

数字媒体技术基础 ——数字音频

王晗

wanghan@bjfu.edu.cn

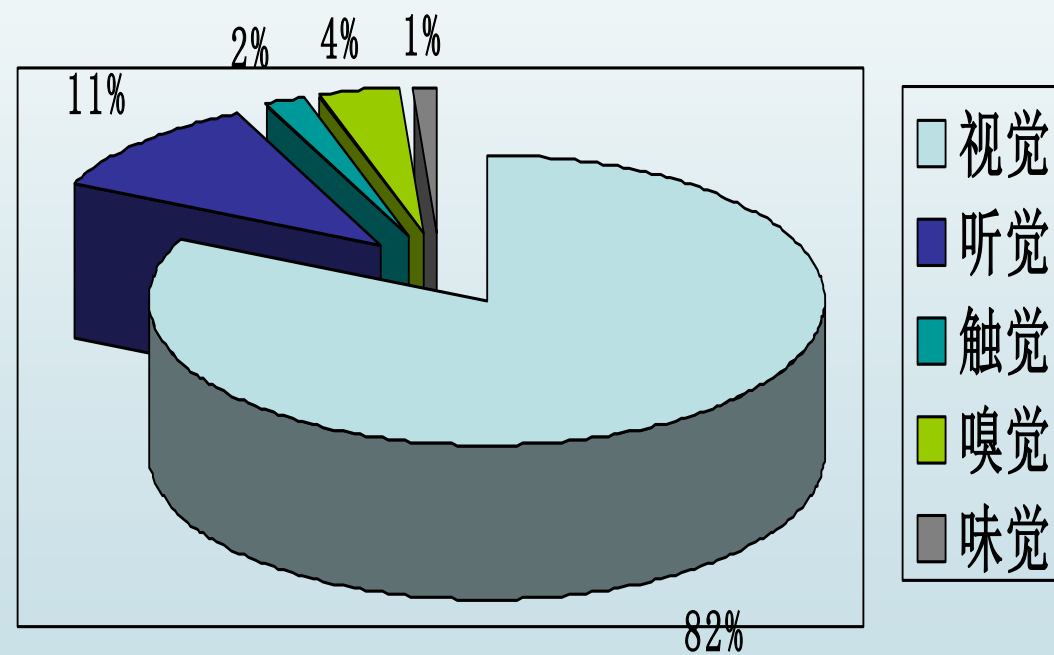


专题提纲

- 音频的概念及特性
- 音频处理设备
- 音频的数字化
- 数字语音处理技术
- 数字音频编辑工具

音频的概念及特性

声音的重要性

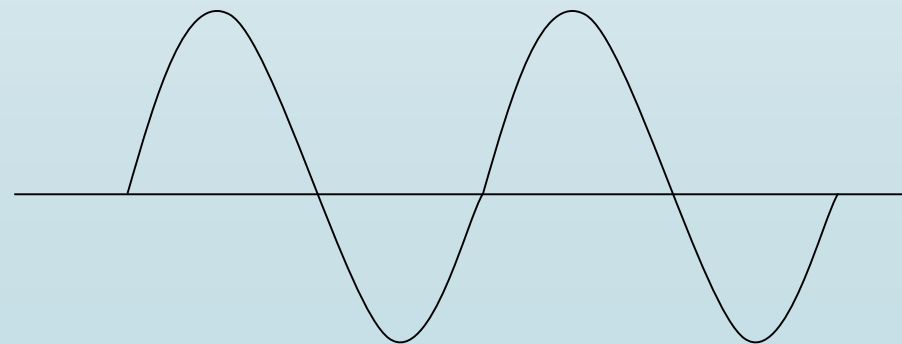


Treicher信息获取实验

音频的概念及特性

■ 声音的产生

- 声音是通过一定介质（如空气、水等）传播的一种连续振动的波。
- 声音必须通过空气或其他的媒质进行传播，形成声波，因此声音具有波形的基本属性，可以用波形图来表示。



音频的概念及特性

- 声音的特性

频率，振幅，周期——物理学上描述

音调，响度，音色——生理学上描述

- 傅里叶定律：在有限频谱内，无论多么复杂的波形，都是由非常多的不同频率振幅的正弦波叠加而成。

音频的概念及特性

■ 声波的频率

声波的频率就是声源振动的频率，即每秒钟内来回往复振动的次数。

频率的单位通常用Hz（赫兹）来表示，简称赫。

一般来说，物体振动越快，频率就越高，人感受到的音调也就越高，反之亦然。

声波的频率对人耳的听觉感受影响很明显。按照声波的频率不同，声音可以分为次声波、超声波、人耳可听声三种。

音频的概念及特性

■ 声波的频率

- ◆ **次声波**：指频率低于20Hz的声波。一般在地震、火山爆发、风暴、枪炮发射等会产生次声波，它对人体能够造成危害。
- ◆ **超声波**：指频率高于20kHz（每秒振动20000次以上）的声波。超声波频率高、方向性好、穿透力强，可以广泛用于测距、测速、清洗、焊接、碎石、医学诊断等领域。
- ◆ **人耳可听声**：频率介于20Hz-20kHz这段范围的声波是人耳可听声。
 - 低频：500Hz以下
 - 中频：500-2000Hz
 - 高频：2000Hz以上

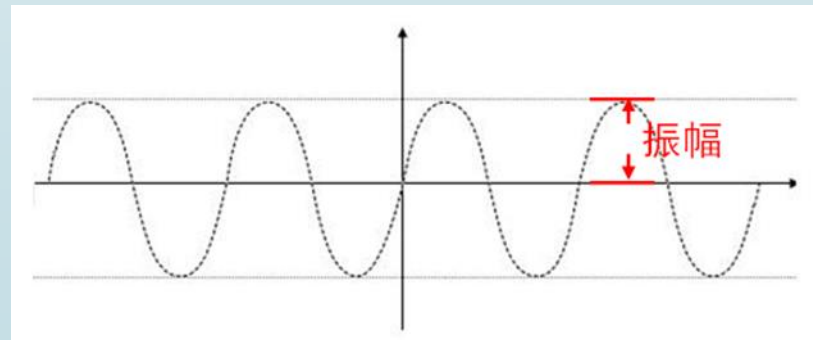
产品参数

品牌		型号	
电压	110-220伏	颜色	冰川绿
刷头	1支ProResult标准牙刷头	其他配件	1个充电器
电池	充电式设计	震频	31000次/分钟

音频的概念及特性

■ 声波的振幅

- 指的是发声物体在振动时偏离中心位置的幅度，代表发声物体振动时的动能、势能的大小。
- 由物体振动时产生声音的能量（声能）或声波压力（声压）大小所决定的。
- 声能或声压越大，引起人耳主观感觉到的响度也越大。



