



# 多媒体技术

# 回顾

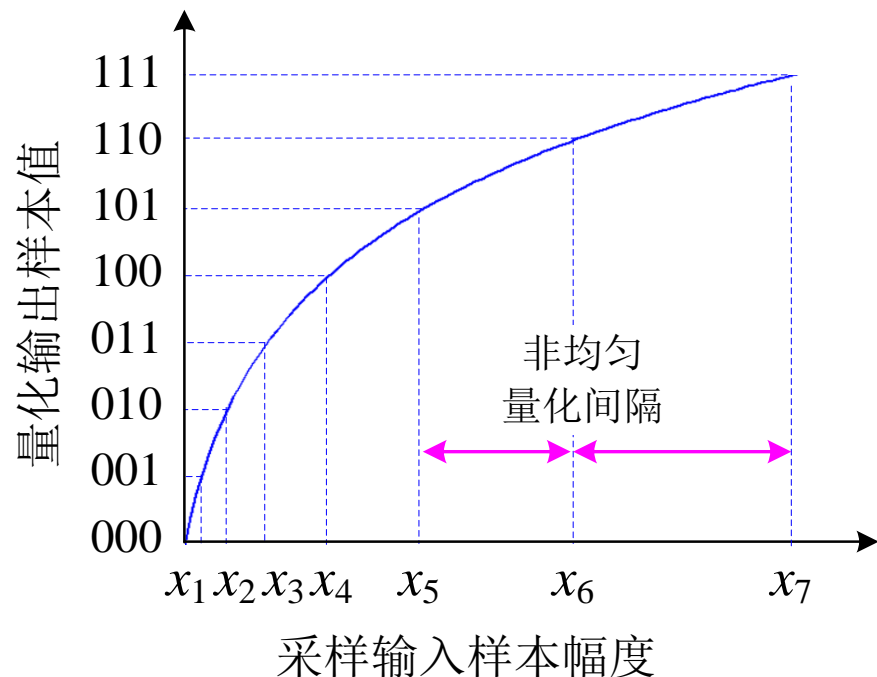
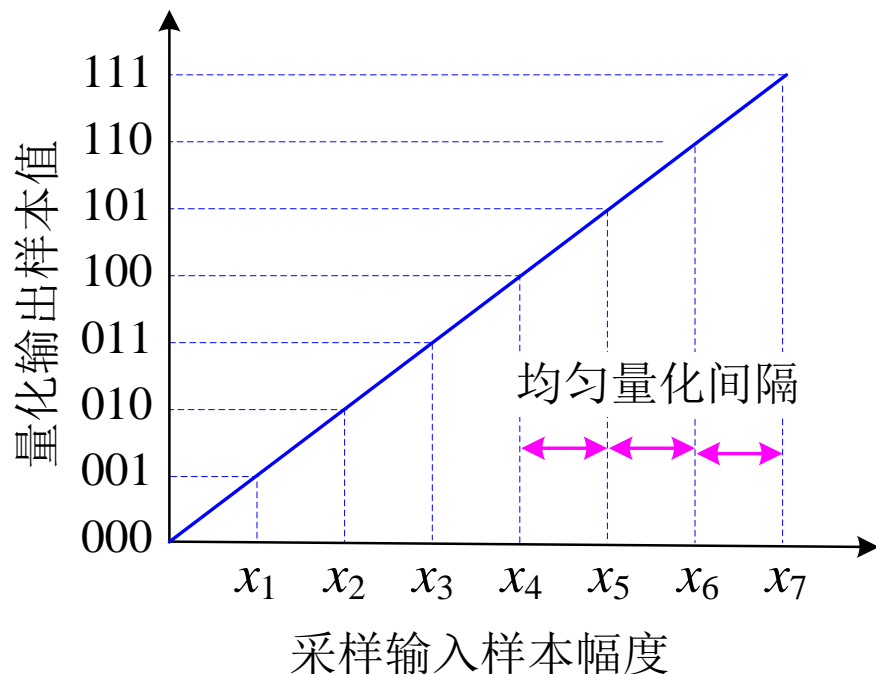
---

