~2)	标签标志。如果存在标签并且正确的话，必须包含'TAG'。
B	30	(3~32)	标题
C	30	(33~62)	艺术家
D	30	(63~92)	唱片集
E	4	(93~96)	年代
F	30	(97~126)	注释
G	1	(127)	流派

有了上述的这些信息，我们就可以自己写代码，从 MP3 文件中抓取信息以及修改文件名了。但是，如果真的想写一个播放软件，还是需要读它的数据帧，并进行解码。

四、解析方法

当你想读取 MPEG 文件的信息时，解析前三个字节，判断是否有 ID3V2 标签，有则根据上面的方法算出 ID3V2 标签的总大小，这样就找到了音频数据帧的第一帧，读取它的头信息，获取比特率、采样率、MPEG 版本号、Layer 描述号等信息，根据上面提供的方法算出每帧的长度和每帧持续时间，对于定比特率的其它帧是相同的，也就是说解析第一帧就达到了目的。但这也不是所有情况。变比特率的 MPEG 文件使用使用所谓比特变换，也就是说每一帧的比特率依照具体内容变化。这时就需要你每一帧都解析。