一文看懂音频原理

笔者最近正好在做和声音处理有关的项目，突然对音频数字化感兴趣，想了解一下基本原理。可网上文章知识都很散、排版也不美观。因此笔者便决定自己写一篇文章，整合一下数字音频的基本知识。

本篇博客为面向大众的科普性文章。涉及声音原理、音频文件属性、音频格式等方面。预计阅读时间为10分钟。

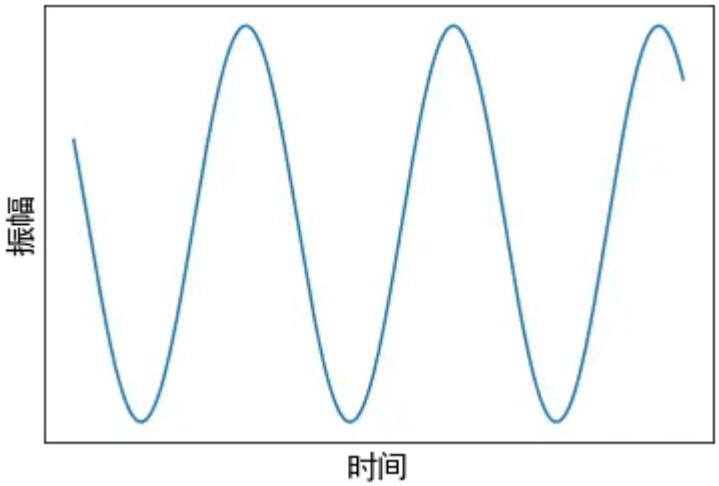
1.何为声音

中学物理中我们知道，声音是物体振动产生的声波。声音通过介质（空气、固体、液体）传入到人耳中，带动听小骨振动，经过一系列的神经信号传递后，被人所感知。

声音是一种波。物体振动时会使介质（如空气）产生疏密变化，从而形成疏密相见的纵波。

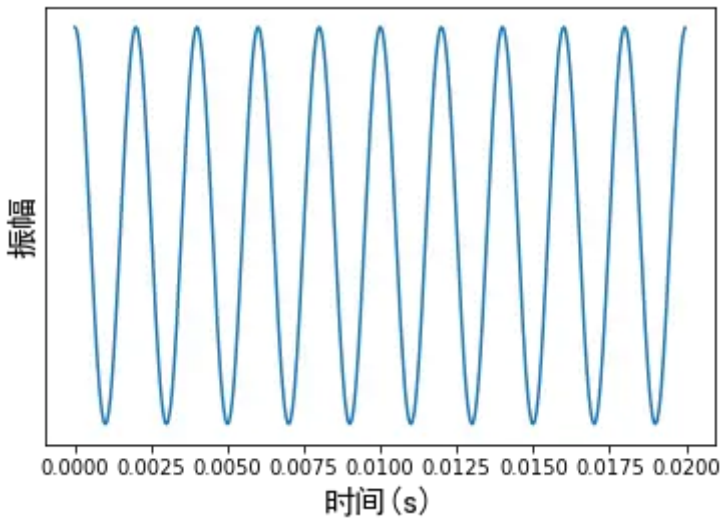
既然声音是波，那么我们就可以用图的形式来表示它。

给定空间中某一点，该点的空气疏密随时间的变化如下：



波形图

下图是一个正弦波，其周期为0.002s，频率为500HZ。



该声音很像视频中的“消音”处理。

频率（音调）：声音1秒内周期性变化的次数

人耳的听觉范围在20Hz-20kHz。 低频的声音沉闷厚重，高频的声音尖锐刺耳。 高于 20kHz的声音为超声波。

振幅（响度）：声音的大小

有的时候，我们用分贝（dB）形容声音大小。值得注意的是，**dB是一个比值，是一个数值，没有任何单位标注。（功率强度之比的对数的10倍）**

1分贝	刚能听到的声音
15 分贝以下	感觉安静
30 分贝	耳语的音量大小
40 分贝	冰箱的嗡嗡声
60分贝	正常交谈的声音
70分贝	相当于走在闹市区
85分贝	汽车穿梭的马路上
95分贝	摩托车启动声音
100分贝	装修电钻的声音
110分贝	卡拉OK、大声播放MP3 的声音
120分贝	飞机起飞时的声音
150分贝	燃放烟花爆竹的声音

