

The background of the title slide features a large, detailed speaker on the left side, emitting blue sound waves. To the right of the speaker is a colorful audio waveform or equalizer display with vertical bars in shades of blue, green, yellow, and red. The title text '多媒体技术基础' is overlaid in the center in a large, red, stylized font.

多媒体技术基础

第二章 媒体处理技术

§2.1 数字音频处理技术



2022年9月20日





授课内容

- ◆ 第一部分 多媒体的计算
 - 第一章 多媒体计算机系统
 - 第二章 媒体处理技术
 - 第三章 多媒体数据压缩
- ◆ 第二部分 多媒体的存储
 - 第四章 多媒体数据的数字存储
- ◆ 第三部分 多媒体信息的分析与处理
 - 第五章 多媒体信息分析与处理
- ◆ 第四部分 多媒体网络应用
 - 第六章 实时多媒体通信





第二章 媒体处理技术

◆ § 2.1 数字音频处理技术

□ § 2.1.1 数字音频在计算机上的应用

- § 2.1.1.1 声音文件

- § 2.1.1.2 电脑音乐

- § 2.1.1.3 计算机上的音频处理

□ § 2.1.2 用比特串记录声音信息

- § 2.1.2.1 声音的基本特性

- § 2.1.2.2 声音信号的波形的数字记录

- § 2.1.2.3 针对声音信号“知识”的数字记录

◆ § 2.2 数字图像处理技术

◆ § 2.3 数码相机技术

◆ § 2.4 数字视频处理技术





音频处理技术概述

- ◆ 声音是携带信息的极其重要的媒体，是多媒体技术研究中的一个重要内容。声音的**种类繁多**，如人的话音、乐器声、动物发出的声音、机器产生的声音以及自然界的声音等。
- ◆ 这些声音有许多**共性**，也有各自的**特性**。在用计算机处理这些声音时，既要考虑它们的共性，又要利用它们的各自的特性。





记录声音: 八音盒



1796年，瑞士的一名钟表匠发明了世界上最古老的八音盒，它被装在一枚纯金的印章内，从此随着八音盒的迅速发展，绚丽的八音盒文化为欧洲的音乐史增添了一抹亮色。

“八音” 《三字经》：“匏土革，木石金。丝与竹，乃八音。”

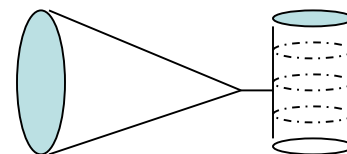
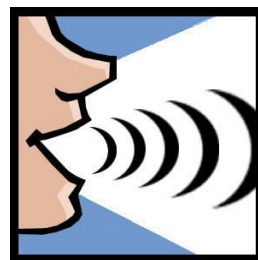
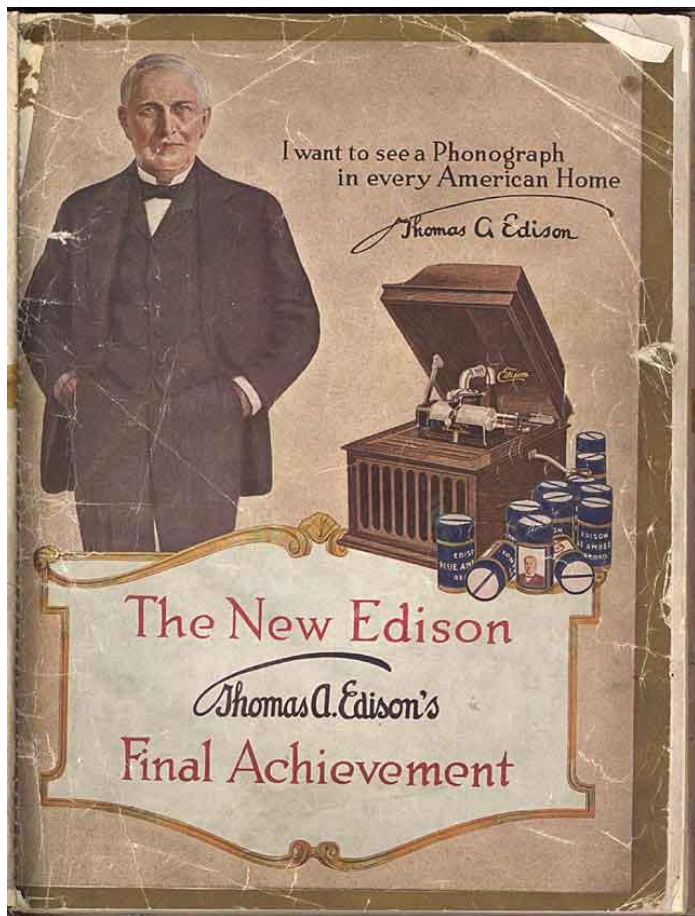


{ynh,cxh}@ustc.edu.cn

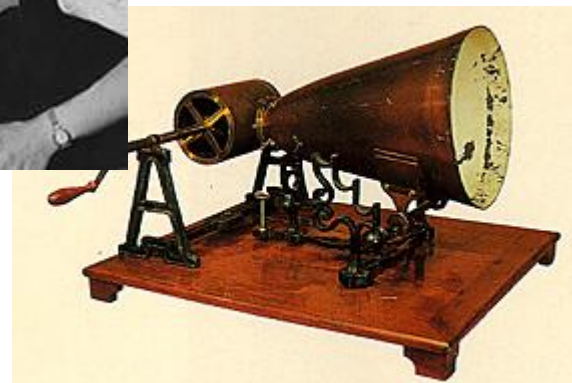




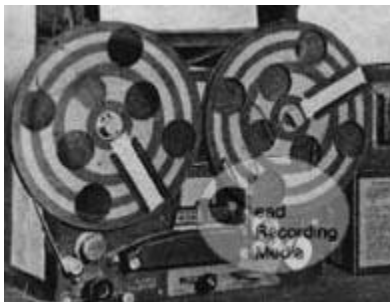
记录声音: 留声机



■ Edison, 1877



记录声音: 磁性录音

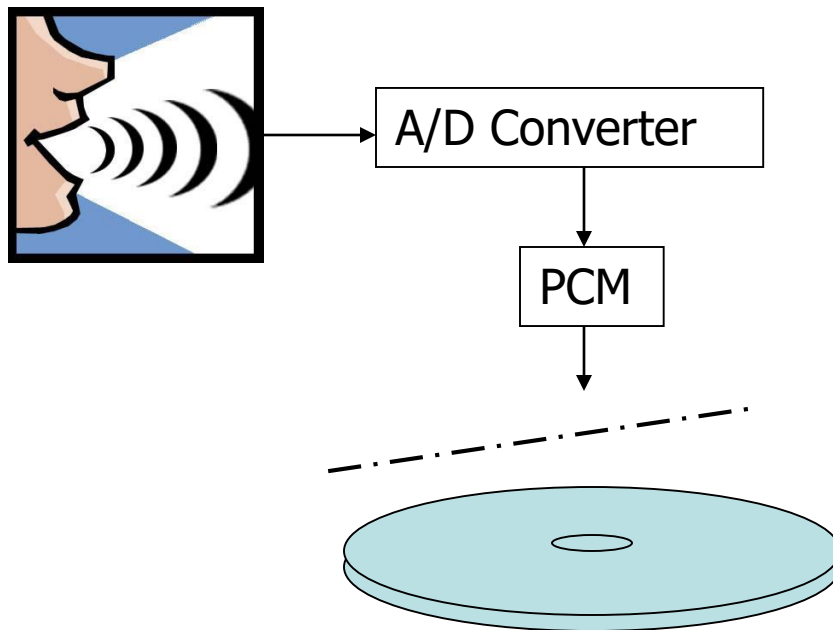


磁性录音，由丹麦发明家Valdemar Poulsen于**1899-1900**年期间发明，被用作录像和录音磁带。





■ Philips, 1978





如今常见声音文件

◆ CD

◆ 业界标准

◆ *.WAV文件

◆ 用于存储

◆ *.MID、*.RMI文件

◆ 用于合成音乐

◆ *.MP3、*.ra、*.rm、
.wma、.m4a、*.amr、
*.ogg文件

◆ 着重于压缩





声音文件的播放

◆ 操作系统自带的媒体播放器

- Windows MediaPlayer
- 各种Linux版本下也有媒体播放器

◆ 著名的音乐播放软件

- 经典：Winamp、RealPlayer等
- 流行：酷狗、酷我、QQ、iTunes、千千静听、KMPLayer等

◆ 声卡或主机驱动中自带的播放软件

- 这种形式现在已经几乎不存在了



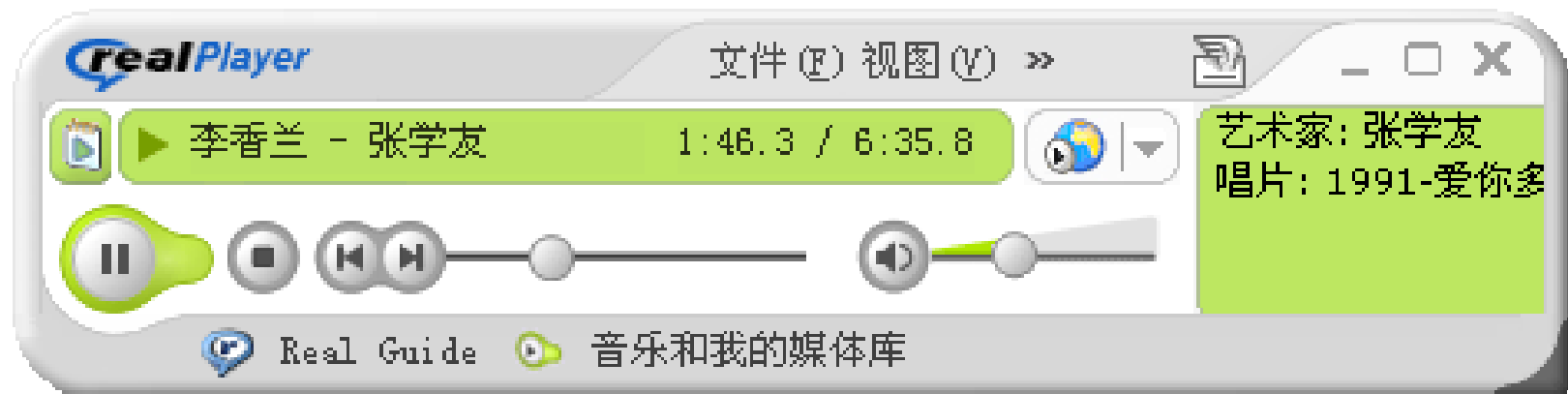


最经典的播放器





曾经流行的压缩音乐mp3播放器





如今的MP3播放器

- 不再是简单的音乐播放
- 云音乐的特征
- 个性化与自动推荐





MP3不仅保存声音数据

歌词搜索

歌曲名: 庄心妍 - 一万个舍不得

搜索

正在播放的歌曲时长为03:28, 蓝色显示为当前已关联歌词。

	歌手	歌曲名	时长	ID	网友评分
1	庄心妍	一万个舍不得	03:28	111442574	★★★
		不得	03:28	13287625	★★☆
		不得	03:28	352958895	★☆☆
		不得	03:28	218964816	☆☆☆
		不得	03:28	48203091	☆☆
		不得	03:28	127281786	☆
		不得	03:28	82475187	☆
		不得	03:28	359984991	☆
		不得	03:28	353306781	☆
		不得	03:28	360376193	☆

选择

关闭

文件属性 (2/2)

属性 专辑封面

文件名: C:\KuGou\庄心妍 - 一万

文件名猜测标签

歌曲来源: 本地歌曲

标签

标题: 一万个舍不得

艺术家: 庄心妍&祁隆

专辑: 一万个舍不得

音轨:

流派:

年代: 2012

备注:

格式

编码: MPEG 1 Layer 3

声道: 立体声

频率: 44100 Hz

比特: 16 Bits

码率: 320 Kbps

长度: 3:28.222

增益:

高级 繁 简 转码 删除 重命名

MP3标签

读取类型优先级: APEv2 > ID3v2 > ID3v1

写入类型: ID3v1 & ID3v2

ID3v2编码类型: ISO-8859-1

☒ ID3v2写入时使用填充数据

上一首

下一首

重新读取文件

保存到文件

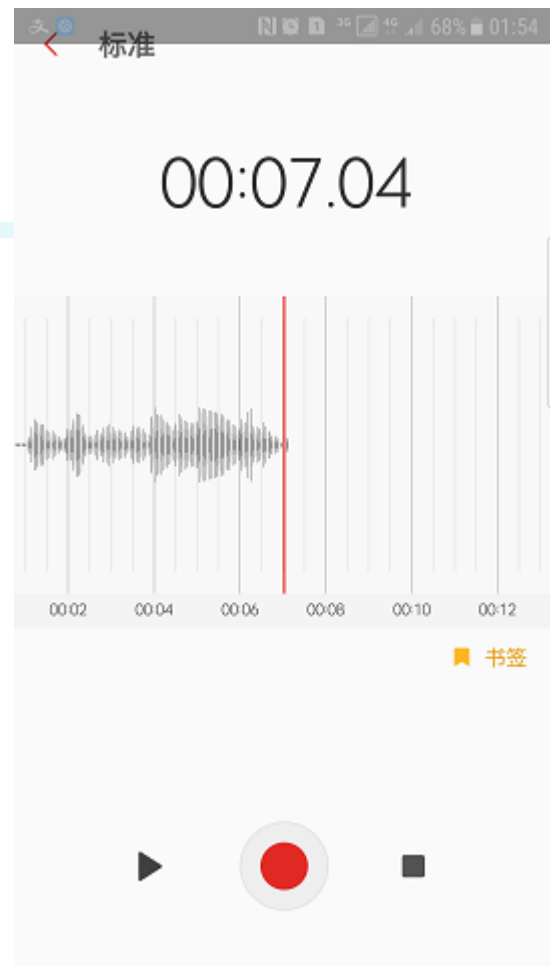
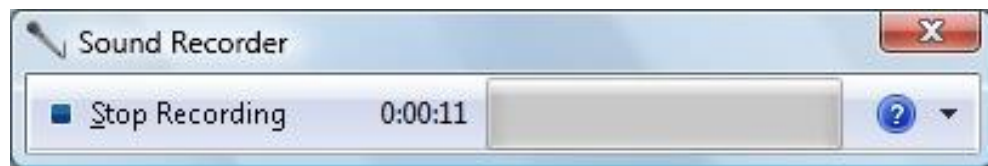
关闭





声音文件的录制

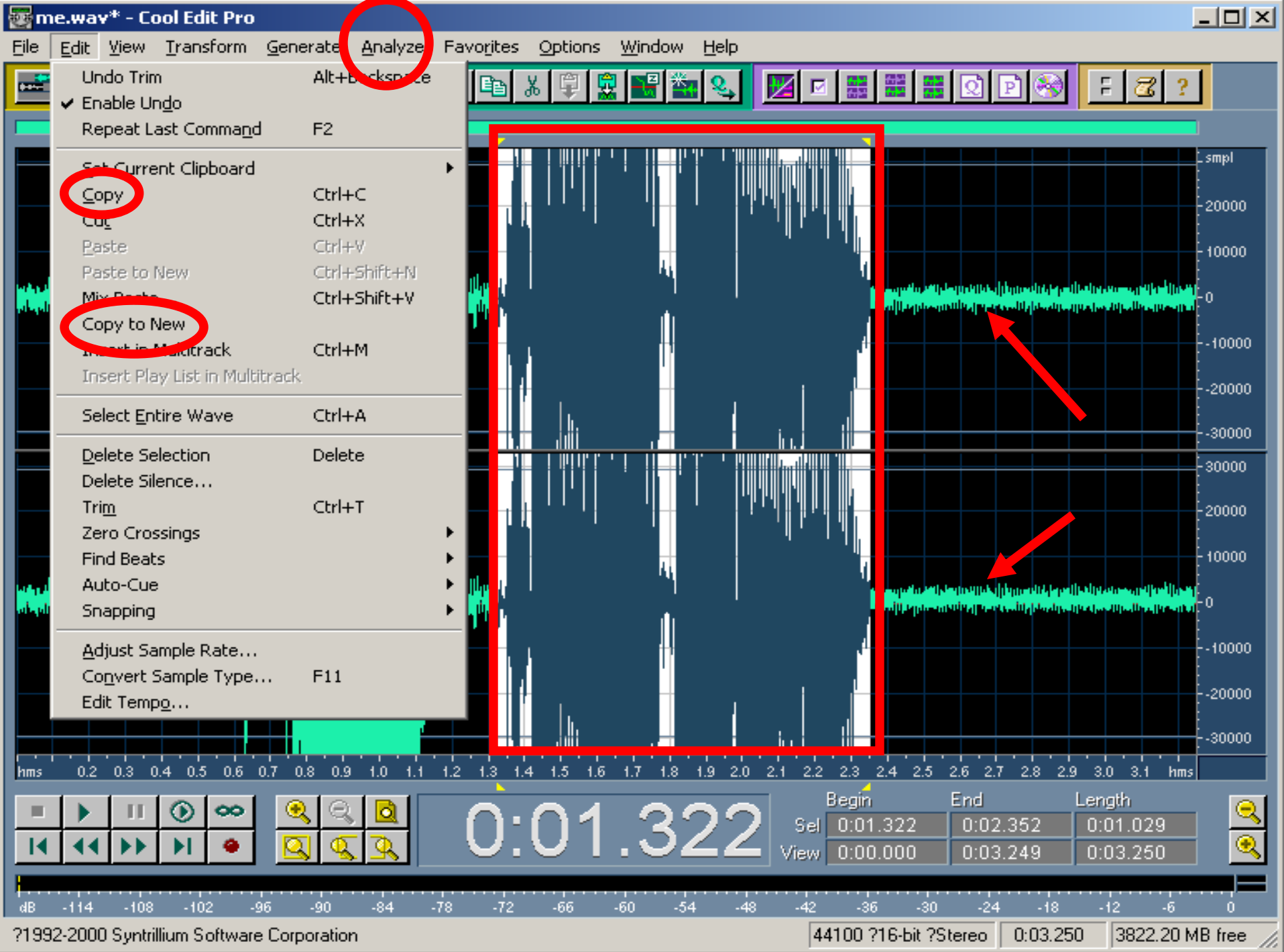
- ◆操作系统自带的录音机
- ◆声卡自带的应用程序
- ◆用于电脑音乐处理的通用软件
- ◆**必要条件**：拥有麦克风
 - 智能手机内置麦克风（阵列）





CoolEdit操作界面





Invert

Reverse

Silence

DirectX

Amplitude

Delay Effects

Filters

Noise Reduction

Special

Time/Pitch

Chorus...

