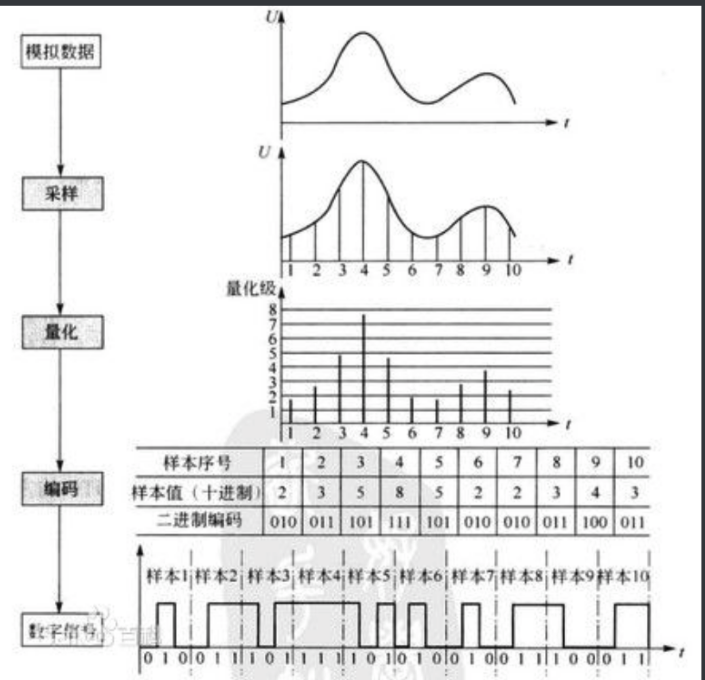
**一、概述**

    PCM [脉冲编码调制](https://baike.baidu.com/item/%E8%84%89%E5%86%B2%E7%BC%96%E7%A0%81%E8%B0%83%E5%88%B6/5143788)是Pulse Code Modulation的缩写。脉冲编码调制是数字通信的[编码方式](https://baike.baidu.com/item/%E7%BC%96%E7%A0%81%E6%96%B9%E5%BC%8F/6451499" \t "_blank)之一。主要过程是将话音、图像等模拟信号每隔一定时间进行取样，使其离散化，同时将抽样值按分层单位四舍五入取整量化，同时将抽样值按一组二进制码来表示抽样脉冲的幅值。

**二、语音编码原理**

    有一定电子基础的都知道传感器采集音频信号是模拟量，而我们实际传输过程中使用的是数字量。而这就涉及到模拟转数字的过程。而[模拟信号](https://baike.baidu.com/item/%E6%A8%A1%E6%8B%9F%E4%BF%A1%E5%8F%B7)数字化必须经过三个过程，即抽样、量化和编码，以实现话音数字化的[脉冲编码调制](https://baike.baidu.com/item/%E8%84%89%E5%86%B2%E7%BC%96%E7%A0%81%E8%B0%83%E5%88%B6)（PCM，Pulse Coding Modulation）技术。



### 1、抽样(Sampling)

    抽样是把[模拟信号](https://baike.baidu.com/item/%E6%A8%A1%E6%8B%9F%E4%BF%A1%E5%8F%B7" \t "_blank)以其信号带宽2倍以上的频率提取样值，变为在时间轴上离散的抽样信号的过程。

**采样率 (sample)**：每秒从连续信号中提取并组成离散信号的采样个数，用赫兹（Hz）来表示.

举例：  
如音频信号采样率为8000hz。  
可以理解上图采样对应图中 那段电压随时间变化的曲线 为1秒 那下面那个1 2 3 …10那就因该有1-8000个点，即将1秒均分为8000份，依次取出来那8000个点时间 对应的电压值。

### 2、量化（quantizing）

    抽样信号虽然是时间轴上离散的信号，但仍然是[模拟信号](https://baike.baidu.com/item/%E6%A8%A1%E6%8B%9F%E4%BF%A1%E5%8F%B7" \t "_blank)，其样值在一定的取值范围内，可有无限多个值。显然，对无限个样值一一给出数字码组来对应是不可能的。为了实现以数字码表示样值，必须采用“四舍五入”的方法把样值分级“取整”，使一定取值范围内的样值由无限多个值变为有限个值。这一过程称为量化。

