

多媒体技术



数字音频与MIDI基础

河南大学计算机与信息科学学院

刘扬

2017/11/28 上午11时7分45秒

目录

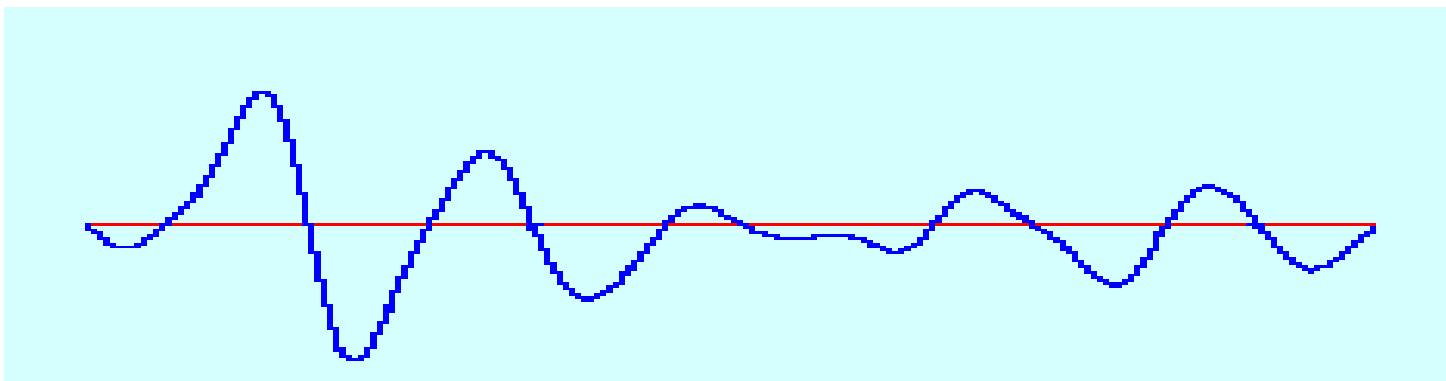
- 1、音频基本概念
- 2、听觉系统的认知心理学感知特性
- 3、音频数字化
- 4、声音重建与声卡
- 5、音频的简单编码
- 6、波形声音的常用处理操作的编程原理
- 7、电子音乐合成与MIDI系统
- 8、语音合成(TTS)
- 9、音频文件格式

1、音频基本概念

■ 声波的基本特性

- 反射 (reflection)、折射 (refraction)、衍射 (diffraction)

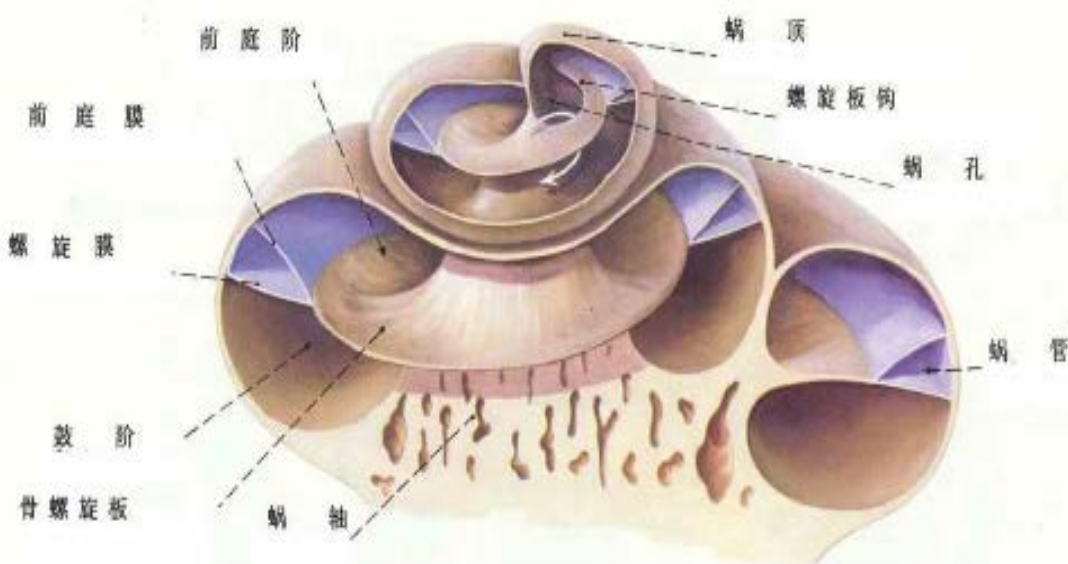
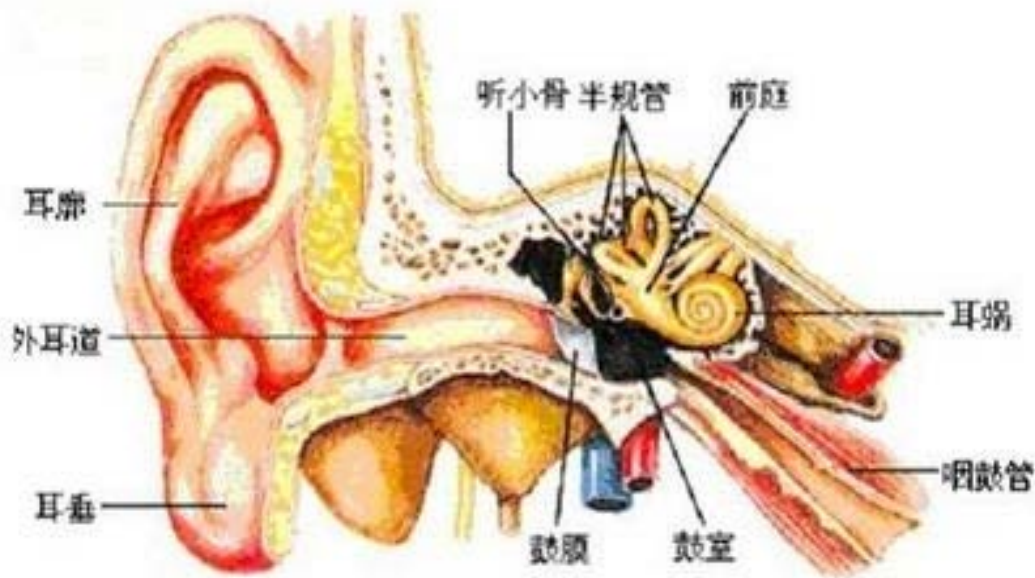
- 音频信息可分为规则音频和不规则声音。其中规则音频又可以分为语音、音乐和音效。规则音频是一种连续变化的模拟信号,可用一条连续的曲线来表示,称为声波。声音的三个要素是**音调、音强和音色**。声波或正弦波有三个重要参数: 频率 ω_0 、幅度 A_n 和相位 ψ_n , 这也就决定了音频信号的特征



听觉系统

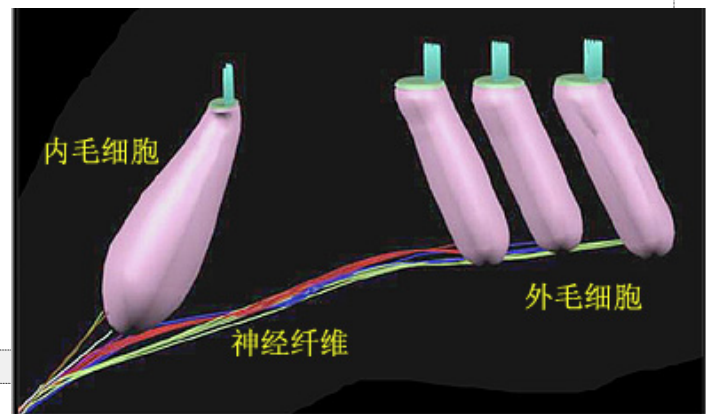
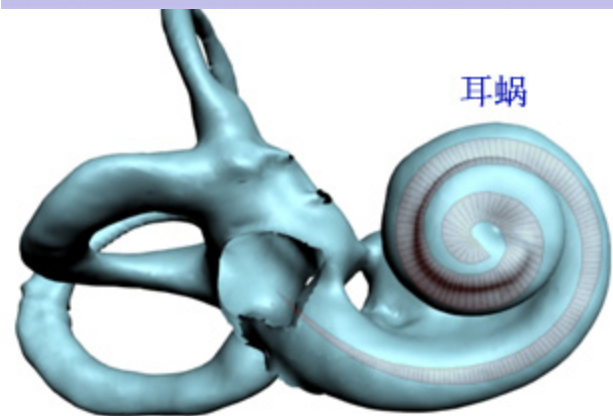
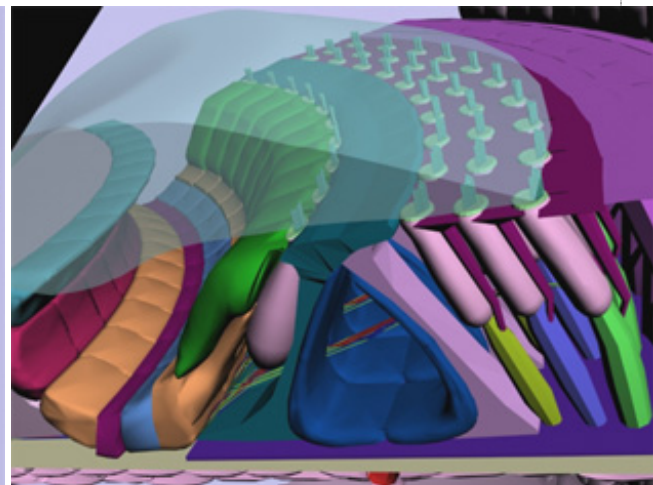
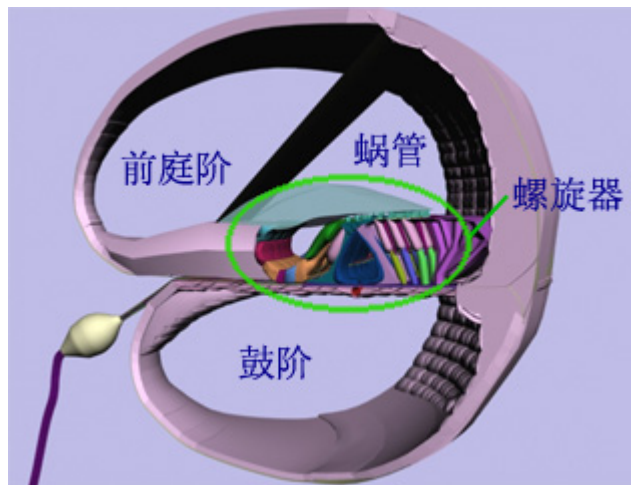
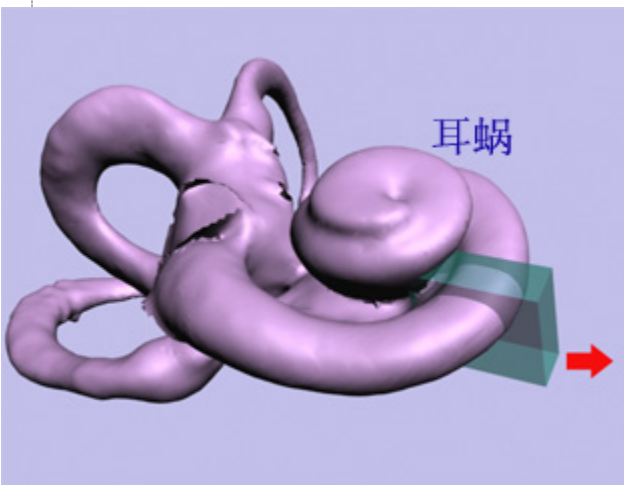
■ 耳蜗对声音频率的分析(美籍匈牙利学者Bekesy1928年听觉的**行波学说**,基底膜产生最大振幅部位)。基底膜长约30mm,靠耳蜗底部较窄,向顶部逐渐加宽。从蜗底到蜗顶,分别感受高低频的声音。**声波频率越高,行波传播越近**,基底膜出现最大振幅的部位越靠近耳蜗**底部**;反之,**声波频率越低,则行波传播越远**,基底膜出现最大振幅的部位越靠近耳蜗**顶部**。振动最大处**毛细胞**受刺激最强,不同来源和组合的听神经纤维的神经冲动传到听觉中枢的不同部位,就可引起不同音调的感觉

■ 内耳耳蜗对声音**强度**的初步分析(声波在基底膜上引起振动的幅度声波强度不同)基底膜上的振动幅度、毛细胞受刺激的强度、兴奋的神经纤维数目及发放的冲动频率也就不同。传入内耳的声波强,在基底膜上引起的振动幅度就大,该部位的毛细胞受刺激强,兴奋的神经纤维数目多,每条神经发放的冲动也多,因而在大脑皮层引起强音的感觉。反之,引起弱音的感觉。



听觉系统

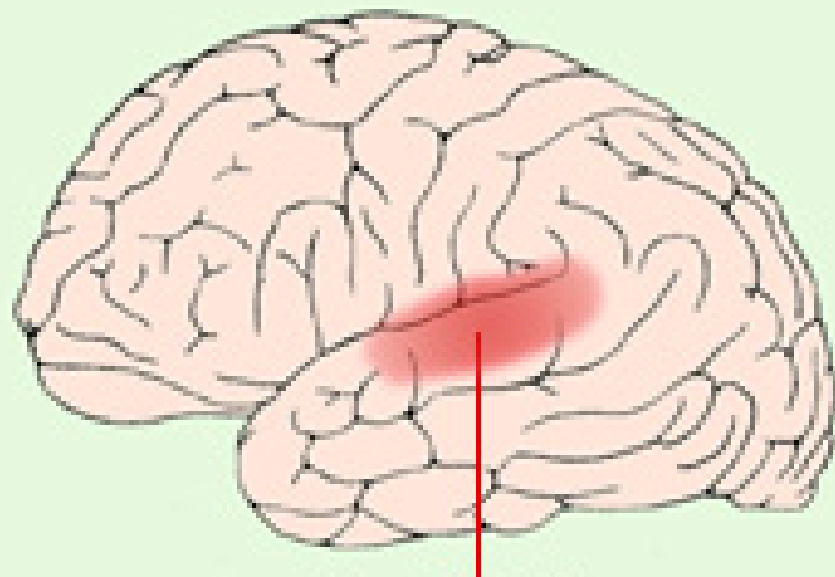
- 基底膜像盘旋上升的山路一样，从蜗轴底部可以拾级而上，直达蜗顶，总长为30~35毫米。这条“山路”并非上下一样宽，靠近底转也就是靠蜗轴底部最窄，宽约40~80微米；沿蜗轴旋转上升时，“山路”越走越宽，到达蜗顶时，基底膜宽达500微米，约增宽10倍。基底膜在底周处较紧密，在蜗顶处较疏松。基底膜约由29000根横行纤维所构成。



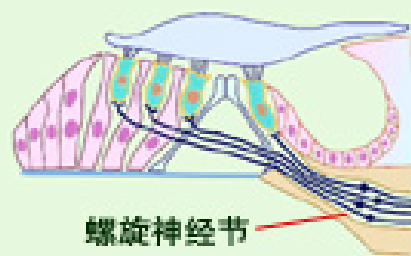
听觉传导通路



听觉系统

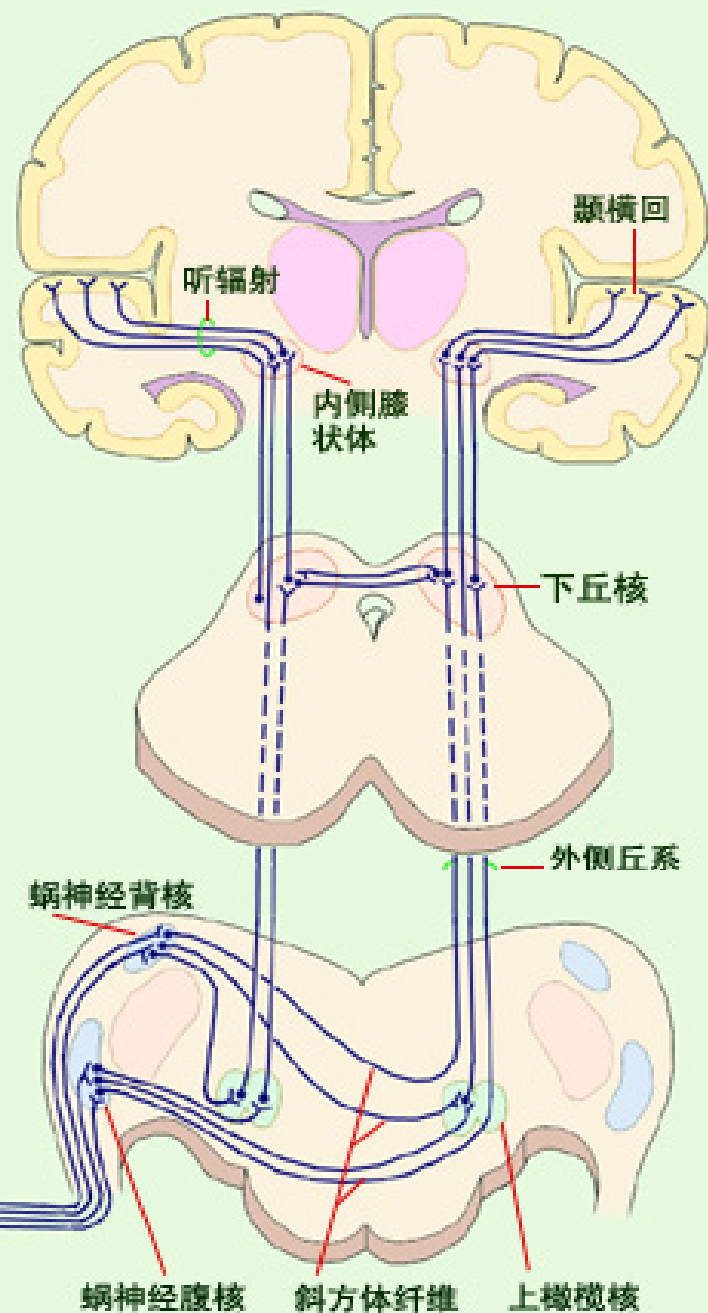


听皮质 (颞横回)



螺旋神经节

蜗神经



听觉传导通路

声波的特性-幅度与音强

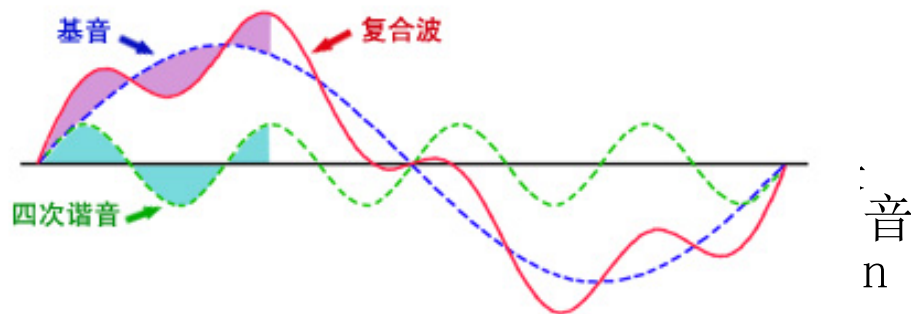
- **基频与音调 (Pitch)**：频率是指信号每秒钟变化的次数。人对声音频率的感觉表现为音调的高低，在音乐中称为音高。音调正是由频率 ω 所决定的。音调与声音的频率相关，频率高则音调高，频率高则音调高。人的听力范围是20Hz~20KHz，低于20Hz称为次声波，高于20KHz称为超声波。
- **幅度与音强 (Loudness)**：声强的大小与声速成正比，与声波的频率的平方、振幅的平方成正比。超声波的声强大是因为其频率很高，炸弹爆炸的声强大是因为振幅大。声音强度由振动幅度的大小决定，以能量来计算称声强，以压力计算表示时称声压。声强 I 与声压 P 的关系为：

$$I = (P^2) / (\rho v) \quad \text{其中 } \rho - \text{介质密度}, v - \text{声速}$$

人耳对于声音细节的分辨只有在强度适中时才最灵敏。人的听觉响应与强度成对数关系。常用音量来描述音强，以分贝为单位。在处理音频信号时，绝对强度可以放大，但其相对强度更有意义，一般用动态范围定义：正常谈话的音强约为60dB，安静的播音室一般为25dB，人能忍受的最大音强为120dB；声压 I 的声强级 L 为：

