

请思考

关于声音能想到什么内容?





声音与听觉器官

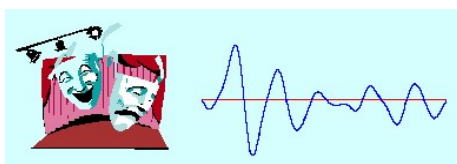
声音知识

■ 有关声音的概念

- 物体在外力的作用下产生振动，对周围的空气产生变化的压强，并以类似波的运动形式在空气中传播，当声波传播到人的耳朵中而引起耳膜振动时，人所感觉到的就是声音。
- **具有普通波的所有特性**，反射等。

■ 声音的几个概念：

- 1)声音的**强弱**：声波压力大小。
- 2)音调的**高低**：声音频率。
- 3)声音的**信号**：分类。



Think

关于声音的“有趣”应用？





Pia!

Pia! 这就是响指!

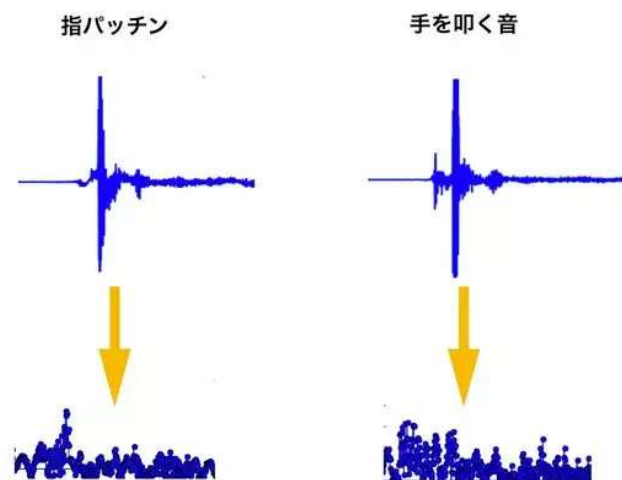


打个响指就让房间灯光变成粉色

- Twitter主上一位叫imajo的日本小哥就做了这样一个有点浪漫的机器学习项目
- 打个响指就让房间灯光变成粉色的视频。



试试频谱分析吧(分辨声音种类)





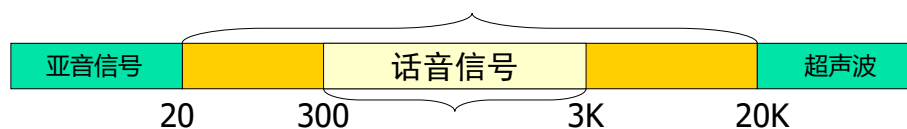
Cymatics

藉由Cymatics使你看见声音

主要参数

声音知识

人能够听到: **20~20 000Hz**, 0~120dB



频率和幅度

亚音信号(次音、次声): 频率<20Hz;

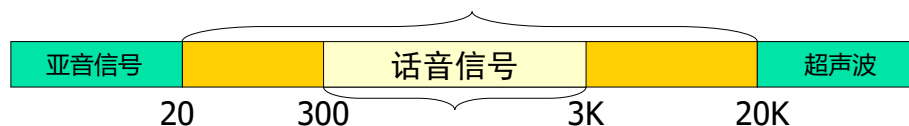
音频信号(可听声): 20Hz~20kHz;

超音频信号(超声波): 频率>20kHz: 超声波探测。



主要参数

声音知识



话音信号: **300~3000Hz** (人说话的频率)

不同种类声音的频带 (带宽、频率范围)

男性语音

100Hz~9kHz

女性语音

150Hz~10kHz

其它

电话语音 200Hz~3.4kHz

调幅广播AM 50Hz~7kHz

调频广播FM **20Hz~15kHz**

宽带音响 20Hz~20kHz



音量与声音类比情况

声音知识

130分贝	喷射机起飞声音
110分贝	螺旋桨飞机起飞声音
105分贝	永久损听觉
100分贝	气压钻机声音
90分贝	嘈杂酒吧环境声音

85
分贝

85分贝及以下 不会破坏耳蜗内的毛细胞

80分贝	嘈杂的办公室
75分贝	人体耳朵舒适度上限
70分贝	街道环境声音
50分贝	正常交谈声音
20分贝	窃窃私语



130分贝



100分贝



保护你的耳朵！

欧盟发出警告 MP3可致永久失聪

欧洲联盟委员会13日发表声明，警告公众若长时间用MP3等播放器大音量听音乐，可能导致永久性听力丧失。

欧盟科学家们向欧盟委员会提出报告说，如果MP3使用者把音量调至89分贝以上，每周收听超过5小时，5年后听力可能永久性丧失。

据新华社

众所周知，长时间戴耳机听音乐会损害听力，近日，英国《每日邮报》载文建议遵循“60/60”原则，即戴耳机听音乐一次不要超过60分钟，音量不要超过最大值的60%，这样就能尽量保护耳朵。

60/60原则

声音调小

在能够听得清楚的情况下，尽量把声音调小一点。



话音编码(声音的数字化)

