

数据率为56 000 bps或64 000 bps。由于每秒只有8000个样本，所以4 kHz以上的频率就丢失了。

音频CD是以每秒44 100个样本的采样率进行数字化的，足以捕获最高达到22 050 Hz的频率，这对于人而言是很好的，但是对于狗而言却是很差的。每一样本在其振幅范围内以16位进行线性量化。注意，16位样本只有65 536个不同的值，而人耳以最小可听度为步长进行测量时的动态范围大约为一百万。所以每个样本只有16位引入了某些量化噪声（尽管没有覆盖全部动态范围，但是人们并不认为CD的质量受到损害）。以每秒44 100个样本、每个样本16位计算，音频CD需要的带宽单声道为705.6 Kbps，立体声为1.411 Mbps（参见图7-2）。音频压缩也许要以描述人类听觉如何工作的心理声学模型为基础。使用MPEG第3层（MP3）系统进行10倍的压缩是可能的。采用这一格式的便携式音乐播放器近年来已经十分普遍。

数字化的声音可以十分容易地在计算机上用软件进行处理。有许许多多的个人计算机程序可以让用户从多个信号源记录、显示、编辑、混合和存储声波。事实上，所有专业的声音记录与编辑系统如今都是数字化的。模拟方式基本上过时了。

7.3 视频压缩

现在我们已经十分清楚，以非压缩格式处理多媒体信息是完全不可能的——它的数据量太大了，惟一的希望是有可能进行大比例的数据压缩。幸运的是，在过去几十年，大量的研究群体已经发明了许多压缩技术和算法，使多媒体传输成为可能。在下面几节中，我们将研究一些多媒体数据（特别是图像）的压缩方法，更多的细节请参见（Fluckiger, 1995；Steinmetz和Nahrstedt, 1995）。

所有的压缩系统都需要两个算法：一个用于在源端对数据进行压缩，另一个用于在目的端对数据进行解压缩。在文献中，这两个算法分别被称为编码（encoding）算法和解码（decoding）算法，我们在本书中也使用这样的术语。

这些算法具有某些不对称性，这一不对称性对于理解数据压缩是十分重要的。首先，对于许多应用而言，一个多媒体文档（比如说一部电影）只需要编码一次（当该文档存储在多媒体服务器上时），但是需要解码数千次（当该文档被客户观看时）。这一不对称性意味着，假若解码算法速度快并且不需要昂贵的硬件，那么编码算法速度慢并且需要昂贵的硬件也是可以接受的。从另一方面来说，对于诸如视频会议这样的实时多媒体而言，编码速度慢是不可接受的，在这样的场合，编码必须即时完成。

第二个不对称性是编码/解码过程不必是100%可逆的。也就是说，当对一个文件进行压缩并进行传输，然后对其进行解压缩时，用户可以期望取回原始的文件，准确到最后一位。对于多媒体，这样的要求是不存在的。视频信号经过编码和解码之后与原始信号只存在轻微的差异通常就是可以接受的。当解码输出不与原始输入严格相等时，系统被称为是有损的（lossy）。所有用于多媒体的压缩系统都是有损的，因为这样可以获得更好的压缩效果。

7.3.1 JPEG标准

用于压缩连续色调静止图像（例如照片）的JPEG（Joint Photographic Experts Group，联合摄影专家组）标准是由摄影专家在ITU、ISO和IEC等其他标准组织的支持下开发出来的。JPEG标准对于多媒体而言是十分重要的，因为用于压缩运动图像的标准MPEG不过是分别对每一帧进行JPEG编码，再加上某些帧间压缩和运动补偿等额外的特征。JPEG定义在10918号国际标准中。它具有4种模式和许多选项，但是我们在这里只关心用于24位RGB视频的方法，并且省略了许多细节。

用JPEG对一幅图像进行编码的第一步是块预制。为明确起见，我们假设JPEG输入是一幅640×480的RGB图像，每个像素24位，如图7-6a所示。由于使用亮度和色度可以获得更好的压缩效果，所以从RGB值中计算出一个亮度信号和两个色度信号，对于NTSC制式，分别将其记作Y、I和Q，对于PAL制式，分别将其记作Y、U和V，两种制式的计算公式是不同的。下面我们将使用NTSC的符号，但是压缩算法是相同的。

对Y、I和Q构造不同的矩阵，每个矩阵其元素的取值范围在0到255之间。接下来，在I和Q矩阵中对由4个元素组成的方块进行平均，将矩阵缩小至320×240。这一缩小是有损的，但是眼睛几乎注意不到，因为眼睛对亮度比对色度更加敏感，然而这样做的结果是将数据压缩了2倍。现在将所有三个矩阵的每