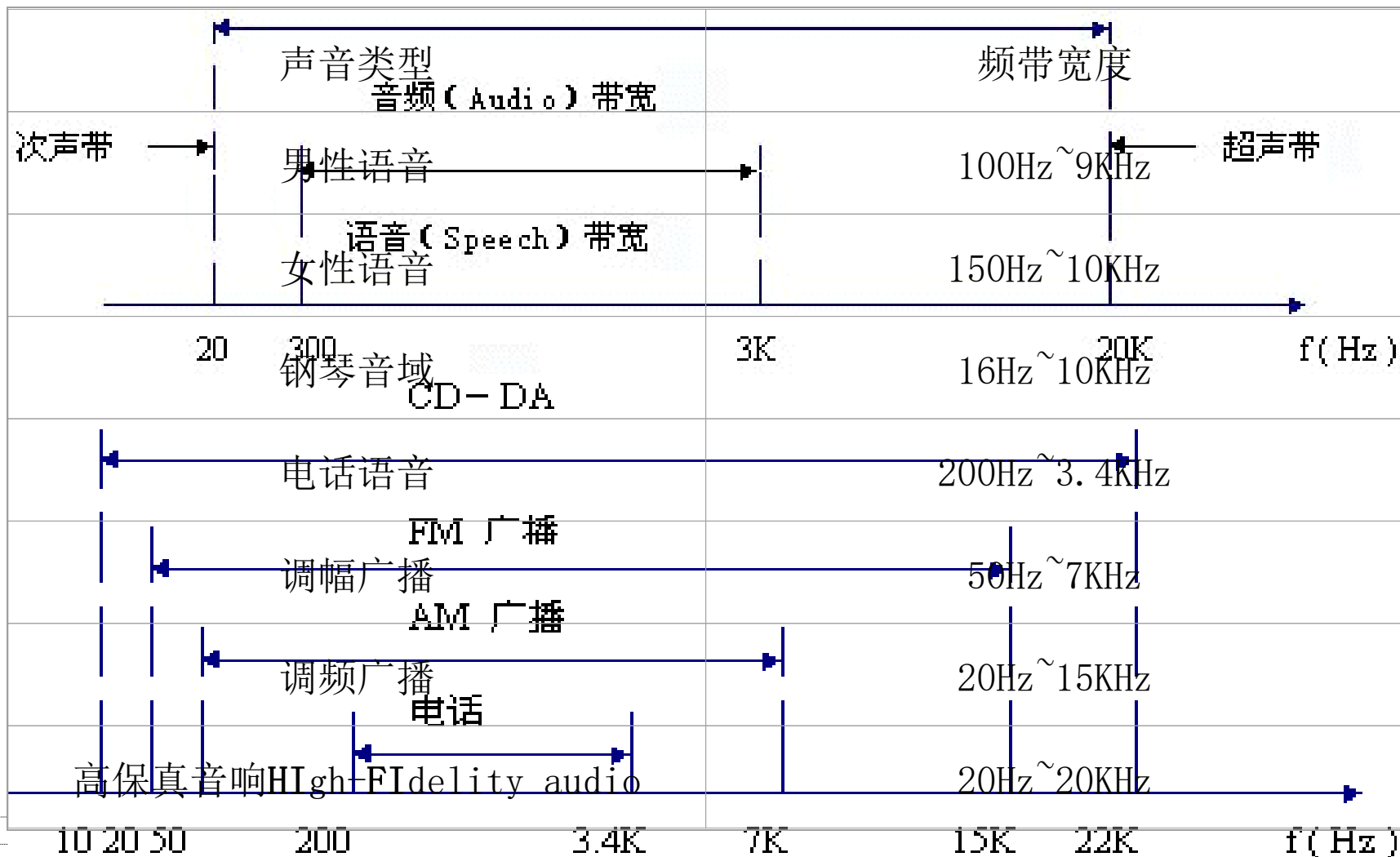
$$L = 10 \times \log_{10}(I/I_0) \text{ (dB)}$$

- **谐波与音色 (Timbre)**：音色是定。 $n \times \omega_0$ 称为 ω_0 的高次谐波。音色就越有明亮感和穿透力。不同音，由此产生各种音色效果。



音宽与频带

- **音宽与频带：** 频带宽度或称为带宽，它是描述组成复合信号的频率范围。



乐音与噪音

乐音的等程音阶(12平均率音阶)

- 音乐中音阶的划分是在频率的对数坐标 ($20 \times \log$) 上取等分而得的:
- $n = 12 \log_2(f_n/f_0)$ n 为半音数, $n=1$ 为1个半音, 即小二度; $n=2$ 为1个全音, 即大二度, 全音。
- `n=1:1000; f0=1;%11025`
- `for i=0:12; b=sin(f0.*2^(i/12)).*n);`
- `wavplay(b); end`

音名	#c'			#d'			#f'			#g'			#a'					
频率(Hz)	377.2			311.1			370.0			415.3			466.2					
F(n)/F(c')	$2^{1/12}$			$2^{3/12}$			$2^{6/12}$			$2^{8/12}$			$2^{10/12}$					
音名	c'	d'	e'	f'	g'	a'	b'	c''										
唱名	1	2	3	4	5	6	7	1										
频率(Hz)	261.6	293.7	329.6	349.2	392.0	440.0	493.9	523.3										
F(n)/F(c')	$2^{0/12}$	$2^{2/12}$	$2^{4/12}$	$2^{5/12}$	$2^{7/12}$	$2^{9/12}$	$2^{11/12}$	$2^{12/12}$										

音名	c'	#c'	d'	#d'	e'	f'	#f'	g'	#g'	a'	#a'	b'	c''
唱名	1		2		3	4		5		6		7	
$\frac{F(n)}{F(c')}$	1	$2^{1/12}$	$2^{1/6}$	$2^{1/4}$	$2^{1/3}$	$2^{5/12}$	$2^{1/2}$	$2^{7/12}$	$2^{2/3}$	$2^{3/4}$	$2^{5/6}$	$2^{11/12}$	2
频率	261.63	377.18	293.66	311.13	329.63	349.23	369.99	392.00	415.30	440.00	466.16	493.88	523.25

乐音与噪音

- 噪声作为一个随机信号，仍然具有统计学上的特征属性。谱密度（功率按频率的分布函数）即是噪声的特征之一，从而人们可以通过不同类型的谱密度来区分噪声。幂律噪声（Power-law noise）在单位频域内的谱密度正比于 $1/f^\alpha$

- 白噪声 ($\alpha = 0$) 表示在全频域内单位频域下都分布有相同的能量密度，在线性空间内它具有平坦的频谱。

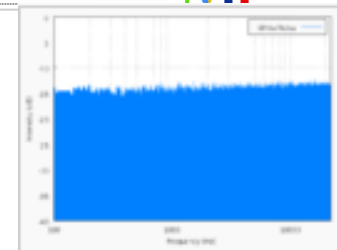
```
white=rand(48000*10,1)*2-1;
wavwrite(white,48000,16,'white noise uniform.wav');
wavplay(white,48000)
hist(white,100,1)
```

- 粉红(pink)噪声 ($\alpha = 1$) 又称作 $1/f$ 噪声，它的频谱在对数空间内是平坦的，“非常悦耳的一种噪声”，也就是说在等比例宽度的频带内具有相等的功率。最常用于声学测试，利用粉红噪声可以模拟出瀑布或者下雨的声音，与脑电图 α 波频谱类似（清醒、安静、闭眼时出现 α 波，波幅呈棱形规律由小到大，再由大到小变化；兴奋时出现 β 波快波；困倦时出现 θ 波；器质性病变时 δ 波）。

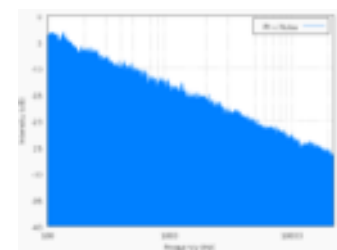
```
pink=powernoise(1,48000*10,'normalize');
wavwrite(pink,48000,16,'pink.wav');
```

- 红噪声 ($\alpha = 2$) 又称作布朗噪声 (Brown noise)，当频率提高为2倍时，它的谱密度都会降低6dB，也就是说红噪声的谱密度是随频率增加而呈衰减的

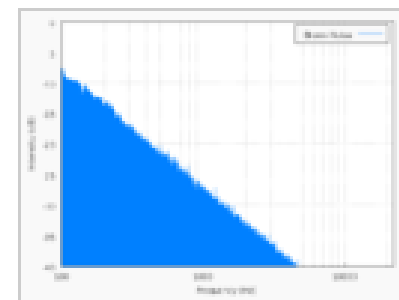
```
red=powernoise(2,48000*10,'normalize');
wavwrite(red,48000,16,'red.wav');
```



白噪声的频谱



粉红噪声的频谱



红噪声的频谱

幂律噪声

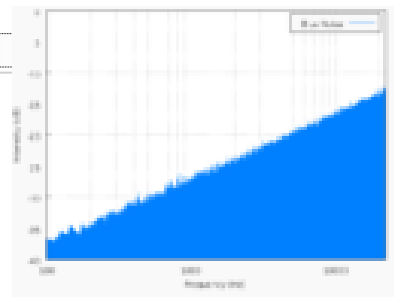
- 蓝噪声又称作天蓝噪声，它随着频率提高2倍谱密度提高3dB，从而频谱与 f 成正比（在有限带宽内）。在计算机图形学中，蓝噪声在对图像进行抖动处理中很有用。

```
blue=powernoise(1,48000*10,'normalize');  
wavwrite(blue,48000,16,'blue.wav');
```

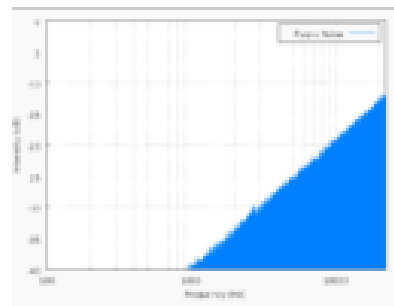
- 紫噪声随着频率提高2倍谱密度提高6dB，从而频谱 f^2 与成正比（在有限带宽内）。

```
white=randn(48000*10+1,1);  
purple=diff(white);  
purple=purple/(max(abs(purple)));  
wavwrite(purple,48000,16,'purple.wav');
```

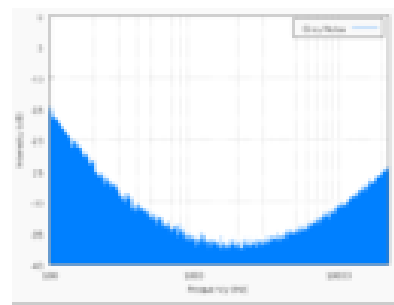
- 灰(gray)噪声是在某一特定频率范围内遵循音质等响度曲线变化的随机粉红噪声，这种噪声能够使人类听觉系统在全频率上感受到同样的响度。



蓝噪声的频谱



紫噪声的频谱



灰噪声的频谱

听觉心理变量与物理变量的关系

心理变量	首要的物理变量	次要的物理变量
响度	声强	声波频率
音调	声波频率	声强
音色	声波复合	—
音量	频率和强度	—
密度	频率和强度	—
谐和（流畅或粗糙）	谐波结构	音乐技巧
噪声	强度	频率组合，各种时间参量
骚扰声	强度	频率组合，无意义

2、听觉系统的认知心理学感知特性

- 20Hz~20kHz整个可听声频率范围内，上下限频率共10个倍频程。

低音频段

- 1 20~40
- 2 40~80
- 3 80~160

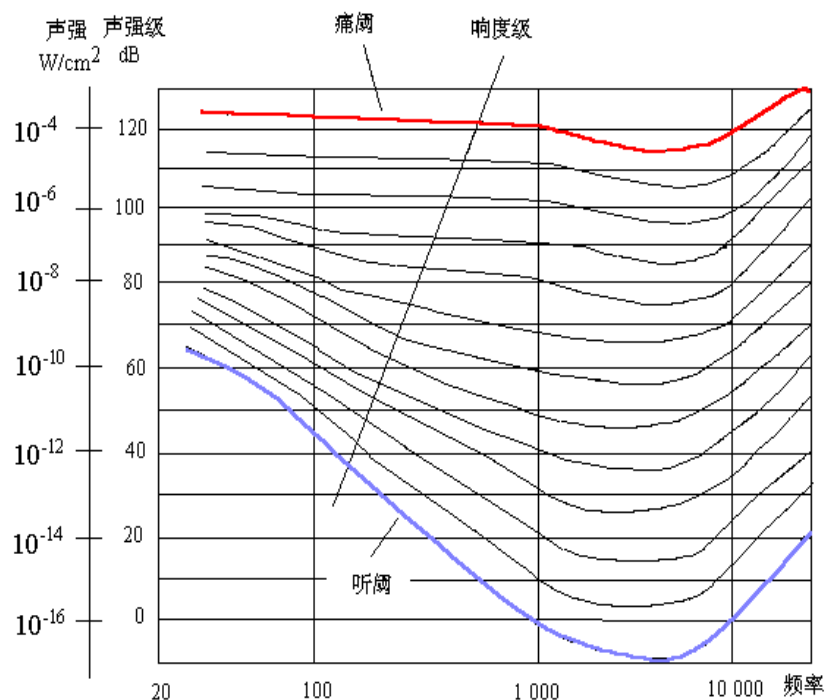
中音频段

- 4 160~320
- 5 320~640
- 6 640~1280
- 7 1280~2500

高音频段

- 8 2500~5000
- 9 5000~10000
- 10 10000~20000

- 人耳对不同频率的敏感程度差别很大，其中对2kHz~4 kHz范围的信号最为敏感，幅度很低的信号都能被人耳听到。而在低频区和高频区，能被人耳听到的信号幅度要高得多。

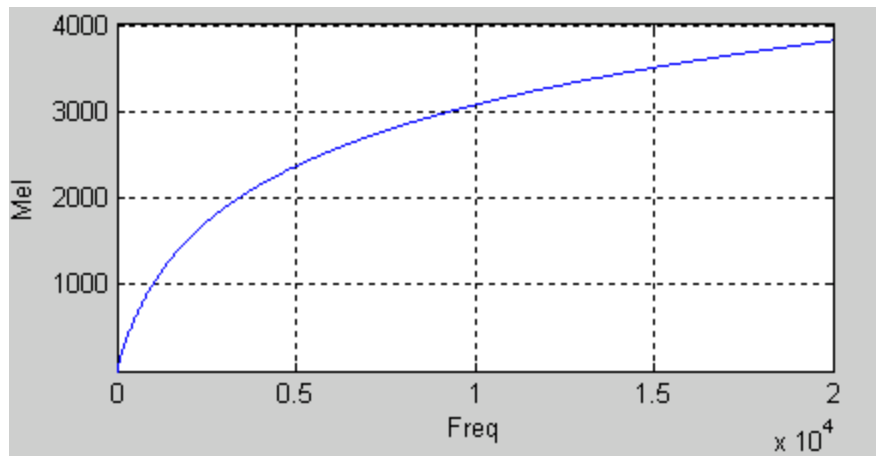


“听阈—频率”曲线

听觉系统的认知心理学感知特性

- 其中f 单位为Hz，这也是两个既不相同又有联系的单位。主观感觉的音高单位则是美尔“Mel”

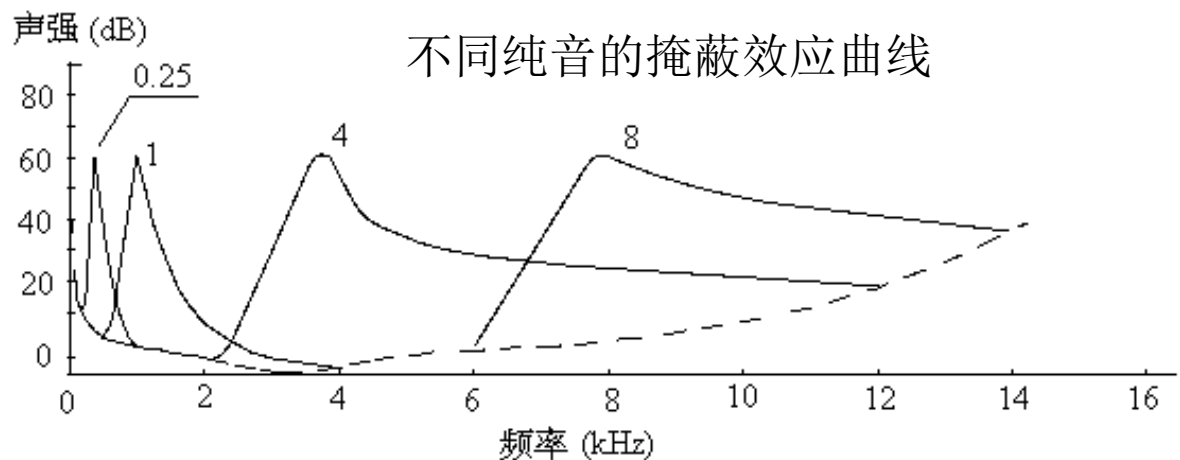
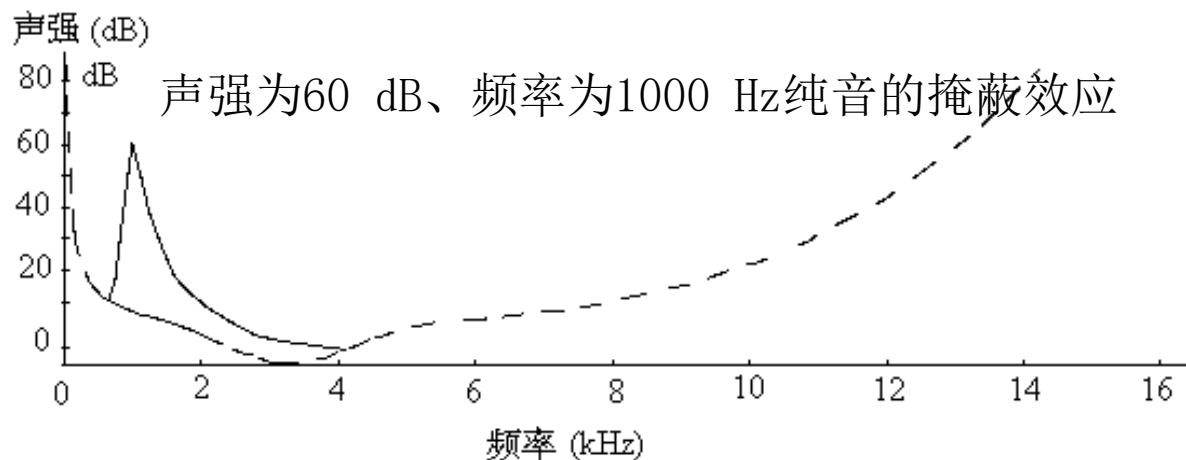
$$\text{Mel} = 2595 \log_{10}(1 + f/700)$$



“音高—频率”曲线

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称为掩蔽效应。前者称为掩蔽声音(masking tone)，后者称为被掩蔽声音(masked tone)。掩蔽分成频域掩蔽和时域掩蔽。



掩蔽效应与临界频带

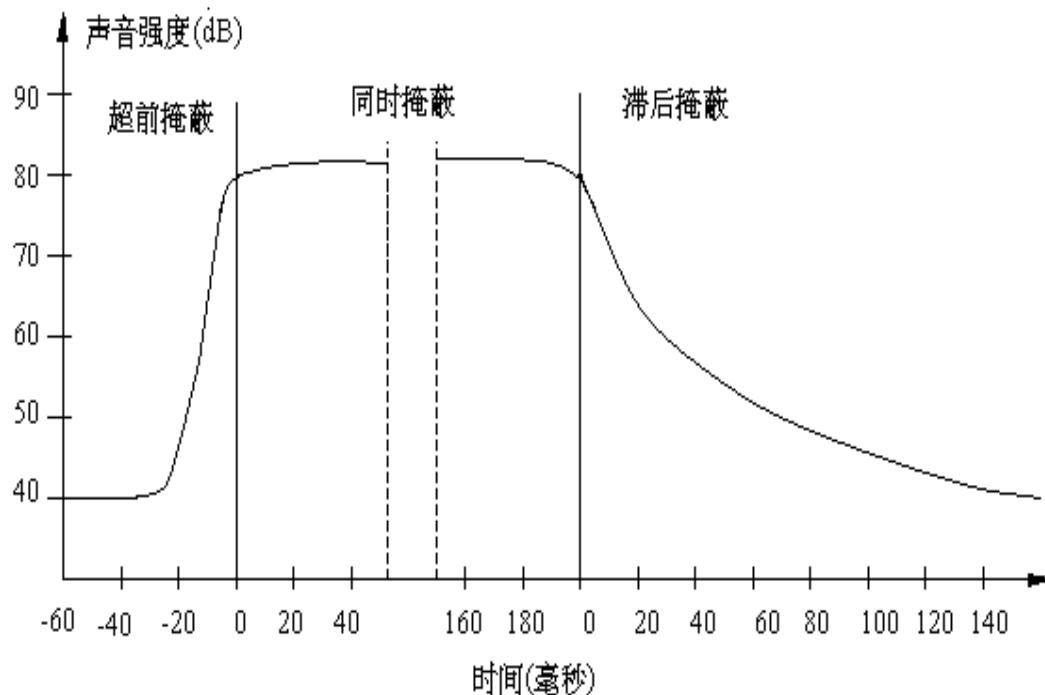
- 临界频带(Critical Band):一个纯音可以被以它为中心频率一定带宽的连续噪音所掩蔽, 如如这频带内的噪声功率等于在噪声中刚能听到的该纯音的功率, 则这频带就称为听觉临界频带.
- 临界频带位置通常不固定, 以任何频率为中心频率都有临界频带, 连续的临界频带序号记为临界频带域(Bark域), 通常20Hz~20KHz分24个Bark域

临界	频率 (Hz)		
频带	低端	高端	宽度
0	0	100	100
1	100	200	100
2	200	300	100
3	300	400	100
4	400	510	110
5	510	630	120
6	630	770	140
7	770	920	150
8	920	1080	160
9	1080	1270	190
10	1270	1480	210
11	1480	1720	240
12	1720	2000	280
13	2000	2320	320
14	2320	2700	380
15	2700	3150	450
16	3150	3700	550
17	3700	4400	700
18	4400	5300	900
19	5300	6400	1100
20	6400	7700	1300
21	7700	9500	1800
22	9500	12000	2500
23	12000	15500	3500
24	15500	22050	6550

