

换言之,运动的平滑性是由每秒不同的图像数决定的,而闪烁则是由每秒刷新屏幕的次数决定的。这两个参数是不同的。一幅静止的图像以每秒20帧的频率显示不会表现出断断续续的运动,但是却会出现闪烁,因为当一帧画面在视网膜上消退时下一帧还没有出现。一部电影每秒有20个不同的帧,在80 Hz的刷新率下每一帧将连续绘制4次,这样不会出现闪烁,但是运动将是断断续续的。

当我们考虑在网络上传输数字视频所需要的带宽时,这两个参数的重要性就十分清楚了。目前许多计算机显示器都采用4:3的纵横比,所以可以使用便宜的并且大量生产的显像管,这样的显像管本来是为电视市场的消费者设计的。显示器常用的配置有640×480 (VGA)、800×600 (SVGA)、1024×768 (XGA) 以及1600×1200 (UXGA)。每像素24位的UXGA显示以及25帧/秒,需要1.2Gbps的带宽,即使VGA显示也需要184Mbps。将这些速率加倍以避免闪烁是没有吸引力的,更好的解决方案是每秒传输25帧,同时让计算机保存每一帧并将其绘制两次。广播方式的电视没有使用这一策略,因为电视机没有存储器,并且模拟信号如果不首先转换成数字形式无论如何也无法存放在RAM中,而模数转换则需要额外的硬件。因此,隔行扫描对于广播方式的电视而言是需要的,但是对数字视频则不需要。

7.2.2 音频编码

音频(声音)波是一维的声(压)波。当声波进入人耳的时候,鼓膜将振动,导致内耳的小骨随之振动,将神经脉冲送入大脑,这些脉冲被收听者感知为声音。类似地,当声波冲击麦克风的时候,麦克风将产生电信号,将声音的振幅表示为时间的函数。

人耳可以听到的声音的频率范围从20 Hz到20 000Hz,而某些动物,特别是狗,能够听到更高频率的声音。耳朵是以对数规律听声音的,所以两个振幅为A和B的声音的比率习惯以dB(分贝)为单位来表示,公式为

$$\text{dB} = 20 \log_{10}(A/B)$$

如果我们定义1 kHz正弦波可听度的下限(压力大约为0.0003 dyne/cm²)为0 dB,那么日常谈话大约为50 dB,而使人感到痛苦的阈值大约为120 dB,动态范围为一百万量级。为避免混淆,上面公式中的A和B是振幅。如果我们使用的是功率水平,则上面公式中对数前面的系数应该为10,而不是20,因为功率与振幅的平方成正比。

音频波可以通过模数转换器(Analog Digital Converter, ADC)转换成数字形式。ADC以电压作为输入,并且生成二进制数作为输出。图7-5a中为一个正弦波的例子。为了数字化地表示该信号,我们可以每隔 ΔT 秒对其进行采样,如图7-5b中的条棒高度所示。如果一个声波不是纯粹的正弦波,而是正弦波的线性叠加,其中存在的最高频率成分为 f ,那么以 $2f$ 的频率进行采样就足够了。1924年贝尔实验室的一位物理学家Harry Nyquist从数学上证明了这一结果,这就是著名的Nyquist抽样定理。更多地采样是没有价值的,因为如此采样可以检测到的更高的频率并不存在。

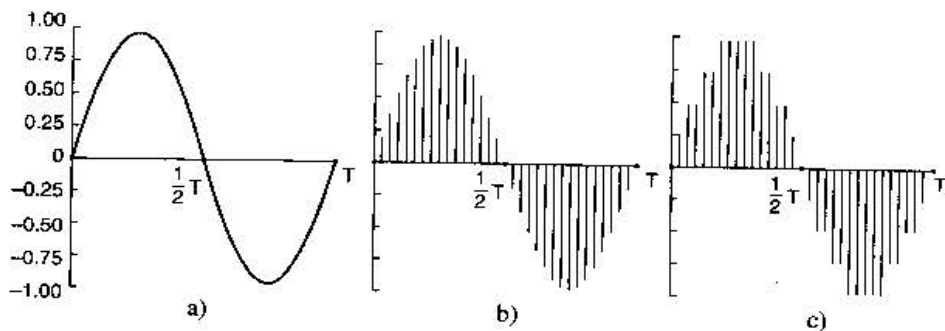


图7-5 a) 正弦波, b) 对正弦波进行采样, c) 对样本进行4位量化

数字样本是不准确的。图7-5c中的样本只允许9个值,从-1.00到1.00,步长为0.25,因此,需要4个二进制位来表示它们。8位样本可以有256个不同的值,16位样本可以有65 536个不同的值。由于每一样本的位数有限而引入的误差称为量化噪声(quantization noise)。如果量化噪声太大,耳朵就会感觉到。

对声音进行采样的两个著名的例子是电话和音频CD。电话系统使用的是脉冲编码调制(pulse code modulation),脉冲编码调制每秒以7位(北美和日本)或8位(欧洲)对声音采样8000次,故这一系统的