    量化后的抽样信号与量化前的抽样信号相比较，当然有所失真，且不再是模拟信号。这种量化失真在接收端还原模拟信号时表现为噪声，并称为量化噪声。量化噪声的大小取决于把样值分级“取整”的方式，分的级数越多，即量化级差或间隔越小，量化噪声也越小。

**采样位数：**指的是描述数字信号所使用的位数。  
    8位(8bit)代表2的8次方=256，16 位(16bit)则代表2的16次方=65536； 采样位数越高，精度越高。

这里引用了采样位数来描述模拟信号之间的最小间隔。  
假设我们采样位数是 8 ，而模拟信号的范围是2，0，那么数字信号之间的间隔最小就是2/2^8= 2/256=1/128;  
同理采样位数是16，那么 数字信号之间的最小间隔就是2/256/256=1/(128\*256)

举例  
如音频传感器采集到的电压范围为0-3.3V，采样位数为8bit(位)  
即我们把3.3V/ 2^8 = 0.0128 即为量化精度。  
我们把3.3v分成0.0128为步进的Y轴，如图3中的1 2 …8就变成了0 0.0128 0.0256 …..3.3 V  
比如某个采样点的电压值为1.652V(128 \* 0.128 与 129 \* 0.128之间) 我们对它四舍五入就为1.65V 对应 量化等级为128 。

### 3、编码（Coding）

    量化后的抽样信号就转化为按抽样时序排列的一串十进制数字码流，即十进制数字信号。简单高效的数据系统是二进制码系统，因此，应将十进制数字代码变换成二进制编码，根据十进制数字代码的总个数，可以确定所需二进制编码的位数，即字长(采样位数)这种把量化的抽样信号变换成给定字长的二进制码流的 过程称为编码。

举例  
接着上面的1.65V 对应 量化等级为128 。对应的2进制为10000000 。即该采样点编码后结果为10000000 。当然这是没有考虑正负值情况下的编码方式，而且编码方式种类非常多，需要具体问题具体分析。（PCM音频格式编码为A律13折线编码）

## 三、PCM 音频编码

    PCM信号未经过任何编码和压缩处理(无损压缩)。与模拟信号比，它不易受传送系统的杂波及失真的影响。动态范围宽，可得到音质相当好的效果。编码上采用A律13折线编码。

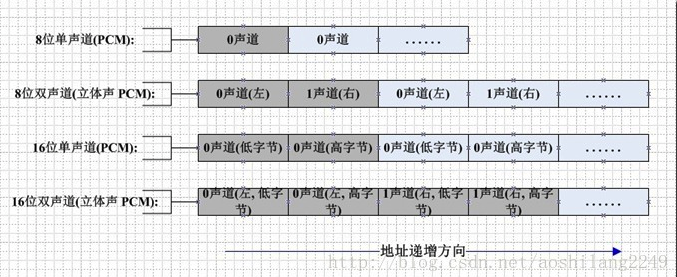
### 1、A律13折线

    A律是P**CM非均匀量化**中的一种对数压扩形式。数字脉冲编码调制（PCM）是目前模拟信号数字化的基本方法，PCM包括采样、量化、编码三个步骤，其中量化是对抽样值的取值离散，根据量化间隔的不同选取分为均匀量化和非均匀量化，非均匀量化可以有效地改善信号的量化信噪比。语音信号的量化常采用ITU建议的两种对数形式的**非均匀量化压缩**特性：A律和μ律，A律编码主要用于30/32路一次群系统， A律PCM用于欧洲和中国。

### 2、声道

    声道可以分为单声道和立体声（双声道）  
    PCM的每个样本值包含在一个整数i中，i的长度为容纳指定样本长度所需的最小字节数。

首先存储低有效字节，表示样本幅度的位放在i的高有效位上，剩下的位置为0，这样8位和16位的PCM波形样本的数据格式如下所示。



### 3、采样频率

    人对频率的识别范围是 20HZ - 20000HZ, 如果每秒钟能对声音做 20000 个采样, 回放时就足可以满足人耳的需求.