- 声音

- 心理变量：音强/响度、音调、音色
- 物理量纲：声强度、频率、频谱分析
- 听觉特性：等响曲线、掩蔽、临界频带、相位、听觉空间
- 质量标准
- 数字化过程
  - 采样、量化、编码

## 2.2.2 音频的数字化和符号化

### • 量化



## 2.2.2 音频的数字化和符号化

---

- 量化

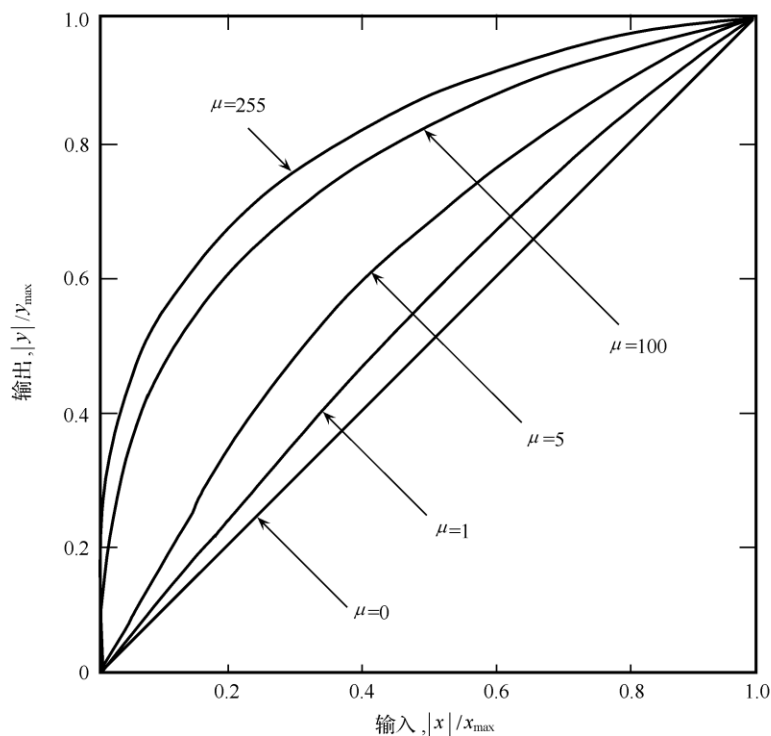
- 采样输入信号幅度和量化输出数据之间定义了两种对应关系：

- $\mu$ 律压扩算法：北美和日本，输入和输出对数关系
    - A律压扩算法：中国大陆和欧洲

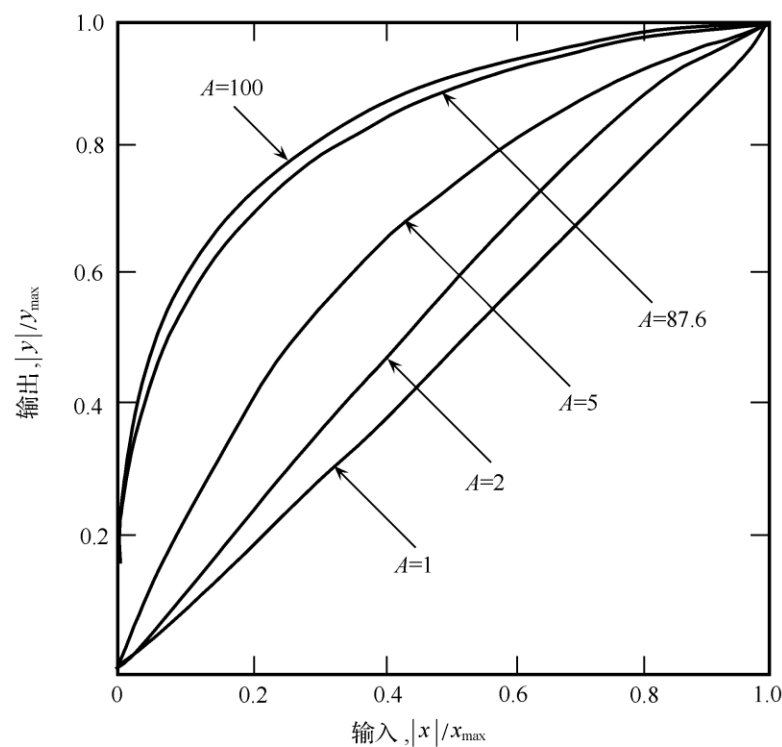
## 2.2.2 音频的数字化和符号化

- 量化

(1)  $\mu$ 律压扩



(2) A律压扩



## 2.2.2 音频的数字化和符号化

### • 量化

(1)  $\mu$  律压扩

$$y = y_{\max} \frac{\ln [1 + \mu(|x|/x_{\max})]}{\ln(1 + \mu)} \operatorname{sgn} x$$

$\mu$  是确定压缩量的参数，0 表示无压缩均匀量化，实用时取255

(2) A律压扩

$$y = \begin{cases} y_{\max} \frac{A(|x|/x_{\max})}{1 + \ln A} \operatorname{sgn} x & \left( 0 < \frac{|x|}{x_{\max}} \leq \frac{1}{A} \right) \\ y_{\max} \frac{1 + \ln [A(|x|/x_{\max})]}{1 + \ln A} \operatorname{sgn} x & \left( \frac{1}{A} < \frac{|x|}{x_{\max}} < 1 \right) \end{cases}$$

$$\operatorname{sgn} x = \begin{cases} +1 & , x \geq 0 \\ -1 & , x < 0 \end{cases}$$