Delay...

Echo...

Echo Chamber...

Flanger...

Full Reverb...

Multitap Delay...

Reverb...

Sweeping Phaser...

幅度上的翻转??

时间上的翻转

静音

在频率域进行处理
有关幅度的处理

关于延时的处理（时间域）

关于噪声的处理



小结：声音文件

◆自然界声音录制而成

- 声音波形的直接记录：*.WAV文件
- 存储压缩后的波形数据：*.MP3文件

◆电脑合成的声音

- 纯粹人造声音：*.MID、*.RMI文件
- 与自然界声音混合拼接

◆声音文件的操作

- 常规操作：录制、播放
- 编辑：音效生成、声音内容篡改

变迁：唱片→磁带→CD光盘→磁盘文件→网络流媒体





第二章 媒体处理技术

◆ § 2.1 数字音频处理技术

□ § 2.1.1 数字音频在计算机上的应用

- § 2.1.1.1 声音文件
- § 2.1.1.2 电脑音乐
- § 2.1.1.3 计算机上的音频处理

□ § 2.1.2 用比特串记录声音信息

- § 2.1.2.1 声音的基本特性
- § 2.1.2.2 声音信号的波形的数字记录
- § 2.1.2.3 针对声音信号“知识”的数字记录

◆ § 2.2 数字图像处理技术

◆ § 2.3 数码相机技术

◆ § 2.4 数字视频处理技术





电脑音乐

◆ 电脑音乐是指利用多媒体**电脑**及与音乐相关的软件为工具**制作**并播放出来的音乐。电脑音乐文件有多种格式，如：wav文件、mp3文件、midi文件、aac文件、wma文件、ra文件等。

◆ 音乐是一种抽象的**艺术**，用音高和节奏来展示艺术的魅力；电脑是一种**科技**工具，把复杂的运算集中在一块小小的集成电路中，用高速的衍变解脱现实的束缚。音乐和电脑结合起来就是电脑音乐，它是数码时代的听觉新艺术，是音乐与科技的完美结合。





电子音乐 (Electronic Music)

电脑音乐 (Computer Music)

◆ 电子音乐 (Electronic Music) 不以物理振动为基本发音原理，完全通过电子设备生成或处理加工音乐

◆ 随着计算机逐步进入电子音乐领域，并逐渐成为了电子音乐的核心，人们将这种以计算机为核心的电子音乐称为电脑音乐 (Computer Music)。← 标志着电子音乐进入数字化的阶段。

□ 从80年代到90年代中期，PC运算能力有限，在音乐制作中主要是用来担任音序编辑的工作，即使是用来合成音色也是非常有限的。而声音的合成、编辑工作则主要是交给专门的设备——各种合成器音源以及采样器音源等。这一时期的电脑音乐制作往往需要多件设备协同运转，MIDI技术规格就是这一时期的重要产物。

□ 从90年代后期开始，PC的功能越来越强大，陆续出现了各类专业级的依靠个人电脑CPU运算的软件合成器、软件采样器以及音频编辑软件，并且软件质量伴随着电脑科技的发展而迅速提高。

◆ 目前使用最多的“电音”，往往不区分电子音乐和电脑音乐

□ 本课程中不对上述术语做区分





让科技与艺术握手

- ◆用鼠标“演奏”“高山流水”
- ◆用舞蹈动作控制音乐
- ◆让歌曲“倒背如流”

- ◆鼓的敲击声 >> 钟摆的声音
- ◆摩托车发动声 >> 狮子的吼声
- ◆自己的声音 >> 沈腾的声音？





改变了艺术创作的流程

◆传统的音乐创作

□ 作曲家谱曲 → 乐谱交给乐队演奏 → 录音师与乐队合作录制，发行

◆一体化音乐制作人

□ 构思旋律谱曲 → 软件模拟乐队演奏 → 软件合成乐曲和其他声音 → 声音文件

低成本导致大量网络红歌诞生





示例：虚拟歌手

◆初音未来（初音ミク）是以Yamaha的VOCALOID 2语音合成引擎为基础开发贩售的虚拟女性歌手软件。软件使用了Yamaha的VOCALOID 2语音合成引擎，把人类的声音录音并合成为酷似真正的歌声。只需输入音调、歌词则可发出声音，亦可以调整震音、音速等的“感情参数”，最多能够16人合唱，亦支持即时演奏。制作完成后会以WAV格式输出，但软件本身只可做出歌唱部分，伴奏音声需要使用其他音乐软件合成。

◆初音未来是世界上第一个使用全息投影技术举办演唱会的虚拟偶像。





示例：虚拟歌手

2013年9月6日，亚洲天王周杰伦在台北小巨蛋举行演唱会，并“邀请”到已故台湾歌手邓丽君与之同台献唱。



虚拟影像重建技术 + 语音合成技术

《明日之子》盛世魔音赛道的二次元歌手赫兹





电脑音乐创作

◆ 电脑音乐创作分为三个步骤:

- 编曲（确定旋律）
- 合成（不同的音乐片断合成等）
- 后期制作（加入人工录音等）

◆ **CakeWalk**、**Vegas Audio**、CoolEdit、AutoScore等

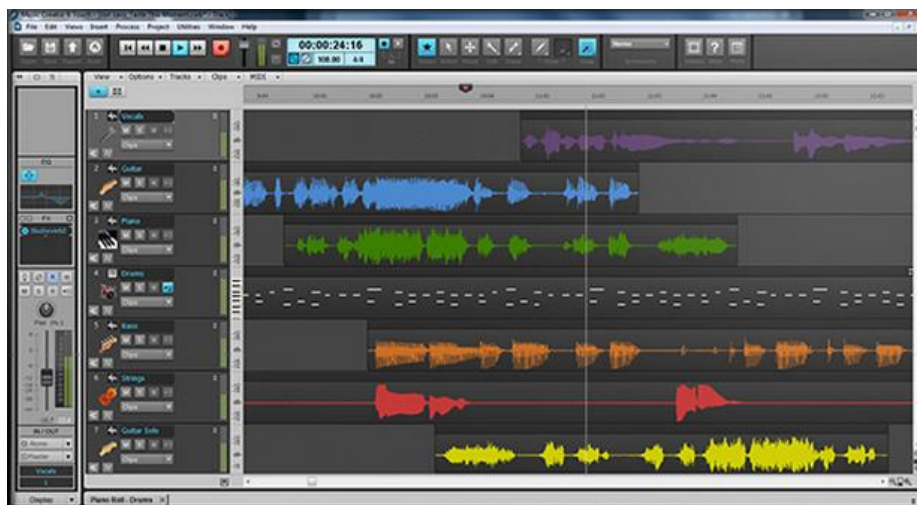




电脑音乐的编曲

◆ **Cakewalk**，由Twelve Tone System公司（现改名为Cakewalk公司）开发。早期，它是专门进行**MIDI制作、处理**的音序器软件。

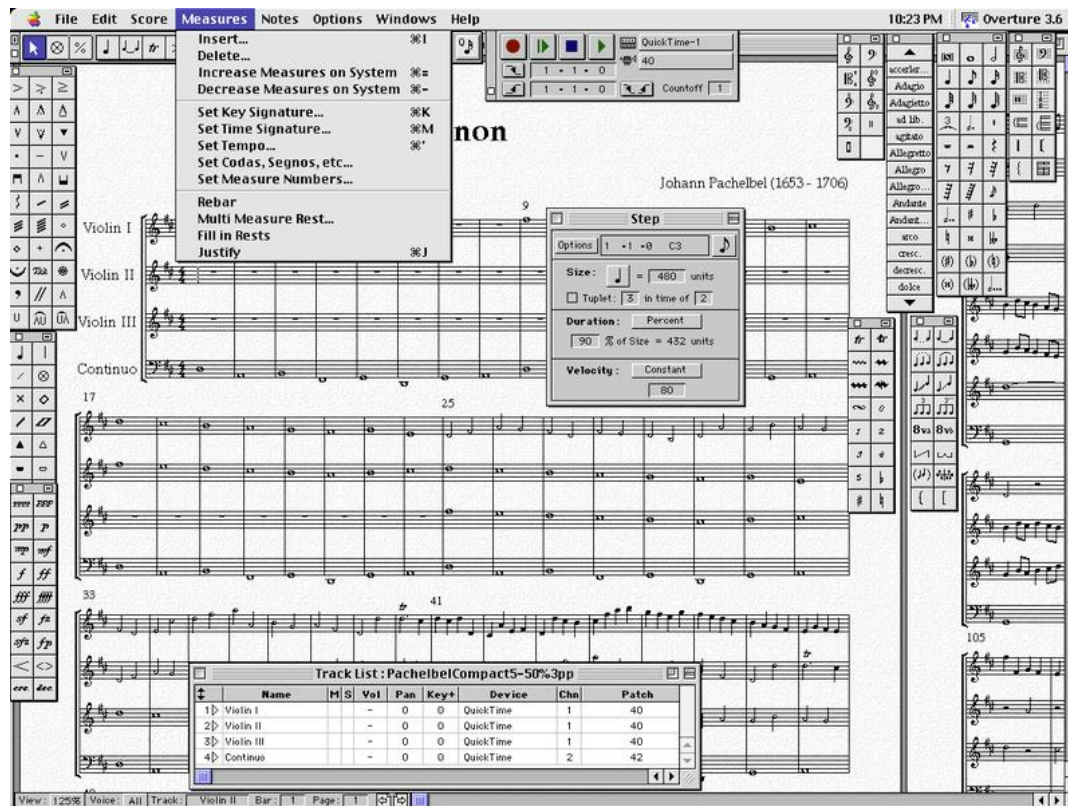
◆ 2009年后该公司发布的新版本软件更名为**SONAR**。Sonar在Cakewalk的基础上，增加了针对软件合成器的全面支持，并且增强了**音频**功能。





打谱软件Overture

Overture was originally published by Opcode Systems. It was later published by Cakewalk Inc. It is, as of 2009, published by Sonic Scores (previously named Geniesoft).



在许多曲谱网上都有大量Ove格式的琴谱







小结：电脑音乐

◆ 电脑音乐创作分为三个步骤：

- 编曲（确定旋律）
- 合成（不同的音乐片断合成等）
- 后期制作（加入人工录音等）

◆ 专业音乐创作环境

- 音乐编曲/编辑软件
- 专业的效果器插件（对人声、乐音进行均衡、压缩、混响等效果美化）
- 专业的音源插件（不同音色乐器的样本库）





第二章 媒体处理技术

◆ § 2.1 数字音频处理技术

□ § 2.1.1 数字音频在计算机上的应用

- § 2.1.1.1 声音文件
- § 2.1.1.2 电脑音乐
- § 2.1.1.3 计算机上的音频处理

□ § 2.1.2 用比特串记录声音信息

- § 2.1.2.1 声音的基本特性
- § 2.1.2.2 声音信号的波形的数字记录
- § 2.1.2.3 针对声音信号“知识”的数字记录

◆ § 2.2 数字图像处理技术

◆ § 2.3 数码相机技术

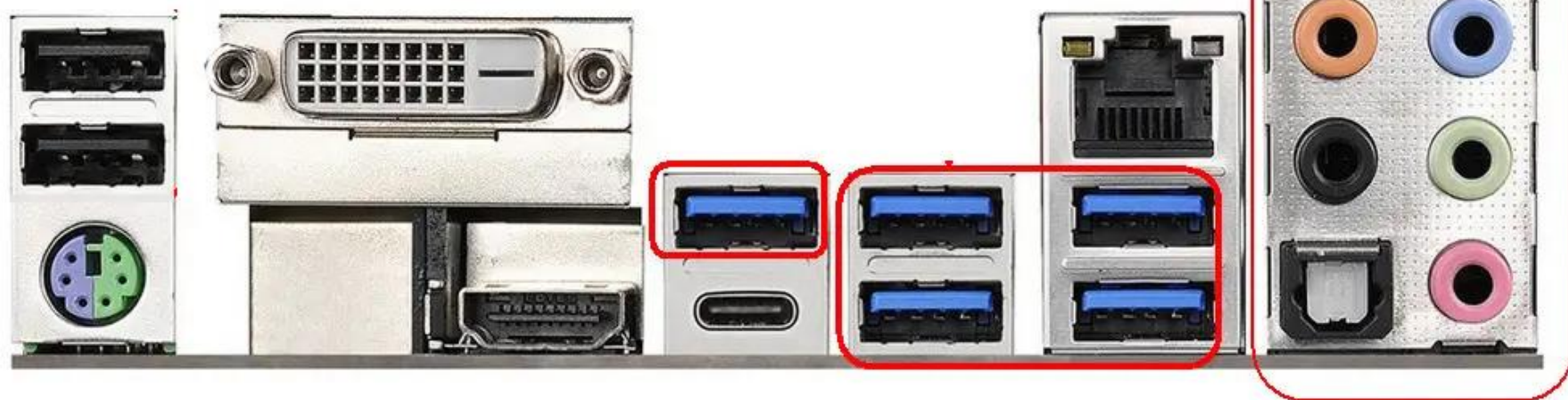
◆ § 2.4 数字视频处理技术





PC上的音频接口（插座）

1. 橙 CTR BASS 中央及重低音输出（C/SW）
2. 黑 REAR 4.1 以上音箱后置音箱接口
3. 灰 SIDE 4.1 以上音箱的边置环绕音箱接口
4. 蓝 Line In （音频输入接口）
5. 绿 Line out （音频输出接口）
6. 粉红 MIC （麦克风）



*3 灰色|光纤S/PDIF数字音频输出接口：7.1声道时，连接侧边环绕喇叭输出。





音频接口(插头)

