

著名的jonny设计了PMP格式，让大家可以使用PSP来欣赏到高质量的影音。但随着版本的推进，出现了PMP1.0, PMP2.0, PMP-AVC, 这些名词和概念可能让大家感到很困惑，现在，我根据我的研究和理解来给大家解释一下这些名词的意思。

【PMP格式】

PMP是jonny设计的视频音频封装格式，如同AVI一样，只是一种封装的容器格式。

PMP格式现在只有两个版本，一个是1.0, 一个是2.0

【PMP 1.0】

PMP1.0是一个比较简陋的版本，只支持封装MP4V流(xvid, divx)和MP3流(而且要求是CBR的,采样率为44100)。

从文件头开始，顺序结构如下：

[\[Copy to clipboard\]](#)

CODE:

4字节：一定是"pmpm"，pmp的标志

4字节：版本号,为0，表示版本是1.0

4字节：视频帧总数

4字节：视频宽度

4字节：视频高度

4字节：视频scale

4字节：视频rate，注意：视频的帧率 $\text{fps} = \text{rate}/\text{scale}$

4字节：最大视频帧的大小

视频帧总数×4个字节：每帧视频的索引，每个索引4个字节，最低一个bit位表示是否关键帧，其余的31位bit表示帧的大小。

视频数据：字节数，由上面的索引计算可以得出。

4个字节：音频帧的数据大小（由于采用的是cbr模式，所有的音频帧数据大小一样，但关键帧的大小比普通帧大小多一个字节）

4个字节：音频帧的总数

音频帧总数×1个字节：每帧音频的索引，每个索引1个字节，关键帧为1，普通帧为

0，注意：每帧的实际大小=音频帧的数据大小+索引值。

音频数据：字节数，由上面的索引计算得出。

从上面的分析来看，PMP1.0格式有很大的不足，没有视频和音频的标志位，也就是说，固定死了视频和音频的编码格式，无法封装各式的流，同时，由于视频和音频数据是非交错存储，播放程序在回放的时候，文件指针在来回地移动，读视频帧的时候移到前面，读音频帧又移到后面。

【PMP 2.0】

PMP2.0开始，jonny估计意识到1.0的不足，重新设计了文件的格式结构，个人认为这个改变很不错：

从文件头开始，顺序结构如下：

[\[Copy to clipboard\]](#)

CODE:

4字节：一定是"pmpm"，pmp的标志

4字节：版本号,为1，表示版本是2.0

4字节：视频格式标志，这是一个改进，为支持封装各种视频流提供保证，0表示MP4V流(xvid, divx)，1表示AVC流(PMP-AVC其实就是PMP2.0格式，只不过封装了AVC流)

4字节：视频帧总数

4字节：视频宽度

4字节：视频高度

4字节：视频scale

4字节：视频rate，注意：视频的帧率fps = rate/scale

4字节：音频格式标志，同样为了以后支持封装各种音频提供保证，现在只支持mp3流，该值为0；

4字节：包含的音频流数量，为支持多音轨封装提供了保证，如果一个pmp中封装了两条音轨，该值就为2

4字节：每帧视频附带的音频帧的最大数，由于PMP2.0采用了视频音频交错存储的方式，一帧视频和相应的几帧音频放在一起，这是一个最大值；

4字节：音频scale,默认为1152

4字节：音频rate,默认为44100

4字节：音频是否立体声，0表示单声，1表示立体声

视频帧总数×4个字节：每帧视频的索引，每个索引4个字节，最低一个bit位表示是否关键帧，其余的31位bit表示帧的大小（注意，这里的帧大小和1.0格式的帧大小不一样，看下面的解释）

视频音频混和数据：这里，jonny做了一个比较有意思的设计，把1个视频帧和其相应的音频帧混合起来，当做一个数据帧；那究竟怎么个混合法呢，我们分析一下：

首先计算每个视频帧的时间戳和每个音频帧的时间戳， $\text{videotime} = \text{videonum} / \text{videofps}$ ， $\text{audiotime} = \text{audionum} / \text{audiofps}$ （videonum和audionum都从0开始）

那么第一个视频帧的时间戳就是 $0/\text{videofps} = 0$ ，而第二个视频帧的时间戳就是 $1/\text{videofps}$ ，然后，把第一个视频帧和所有时间戳小于 $1/\text{videofps}$ 的音频帧接在一起，成了一个数据帧，如此类推下去；

然后在每个数据帧的前面再加上n个字节， $n = 1 + 4 + 4 + 4 + 4 \times \text{每音轨被混合的帧数} \times \text{音轨数}$ ，

其中这n个字节的意义如下：

1字节：本数据帧中，每音轨被混合的帧数；

4字节：被混合的第一个音频帧和视频帧之间的时间差；

4字节：被混合的最后一个音频帧和视频帧之间的时间差；

4字节：视频帧的大小；

$4 \times \text{每音轨被混合的帧数} \times \text{音轨数}$ ：每个被混合的音频帧的大小