←..... 符号函数

A是压缩系数，1 表示无压缩，实用87.6

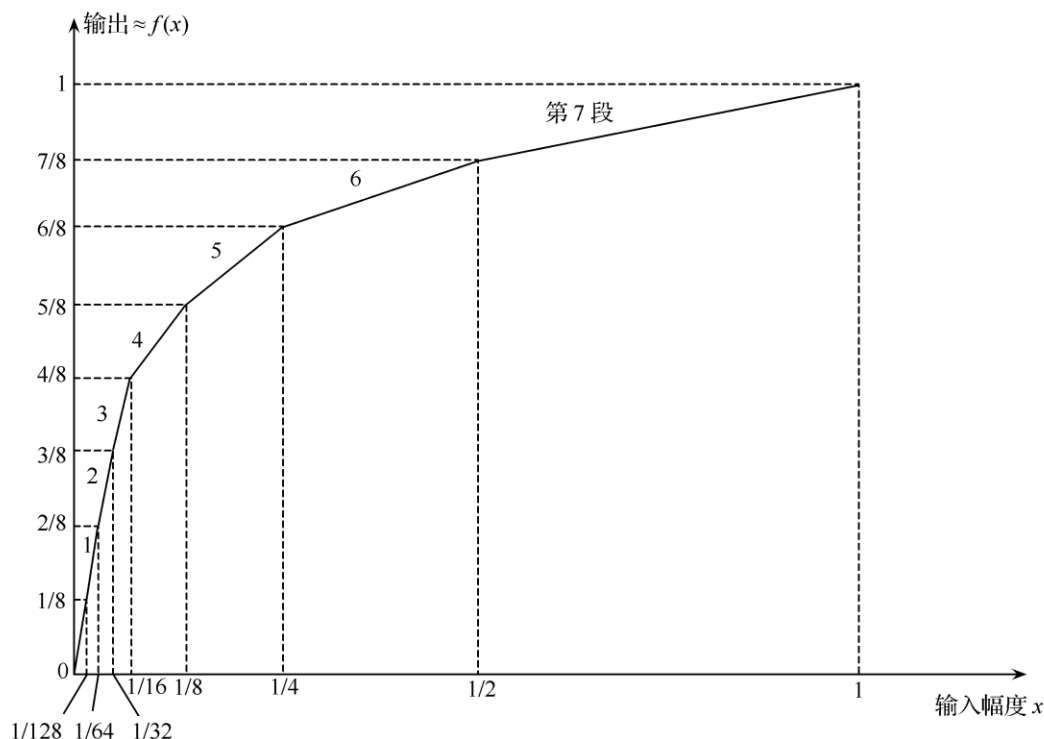
## 2.2.2 音频的数字化和符号化

### • 量化

#### 1) 13折线 实现A律

电路实现连续曲线函数复杂，一般利用数字电路形成折线近似

先非均匀量化，再在同一折线的小范围内对信号进行均匀量化，如分成16个量化级



## 2.2.2 音频的数字化和符号化

### • 量化

线段8斜率:  $1/8 \div 1/2 = 1/4$