2.声音采集与存储

采样，指把时间域或空间域的连续量转化成离散量的过程 。

对声音的采样常用麦克风等设备将声音信号转换成电信号，再用模/数转换器将电信号转换成一串用1和0表示的二进制数字（数字信号）。

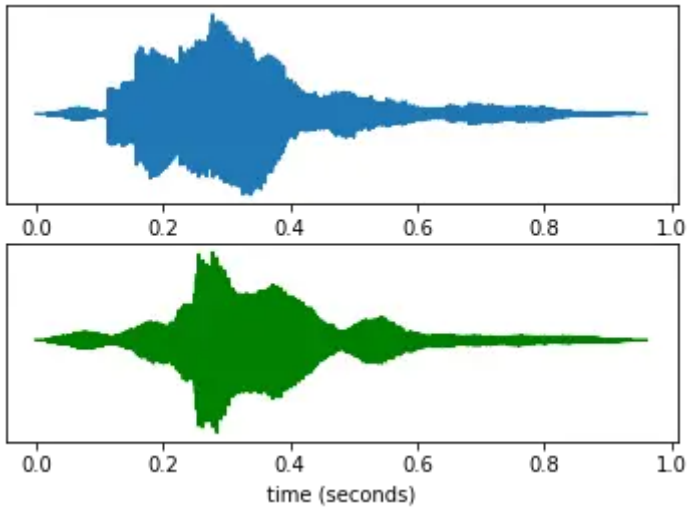
我们每秒对声音采样上万次，获得上万个按照时间顺序排列的二进制数字。于是，我们就将连续变化不断的声音转化成了计算机可储存并识别的二进制数字。

如win10的关机音效：

该声音由84700个不同的数字组成。 其中的一段数字如下：（二进制数字已转换为十进制）

... 413, 263, 137, 15, -124, -253, -369, -463, -511, -545, -587, -632, -678, -701, -687, -659, -623, -579, -539, -473, -380, -282, -162, -35, 78, 211, 341, 430, 499, 548, 551, ...

如果用图像的形式表示该音频，则图像如下：（横轴是时间，纵轴为振幅，两个图像分别代表左右声道。由于声音频率较大，所以在图像中的信号不是“正弦”，而是实心的。）



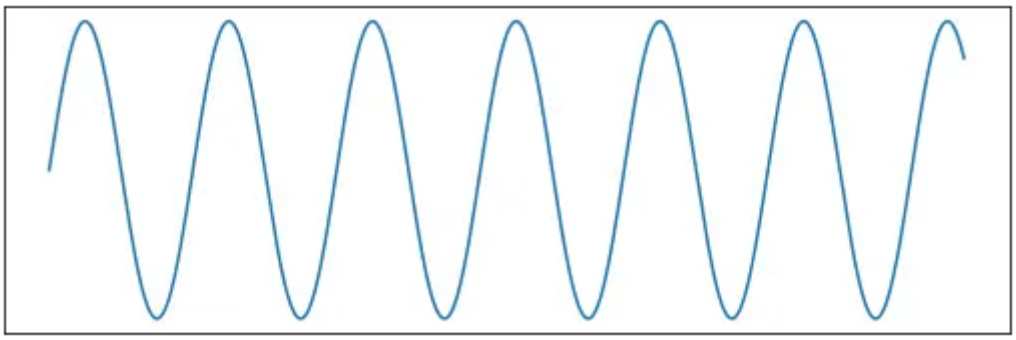
2.1 采样频率

采样频率指录音设备在一秒钟内对声音信号的采样次数。采样频率越高，声音的还原就越真实越自然。

目前主流的采样频率有22.05KHz、44.1KHz、48KHz三种。

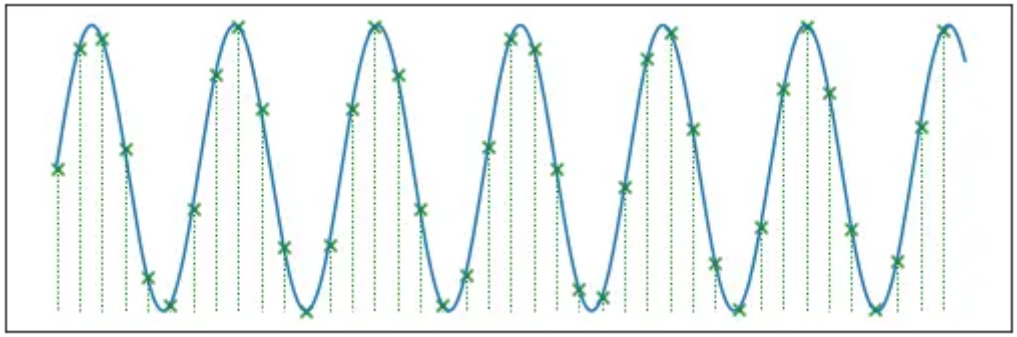
22.05 KHz为FM广播的声音品质，44.1KHz为理论上的CD声音品质。48KHz为人耳可辨别的最高采样频率。