音频的概念及特性

分贝(decibel): 振幅类物理量的度量单位。在声学领域中，分贝的定义是声源功率与基准声功率比值的对数乘以10的数值。

$$\text{分贝值(dB)} = 10 \lg \text{声源功率} / \text{基准声功率}$$

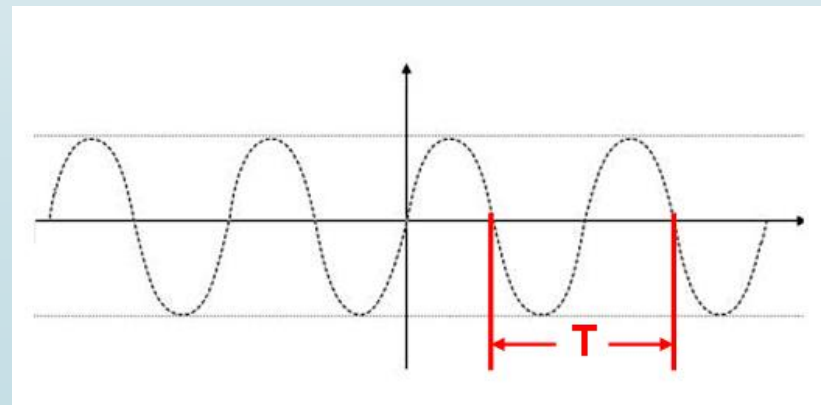
零分贝的设定，是根据听力正常的年轻人所能听到的最小声音所得到的。每增加10分贝等于强度增加10倍，增加20分贝增加100倍，30分贝则增加1000倍。

相对于0分贝的，一般的耳语大约是20分贝，极安静的住宅区40分贝，一般公共场所50分贝，交谈约60分贝，交通繁忙地区85分贝，飞机场跑道120分贝。

音频的概念及特性

■ 声波的周期

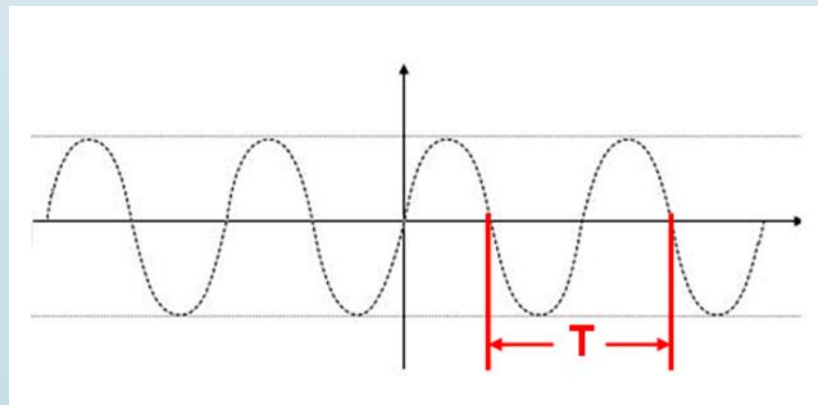
一个声波完成一次振动所需要的时间称为周期，用符号 T 表示，单位通常为秒（s）。周期与频率是互为倒数关系。



音频的概念及特性

■ 声波的波长

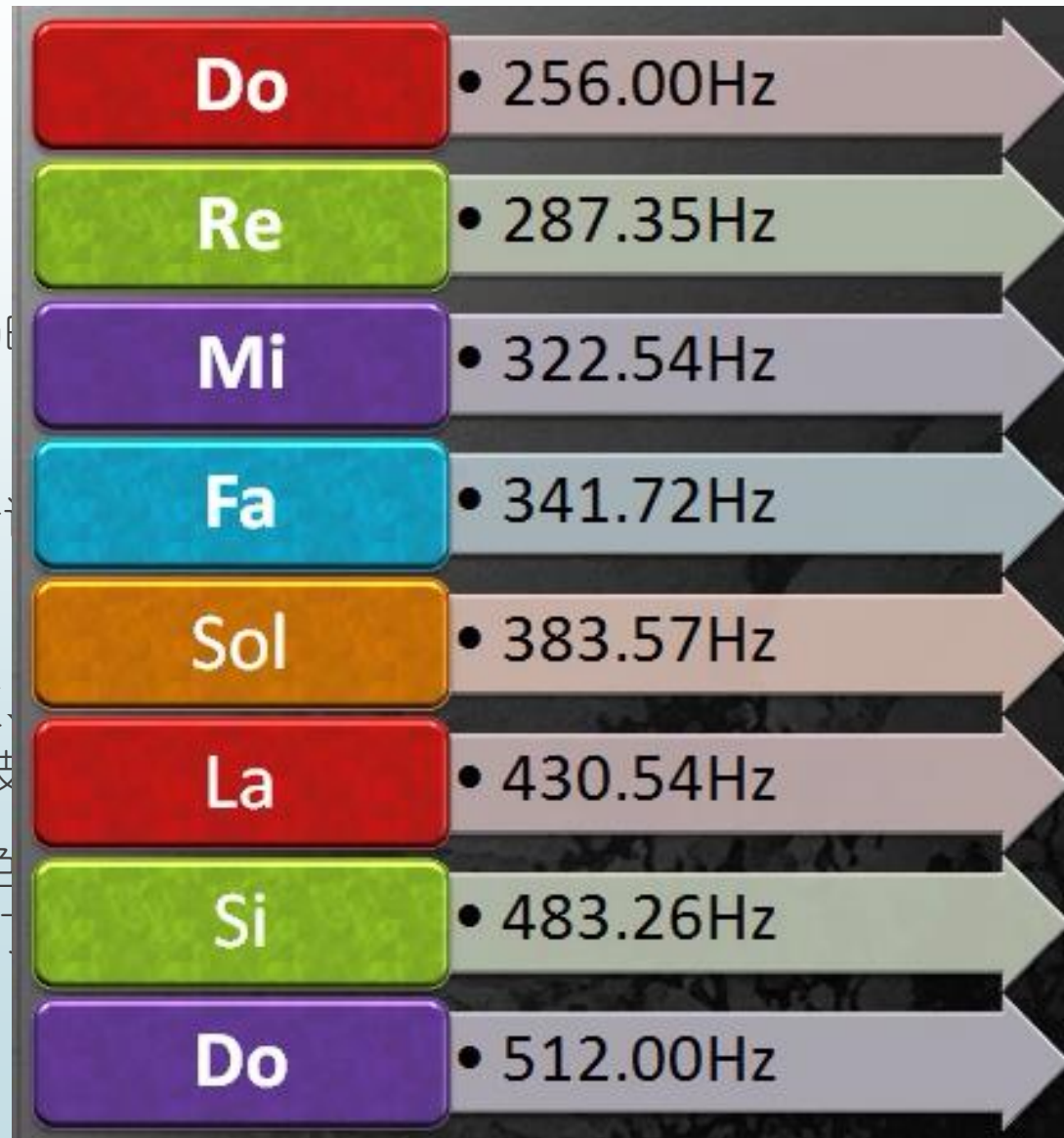
- 声波的波长是指声波在一个周期的时间内传播的距离。
- 语音中的低频部分，由于其波长较长，能向各个方向均匀地传播
- 语音中的高频部分仅向前直射。
- 声波遇到障碍物时存在衍射现象。