掩蔽效应

- 除了同时发出的声音之间有掩蔽现象之外，在时间上相邻的声音之间也有掩蔽现象，并称为时域掩蔽。时域掩蔽分为超前掩蔽(pre-masking)和滞后掩蔽(post-masking)。

- 产生时域掩蔽主要原因是人的大脑处理信息需要花费一定的时间。一般来说，超前掩蔽很短，只有大约5~20 ms，而滞后掩蔽可持续50~200 ms。



■ 相位

- 从声音的波形来看，声音的起点和方向也要反映声音的特性，这就是声音的相位。当两个声音相同相位完全相反时，它们将相互抵消；当两个声音相同而且相位也相同时，声音就会得到加强。相位的确定对于多声道声音系统的设计非常重要，其可以应用在回声的消除、会议系统的声音设计上。

■ 自然声音的时变现象

- 声音的音调分成三个区域：起始区、稳定状态区、延迟区。研究表明，音调的频谱分量随时间改变。在稳定状态区，频谱保持固定。在起始区，振幅频谱随时间变化。因此自然声音的起始部分是非常难识别的。例如刚听了一小节音调后要识别乐器，专家也会觉得较难。时变现象用于数字系统中，说明声音中的某些错误是不太容易发现的，但如果出现停顿就很容易引起人的注意。

■ 听觉空间

- 人耳可听到来自各个方向的声音，并用不同的因素来判定声源的位置。声源的位置不论对于增进人们的感受还是增进对声音的理解，都是非常重要的。通过声音的精确再现，就可以构造出听觉空间。方位的线索是各种声音到达两耳的精确时间和强度。（有关音频媒体的三维化处理在VR技术讲解）

■ 听觉的频谱特性

- 声音是时间函数，通过傅里叶变换可做出其频谱图。人耳对频谱成分的波峰和波谷是非常敏感的。在语言中，元音很少有频谱变速变化的区域。基频改变，人耳是很敏感的。例如：快进的录像，音调会发生变化。音色非常复杂，目前尚在研究中。音色的处理将使我们能识别音源，音色也代表和声音有关的主观质量。

■ 声音的心理模拟

- 通过人工真实的方法，可以对视觉空间的景物进行再造或虚构，同样也可以对听觉空间的声音进行心理的模拟，这就是所谓的可听化（audiolization）。用声音可以表达出一些声音的效果。

3、音频的数字化和符号化

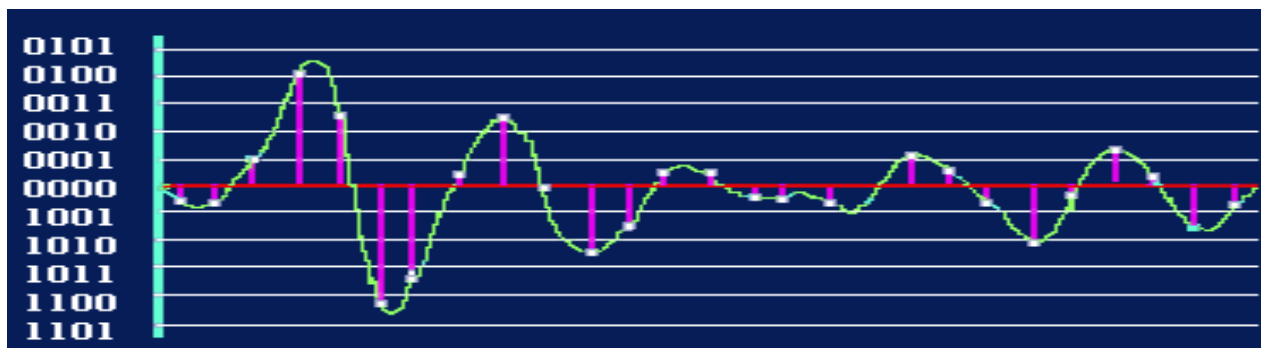
- 数字音频一般分为波形音频和MIDI（Musical Instrument Digital Interface乐器数字化接口）音频
- 波形音频数字化：声音进入计算机的第一步就是数字化，数字化实际上是将模拟音频信号经过A/D转换形成的数据。其质量取决于采样频率、量化位数和声道数。
- 波形音频的存储量

$$M = (\text{采样频率} \times \text{采样位} \times \text{声道数}) / 8 (\text{字节/秒})$$

音频数字化过程

1. 取样 (sampling)
2. 量化 (quantization, AD conversion)
3. 编码 (encoding)

8K8b单声道44KB
22K16b立体声475KB



模拟
声音
信号

Sampling

ADC

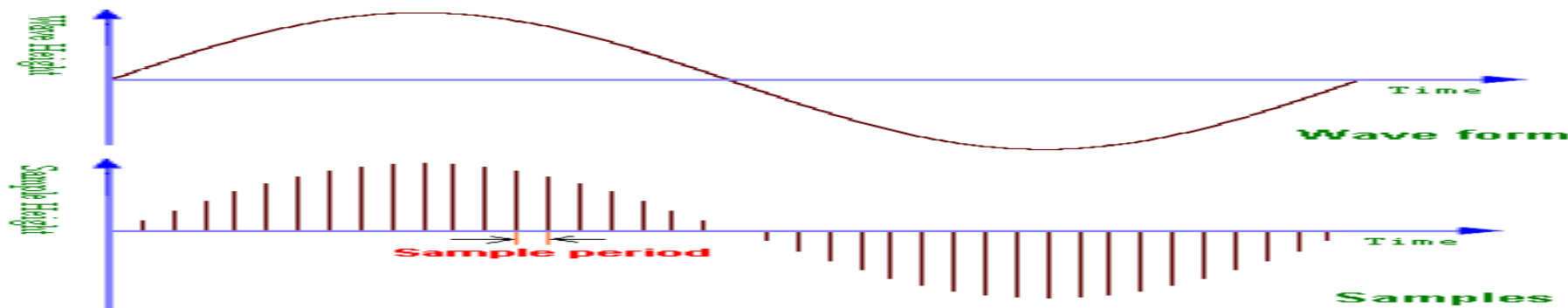
Encoding

数字
声音



音频数字化

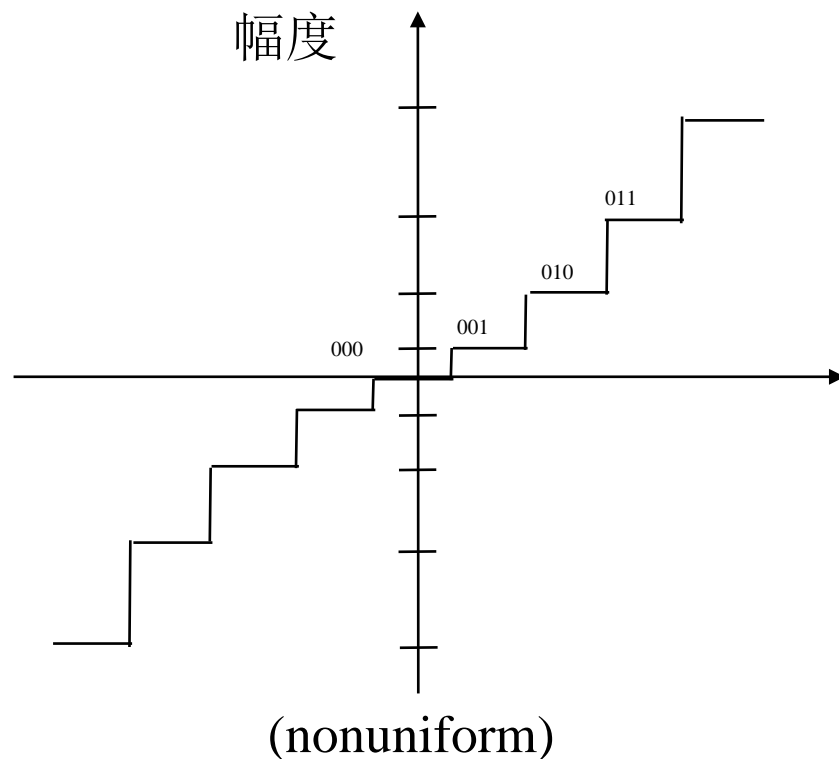
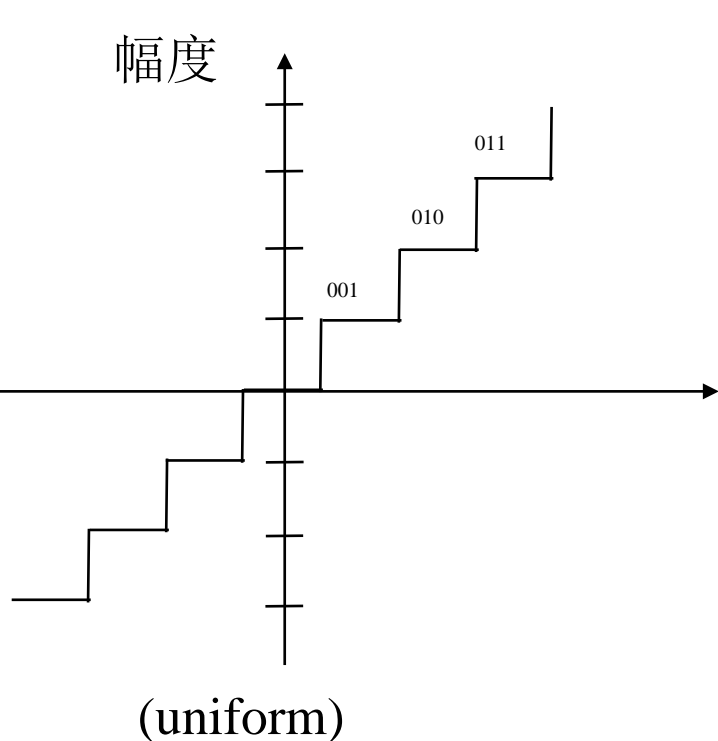
- 采样：在某些特定的时刻对这种模拟信号进行测量叫做采样(sampling)，由这些特定时刻采样得到的信号称为离散时间信号。每隔相等的一小段时间采样一次，这种采样称为均匀采样(uniform sampling)；
- 采样频率与采样定理
 - 奈奎斯特理论(Nyquist theory) 无损数字化(lossless digitization) $T \leq 1/2f_c$ 或 $f_c \leq 1/2T$
 - 采样频率标准有8KHz, 11.025KHz (AM), 22.05KHz (FM) 和 44.1KHz (CD) 三种, 根据奈奎斯特采样理论, 为了不让声音失真, 采样频率应该 $f \geq 40\text{KHz}$;



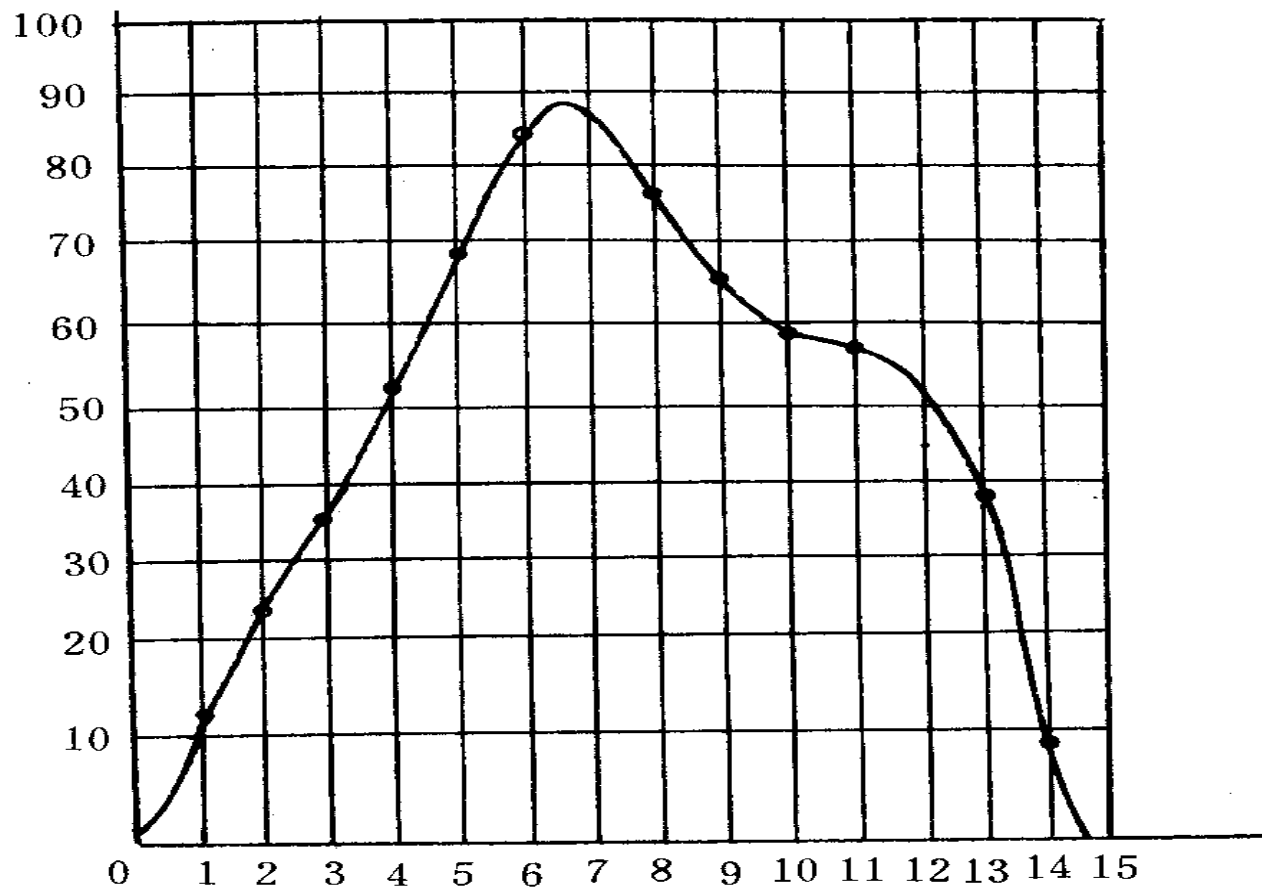
量化与量化精度

■ 量化与量化精度

- 连续幅度的离散化通过量化(quantization)来实现,就是把信号的强度划分成一小段一小段,如果幅度的划分是等间隔的,就称为线性量化,否则就称为非线性量化,采样精度可以用信噪比(signal-to-noise ratio, SNR)表示。



声音的采样以及量化图



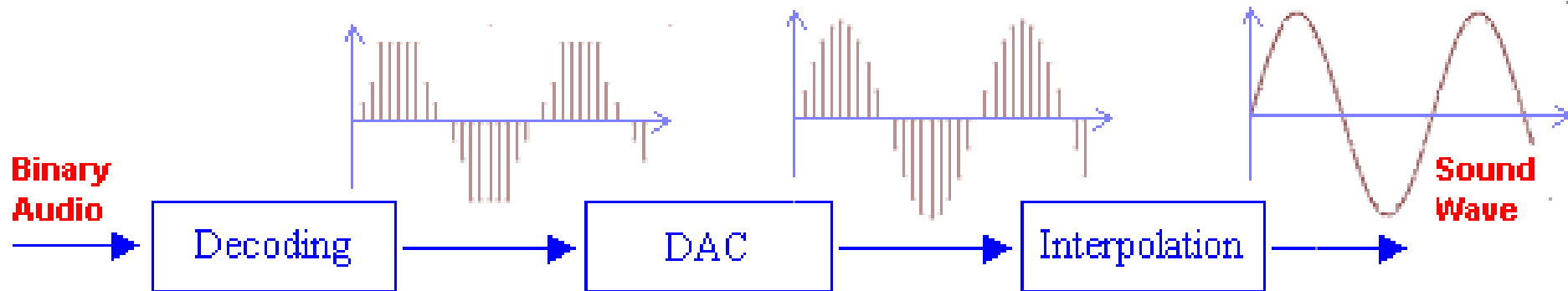
幅度
样点

12	22	35	52	69	85	87	76	66	59	58	54	38	09	0
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

采样及量化

4、声音重建与声卡 Reconstruction of Sound

- 1. 解码 Decoding
- 2. D/A转换 Dequantization (D/A conversion)
- 3. 滤波 Interpolation