绿联



HiFi级光纤音频线
电视/游戏机接音响

绿联



京东
配送

3.5转双6.5音频线
高保真设计
稳定传输

JINGHUA
晶华



电脑音频
电脑主机
笔记本
供电
显示设备
显示器
投影仪等

VGA转HDMI转换线 高清转换音视频同步



HDMI输出指示灯



HDMI
输出接口

红白L/R
音频输出接口

SPDIF光纤
音频输出接口

接显示器

接音响

接功放



3.5转双莲花 1分2音频线



手机(嵌入式计算机)音频接口



3.5mm



蓝牙



DAC Chipset



Type-c to 3.5mm Female

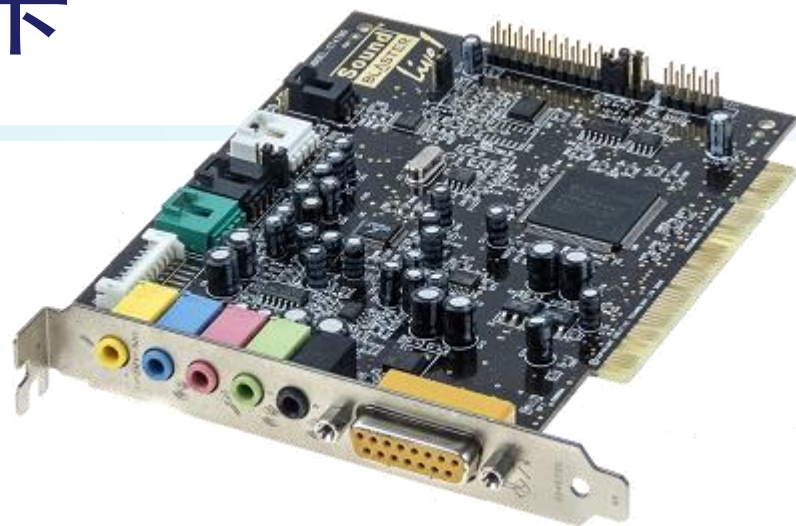
Type-C



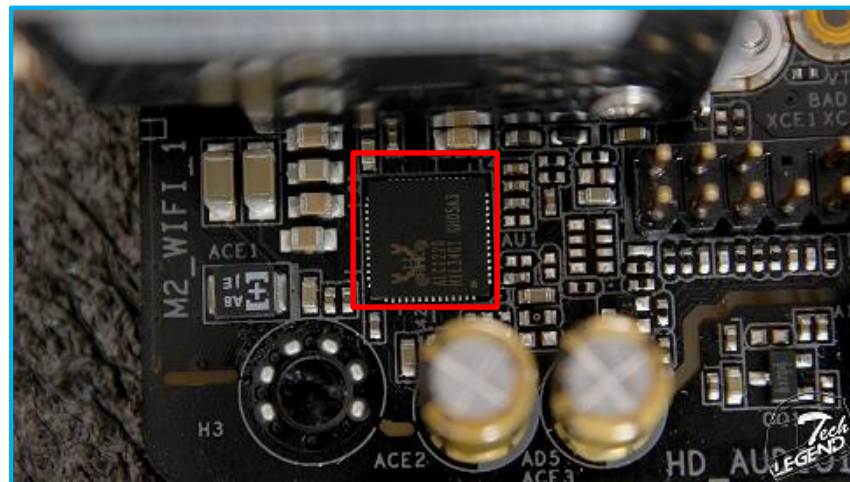


计算机上的声卡

早期是独立声卡，通过扩展槽连接PC主板



7.1 CH HD Audio
(Realtek ALC1220 Audio Codec)

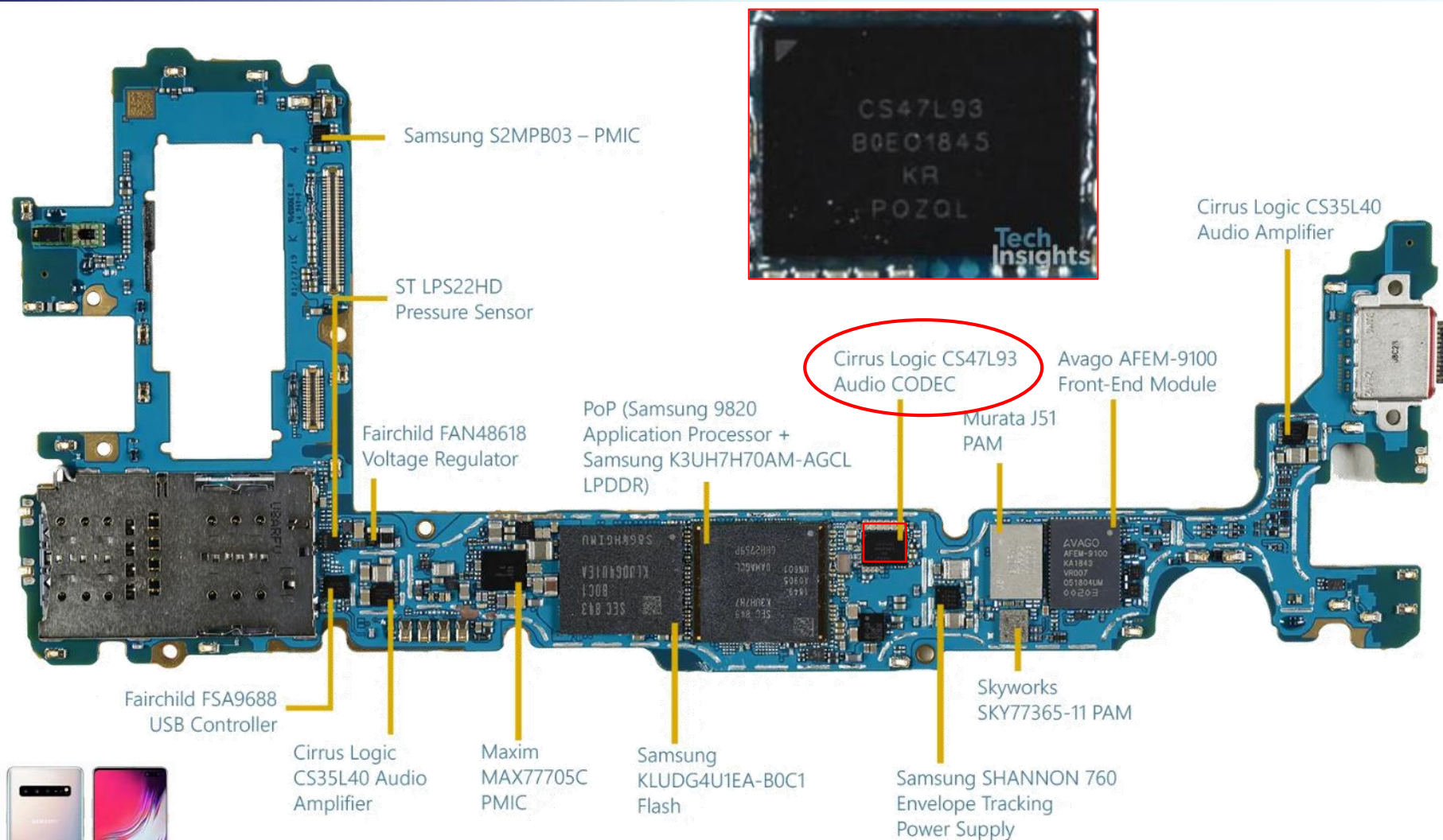


现在，各类PC几乎都是集成声卡





音频编解码芯片Cirrus logic CS47L93@Samsung Galaxy S10+



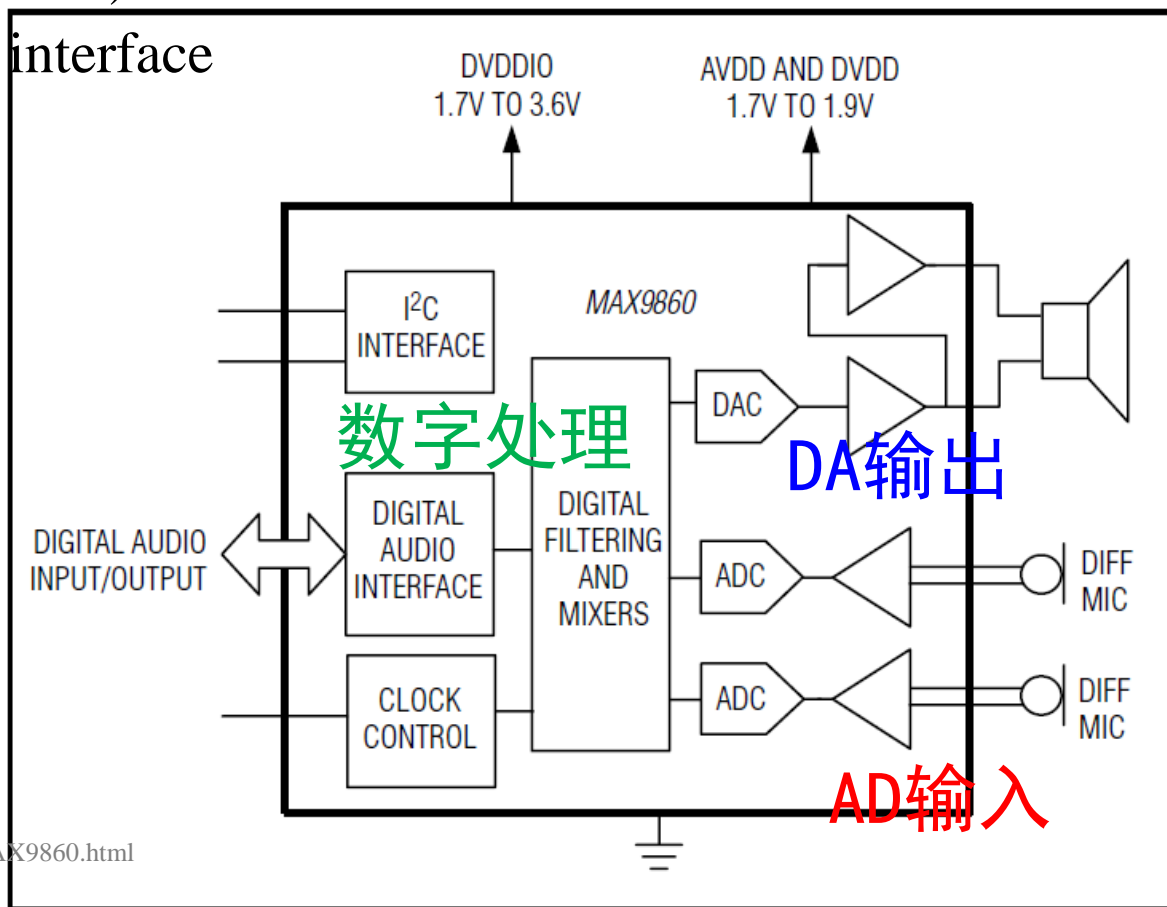


单声道音频编解码器 MAX9860

16-Bit Mono Audio Voice Codec

- ◆ Digital Highpass Elliptical Filters with Notch for 217Hz (GSM)
- ◆ 90dB DAC DR ($f_S = 48\text{kHz}$) **Dynamic Range (DR)**
- ◆ 81dB ADC DR ($f_S = 48\text{kHz}$)
- ◆ 2-wire, I2C-compatible interface

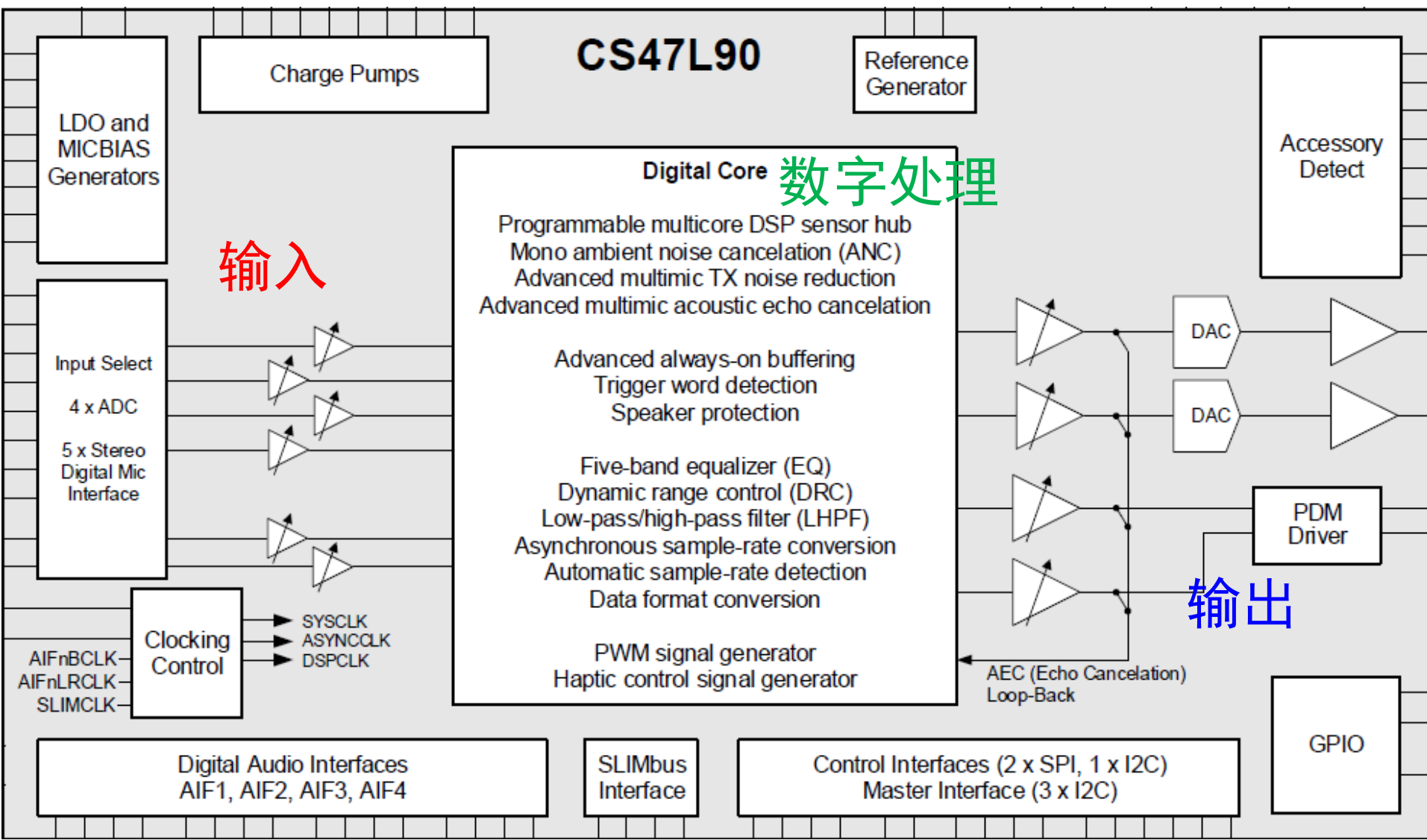
MAX9860是一种低功耗，语音频带，单声道音频编解码器。旨在提供完整的音频解决方案，适用于无线语音耳机和其他单声道语音音频设备。





多通道音频编解码芯片

Cirrus logic CS47L93





多通道音频编解码芯片

Realtek ALC1220

The ALC1220-VB is a high performance multi-channel audio codec with bilingual interface supporting High Definition Audio 1.0a and industrial standard I2S+I2C mode.

The ALC1220-VB provides **ten DAC channels** that simultaneously support 7.1-channel playback, plus 2 channels of independent stereo output (multiple streaming) through the front panel stereo output with up to 120dB Signal-to-Noise (SNR).

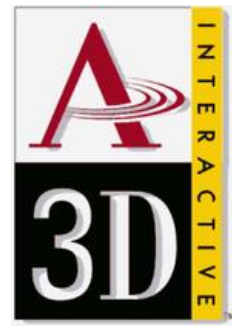
Three stereo ADCs are integrated and can support multiple analog audio inputs, including a 110dB SNR stereo line level input and microphone array with software features **Acoustic Echo Cancellation (AEC)**, **Beam Forming (BF)**, **Noise Suppression (NS)** and **Far Field voice Pick up (FFP)** technologies.





Aureal 3D环绕声(A3D Surround)

- ◆ A3D (Aureal 3-Dimensional)技术由 Aureal 半导体公司开发，用在其PC声卡芯片中。能够在两个甚至四个说话者之间通过耳机传输三维声音
- ◆ A3D Surround吸收了A3D技术和环绕声解码技术(如Dolby的ProLogic和 AC-3)，**创建一个围绕听者的5组音频流的声场**，即产生五个“虚拟音箱”，它实际上是经过A3D Surround 处理后用两个音箱播放出来的。也就是说这5组音频流并不像传统的“家庭影院”那样需要用5个实际的音箱进行回放





3D声还原系统

SRS(Sound Retrieval System)

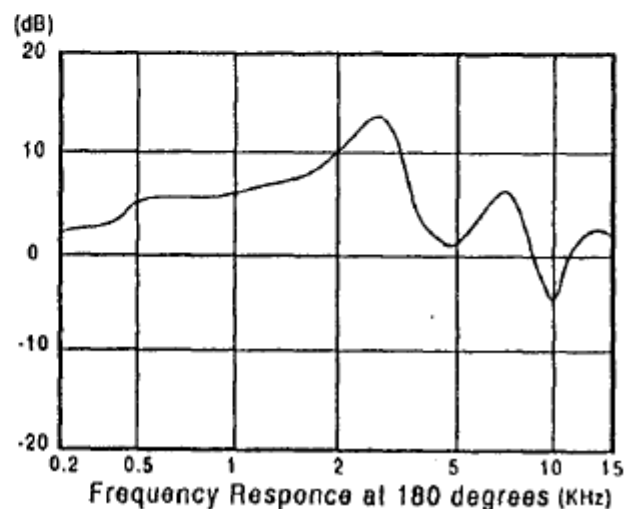
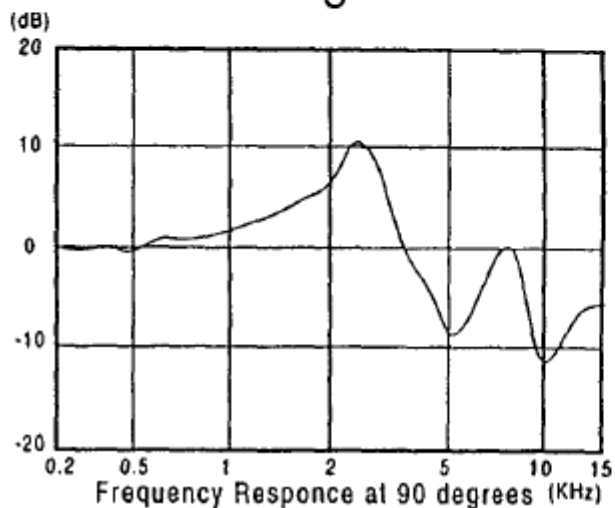
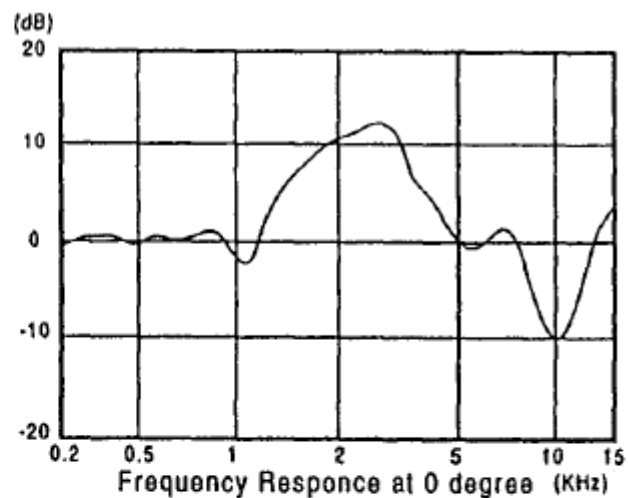
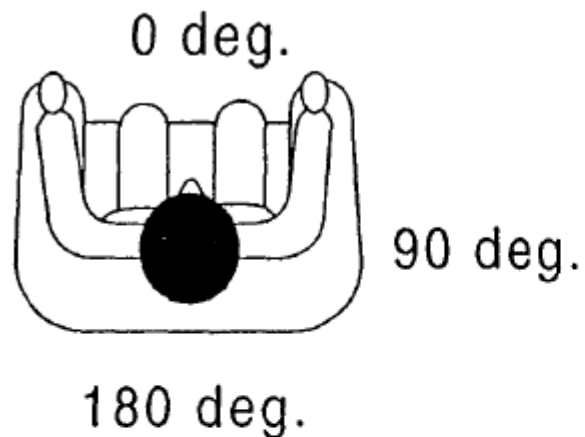
◆利用仿声学原理，根据人耳对各空间方向声音信号函数的反映不同，对双声道立体声信号中的反射声、回声等信号提取出后进行技术处理。尽管这些信号仍来自前方，但给人的错觉是来自四面八方。其过人之处是只使用两只普通音箱，产生出仿3D环绕声五声道的放音效果，有如音乐厅的身临其境感觉。Vivid 3D Pro就是SRS技术的典型应用。

□ SRS发明人是阿诺德·克雷曼（Arnold Klayman）是杰出的科学研究者，发明家，也是一个音乐爱好者。在从事心理声学 and 音频电子研究时，发现了人类听觉系统对三维空间的声源的定位机制，从而产生了SRS技术的雏形。





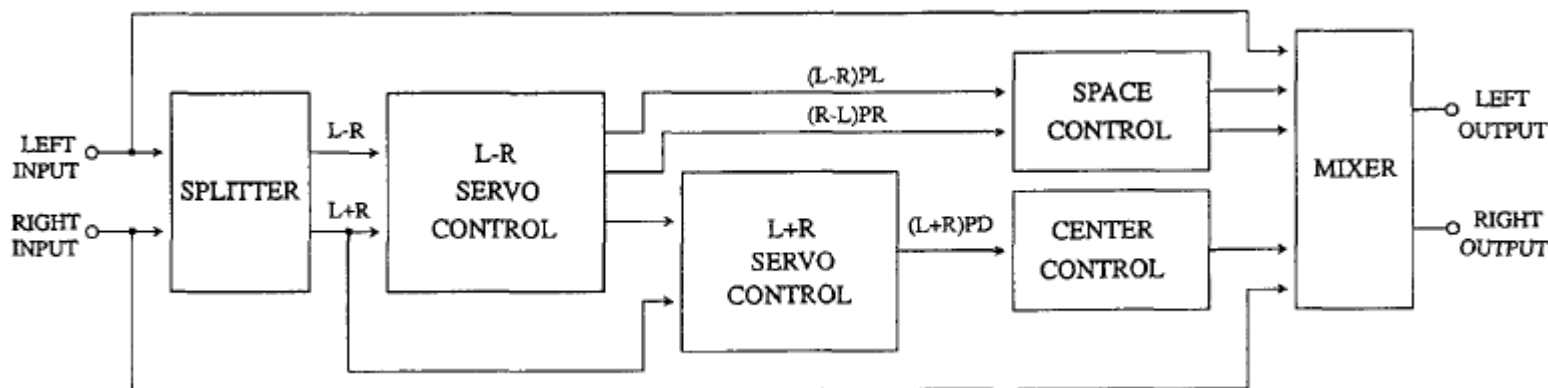
外耳对声波的作用





SRS如何用双声道实现3D

A **head-related transfer function (HRTF)** is a response that characterizes how an ear receives a sound from a point in space; a pair of HRTFs for two ears can be used to synthesize a binaural sound that seems to come from a particular point in space. **It is a transfer function**, describing how a sound from a specific point will arrive at the ear (generally at the outer end of the auditory canal).