线段7斜率:  $1/8 \div 1/4 = 1/2$

线段6斜率:  $1/8 \div 1/8 = 1$

线段5斜率:  $1/8 \div 1/16 = 2$

线段4斜率:  $1/8 \div 1/32 = 4$

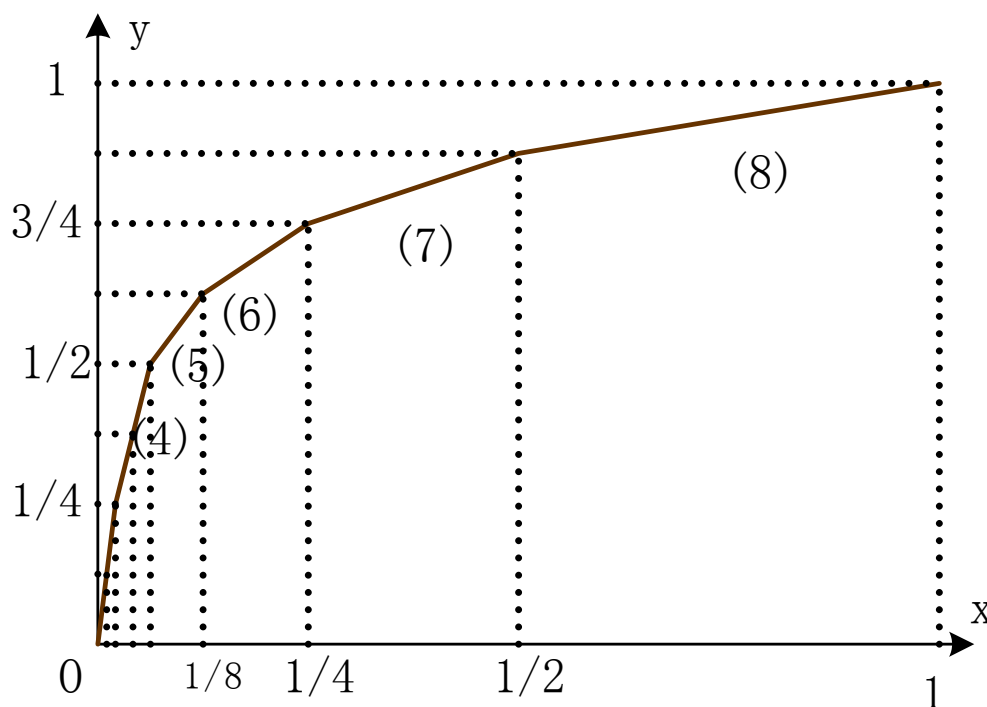
线段3斜率:  $1/8 \div 1/64 = 8$

线段2斜率:

$1/8 \div 1/128 = 16$

线段1斜率:

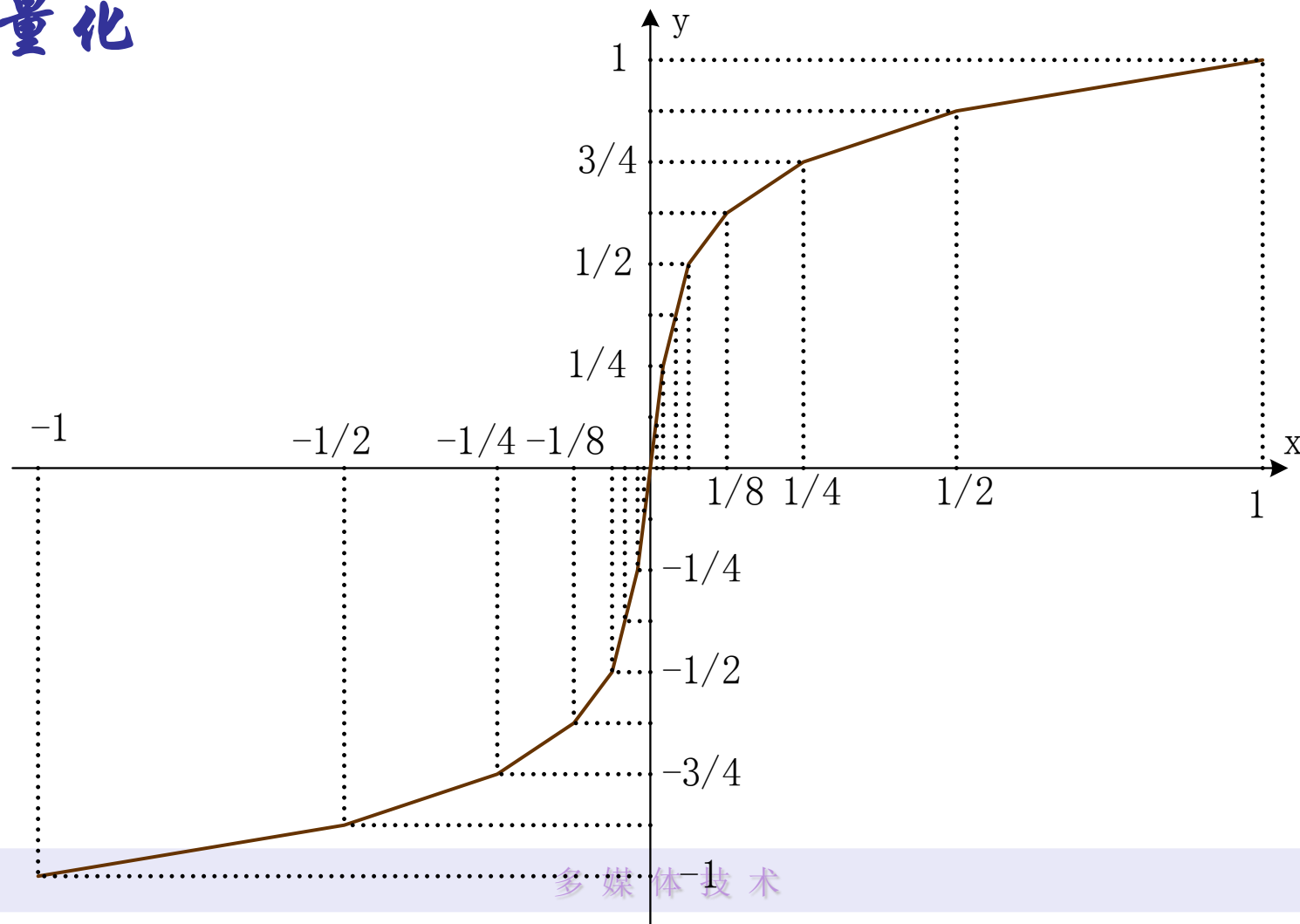
$1/8 \div 1/128 = 16$





## 2.2.2 音频的数字化和符号化

- 量化



## 2.2.2 音频的数字化和符号化

- 量化

A=87.6与 13 折线压缩特性的比较

y	0	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{3}{8}$	$\frac{4}{8}$	$\frac{5}{8}$	$\frac{6}{8}$	$\frac{7}{8}$	1
x	0	$\frac{1}{128}$	$\frac{1}{60.6}$	$\frac{1}{30.6}$	$\frac{1}{15.4}$	$\frac{1}{7.79}$	$\frac{1}{3.93}$	$\frac{1}{1.98}$	1
按折线分段时的x	0	$\frac{1}{128}$	$\frac{1}{64}$	$\frac{1}{32}$	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{2}$	1
段落	1	2	3	4	5	6	7	8	
斜率	16	16	8	4	2	1	$\frac{1}{2}$	$\frac{1}{4}$	