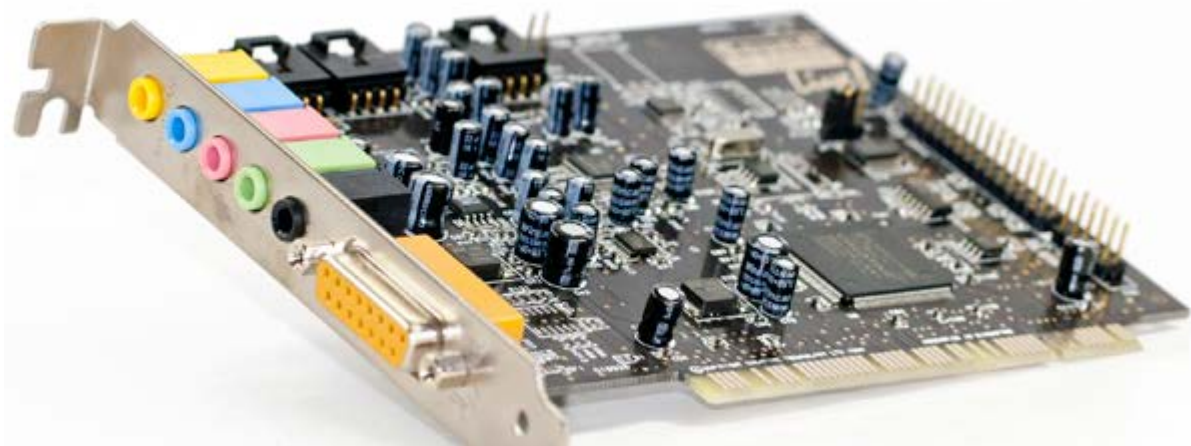
■ 声音硬件的类型

- 自然声音的获取设备
 - 声音卡
 - 麦克风
- 声音重建与播放设备
 - 声音卡
 - 音箱
- 合成声音的合成器
 - 声音卡（MIDI合成器）

音频卡的工作原理

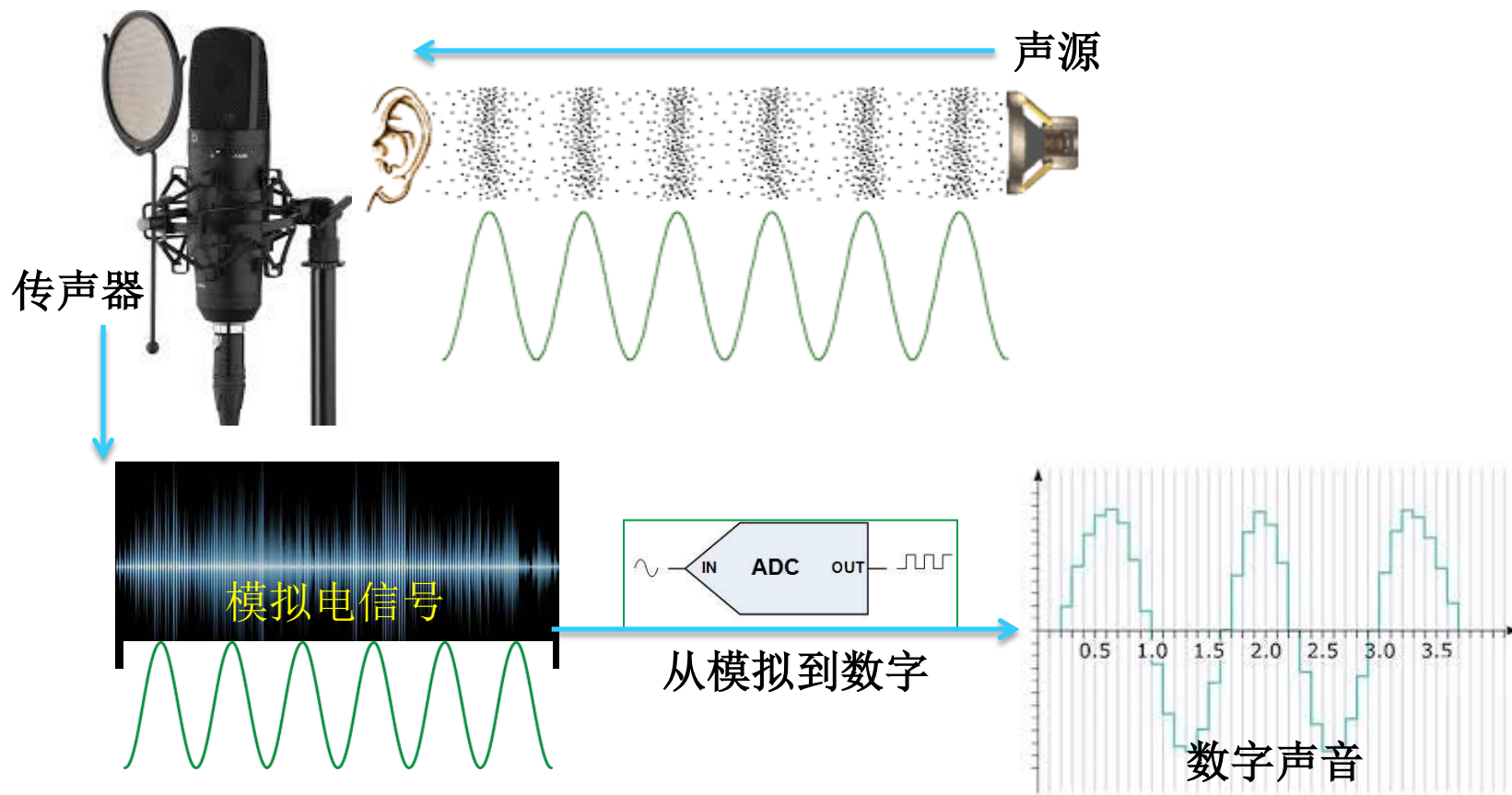
■ 音频卡

- 处理音频信号的PC插卡是音频卡（Audio Card），又称声音卡，声音卡处理的音频媒体有数字化声音（Wave）、合成音乐（MIDI）、CD音频。



音频卡的工作原理

■ 声音是怎样工作的？



音频卡的工作原理

■ 音频卡的功能

- 音频的录制与播放
- 编辑与合成
- MIDI接口
- 文 - 语转换
- CD-ROM接口
- 游戏接口
- 支持全双工功能

■ 用途：

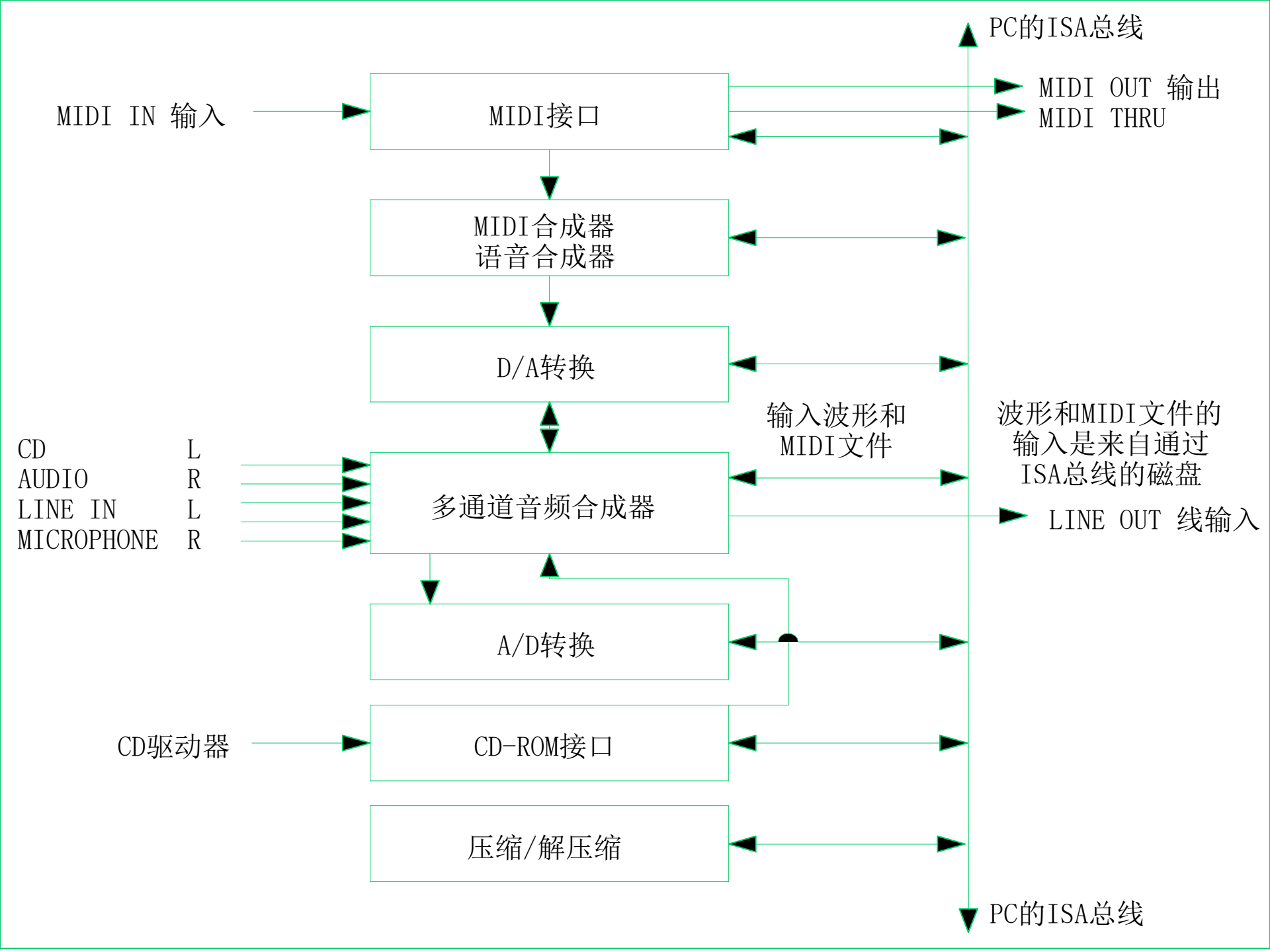
- 波形声音的获取、编码；
- 波形声音的重建、播放；
- MIDI声音的输入；
- MIDI声音的合成、播放；
- （CD-ROM 驱动器的控制，CD-DA声音的播放。）

音频卡的连接方式



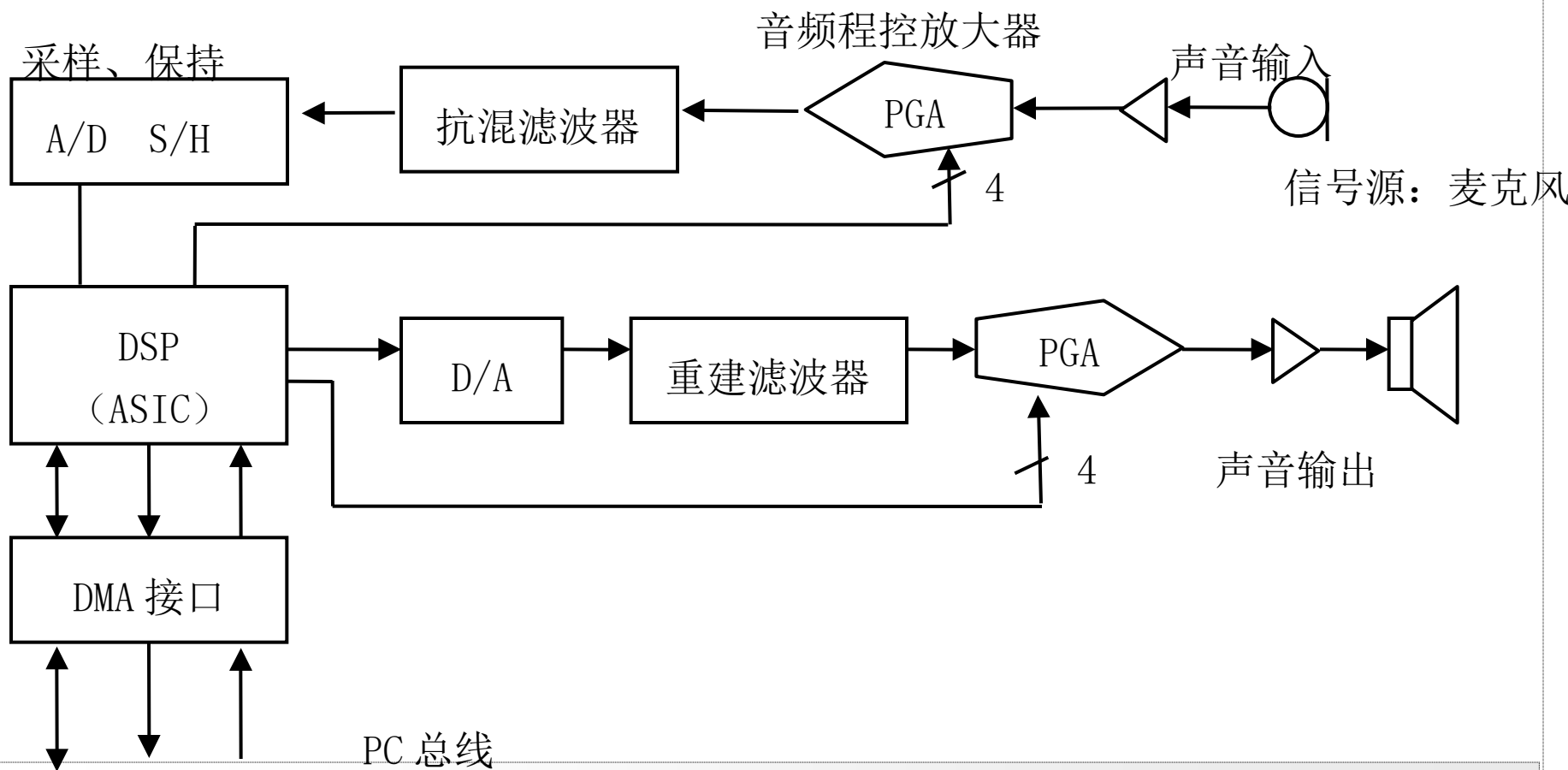
■ 音频卡的体系结构

- 音频卡由下列部件组成：
 - MIDI输入/输出电路；
 - MIDI合成器芯片；
 - 用来把CD音频输入与线输入相混合电路；
 - 带有脉冲编码调制电路的模数转换器，用于把模拟信号转换为数字信号以生成波形文件；
 - 用来压缩和解压音频文件的压缩芯片；
 - 用来合成语音输出的语音合成器；
 - 用来识别语音输入的语音识别电路；
 - 输出立体声的音频输出或线输出的输出电路等。

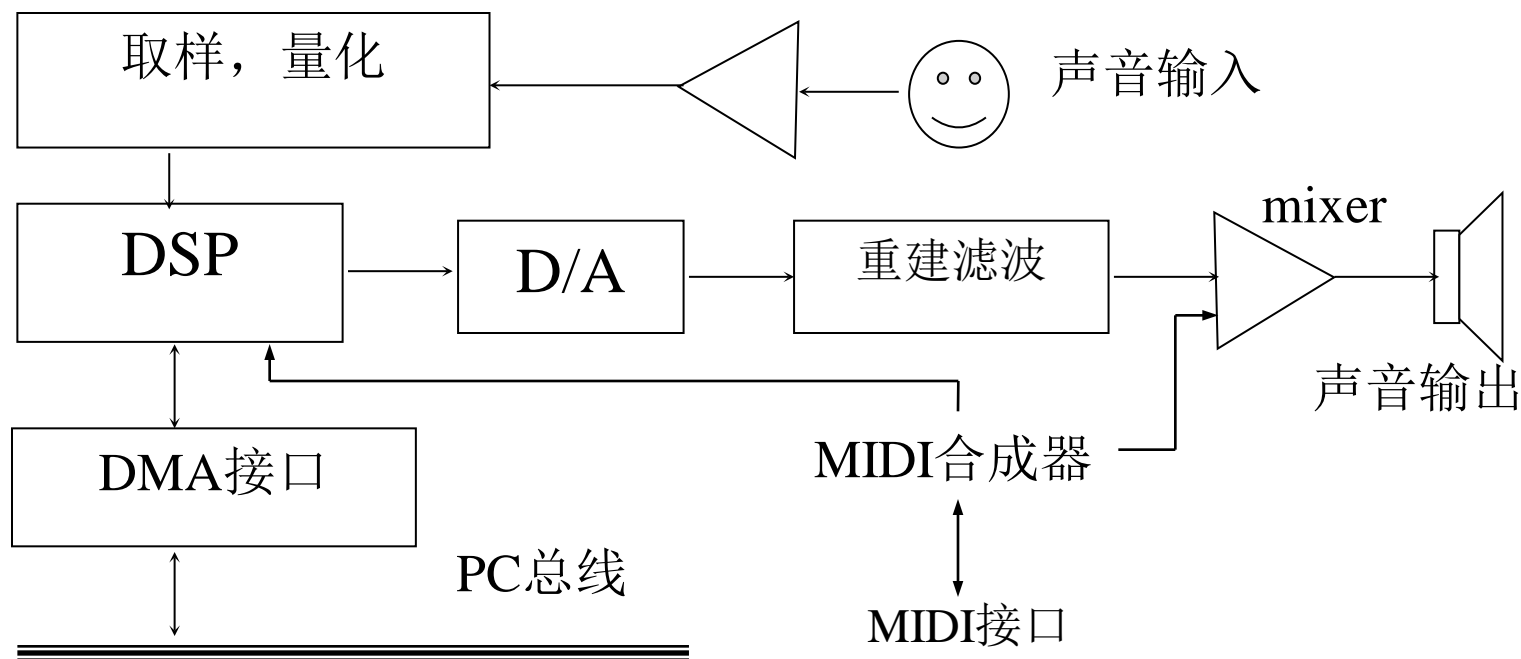


音频卡的工作原理

■ 数字化声音处理



以DSP为核心的声音卡原理图

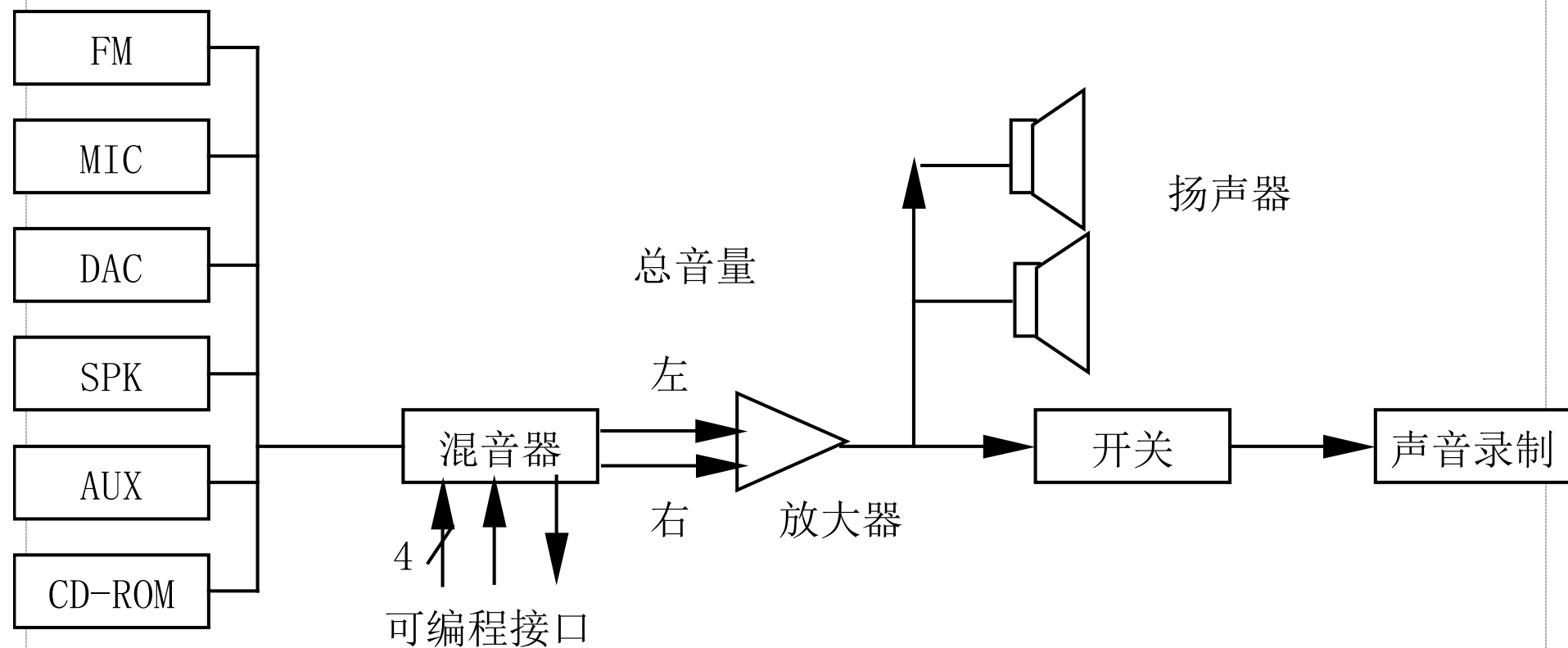


数字信号处理器DSP (Digital Signal Processor)

- 是一种专门用于信号处理的微处理器；能高速执行“乘加”操作，流水线处理方式，适合矩阵运算；通用，可编程；具体可分为：16位/32位/64位和定点/浮点
- 多个DSP组成并行处理系统，满足实时信号处理的要求。
- DSP的功能
 - 完成8位/16位、单声道/双声道的数字化声音的获取与重建
 - 完成不同压缩比的ADPCM编码；
 - 控制采样频率；
 - 提供数/模转换器(DAC)的控制；
 - 解释MIDI命令；
 - 控制不同方式的DMA数据传送。
 - 外置声卡上的DSP效果器和软件效果器插件功能类似于软音源和合成器音源。

音频卡的工作原理

■ 混音器



声卡的选择

- 支持16位、48kHz采样。
- 支持全双工，放音时还能录音，适合Internet电话。
- 输出功率（1-10W）。
- 支持波表合成（硬波表2MB-4MB/软波表）。
- 支持32或64个复音。
- 支持3D音效增强技术，模拟三维空间声效果（SRS, Qsound, APX等技术）。
- 与通信结合，如 Modem/Fax 等。

声音播放设备-----音箱

■ 分为无源音箱和有源音箱两大类

■ 有源音箱的组成：

- 扬声器
- 功率放大器
- 外壳

■ 主要性能指标：

- 功率；频率响应范围；
- 失真度；磁屏蔽；
- 静态噪声；扩展功能等.



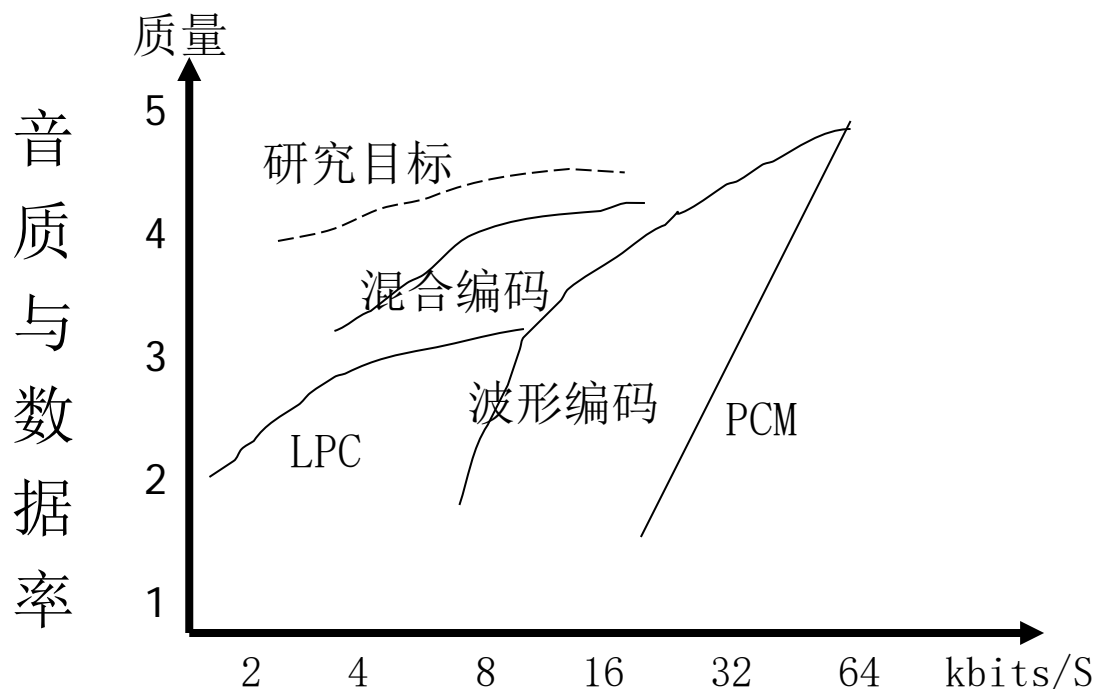
“创新多音箱环绕”
(Creative MultiSpeaker Surround)



5、音频的基本编码

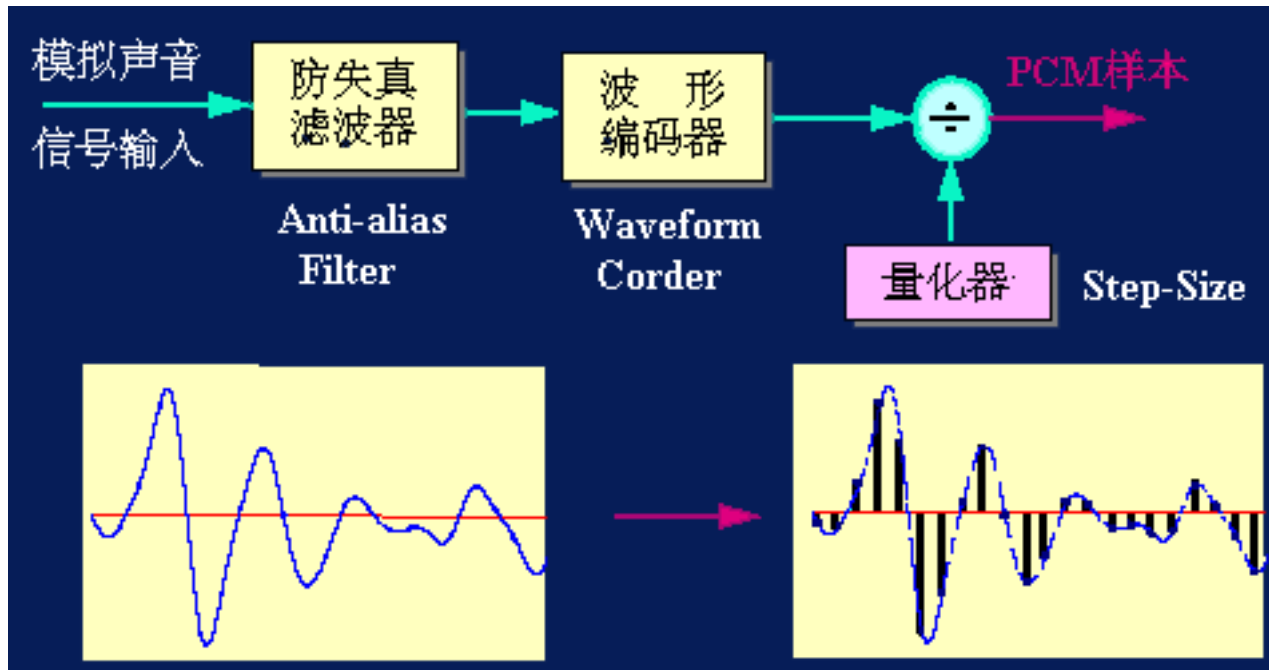
■ 音频编解码器可以分三种类型：

- 波形编解码器 (waveform codecs)：音质好，但数据率也很高；
- 音源编解码器 (source codecs)：数据率很低，产生的合成语音的音质有待提高；
- 混合编解码器 (hybrid codecs)：使用音源编解码技术和波形编解码技术，数据率和音质介于它们之间。