小结：计算机的音频处理

◆ 声卡（独立/集成）的作用

- 模拟声音的输入，ADC
- 数字化的处理：滤波、特效.....
- 模拟声音的输出，DAC

◆ 音频相关的处理

- Digital Highpass Elliptical Filters with Notch for 217Hz (GSM)
- Acoustic Echo Cancellation
- Beam Forming (BF)
- Noise Suppression (NS)
- Far Field voice Pick up (FFP)
-

◆ 单声道/立体声/多声道

- 2个声道实现环绕声



◆ § 2.1 数字音频处理技术

□ § 2.1.1 数字音频在计算机上的应用

- § 2.1.1.1 声音文件
- § 2.1.1.2 电脑音乐
- § 2.1.1.3 计算机上的音频处理

□ § 2.1.2 用比特串记录声音信息

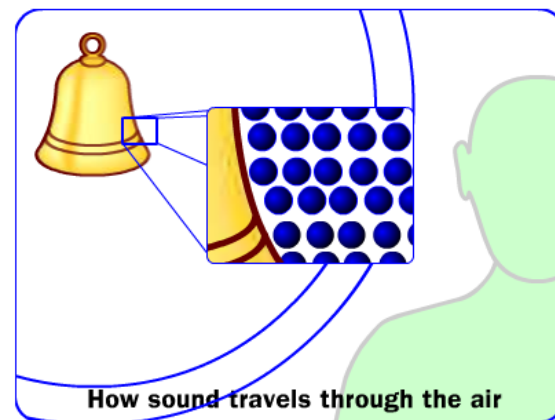
- § 2.1.2.1 声音的基本特性
- § 2.1.2.2 声音信号的波形的数字记录
- § 2.1.2.3 针对声音信号“知识”的数字记录

◆ § 2.2 数字图像处理技术

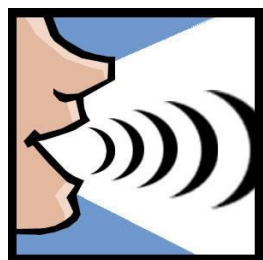
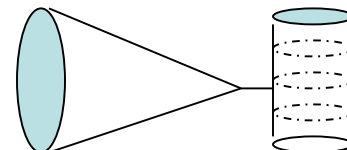
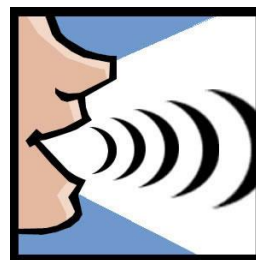
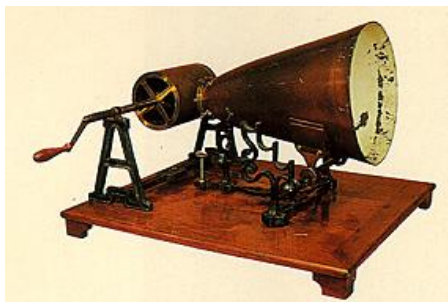
◆ § 2.3 数码相机技术

◆ § 2.4 数字视频处理技术

- ◆ 声音是通过空气传播的一种**连续**的波，叫声波
- ◆ 音频信号是**时间依赖**的连续媒体
- ◆ 模拟人耳的两个通道
- ◆ **语音**处理包含情感和意向，涉及语言学、社会学和声学

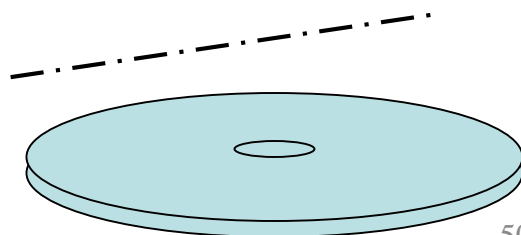


Edison, 1877



A/D Converter

PCM



Philips, 1978



声音信号的属性-频率

- ◆一般来说，人的听觉器官能感知的声音频率大约在20~20000 Hz之间，不同的人耳存在差异
- ◆**亚音信号**：频率小于20 Hz的信号，或称为次音信号(subsonic)
- ◆**音频(Audio)信号**：频率范围为20 Hz~20 kHz的信号
- ◆**超音频信号**：频率高于20 kHz的信号，或称超声波(ultrasonic)信号

人的发音器官发出的声音频率大约是80~3400 Hz

人说话的信号频率通常为300~3000 Hz，这种频率范围的信号称为**话音(speech)信号**





关于亚音信号 人体的固有频率

- ◆由科学测试知道人体各部位有不同的固有频率，如眼球的固有频率最大约为60赫兹，颅骨的固有频率最大约为200赫兹等；把人体作为一个整体来看，如水平方向的固有频率约为3—6赫兹，竖直方向的固有频率约为48赫兹。
- ◆有关部门规定，要求用手工操作的各类振动机械的频率必须大于20赫兹。
- ◆次声波枪和次声波炸弹利用频率为16—17赫兹的次声波，与人体内的某些器官发生共振，使受振者的器官发生变形、位移或出血，从而达到杀伤敌方的目的。

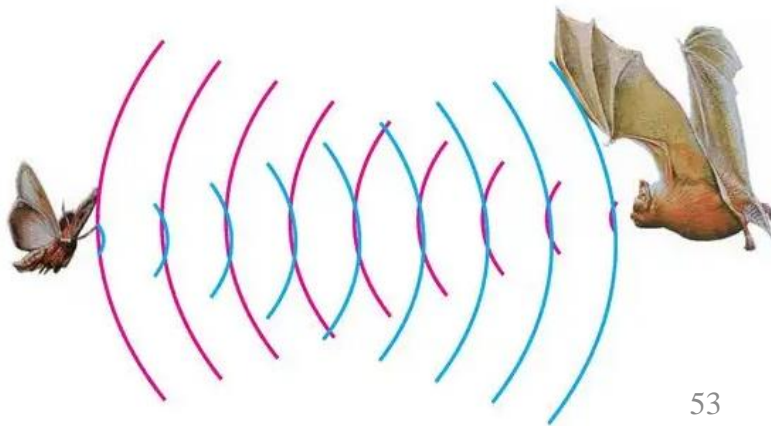




关于超音频信号 相关应用



B超



超声波测距电路模块



蝙蝠靠超声波探测

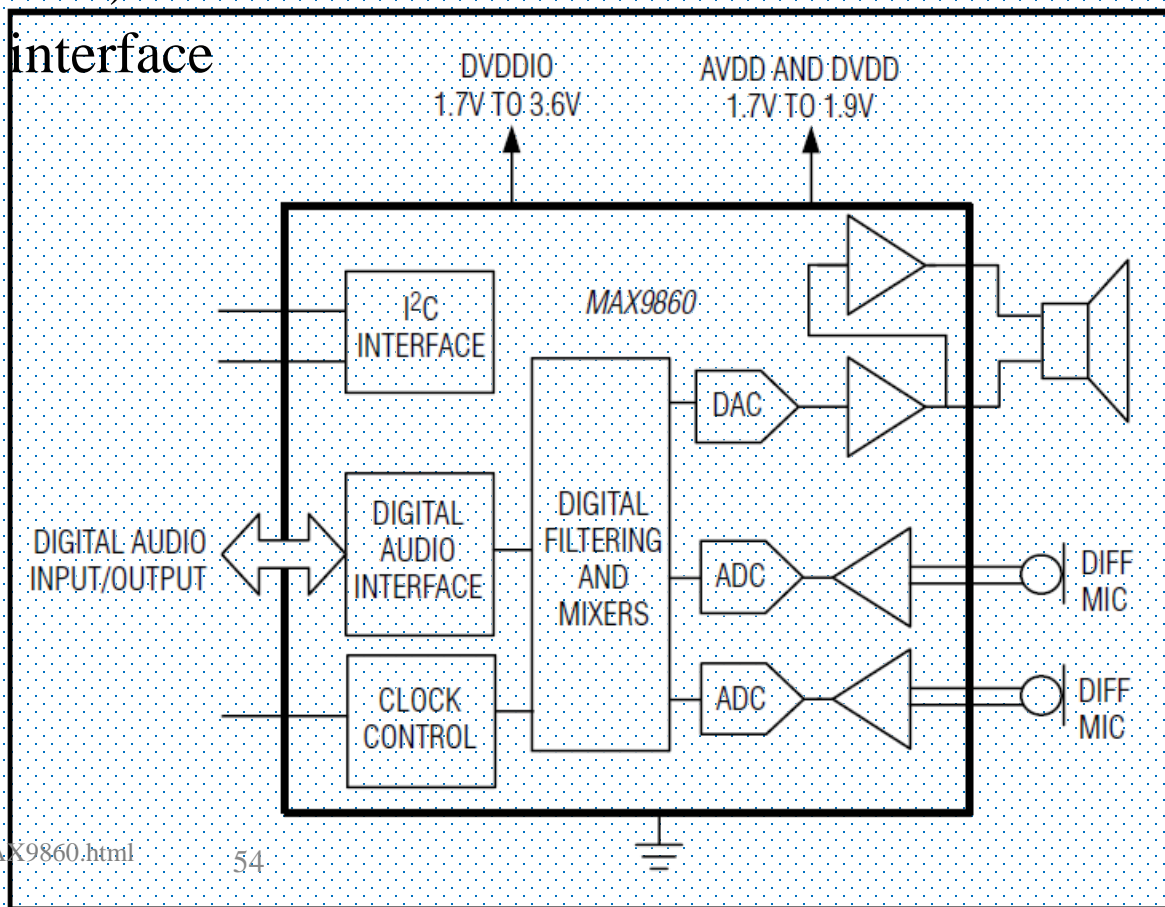


单声道音频编解码器 MAX9860

16-Bit Mono Audio Voice Codec

- ◆ Digital Highpass Elliptical Filters with Notch for 217Hz (GSM)
- ◆ 90dB DAC DR ($f_S = 48\text{kHz}$) *Dynamic Range (DR)*
- ◆ 81dB ADC DR ($f_S = 48\text{kHz}$)
- ◆ 2-wire, I2C-compatible interface

MAX9860是一种低功耗，语音频带，单声道音频编解码器。旨在提供完整的音频解决方案，适用于无线语音耳机和其他单声道语音音频设备。





MAX9860电气特性

数字低通滤波器

<https://www.maximintegrated.com/en/products/analog/audio/MAX9860.html>
<https://datasheets.maximintegrated.com/en/ds/MAX9860.pdf>

◆ Passband Cutoff

◆ Stopband Cutoff

◆ Passband Ripple

◆ Stopband Attenuation

PARAMETER	SYMBOL	CONDITIONS	MIN	TYP	MAX	UNITS
DAC LOWPASS DIGITAL FILTER						
Passband Cutoff	f_{PLP}	With respect to f_S within ripple; $f_S = 8\text{kHz}$ to 48kHz		$0.448 \times f_S$		Hz
		-3dB cutoff		0.451		f_S
Passband Ripple		$f < f_{PLP}$		± 0.1		dB
Stopband Cutoff	f_{SLP}	With respect to f_S ; $f_S = 8\text{kHz}$ to 48kHz		$0.476 \times f_S$		Hz
Stopband Attenuation		$f > f_{SLP}$, $f = 20\text{Hz}$ to 20kHz		75		dB
ADC LOWPASS DIGITAL FILTER						
Passband Cutoff	f_{PLP}	With respect to f_S within ripple; $f_S = 8\text{kHz}$ to 48kHz		$0.445 \times f_S$		Hz
		-3dB cutoff		0.449		f_S
Passband Ripple		$f < f_{PLP}$		± 0.1		dB
Stopband Cutoff	f_{SLP}	With respect to f_S ; $f_S = 8\text{kHz}$ to 48kHz		$0.469 \times f_S$		Hz
Stopband Attenuation		$f > f_{SLP}$		74		dB



MAX9860电气特性

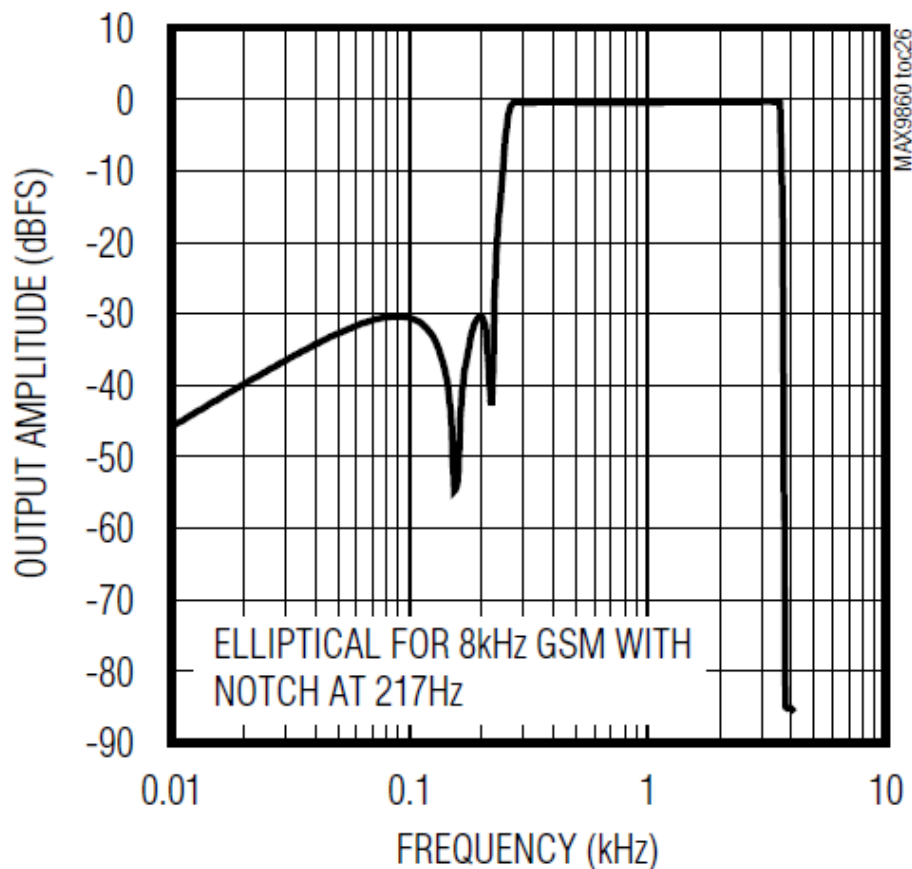
数字高通滤波器

<https://www.maximintegrated.com/en/products/analog/audio/MAX9860.html>

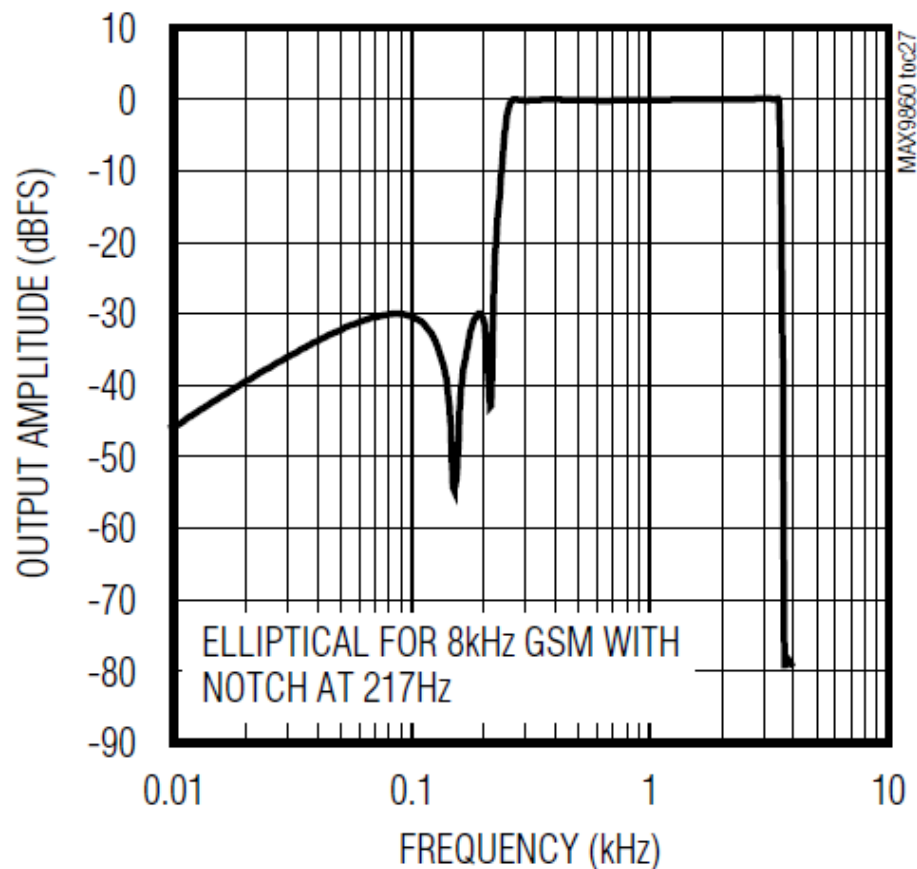
<https://datasheets.maximintegrated.com/en/ds/MAX9860.pdf>

