

本节课内容：数字音频基础

- 1、声音数字化
- 2、数字声音文件的存储
- 3、声音工具
- 4、声音的合成

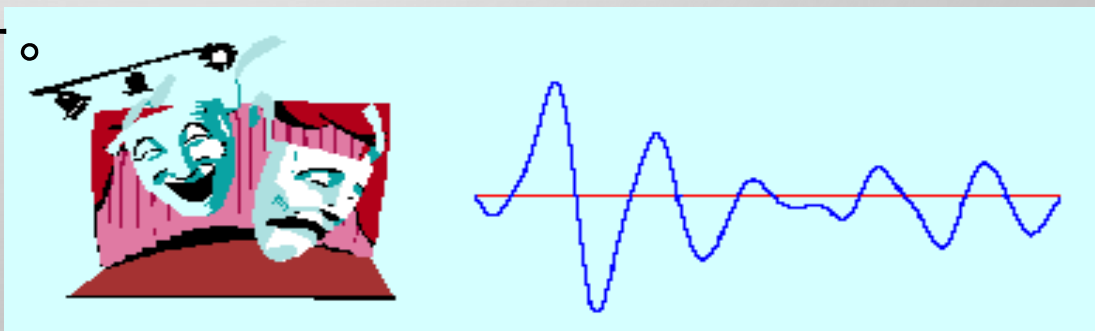
林福宗：c2：数字声音及MIDI简介

Ze-Nian Li：c6：数字音频基础

声音数字化

1.1 什么是声音

- ▶ 声音是听觉器官对声波的感知，而声波是通过空气或其他媒体传播的连续振动
- ▶ 声音的强弱体现在声波压力的大小上，音调的高低体现在声音的频率上
- ▶ 声波具有普通波所具有的特性，例如反射(reflection)、折射(refraction)和衍射(diffraction)等
- ▶ 声音用电表示时，声音信号在时间和幅度上都是连续的模拟信号，如图所示。它是由许多频率不同的分量信号组成的复合信号。



加州旅馆(母带版).wav - Cool Edit Pro

File Edit View Effects Generate Analyze Favorites Options Window Help

Files Effects Favorites

加州旅馆(母带版).wav

Show File Types: Sort By:

☒ Wave ☒ MIDI ☒ Video

Recent Auto-Play Full Paths

0:00.000

Begin End Length

Sel 0:00.000 0:00.000 0:00.000

View 0:00.000 6:30.923 6:30.923

Opened in 140.00 seconds

L: 0dB @ 3:09.729 192000 732-bit ?Stereo 586 MB 146 MB free

开始 网络教学综合... 4 Microsoft Pow... 加州旅馆(母...

16:07

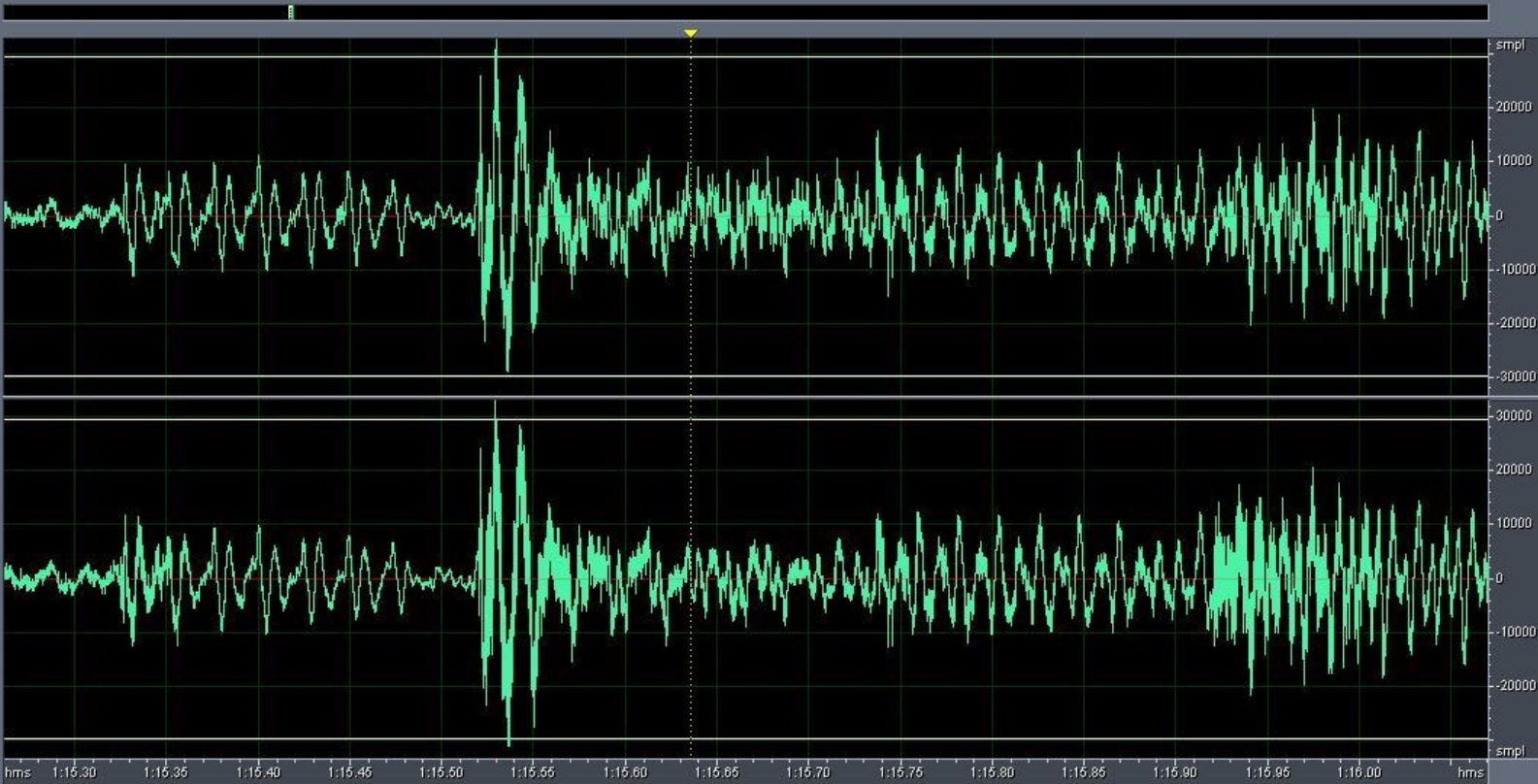
加州旅馆(母带版).wav - Cool Edit Pro

File Edit View Effects Generate Analyze Favorites Options Window Help



Files Effects Favorites

加州旅馆(母带版).wav

Show File Types: Sort By: Wave MIDI Video Recent Auto-Play Full Paths



hms 1:15.30 1:15.35 1:15.40 1:15.45 1:15.50 1:15.55 1:15.60 1:15.65 1:15.70 1:15.75 1:15.80 1:15.85 1:15.90 1:15.95 1:16.00 hms



1:15.635

	Begin	End	Length
Sel	1:15.635		0:00.000
View	1:15.261	1:16.069	0:00.808

Opened in 140.00 seconds

L: -13.2dB @ 1:15.807 192000 ?32-bit ?Stereo 586 MB 146 MB free

开始 网络教学... 4 Microsoft... 加州旅馆... 加州旅馆... NVIDIA ... 16:17

声音的频率

- 高保真声音(high-fidelity audio): 10~20000Hz
- 频率为**20Hz—20kHz**的信号称为音频（audio）信号，可以被人的耳朵感知。
- 人的发音的频率范围：**80Hz—3.4kHz(speech)**
- 人说话的频率范围：**300Hz—3kHz**