* 8000hz 为电话采样。
* 22050 的采样频率是常用的。
* 44100已是CD音质, 超过48000的采样对人耳已经没有意义

对采样率为44.1kHz的AAC（Advanced Audio Coding）音频进行解码时，一帧的解码时间须控制在23.22毫秒内。通常是按1024个采样点一帧。

**为什么这里需要说下音频帧呢？**  
音频的帧的概念没有视频帧那么清晰，几乎所有视频编码格式都可以简单的认为一帧就是编码后的一副图像。但音频帧跟编码格式相关，它是各个编码标准自己实现的。因为如果以PCM（未经编码的音频数据）来说，它根本就不需要帧的概念，根据采样率和采样精度就可以播放了。比如采样率为44.1kHZ，采样精度为16位的音频，你可以算出bitrate（比特率）是4410016kbps，每秒的音频数据是固定的4410016/8 字节。  
但是我们不希望每一次采样都返回给我们进行处理，我们希望的是返回一段时间内的所有采样数据。这里的音频帧就是每次返回给我们多少个采样数据，一般情况是下返回2048个采样数据。  
那么单声道 采用16位采样位数 2048个采样数据的大小是多少呢 2048\*16/8 = 4096字节。

### 4、采样位数

    每个采样数据记录的是振幅, 采样精度取决于储存空间（采样位数）的大小:

* 1 字节(也就是8bit) 只能记录 256 个数, 也就是只能将振幅划分成 256 个等级
* 2 字节(也就是16bit) 可以细到 65536 个数, 这已是 CD 标准了;
* 4 字节(也就是32bit) 能把振幅细分到 4294967296 个等级, 实在是没必要了

    如果是双声道(stereo), 采样就是双份的, 文件也差不多要大一倍.

### 5、比特率

    码率是指经过编码后的音频数据每秒钟需要用多少个比特来表示.

### 6、有损和无损

    对于我们最常说的“无损音频”来说，一般都是指传统CD格式中的16bit/44.1kHz采样率的文件格式，而知所以称为无损压缩，也是因为其包含了20Hz-22.05kHz这个完全覆盖人耳可闻范围的频响频率而得名。

这里我有个混淆的地方是 声道和采样频率之间的关系？刚开始假设采样频率是44100,要是使用双声道的话，那么每个声道的采样率就是22100了。其实这是错误的，采样频率是在每个声道上的采样速率，不是在所有频道的采样速率。  
因此，要是采样速率是44100 ，那么双声道，采集的样本个数应该是88200个。