直观理解：一段连续的声音如下

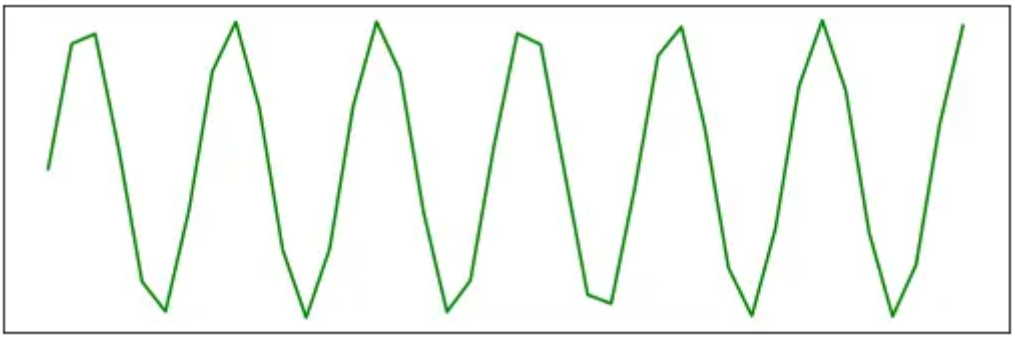


一段声音

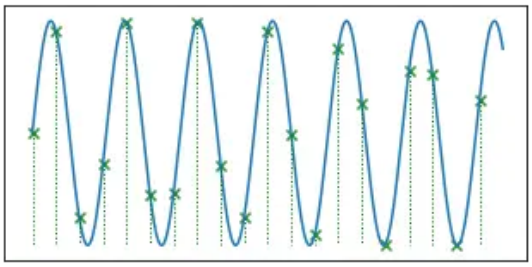
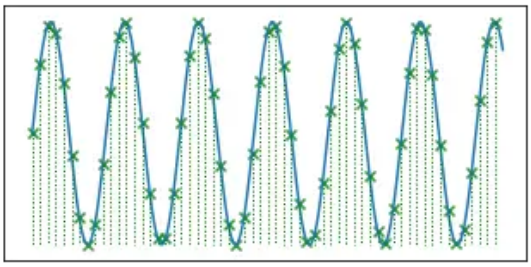
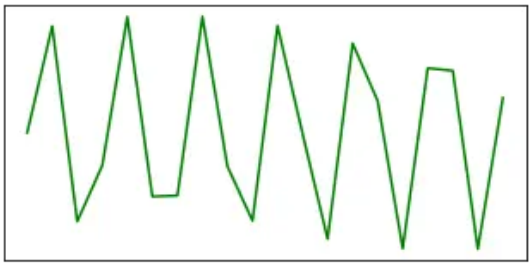
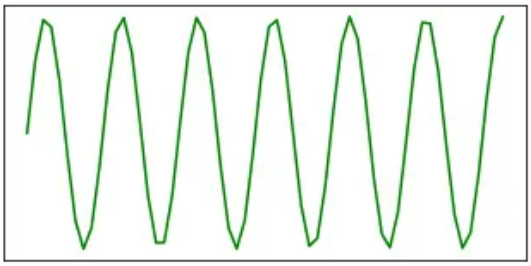
我们等间隔地对其采样