声音(audio):	20~20 000Hz
话音(speech):	300~3000/3400 Hz
亚音/次音(subsonic):	<20 Hz
超声(ultrasonic):	>20 000 Hz

声音的三要素

音强、音调、音色

- **音强**（声音的强弱，响度）取决于声音的幅度。
- **音调**取决于声音的频率。
- 一般的声音都是由发音体发出的一系列频率、振幅各不相同的振动复合而成的。这些振动中有一个频率最低的振动，由它发出的音就是基音，其余为泛音。
- **音色**是由混入基音的泛音所决定的，高次谐波越丰富，音色越有明亮感和穿透力。

音质

- 声音聆听效果的好坏。噪声信号强的声音比噪声信号弱的声音音质要差。
- 谈论某音响的音质好坏，主要是**衡量声音的三要素是否达到一定的水准**，即相对于某一频率或频段的**音高是否具有一定的强度**，并且在要求的频率范围内、同一音量下，各频点的**幅度是否均匀、均衡、饱满**，频率响应**曲线是否平直**，声音的音准是否准确，既忠实地呈现了音源频率或成分的原来面目，频率的畸变和相移又符合要求。声音的泛音适中，谐波较丰富，听起来音色就优美动听。

声音的听觉特性

- 一般以**1kHz纯音**为准进行测量，人耳刚能听到的声压为 0dB。分贝值是呈指数形式上涨的。每上升10，表示音量增加10倍。
- **分贝**（decimal Bel）：两个电压、功率、声压之比。简称dB。
- 时钟滴答声约为15dB；人低声耳语约为30dB；冰箱、电风扇的声音为40~70dB；大声说话为60~70dB；汽车噪音为80~100dB；电视机伴音可达85dB；电锯声是110dB；喷气式飞机的声音约为130dB.....
- 国家规定的办公室内噪音标准为45dB~60dB。

声音数字化

1.2 声音质量的度量

- 声音的质量与声音的带宽有关，一般来说频率范围越宽，声音质量也就越高。

声音类型	带 宽
电话语音	200Hz—3.4kHz
调幅广播	50Hz—5.5kHz
调频广播	20Hz—11kHz
CD	5Hz—20kHz

- 信噪比（**SNR**）、主观平均判分法（**MOS**）

声音质量评分标准

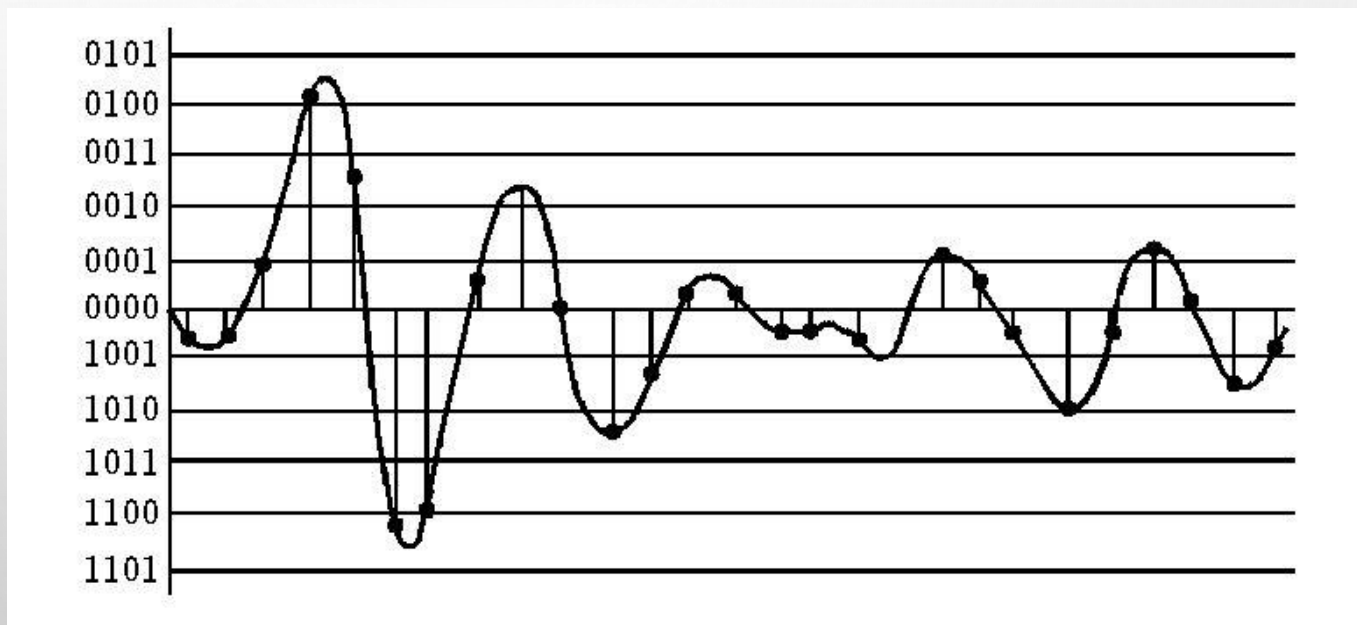
- 主观评价评分标准 **MOS/DMOS**

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)

声音数字化

1.3 如何数字化声音

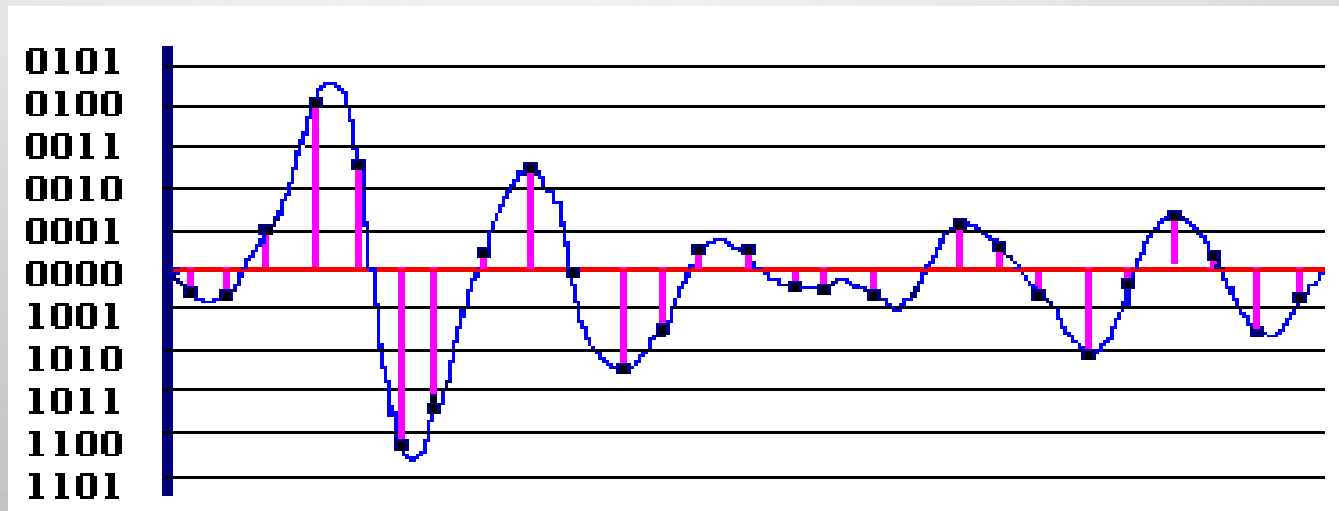
- 声音用电表示时，声音信号在时间和幅度上都是连续的模拟信号。它是由许多频率不同的分量信号组成的复合信号。



用数字表示声音波形，即得到了数字音频，这个过程成为数字化。

声音的数字化

- 连续时间的离散化通过采样来实现，就是每隔相等的一段时间采样一次，这种采样称为**均匀采样(uniform sampling)**
- 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。
- **编码**：把已经量化后的采样值用二进制数码表示出来