数字化声音的基本方法

声音信号数字化

声音知识

■ 模拟信号

- 幅度取值是连续的

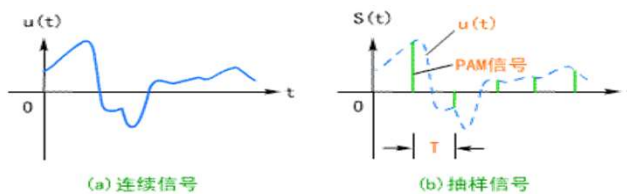


图 0-1 模拟信号

■ 数字信号

- 幅度取值是离散的

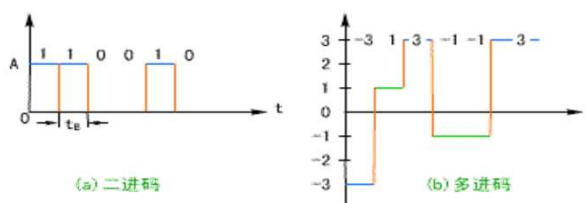


图 0-2 数字信号



声音信号数字化

声音知识

■ 声音的数字化的定义

将具有一定振幅和频率的连续变化的模拟声音信号，通过A/D转换器，以一定的频率对模拟音频信号截取一个振幅值，并用指定字长的二进制位表示，从而将连续的模拟音频信号转变成能被计算机处理的，离散的数字音频信号。

■ 为什么要使用数字信号？

- 1、数字信号计算是一种精确的运算方法，它不受时间和环境变化的影响；
- 2、表示部件功能的数学运算相对容易实现；
- 3、对数字运算部件进行编程，如欲改变算法或改变某些功能，还可对数字部件进行再编程。

数字化过程

声音数字化过程包括采样(抽样)和量化两个步骤。

■ 采样：

- 截取模拟声音信号振幅值的过程称为采样，所得到的振幅值称为采样值。

■ 量化：

- 采样值用二进制位表示称为量化。



奈奎斯特理论(定理):

采样频率须大于等于声音信号**截止频率(最高频率)**的**2**倍, 这样就能把以数字表达的声音还原成原来的声音, 称为**无损数字化**。

量化方法

量化阶距的特点

■ 均匀量化

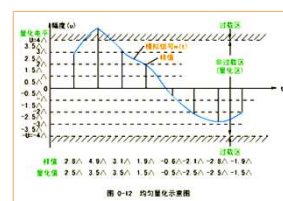
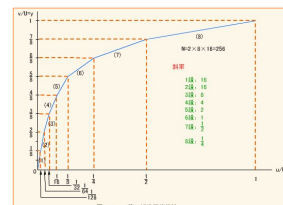


图 0-12 均匀量化示意图

■ 非均匀量化





声音质量与数据率

声音知识

与声音质量有关的重要因素

- **1)采样频率**: 每秒钟内所采样的次数

采样频率越高, 则单位时间内获取的样本数越多, 数字化后的音频信号就越逼真, 但所需要的存储量也越大。

- **2)量化位数**: 表示每个采样样本的二进制位数

量化位数多, 噪声低, 但所需要的存储量也越大。

- **3)声道数**: 采样时声音通道的个数

立体声效果更真实, 但数据量将成倍增加。

声音质量的度量

声音知识

声音质量度量的基本方法有两种:

- **1、客观质量度量**: 用信噪比(SNR) 度量。



信噪比(SNR)

- **2、主观质量度量**: 主观平均判分法(MOS)。



MOS

由若干实验者对声音质量进行评分, 以平均值作为其质量的度量值。

分数	质量级别	失真级别
5	优 (Excellent)	无察觉
4	良 (Good)	(刚)察觉但不讨厌
3	中 (Fair)	(察觉)有点讨厌
2	差 (Poor)	讨厌但不反感
1	劣 (Bad)	极讨厌 (令人反感)



声音质量等级与数据率 (P11)



数据率? (kb/s)

■ 通常分为5级:

- 电话
- 调幅广播
- 调频广播
- 光盘(CD)
- 数字录音带
 - (DAT, Data Audio Tape)

质量	采样频率 (kHz)	样本精度 (bit/s)	单道声/立体声	数据率(未压缩) (kb/s)	频率范围
电话*	8	8	单道声	64	200~3 400 Hz
AM	11.025	8	单道声	88.2	50~7 000Hz
FM	22.050	16	立体声	705.6	20~15 000Hz
CD	44.1	16	立体声	1411.2	20~20 000 Hz
DAT	48	16	立体声	1536.0	20~20 000 Hz

* 电话使用μ律编码，动态范围为13位，而不是8位(详见第3章)

声音文件的数据量(课本补充)

声音知识

■ 计算公式:

■ 声音文件的数据量=(采样频率 × 量化精度 × 声道数 × 时间)/ 8

采样频率单位: - Hz

时间单位: - 秒

数据量单位: - 字节

重要



“坑”在哪里？

Quiz

例：有1声音文件保存了5分钟双声道、16位量化位数、44.1kHz采样频率的未压缩声音数据，问其数据量是多少？

$$\begin{aligned}\text{数据量} &= (\text{采样频率} \times \text{量化精度} \times \text{声道数} \times \text{时间}) / 8 \\ &= (44.1 \times 1000 \times 16 \times 2 \times 5 \times 60) / 8 \\ &= 529240000 \text{ 字节} \\ &= 529240000 / (1024 \times 1024) \approx 50.468 \text{ MB}\end{aligned}$$

