



第五章 声音篇

合肥工业大学 计算机与信息学院

数字媒体方向课程组

目录

CONTENTS

01

声音概念

02

声音的数字化

03

MIDI 音乐

04

语音

05

小结

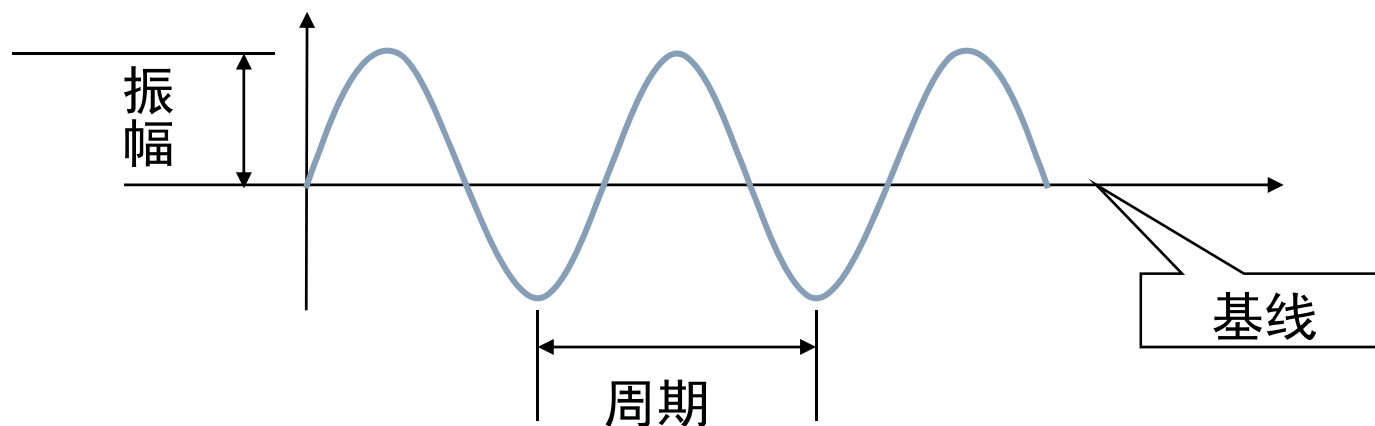
声音处理的历史

耳朵听→留声机（机械方式记录）→唱片、磁带（电模拟方式记录）→光碟、计算机（数字信号）



声音的产生

声音是由空气中分子的振动而产生的。自然界的声音是一个随时间而变化的连续信号，可近似地看成是一种周期性的函数。通常用模拟的连续波形描述声波的形状，单一频率的声波可用一条正弦波表示。



声音三大要素

振幅——每个声音都有振幅，用来表示声音听起来的大小程度的。声音的振幅用来衡量气压波在其初始位置或静止位置上的偏移大小。

频率——决定音调的高低。声音的频率是周期数的倒数。它表示每秒钟的周期数，其单位是赫兹（Hz）或每秒周期数（cps）。

泛音——决定音的色彩。



声音的频率范围

次声波	0~20Hz
人耳能听到的声音	20Hz~20kHz
超声波	20kHz~1GHz
高超声波	1GHz~10TH

目录

CONTENTS

01

声音概念

02

声音的数字化

03

MIDI 音乐

04

语音

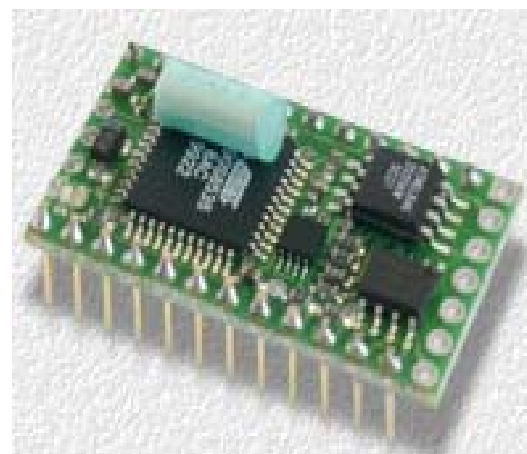
05

小结

模数转换

声波是随时间而连续变化的物理量，通过**能量转换装置**，可用随声波变化而改变的电压或电流信号来模拟。以模拟电压的幅度来表示声音的强弱。

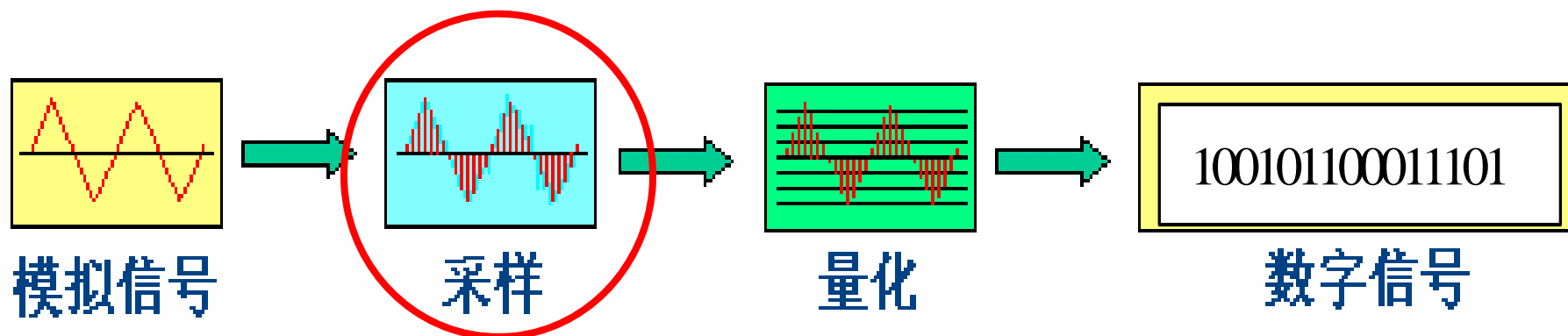
为使计算机能处理音频，必须对声音信号模数转换，要经过**采样**、**量化**和**编码**三个步骤。