①采样

采样频率

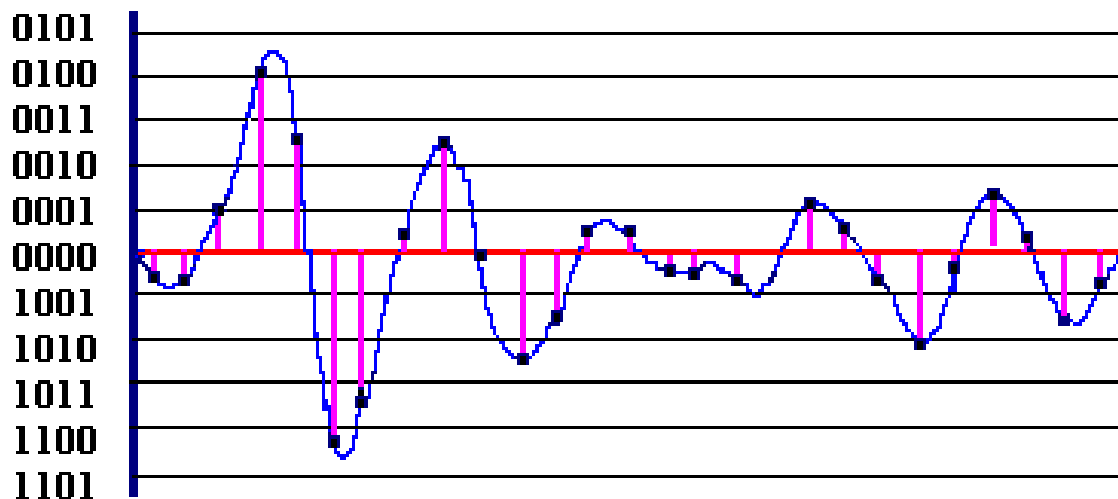
奈奎斯特理论指出：采样频率不应低于声音信号最高频率的两倍，这样就能把以数字表达的声音还原成原来的声音，称为无损数字化。

$$f_s \geq 2f_{\max}$$

话音信号最高频率约为3.4kHz，所以最低的采样频率取为8kHz（电话音质）。

- 多媒体中最常用的采样频率分别是44.1kHz（CD音质）、22.05kHz（FM）和11.025kHz（AM）。

- **量化率**
(量化精度)
(采样精度)



- ▶ **量化**：在振幅维的采样
- ▶ **量化精度**：度量声音波形幅度的精确程度，用每个声音样本的位数(即bit)表示。
- ▶ 典型的均匀量化率是8位和16位。
 - ▶ 8位量化把纵轴分成256个区间。
 - ▶ 16位量化把纵轴分成65536个区间。
- ▶ 样本位数的大小影响到声音的质量

采样精度

质量	采样频率 (kHz)	样本精度	单道声/ 立体声	数据率 (kb/s)	频率范围 (Hz)
电话	8	8	单道声	64	200~3 400
AM	11.025	8	单道声	88	50~5 500
FM	22.050	16	立体声	705.6	20~11 000
CD	44.1	16	立体声	1411.2	5~20 000
DAT	48	16	立体声	1536	5~20 000
DVD音频	192 (最大)	24 (最大)	最高为6声道	9600	0~96 000

电话使用 μ 律编码，动态范围为13位，压缩后的样本精度为8位¹⁵

量化噪声

- 信号噪声比/信噪比(**signal-to-noise ratio, SNR**):衡量信号质量

$$SNR = 10 \log_{10} \frac{V_S^2}{V_N^2} = 20 \log_{10} \frac{V_S}{V_N} \text{ (单位 } dB \text{)}$$

- **信号量化噪声比 (SQNR)**

- 量化（离散化）会带来误差→量化噪声、量化误差
→某个采样时间点的模拟值和最近的量化值之间的差值。

$$SQNR = 20 \log_{10} \frac{V_S}{V_q} (dB)$$

信号量化噪声比 (SQNR)

- 假设每个采样点的量化精度为N位，数字信号的取值范围是 $-2^{(N-1)} \sim 2^{(N-1)} - 1$ 。设实际模拟信号的范围是

- $V_{\max} \sim +V_{\max}$ ，那么每个量化级代表的电压是 $2V_{\max}/2^N$ 或者 $V_{\max}/2^{(N-1)}$ 。当信号取值为 V_{\max} 时，

$$SQNR = 20 \log_{10} \frac{V_s}{V_q} = 20 \log_{10} \frac{V_{\max}}{0.5 * V_{\max} / 2^{N-1}} = 6.02N (dB)$$

采样点中每一位增加6dB的分辨率

最大的量化误差为量化间隔的一半。
很多时候量化误差要小于该值。

信号量化噪声比 (SQNR)

- 信号为正弦波时，SQNR（证明见书中参考文献）：

$$SQNR = 20 \log_{10} \frac{V_s}{V_q} = 6.02N + 1.76(dB)$$

- 16bit量化时， $SNR = 6.02 \times 16 + 1.76 = 98.08(dB)$ ，
8bit量化时， $SNR = 6.02 \times 8 + 1.76 = 49.92(dB)$

声音

如果192k采样率，24位精度的立体声，音乐总长6分30秒呢？

- 采样率决定了录音时提取样本的频率，较高的采样率（44.1kHz）能更精确地录制声音的高频分量。
- 采样精度（16位或8位）决定了数字化声音时的准确度。较大的采样精度会使录音更接近原始的声音。
- 音质越高，文件尺寸也越大。
- 举例：一分钟的采样率为44.1kHz采样精度为16位立体声的数字音频录音的文件大小为：

$$\frac{44100 \times 60 \times 2 \times 16}{1024 \times 1024 \times 8} \approx 10.1MB$$

再回头看看量化

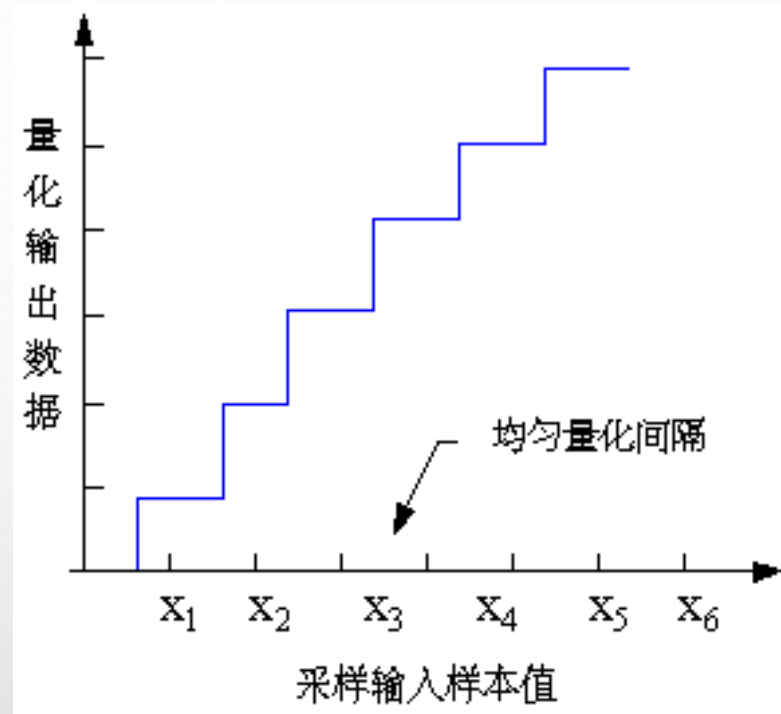
- 量化的方法

- 主要有均匀量化和非均匀量化

- 均匀量化

- 采用相等的量化间隔/等分尺度量采样得到的信号幅度，也称为**线性量化**。量化后的样本值 Y 和原始值 X 的差

$E=Y-X$ 称为量化误差或量化噪声



非均匀量化

- 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
- 可在满足精度要求的情况下用较少的位数来表示
- 声音数据还原时, 采用相同的规则
- 采样输入信号幅度和量化输出数据之间定义了两种对应关系