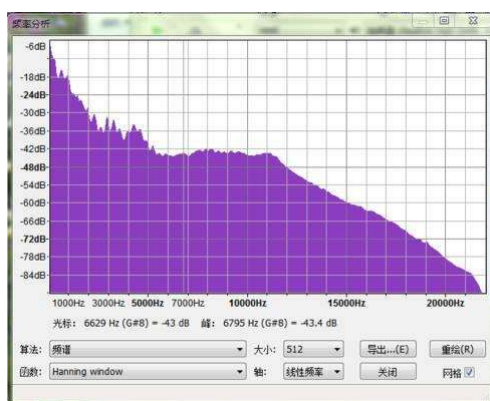
Think

如果要计算“速率”呢？

质量	采样频率 (kHz)	样本精度 (bit/s)	单道声/立体声	数据率(未压缩) (kb/s)	频率范围
电话*	8	8	单道声	64	200~3 400 Hz
AM	11.025	8	单道声	88.2	50~7 000Hz
FM	22.050	16	立体声	705.6	20~15 000Hz
CD	44.1	16	立体声	1411.2	20~20 000 Hz
DAT	48	16	立体声	1536.0	20~20 000 Hz

* 电话使用μ律编码，动态范围为13位，而不是8位(详见第3章)





无损的音乐



320K的音乐



这个杀手非常不冷

我只能偶尔听到中微子穿过身体的声音，没办法，宇宙微波背景辐射还是太吵了。

07-18 23:47

👍 120



夏雪14384321

320MP3跟无损格式在听感方面是有区别的，MP3在细节方面丢失不少。听一些古典音乐就容易听出来了。真正喜欢音乐追求音乐的人，无损是他们的信仰，平时MP3随便听听就可以的人跟它们多解释也没用。

07-16 17:10

👍 104



HiFi

HiFi是High-Fidelity的缩写



蓝牙的带宽？

- 安卓常见的APTX蓝牙编码，经过传输过后的音频文件码率只能达到352kbps，在音质方面也只是略好于320kbps的MP3格式，远达不到无损的标准。
- 随着未来蓝牙协议的升级，带宽提上去了，HiFi格式播放也指日可待。



比较常见的编码-SBC

- 常见的编码：
- aptX、AAC、SBC、LDAC、LHDC和HWA格式。
- **SBC (Sub-band coding, 子带编码)**
 - 挺早的蓝牙音频编码，也是运用最广泛的编码格式。
 - 采用SBC编码的蓝牙耳机，一般都有比特率较低、压缩率较高的问题，用这一编码传输之后的音乐细节基本上大量损失，音乐听感也很差。



比较常见的编码- AAC

- 常见的编码：
- aptX、AAC、SBC、LDAC、LHDC和HWA格式。
- **AAC (Advanced Audio Coding, 高级音频编码)**
 - AAC编码的出现，目的是为了取代MP3格式，在同一码率下AAC的听感要优于MP3。目前市面上主流AAC播放的设备，大部分都来自于苹果。



比较常见的编码-aptX

- 常见的编码：
- aptX、AAC、SBC、LDAC、LHDC和HWA格式。
- **aptX**
 - aptX有着低延时的特性，在其母公司被高通收购之后，广泛的应用于安卓手机平台上；
 - aptX编码还有一些细分的格式，其中一种名为aptX HD的编码格式可以达到24 bit 48KHz的音频格式，基本达到“CD”级别的聆听体验。



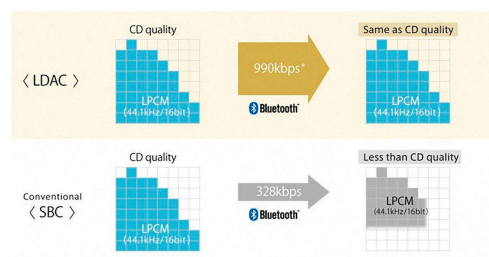


比较常见的编码-LDAC

- 常见的编码：
- aptX、AAC、SBC、LDAC、LHDC和HWA格式。

- **LDAC**

- LDAC是索尼推出的无线音频编码技术，理论传输速率最高可达990kbps，高出SBC三倍之多，这也是目前最接近无损（无损音质：1411kbps）的编码格式。



比较常见的编码-HWA

- 常见的编码：
 - aptX、AAC、SBC、LDAC、LHDC和HWA格式。
- **HWA (HiRes Wireless Audio)**
 - 华为提出的HWA是继aptX HD、LDAC之后，第三家蓝牙高音质协议。
 - HWA提供三种码率模式，分别是400kbps、500kbps/560kbps与900kbps。





声音文件的存储格式

声音知识

■ 常见声音文件的格式

- Wave文件(. WAV)
- Voice文件(.VOC)
- Audio文件(.AU)
- MIDI文件(.MID)
- Sound文件(.SND)
- mp3
- WMA

【Wave文件(. WAV)】

WAV: 波形文件格式,

支持: 各种采样频率和样本精度

声音数据的压缩

格式块(format chunk)

声音数据块(sound data chunk)

```
groupID = 'RIFF'
riffType = 'WAVE'
```

```
FormatChunk
ckID = 'fmt'
```

```
SoundDataChunk
ckID = 'data'
```



