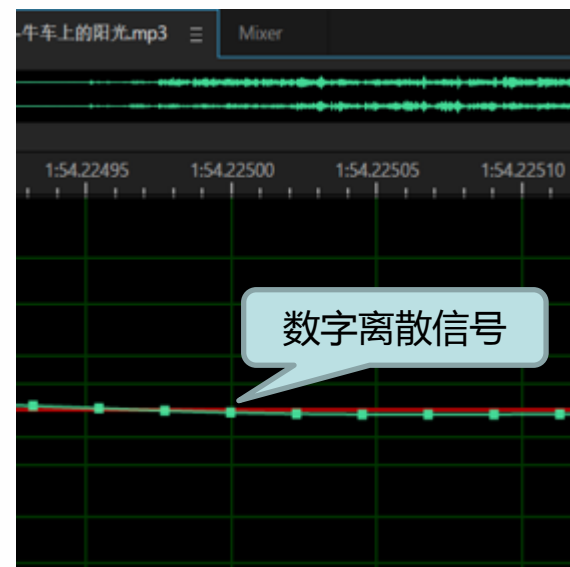
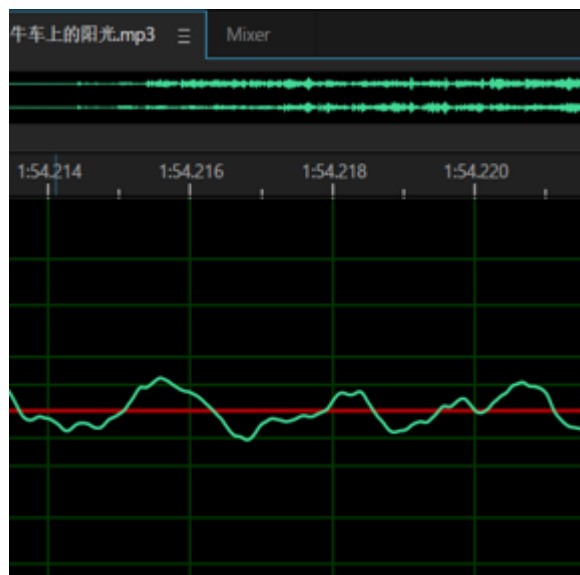
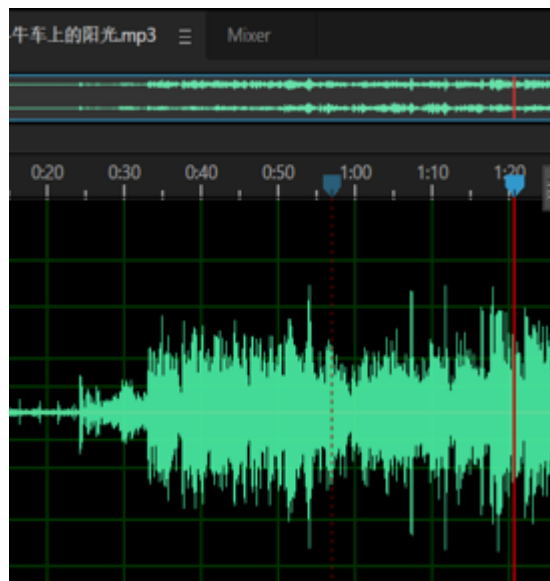




自然界声音信号数字化

背景:

我们用录音软件实际采集声音并保存成计算机中的MP3媒体信息: 实际是一种**数字离散信号**



在声音编辑软件中拉伸

继续拉伸

为了更加形象, 我们继续看下面一页的动态拉伸过程



背景:

我们用录音软件实际采集声音并保存成计算机中的MP3媒体信息: 实际是一种**数字离散信号**



1 Second

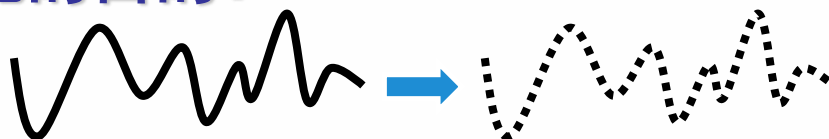




数字离散信号

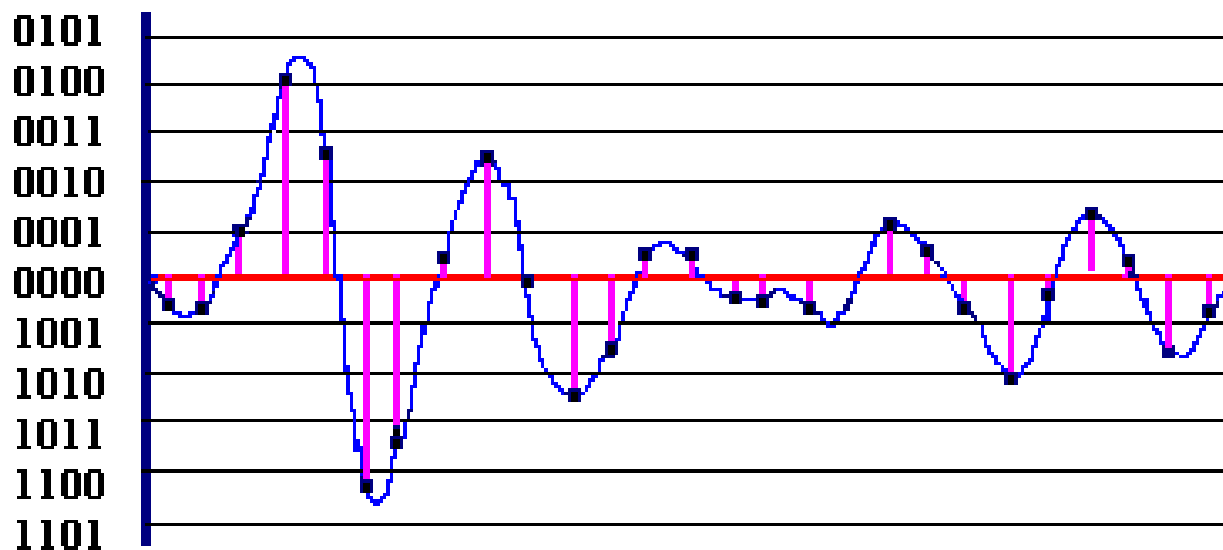
自然界声音数字化的目的:

自然界模拟信号(连续)



计算机中数字信号(离散)

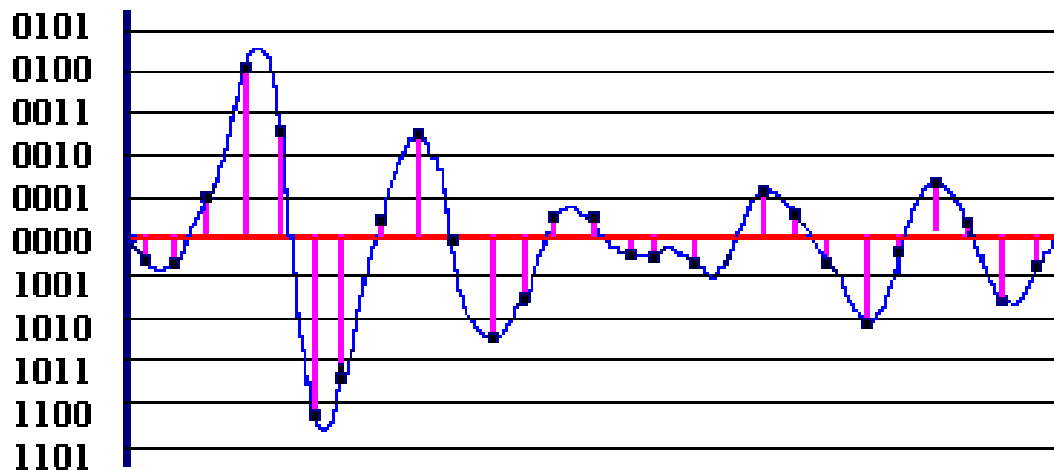
自然界声音数字化的过程:



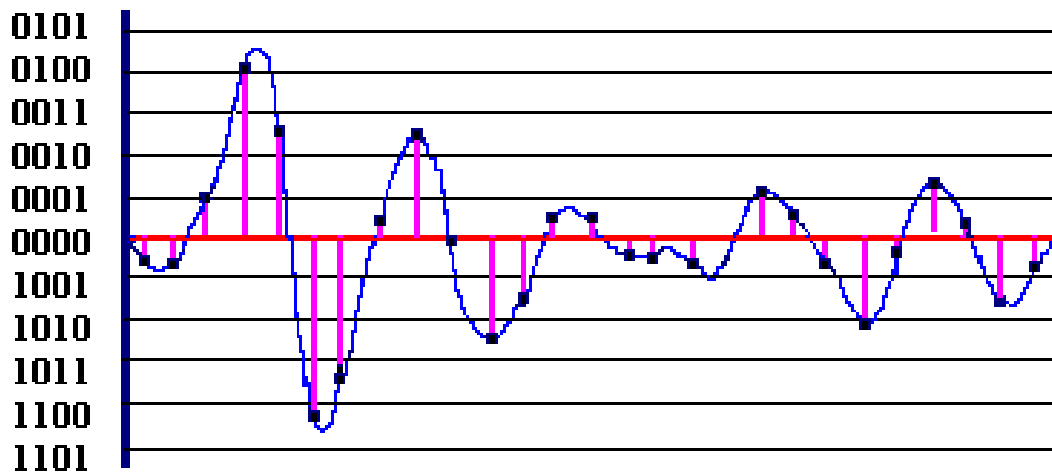


对模拟音频信号进行采样量化编码后，得到数字音频。**数字音频的质量**取决于

- 采样频率
- 量化位数
- 声道数



我们先分别介绍这3个概念，然后解释为什么数字音频的质量取决于这3个参数？



- **采样频率**是指一秒钟时间内采样的次数。
- 在计算机多媒体音频处理中，采样频率通常采用三种：**11.025KHz**(语音效果)、**22.05KHz**(音乐效果)、**44.1KHz**(高保真效果)。常见的**CD**唱盘的采样频率即为**44.1KHz**。



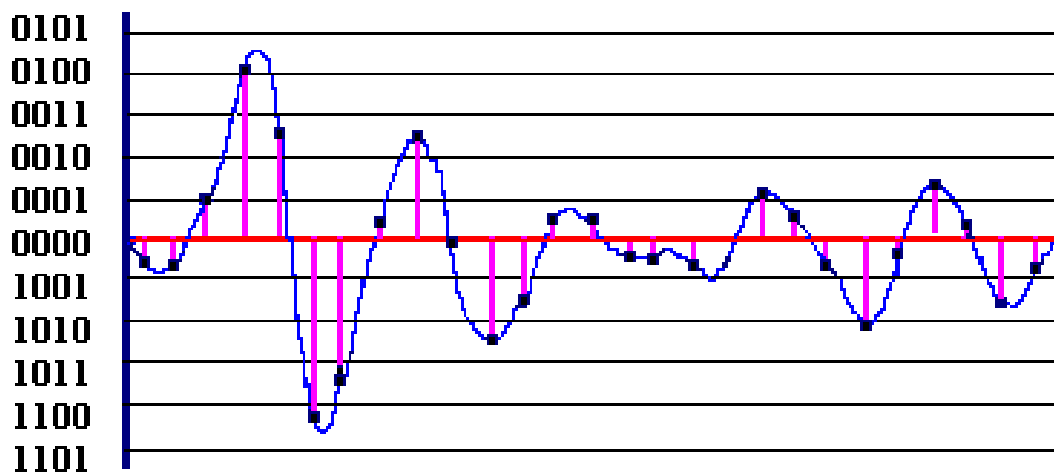
digitization

英 [ˌdɪdʒɪtəlˈzeɪʃn]

美 [ˌdɪdʒɪtəˈzeɪʃn]

采样定律用公式表示为 $f_s \geq 2f$

其中 f 为被采样信号的最高频率。



- **量化位数**，是描述每个采样点样值的二进制位数。
- 例如，**8位量化位数**表示每个采样值可以用 2^8 即**256**个不同的量化值之一来表示，而**16位量化位数**表示每个采样值可以用 2^{16} 即**65536**个不同的量化值之一来表示。常用的量化位数为**8位、12位、16位**。