(“压大补小”)

- μ 律压扩算法 (美日采用)
- A 律压扩算法 (欧洲、中国大陆采用)

注: 压扩(COMPANDING)

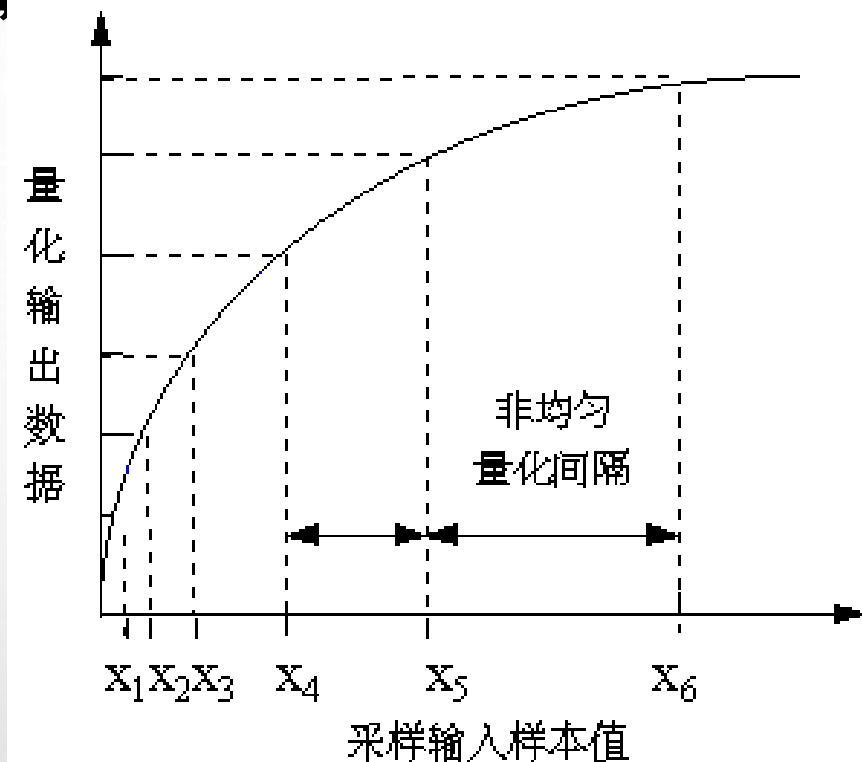
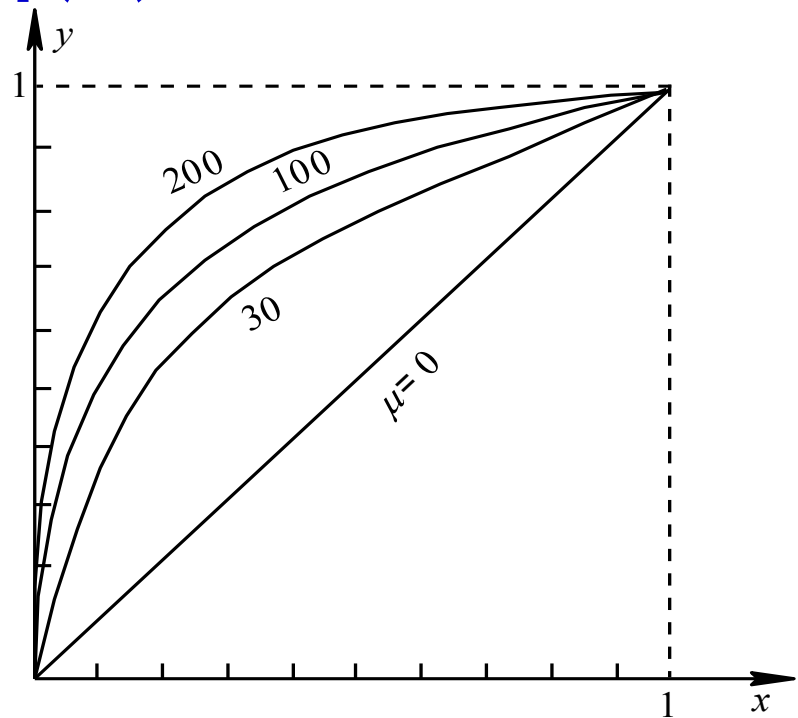


图 非均匀量化

μ 律压扩算法

$$y = \frac{\ln(1 + \mu x)}{\ln(1 + \mu)}, \quad 0 \leq x \leq 1$$

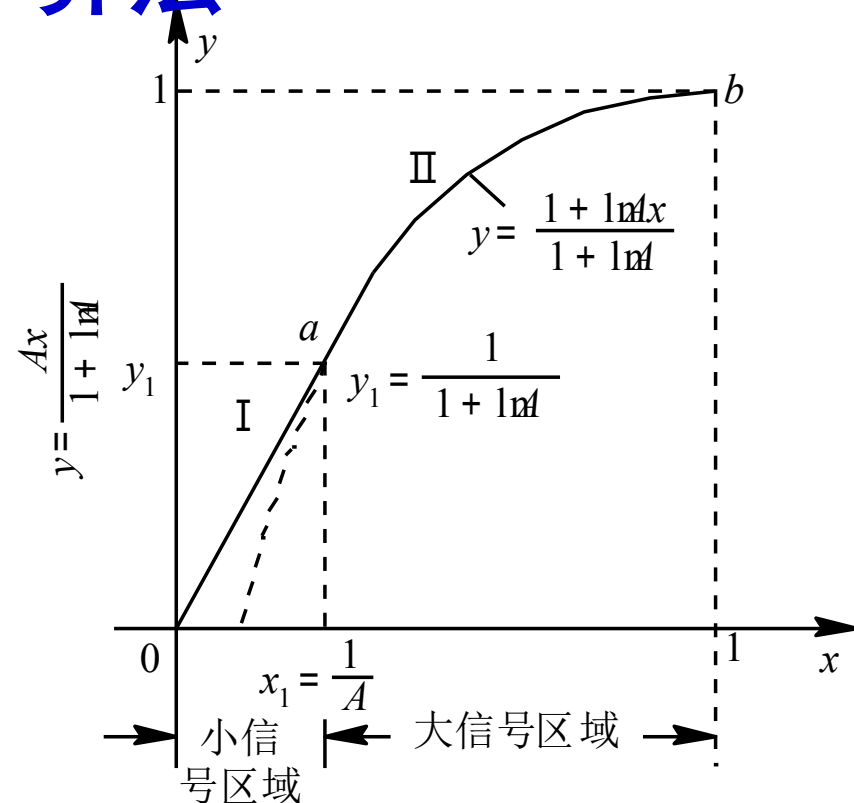


$$\left. \left(\frac{dy}{dx} \right) \right|_{x \rightarrow 0} = \left. \frac{\mu}{(1 + \mu x) \ln(1 + \mu)} \right|_{x \rightarrow 0} = \frac{\mu}{\ln(1 + \mu)} = 21.6$$

$$\left. \left(\frac{dy}{dx} \right) \right|_{x \rightarrow 1} = \left. \frac{\mu}{(1 + \mu x) \ln(1 + \mu)} \right|_{x \rightarrow 1} = \frac{100}{(1 + 100) \ln(1 + 100)} = 0.214$$