8位可编程A/D转换芯片

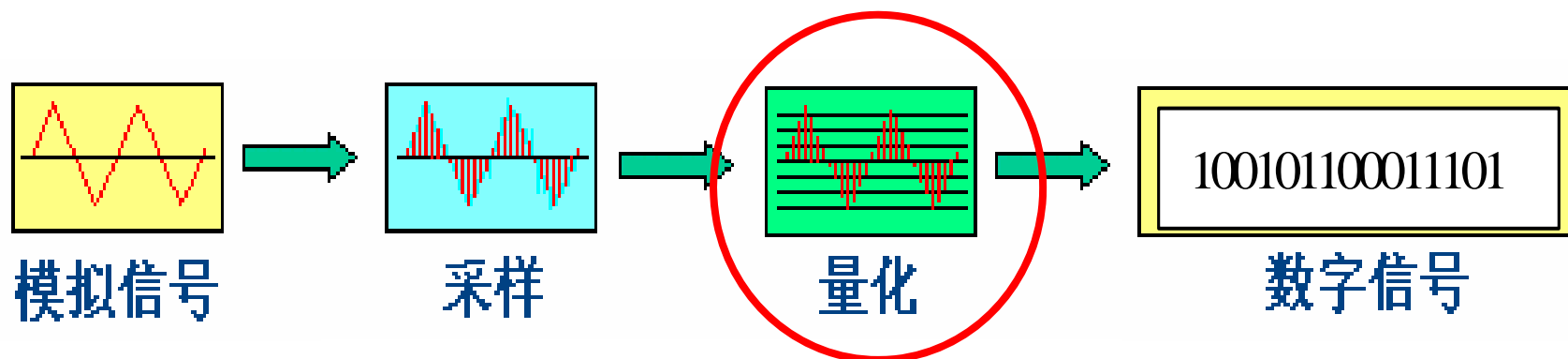
采样

将在时间轴上连续的声音波形进行时间轴上的离散化，具体做法是每隔一个时间 t ，采集一个波形数据，连续波形进行采样的频率（时间 t 的倒数）叫**采样率**。



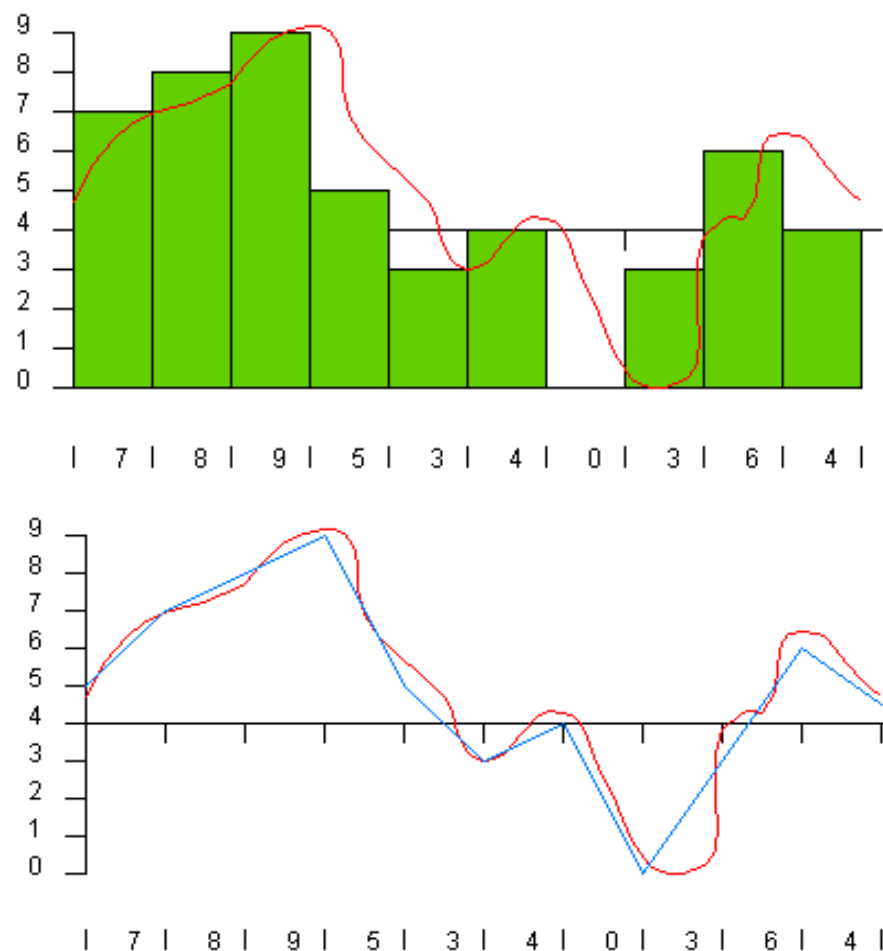
量 化

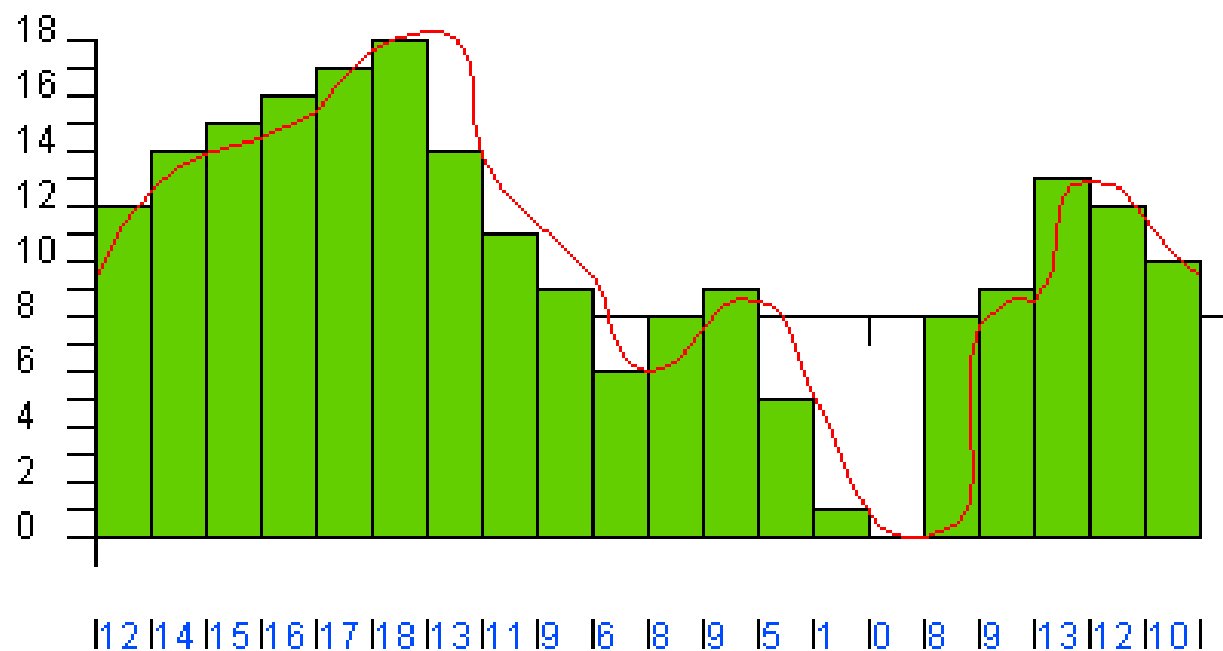
即用多少**二进制**来表示声音波形的高度，bit数越多 声音质量越好。这实际上是在振幅轴上的离散化。



波形还原

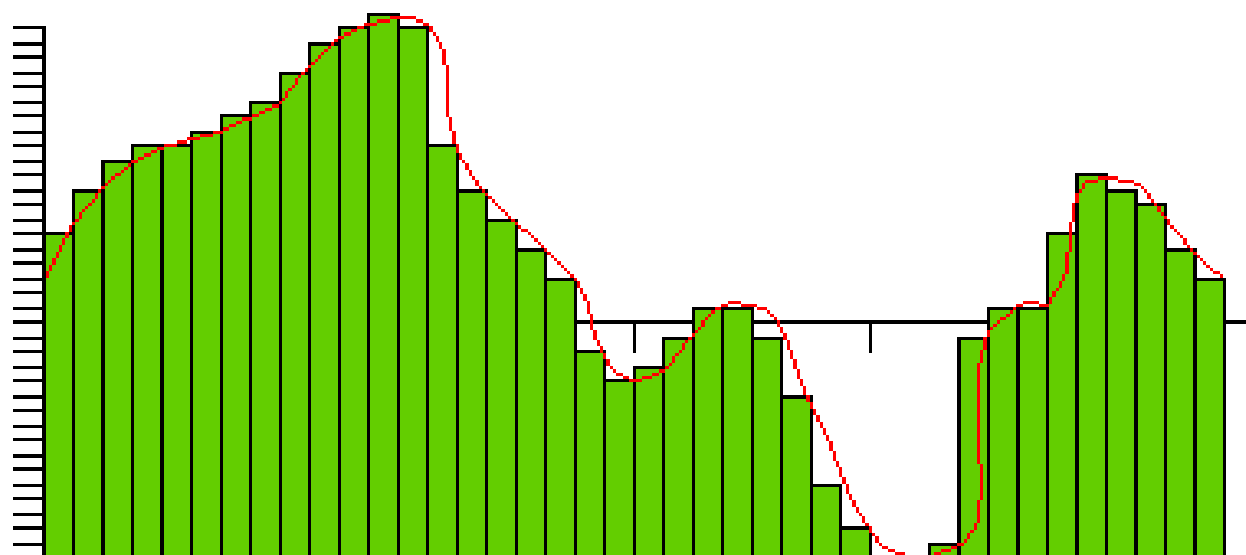
当D/A转换器从上图得到的数值中重构原来信号时，得到下图蓝色(直线段)线段所示的波形。重构后的信号波形有较大的失真。