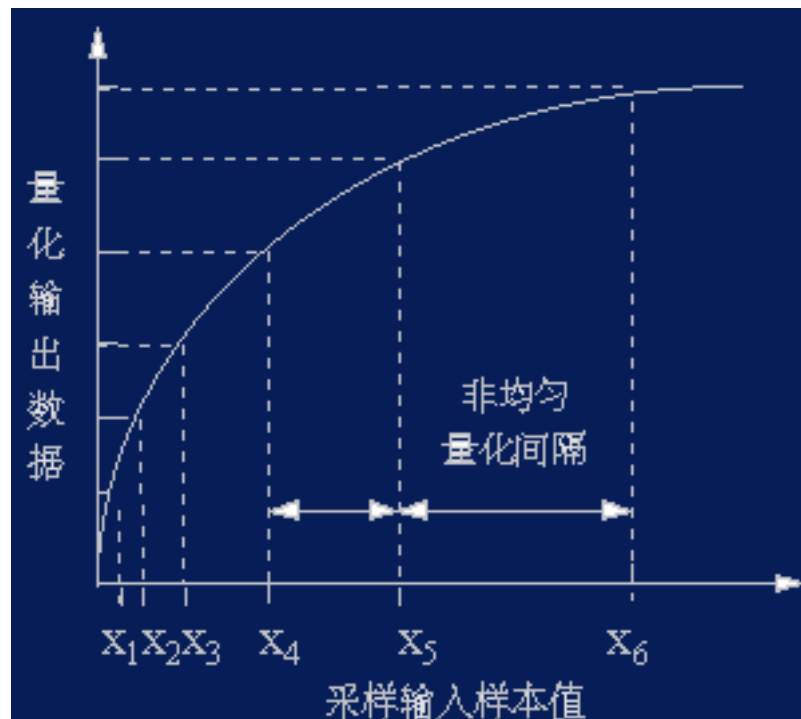
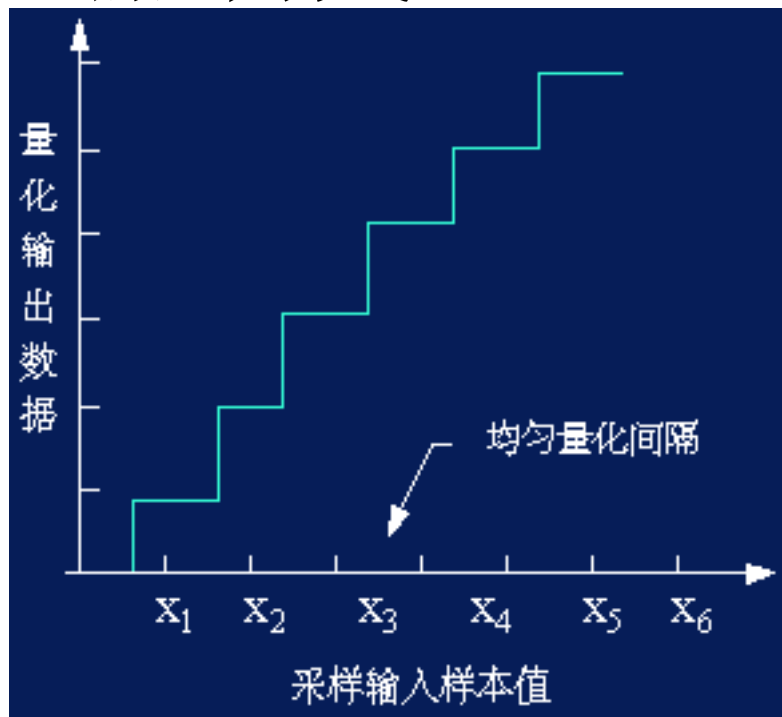
波形基本编码

■ 脉冲编码调制 (Pulse-code modulation, PCM)



波形的基本编码

■ PCM的量化方式



- 非线性量化的基本想法是，对输入信号进行量化时，大的输入信号采用大的量化间隔，小的输入信号采用小的量化间隔





6、波形声音的常用编辑操作的编程原理

- 从人与计算机交互的角度看，音频信号的处理包括下述3点：
 - 人与计算机通信，也就是计算机接收音频信号。包括音频获取、语音的识别和理解。
 - 计算机与人通信，也就是计算机输出音频。包括音乐合成、语音合成、声音的定位以及音频视频的同步。
 - 人-计算机-人通信。人通过网络与异地的人进行语音通信，相关的音频处理有语音采集、音频的编码和解码、音频的存储、音频的传输、基于内容的检索等。
- 音频的常用的处理实质是对一维或（多个一维）数组或矩阵的操作。常见的音频的操作有录音、分段与拼接、音量调整、回声、混响、延时、混音、淡入淡出等。

■ 录音

```
Fs=8000;  
n=5*Fs;  
y=wavrecord(n, Fs);  
wavwrite(y, Fs, ' test.wav' );
```

■ 播放、变调

```
a=wavread('sgyy.wav');  
 wavplay(a, 8000); wavwrite(a, 10000, 'sgyy2.wav');  
 wavplay(a, 10000);  
a=wavread('kw.wav');  
 wavplay(a, 8000); wavwrite(a, 6000, 'kw2.wav');  
 wavplay(a, 6000);
```

■ 分段与拼接(splicing & assembly)

🔊 `a=wavread('pray1.wav');`

🔊 `b=wavread('pray2.wav');`

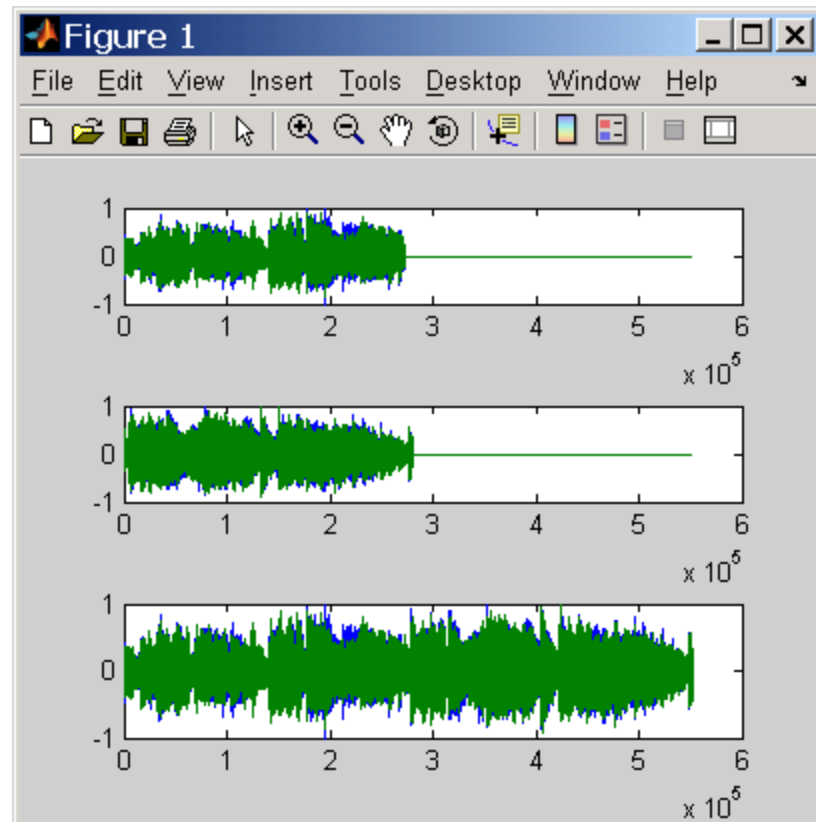
`c=[a;b];`

🔊 `wavplay(c, 22050);`

`subplot(311);plot(a);`

`subplot(312);plot(b);`

`subplot(313);plot(c);`



波形声音的常用编辑处理的编程原理

■ 音量调整(volume adjustment)

```
a=wavread(' am. wav' );
```

```
b=a.*2;
```

```
c=a./4;
```

```
Fs=22050;
```

🔊 `wavplay(a, 22050);`

🔊 `wavplay(b, 22050);`

🔊 `wavplay(c, 22050);`

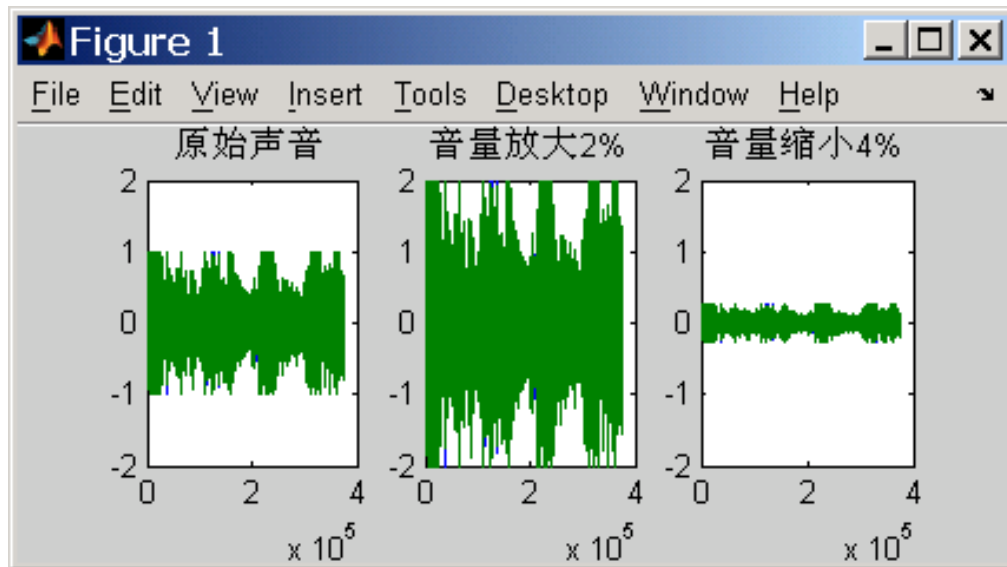
```
wavwrite(b, Fs, ' amb. wav' );
```

```
wavwrite(c, Fs, ' amc. wav' );
```

```
subplot(131);plot(a); title(' 原始声音' );
```

```
subplot(132);plot(b); title(' 音量放大2%' );
```

```
subplot(133);plot(c); title(' 音量缩小4%' );
```



■ 回声、混响、延时、混音

```
a=wavread('echo.wav');
```

```
Fs=22050;%n=0.05;t=f*n;
```

```
y0=[a;zeros(22.05*400,1)];
```

```
y1=[zeros(22.05*200,1);
```

```
    a.*0.5;zeros(22.05*200,1)];
```

```
y2=[zeros(22.05*400,1);a.*0.2];
```

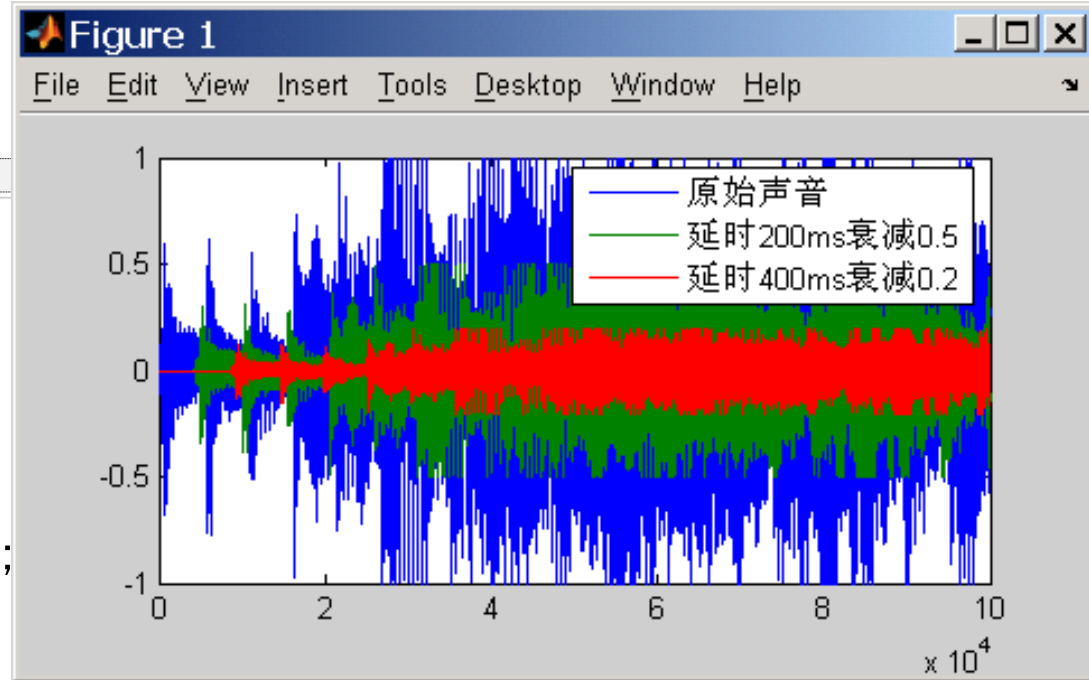
```
c=(y0+y1+y2);
```

🔊 `wavplay(c,Fs);L=size(y0,1);`


```
wavwrite(c,Fs,'echos.wav');
```


🔊 `plot(1:L,y0,1:L,y1,1:L,y2);`

```
legend('原始声音','延时200ms衰减0.5','延时400ms衰减0.2');
```



■ 淡入淡出

 `a=wavread('d1.wav'); b=wavread('d2.wav');`

 `x=[0:0.00001:1]'; y=1-x; Fs=11025; cz=[a;b];
ac=[a(1:size(a,1)-size(x,1));`

`y.*a(size(a,1)-size(x,1)+1:size(a,1))];`

`bc=[x.*b(1:size(y,1));b(size(y,1)+1:size(b,1))];`

`cf=[ac(size(ac,1)/2:size(ac,1));bc(1:size(bc,1)/2)];`

`wavwrite(cz,Fs,'cz.wav');wavwrite(cf,Fs,'cf.wav');`

 `wavplay(cz,Fs);`

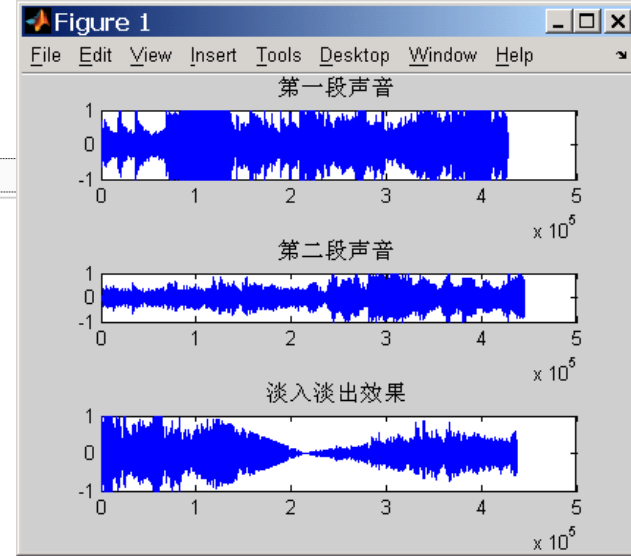
 `wavplay(cf,Fs);`

`subplot(311);plot(a);title('第一段声音');`

`subplot(312);plot(b);title('第二段声音');`

`subplot(313);plot(c);title('淡入淡出效果');`

%此外还有滤噪、向下取样、均衡化、时间扩展、颤音等



7、电子音乐合成与MIDI系统



- 电子乐器数字接口(Musical Instrument Digital Interface, MIDI) 是数字音乐的国际标准。用于在音乐合成器(synthesizers)、乐器(musical instruments)和计算机之间交换音乐信息的一种标准协议，它是音乐的符号化一种表示方式，MIDI的音乐符号化过程实际上就是产生MIDI协议信息的过程。音乐合成器是计算音乐系统中最重要的设备之一
- MIDI是一种电子乐器之间以及电子乐器与电脑之间的统一交流协议。从80年代初问世至今，它经历了长时间的发展，现已成为电脑音乐的代名词。可以从广义上将其理解为电子合成器、电脑音乐的统称，包括协议、设备等等相关的含义。
- MIDI标准的成熟出现各种电子乐器：键盘式（合成器、主控键盘）、弦控式（MIDI吉他）、敲击式（鼓机）、吹奏式（呼吸控制器）、音源模块（没有键盘的电子合成器）。

■ 音乐基础知识

- 声音分类：乐音和噪音
- 乐音的音质：基音和泛音
- 乐音的四要素：音高、音色、响度、时值

■ MIDI的形成

- 1982年，国际乐器制造者协会通过了美国Sequential Circuits公司的大卫.史密斯提出的“通用合成器接口”的方案，并改名为“音乐设备数字接口”，即“Musical Instrument Digital Interface”，缩写为“MIDI”，公布于世。
- 1985年11月，国际乐器制造者协会公布了《MIDI 1.0版的细节规定》。

MIDI的修订

- Though the MIDI Specification is still called "MIDI 1.0" there have been many enhancements and updates made since the original specification was written in 1984. Besides the addition of new MIDI messages such as the MIDI Machine Control and MIDI Show Control messages, there have also been improvements to the "basic" protocol, adding features such as Bank Select, All Sound Off, and many other new controller commands.
- Until 1995 there were five separate documents covering the basic MIDI specification, the additions (MSC & MMC), plus Standard MIDI Files and General MIDI. The "95.1" version -- published in January of 1995 -- compiled the latest versions of these documents together. The version of the basic MIDI specification (called the Detailed Specification) was version 4.2 prior to that time, which was itself a compilation of the "Detailed Specification v4.2" document and the "4.2 Addendum". Version 95.1 integrated the existing documents and fixed some minor errors in the various documents.
- Version 95.2 - September 1995
 - Added text for redefinition of Device-ID proposal (MMA-0015) which was approved since 4.2 and was missing from 95.1
 - Rewrote Universal SysEx ID description which was unclear
 - Moved EOF message (MMA-0011) from p44 into Sample Dump Standard Generic Handshaking Messages (P35-36)
 - Rewrote File Dump Handshaking Flags (p42-44) so as not to duplicate Sample Dump text on same message
 - Replaced all names referring to the Device ID message with the correct name "<Device ID>"
 - Moved MIDI Implementation Chart from back of section to before Tables
 - Rewrote the notes to Table 7 (SysEx Messages) to be more clear
 - Updated Table of Manufacturer ID's
- Current Version 96.1 - March 1996:
 - Changed Table 7 (SysEx Messages) to include reference to Universal SysEx messages and correct ID assignments
 - Fixed omission on Page 35 re: number of Generic Handshaking Messages
 - Added clarifications to SMF text on MIDI timing
 - Updated MSC Specification from 1.0 to 1.1

■ 音色排列方式的标准GS、GM和XG

- GS标准：是日本ROLAND公司制定并推出的。其生产开发的电子键盘、MIDI音源以及软波表都享有盛誉。所以GS颇具权威性，它完整的定义了128种乐器的统一排列方式，并规定了MIDI设备的最大复音数不可少于24个等详尽的规范。
- GM标准：则是在GS的基础上，加以适当简化而成的。由于它比较符合众多中小厂商的口味，成为了业界广泛接受的标准。
- XG标准：YAMAHA公司推出的标准——XG。与GM、GS相比XG提供了更为强劲的功能和一流的扩展能力，并且完全兼容以上两大标准。而且凭借YAMAHA公司在电脑声卡方面的优势，使得XG在PC上有着广阔的用户群。

■ MIDI的物理接口标准

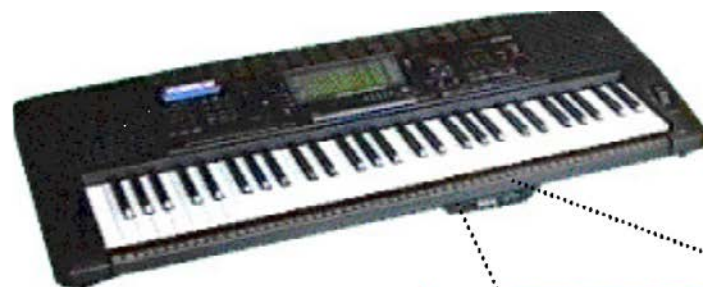
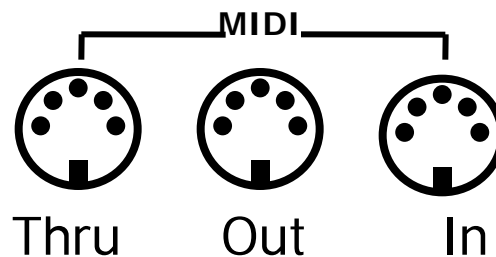
- 各个MIDI设备通过专用的串行电缆(MIDI线)连接, 并以31.25 KBPS 的速度传送着数字音乐信息。MIDI电缆使用5针DNI连接器, 其中仅用了3针(屏蔽线、当前循环线、数据传输线)