声学模型

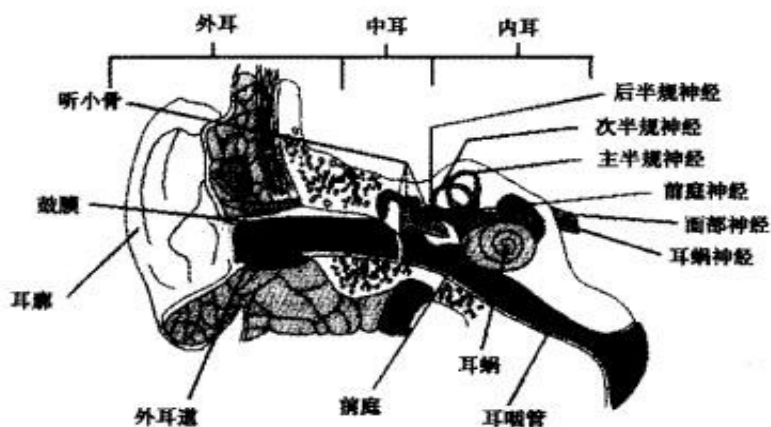
人耳特点及心理声学模型



感知基础—人的听觉器官

MPEG声音

- 人主要用耳朵接受声音。



心理声学

对听觉心理的研究称为心理声学。

- 有趣的结果(人类在不损伤听觉器官的情况下听到的声音):

- 响度之比为1: 1000, 000, 000, 000。
- 所能听到的最低耳语声和喷气式飞机起飞时的声音响度的**万亿分之一**。
- 听到最轻微的声音时, 耳鼓膜的振动**幅度**比单个的**氢原子直径都小**。
- 如果此灵敏度再有任何微小的提高, 我们就会由于周围空气中氮原子的随机运动而无法入睡。



感知声音编码

(Perceptual Audio Coding)

	算 法	名 称	数据率	标准	应用	质量
波形编码	PCM	脉冲编码调制			公共网 ISDN配音	4.0~4.5
	μ -law, A-law	μ -律, A-律	64kbps	G.711		
	APCM	自适应脉冲编码调制				
	DPCM	差分脉冲编码调制				
	ADPCM	自适应差分脉冲编码调制	32kbps	G.721		
	SB-ADPCM	子带-自适应差分脉冲编码调制	64kbps 5.3kbps 6.3kbps	G.722 G.723		
参数编码	LPC	线性预测编码	2.4kbps		保密话声	2.5 ~ 3.5
混合编码	CELP	码激励LPC	4.6kbps		移动通信	4.0~3.7
	VSELP	矢量和激励LPC	8kbps		语音邮件	
	RPE-LTP	规则码激励长时预测	13.2kbps		ISDN	
	LD-CELP	低延时码激励LPC	16kbps	G.728 G.729		
	MPEG	多子带, 感知编码	128kbPs		CD	5.0
	Dolby AC-3	感知编码			音响	5.0



感知声音编码

(Perceptual Audio Coding)

- 根据波形本身得相关行和**模拟人的发音器官的特性**进行编码
- 利用人的**听觉系统的特性**达到压缩声音数据的目的



听觉系统的感知三个特性(重点)

听觉系统的感知特性



响度

声音的强弱

音高

用频率表示声音的音高

掩蔽效应

一种频率的声音阻碍听觉系统感受另一种频率的声音的现象。

响度

声音的强弱

- (1) 物理度量：用客观测量单位来度量。
- (2) 心理度量：用心理上主观感觉的声音强弱度量。
 - 方(Phon) 或 宋(sone)
- 响度的相对量称为响度级，表示某响度与基准响度比值的对数值。
 - 1kHz纯音的声强达到 $10^{-16}\text{W}/\text{cm}^2$ (即0dB声强)时，**人耳刚好能听到，响度就定义0方。**
 - 1kHz纯音的声强达到 $10^{-4}\text{W}/\text{cm}^2$ (即120dB声强)时，**人耳感到疼痛，响度定义为120方。**

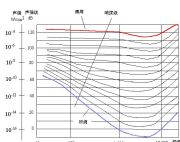


Think

如何界定听觉的范围？



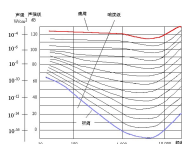
听阈



- 定义：
- 当声音弱到人的耳朵刚好可以听到时的声音强度称为听阈 yù。
- 听阈是随频率变化。
- 听阈-频率曲线



痛阈



- 定义：
- 当声音强到人的耳朵感到疼痛时的声音强度称为痛阈。

- 痛阈也随频率变化。
- 痛阈-频率曲线

特点

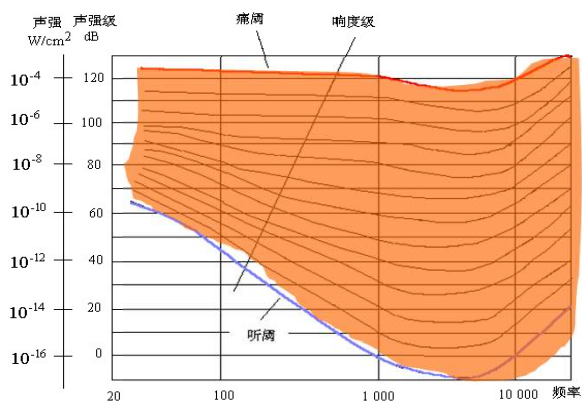
■ 听觉范围

- “听阈-频率”曲线与“痛阈-频率”曲线之间的区域。

■ 主观感觉与频率紧密相关

- 2kHz ~ 4kHz范围的声音最敏感，幅度很低的声音信号都能听到；
- 低频信号和高频信号则幅度要高得多才能听到。
- 语音信号能量主要集中在500 ~ 1000Hz范围左右。