采样率2000Hz
量化等级为20

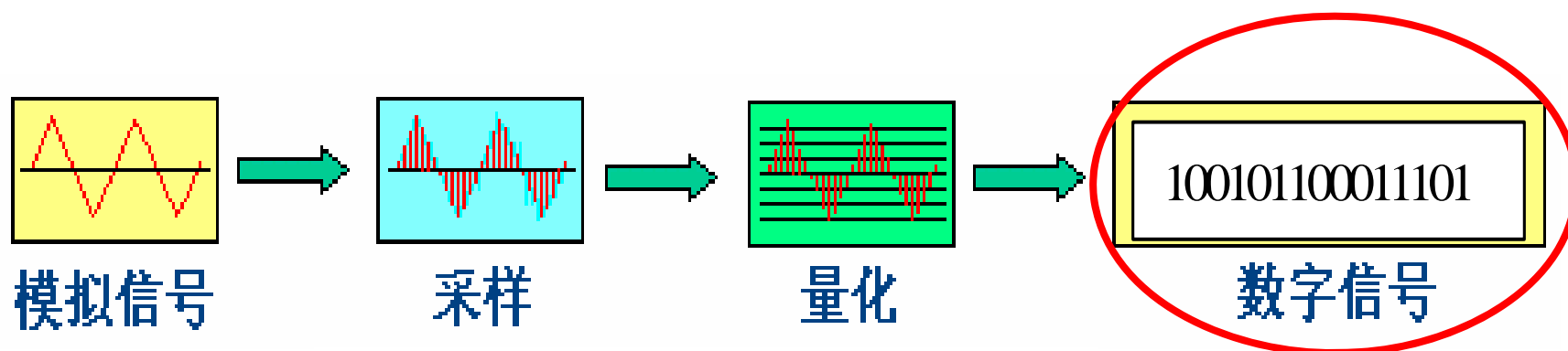
采样和量化哪
一个对声音的
数字化效果影
响大？



采样率4000Hz
量化等级为40

编 码

将采样后量化好的数据按照一定的数据格式编排（含数据的压缩）存放于计算机中。



常见采样与量化指标

	语音效果	音乐效果	高保真效果
采样频率kHz	11.025	22.05	44.1
量化精度bit	8	16	16
数据率kb/s	88.2	352.8	705.6
字节率kB/s	11.025	44.1	88.2

补充：48khz：DVD、数字电视中使用。

96khz-192khz：DVD-Audio、蓝光高清等使用。

SONY公司的音频格式Hi-Res Audio，192kHz / 24bit

还有一个参数

通道数

单声道 双声道 . . .