■ MIDI接口

- MIDI In (输入口)
 - 接收从其他MIDI装置传来的消息。
- MIDI Out (输出口)
 - 发送某装置生成的原始MIDI消息。向其他设备发送MIDI消息。
- MIDI Thru (转发口)
 - 传送从输入口接收的消息到其他MIDI装置。向其他设备发送MIDI消息。

音频合成和MIDI接口规范

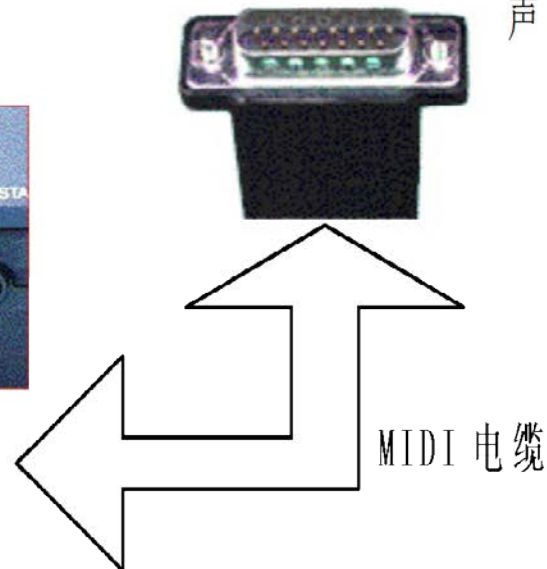
■ MIDI接口



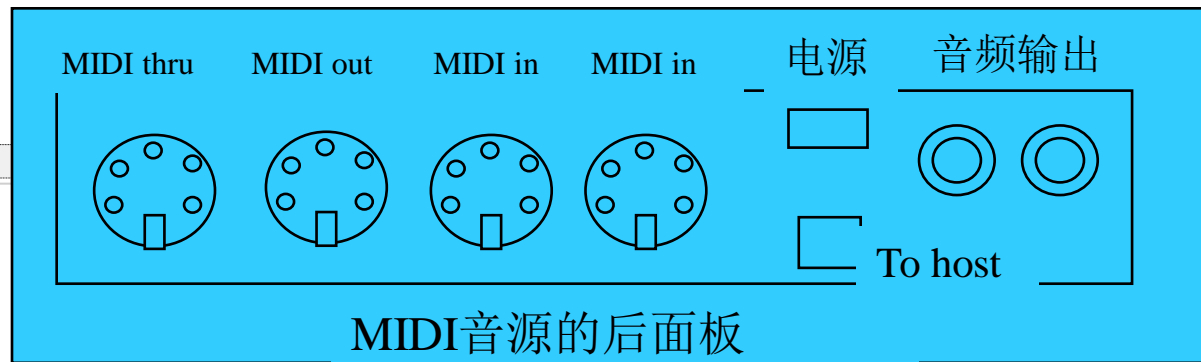
电子琴



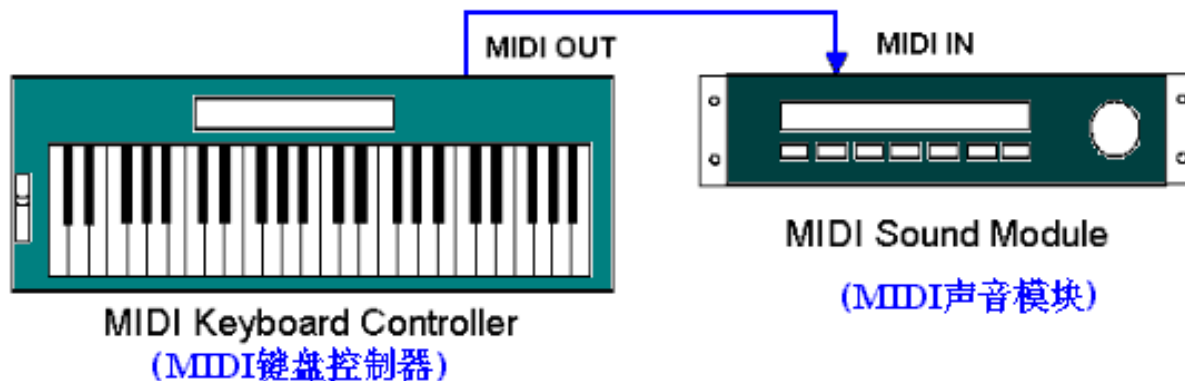
声音卡



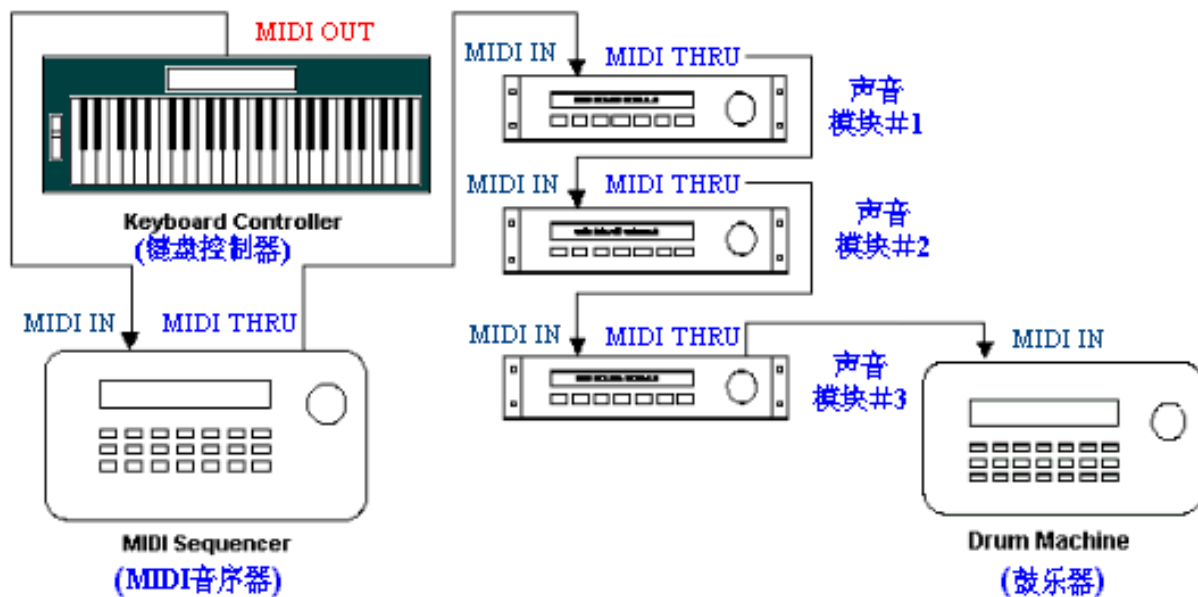
MIDI设备的连接



■ 简单连接



■ 复杂连接

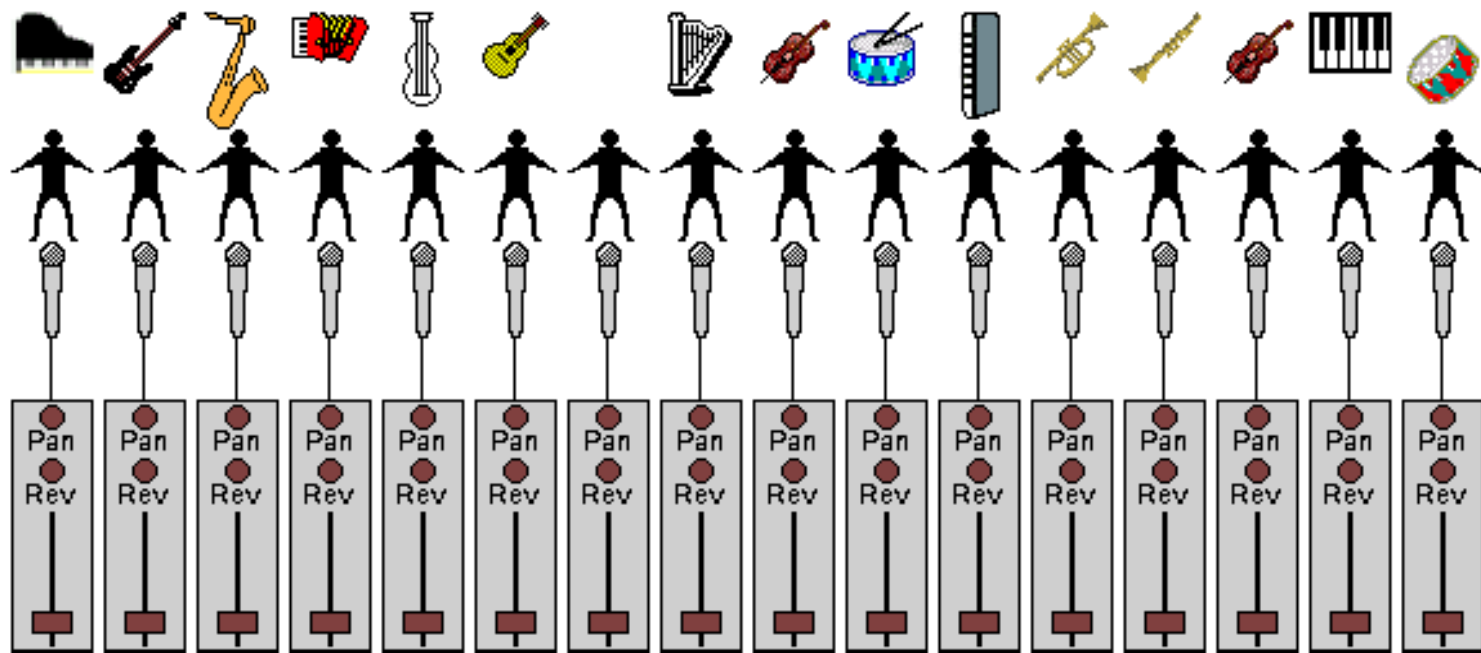


MIDI术语

- 音乐合成器 (Musical Synthesizer)：用来产生并修改正弦波形并叠加，然后通过声音产生器和扬声器发出特定的声音。泛音的合成决定声音音质。
- 复调声音：简称为复音 (Polyphony)，指合成器同时演奏若干音符时发出的声音。它着重于同时演奏的音符数。
- 多音色 (Timbre)：指同时演奏几种不同乐器时发出的声音。它着重于同时演奏的乐器数。
- MIDI电子乐器：能产生特定声音的合成器，其数据传送符合MIDI通信约定
- MIDI消息 (message) 或指令：乐谱的一种记录格式，相当于乐谱语言
- MIDI接口 (interface)：MIDI硬件通信协议。
- MIDI通道 (channel)：MIDI标准提供了16个通道，每种通道对应一种逻辑的合成器。
- MIDI文件：由控制数据和乐谱信息数据构成。
- 音序器 (Sequencer)：用来记录、编辑和播放MIDI文件的软件。

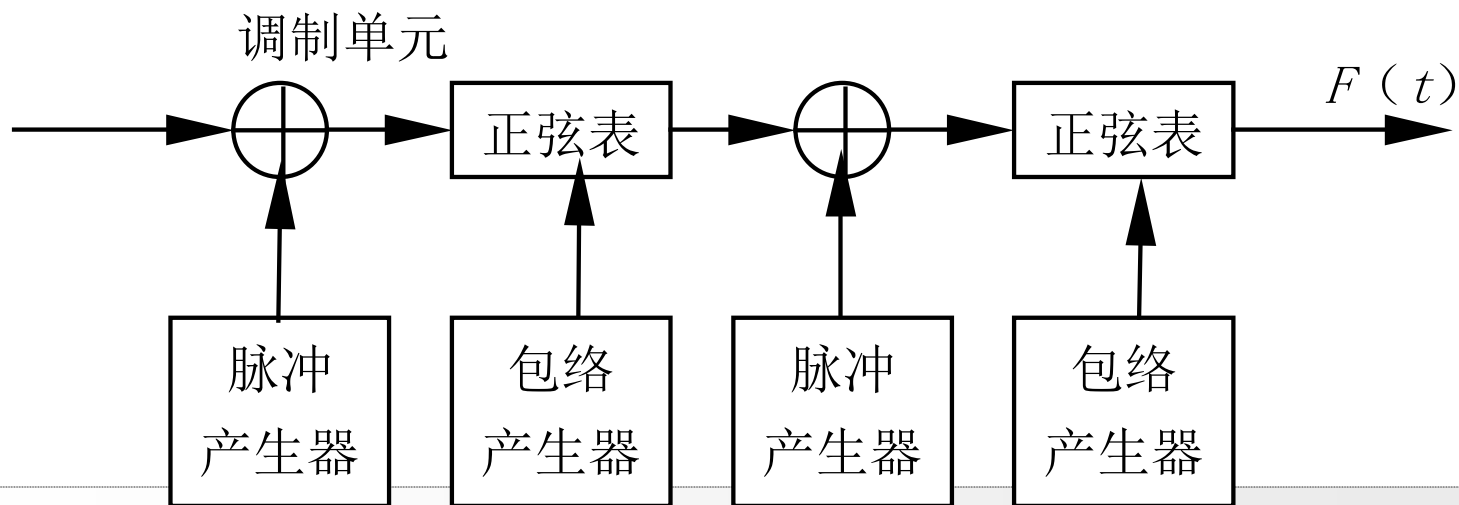
MIDI的通道概念

- 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, 如图2-11所示。在MIDI信息(MIDI messages)中, 用4个二进制位来表示这16个逻辑通道。音乐键盘可设置在这16个通道之中的任何一个, 而MIDI声源或者声音模块可被设置在指定的MIDI通道上接收。



音频卡的MIDI合成的工作原理

- **音乐合成技术：**产生乐音的方法很多，现在用得较多的方法有两大类：
 - 模拟合成法： 减法合成如滤波器，加法合成
 - 数字合成法： FM频率合成、 Wavetable波表合成、 LA线形合成、 AI先进集成式合成、 AV先进向量合成、 VAST可变结构合成技术等
- **合成器： 波形表（Wave Table）合成与频率调制FM合成**



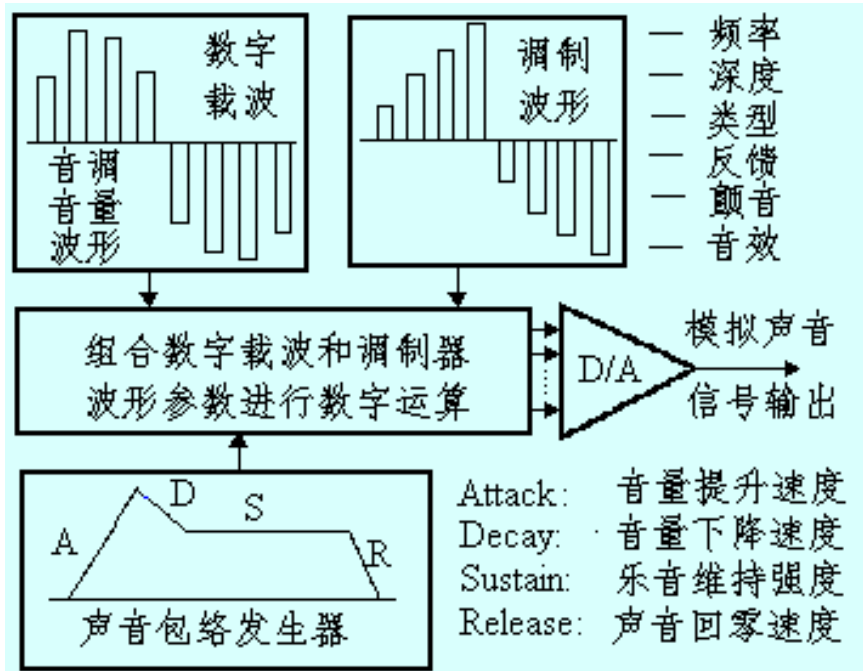
数字音乐合成技术

■ FM Synthesis, 调频合成法

- FM英文全名为Frequency Modulation（调频），这是早期所使用的技术。使用简单的硬件电路，利用几个乐器所产生不同的波形，定出取样频率、振幅，通过封装波形产生器和累加器，组合而成所需要的声音。不过由于取样的频率较少，加上乐器的数量有限制，所产生的声音效果并不好，以播放的感觉来说，仅有乐器发出的乐声间频率和音调上不同，很难去分辨之间的差别。

- 由五部分组成：

- 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

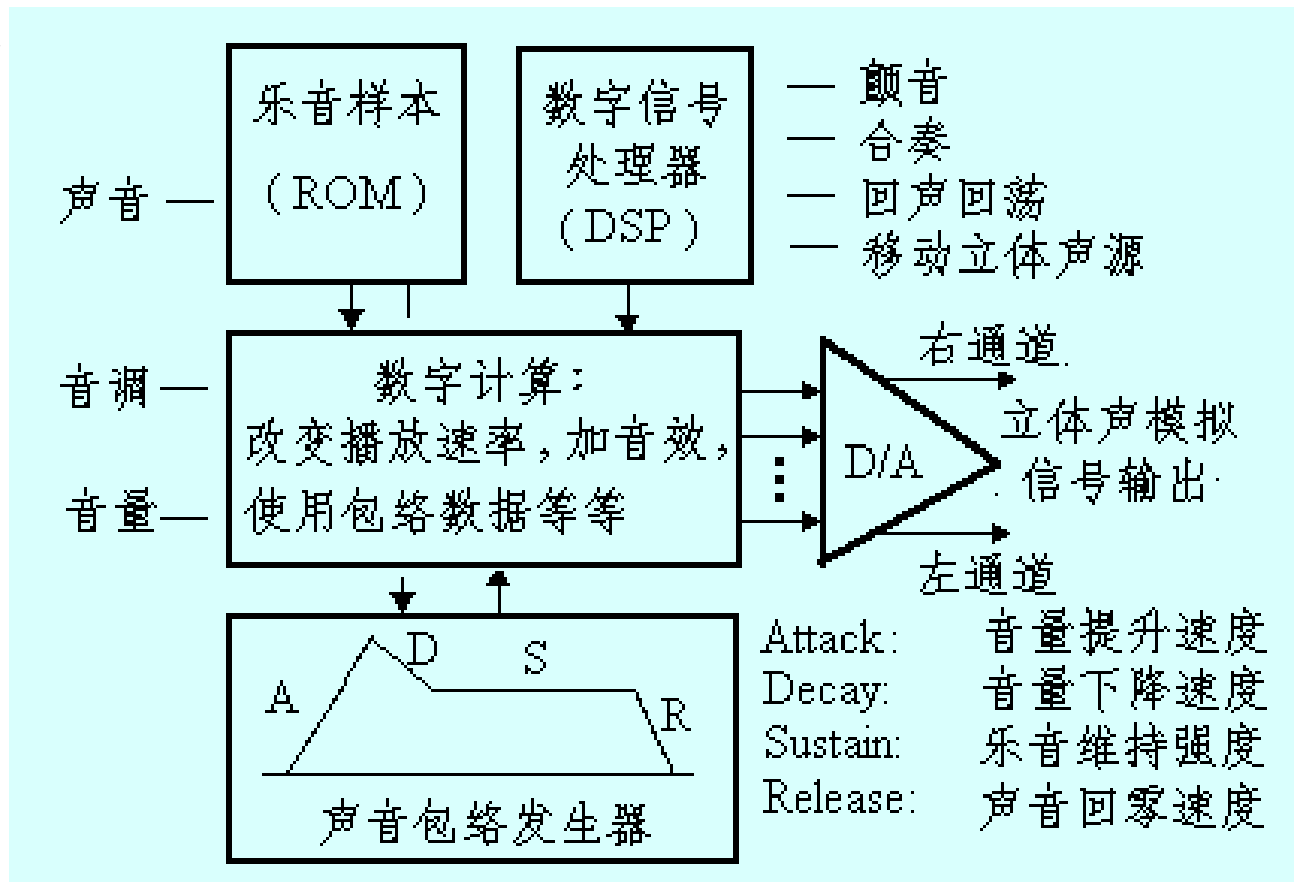


FM声音合成器的工作原理

- 乐音样本合成法，也称为Wavetable Synthesis，波表合成法
 - Wavetable的技术，是将每种乐器的声音录制取样一个或多个循环，加以适当的处理后存储成音色文件，记录在合成器的内存当中，当需要播出某个乐器的声音时，合成器就从内存中找出音色，同时并播放出来。Wavetable的声音播放会比较真实，而且可以做比较多的音效变化。
 - Wavetable的衡量标准：波表库容量、复音数、特殊效果
- 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。

