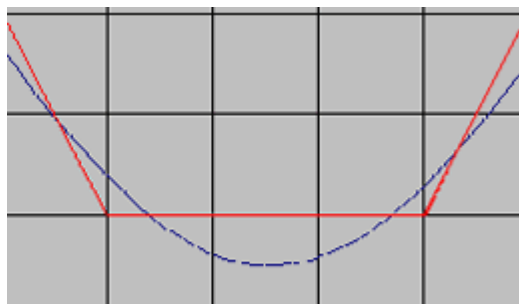


将音频数字化，其实就是将声音数字化。最常见的方式是透过脉冲编码调制PCM(Pulse Code Modulation)。运作原理如下。首先我们考虑声音经过麦克风，转换成一连串电压变化的信号，如图一所示。这张图的横座标为秒，纵座标为电压大小。要将这样的信号转为 PCM 格式的方法，是使用三个参数来表示声音，它们是：**声道数**、**采样位数**和**采样频率**。

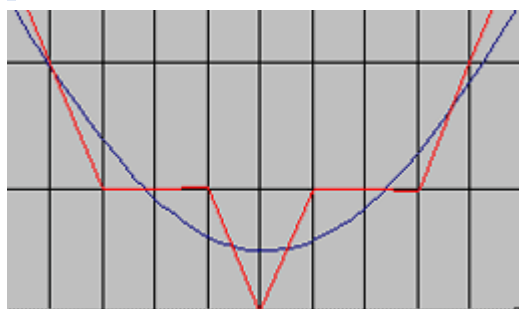
采样频率：即取样频率,指每秒钟取得声音样本的次数。采样频率越高,声音的质量也就越好,声音的还原也就越真实，但同时它占的资源比较多。由于人耳的分辨率很有限,太高的频率并不能分辨出来。在16位声卡中有22KHz、44KHz等几级,其中，22KHz相当于普通FM广播的音质，44KHz已相当于CD音质了，目前的常用采样频率都不超过48KHz。

采样位数：即采样值或取样值（就是将采样样本幅度量化）。它是用来衡量声音波动变化的一个参数，也可以说是声卡的分辨率。它的数值越大，分辨率也就越高，所发出声音的能力越强。

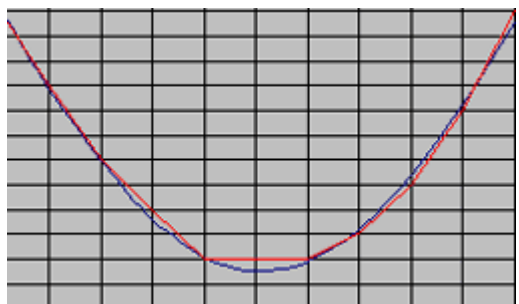
声道数很好理解，有单声道和立体声之分，单声道的声音只能使用一个喇叭发声（有的也处理成两个喇叭输出同一个声道的声音），立体声的pcm可以使两个喇叭都发声（一般左右声道有分工），更能感受到空间效果。



声波曲线1



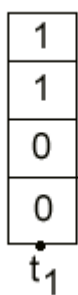
声波曲线2



声波曲线3

下面再用图解来看看采样位数和采样频率的概念。让我们来看看这几幅图。图中的黑色曲线表示的是pcm文件录制的自然界的声波，红色曲线表示的是pcm文件输出的声波，横坐标便是采样频率；纵坐标便是采样位数。这几幅图中的格子从左到右，逐渐加密，先是加大横坐标的密度，然后加大纵坐标的密度。显然，当横坐标的单位越小即两个采样时刻的间隔越小，则越有利于保持原始声音的真实情况，换句话说，采样的频率越大则音质越有保证；同理，当纵坐标的单位越小则越有利于音质的提高，即采样的位数越大越好。

在计算机中采样位数一般有8位和16位之分，但有一点请大家注意，8位不是说把纵坐标分成8份，而是分成2的8次方即256份；同理16位是把纵坐标分成2的16次方65536份；而采样频率一般有11025HZ (11KHz) ， 22050HZ (22KHz) 、 44100Hz (44KHz) 三种。



样点	t1	t2	t3	t4	t5	t6	t7	...	t1
幅值	0011	0101	0111	1001	1011	1101	1110	...	01

那么，现在我们可以得到pcm文件所占容量的公式：

存储量=(采样频率*采样位数*声道)*时间/8(单位：字节数)

例如，数字激光唱盘(CD - DA, 红皮书标准)的标准采样频率为44.1kHz, 采样数位为16位, 立体声(2声道), 可以几乎无失真地播出频率高达22kHz的声音, 这也是人类所能听到的最高频率声音。激光唱盘一分钟音乐需要的存储量为:

$$(44.1 \times 1000 \times 16 \times 2) \times 60 / 8 = 10,584,000 \text{ (字节)} = 10.584 \text{ MBytes}$$

这个数值就是pcm声音文件在硬盘中所占磁盘空间的存储量。

计算机音频文件的格式决定了其声音的品质, 日常生活中电话、收音机等均为模拟音频信号, 即不存在采样频率和采样位数的概念, 我们可以这样比较一下:

44KHz, 16BIT的声音称作: CD音质;

22KHz、16Bit的声音效果近似于立体声 (FM Stereo) 广播, 称作: 广播音质;

11kHz、8Bit的声音, 称作: 电话音质。

微软的WAV文件就是pcm编码的一种。