5.1声道： 1992年，DOLBY实验室提供了五个声道的从20Hz到20kHz的全通带频响，即正前方的左（L）、中（C）、右（R），后边的两个独立的环绕声通道左后（LS）和右后（RS）。另外，为了弥补低音的不足，还同时提供了一个100Hz以下的超低音声道供用户选用，因为这只是一个辅助通道，所以定为0.1通道，这样就构成了AC-3的5.1通道。



数字化容量计算

数字音频的存储量:

存储量=采样频率×量化位数/8 ×声道数×时间

数据率=采样频率×量化位数/8

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

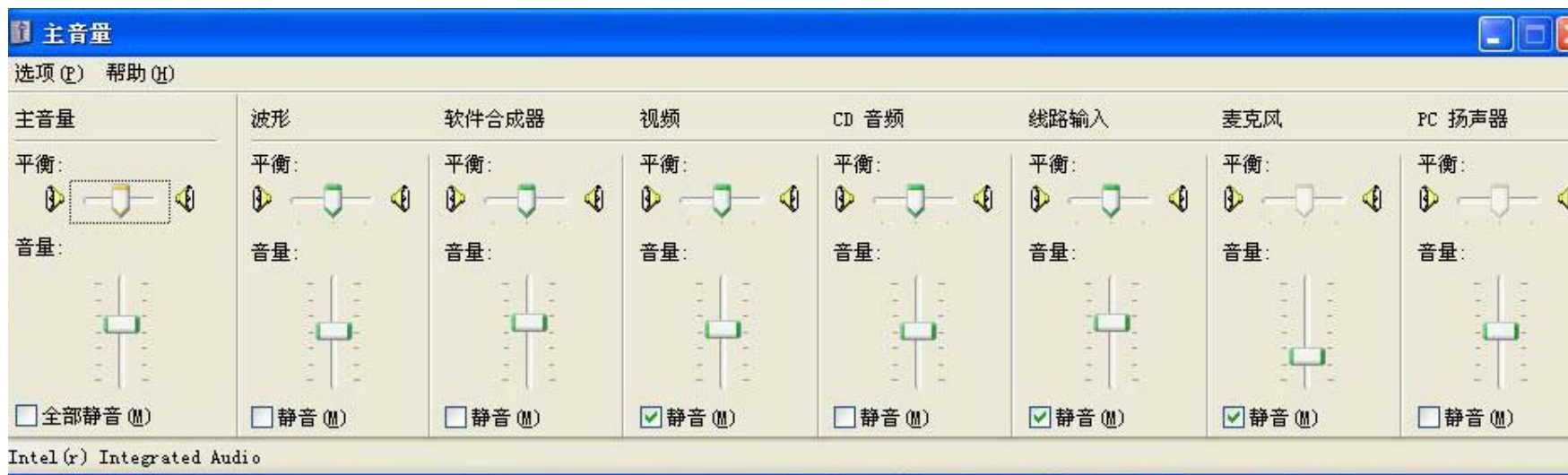
$$44.1 \text{ K} \times 16 \times 2 \times 60 \div 8 = 10584 \text{ KB}$$

音频通用的应用与研究内容

- ◆ 高品质的采集技术
- ◆ 音频的编码与解码
- ◆ 音频数据实时传输
- ◆ 音视频的同步技术
- ◆ 音频效果与编辑技术

音频采集

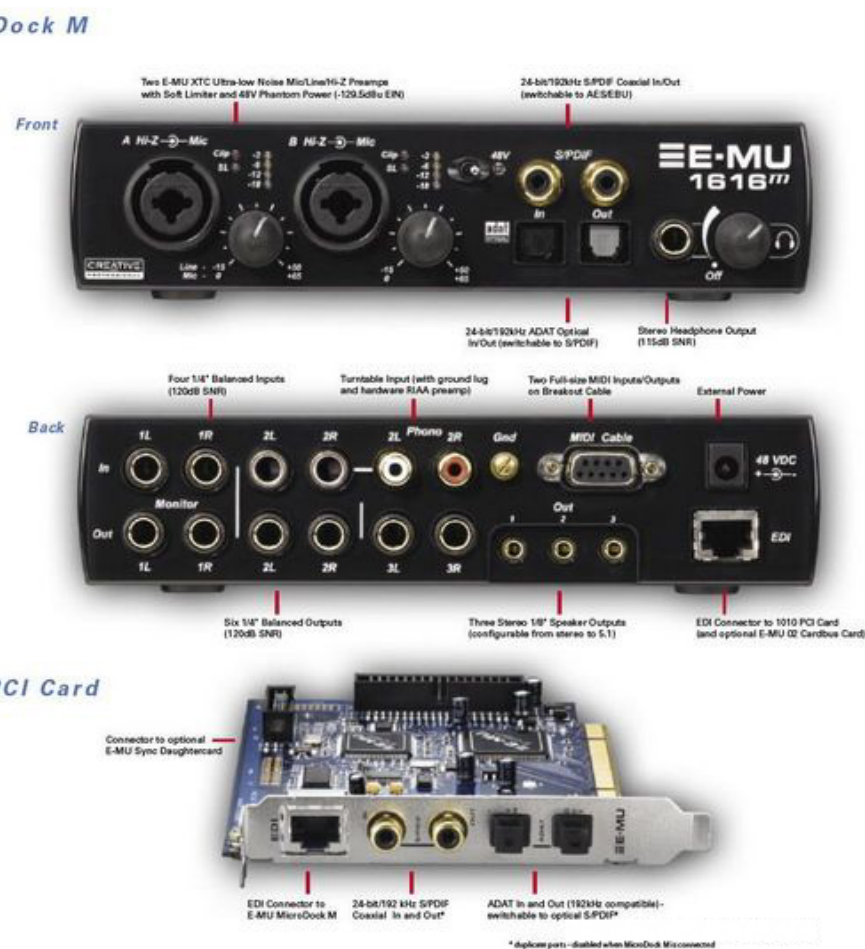
音频的采集在于拾音技术的提高，与采集设备（如声卡、话筒等）关系非常大。



电脑采集声音的通道

声卡

声卡是多媒体计算机必不可少的一种类型数字化声音信息。PC是英国的ADLIB AUDIO公司生产的ADLIB声卡。1989年，Creative Labs SoundBlaster