## 2.2.2 音频的数字化和符号化

---

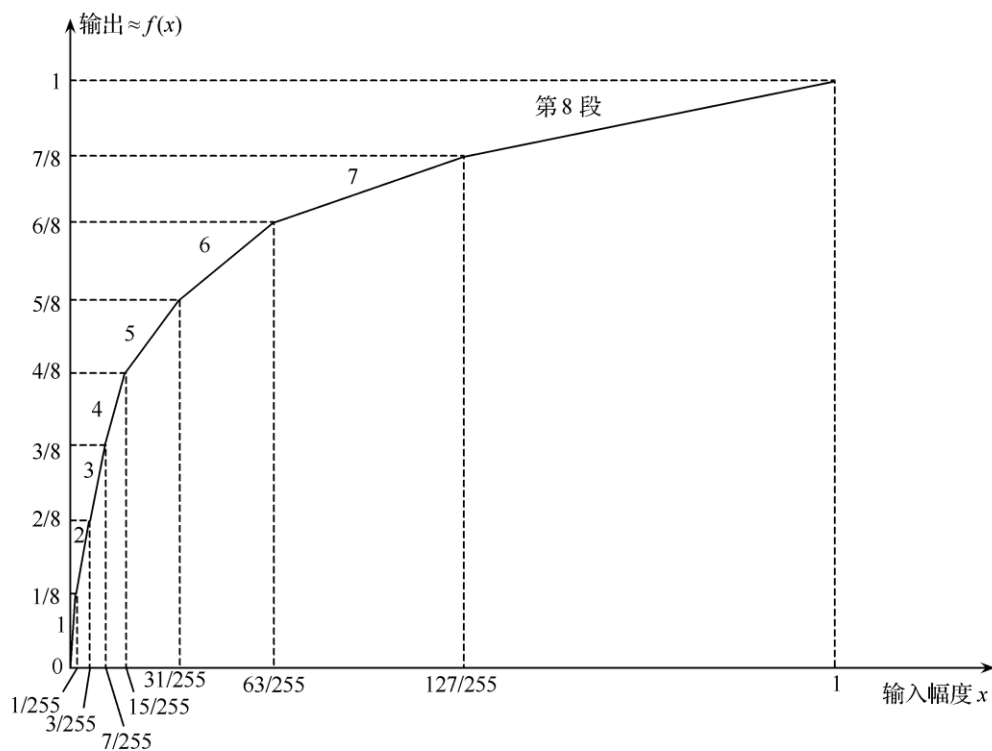
- 量化

- 在A律特性中，选A等于87.6有两个目的：
  - 使曲线在 origin 附近的斜率等于16，使16段折线简化成仅有13段；
  - 使13折线的转折点上曲线的横坐标x接近 $1/2^i$ 。

## 2.2.2 音频的数字化和符号化

- 量化

2) 15 折线 实现  $\mu$  律!



## 2.2.2 音频的数字化和符号化

### • 量化

- 把y轴均分8段;
- 对应于y轴分界点 $i/8$ 处的x轴分界点根据下式计算:

$$y = \frac{\ln(1 + 255x)}{\ln(1 + 255)} \Rightarrow x = \frac{256^y - 1}{255} = \frac{256^{i/8} - 1}{255} = \frac{2^i - 1}{255}$$

$i$	0	1	2	3	4	5	6	7	8
$y = i/8$	0	1/8	2/8	3/8	4/8	5/8	6/8	7/8	1
$x = (2^i - 1) / 255$	0	1/255	3/255	7/255	15/255	31/255	63/255	127/255	1
斜率 $\times 255$	1/8	1/16	1/32	1/64	1/128	1/256	1/512	1/1024	
段号	1	2	3	4	5	6	7	8	