音频的概念及特性

声音的心理特性

- (1) 音量（也称响度）：人耳对声音强弱的感觉
——与声波的振幅有关
- (2) 音调：人耳对声音高低的感受称为音调
——与声波的频率有关
- (3) 音色：音色是人们区别具有同样响度、同样音调的声音或说是人耳对各种频率、各种强度的声波的感受，或称为音品、音质。
——与声波的频谱结构有关。音色与发声体的材料、结构、发声方式、发声环境等有关。音色的不同使我们能区分不同的声音，如我们能根据木琴的声音分辨出木琴声，根据小号的声音分辨出小号声。



音频处理设备

- 声音记录设备
 - 机械留声机
 - 钢丝录音机
 - 磁带录音机
 - 激光唱机
 - MP3随身听
 -
- 音频制作设备
 - 话筒
 - 音箱
 - 模拟调音台
 -



音频处理设备

► 音频处理设备

声卡

耳机和音箱

麦克风

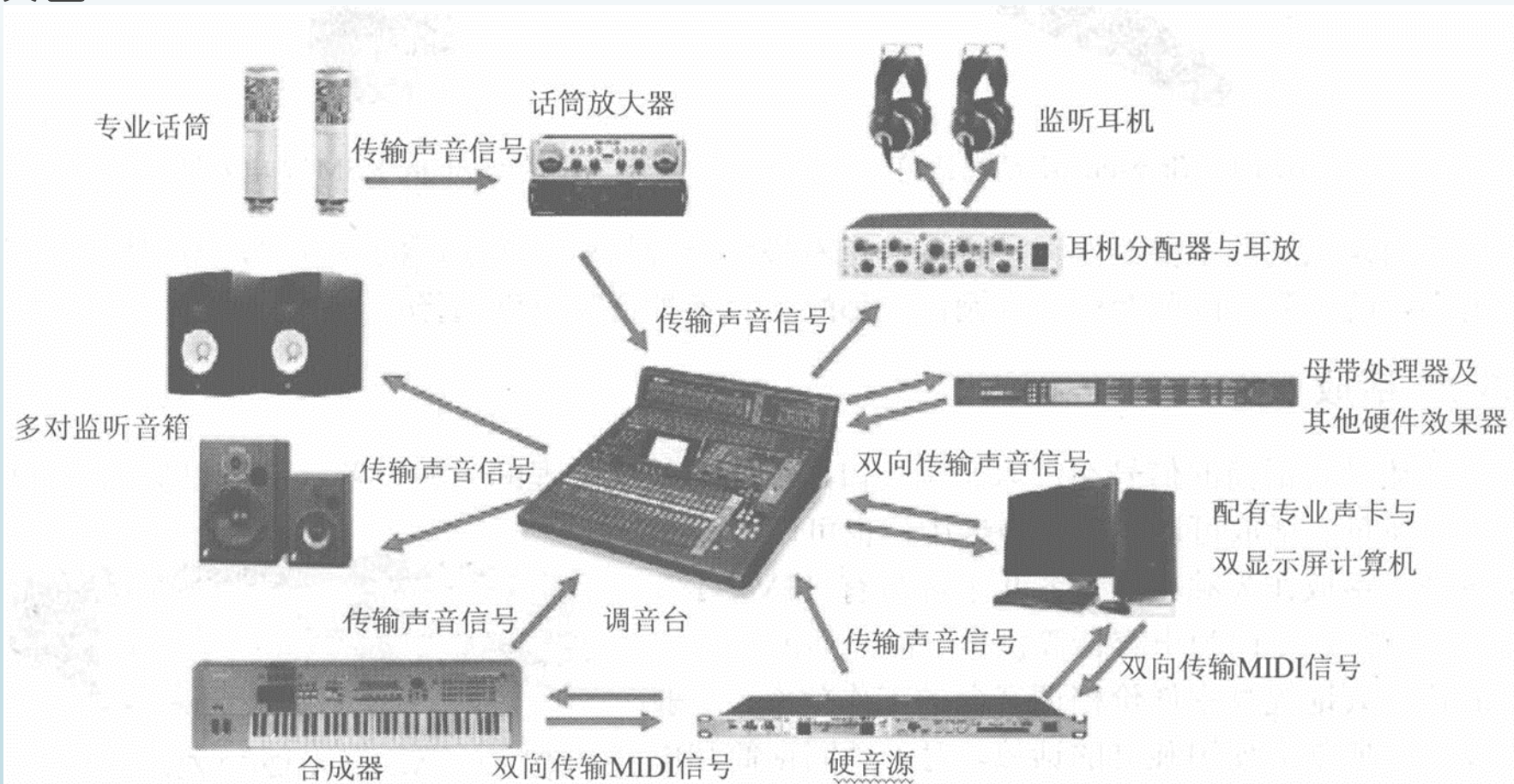
MIDI键盘

调音台

录音室

硬音源

信号处理设备



声卡

- 声卡，也叫音频卡，是多媒体电脑中用来处理声音的接口卡。
- 作用：把来自话筒、收录音机、激光唱机等设备的语音、音乐等声音变成数字信号交给电脑处理，并以文件形式存盘，还可以把数字信号还原成为真实的声音输出。