Wavetable

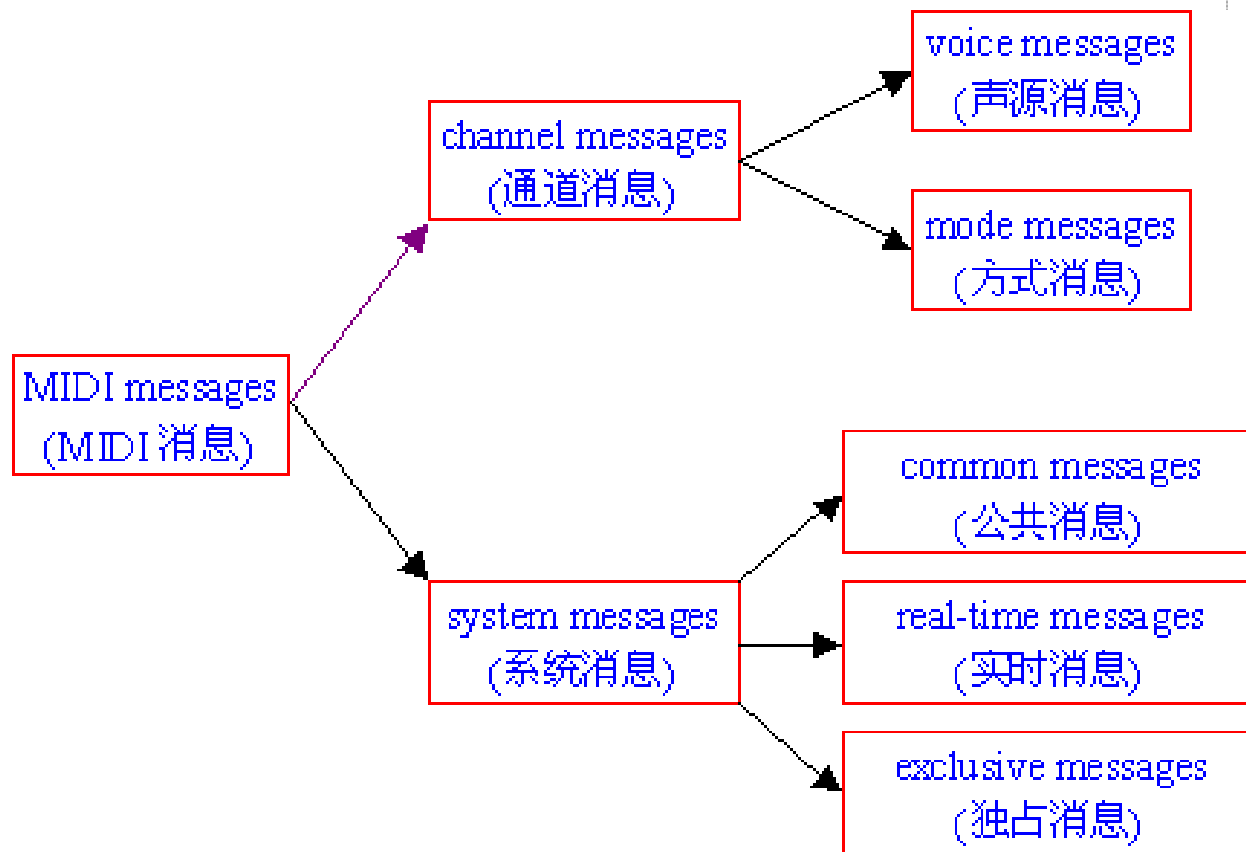
- 电子合成器 (Synthesizer), 是利用波形组合或声音取样来发音的键盘乐器, 它可以将其中的声音加以改造、重组并创造新的音色。



Wavetable合成器的工作原理

MIDI的数据标准

- MIDI设备使用的一系列MIDI音符，可被认为是告诉音乐合成器如何播放一小段音乐的指令。因为MIDI数据是一套音乐符号的定义，而不是实际的音乐声音，因此MIDI文件的内容被称为MIDI消息(MIDI messages)。一个MIDI消息由1个8位的状态字节并通常跟着2个数据字节组成。在状态字节中，最高有效位设置成“1”，低4位用来表示这个MIDI消息是属于哪个通道，4位可表示16个可能的通道，其余3位的设置表示这个MIDI消息是什么类型的消息。



MIDI消息

MIDI Message

- 任何电子乐器，只要有处理MIDI消息的微处理器和合适的硬件接口，就构成了一个MIDI设备。当一组MIDI消息通过音乐合成芯片处理时，合成器能解释这些符号并且产生音乐。
- MIDI消息作用： 描述乐曲的乐谱及演奏要求，控制MIDI音源进行演奏
- 数据格式：
 - 1个状态字节(最高位为“1”)+ n个数据字节(最高位为“0”)
- 分类：
 - Channel message (带channel号，只对指定通道起作用)
 - voice message 实际的演奏数据, 控制乐器的发声
 - mode message 决定乐器对声音消息如何响应
 - System message (对全部通道都起作用)

MIDI message举例

■ note on :

90	3c	40
----	----	----

(音符开始)

通道号

键号

速度

■ note off :

90	3c	00
----	----	----

(音符结束)

CH#

key#

speed

■ 音符# 0 12 24 36 48 60 72 84 96 108 120 127

■ 音阶 C₋₁ C₀ C₁ C₂ C₃ C₄ C₅ C₆ C₇ C₈ C₉ C₁₀

■ 击键力度 0 1 , , , , , , , 64 , , , , , , , 127

■ off ppp pp p mp mf f ff fff

MIDI的特点

- 与波形声音相比，MIDI不是声音数据而是指令，所以数据量要少得多。30分钟的音乐，用MIDI文件记录只需200KB，用16位CD品质的未压缩WAV文件记录需317MB
- MIDI可以与其他波形声音配合使用，形成伴乐的效果。而两个波形声音一般是不能同时使用的
- 对MIDI的编辑也很灵活，用户可以自由地改变音调、音色等属性，直到自己想要的效果
- MIDI在音质上还不能与真正的乐器完全相似。无法模拟自然界中其它非乐曲类声音
- MIDI音频文件协议以及音频形态与存储方法。MIDI音频文件中不包含音频的波形，存储空间小，MIDI音频文件包括描述声音的键、通道号、音量、音色、音长和击键强度等。

8. 语音合成 (Speech synthesis)

- 语音合成，即计算机言语输出，它是一门跨学科的前沿技术，涉及到下列彼此相关的各个领域：自然的语言理解、语言学、信号处理、心理学及声学等。语音合成也是非常重要的智能接口技术，特别是通过文语转换技术可以让计算机朗读文章，因而受到了很大重视。可用于使用听觉媒体补充和部分替代显示输出等视觉媒体；解决语音信号传输的大数据量问题。其关键性能有两个：
 - **正确**：指文字的读音要正确，保证这一点的难度在于一个字常常有几个读音，到底那个读音正确要根据组词甚至前后文来判断。例如不能将“银行”的“行”读成“xing2”。为保证正确性，必须先对句子进行分词。这一点西文有着得天独厚的优势，因为词与词间有空格分离，而对汉语的句子进行分词却不是简单的事。
 - **自然**：合成的语音要让人能听得懂听得舒服，还必须要有较高的自然度。即读出来的文章韵律和节奏要比较准确。要做到自然，常常需要对句子进行分析和理解，知道哪儿重、哪儿轻、何时急、何时缓。

■ 合成方法

- 发音器官参数语音合成：对人的发音过程进行直接模拟
- 声道模型参数语音合成：基于声道截面积函数或声道谐振特性合成语音
- 波形编辑语音合成技术：波形编辑语音合成技术是直接把语音波形数据库中的波形相互拼接在一起，输出连续语流，PSOLA (Pitch Synchronous Overlap Add) 方法

■ 语音基元数据库的构建

- 基元的选择
 - 选择音节
 - 选择双音素和三音素
- 语音数据的存储形式
 - 波形存储方式存储：数字化的语音波形数据
 - 参数存储方式存储：从语音信号中提取的参数，常用的有LPC参数、LSP (LSF)、共振峰参数等

■ 韵律模拟

- 自然语言中的韵律特征
 - 语调、节奏和重音等能表达说话者的语义和感情，是自然语流的重要组成部分
- 韵律合成及方法
 - 超音段特征（音高、音长、音强及频率分布的变化）的修改构成了韵律合成的基础
 - 方法：修改基频模式、共振峰模式、PSOLA算法等
- 韵律模拟的问题
 - 需解决韵律规则、韵律描述、计算模型和修改算法等问题

■ 几个关键问题

- 语音基元数据库的构建；
- 自然语流中韵律的模拟；
- 语言理解与语音生成的结合；
- 输出言语的自然度评价标准

- 音素(phoneme)是语音的最小单位。音素分为：
 - 元音(vowel)（浊音），不受声道阻碍的音。
 - 辅音(consonant)（浊音或清音），受声道阻碍的音。
- 不同的音素各有其不同的参数。
 - 基频、3~5个共振峰(formant)；（共振峰是语音信号频谱包络线的峰值，从低频到高频方向记为F1, F2, F3...。）
- 英语语音
 - 每字(词)一个或几个音节(syllable)（多音节字）；音节由一个或几个音素组成；英语的音素（元音20个，辅音28个）
- 汉语语音
 - 每字一个音节(syllable)（单音节字）；音节由一个或几个音素组成
 - 汉语的音素：元音42个(单元音13，复元音13，复鼻尾音16)；辅音22个；（或者分为：声母21个，韵母39个）

■ 汉语语音的三要素：

- 声母(21)
- 韵母(39)
- 音调(4个：阴平、阳平、上声和去声)

■ 汉语语音的数目：

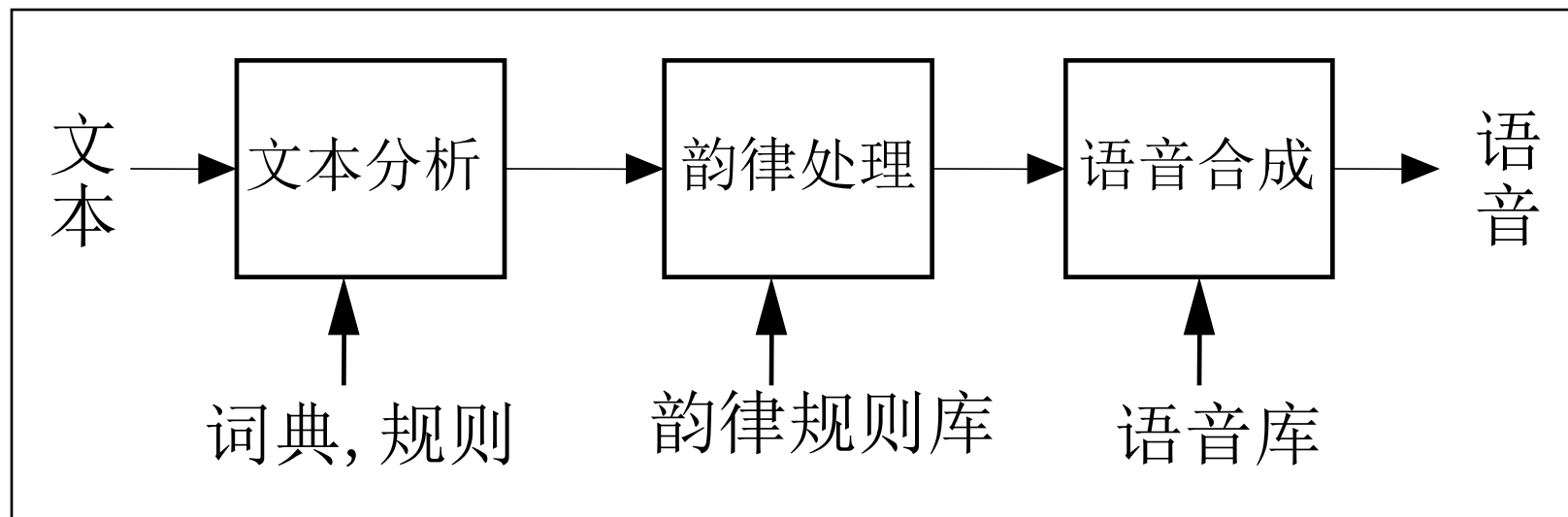
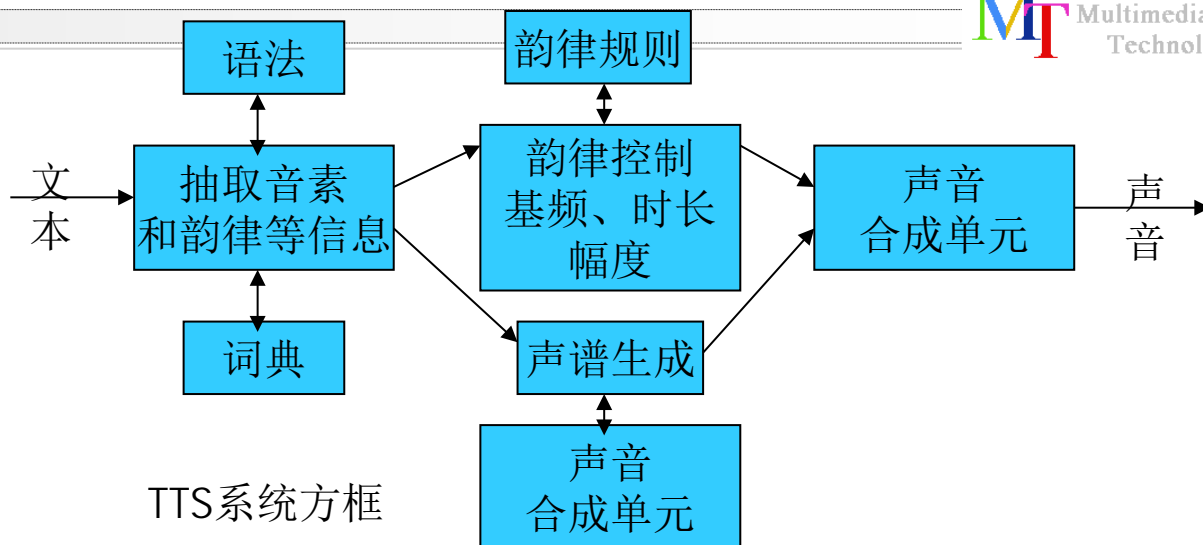
- 无声调的音节数目：412个
- 带声调的音节数目：1282个

■ 汉语语音的特点

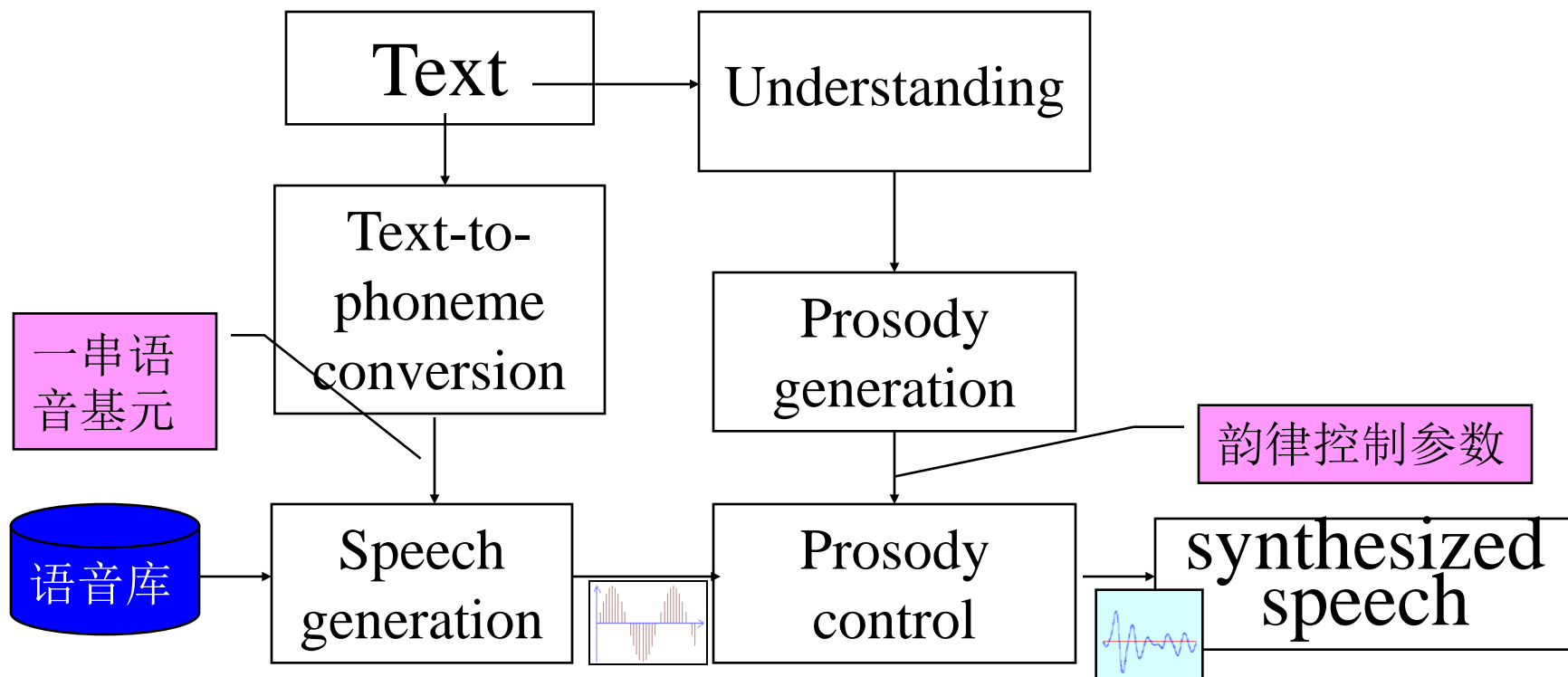
- （1）音系简单。即音节少、音素少。汉语普通话每个字的语音都是单音节字，每个字音虽为多音素，但一般只含1～4个音素。音素是语音的最小单位。
- （2）听感清亮、柔和。这是因为：清辅音多，所以没有快促之感；没有入声短促发音，所以使高频成份较多；开口音节多，所以听感好
- （3）含有鲜明的轻重音和儿化韵。从而使字词分隔清楚，语言表达准确而丰富。

文语转换

- 第1步：文本分析，
- 第2步：韵律处理，
- 第3步：语音合成。



文语转换过程



In 1992, CNET in France, pitch-synchronous overlap-and-add (PSOLA) method.

In 1996, ATR in Japan developed the CHATR speech synthesizer.

■ 第1步：文本分析

- (1) 将输入的文本规范化。查找拼写错误，过滤掉文本中出现的一些不规范或无法发音的字符。
- (2) 分析文本中词或短语的边界，确定文字的读音，同时分析文本中出现的数字、姓氏、特殊字符、专有词语以及各种多音字的读音方式
- (3) 根据文本的结构、组成和不同位置上出现的标点符号，确定发音时语气的变换以及不同音的轻重方式。(统计学方法及人工神经网络技术)

■ 第2步：韵律处理

- 分析并决定各个音节的声调、语气和停顿方式，发音的轻重、长短等，这些都属于韵律特征。早期的韵律生成方法采用基于规则的方法。目前通过神经网络或统计驱动的方法进行韵律生成已获成功。

■ 第3步：语音合成

- 主要功能：根据韵律控制参数，从原始语音库中取出相应的语音基元，利用特定的语音合成技术对语音基元进行韵律特性的调整和修改，最终合成出符合要求的语音。

(I) 参数合成法

- 根据语音生成的“声道—滤波器”模型，控制激励源和滤波器的参数（一般每隔10ms-30ms一组参数），就能灵活地合成出各种语音。
- 预先录制涵盖所有可能的读音；然后提取出这些声音的声学参数，并整合成一个完整的音库。
- 在发音过程中，先从音库中选择合适的声学参数，再根据韵律参数，通过合成算法产生语音。
- 代表性系统：美国DEC公司的DECtalk（1987 年），发音清晰，可产生7种不同音色的声音。
- 优点：音库一般较小，能适应的韵律特征的范围较宽。
- 缺点：准确提取共振峰参数比较困难，合成语音的音质难以达到实用要求。