## 2.2.2 音频的数字化和符号化

---

- 量化

- 用13折线法中的（第一和第二段）**最小量化间隔**作为均匀量化时的量化间隔：

- （1）若对13折线法中第1至8段进行均匀量化时共多少个量化间隔，需要比特数是多少？

- （2）13折线法中第1至8段进行非均匀量化：共多少个量化间隔，需要比特数是多少？

## 2.2.2 音频的数字化和符号化

- 量化

- 用13折线法中的（第一和第二段）**最小量化间隔**作为均匀量化时的量化间隔：

- 对于13折线法，第一至第八段包含的均匀量化间隔数分别为16、16、32、64、128、256、512、1024，共有**2048**个均匀量化间隔；
    - 对于**非均匀量化**，只需**128**个不同的量化间隔。

- 在保证小信号量化间隔相等的条件下：

- 均匀量化需要11比特编码；
    - 非均匀量化只要7比特就够了。

## 2.2.2 音频的数字化和符号化

---

- 编码

- 根据一定的协议或格式把模拟信息转换成二进制比特流的过程。

- 最简单的：直接用量化后的二进制数作为输出的数字信号，叫脉冲编码调制（PCM）编码。

- 优点：信号质量好，最大限度的保留了信号的原始信息。

- 缺点：数据量非常大，需要大量的存储空间。



## 2.2.2 音频的数字化和符号化

- 编码

- 自然码

- 无符号的二进制代码;

- 折叠码

- 在自然码最高位增加了一个符号位而构成，用以表示信号极性，一般用“0”表示负，而用“1”表示正。

- 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

## 2.2.2 音频的数字化和符号化

电平 序号	自然码 NBC				折叠码 FBC				格雷码 RBC			
	$b_1$	$b_2$	$b_3$	$b_4$	$b_1$	$b_2$	$b_3$	$b_4$	$b_1$	$b_2$	$b_3$	$b_4$
15	1	1	1	1	1	1	1	1	1	0	0	0
14	1	1	1	0	1	1	1	0	1	0	0	1
13	1	1	0	1	1	1	0	1	1	0	1	1
12	1	1	0	0	1	1	0	0	1	0	1	0
11	1	0	1	1	1	0	1	1	1	1	1	0
10	1	0	1	0	1	0	1	0	1	1	1	1
9	1	0	0	1	1	0	0	1	1	1	0	1
8	1	0	0	0	1	0	0	0	1	1	0	0
7	0	1	1	1	0	0	0	0	0	1	0	0
6	0	1	1	0	0	0	0	1	0	1	0	1
5	0	1	0	1	0	0	1	0	0	1	1	1
4	0	1	0	0	0	0	1	1	0	1	1	0
3	0	0	1	1	0	1	0	0	0	0	1	0
2	0	0	1	0	0	1	0	1	0	0	1	1
1	0	0	0	1	0	1	1	0	0	0	0	1
0	0	0	0	0	0	1	1	1	0	0	0	0