声卡的组成

- 1). **Wave合成器** Wave合成器的模/数转换和数/模转换是声卡上数据处理器件。
- 2). **MIDI合成器** 标准的多媒体计算机通过MIDI合成器播放MIDI文件
- 3). **混音器** 声卡上的混音器芯片可以对以下音源进行混合：数字化声音(DAC)，调频FM合成音乐(FM)CD音频(CD-ROM)，线路输入(AUX)，话筒输入(MIC)及PC声音输出(SPK)。
- 4). **MIDI接口** 声卡能够接收、录制及输出MIDI信号。
- 5). **CD-ROM接口** CD-ROM接口提供了从CD-ROM的CD-DA的输出信号到声卡音源输入的通路，CD-ROM将音频信号直接通过声卡的功发送到扬声器。
- 6). **DSP数字信号处理器** 用作对数字音频信号的实时压缩和解压缩，以及用于语音朗读、语音识别等特殊音频信号的处理。
- 7). **SPDIF数字音频接口** SPDIF的输出端可用于接驳专门的数字录音设备如DAT和MD等，而将SPDIF的输入端接到光驱的Digital Out或DVD解压卡的相应输出端子则可得到比使用模拟音频输入要好得多的音质。

主要音频格式

WAV 该格式记录了声音的波形，它不对数据进行压缩，所以文件很大，但只要“拾音”设备足够好，制作的声音质量可以达到专业级水平。

MOD 该格式存放乐谱和乐器使用的各种音色样本，所以该文件不大，但声音回放效果很好，对回放设备要求比较低，而且音色种类可以用无止境。

MP3 采用MPEG Audio 的Layer-3压缩方案，压缩比达到14-12:1，用一张MP3CD可以容纳10张唱片的歌曲。是目前网上常用的音乐格式。

RA格式 这是RealNetwork公司的RealAudio格式，它的压缩比很大非常适合在网上实时播放，是现在网络实时播放的主要格式之一。

主要音频格式

CDA

就是激光唱片的格式，记录的是波形流，纯正、HIFI。采样位数多为16位，现在一经常看到20位、24位、36位采样录音CDA。缺点是无法编辑和文件长度太大。

MIDI

作为音乐工业的数据通信标准，MID是一种专业性的语言，他的科学性、兼容性、复杂程度等各方面都是最高的。回放效果也是最好的。

WMA

是微软WMT 4.0(Windows Media Technology)的一个组件。Windows中提供的媒体播放器Windows Media Player 7/8支持播放。

从HIFI的角度，声音质量由高到低顺序：

原声乐器演奏>MIDI>CD唱片>MOD>所谓声卡上的MIDI>CMF

无损格式

- **Wav** PCM | 1411 kbps | 44100 Hz | 立体声
- **flac** FLAC | 860 kbps | 44100 Hz | 立体声

Free Lossless Audio Codec。FLAC是免费的并且支持大多数的操作系统

- **ape** Monkey's Audio | 980 kbps | 44100 Hz | 立体声

APE是流行的数字音乐无损压缩格式之一，是以更精炼的记录方式来缩减体积，还原后数据与源文件一样，从而保证了文件的完整性。压缩率约为55%，比FLAC高。

目录

CONTENTS

01

声音概念

02

声音的数字化

03

MIDI 音乐

04

语音

05

小结

MIDI的概念

- MIDI是Musical Instrument Digital Interface的缩写，可以把MIDI理解成是一种协议、一种标准、或是一种技术，但不要把它看作是某个硬件设备。它的出现解决了各个不同厂商之间的数字音乐乐器的兼容问题！

容量小 音质好 适用广



MIDI文件

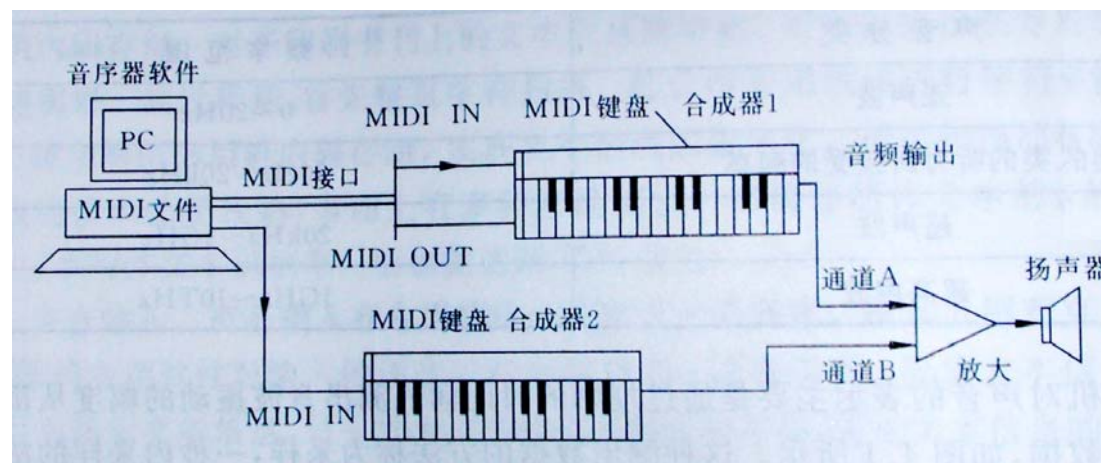
MIDI文件中存放的是一组时序命令，而不是音频信号，是一些对音乐表现的描述，比如音高、音符开始与结束的时间、强度、是否颤音等等。

```
0 FF 58 04 04 02 24 08  
0 FF 51 03 500000  
0 C0 5 通道1, 音色5  
0 C1 46 通道2, 音色46  
0 C2 70 通道3, 音色70  
0 92 48 96 通道3 开始弹奏C2, 用力  
0 92 60 96 通道3 开始弹奏C3, 用力  
96 91 67 64 通道2 开始弹奏G3, 用力  
96 90 76 32 通道1 开始弹奏E4, 钢琴  
192 82 48 64 通道3 停止弹奏C2, 标准  
0 82 60 64 通道3 停止弹奏C3, 标准  
0 81 67 64 通道2 停止弹奏G3, 标准  
0 80 76 64 通道1 停止弹奏E4, 标准  
0 FF 2F 00 结束
```