- 功能：
 - 音乐合成发音功能
 - 混音器功能和数字声音效果处理器功能
 - 模拟声音信号的输入和输出功能



耳机和音箱

- 耳机和音箱都可以称为扬声器，是一种电声换能器件，能够将音频信号变换为声音。



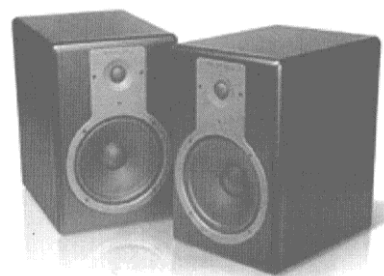
耳机和音箱

- 音箱就是放置扬声器的箱子，即音箱主机箱体或低音炮箱体内自带功率放大器，对音频信号进行放大处理后由音箱本身回放出声音，能增强音响效果。两只音箱可以拆开安放在最佳位置。

音箱

有源音箱：又称为“主动式音箱”，其扬声器箱体内已经内置有功率放大器，用户不必考虑它与放大器匹配的问题。

无源音箱：又称为“被动式音箱”，需要连接功放才能使用。



耳机和音箱

耳机从结构上分：

- **开放式耳机**：外壳是开放的，耳机质量轻，声音自然，无压迫感。外界声音不会完全隔绝，佩戴舒适，但声音有外泄。
- **半开放式耳机**：耳机的开放与频率或方向相关，即只对某些频率开放，或是在一定方向上是开放的。
- **封闭式耳机**：耳机的外壳是封闭的，通过软音垫包裹，外界声音隔绝，声音几乎无外泄，长时间佩戴有不适感，在专业监听中使用较多。

全封闭

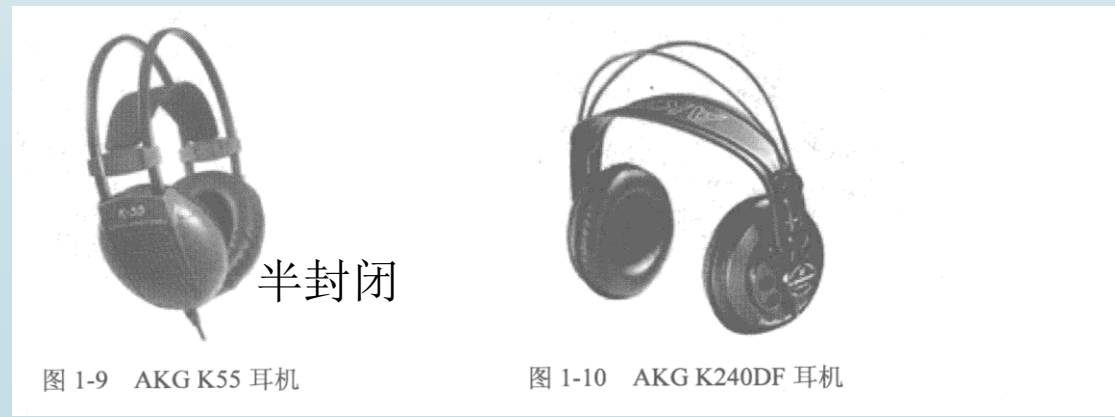


图 1-9 AKG K55 耳机

图 1-10 AKG K240DF 耳机

麦克风

- 麦克风，学名为传声器，由**Microphone**翻译而来，它是将声音信号转换为电信号的能量转换器件。
- 按工作原理分：电动式话筒和电容式话筒。
- 电动式话筒：
 - 原理：利用磁场中导体运动产生电动势而工作
 - 优点：中频特性好，抗震能力强.多用于卡拉OK和移动音频工作站。
 - 缺点：灵敏度较低、频响不够宽和容易产生磁感应噪声。
- 电容式话筒：
 - 原理：应用静电感应原理，靠声波振动来使电容一端移动，造成电容两极间距离变化，从而产生电压差，形成声音的电信号。
 - 优点：灵敏度高，频响范围宽，音质好。
 - 缺点：对环境要求高，受潮或长时间受热会不稳定。



图 1-6 Samson Q2 动圈话筒

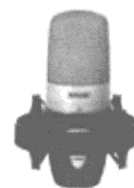


图 1-7 Shure KSM27 电容话筒

MIDI键盘

- MIDI (Music Instrument Digital Interface) 键盘外观上与电子琴很相似，本身不能发声，一般与电脑相连接使用。

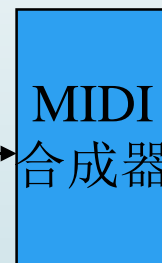
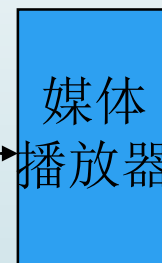