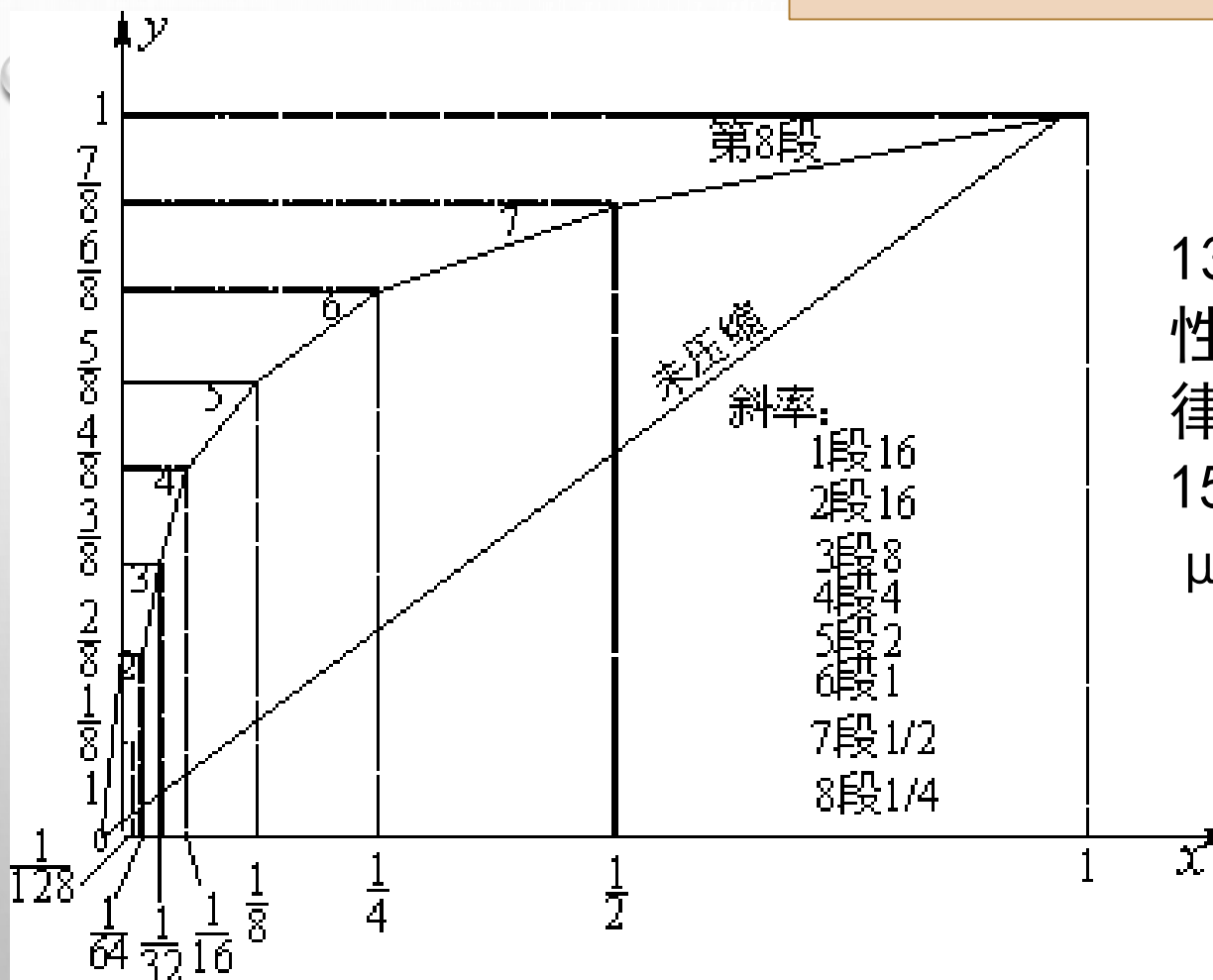
A律压扩算法

$$y = \begin{cases} \frac{Ax}{1 + \ln A} & , 0 < x \leq \frac{1}{A} \\ \frac{1 + \ln Ax}{1 + \ln A} & , \frac{1}{A} < x \leq 1 \end{cases}$$



$$\frac{dy}{dx} = \begin{cases} \frac{A}{1 + \ln A} = 16 & , 0 < x \leq \frac{1}{A} \\ \frac{A}{(1 + \ln A)Ax} = \frac{0.1827}{x} & , \frac{1}{A} < x \leq 1 \end{cases}$$

早期的A律和 μ 律压扩特性是用非线性模拟电路获得的



13折线A律，它的特性近似 $A=87.6$ 的A律压扩特性。

15折线 μ 律，近似 $\mu=255$ 的压扩特性。

①采样 ②量化 ③编码

编码就是把已经量化后的采样值用二进制数码表示出来

(1)自然码

无符号的二进制代码

(2)符号 - 数值码

在自然码最高位（MSB）增加了一个符号位而构成，用以表示数值的正负，一般用“0”表示正，而用“1”表示负，一般用来表示双极性信号。

(3)2的补码（正数与原码相同，负数取反+1,符号位除外）

(4) 格雷码

每增加1个数值时只有一个码元变化的码

000	001	011	010	110	111	101	100	格雷码
-----	-----	-----	-----	-----	-----	-----	-----	-----

0	1	2	3	4	5	6	7	十进制数
---	---	---	---	---	---	---	---	------

本节课内容：数字音频基础

- 1、声音数字化
- **2、数字声音文件的存储**
- 3、声音工具
- 4、声音的合成
- 5、MIDI：乐器数字化接口

2 音频文件的存储（格式）

- **au** Sun和NeXT公司的声音文件存储格式
- **aif** Audio Interchange File Format
- **mid/midi** MIDI文件存储格式
- **mp3** MPEG-1 Layer 3
- **wma** Windows Media Audio微软的音频格式
- **ra** RealNetworks的流式声音文件格式
- **wav** Windows上的声音文件存储格式
- **snd** Macintosh上的声音文件存储格式
- **mod** Amiga Music Module 文件
- **cmf** creative声霸卡带的MIDI文件存储格式
- **voc** creative声霸卡带的声音文件存储格式

WAVE文件格式剖析

- 多媒体中使用的声波文件格式之一，它是以RIFF格式为标准的。
- RIFF是英文Resource Interchange File Format的缩写，每个WAVE文件的头四个字节便是“RIFF”。
- WAVE文件由文件头和数据体两大部分组成。其中文件头又分为RIFF / WAV文件标识段和声音数据格式说明段两部分。

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

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

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


偏移	字节数	数据类型	内容
00H	4	char	"RIFF"标志
04H	4	long int	自08H起的文件长度
08H	4	char	"WAVE"标志
0CH	4	char	"fmt"标志
10H	4	long int	"fmt" chunk的长度（自14H起）
14H	2	int	格式类别（1为PCM形式的音频数据）
16H	2	int	通道数，单声道为1，双声道为2
18H	4	int	采样率（每秒样本数）
1CH	4	long int	波形音频数据传送速率，其值为通道数×每秒数据位数×每样本的数据位数 / 8。播放软件利用此值可以估计缓冲区大小。
20H	2	int	数据块的调整数（按字节计算），其值为通道数×每样本的数据位数 / 8。
<pre> 0 1 2 3 4 5 6 7 8 9 a b c d e f 00000000h: 52 49 46 46 D6 DA 00 00 57 41 56 45 66 6D 74 20 ; RIFF误..WAVEfmt 00000010h: 12 00 00 00 01 00 01 00 40 1F 00 00 80 3E 00 00 ;@...€>.. 00000020h: 02 00 10 00 64 61 74 61 C0 DA 00 00 00 00 00 00 ;data磊..... </pre>			
24H	4	char	数据标记符 "data"
28H	4	long int	音频数据块的长度