MIDI系统

- MIDI端口
- MIDI键盘
- 音序器
- 合成器



MIDI系统实际就是一个作曲、配器、电子模拟的演奏系统。从一个MIDI设备转送到另一个MIDI设备上去的数据就是MIDI消息。MIDI数据不是数字的音频波形，而是音乐代码或称电子乐谱。

乐器种类：合成器、编曲机、MIDI设备、吉他、效果器、音箱、调音表、节拍器

音序器

MIDI作曲和配器系统核心部分,序列器实际上是一个音乐词处理器 (word porcessor), 应用它可以记录、播放和编辑各种不同MIDI乐器演奏出的乐曲, 将其存入MIDI文件。**序列器可以是硬件, 也可以是软件**, 它们作用过程完全与专业录音棚里多轨录音机一样。



MIDI键盘

MIDI键盘是用于MIDI乐曲演奏的，MIDI键盘本身并不发出声音，当作曲人员触动键盘上的按键时，就发出按键信息，所产生的仅仅是MIDI音乐消息，从而由音序器录制生成MIDI文件。



键盘式的（合成器、主控键盘）、弦控式的（MIDI吉他）、敲击式的（鼓机）、吹奏式的（呼吸控制器）

合成器

解释MIDI文件中的指令符号或者MIDI消息，生成所需要的声音波形，经放大后由扬声器输出，声音的效果比较丰富。

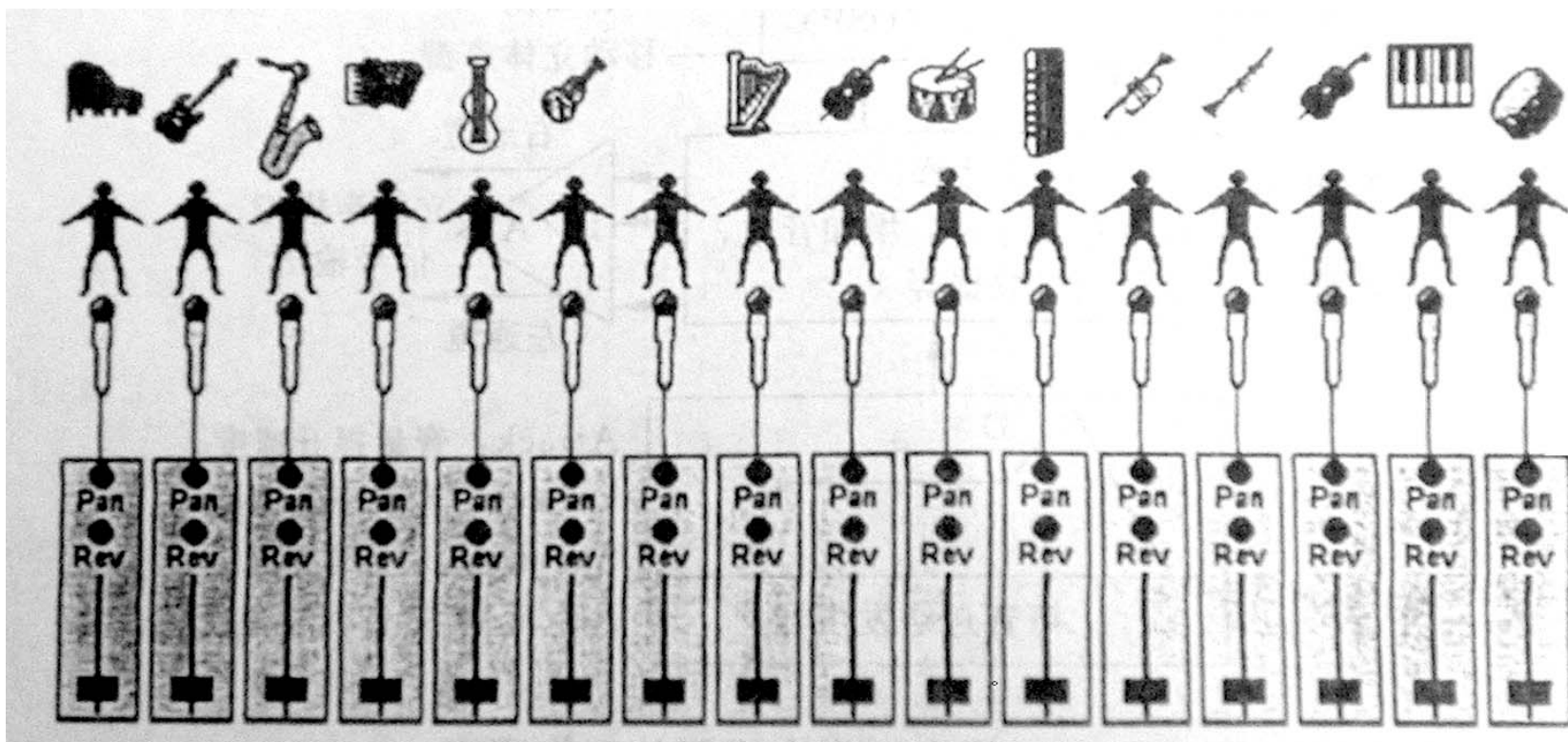
调频合成(FM)：运用声音振荡的原理对MIDI进行合成处理，由于技术本身的局限，效果很难令人满意。