◆ Digital Highpass Elliptical Filters with Notch for 217Hz (GSM)

DAC DIGITAL FILTER
FREQUENCY RESPONSE, 8kHz



ADC DIGITAL FILTER
FREQUENCY RESPONSE, 8kHz





声音信号的属性-带宽

◆对声音信号的分析表明，声音信号由许多频率不同的信号组成，这类信号称为复合信号，而单一频率的信号称为分量信号。带宽是声音信号的一个重要参数，它用来描述组成复合信号的频率范围。如高保真音信号(high-fidelity audio)的频率范围为10 Hz~20000 Hz，它的带宽约为20 kHz（20000-10）。





声音信号的属性-响度

- ◆ 声音的响度就是声音的强弱。
- ◆ **听阈**：当声音弱到人的耳朵刚刚可以听见时，我们称此时的声音强度为“听阈”。
- ◆ **痛阈**：声音强到使人耳感到疼痛时，这个阈值称为“痛阈”。
- ◆ 注意：听阈和痛阈都是和频率相关的。



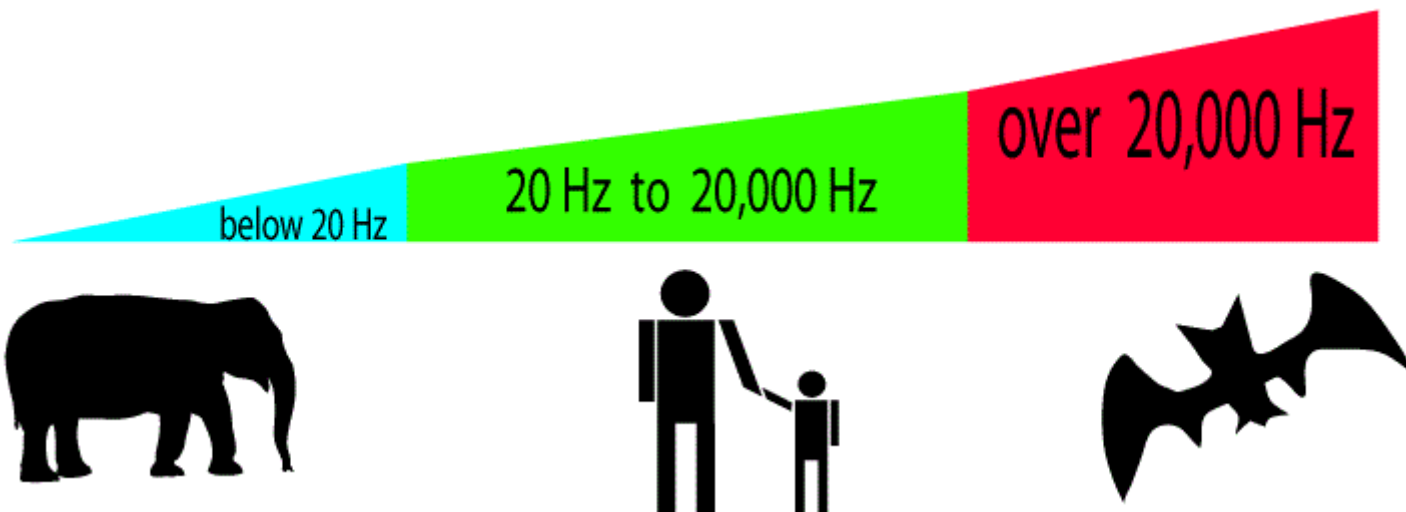


小结：声音信号的属性

◆ 各种对声音的处理方法都是基于声音的基本特性，学习对音频信号进行处理就必须熟悉这些基本特性

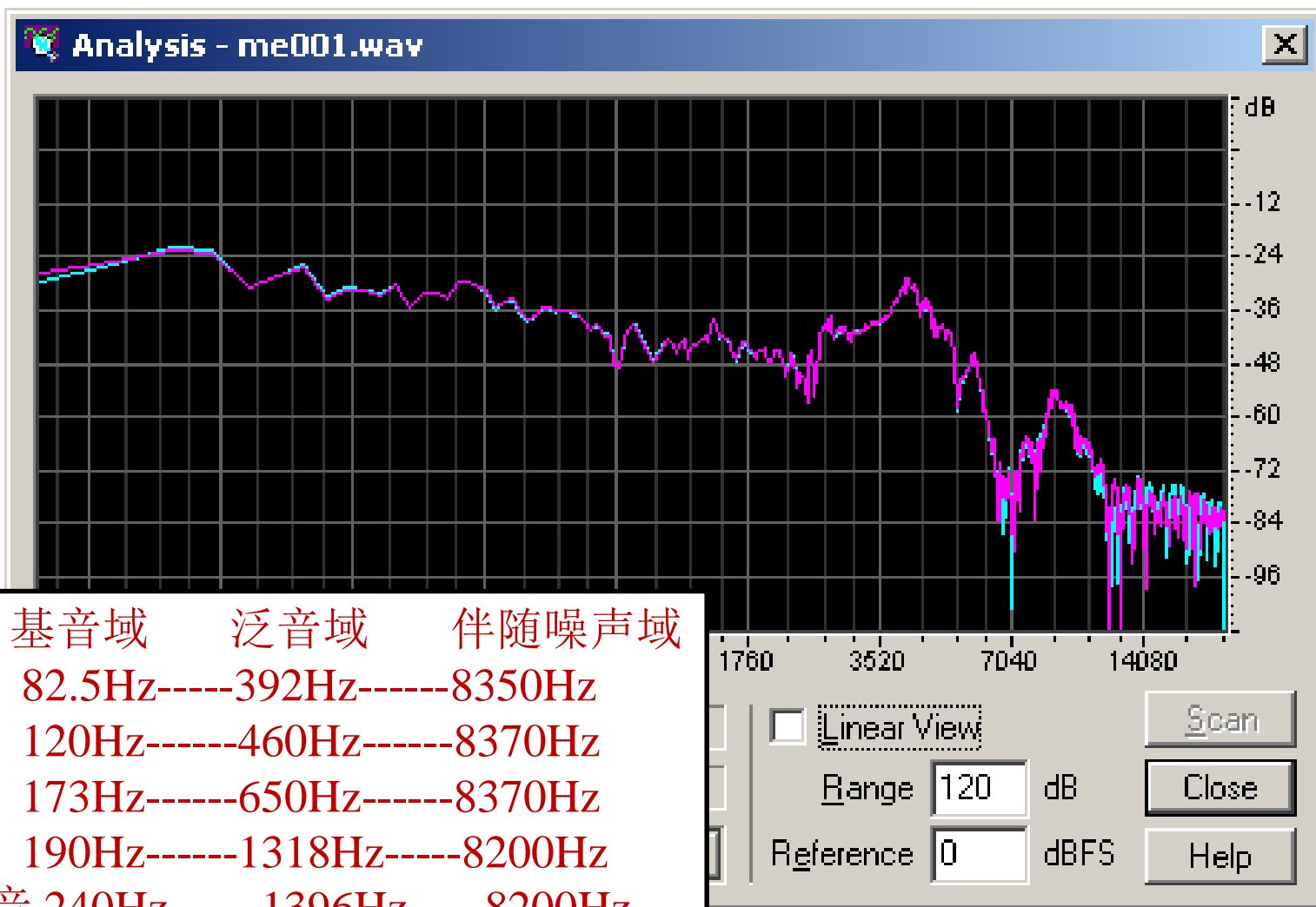
□ 频率、带宽、响度

◆ **频率**特性是非常重要的属性，很多对音频信号的处理都和频率相关，如电路中的滤波器





话音频谱示例



基音域 泛音域 伴随噪声域

男低音 82.5Hz-----392Hz-----8350Hz

男中音 120Hz-----460Hz-----8370Hz

男高音 173Hz-----650Hz-----8370Hz

女低音 190Hz-----1318Hz-----8200Hz

次女高音 240Hz-----1396Hz-----8200Hz

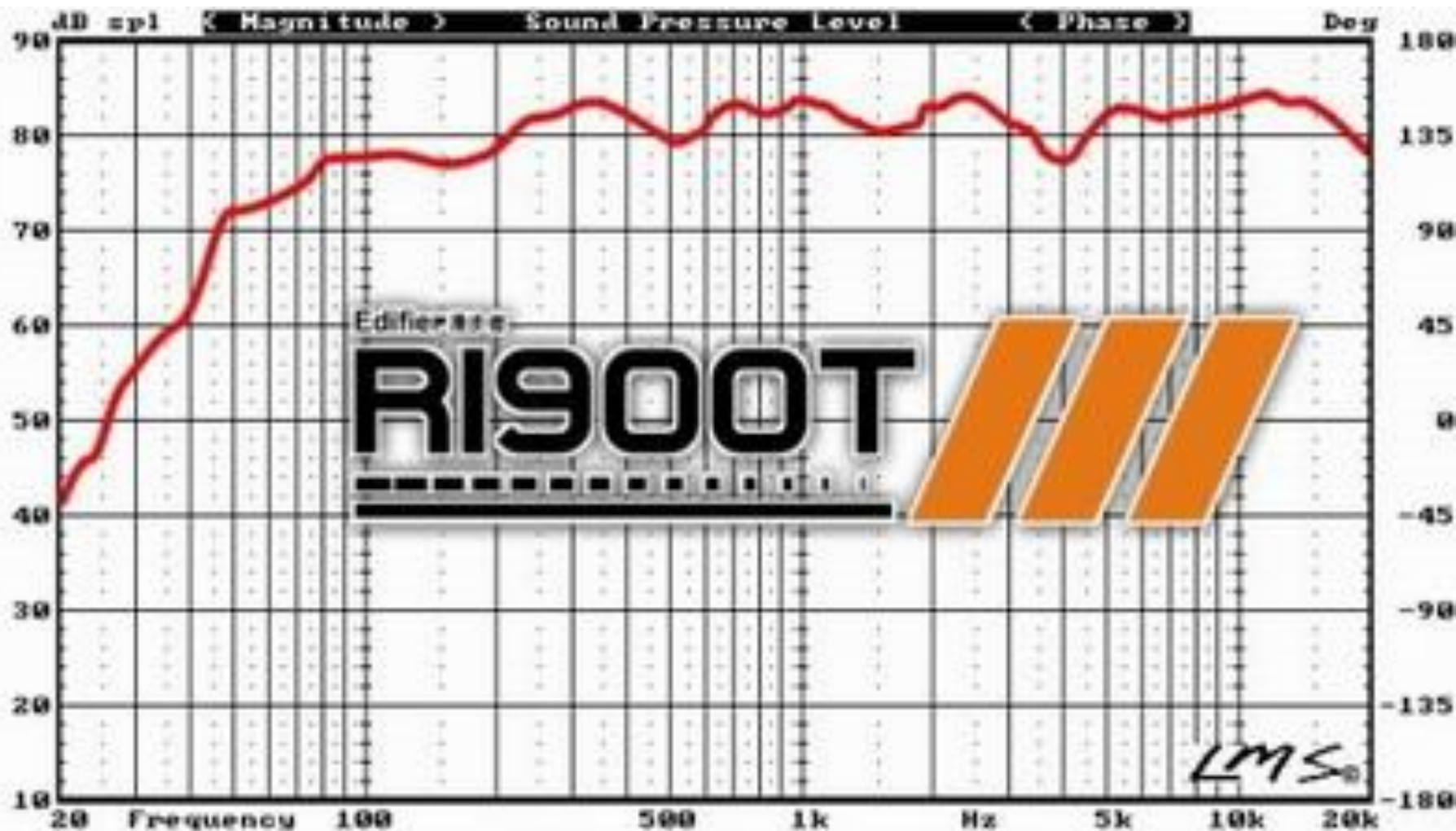
女高音 260Hz-----1590Hz-----8820Hz

{ynh,cxh}@ustc.edu.cn





用频率响应曲线表示音响特性





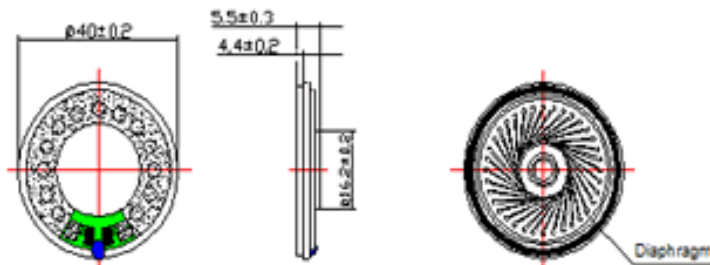
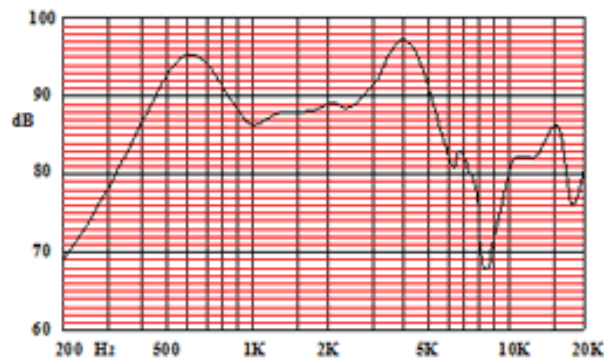
音箱扬声器的频率响应

Mylar Speaker



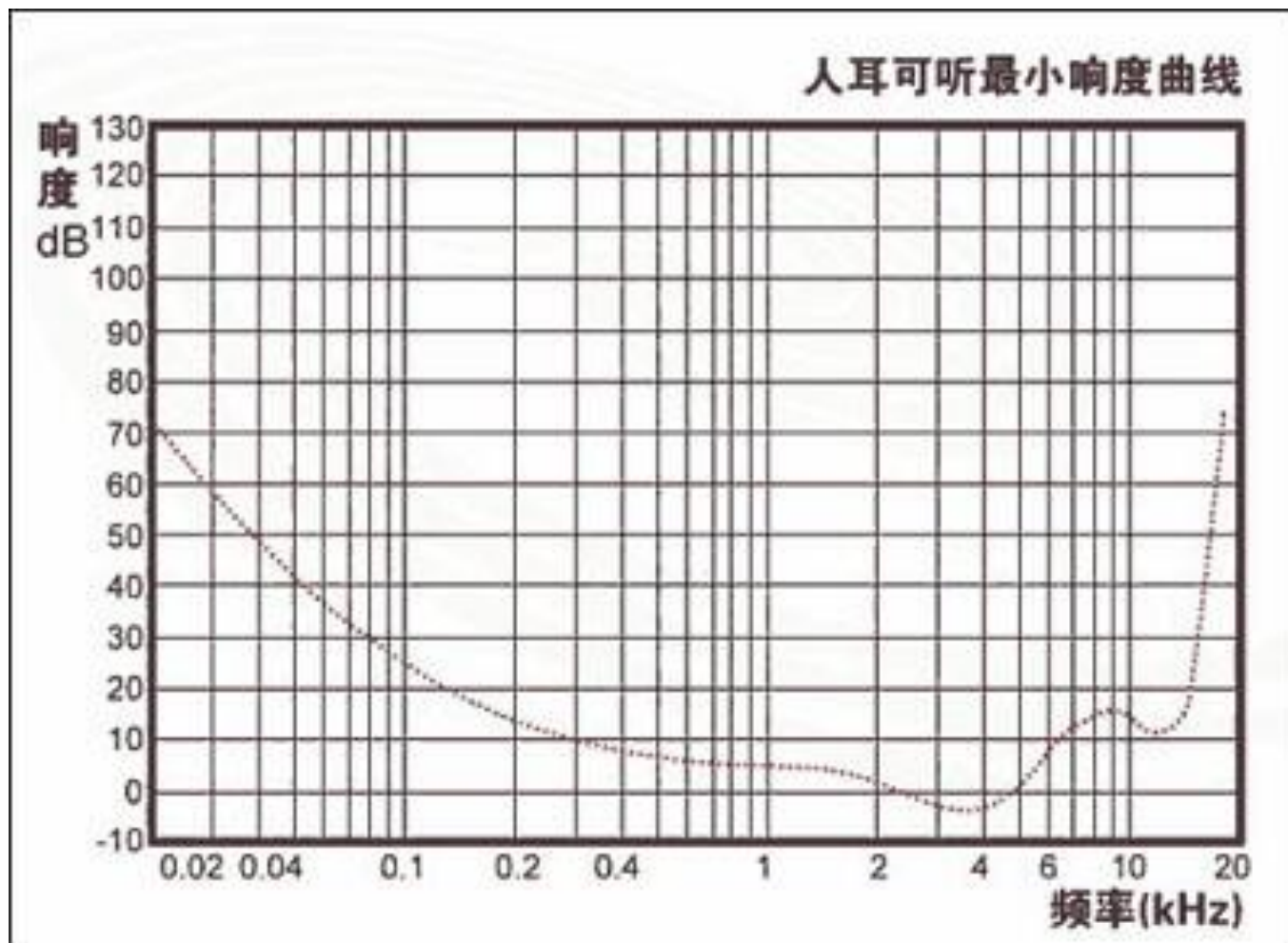
TE-DM40A-8H5.5

Dimensions	40x5.5mm
AC Impedance	$8 \pm 15\% \Omega$ at 2000Hz
Rated Input Power	1.0W
Max. input power	1.2W
Resonance Frequency	$600 \pm 20\% \text{Hz}$
Output Sound Pressure Level	$97 \pm 3 \text{dB}/0.1\text{M } 0.1\text{W}$ at 0.8, 1.0, 1.2, 1.5KHz Average
Frequency Response	Fo~6KHz
Operating Temperature	-20~+65°C
Storage Temperature	-30~+70°C





子系统的频率响应（人耳）





理解“频带”概念（举例）

◆全频带音箱，低音音箱和超低音音箱。

- 所谓全频带音箱是指能覆盖低频，中频和高频范围放音的音响。全频带音箱的下限频率一般为30Hz-60Hz，上限频率为15KHz-20KHz。低音音箱和超低音音箱一般是用来补充全频带音箱的低频和超低频放音的专用音箱。