本节课内容：数字音频基础

- 1、声音数字化
- 2、数字声音文件的存储
- **3、声音工具**
- 4、声音的合成
- 5、MIDI：乐器数字化接口

3 声音工具

- 声音工具(**audio tools**)是用来录制、播放、编辑和分析声音文件的软件，在音乐节目的后期制作、多媒体音效制作和视像配音等方面有重要作用
- 声音工具使用相当普遍，可供选择的聲音工具相当多，它们的功能相差很大。如**Windows**自带的声音工具，**Cool Edit**和 **Adobe Audition**等等。

声音工具(续1)



(a) Windows录音机的录放界面



(b) Windows Media Player播放音乐时的界面

图 Windows自带的声音工具

音频软件开发技术

- 简单声音的呈现
- **API函数**呈现声音 (**PlaySound**)
- 多媒体控件**MCI32**播放音频
- 微软媒体播放器控件**wmplayer**播放音频
- 使用**DirectSound**播放wav文件

DIRECTX简介

- 一种图形应用程序接口(API)，简单的说它是一个辅助软件，一个提高系统性能的加速软件，由微软创建开发的，微软将定义它为“**硬件设备无关性**”。
- 它提供了一整套的多媒体接口方案。
 - **Direct Graphics(Direct 3D+Direct Draw)**
 - **Direct Input、**
 - **Direct Play、**
 - **Direct Sound、**
 - **Direct Show、**
 - **Direct Setup、**
 - **Direct Media Objects**

本节课内容：数字音频基础

- 1、声音数字化
- 2、数字声音文件的存储
- 3、声音工具
- **4、声音的合成**
- 5、MIDI：乐器数字化接口

4 合成声音

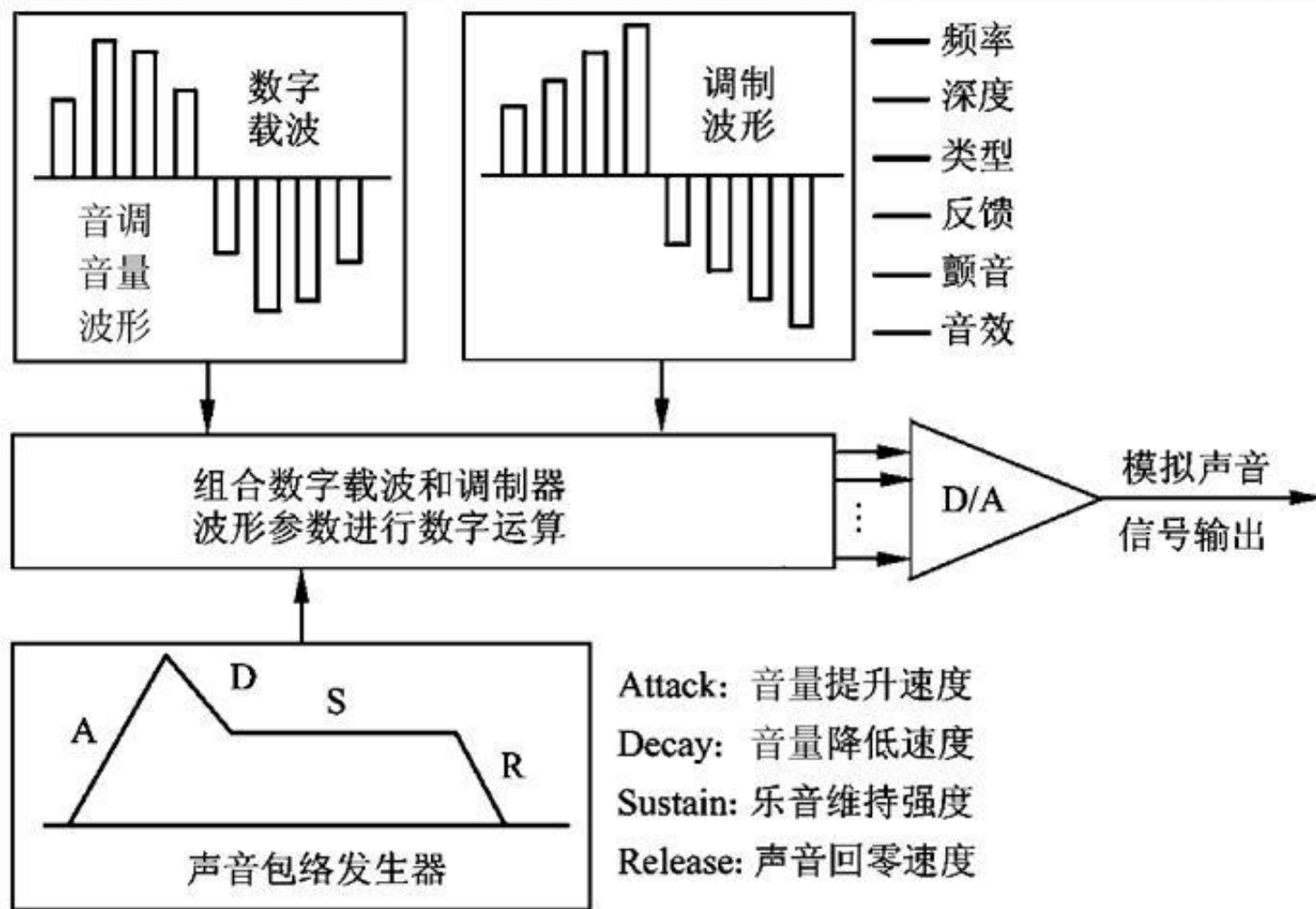
- 4.1 调频 (FM)

- 在载波正弦信号中加入一个设计调频信号的项，使原来的正弦信号发生改变，以此来模拟声音的波形。

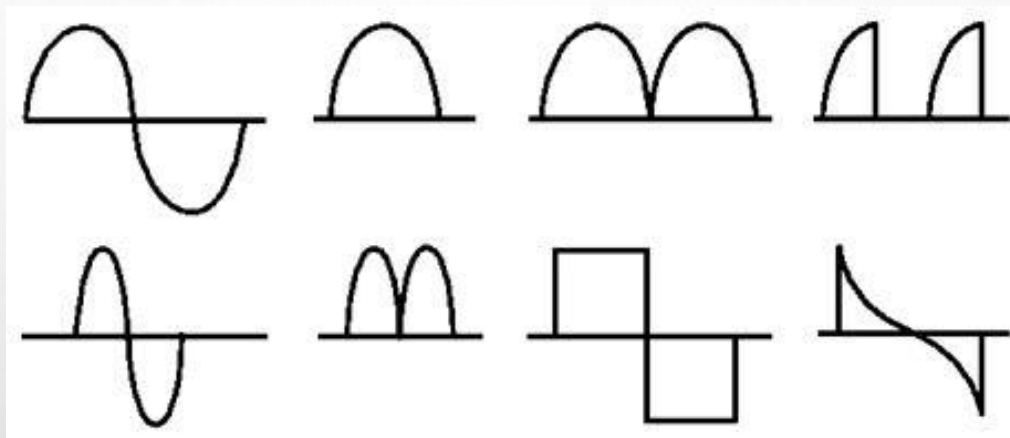
- 4.2 波形表法 (wave table)

- 把真实乐器发出的声音以数字的形式记录下来（波形表），播放时改变播放速度，从而改变音调周期，生成各种音阶的音符。

4.1 频率调制 (FM) 合成声音 (P94)