MPEG声音





听(痛)阈是随频率变化

响度

声音的强弱

音高

用频率表示声音的音高

掩蔽效应

一种频率的声音阻碍听觉系统感受另一种频率的声音的现象。



10.1.2 听觉系统对音高的感知特性

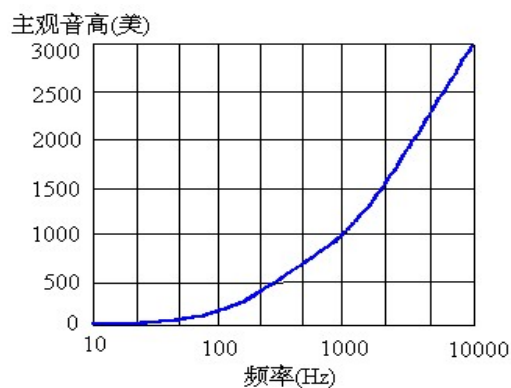
MPEG声音

■ 音高的客观度量与主观度量

- 客观度量：
 - 振动速度(频率)表示声音的音高，单位Hz
- 主观度量：
 - 用人耳对音高的主观感觉确定，单位Mel(美)
 - $Mel = 1000 \log_2(1 + f)$

■ 2.客观音高(频率)与主观音高的关系

- 主观音高与频率(客观音高)
 - 不是线性关系，“音高-频率”
 - 对数曲线



响度

声音的强弱

音高

用频率表示声音的音高

掩蔽效应

一种频率的声音阻碍听觉系统感受另一种频率的声音的现象。



掩蔽效应

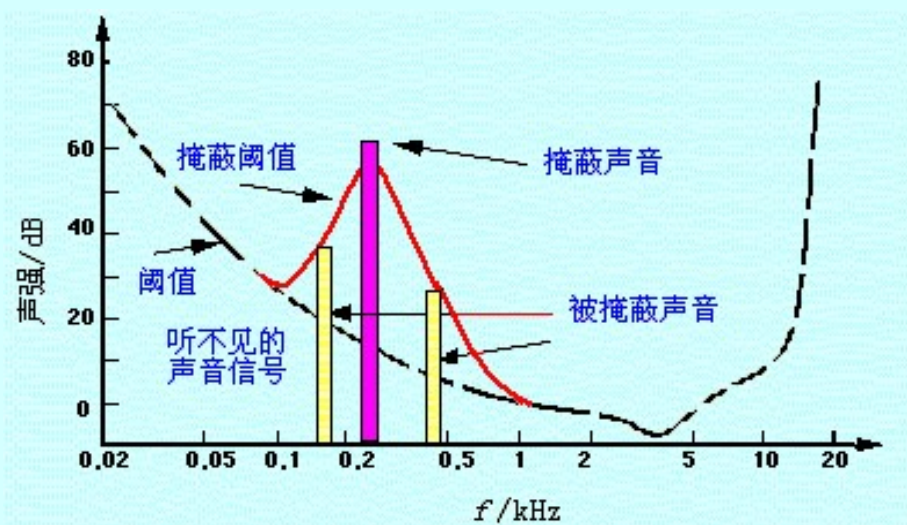
一种频率的声音阻碍听觉系统感受另一种频率的声音的现象。

■ 掩蔽分类

- 频域掩蔽
- 时域掩蔽

掩蔽效应

MPEG声音





1. 频域掩蔽

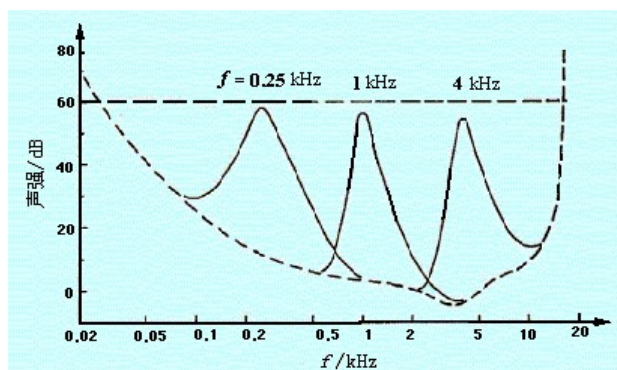
MPEG声音

■ 定义：

- 一个**强纯音**会**掩蔽**在其附近同时发声的**弱纯音**。也称同时掩蔽。

■ 特点：

- 弱纯音离强纯音越近，强纯音对其的掩蔽效果越明显。
- 低频纯音可以有效掩蔽高频纯音，而高频纯音对低频纯音的掩蔽效果不明显。



2. 时域掩蔽

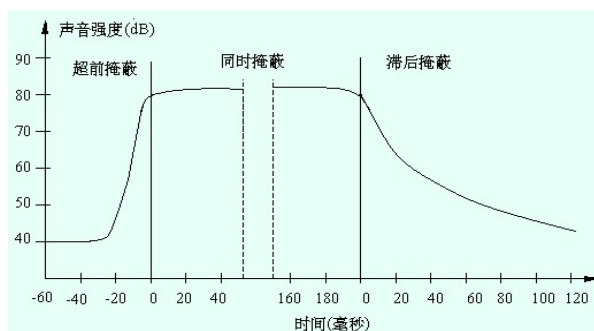
MPEG声音

■ 定义：

- 在时间上相邻声音之间的掩蔽现象称为时域掩蔽。