◆杜比数码环绕声系统（Dolby Stereo digital）

- 该系统称作AC-3，其AC是指Audio Coding。该系统设置互相独立的6个声道，它们是全频带的左、右、中置、左环绕、右环绕，再加上1个120HZ以下的超低音声道，因而又称作5.1声道。

音响系统可以等效为一个通信系统（带通）





第二章 媒体处理技术

◆ § 2.1 数字音频处理技术

□ § 2.1.1 数字音频在计算机上的应用

- § 2.1.1.1 声音文件
- § 2.1.1.2 电脑音乐
- § 2.1.1.3 计算机上的音频处理

□ § 2.1.2 用比特串记录声音信息

- § 2.1.2.1 声音的基本特性
- § 2.1.2.2 声音信号的波形的数字记录
- § 2.1.2.3 针对声音信号“知识”的数字记录

◆ § 2.2 数字图像处理技术

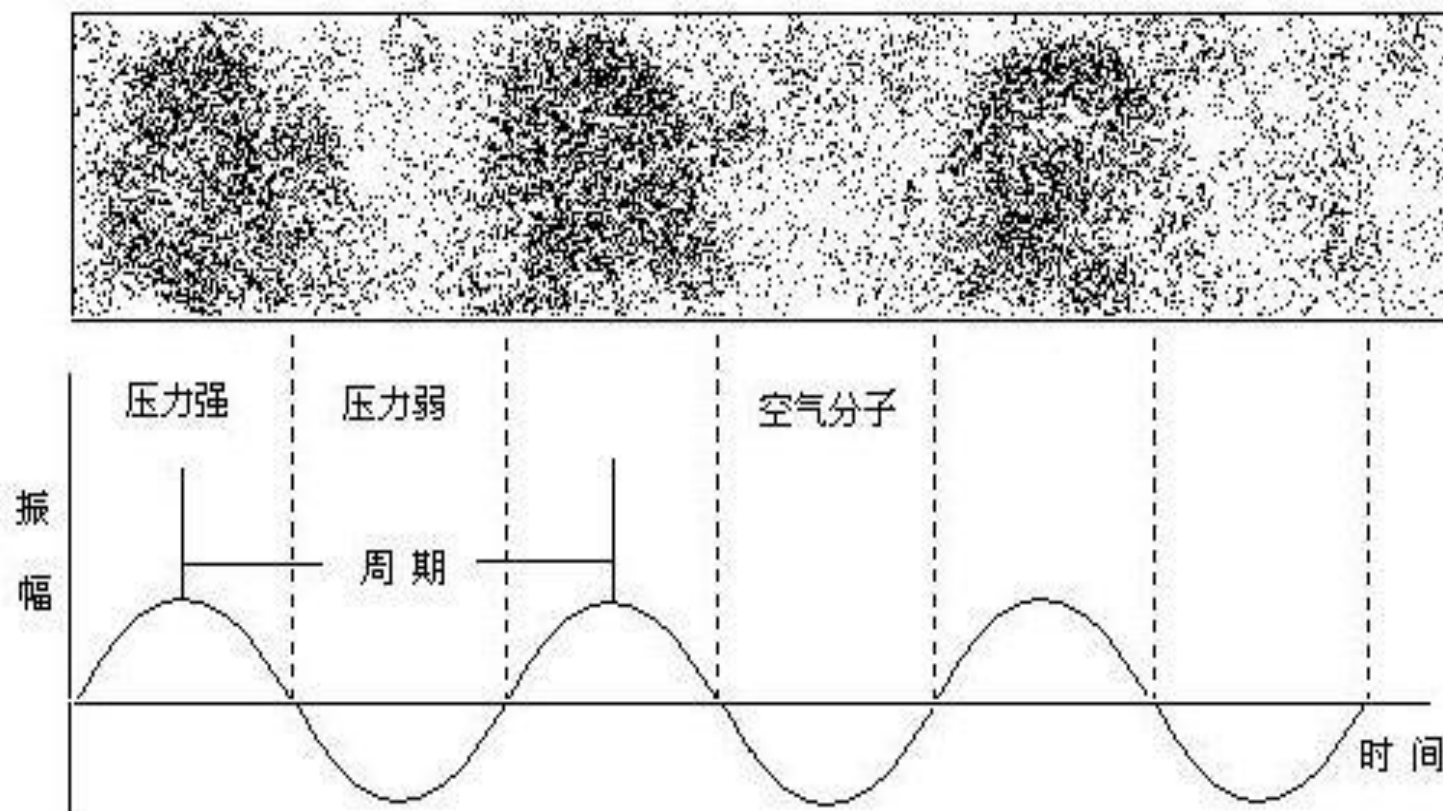
◆ § 2.3 数码相机技术

◆ § 2.4 数字视频处理技术



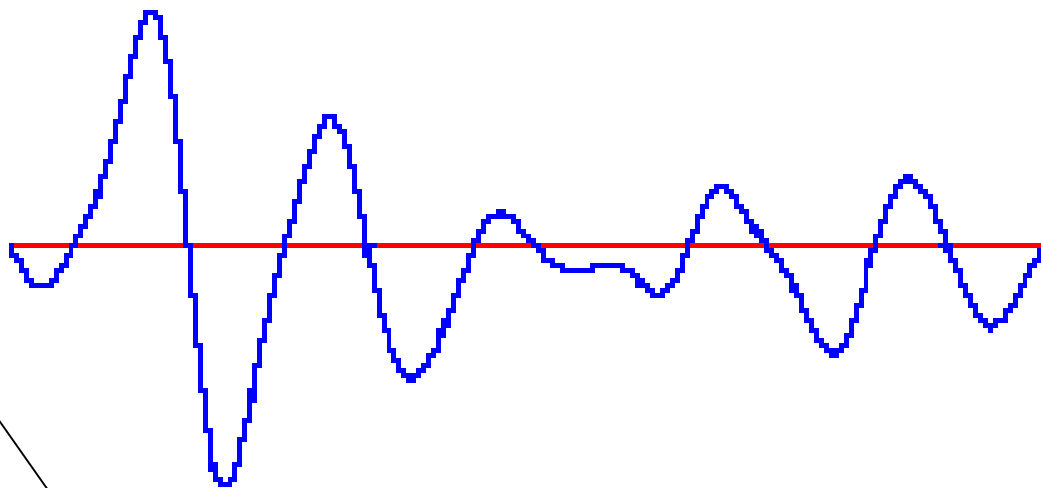


声波的变化特性





音频信号是模拟信号



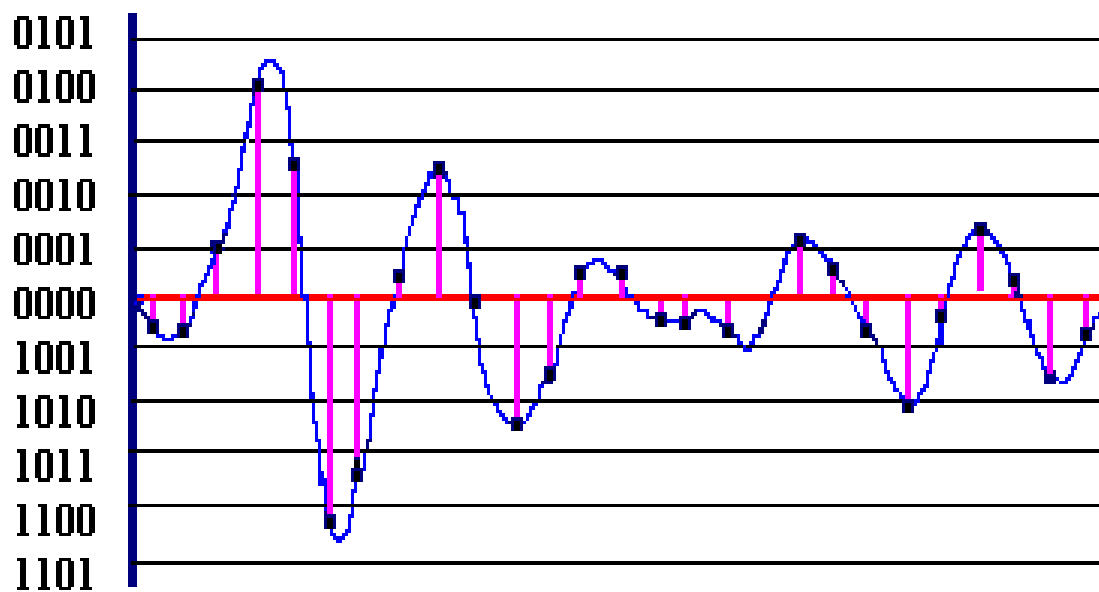
语音信号是典型的连续信号，不仅在时间上是连续的，而且在幅度上也是连续的。在时间上“连续”是指在一个指定的时间范围里声音信号的幅值有无穷多个，在幅度上“连续”是指幅度的数值有无穷多个。我们把在时间和幅度上都是连续的信号称为模拟信号。



模拟音频信号的采样和量化

◆**采样**：在某些特定的时刻对模拟信号测量叫做采样 (sampling)，得到一个离散的序列

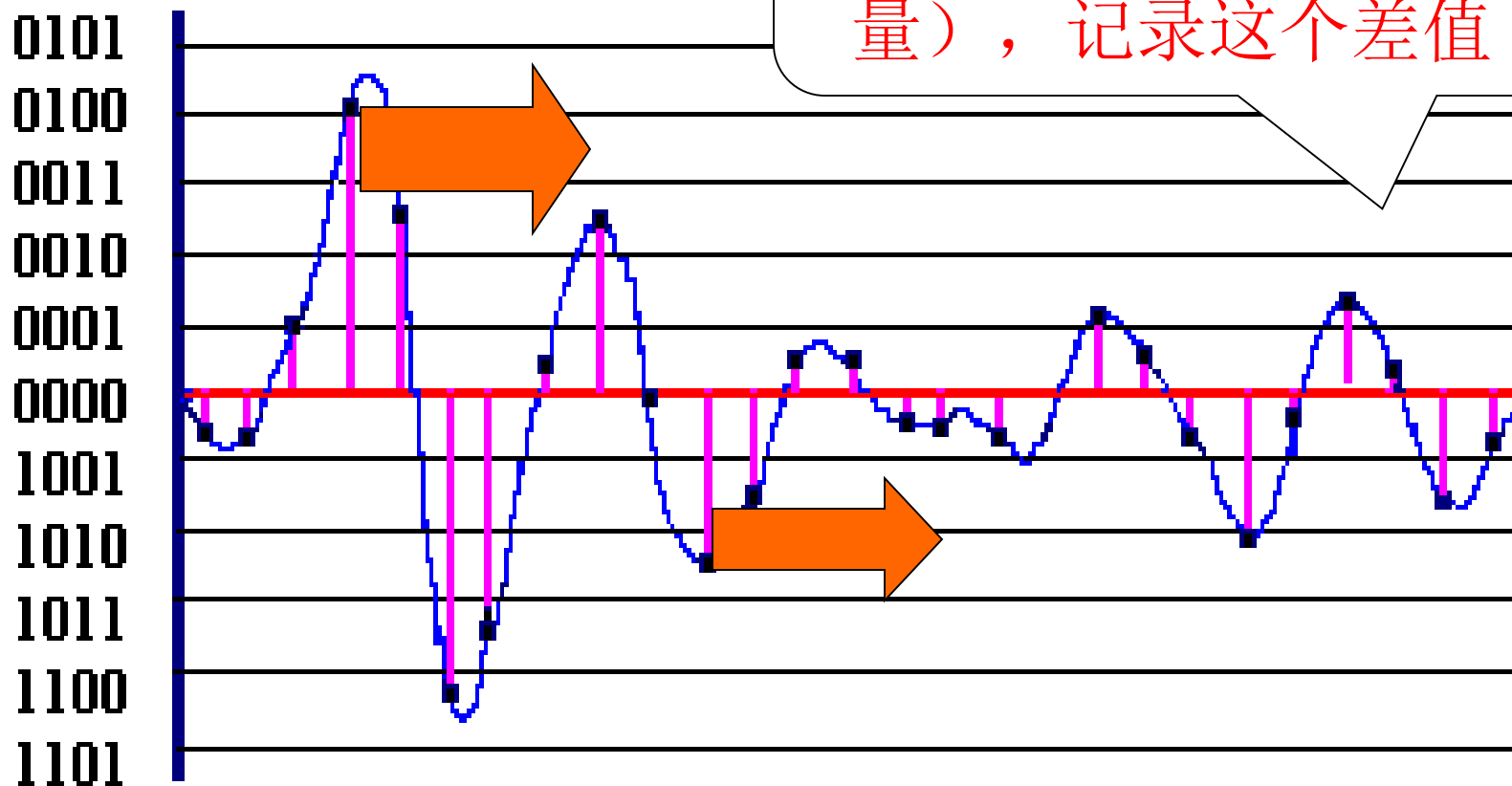
◆**量化**：把信号幅度取值的数目加以限定，形成的离散信号的幅度只具备有限个数的数值





增量

箭头的宽度就是相邻两个点之间的差值（增量），记录这个差值





采样频率、样本精度

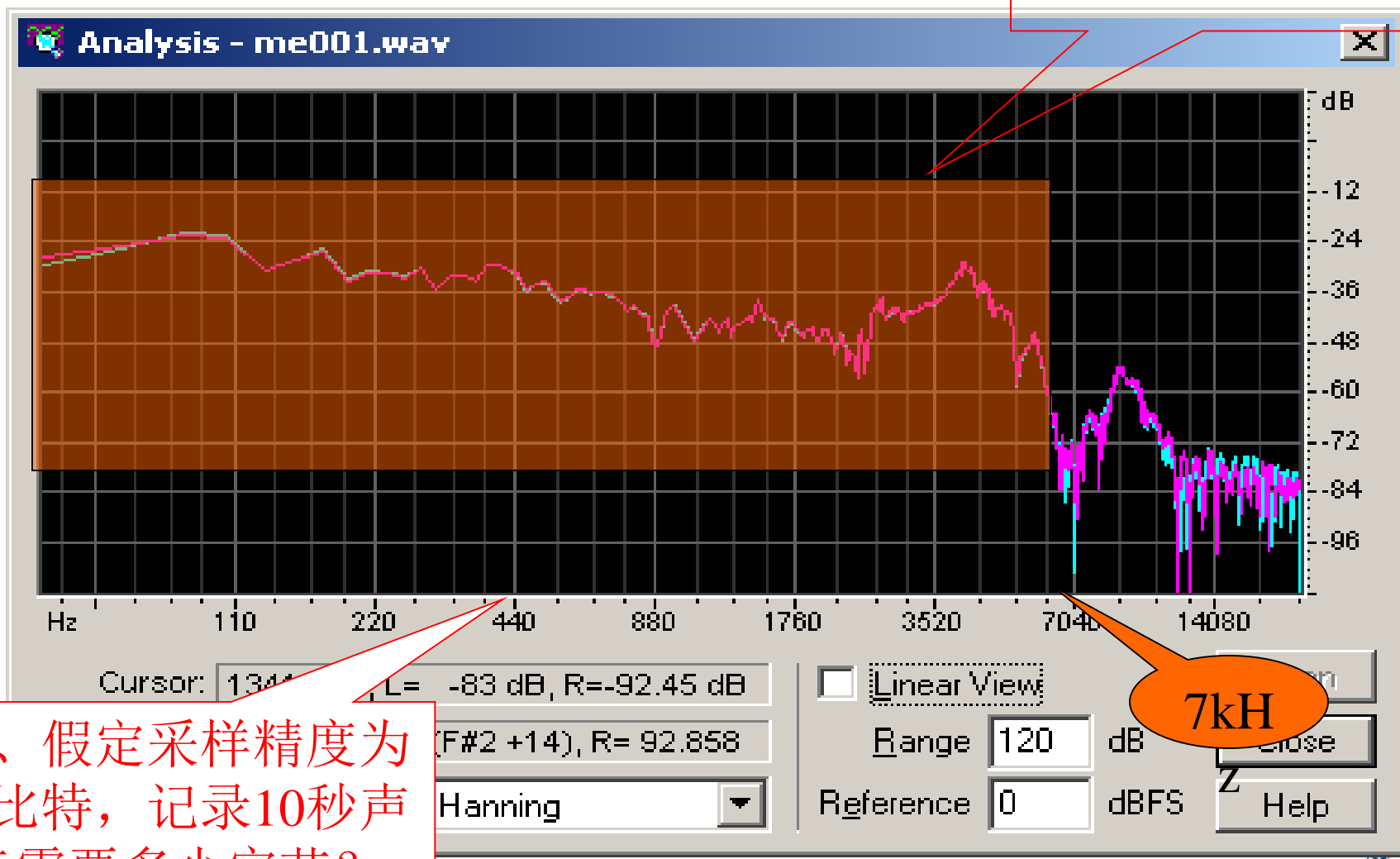
- ◆ 声音数字化需要回答两个问题：
- ◆ ①每秒钟需要采集多少个声音样本，也就是采样频率(fs)是多少
- ◆ ②每个声音样本的位数(bit per sample, bps)应该是多少，也就是样本精度





采样频率&样本精度

1、为保证采样不造成失真，采样频率最小应该取多少？



2、假定采样精度为8比特，记录10秒声音需要多少字节？





奈奎斯特采样定律

◆ 采样频率不应低于声音信号最高频率的两倍

◆ 可以这样来理解奈奎斯特理论：声音信号可以看成由许许多多正弦波组成的，一个振幅为 A 、频率为 f 的正弦波至少需要两个采样样本表示，因此，如果一个信号中的最高频率为 F_m ，采样频率最低要选择 $2 * F_m$ 。例如，电话语音的信号频率约为3.4 kHz，采样频率就选为8 kHz。





声音质量与数据率

质量	采样频率(kHz)	样本精度(bit/s)	单声道/立体声	数据率(kB/s)	频率范围
电话	8	8	单声道	8	200~3400 Hz
AM	11.025	8	单声道	11.0	20~15000Hz
FM	22.050	16	立体声	88.2	50~7000Hz
CD	44.1	16	立体声	176.4	20~20000 Hz
DAT	48	16	立体声	192.0	20~20000 Hz
HDCD	96	24	立体声	384	20~20000 Hz