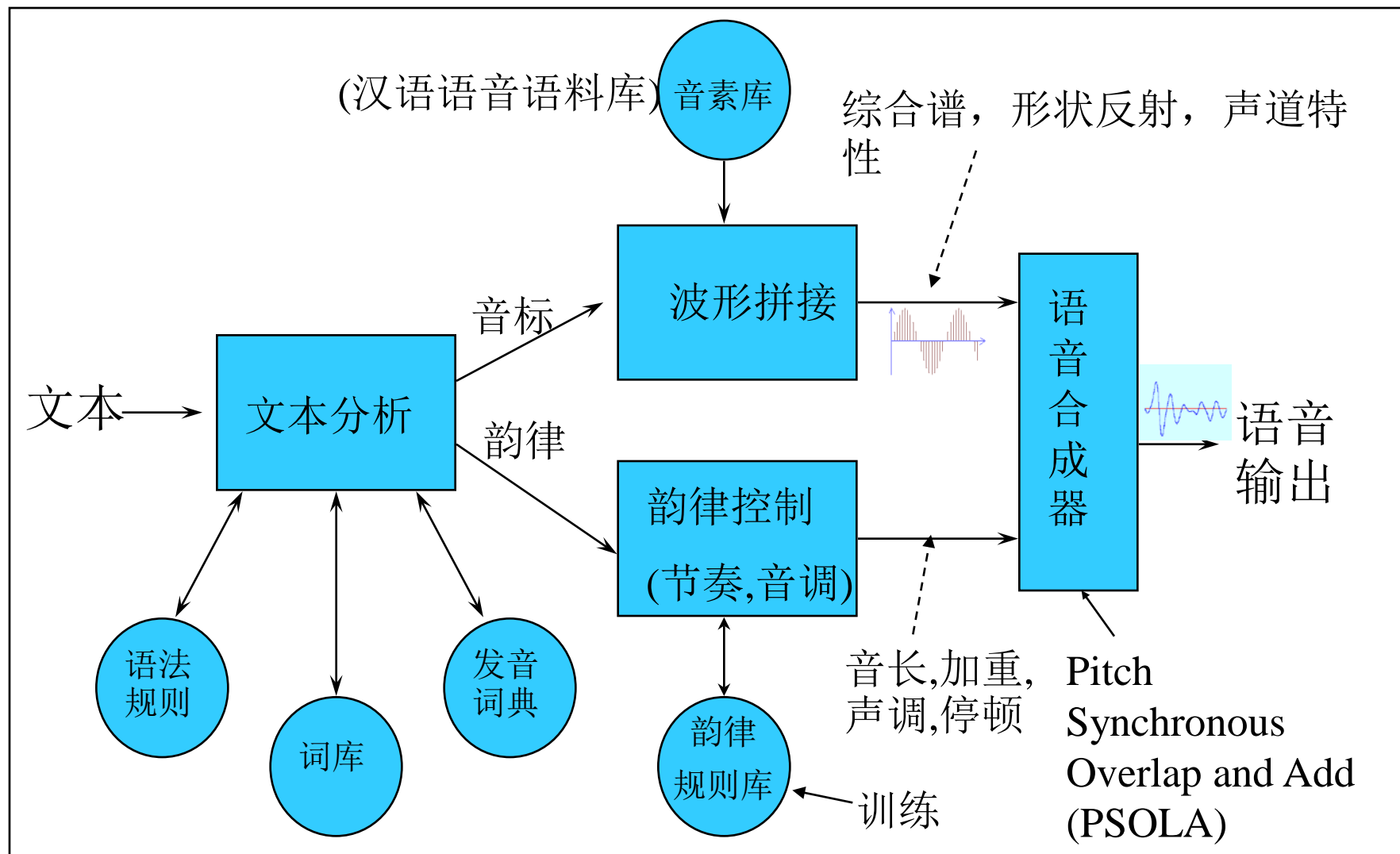
(II) 波形拼接法

- 基本思想：预先存储语音的基元（单音或词组的波形），合成时读取基元，进行拼接和韵律修饰，然后输出连续语流。
- 优点：由于语音基元取自自然语音的词或句子，它隐含了声调、重音、发音速度变化时的细微特性，合成的语音清晰自然，其质量普遍高于参数合成法。
- 缺点：韵律参数修改范围受限。
- 波形拼接法需考虑的问题
 - 语音基元的选择。语音基元是指拼接的基本单位。它可能是音素、双音子（Diphone）、三音子（Triphone）、半音节（首音、尾音）、音节、词语、语句等。基元越小，语音数据库越小，拼接越灵活，韵律修饰的规则就越复杂。
 - 语音基元的样板数。对于同一个基元，由于语境不同和重音表现不同，其声学特征有很大差别。为了减小韵律修饰的负担，可以建立多样板语音数据库。

基音同步叠加技术(PSOLA)

- 波形拼接法—— PSOLA技术：1992年提出基音同步叠加技术（Pitch Synchronous OverLap and Add, PSOLA）：
 - 首先在语音库中选择最合适的语音单元，并在选音过程中采用多种复杂的技术，包括统计学方法（如HMM）或ANN技术，
 - 拼接时，对拼接单元的韵律特征进行调整，使合成波形既保持了原始发音的主要音段特征，又能使拼接单元的韵律特征符合上下文的要求，从而获得很高的清晰度和自然度。

文语转换器框图---PSOLA方法



(III) 基于数据驱动的语音合成

- 方法:语音数据库非常大(包括各种可能语境下的语音单元), 以尽量多的语音基元样板来满足韵律的需求. 语音合成时, 从庞大的语音数据库中进行挑选, 不需要韵律修饰功能。
- 优点:只要语音数据库足够大, 就有可能拼接出任何语句。由于合成的语音基元都是来自自然的原始发音, 合成语句的清晰度和自然度都将非常高。
- 例:中国科大KD-86, KD-2000汉语文语转换系统.

(IV) 基于深度学习的语音合成

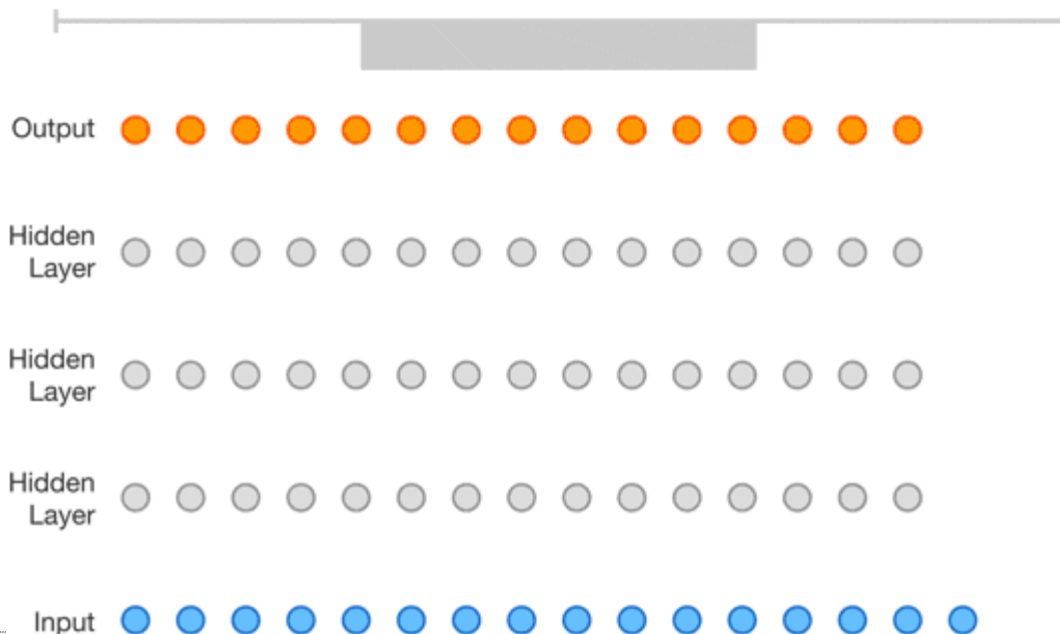
■ DeepMind基于深度学习的原始语音生成模型(WaveNet)

WaveNets是一种卷积神经网络，能够模拟任意一种人类声音，生成的语音听起来比现存的最优文本-语音系统更为自然，将模拟生成的语音与人类声音之间的差异降低了50%以上



WaveNet音频建模

1 Second



WaveNet模型的内部结构

(V) 可视语音合成

- 可视语音合成(Visual Speech Synthesis, VSS)是指人们在使用语言交流时所表达出的面部表情和动作，它能在一定程度上传达人们想要表达的意思，并能帮助人们加深对语言的理解。
- 分类：
 - 基于文本驱动的可视语音合成 (Text to Visual Speech)
 - 基于语音驱动的可视语音合成 (Speech driven Face Animation) 。
- 应用:虚拟现实、虚拟主持人、虚拟会议、电影制作、游戏娱乐等。
- 重点难点:语音与人脸的同步映射模型的建立

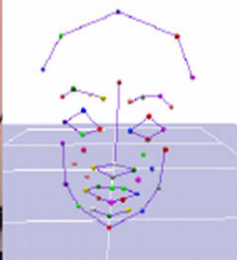
VSS的构建步骤

■ 建立一个audio-visual 多模态数据库;

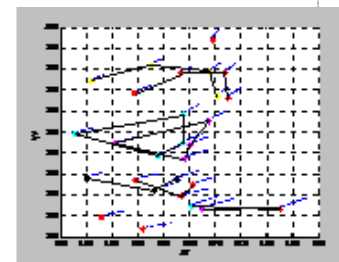
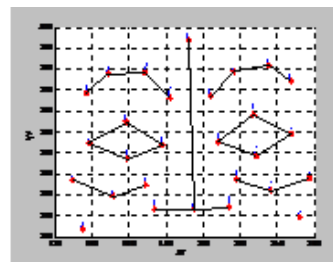
- 1) 录制二维连续视频, 从中提取感兴趣区域;
- 2) 通过三维激光扫描仪截取静态视位;
- 3) 从标记特征点的人脸上获取三维特征点坐标运动信息, 直接使用运动实时捕获设备获取标记特征点的三维运动坐标



运动捕获器

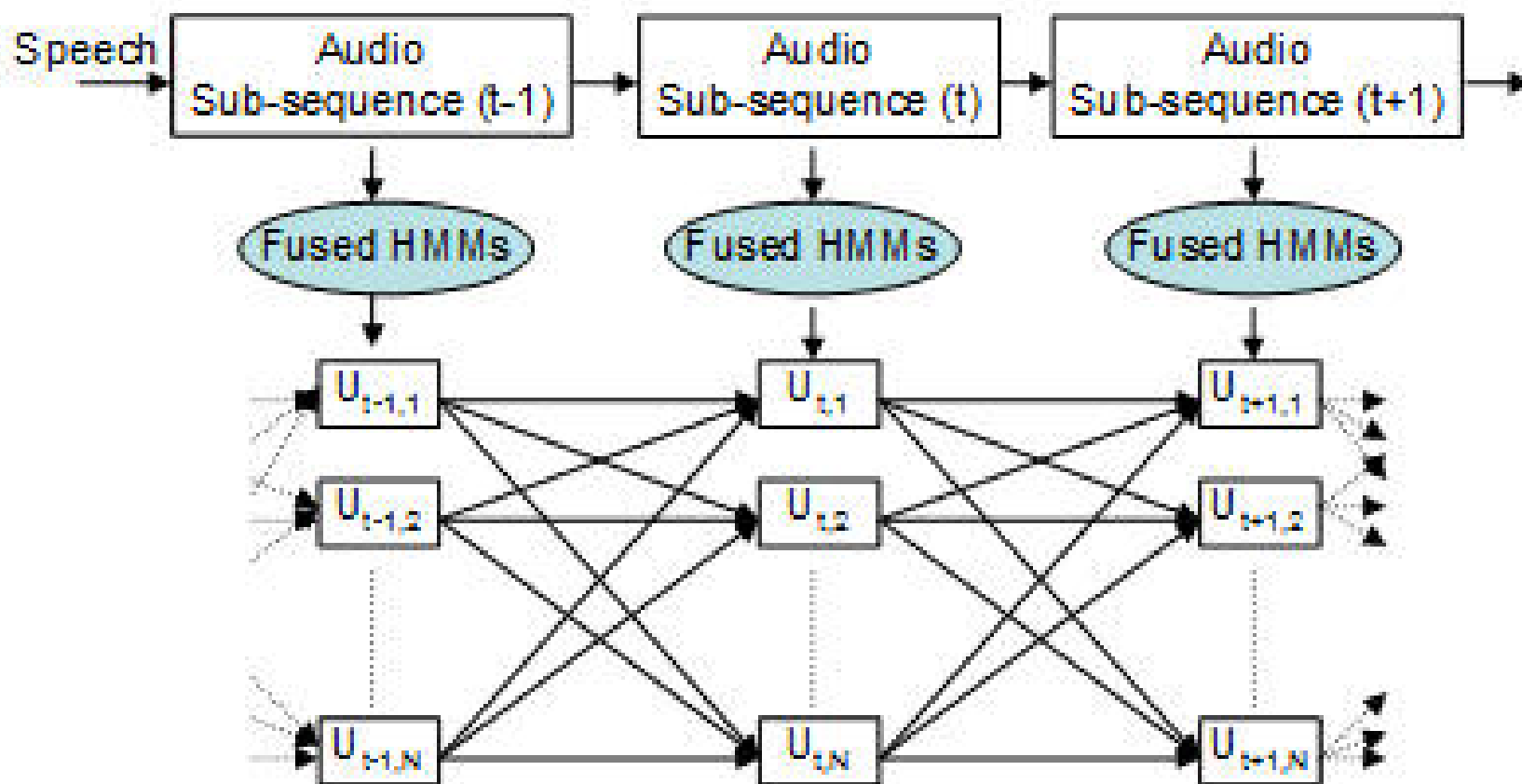


人脸标记 人脸运动捕获

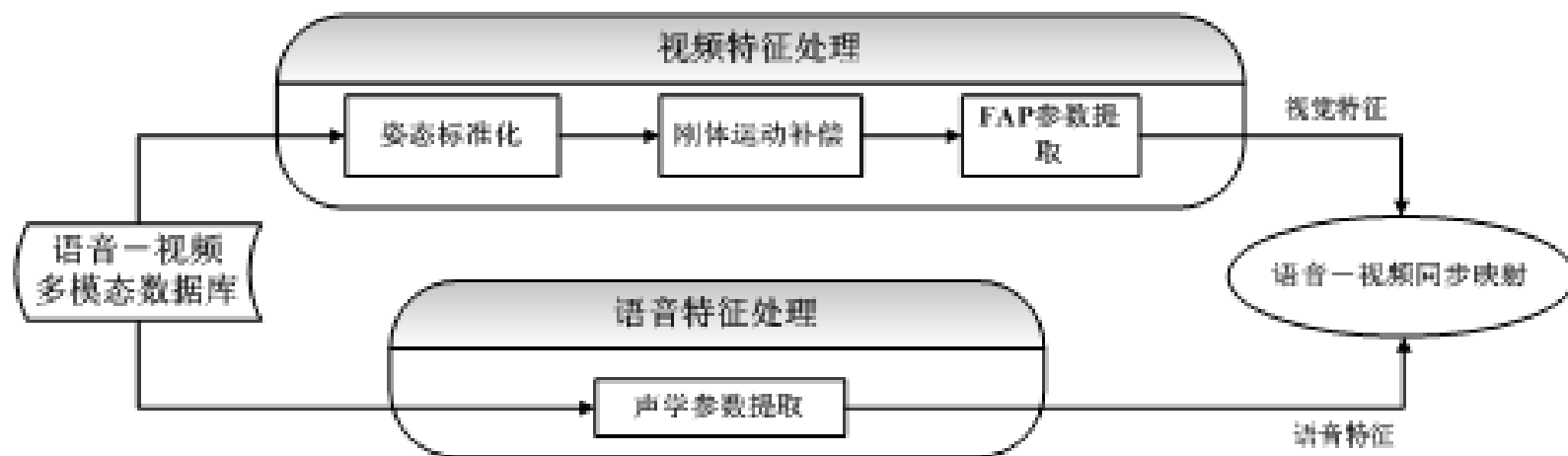


二维人脸表情实时捕获

- 提取语音视频特征表示;
- 描述audio-visual特征间关联关系;



训练过程

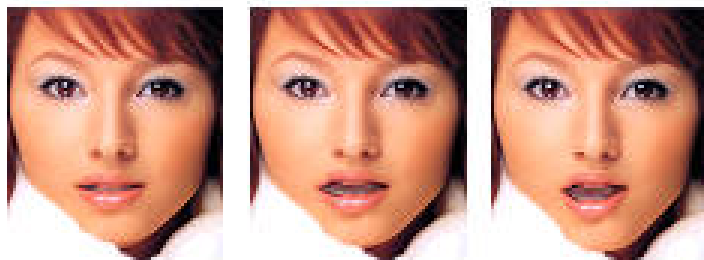


动画过程

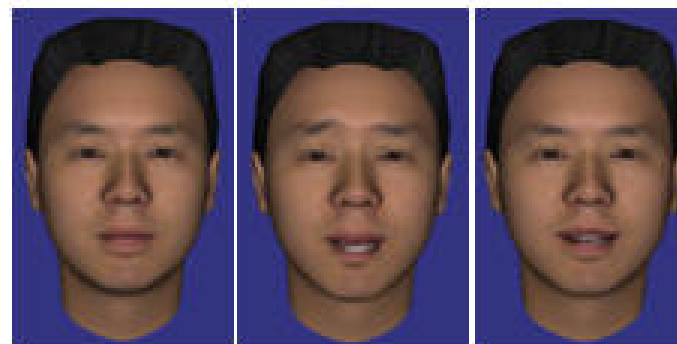


基于动态基元选取可视语音合成

■ 根据语音序列进行可视语音合成



二维可视语音合成结果



三维可视语音合成结果

■ 进一步的发展方向

- 提高合成语音的自然度
- 丰富合成语音的表现力
- 多语种文语合成(multi-language TTS).
- 语音是构成人类语音信号的各种声音。在采集和存储上可以与波形声音一样，但由于语音是由一连串的音素组成。“一句话”中包含许多音节以及上下文过渡过程的连接体等特殊的信息，并且语音本身与语言有关，所以要把它作为一个独立的媒体来看待

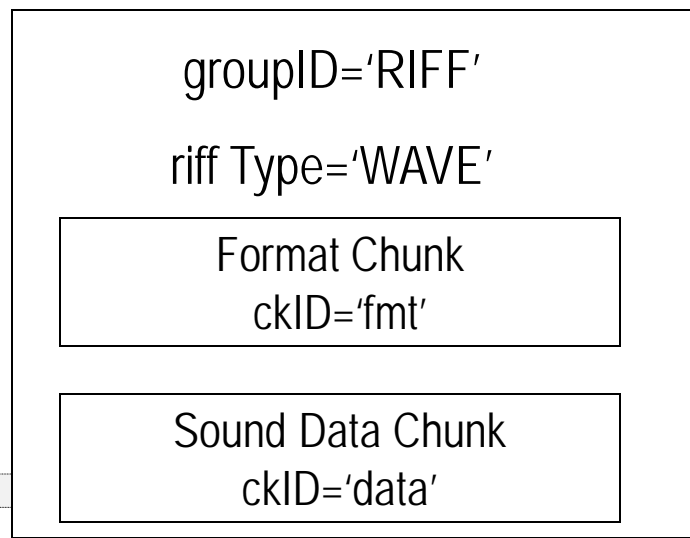
■ 技术展望

- 特定应用场合的计算机言语输出系统
- 韵律特征的获取与修改
- 语言理解与语言合成的结合
- 计算机言语输出与言语识别的结合
- 计算机言语输出与图像处理的结合(可视语音合成)

9、 音频文件格式

■ 1. 波形声音文件

- <1>Wave文件(.WAV):Microsoft公司的声音格式符合RIFF(Resource Interchange File Format资源交换文件:是多媒体扩展支持的一种带标记的文件结构)文件规范。应用于Windows平台。
- WAV文件格式支持存储多种采样频率和量化精度的声音数据,支持声音数据的压缩。文件有许多不同类型文件构造块组成,其中最主要是格式块(Format Chunk)和声音数据块(Sound Data Chunk)
 - 格式块包含波形的参数
 - RIFF中的其他文件块是可选的



波形文件结构



偏移量	字节	数据
0000	4	"RIFF"
0004	4	波形块的大小(文件大小减8)
0008	4	"WAVE"
000C	4	"fmt"
0010	4	格式块大小(16字节)
0014	2	wf.wFormatTag=WAVE_FORMAT_PCM=1
0016	2	wf.nChannels
0018	4	wf.nSamplesPerSec
001C	4	wf.nAvgBytesPerSec
0020	2	wf.BlockAlign
0022	2	wf.wBitsPerSample
0024	4	"data"
0028	4	波形文件大小
002C		波形数据开始

音频文件格式

- <2>AIFF文件(. AIF):Apple的声音格式符合AIFF(Audio Interchange File Format音频交换文件)文件规范。应用于Macintosh平台.
- <3>Audio文件(. AU):Sun Microsystems的数字音频格式, 是Internet常用的音频文件格式.
- <4>Sound文件(. SND):NeXT Computer的数字音频格式.
- <5>Voice文件(. VOC):Creative Labs的数字音频格式, 用于保存其Creative Sound Blaster系列声卡的音频文件格式.
- <6>MPEG文件(. MP1/. MP2/. MP3):MPEG标准的音频层, 其压缩方法采取的是一种有损压缩, 根据压缩质量和编码复杂程度可分为三层, MP1:压缩率4:1, MP2:压缩率6:1~8:1, MP3:压缩率10:1~12:1, 使用比较多.
- <7>RealAudio文件(. RA/. RM/. RAM):是RealNetworks的流式音频文件格式, 用于低速的广域网实时传输音频数据. 并可以根据网络的速率, 获得不同质量的音质(14. 4Kb/s:AM 28. 8Kb/s:FM 56Kb/s:CD)

2、标准MIDI文件（SMF）

- MIDI文件(.MID/.RMI):MIDI文件定义产生声音的指令(音色、强弱、时长等), MIDI声音效果取决于合成器的质量和波表的质量。
- 1988年被MMA采用, 扩展名为.MID, 用作MIDI音乐的文件交换标准, 也是音乐作品发行的标准。
- 一个MIDI文件包含1个标题块和若干音轨块。
- 标题块指出:标识符, 长度, 音轨块数目, MIDI 格式(格式0, 1, 2), 时间格式(PPQN及SMPTE)等。
 - PPQN (pulses per 1/4 note), 单位: beats/minute(BPM)
 - SMPTE (Society of Motion Picture and Television Engineers)
- hh:mm:ss:ff:bb 单位: frames/second (fps)
- 音轨块用于记录MIDI数据, MIDI数据由一系列的MIDI message组成。

知识点小结

- 音频基本概念
- 听觉与声音认知心理学感知特性
- 音频数字化
- 电子音乐合成与MIDI系统
- 声音属性与编码
- 声音重建与声卡
- 波形声音编程原理
- 音频文件格式