MIDI音乐的制作与播放

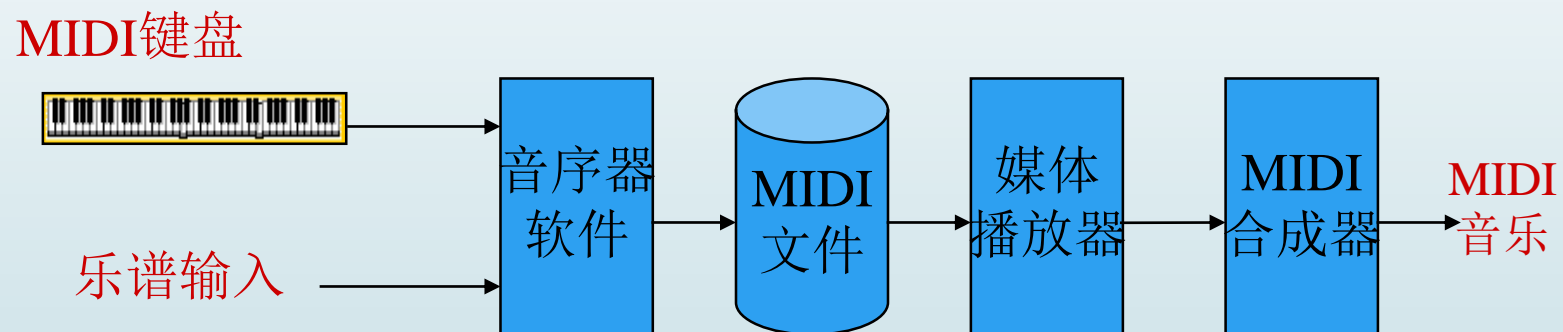
MIDI键盘



乐谱输入



MIDI
音乐



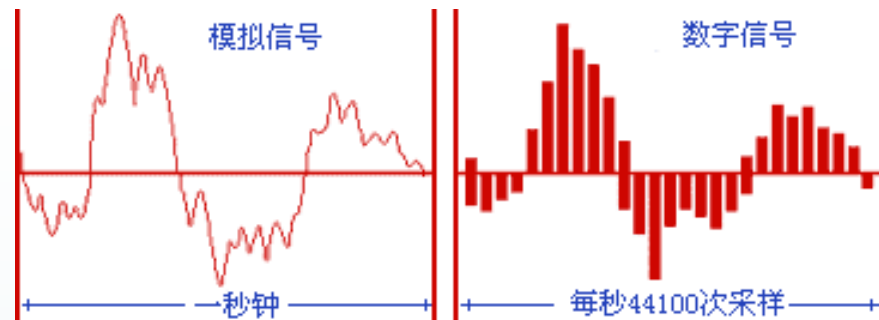
调音台

调音台的功能

- 调音台又称调音控制台，它将多路输入信号进行放大、混合、分配、音质修饰和音响效果加工，具有电平控制、音色调整、音质均衡与补偿的功能。
- 它负责计算机与话筒、音箱、音源和合成器等音频设备之间声音信号的传递，拥有各种辅助输出、监听输出、左右立体声输出和混合输出等输出方式，因此它在音频系统中占有核心地位，是音频制作与编辑、信号处理的中心。



音频的数字化



(1) 模拟音频：

- 通过各种拾音器，如话筒，将物体振动所产生的声音录制下来，存储在磁带、磁盘等电磁存储介质上，这种声音是以电信号存在的，称为模拟音频。
- 模拟音频技术中以模拟电压的幅度表示声音强弱，模拟声音在时间上是连续的；

(2) 数字音频：

- 通过计算机设备及相关音频软件将模拟音频转换为计算机可以识别的以二进制存在的声音信号，就形成了数字音频。
- 而数字音频是一个数据序列，在时间上是断续的。

音频的数字化

(3) 音频数字化

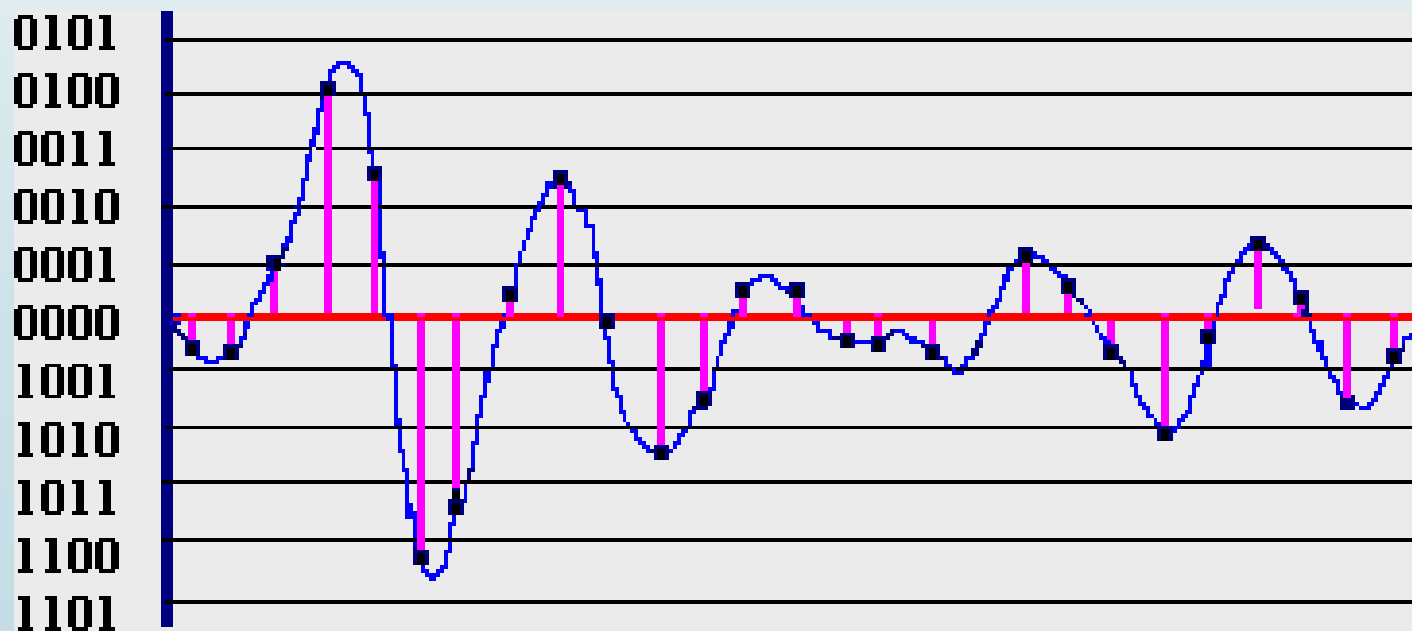
- 数字音频是通过采样和量化，把模拟量表示的音频信号转换成由许多二进制数1和0组成的数字音频信号。
- 音频的数字化分为三个步骤：
 - 采样：在时间轴上对信号数字化；
 - 量化：在幅度轴上对信号数字化；
 - 编码：按一定格式记录采样和量化后的数字数据。