注：* 电话使用 μ 律编码，动态范围为13比特，而不是8比特





数字音频文件大小计算示例

- ◆ 数据量 Byte=采样频率Hz
- ◆ $\times (\text{采样位数}/8)$
- ◆ $\times \text{声道数}$
- ◆ $\times \text{时间s}$
- ◆ 如果采样频率为44.1kHz，分辨率为16位，立体声，录音时间为10s，符合CD音质的声音文件的大小是多少？
- ◆ $44100\text{Hz} \times (16/8) \times 2 \times 10\text{s} = 1764\text{KByte}$





SNR表示量化精度

$$\begin{aligned} SNR &= 10 \lg [(V_{\text{signal}})^2 / (V_{\text{noise}})^2] \\ &= 20 \lg (V_{\text{signal}} / V_{\text{noise}}) \end{aligned}$$

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 $V_{\text{noise}}=1$ ，量化精度为1比特表示 $V_{\text{signal}}=2^1$ ，
它的信噪比 $SNR=6$ 分贝。

例2：假设 $V_{\text{noise}}=1$ ，量化精度为16比特表示 $V_{\text{signal}}=2^{16}$
它的信噪比 $SNR=96$ 分贝。





关于采样频率的思考？

- ◆ Why speech sampling @8k
- ◆ Why CD sampling @44.1k
- ◆ Why HDCD sampling @96k





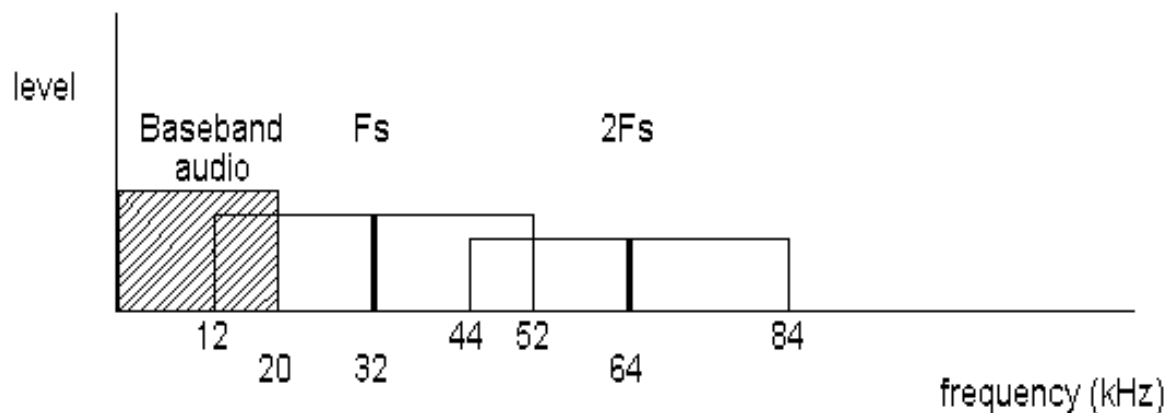
Why 44.1

◆ In 625 line 25 Hz (PAL) systems the line frequency is 15.625 kHz and 588 out of the 625 lines are "active" for carrying video information. If three samples are recorded per line the sample rate is

◆ $15.625 \times 588 / 625 \times 3 = 44.1$ kHz.

◆ In 525 line 30 Hz systems the line frequency is 15.75 kHz and 490 out of the 525 lines are "active" for carrying video information. If three samples are recorded per line the sample rate is

◆ $15.75 \times 490 / 525 \times 3 = 44.1$ kHz.

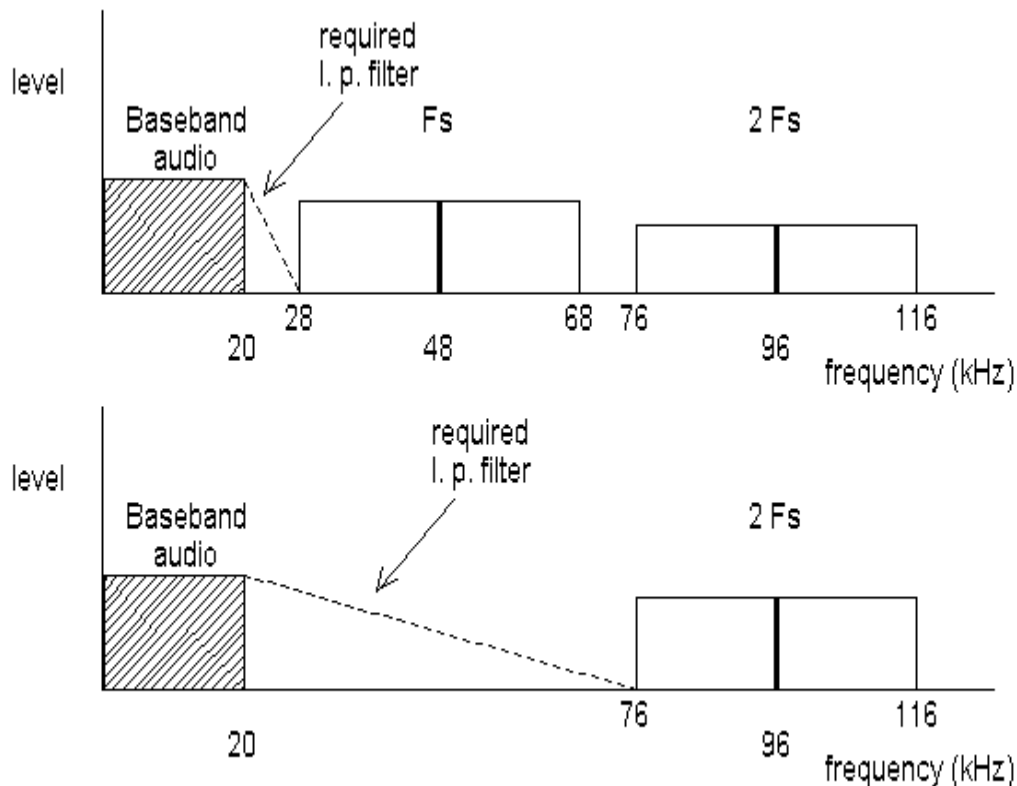




Why 48/96

◆ To prevent aliasing the audio band must pass through a filter with a very **rapid roll off** above 20 kHz. Such filters can introduce **irregularities to the frequency response below 20 kHz** and many designs can also create **undesirable phase shifts**.

◆ Using a sample frequency which is double the one finally required moves the side bands further away from the audio band and allows **simpler filters** to be used.





小结：声音信号波形的记录

- ◆研究声卡的工作原理（或者说是研究声音信号数字化的过程）
- ◆需要对声音信号**采样**和**量化**的过程有了解
- ◆掌握利用**采样频率**、**样本精度**计算数字音频文件的方法
- ◆对“**带宽**”的理解



◆ § 2.1 数字音频处理技术

□ § 2.1.1 数字音频在计算机上的应用

- § 2.1.1.1 声音文件
- § 2.1.1.2 电脑音乐
- § 2.1.1.3 计算机上的音频处理

□ § 2.1.2 用比特串记录声音信息

- § 2.1.2.1 声音的基本特性
- § 2.1.2.2 声音信号的波形的数字记录
- § 2.1.2.3 针对声音信号“知识”的数字记录

◆ § 2.2 数字图像处理技术

◆ § 2.3 数码相机技术

◆ § 2.4 数字视频处理技术



什么是知识？

- ◆ 知识是一些断言，是一些命题，这些命题是关于某些“材料”的（data），或者说，由某些材料可以分析出一定的命题，而这些命题是对那些材料的解释。
- ◆ 根据声音的基本特性，可以对已有的一段音乐进行总结归纳，得出某些抽象的知识（这段音乐的各种属性）



利用声音的“知识”



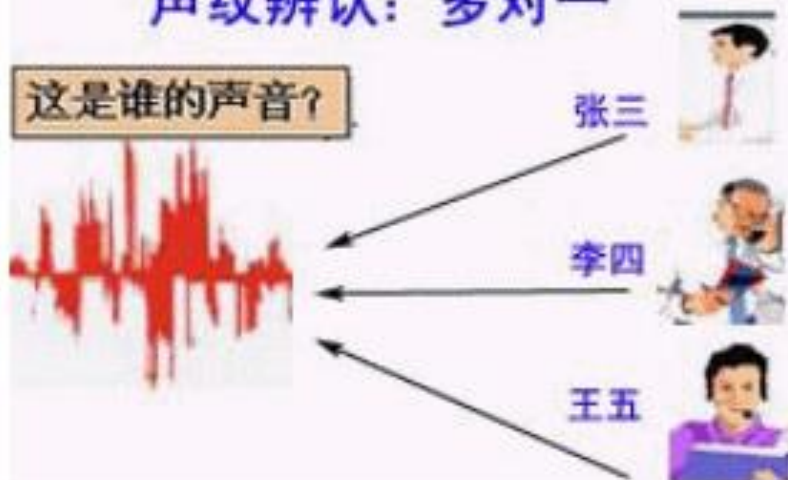
声纹确认：一对一

这是张三的声音吗?



声纹辨认：多对一

这是谁的声音?





分析 VS. 综合

- ◆ 音频信号的**分析**：得出有关音频信号的属性信息，这些信息可以表述、纪录、用于传输
- ◆ 音频信号的**综合**：根据已经提供的关于音频的具体特性、信息，人工合成音频并尽量恢复成和原来的一样





乐音三要素

◆ 一个乐音必备的三要素是：音高、音色和响度。若把一个乐音放在运动的旋律中，它还应具备时值——持续时间。

◆ 音高

□ 音高指声波的基频。基频越低，给人的感觉越低沉。知道了音高与频率的关系，我们就能够设法产生规定音高的单音了。

◆ 音色

□ 具有固定音高和相同谐波的乐音，有时给人的感觉仍有很大差异。音色是由声音的频谱决定的：各阶谐波的比例不同，随时间衰减的程度不同，音色就不同。

◆ 响度

□ 响度是对声音强度的衡量，它是听判乐音的基础。





傅立叶分解—谐振分析

◆任何一周期函数都可表示为简谐函数的合成。也就是说，任何一个复杂的周期振动都可以分解为一系列简谐振动之和。

$$\begin{aligned} F(t) &= A_0 + \sum_{k=1}^{\infty} A_k \cos k\omega_0 t + \sum_{k=1}^{\infty} B_k \sin k\omega_0 t \\ &= A_0 + \sum_{k=1}^{\infty} A_k \cos[(k\omega_0 t + \varphi_k)] \end{aligned}$$

称为周期函数 $F(t)$ 的傅里叶级数，
而 A_0, A_k, B_k 和 A_k, φ_k 称为傅里叶系数





基频和谐频

◆ 这些分振动中频率最低的称为基频振动，它的频率就是原周期函数的频率。称为基频

其它分振动的频率都是基频的整数倍，称为谐频。

◆ 拍频：同向进行之波列频率接近时，其干涉作用形成定时强弱的规律。它每秒间的拍数称为“拍频”，即为二波频率之差。

□ 1863年，赫尔姆霍茨《声音的感的理论》

□ 两个单音形成的拍频如果是每秒三十三次，则产生难以忍受的不谐和音；拍频小于或大于这个数时就稍好一些；当拍频超过每秒一百三十二次时，不愉快的感觉则完全消失





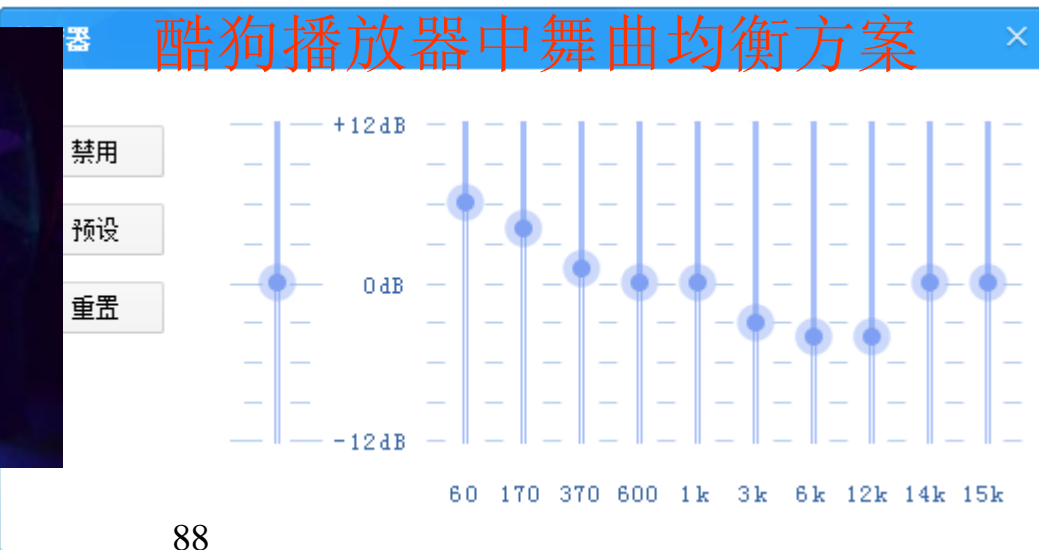
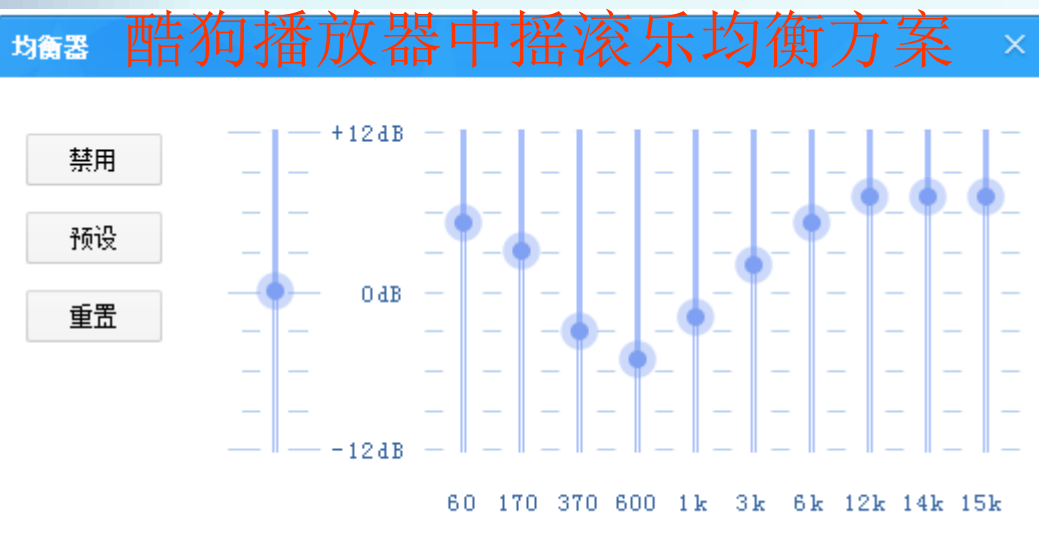
乐器的基音和泛音

- ◆小提琴基波频率约200Hz~2.5kHz，泛音>10kHz。
- ◆低音提琴基波频率约40hz~200Hz，泛音> 5kHz。
- ◆单簧管基音频率约160Hz~1.4kHz，泛音>15kHz。
- ◆短笛基音频率600hz~4kHz，泛音> 20kHz。
- ◆钢琴基音频率30Hz~4.2kHz，泛音> 20kHz。
- ◆乐器中基波频率低于80Hz的有钢琴、低音提琴、大提琴、竖琴等。而泛音频率超过15kHz的有钢琴、短笛、长笛、单簧管等。民族乐器中的二胡、笛子、锁呐、笙的泛音频率也是相当高的。





通过滤波改变声音效果





四段均衡器的频率特性

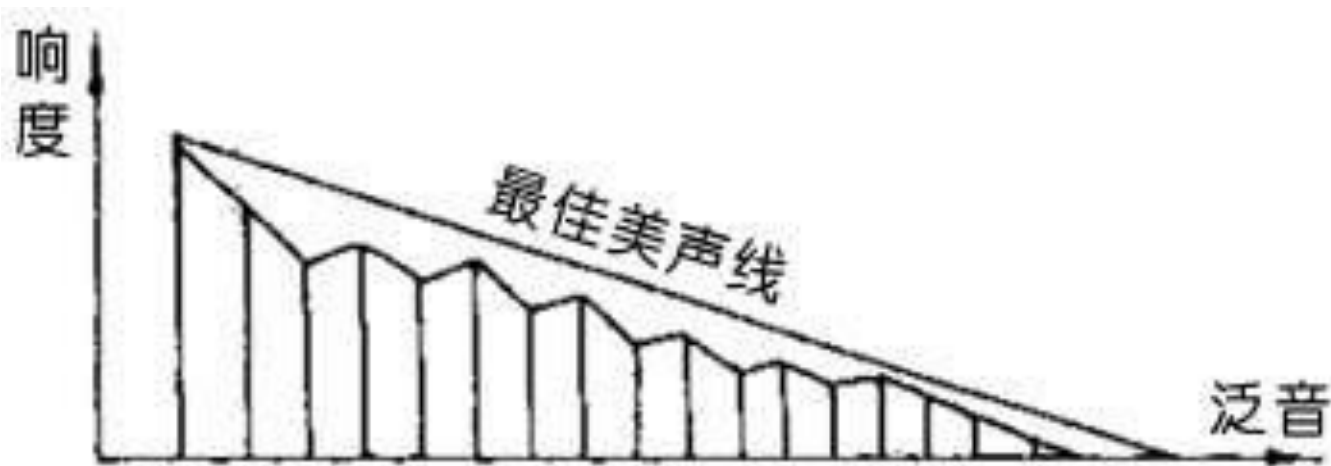
频段	人耳的听觉感受		
	过低	丰满	过高
6-20kHz	韵味失落	色彩鲜明富于表现力	尖噪、嘶哑刺耳
600Hz-6kHz	暗淡、朦胧	明亮、清晰	呆板
200-600Hz	空虚无力	圆润有力	生硬
20-200Hz	苍白单薄	丰满、混厚深沉	浑浊不清