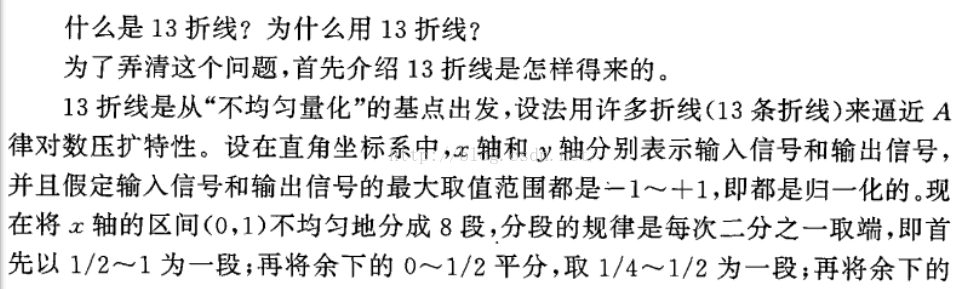
## 四、A律13折

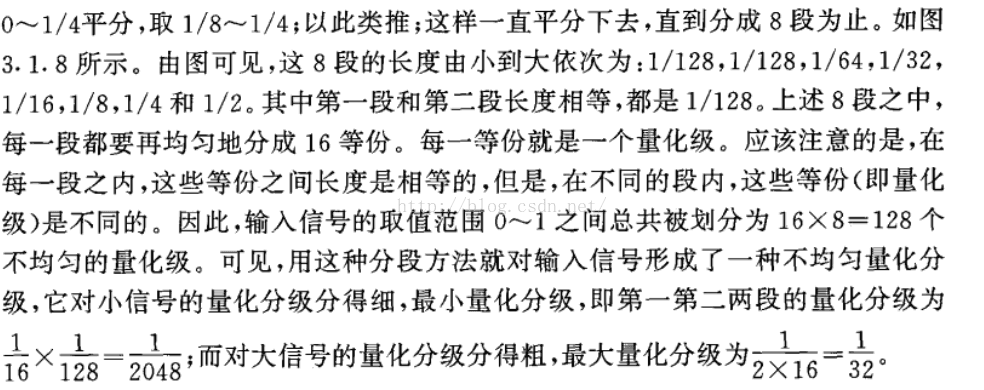
归纳：

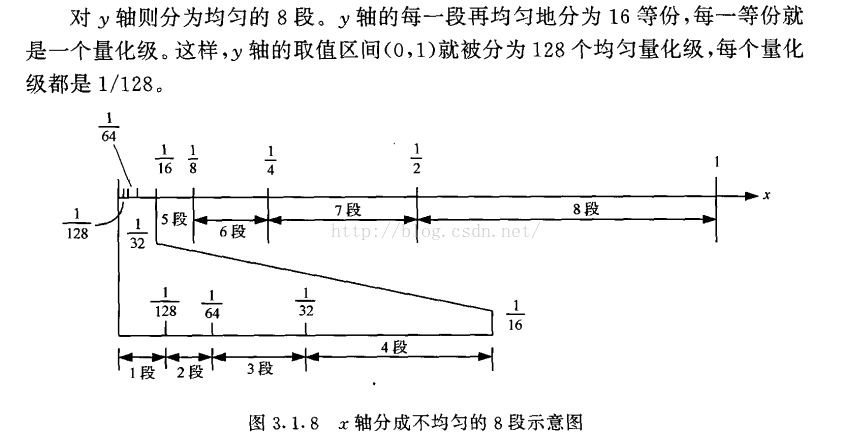
（1）y轴等分成8分，而x轴按1/2比逐次缩小化取值，总共连成8个转折点，由于第一第二个转折点的斜率相等，因此只需考虑7个点。