音频的数字化

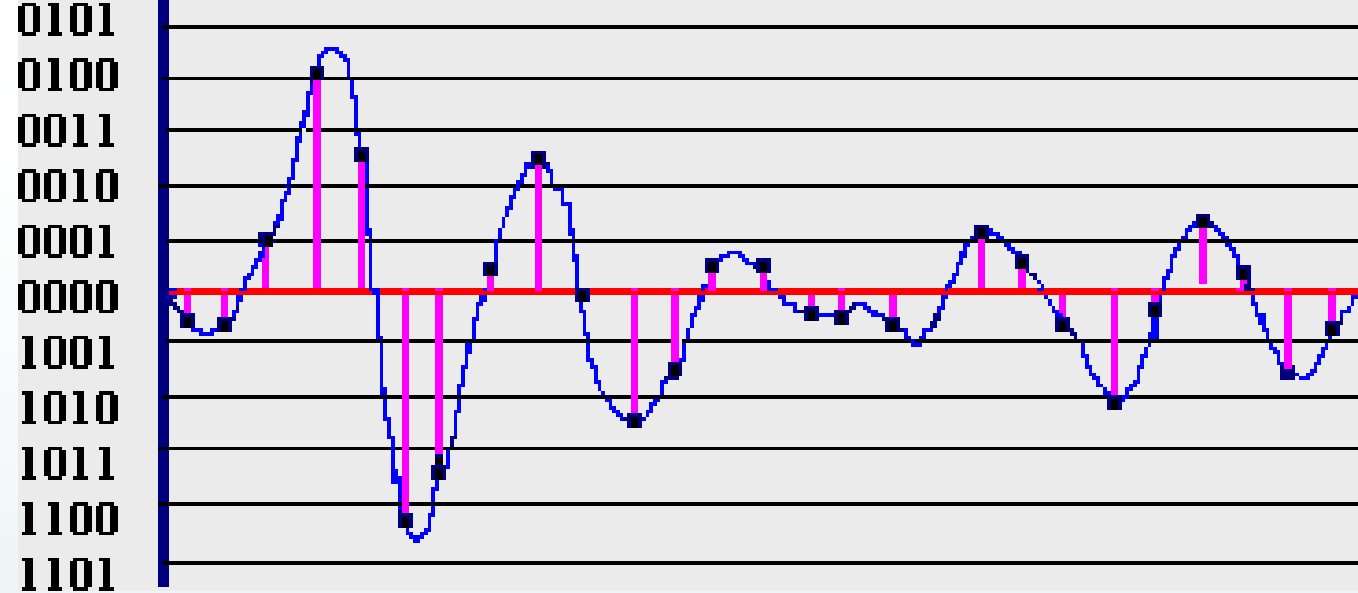
(3) 音频数字化

- 首先我们考虑声音经过麦克风，转换成一连串电压变化的信号，如下图所示。这张图的横座标为秒，纵座标为电压大小。对连续信号按一定的时间间隔在模拟声波上截取一个振幅值。



音频的数字化

(3) 音频数字化



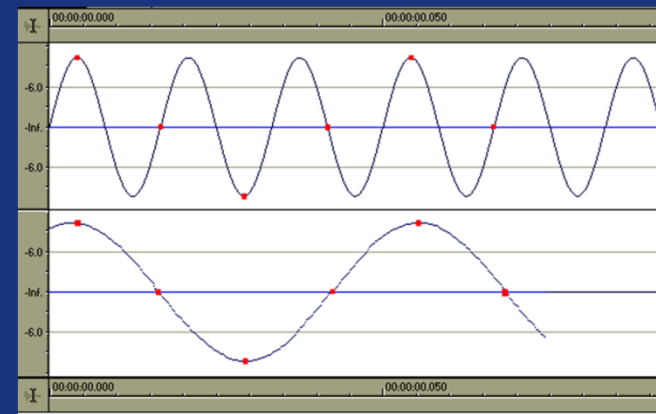
- 我们把分割线与信号图形交叉处的座标位置记录下来，可以得到如下资料，
(0.01,11.65) (0.02,14.00) 、 (0.03,16.00) 、 (0.04,17.74)
(0.18,15.94) 、 (0.19 ,17.7) (0.20,20) 。我们现在已经把这个波形以数字记录下来了。由于我们 已经知道时间间隔是固定的 0.01 秒，因此我们只要把纵座标记录下来就可以了，得到的结是 11.65 14.00 16.00 这一数列。这一串数字就是 将以上信号数字化的结果。

音频的数字化

(3) 音频数字化

- **采样**：每间隔一个时间段采集一次声音信号的幅度样本的过程。
- **采样频率**：指每秒钟所采集的声波幅度样本的次数。（单位是Hz）
- 采样频率越高，声音的保真度就越高，但用于存储音频的数据量也越大。
- **采样定理（奈奎斯特定理）**：
在进行模拟/数字信号的转换过程中，当采样频率 f_s 大于信号中最高频率 f_{max} 的2倍时，即： $f_s \geq 2f_{max}$ ，则采样之后的数字信号完整地保留了原始信号中的信息。

采样频率与声音质量



如果采样频率不够大，数字化的音频波形就与原始声波的形状相差很远，音质就会与原来不同。

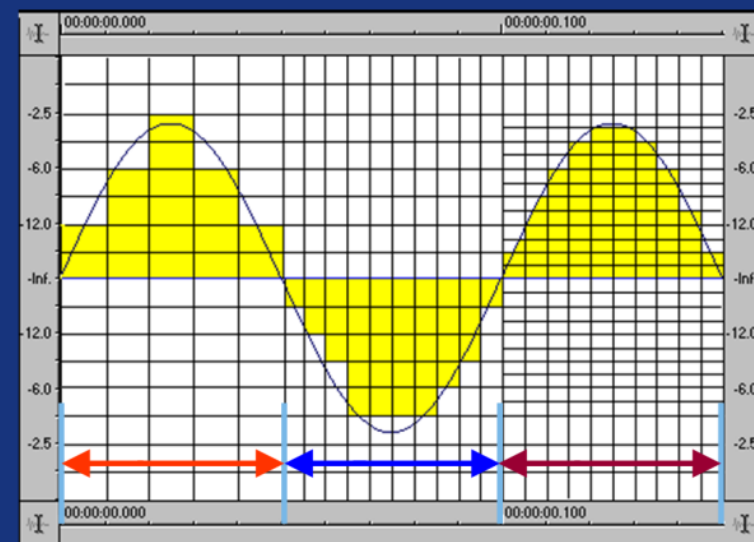
音频的数字化

常用的采样率：

- 11,025 Hz AM 广播质量 可用于长的音乐,高质量的语音,音效
- 22,050 Hz FM 广播质量 可用于短的高质量音乐
- 32,000 Hz 接近CD 质量 可用于专业或业余数字产品
- 44,100 Hz CD 质量 可用于高保真声音和音乐
- 48,000 Hz 数字音频磁带音质(DAT 质量)
- 96,000 Hz DVD 质量
- 192000Hz DVD 质量