- 时域掩蔽分超前掩蔽和滞后掩蔽。

■ 产生时域掩蔽的原因

- 大脑处理信息需要一定的时间。





掩蔽效应是心理声学模型的基础

特点

听觉范围

- “听阈-频率”曲线与“痛阈-频率”曲线之间的区域。

主观感觉与频率紧密相关

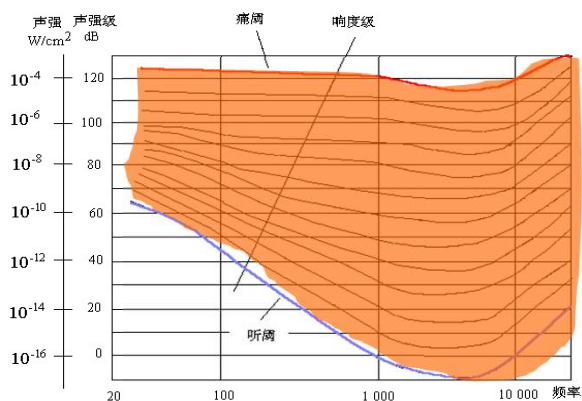
- 2kHz ~ 4kHz范围的声音最敏感，幅度很低的声音信号都能听到；
- 低频信号和高频信号则幅度要高得多才能听到。
- 语音信号能量主要集中在500 ~ 1000Hz范围左右。



人耳的听觉范围



听(痛)阈是随频率变化





作用

■ 人类在利用自身的听觉系统的特性来压缩声音数据方面取得了很大的进展，先后制定了 MPEG-1 Audio, MPEG-2 Audio和MPEG-2 AAC等标准。

- 听觉系统的感知特性，掩蔽效应
- MPEG Audio与感知特性
- MPEG-1 Audio
- MPEG-2 Audio
- MPEG-2 AAC(MPEG2的新扩展)
- MPEG-4 Audio

MPEG Audio与感知特性

MPEG声音

■ MPEG Audio是指MPEG1Audio, MPEG2Audio和MPEG-2AAC, 处理10~20000kHz范围内的声音数据。

■ 一、MPEG Audio与心理声学模型

- MPEG Audio使用心理声学模型来实现压缩声音数据的目的。

■ 二、心理声学模型的要点

- 1.考虑了听觉阈值的概念：听觉系统中存在一个听觉阈值电平，低于这个电平的声音信号人就听不到，因此可以去掉这部分信号而不会影响人对声音信息的获取。
- 2.利用了听觉的掩蔽特性：人的听觉系统感觉不到强声中的轻声。所以，当多种频率声音存在时，凡是被掩蔽的声音均可以去掉。