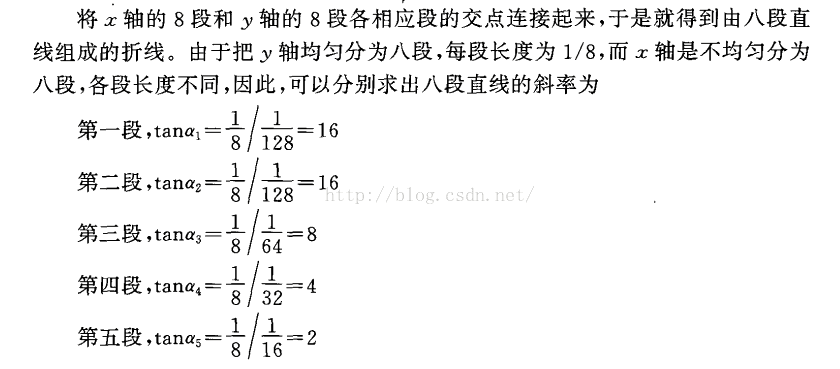
（2）由于有正负之分，所以在负端也会有7个点。也就是正负端应有14个点。

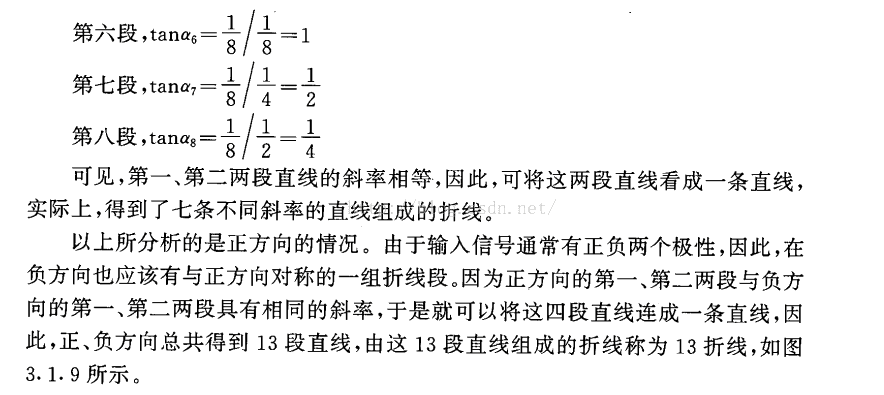
（3）由于正方的一二段和负端的一二段斜率一样，因而正负端的两个点可只考虑一个点的斜率，因而只需考虑13点。这就是13点折线的来历。