音频的数字化

- **量化**：就是把采样得到的每一个样本值从模拟量转换为二进制的数字值的过程。
- 通常，把对声波波形幅度的数字化表示称为量化（Quantization）。
- **量化位数**：又称作量化精度或采样位数，简单地说就是描述声音波形的数据是多少位的二进制数据。
- 量化位数也是衡量数字声音质量的重要指标，在相同的采样频率下，量化位数越高，声音的质量越好。
- 量化方法:均匀量化与非均匀量化
- 常用的量化位数：
 - 24 bit DVD 质量
 - 16 bit CD 质量 可用于高保真声音和音乐
 - 12 bit 接近CD 质量 可用于专业或业余数字产品
 - 8 bit FM 广播质量 可用于短的高质量音乐
 - 4 bit AM 广播质量 可用于长的音乐,高质量的语音,音效



横坐标的疏密代表采样频率的大小

纵坐标的疏密代表量化位数的大小

音频数字化的两个重要指标

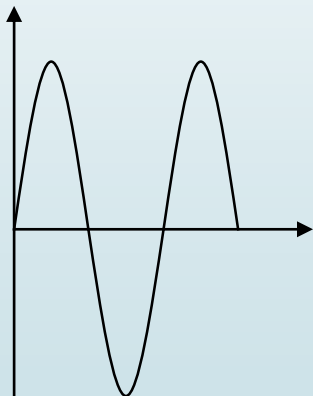
- **采样频率**：即采样点之间的时间间隔。
- **量化深度**（或称为量化分辨率）：单位电压值的可分等级数；
- 两者与音质还原的关系是：采样频率越高，量化深度越大，声音质量越好。
- 当频率越小（时间间隔越短），量化深度（量化分辨率）越大，二者的轮廓越吻合，这也说明数字化的信号能更好的保持模拟音频信号的形状，有利于保持原始声音的真实情况。

音频的数字化

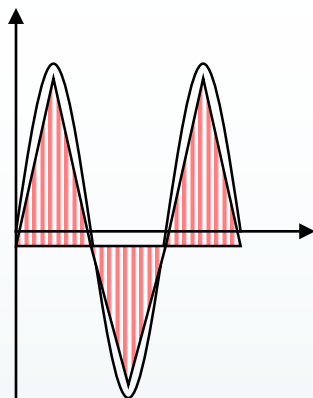
- **编码**：将经过采样和量化后的数字化声音信号（脉冲数字信号）按一定的数据格式以二进制形式表示，这个过程称为编码。
- **波形编码**：基于音频数据的统计特性进行编码。经波形编码的音频质量好，但压缩比不大，且为有损压缩，主要适用于高保真度语音和音乐的压缩技术。
- **参数编码**：基于音频的声学特性进行的编码，其方法是将音频信息以某种模型表示，再抽出合适的模型参数和参数激励信息进行编码，声音重放时再根据这些参数重建还原声音即可。该编码方式压缩比高，但还原信息的质量较差，自然度低，计算量大，而且保真度不高。
- **感知编码**：基于人的听觉特性进行的编码。从人的听觉系统出发，设计心理声学模型，从而实现更高效的数字音频的压缩。它是有损压缩，但人耳基本上分辨不出声音的失真。

为了不失真取样
频率大于声音信
号的2倍

模拟
声音
信号



取 样

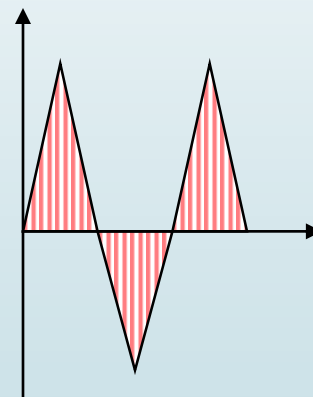


量 化

量化精度高，
声音保真度好

编 码

数字
声音

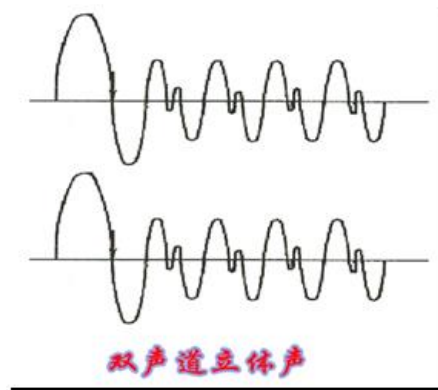
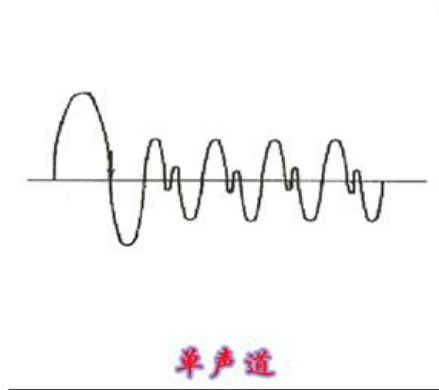


011010011101.....

- 进行数据压缩
- 减少存储量

音频的数字化

- 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的个数。
- 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，也称为立体声。
- 随着声道数的增加，所占用的存储容量也成倍增加。



音频的数字化

➤ 数字音频信息的质量与数据量

- 采样、量化和编码技术是声音数字化的关键技术。
- 采样频率、每个采样值的量化位数以及声音信息的声道数目，是影响数字化声音信息质量和存储量的三个重要因素。
- 采样频率越高、量化位数越大、声道数目越多，声音的质量就越高，但存储量就越大。