样本大小是用每个声音样本的位数bit/s (即bps)表示的，它反映度量声音波形幅度的精度。例如，每个声音样本用16位(2字节)表示，测得的声音样本值是在0~65535的范围里，它的精度误差上限就是输入信号的1/65536。样本位数的大小影响到声音的质量，位数越多，声音的质量越高，而需要的存储空间也越多；位数越少，声音的质量越低，需要的存储空间越少。

这与前面讲的图像媒体数字化中的量化位数是一致的。。。



$\text{SNR} = 3 \log_2 [(V_{\text{signal}})^2 / (V_{\text{noise}})^2] = 6 \log_2 (V_{\text{signal}} / V_{\text{noise}})$

其中， V_{signal} 表示信号电压， V_{noise} 表示噪声电压；SNR的单位为分贝(db)

例1：假设 $V_{\text{noise}}=1$ ，采样量化位数为1位表示 $V_{\text{signal}}=2^1$ ，它的信噪比 $\text{SNR}=6$ 分贝。

假设 $V_{\text{noise}}=1$ ，采样量化位数为16位表示 $V_{\text{signal}}=2^{16}$ ，它的信噪比 $\text{SNR}=96$ 分贝。



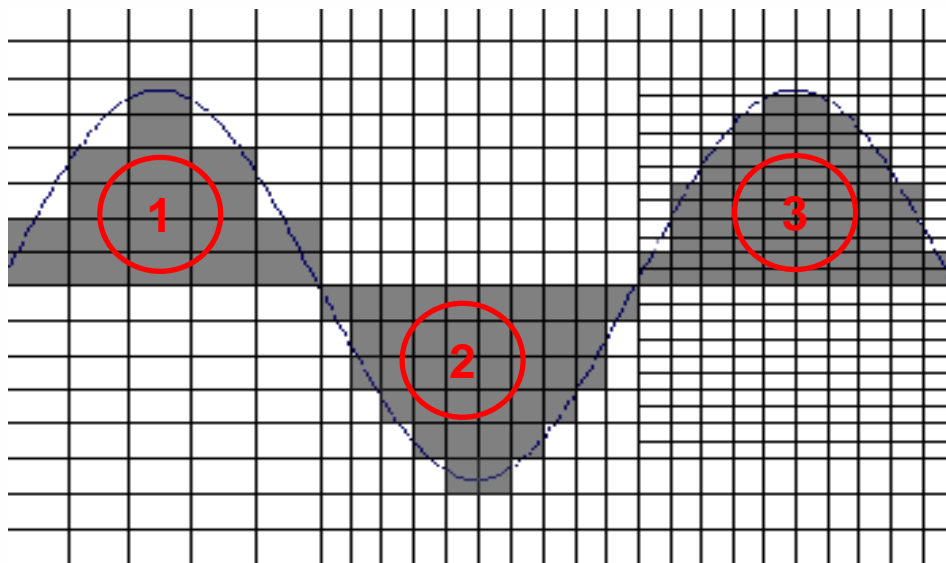
- 声音通道的个数称为声道数，是指一次采样所记录产生的声音波形个数。
- 记录声音时，如果每次生成一个声波数据，称为单声道；每次生成两个声波数据，称为双声道（立体声）。随着声道数的增加，所占用的存储容量也成倍增加。



为什么数字音频的质量取决于

- 采样频率
- 量化位数
- 声道数

这3个参数呢？



先思考这个问题：对于图中的正弦波半波形，1,2,3中哪块阴影区域是其最佳近似？

显然是阴影区域3，因为3具有更高的采样频率和更高的量化位数。
但存储它却需要最大的存储空间，见下一页。。。



采样频率 (1/s) × 量化位数 (bit) × 声道数

数据率 (bit/s) × 持续时间 (s) / 8

推导出： 存储量=采样频率×量化位数/8×声道数×时间

例：采样率11.025KHz、量化位8位，采集1分钟，

则：音频数据率=11.025 (KHz) × 8 (bit)

$$= 88.2 \text{ (Kbit/s)}$$

音频数据量 = 11.025 (KHz) × 8 (bit)

$$\times 60 \text{ (s)} / 8 = 0.66 \text{ (MByte)}$$



- $$16 \times 44.1 \times 1000 \times 2 \times (3.5 \times 60) \div 8 = 37044000 \text{ (B)}$$