波形表合成(Wave Table)：首先将各种真实乐器所能发出的所有声音（包括各个音域、声调）进行取样，存储为一个波表文件。在播放时，根据MIDI文件记录的乐曲信息向波表发出指令，从“表格”中逐一找出对应的声音信息，经过合成、加工后回放出来。

波表合成又常被分为**软波表**和**硬波表**

MIDI通道

标准一共**16个逻辑通道**，每个通道访问一个独立的能产生特定声音的逻辑合成器。最多可以连接16个外部音源设备，但并不是最多只能有16个声部演奏。



MIDI端口

- **MIDI In**: 接收来自其它MIDI设备的MIDI信息。
- **MIDI Out**: 发送本设备生成的MIDI信息到其它设备。
- **MIDI Thru**: 将从MIDI In端口传来的信息转发到相连的另一台MIDI设备上。

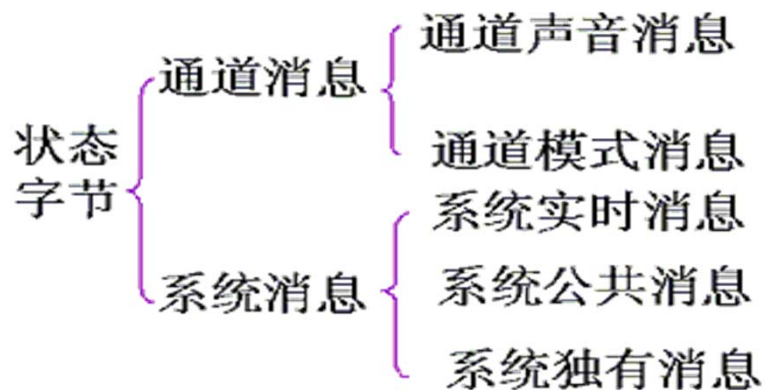


用于连接各种MIDI设备所用的电缆为5芯电缆，通常人们也把它称为MIDI电缆。

MIDI消息

- 通道声音消息在设备间传送实际的表演数据，描述键盘动作、控制器动作和控制盘改变。它们描述音乐的音调、音量、音色、时长和其他的声音质量。

MIDI消息=状态字节+数据字节



“一个人的乐队”——电脑音乐

硬件篇

- ✓ PC电脑
- ✓ 具有较高信噪比的声卡
- ✓ MIDI输入接口和设备
- ✓ 真实的回放音箱

<http://www.midifan.com>

硬件

- MIDI键盘
- MIDI控制器
- MIDI接口
- 音频接口
- 模拟调音台
- 数字调音台
- 话筒
- 耳机
- 话放
- 耳放
- 监听音箱
- DSP处理卡
- 数字效果器

硬件

- 合成器
- 模拟设备
- DJ
- VJ
- 数字钢琴
- 吉他效果器
- AD/DA
- 动态/均衡
- 数字工作站
- 录音机
- 采样器
- 舞曲采样器
- 音序器

音序器软件

Cakewalk曾今最流行的音序器软件，后来整合的功能越来越多。音乐工作站的未来发展方向是MIDI、音频、音源（合成器）一体化制作最先实现这个方式的是著名的Cubase软件。

Cakewalk公司推出了新一代的音乐工作站——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，使之成为新一代全能型超级音乐工作站。。



效果与后期

- **MIDI效果器**（如Style Enhancer, Guitar Pro）
为MIDI信号添加逼真、富于变化的人性化的真实效果
- **音频效果器**（如Waves）
作用于音频Wav事件的效果设置，优化音频效果
- **后期处理软件**（如T-Racks）
利用自身配备的EQ均衡器、电子管压缩器和限制器起到最好的暖声效果，改变“数码之声”的冰冷、机械化

目录

CONTENTS

01

声音概念

02

声音的数字化

03

MIDI 音乐

04

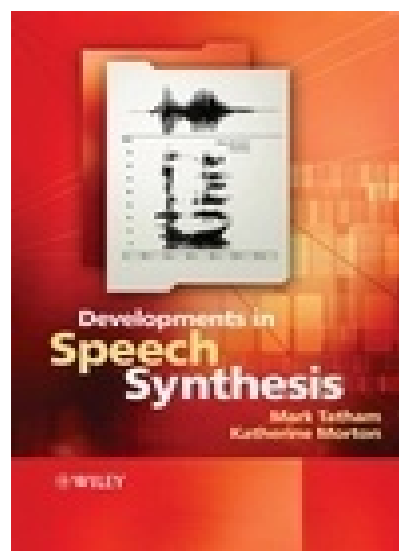
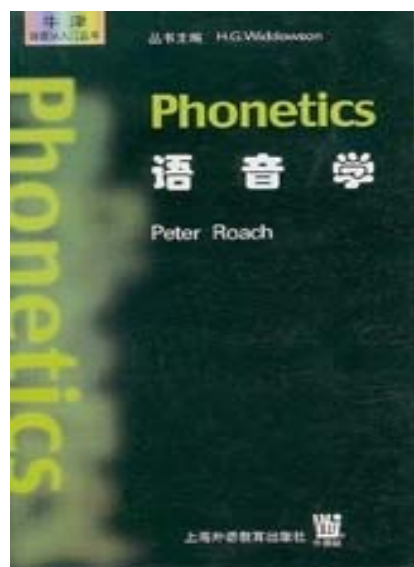
语音