音频的数字化

► 数字音频信息的质量与数据量

声音信息的数据量

确定了数字声音的采样频率、量化位数和声道数就可以计算出声音信息的数据量，其计算公式为：

$$S=R \times r \times N \times D/8$$

其中：

- S表示文件的大小，单位是Byte；
- R表示采样频率，单位是Hz；
- r表示量化位数，单位是bit；
- N表示声道数；
- D表示录音时间，单位是s。

数字音频的存储量：

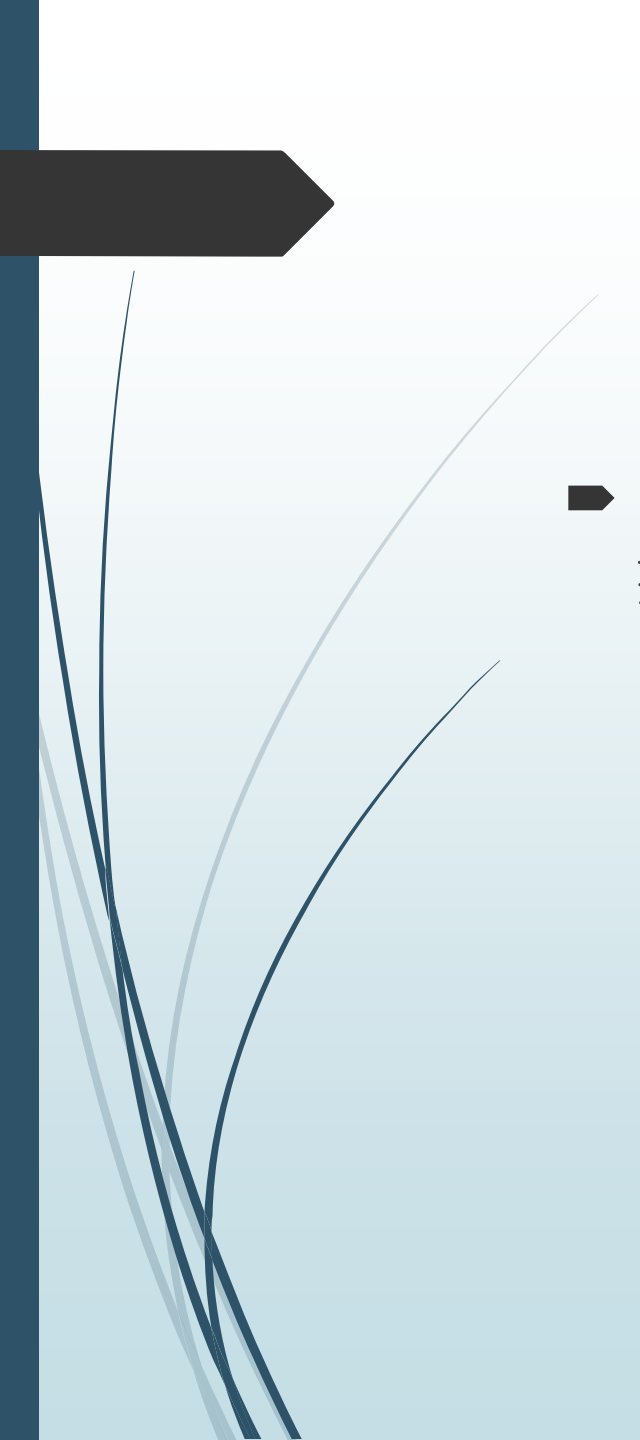
可用以下公式估算声音数字化后每秒所需的存储量（未经压缩的）

$$\text{存储量} = \text{采样频率} \times \text{量化位数} \div 8$$

若使用双声道，存储量再增加一倍

例如，数字激光唱盘（CD-DA）的标准采样频率为44.1 kHz，量化位数为16位，立体声。一分钟 CD-DA 音乐所需的存储量为

$$44.1 \text{ K} \times 16 \times 2 \times 60 \div 8 = 10584 \text{ KB}$$



► 如果要衡量一个数字音频的音质好坏的话，通常可以参考以下指标：

- **采样频率**：即采样点之间的时间间隔，采样间隔时间越短，音质越好。
- **量化深度**(量化分辨率)：是指单位电压值和电流值之间的可分等级数，可分等级越多，音质越好。
- **音频流码率**：数字化后，单位时间内音频数据的比特容量，流码率越大音质越好。
- 以上三个方面的指标中，前面两个指标是绝对指标，而音频流码率是一个相对指标，可以间接用来考察音频的质量。

音频的数字化

► 数字音频信息的质量与数据量

声音质量的评价

声音质量的评价是一个很困难的问题，也是一个值得研究的课题。

目前声音质量的度量有两种基本方法，一种是客观质量度量，另一种是主观质量的度量。

(1) 客观质量的度量

对声波的测量包括评价值的测量、声源的测量和音质的测量，其测量与分析工作，是使用带计算机处理系统的高级声学测量仪器来完成。

(2) 主观质量的度量

采用客观标准方法很难真正评定编码器的质量，在实际评价中，主观的质量度量比客观质量的度量更为恰当和合理。

声音信号的存储格式

在多媒体音频技术中，存储声音信息的文件格式有多种，如WAV、MIDI、MP3、RM及CD唱盘数字音频。

- WAV格式
- MIDI格式
- MP3格式
- RM格式

声音信号的存储格式

WAV格式是Windows使用的标准数字音频格式，也称为波形文件，文件的扩展名为.wav。

利用声音卡和相应的软件可以通过录音创建波形文件，需要时可以方便地将它播放出来，也可以通过适当的软件对文件中数字化的音频信号进行编辑处理。

WAV波形文件比较大，实际使用时常常要将它进行压缩处理。

声音信号的存储格式

MIDI文件的扩展名为“**.MID**”。**MIDI**是(乐器数字接口)的缩写。它是一个通信标准，规定计算机音乐程序、电子合成器和其他电子设备之间交换信息与控制信号的。

MIDI文件占用存储空间小，但文件的录制比较复杂。需使用专用设备录制，如键盘合成器等。

与**WAV**文件不同，**MIDI**文件记录的不是声音本身，而是将每个音符记录为一个数字，因此比较节省空间，可以满足长时间音乐的需要。



声音信号的存储格式

MP3格式采用**MPEG Layer 3**标准对**WAVE**音频文件进行压缩而成，以达到**CD**唱盘的音质。

RM格式采用音频/视频流和同步回放技术来实现在互联网上提供优质的多媒体信息。

随堂实验

- 使用Adobe Audition中对音频处理的基本方法，从网上下载5段不同的**包含**鸟鸣声的音频，通过读取鸟鸣声波形图，分析不同鸟鸣声的区别，以及如何分割鸟鸣与环境噪音
- 提交材料：1.完整音频.wav/mp3， 2.分割出的鸟鸣声.wav/mp3， 3.对于波形内容的分析
- 提交截止时间：4月7日；
- 提交方式：班级负责人搜集
- 命名要求：学号_姓名.rar
- 注：对于音频内容的分析没有具体要求，打破思维的墙，可以对自己得到的任意观测结果进行分析