- 数字式FM合成器采用的基本波形：



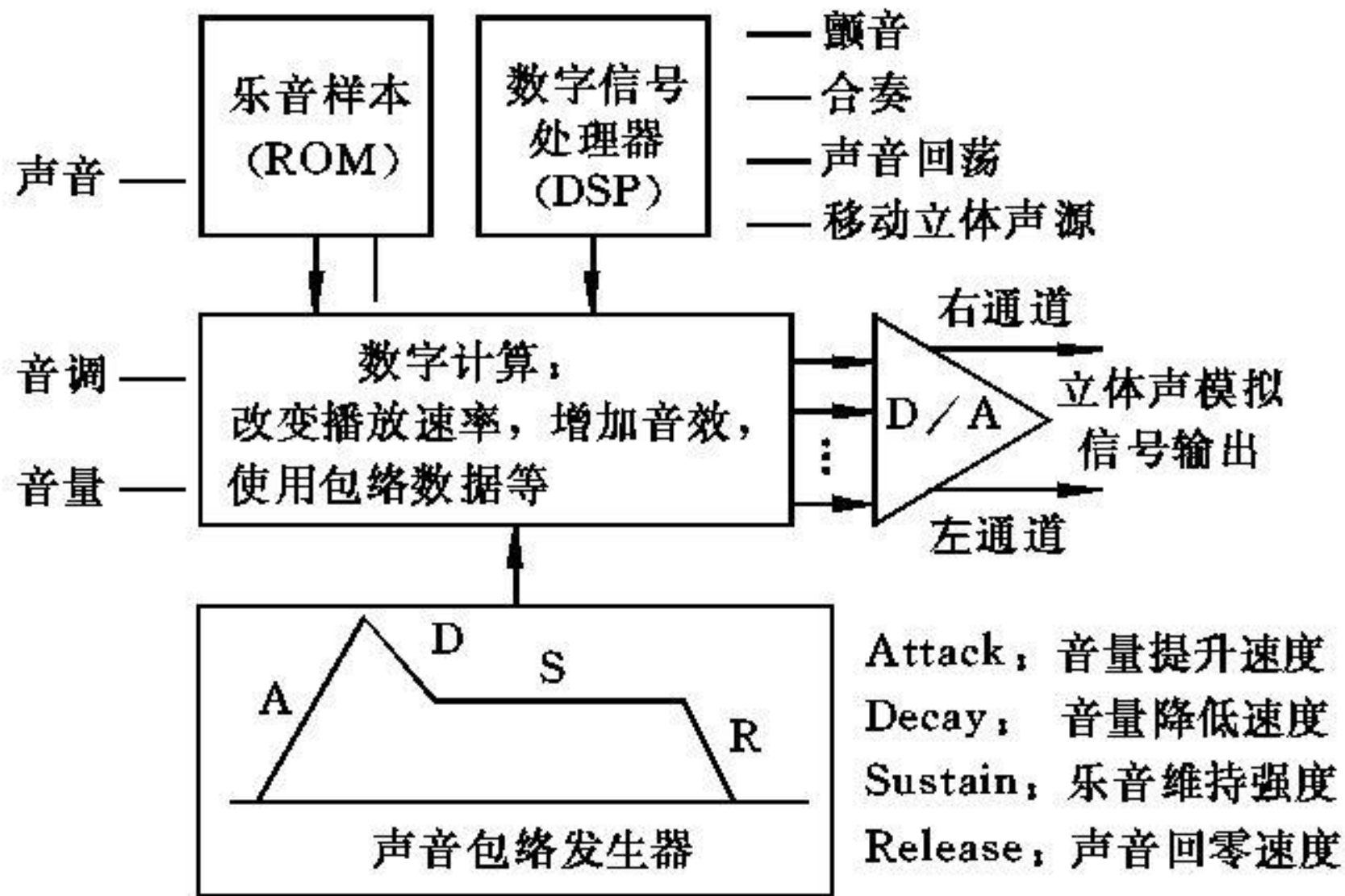
- 用什么样的波形作为数字载波波形、用什么样的波形作为调制波形、用什么样的波形参数去组合才能产生所希望的乐音，这就是FM合成器的算法。

可改变的参数:

- (1) 改变数字载波频率可以改变乐音的音调，改变它的幅度可以改变它的音量。
- (2) 改变波形的类型，如用正弦波、半正弦波或其他波形，会影响基本音调的完整性。
- (3) 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。
- (4) 改变反馈量，就会改变正常的音调，产生刺耳的声音。
- (5) 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。

4.2 波形表法

- 这种方法就是把真实乐器发出的声音以数字的形式记录下来（波形表，一般存储在声卡的存储器中），播放时改变播放速度，从而改变音调周期，生成各种音阶的音符。
- 乐音样本的采集：音乐家在真实乐器上演奏不同的音符，选择44.1kHz的采样频率、16bit的乐音样本，这相当于CD-DA的质量，把不同音符的真实声音记录下来。