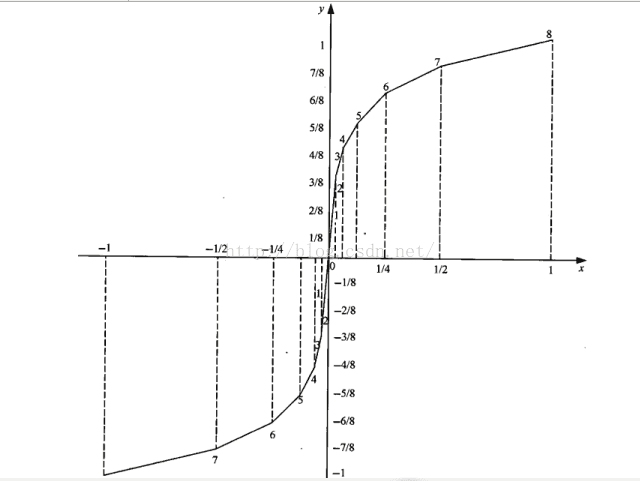












### AAC 格式详解

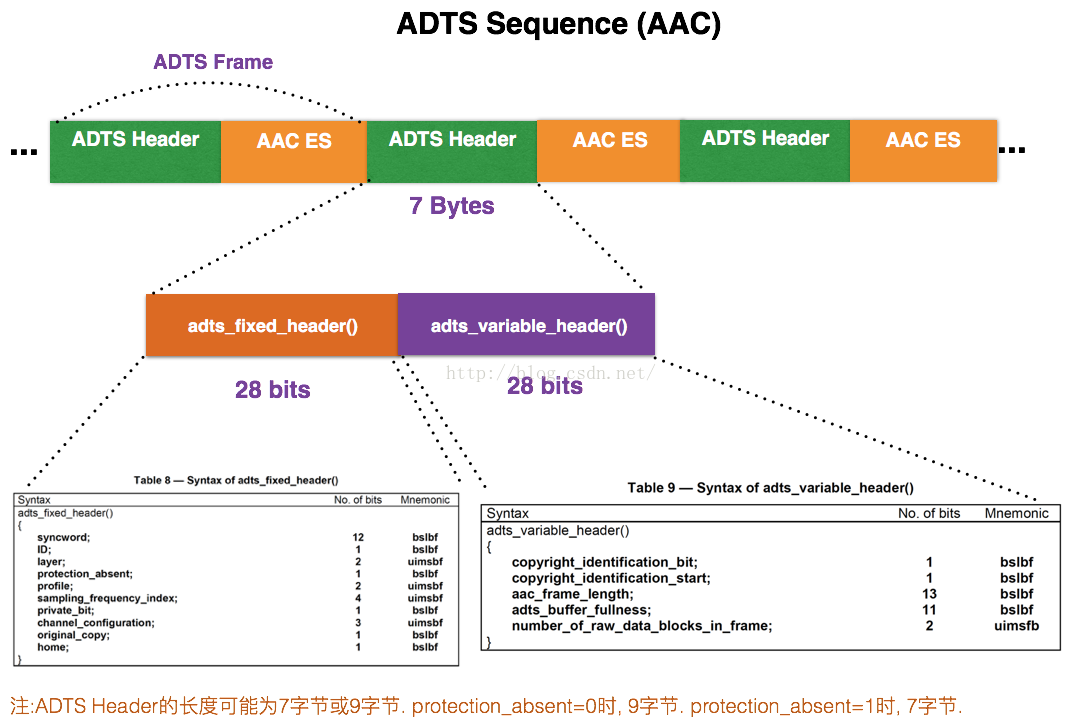
1. ADTS 头 文件信息介绍

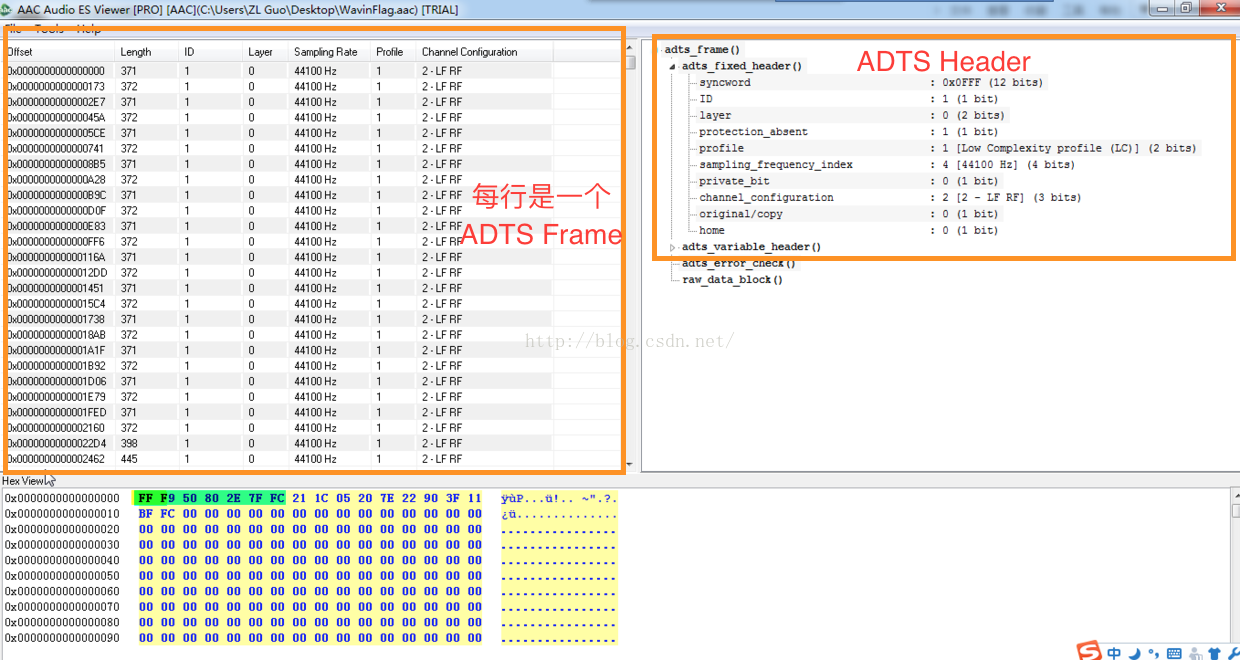
ADTS：Audio Data Transport Stream，是AAC音频文件常见的传输格式。