37044000/1024/1024=35.33MB



以上都是有代表性的声音形式



编码算法与音频数据压缩比

- 未压缩的音频数据量非常大，因此在编码的时候常采用压缩的方式。
- 即采用一定的算法来压缩数据以减少存储空间和提高传输效率。
- 压缩编码的基本指标之一就是压缩比，一般为数据压缩前后的数据量之比：

$$\text{音频数据压缩比} = \frac{\text{压缩前的音频数据量}}{\text{压缩后的音频数据量}}$$



举例1

例1. 高保真立体声数字音频的量化位数为16，试计算其数据传输率。

解：高保真立体声数字音频采样频率为44.1kHz，双声道，其数据传输率为：

$$\text{数据传输率} = 44.1 (\text{kHz}) \times 16 (\text{bit}) \times 2 (\text{声道}) = 1411.2 (\text{kbps})$$

- 数字音频文件所占用的空间可用如下的公式计算：

$$\text{音频数据量 (Byte)} = \text{数据传输率} \times \text{持续时间} / 8 (\text{bit/Byte})$$

其中数据量以字节(Byte)为单位；数据传输率以每秒比特(bps)为单位；持续时间以秒(s)为单位。



举例2

例2. 计算1分钟未经压缩高保真立体声数字声音文件大小。

- 解：高保真立体声数字音频采样频率为44.1kHz，16位量化，双声道，其数据传输率为：

数据传输率 =

$$44.1 (\text{kHz}) \times 16 (\text{bit}) \times 2 (\text{声道}) = 1411.2 (\text{kbps})$$

- 1分钟这样的声音文件的大小为：

音频数据量 =

$$1411.2 (\text{kb/s}) \times 60 (\text{s}) / 8 (\text{bit/Byte}) = 10584 \text{ kB}$$

因而：未经压缩的4分钟的歌曲文件约42M数据，128M的MP3播放器只能存放3首这样的歌曲。



数字音频文件格式

- 数字声音文件格式是数字音频在磁盘文件中的存放形式，相同的数据可以有不同的文件格式，而不同的数据也可以有相同的文件格式。
 - **WAVE 格式** （无损压缩）
 - **MP3 格式** （有损压缩）
 - **RA 格式** （高压缩率有损压缩）



WAVE文件格式

- WAVE文件是一种通用的音频数据文件，文件扩展名为“.WAV”，Windows系统和一般的音频卡都支持这种格式文件的生成、编辑和播放。
- CD激光唱盘中包含的就是WAVE格式的波形数据。一般说来，声音质量与其WAVE格式的文件大小成正比。
- WAVE文件的**特点**是易于生成和编辑，但在保证一定音质的前提下压缩比不够，不适合在网络上播放。



MP3文件

- MP3文件是采用MP3算法压缩生成的数字音频数据文件，以“.MP3”为文件后缀。
- MP3利用MPEG制定的MPEG-1 Audio layer 3的压缩标准，将音频信息用10:1甚至12:1压缩率变成容量较小的数据文件。
- MP3是一种利用了人类心理声学特性的有损压缩，人耳基本不能分辨出失真，音质几乎达到了CD音质标准。按照这种算法，10张原始未压缩CD的内容可以压缩到1张CD-ROM中，而且声音质量好。



RA文件

- Real Audio是Real networks推出的一种音乐压缩格式，它的压缩比可达到96:1，因此在网上比较流行。
- 经过压缩的音乐文件可以在通过速率为14.4kb/s的MODEM上网的计算机中流畅回放。其最大特点是可以采用流媒体的方式实现网上实时播放，即边下载边播放。



声音的应用

应用场合	要求	
	采样频率	量化位数
国际互联网 (语音、简单乐曲)	11,025 Hz	8 bit
游戏 (效果音、效果音乐)	22,050 Hz	8 bit
多媒体自学读物 (提示音)	11,025 Hz	8 bit
电子教案 (语音、效果音)	11,025 Hz	8 bit
多媒体宝典、大全 (乐曲、语音)	22,050 Hz	8 bit
多媒体音乐鉴赏 (音乐、解说)	44,100 Hz	16 bit

- 应根据使用场合和要求转换适当的聲音采样频率。
- 采样频率的转换须使用相应的软件进行。