最佳音色

◆根据意大利美声学的观点，就是将基音到第16个泛音的强度在坐标上连成一条直线，这条直线就被称为最佳美声线





用知识表达的音乐——MIDI

◆使用计算机利用声音合成技术生成音乐的技术称为**电子音乐合成**

◆**MIDI**是Musical Instrument Digital Interface的首写字母组词，可译成“**电子乐器数字接口**”。用于在音乐合成器(music synthesizers)、乐器(musical instruments)和计算机之间交换音乐信息的一种标准协议

◆1982年，国际**乐器制造者协会**的十几家厂商（其中主要是美国和日本的厂商）会聚一堂，会议通过了美国Sequential Circuits公司的大卫.史密斯提出的“通用合成器接口”的方案，即“Musical Instrument Digital Interface”，缩写为“MIDI”，公布于世





电子乐器数字接口（MIDI）

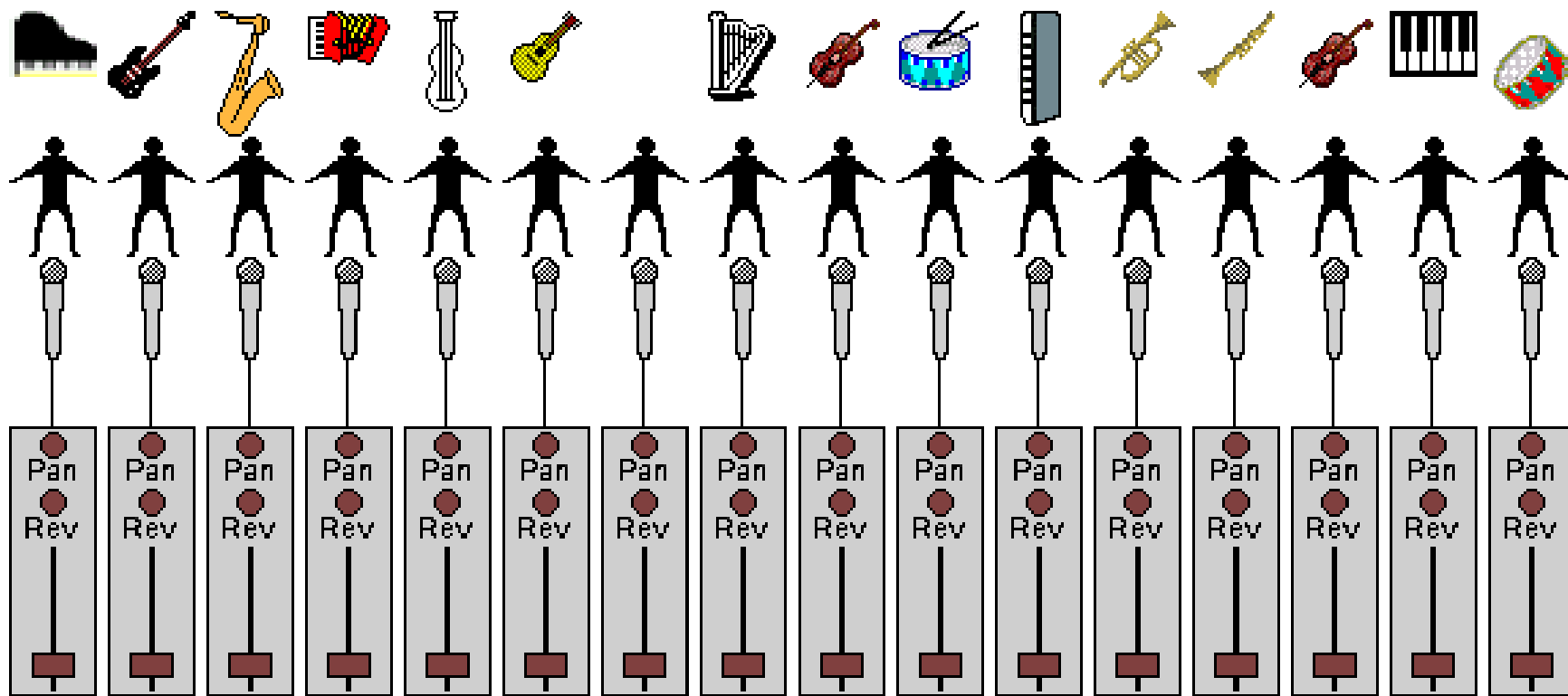
- ◆ 音乐合成器（music synthesizers）、乐器（musical instruments）、计算机三者交换音乐信息的**标准协议**
- ◆ MIDI是一套**命令集合**，指示计算机中的MIDI设备的所有动作（演奏乐符、加大音量、生成音响效果）
- ◆ MIDI是一个数据传输速率在31.23千波特的串行接口，可在同一时刻进行双向传输。连接采用MIDI 缆线（Cable）





MIDI的通道概念

◆ 单个物理MIDI通道(MIDI channel)分成16个逻辑通道，每个逻辑通道可指定一种乐器





合成器 (synthesizer)

利用数字信号处理器或其他芯片来产生音乐或声音的电子装置。

合成器 名称	旋律 乐器声		打击 乐器声	
基本 合成器	3种音色	6个音符	3种音色	3个音符
扩展 合成器	9种音色	16个音符	8种音色	16个音符

多音色 multi-timbre

多音调 multi-notes/polyphony





曾经的和弦手机

◆ Question

□ 什么是和弦铃声？

◆ Yamaha(2000年)

- YMU759 offers a standard library of **256 instruments**, including the 128 instruments specified by the GM standard, as well as an unlimited range of voices.
- YMU759 offers **16-note polyphony**- a feature used either to create **16-note harmonies** or notes with **16 different timbres**.





科大交响乐团设备

OASYS



Korg OASYS

Manufacturer	Korg
Dates	2005–2009
Price	\$8,000–\$8,500

Technical specifications

Polyphony	172 for HD-1, CX-3, & PolysixEX; 96 for AL-1; 52 for MOD-7; 48 for STR-1 & MS-20EX (all maximum; may vary depending on other sounding voices and effects)
Timbrality	16
Oscillator	depends on synthesis engine
Synthesis type	PCM, Wave Sequencing, Vector, Physical Modelling, Sampler, Analog Modelling

Korg Inc. 株式会社コルグ

KORG

Formerly	Keio Electronic Laboratories
Industry	Electronics, Musical Instruments
Founded	1962; 56 years ago (as Keio Electronic Laboratories)
Founder	Tsutomu Katoh Tadashi Osanai
Headquarters	Inagi, Tokyo, Japan
Products	Keyboards, electronic organ
Subsidiaries	Vox
Website	www.korg.com

Korg

Manufacturer	Korg N364/264
Dates	1996 ^[1]

Technical specifications

Polyphony	64 voices
Timbrality	16 parts
LFO	3 modulators
Synthesis type	PCM rompler
Filter	Lowpass
Aftertouch expression	Yes
Velocity expression	Yes
Storage memory	200 ROM patches + 136 ROM GS/GM patches, 4 ROM 4 drum kits + 8 ROM GM drum kits,
Effects	2 DSP generators, 12 algorithms (reverb, chorus, flanger, distortion, EQ, autopan and aural exciter).
Input/output	
Keyboard	76 keys (N264), 61 keys (N364)



MIDI文件和WAV文件的区别

- ◆生成方式
- ◆文件大小
- ◆声音质量
- ◆易编辑度
- ◆用途





产生MIDI音乐的方法

◆ FM合成法：频率调制合成法

- 由美国斯坦福大学某研究生发明。其基本原理是：由一些基本的乐音波形组合在一起，生成新的音频信号
- 对于基本乐音的组合包括：选用不同的基本波形，改变基本波形的频率，改变载波频率等等
- 1975年，日本山叶公司由约翰·乔宁获得调频合成器（FM合成器）算法的授权。1980年，山叶终于发表了第一部FM数位合成器。

◆ 乐音样本合成法：波形表合成法（wavetable）

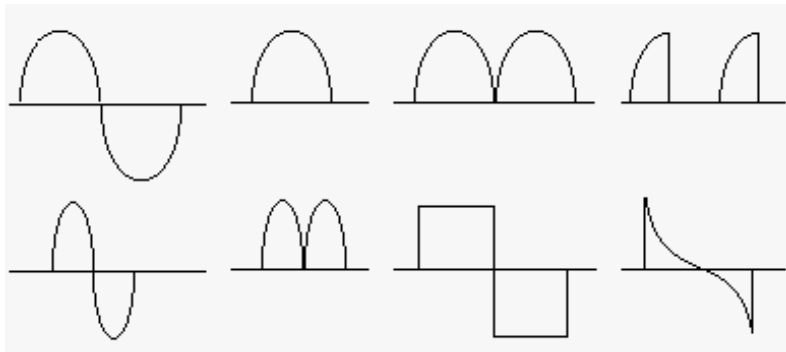
- 把真实乐器发出的声音以数字的形式记录下来，播放的时候改变播放速度从而改变音调周期，生成各种音阶的音符





Yamaha OPL-III数字式FM合成器采用的波形

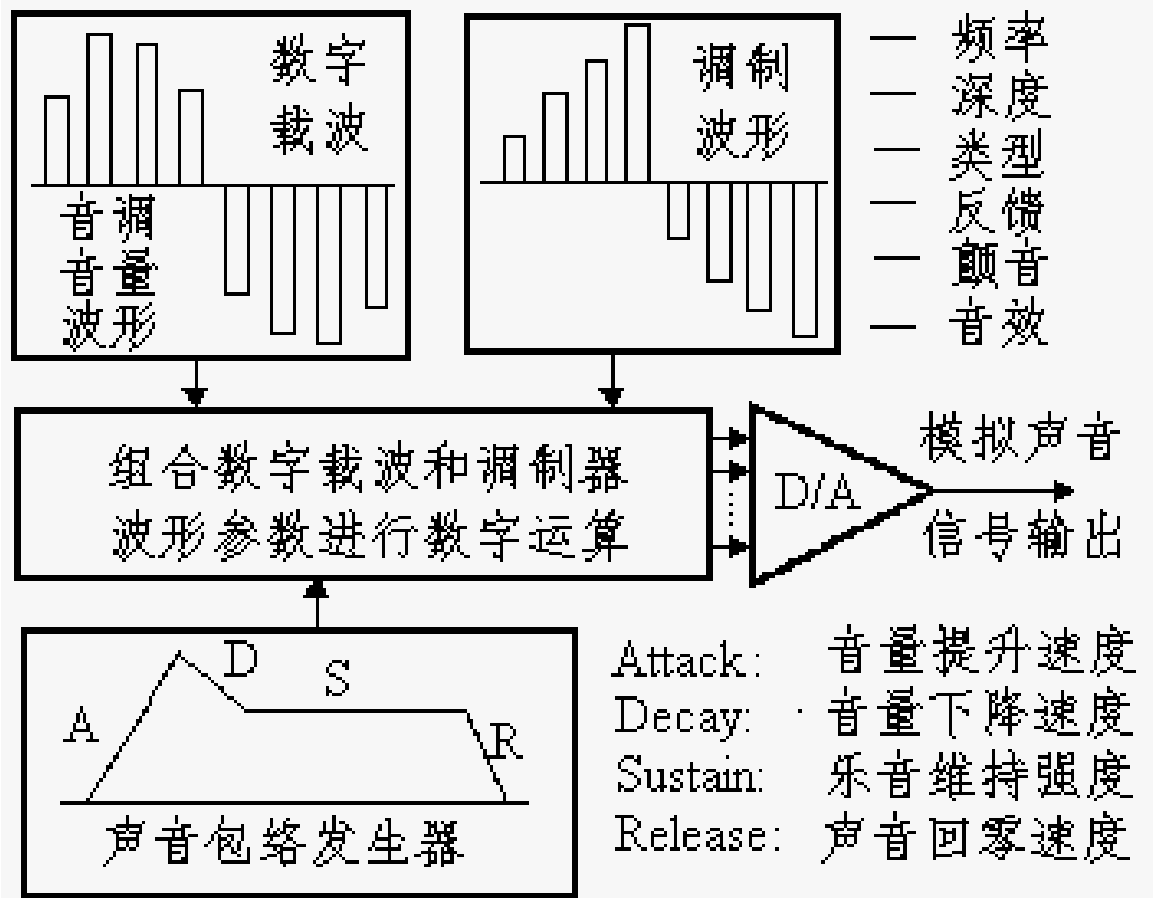
- ◆改变数字载波**频率**可以改变乐音的音调。
- ◆改变数字载波的**幅度**可以改变它的音量。
- ◆改变波形的**类型**，如用正弦波、半正弦波或其它波形，会影响基本音调的完整性。
- ◆**快速改变**调制波形的频率(即音调周期)可以改变颤音的特性。
- ◆**改变反馈量**，就会改变正常的音调，产生刺耳的声音。
- ◆选择的算法不同，载波器和调制器的相互作用也不同，生成的音色也不同。





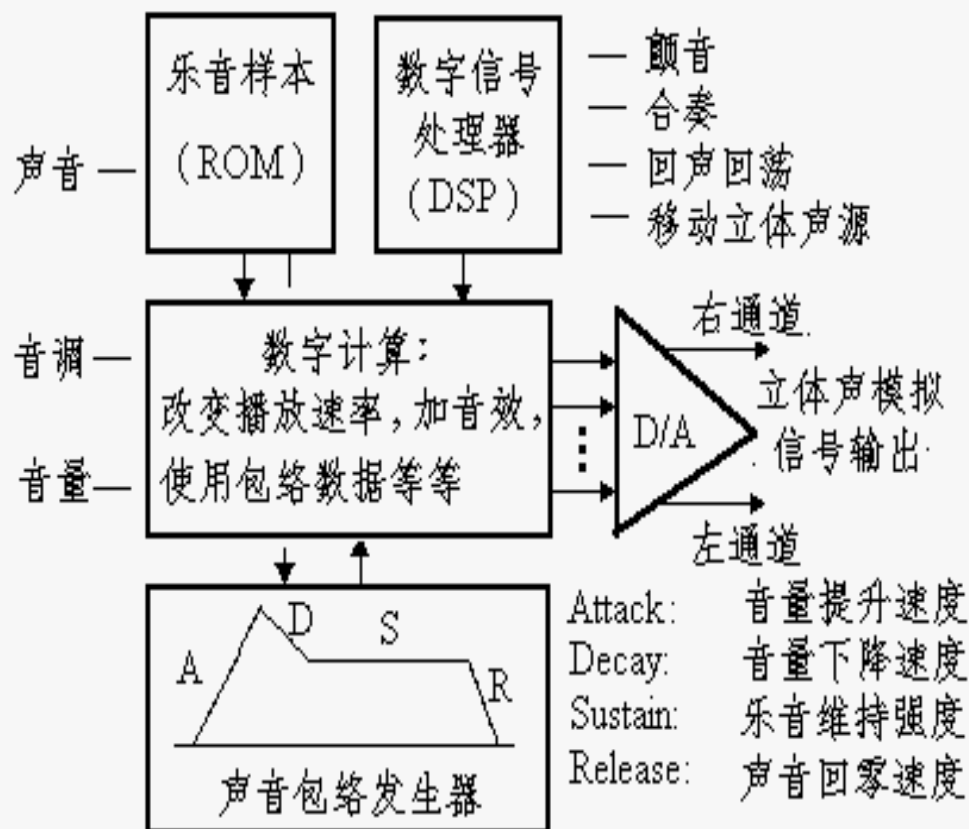
FM合成器工作原理

- ◆ 数字载波器
- ◆ 调制器
- ◆ 声音包络发生器
- ◆ 数字运算器
- ◆ 模数转换器





Wavetable合成器的工作原理



Wavetable合成器所需要的输入控制参数比较少, 可控的数字音效也不多, 大多数采用这种合成方法的声音设备都可以控制声音包络的ADSR参数, 产生的声音质量比FM合成方法产生的声音质量要高。





FM vs. 波形表

- ◆ 都需要录制和播放的双方有先验知识（知道各种基本波形、拥有各种乐器发出声音的样本）
- ◆ 只需要纪录少量的控制信息
- ◆ 用FM合成法产生各种逼真的声音是非常困难的，有些乐音几乎不能产生





MIDI的用途

- ◆ 因为MIDI文件很小，十分适合作为文档、网页等的背景音乐
- ◆ 专业作曲系统





CakeWalk编辑不同的声音通道

Cakewalk Home Studio - [Don't Matter Audio and Midi Demo (Lite).bun - Track]

File Edit View Insert Realtime Go Track Tools Options Window Help

9:01:003 00:00:16:26

1:01:000 1:01:000

1:01:000 1:01:000

2:2 113.85

Name	Source	Key
2 Drums Left	M S R	n/a
3 Bass	M S R	n/a
4 Guitar	M S R	n/a
5 Horns	M S R	n/a
6 Tambourine	M S R	MIDI

Intro Loop Start Loop End

1 2 3 4 5 6 7 8 9 10 11 12 13 14

Drums Left

Bass

Guitar

Horns Left

Tambourine

Shaker

For Help, press F1

Disk space: 1999MB (22%) CPU: 00% Disk: 00%



个人观点

- ◆利用抽象的知识来描述音频媒体的信息是“分析”的结果；而通过对这些知识的“综合”可以还原音频信号，对其他媒体的处理具有**启示作用**
- ◆**如果**用FM方法可以获取完全逼真的信息，意味着人们对声音的认识已经完全透彻（可以用数学来完美描述）
- ◆WAV文件和MIDI文件对声音信号的纪录是人们对“媒体信息”进行存储和再现的两种不同思路
- ◆不同的思路也导致在多媒体数据压缩以及其他相关技术领域内不同的学科分支
- ◆为获取同一目标的不同的出发点产生？？？





本章小结：音频信号处理

◆ 声音存储的变迁

□ 唱片 → 磁带 → CD光盘 → 磁盘文件 → 流媒体

◆ 电脑音乐

□ 编曲 → 合成 → 后期制作

◆ 声音的特性

□ 音高/音色/响度

◆ 声音数字化

□ 采样/量化/量化精度

□ WAV/MIDI/FM/WaveTable





多媒体技术基础

第二章 媒体处理技术

§2.1 数字音频处理技术

谢谢！