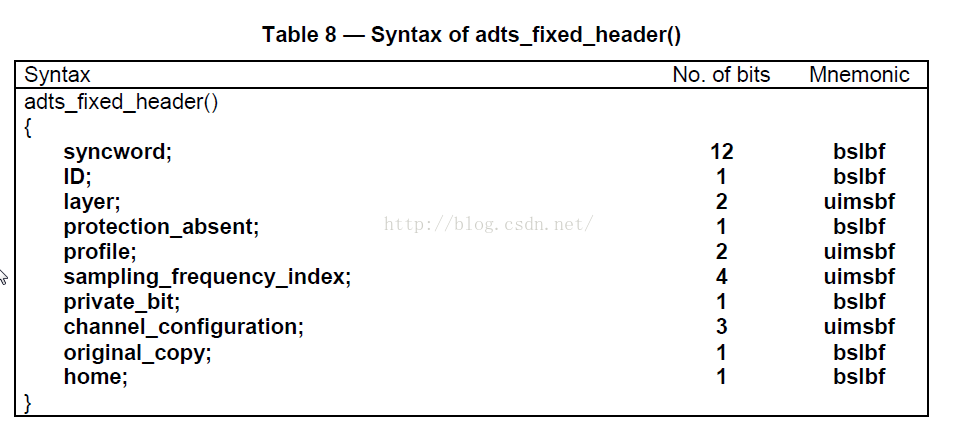
有的时候当你编码AAC裸流的时候，会遇到写出来的AAC文件并不能在PC和手机上播放，很大的可能就是AAC文件的每一帧里缺少了ADTS头信息文件的包装拼接。只需要加入头文件ADTS即可。 一个AAC原始数据块长度是可变的，对原始帧加上ADTS头进行ADTS的封装，就形成了ADTS帧。

下面分析下ADTS头文件结构和信息：

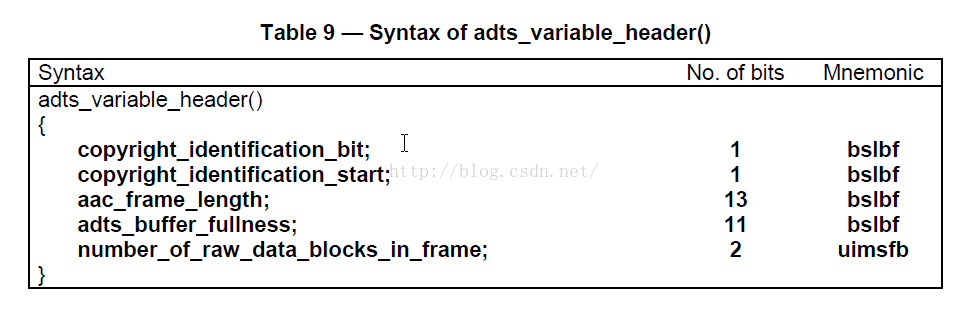
AAC音频文件的每一帧由ADTS Header和AAC Audio Data组成。







* **syncword ：**总是0xFFF, 代表一个ADTS帧的开始, 用于同步.  
  解码器可通过0xFFF确定每个ADTS的开始位置.  
  因为它的存在，解码可以在这个流中任何位置开始, 即可以在任意帧解码。
* **ID：**MPEG Version: 0 for MPEG-4，1 for MPEG-2
* **Layer**：always: '00'
* **protection\_absent：**Warning, set to 1 if there is no CRC and 0 if there is CRC
* **profile：**表示使用哪个级别的AAC，如01 Low Complexity(LC) -- AAC LC  
  profile的值等于 Audio Object Type的值减1.  
  profile = MPEG-4 Audio Object Type – 1
* **sampling\_frequency\_index：**采样率的下标
* **channel\_configuration：**声道数，比如2表示立体声双声道



* **aac\_frame\_length：**一个ADTS帧的长度包括ADTS头和AAC原始流。frame length, this value must include 7 or 9 bytes of header length:  
  aac\_frame\_length = (protection\_absent == 1 ? 7 : 9) + size(AACFrame)  
    
  protection\_absent=0时, header length=9bytes  
  protection\_absent=1时, header length=7bytes
* **adts\_buffer\_fullness：**0x7FF 说明是码率可变的码流。
* **number\_of\_raw\_data\_blocks\_in\_frame：**表示ADTS帧中有number\_of\_raw\_data\_blocks\_in\_frame + 1个AAC原始帧。  
  所以说number\_of\_raw\_data\_blocks\_in\_frame == 0 表示说ADTS帧中有一个AAC数据块。  
  (一个AAC原始帧包含一段时间内1024个采样及相关数据)