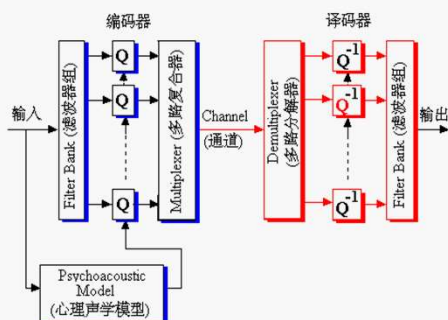


MPEG Audio 采用两种感知编码

MPEG声音

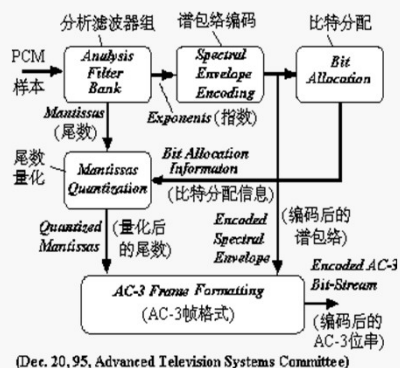
■ 感知子带编码：

- 通过滤波器组后分割成许多子带
- 每个子带对应一个编码器，根据心理声学模型进行编码。



■ 杜比实验室开发的DolbyAC-3编码

- 一种多通道(multichannel)音乐信号压缩技术，它可支持5个3~20000Hz频率范围的通道。



1

感知子带编码

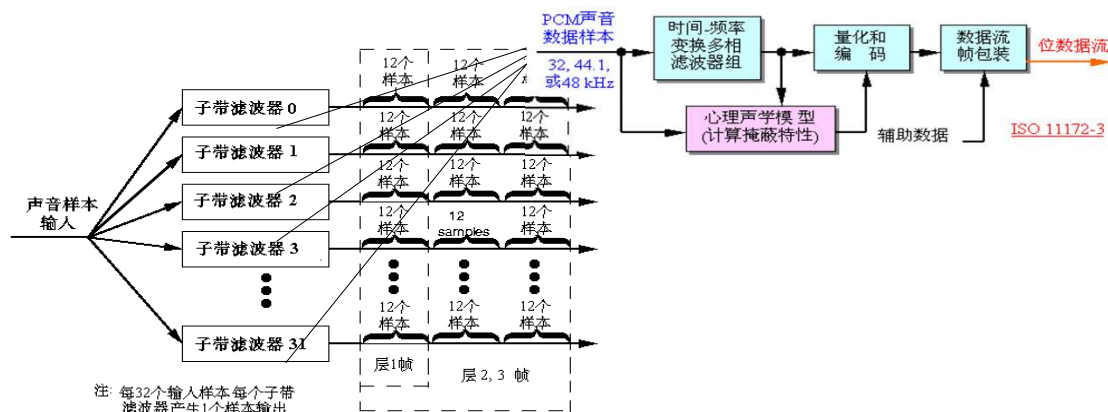
每个子带对应一个编码器，根据心理声学模型进行编码。



MPEG-1 Audio的编码层

MPEG声音

- MPEG Audio压缩编码定义了**3个层次**，3个层次的基本模型相同。
- 层1最基础，每个后继层次比前面的层次压缩比更高，同时编码解码器也更复杂。



MPEG-1 Audio各层压缩所用听觉系统特性

层1

频域掩蔽

层2

频域掩蔽

时域掩蔽

层3

频域掩蔽

时域掩蔽

立体声/声道数量



2

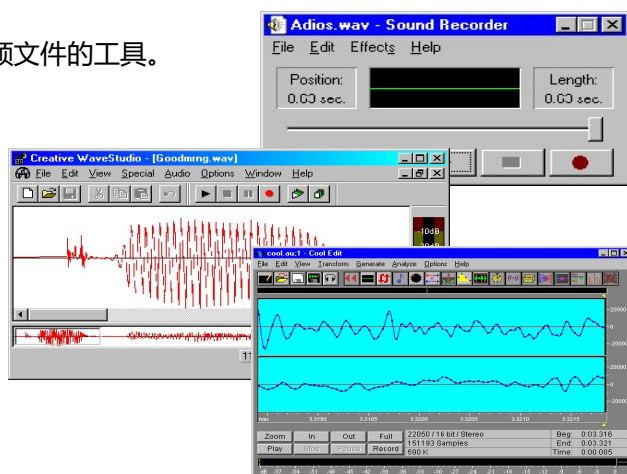
杜比全景声(Dolby Atmos)

杜比音效试听

声音工具

声音知识

- Windows的录音机
用于录制、播放、编辑wave波形音频文件的工具。
- 声音卡附带的工具
- 网络工具



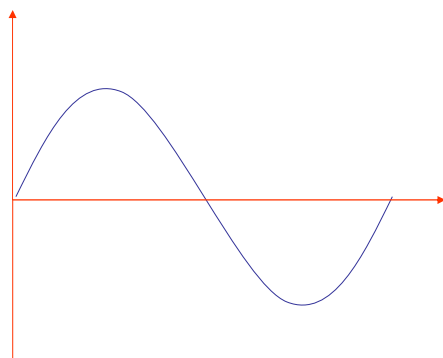


MIDI声音

Think

如何表现这个波形?

+127
...
+64
+32
0
-32
-64
...
-128



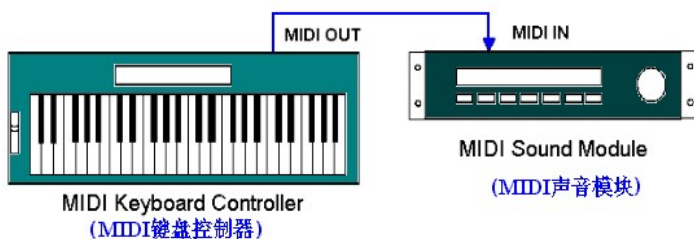


电子乐器数字接口MIDI

声音知识

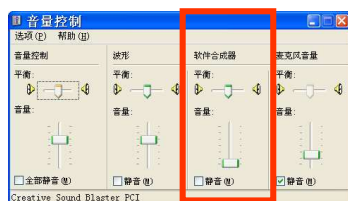
■ 一套**指令**的约定，不是声音信号。

- MIDI标准于1983年8月5日正式制定(数字音乐国际标准)。
- 规定了音乐设备之间、音乐设备与计算机之间的连接电缆和硬件，还规定了音乐设备之间、音乐设备与计算机之间的**通信协议**。
- 各种音乐设备只要遵从MIDI标准配备MIDI信息处理器和MIDI接口，就能相互连接和通信。



优点

- 1) 文件小
- 2) 易于编辑
- 3) 背景音乐



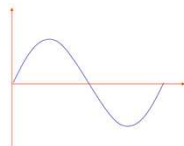
周华键:风雨无阻15K.mid(文件大小 15K)

邵正宵:干纸鹤15K.mid(文件大小 14K)

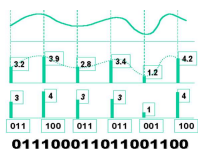




波形音频

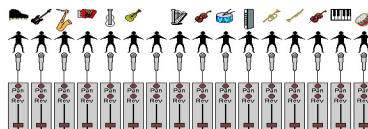


- 数字化的声音波形
- 音质决定存储空间
- 难于编辑



MIDI音频

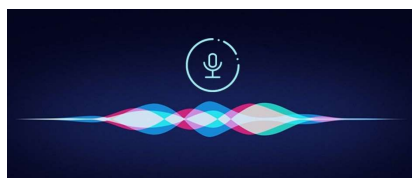
- 音乐信息的合成器指令
- 节省存储空间
- 灵活编辑(借助音序器)
- 处理话音的能力较差



Think

指令如何表现/还原声音?