最终，我们真正采样到的音频如下



如下图可见，采样频率越高，我们获得的声音品质越好。

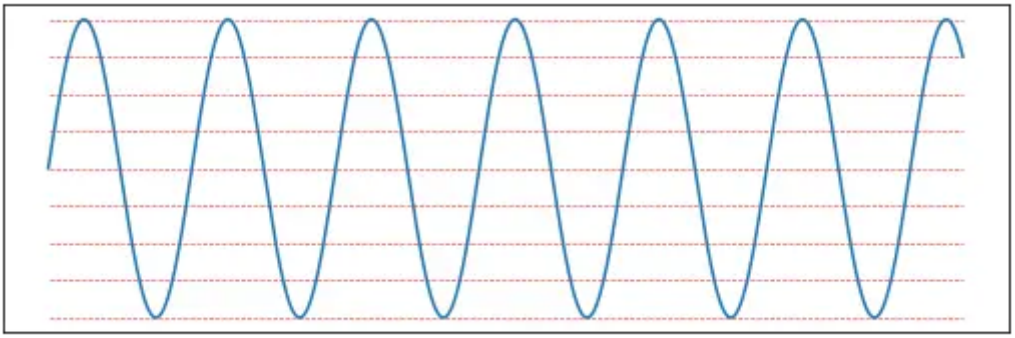
2.2 量化位数

我们不可能获得所有时间下声音的强度，因此声音是等时间间隔、离散采样的。同样，采样获得的数据不可能无限的精确，如数字为63.2222222....，这无法在计算机中储存。因此，采样获得的数据同样也是离散的。

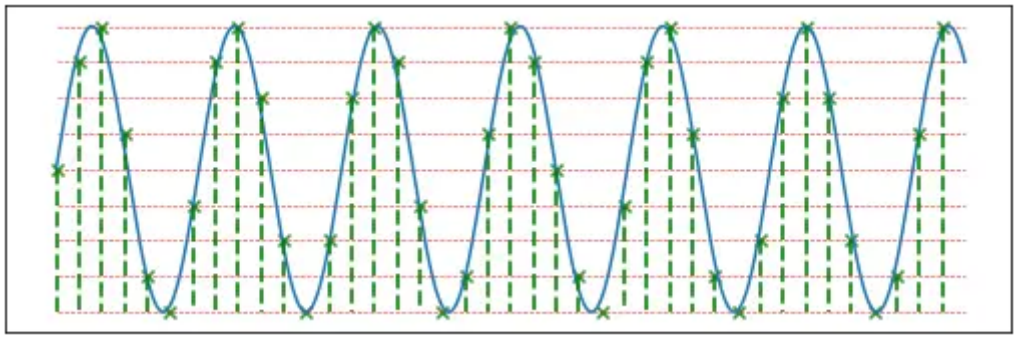
量化位数是音频文件的另一个参数。量化位数越大，声音的质量越高。常用的量化位数有8位、16位和32位。

量化位数指用几位二进制数来存储采样获得的数据。量化位数为8即指用8位二进制数来存储数据，如00010111

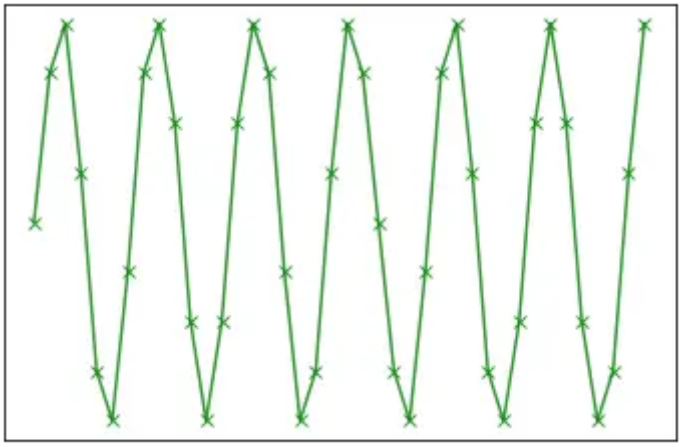
还是之前的例子，有一段正弦声波，假设量化位数为3，即存储的数据只有000/001/010/011/100/101/110/111这8种可能。