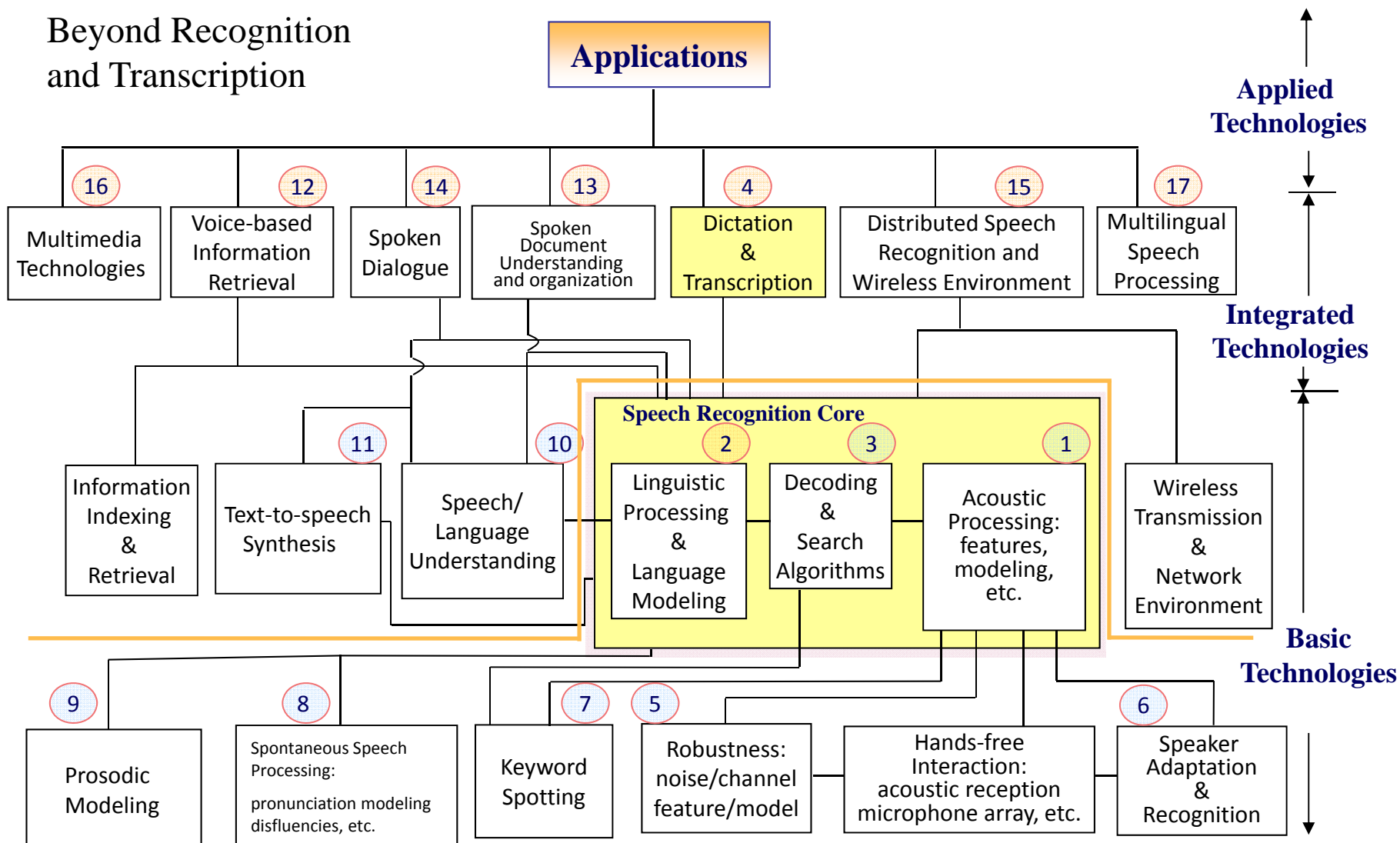
05

小结

什么是语音？

语音也是一种波形，所以和波形声音的文件格式相同。但是因为人类语音的生物学特性，它有着与一般声音不同的特征以及实际的应用价值。





Recognition and Transcription

语音传输

- 传输方式的变化

基带模拟传输→频带模拟传输→数字传输

电话网：64kb/s

- IP电话

通过Internet打电话的形式有四种：计算机到计算机、计算机到电话、电话到计算机、电话到电话。

- Internet上的实时音频广播

Web上的优秀音频技术依赖于缓冲、编码/解码和音频流技术。

语音合成

语音合成是通过机械的、电子的方法产生人造语音的技术，又称文语转换（Text to Speech）技术。

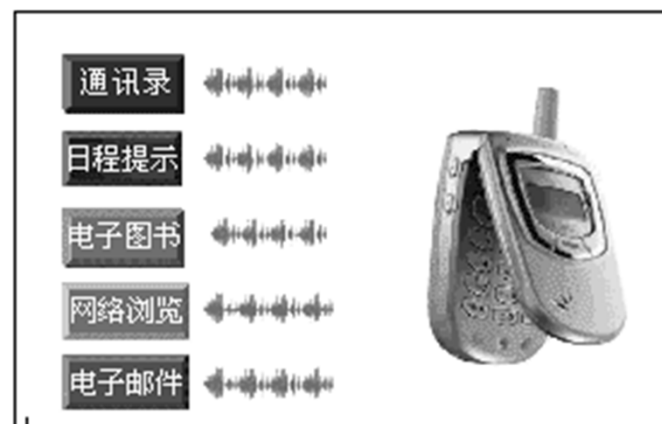
A. 语言学处理

B. 韵律处理

C. 声学处理

参数合成

波形拼接



语音合成已经进入产业化时期

语音分析

- 声音指纹（**WHO**）：根据语音识别说话的人
共振峰特点，稳定周期性
- 语音识别与理解（**WHAT**）：识别和理解语音所表达的内容
①词汇表范围②说话人限制③训练要求④正确识别率
- 说话的心理分析（**HOW**）：分析说话人的心理状况→测谎仪器

目录

CONTENTS

01

声音概念

02

声音的数字化

03

MIDI 音乐

04

语音

05

小结

- MIDI已经并将长期成为“电脑音乐”的代名词
- 语音研究还缺乏坚实的理论基础，目前还没有弄清楚语音的产生“发生”和语音的感知“接受”的本质，而只是做了各种工程上的处理。
- 数字声音的应用面极广，开放资源很多，可以切入各种应用
- 研究上经常将其与视频联系起来