人的耳朵只能听到波形震动 (模拟声音) ?





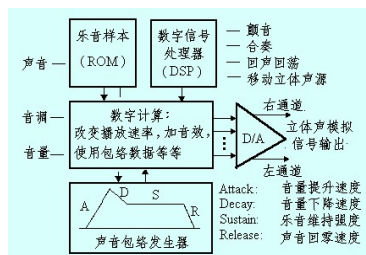
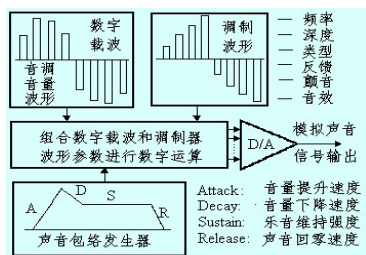
调频合成

(频率调制(FM)合成)

VS

波表合成

(乐音样本合成)



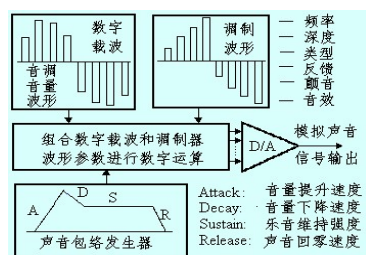
调频合成

(频率调制(FM)合成)

VS

波表合成

(乐音样本合成)



【调频(FM)合成声音】

乐器的声音由两种、三种或四种不同的频率的波形叠加合成。

从理论上讲, FM合成方法可以产生任何乐音, 但是, 这种“物理课式”的合成方法合成出来的声音不够真实。



调频合成

(频率调制(FM)合成)

VS

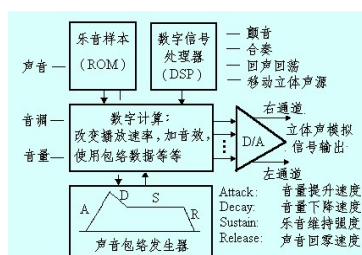
波表合成

(乐音样本合成)

【乐音样本合成】

波表合成采用“波表查找”技术来产生音乐。波表技术以高精度录下真实乐器的声音，并将录制的各种实际乐器的数字化声音存储在只读存储器(ROM)中。

还原的声音质量非常高，所产生的音响效果比FM合成器合成的音乐质量高得多。



MIDI系统(产品)



音序器



为MIDI作曲而设计的计算机程序。

控制器



电子乐器的一种装置(乐器键盘)等。

合成器 (Synthesizer)

合成器是利用数字信号处理器(DSP)或其他芯片来产生声音的电子装置。合成器产生声波, 并把声波通过声音发生器送往扬声器。



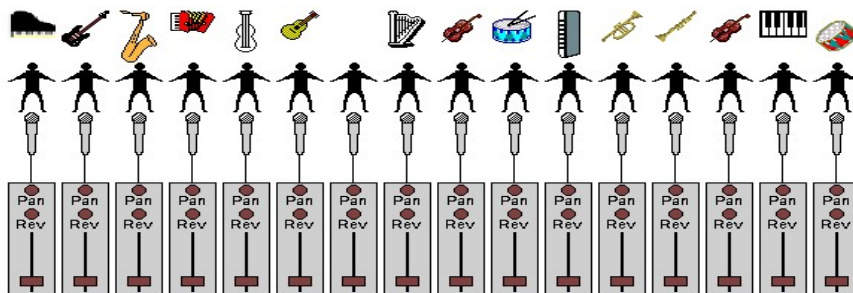
决定质量!!!

MIDI通道

声音知识

■ 通道:

- 一个通道指一条独立的信息传输路线。单个物理MIDI通道分成16个逻辑通道, 每个通道访问一个独立的逻辑合成器, 可充当一种乐器的角色。





英国SSL公司的模拟调音台

声音知识



硬件到软件



参照SSL 4000的硬件制作，Waves的工程师花了一年时间进行物理建模等分析和设计，最终使得这套插件的声音几乎与SSL 4000硬件几乎完全相似。

减少存储空间，提高数据传输率？

音频编码



话音编码(自学?)

Multi-AI-B005 多媒体技术-基础理论技术-话音编码.pptx



关于声音的其它研究领域

Multi-AI-B006 多媒体技术-基础理论技术-声音各领域的研究情况.pptx



本章小结



数字声音

声音知识

■ 两个参数:频率和幅度

- 1)人能够听到的是20~20 000Hz, 0~120dB
- 2)话音信号:300~3000Hz(人说话的频率)

■ 2)声音数字化过程

- 声音数字化过程包括采样(抽样)和量化两个步骤。

■ 3)奈奎斯特理论:

- 采样频率须大于等于声音信号截止频率(最高频率)的2倍, 这样就能把以数字表达的声音还原成原来的声音, 称为无损数字化。



数字声音

声音知识

■ 量化方式

- 均匀量化和非均匀量化。

■ 声音质量与数据率

- 1)与声音质量有关的重要因素
- 2)声音文件的数据量=(采样频率× 量化精度× 声道数× 时间)/ 8

■ MIDI

- MIDI音频和波形音频的区别