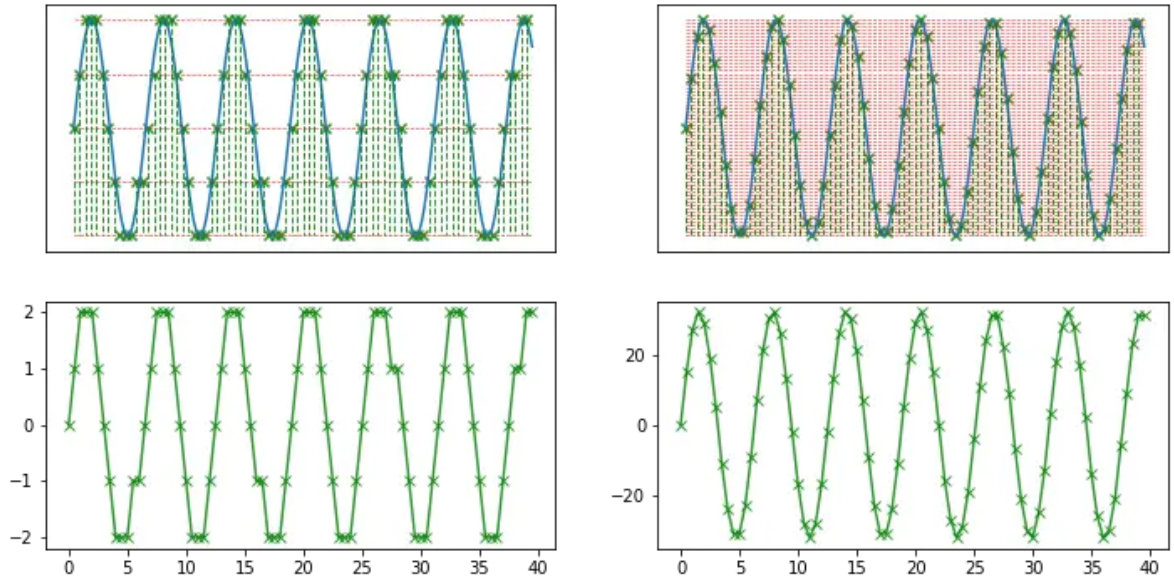
现在，还是等距离采样，不过采样的点只能落在最近的红线上。



此时，每个点纵坐标的取值只有二的三次方，即只有8中可能。



由下图可见，量化位数越大，声音效果越好。



另外值得注意的是，不同量化位数存储的数据不可直接比较。

如4位量化位数存储的1111，其十进制是15，8位量化位数存储的10000000，其十进制是64。不是因为64>15，所以后者对应的声音比前者大。而是应该二者分别除以其总取值范围后在比较。

$$\frac{15}{2^4} > \frac{64}{2^8}$$

前者对应的声音比后者大。

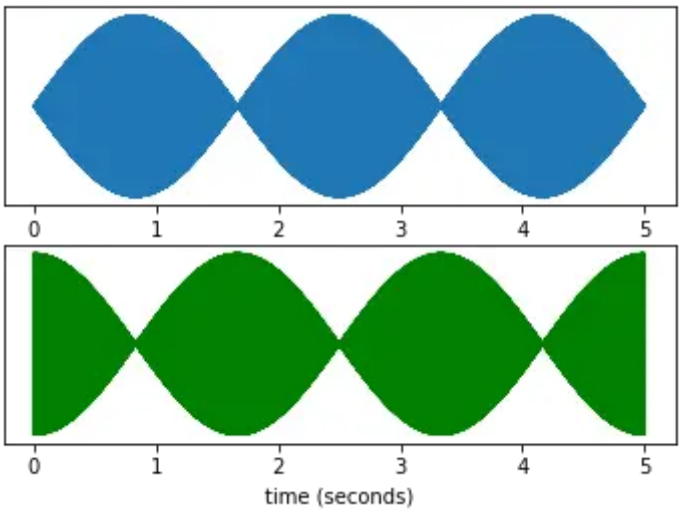
2.3 声道数

声道分为单声道与双声道。

单声道即为左右耳听到的声音相同。

双声道两耳听到的信息不同。相同的声音时间、采样频率和比特率的情况下，双声道文件的存储空间是单声道的两倍。但其会给人空间感，游戏和电影中常采用双声道，可达到“听声辨位”的效果。

示例声音如下：



3.音频格式

常见的音频格式有WAV,MP3,ACC等

3.1 WAV音频格式

WAV是微软开发的音频格式，支持音频压缩 ，但其常用来存放未经压缩的无损音频。由于未压缩，文件尺寸往往比较大，多用于存储简短的声音片段。

3.2 MP3音频格式

MP3是一种音频文件的有损压缩技术，用来大幅度地降低音频数据量。其可在没有明显声音品质受损的情况下，将音频文件压缩成其原文件的十分之一甚至是十二分之一。

3.3 AAC音频格式

相对于MP3，AAC格式的音质更佳，文件更小。可压缩至原文件的十八分之一。