本节课内容：数字音频基础

- 1、声音数字化
- 2、数字声音文件的存储
- 3、声音工具
- 4、声音的合成
- **5、MIDI：乐器数字化接口**

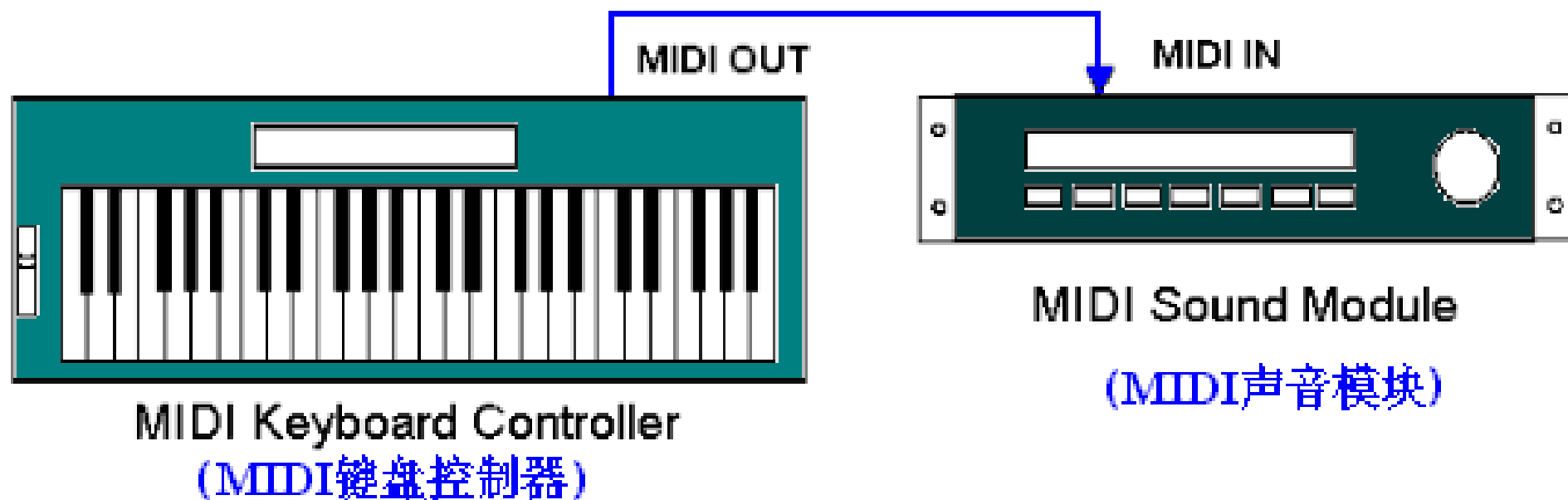
5 MIDI乐器数字接口

- **MIDI**

The **M**usical **I**nstrument **D**igital **I**nterface

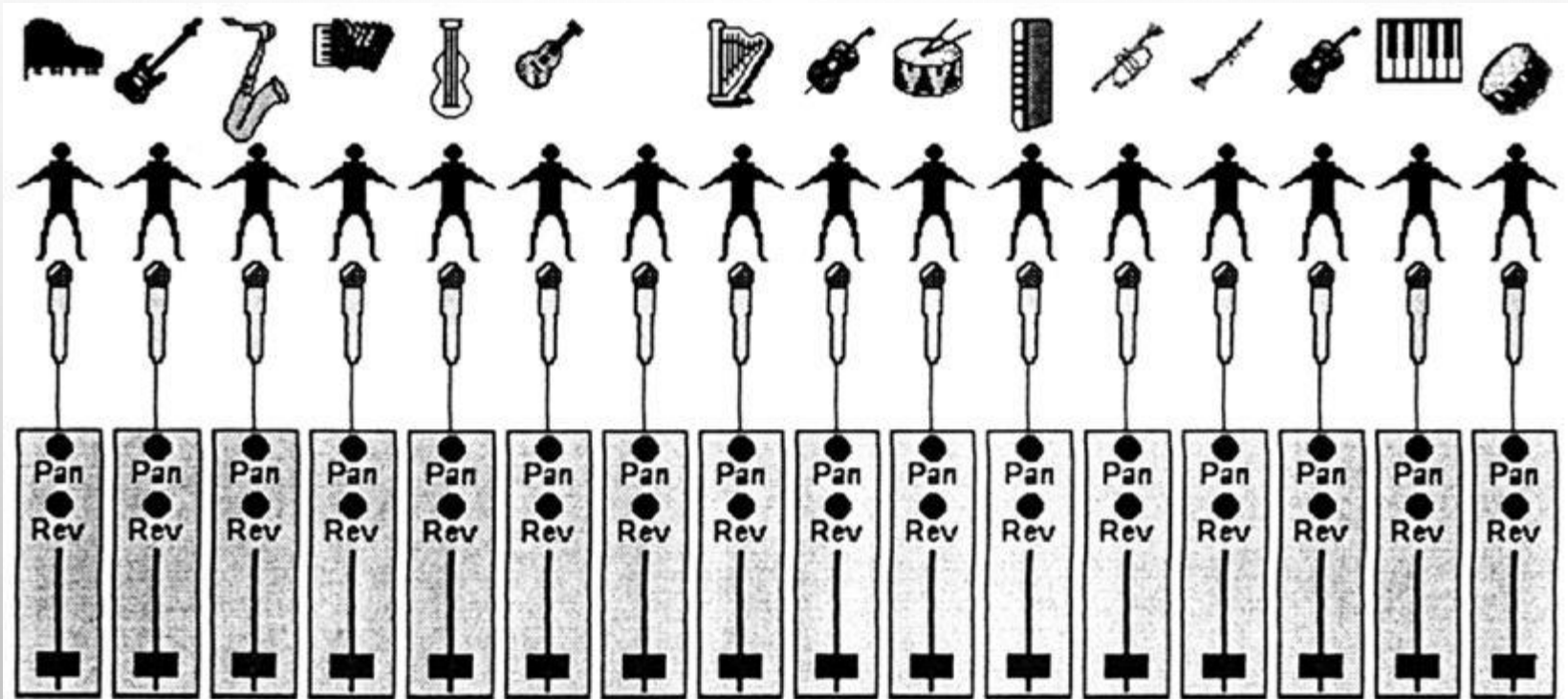
- 用于在音乐合成器(music synthesizers)、乐器(musical instruments)和计算机之间交换音乐信息的一种标准。
- MIDI是乐器和计算机使用的标准语言，是一套指令(即命令)的约定，它指示乐器(即MIDI设备)要做什么，怎么做，如演奏音符、加大音量、生成音响效果等。
- **MIDI优点**：生成的文件比较小，容易编辑，可以作背景音乐。

MIDI控制器和MIDI声音模块

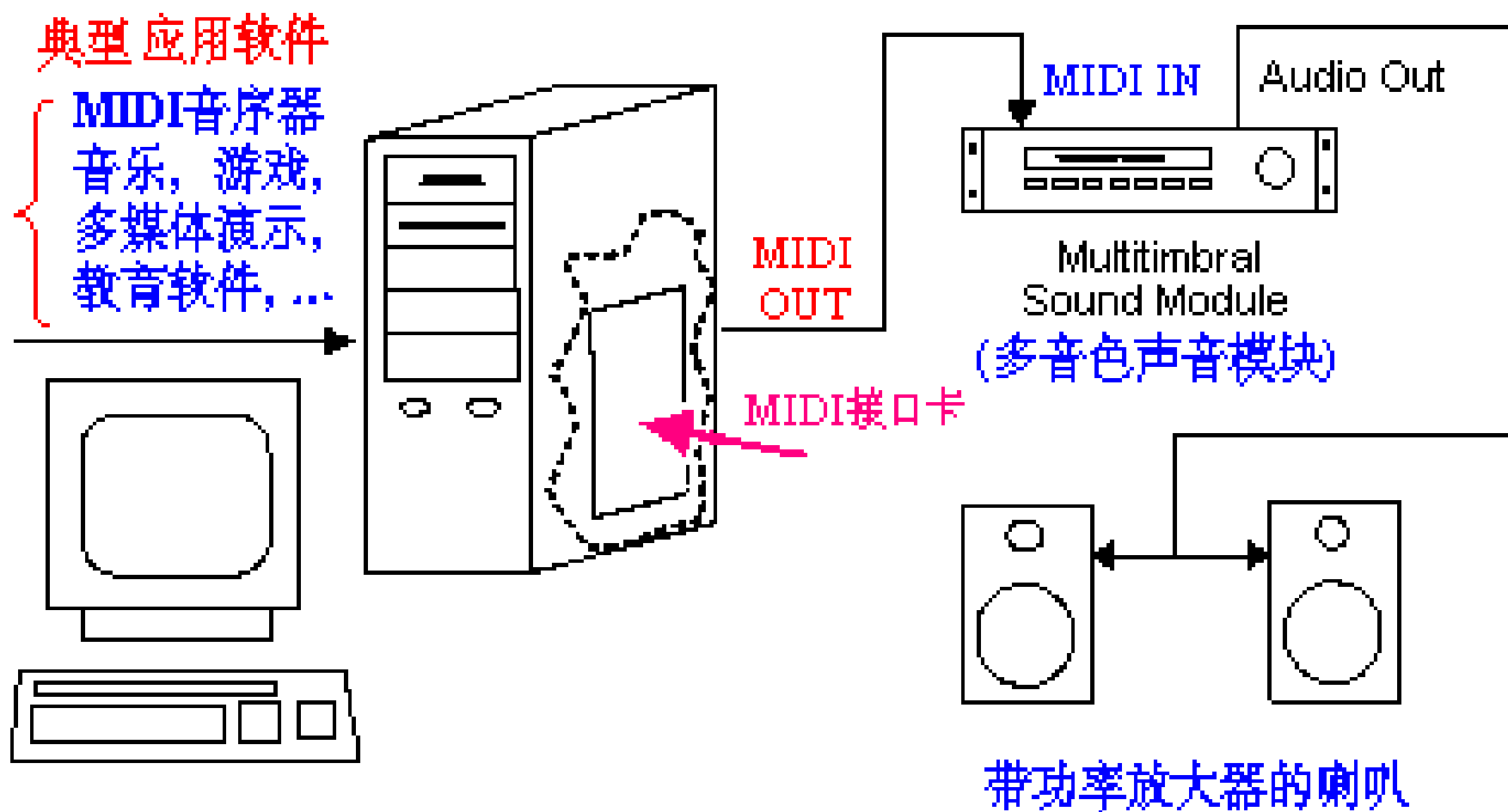


MIDI通道

16个逻辑通道可以指定16种乐器



基于PC的MIDI系统



作业：

- 什么叫采样、量化、线性量化、非线性量化？
- 选择采样频率为44.1kHz和样本精度是16位的录音参数，在不使用任何压缩技术的情况下，计算录制2分钟的立体声音频文件需要多少存储空间？
- 如果一个耳塞能把噪声降低20dB，那么它一共降低了多少强度（能量）？
- * 选做练习：对wav文件格式进行结构的详细分析。
 - 软件工具：UltraEdit