

# 音视频学习从零到整(1)

## 音视频学习笔记出自:CC老师

- 音视频学习,越到实用环节,你会发现一些细节部分知识的缺失,可能会导致整个案例的理解出现很严重的错误.
- 希望大家学习过程,对技术保持一种匠人精神.沉淀下来,学习.

## 一.音频基础复习

### 1.1 声音的产生

相对于视频,可观察这个现象.音频在学习过程,就缺乏了想象的空间.但是如果从原理出发,就不会那么难了.

#### 声音是什么?

声音是波,靠物体的振动产生

### 1.2 声波的3要素

声波的三要素,是频率,振幅,波形.频率代表音阶的高低,振幅代表响度,波形则代表音色.

- 频率越高,波长就会越短.而低频声响的波长则较长.所以这样的声音更容易绕过障碍物,能量衰减就越小.声音就会传播的越远.
- 响度,就是能量大小的反馈.用不同的力度敲打桌面,声音的大小势必发生变换.在生活中,我们用分贝描述声音的响度.

#### 小贴士

- **分贝(decibel)**,是度量声音的强度单位,常用dB表示.是由美国发明家亚历山大.格雷厄姆.贝尔 名字命名的.

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
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

长期在夜晚接受50 分贝的噪音， 容易导致心血管疾病； 55 分贝， 会对儿童学习产生负面影响； 60分贝， 让人从睡梦中惊醒； 70 分贝， 心肌梗死的发病率增加30%左右； 超过110 分贝， 可能导致永久性听力损伤。

- 音色,在同样的频率和响度下,不同的物体发出的声音不一样.比如钢琴和古筝声音就完全不同.波形的形状决定了声音的音色.因为不同的介质所产生的波形不同.就会产生不一样的音色.

### 1.3 声音传播

声音的发生,来源于振动.人类说话,从声带振动发生声音之后,经过口腔,颅腔等局部区域的反射,在经过空气传播到别人耳朵中.这是我们说话到听到的过程.

声音的传播,可以通过空气,液体,固定传播.介质不同,会影响声音的传播速度.

- 吸音棉:通过声音反射而产生的嘈杂感,吸音材料选择使用可以衰减入射音源的反射能量,从而对原有声音的保真效果.比如录音棚墙壁上就会使用吸音材质
- 隔音:主要解决声音穿透而降低主体空间的吵闹感,隔音棉材质可以衰减入射声音的透射能量.从而达到主体空间安静状态,比如KTV墙壁上就会安装隔音棉材料.

## 二.数字音频

## 2.1 模拟信号数字化过程

将模拟信号转换为数字信号的过程,分别是采样,量化和编码.

### 音频采样

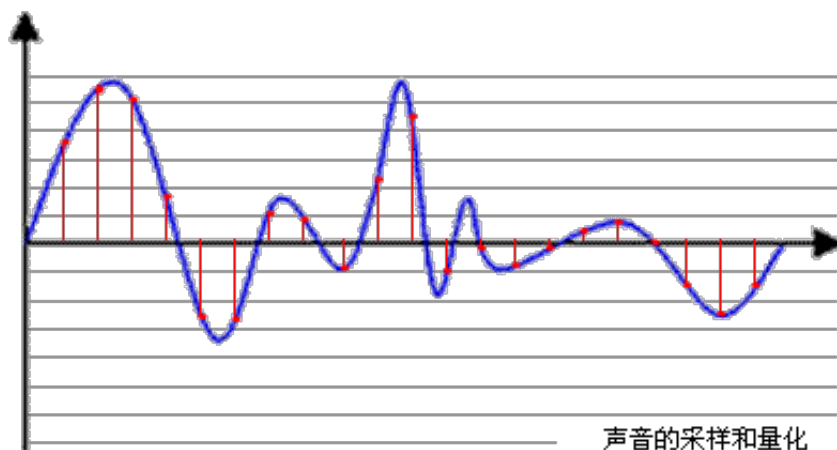
对模型信号进行采样,采样可以理解为在时间轴上对信号进行数字化.而,根据奈斯特定理(采样定理),按比声音最高频率高2倍以上的频率对声音进行采样.这个过程称为AD转换.

比如,前面提到高质量音频信号,其频率范围是20Hz-20KHz.所以采样频率一般是44.1KHz.这样可以保证采样声音达到20KHz也能被数字化.而且经过数字化处理后的声音,音质也不会降低.44.1KHz,指的是1秒会采样44100次

[奈斯特定理\(采样定理\) 资料](#)

### 量化

**量化,指的是在幅度轴上对信号进行数字化.**简单的说,就是声音波形的数据是多少位的二进制数据,通常用bit做单位.比如16比特的二进制信号来表示声音的一个采样.它的取值范围 $[-32768, 32767]$ .一共有65536个值.如16bit、24bit。16bit量化级记录声音的数据是用16位的二进制数,因此,量化级也是数字声音质量的重要指标。我们形容数字声音的质量,通常就描述为24bit (量化级)、48KHz采样,比如标准CD音乐的质量就是16bit、44.1KHz采样.



既然每个量化都是一个采样,那么声音这么多采样,该如何将这些数据存储起来?

### 编码

## 什么叫编码?

按照一定格式记录采样和量化后的数据.

音频编码的格式有很多种,而通常所说的音频裸数据指的是脉冲编码调制(PCM)数据.

如果想要描述一份PCM数据,需要从如下几个方向出发:

- 量化格式(sampleFormat)
- 采样率(sampleRate)
- 声道数(channel)

### 举例:

以CD音质为例,量化格式为16bite,采样率为44100,声道数为2.这些信息描述CD音质.那么可以CD音质数据,比特率是多少?

$$44100 * 16 * 2 = 1378.125\text{kbps}$$

那么一分钟的,这类CD音质数据需要占用多少存储空间?

$$1378.125 * 60 / 8 / 1024 = 10.09\text{MB}$$

如果sampleFormat更加精确或者sampleRate更加密集,那么所占的存储空间就会越大,同时能够描述的声音细节就会更加精确.

存储在这些二进制数据即可理解为将模型信号转化为数字信号.那么转为数字信号之后,就可以对这些数据进行存储\播放\复制获取其他任何操作.

## 小知识

### 麦克风是如何采集声音?

麦克风里面有一层碳膜,非常敏感.声音用波的形式,压缩空气也会压缩这层碳膜.碳膜在手动挤压的情况下,也会发生振动.在碳膜下方会有一个电极,碳膜在振动时就会触发电极.接触时间的长短和频率与声波的振动幅度和频率有关.这样就完成了声音信号到电信号的转换.在经过放大电路处理,就可以来到我们所说的量化处理了.