## 2.2.2 音频的数字化和符号化

---

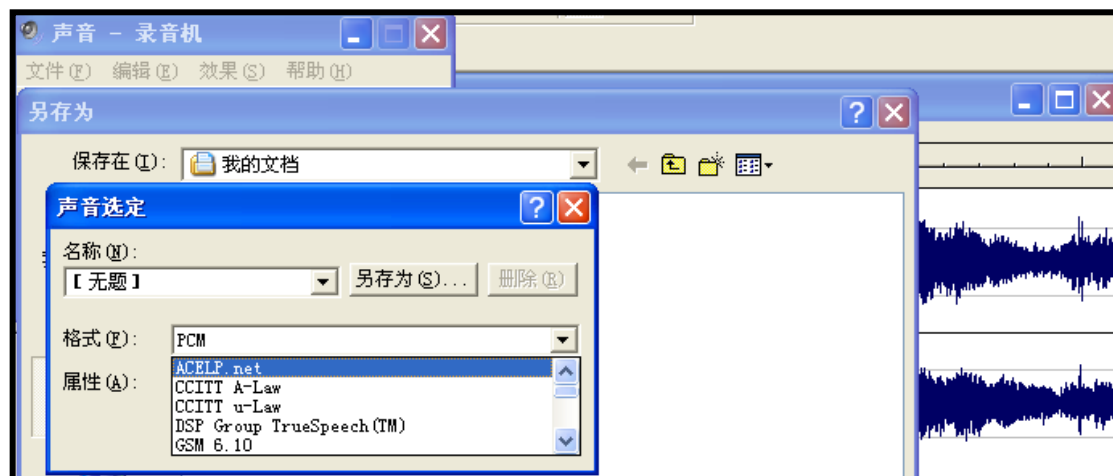
- 编码

- 多媒体信息的特点之一是存在冗余，可以采用不同的压缩技术进行压缩。
- 编码/压缩技术不同->数字音频文件格式不同->播放文件时需要不同的解码技术。

## 2.2.2 音频的数字化和符号化

- 例子

- 对于采样频率为8kHz，样本精度为13位、14位或者16位的输入信号，使用 $\mu$ 律压扩编码或者使用A律压扩编码，经过PCM编码器之后每个样本需8位二进制存储，输出的数据率为64 kb/s。
- 这个数据就是CCITT推荐的**G.711标准：话音频率脉冲编码调制。**



## 2.2.2 音频的数字化和符号化

与数字音频相关的重要特性:

### ✴ 采样频率

采样频率与声音的质量关系最为紧密。采样频率越高，声音质量越接近原始声音，所需的存储量便越多。标准的采样频率有三个：44.1KHz，22.05kHz，和11.025kHz。

### ✴ 采样位数

存放一个采样点所需的比特数。一般的采样位数为8位或16位，即把声音采集为256等份或65536等分。

## 2.2.2 音频的数字化和符号化

### ★ 声道数

有单声道、双声道和多声道。如多种语言音频混存时，需要多声道

### ★ 数据量

(采样频率 × 每点采样位数 × 声道数)

数据量 =  $\frac{\text{(采样频率} \times \text{每点采样位数} \times \text{声道数)}}{8}$  (字节/秒)

8

## 2.2.2 音频的数字化和符号化

质量	采样率 ( KHz )	样本精 度 ( bit )	单声道/ 立体声	数据率 ( 未压缩 ) ( KB/s )	频率范围 ( Hz )
电话	8	8	单声道	8	200-3400
AM	11.025	8	单声道	11.0	50-7000
FM	22.050	16	立体声	88.2	20-15000
CD	44.1	16	立体声	176.4	20-20000
DAT	48	16	立体声	192.0	20-20000
DVD	192 ( 最大 )	24 ( 最大 )	最高6声道	1200.0 ( 最大 )	0-96000 ( 最大 )

数字录音带

## 2.2.2 音频的数字化和符号化

### ★ 信噪比SNR(Signal Noise Ratio)

#### 信号量化噪声比SQNR

- 量化噪声：采样点的**模拟值**和最近的**量化值**之间的**差**。
- 误差最大可以达到离散间距的一半。

$$\begin{aligned} SQNR &= 20 \log_{10} \frac{V_{signal}}{V_{quan\_noise}} = 20 \log_{10} \frac{2^{N-1}}{\frac{1}{2}} \\ &= 20 \times N \times \log 2 = 6.02 N(\text{dB}) \end{aligned}$$

- 量化精度为  $N$  位，数值信号的取值范围  $-2^{N-1}$  到  $2^{N-1}-1$ ；
- 考虑峰值情况， $V_{signal}$  取为  $2^{N-1}$ ， $V_{quan\_noise}$  取  $1/2$ 。



## 2.2.2 音频的数字化和符号化

✴ 信噪比SNR(Signal Noise Ratio)

信号量化噪声比SQNR

$$\begin{aligned} SQNR &= 20 \log_{10} \frac{V_{signal}}{V_{quan\_noise}} = 20 \log_{10} \frac{2^{N-1}}{\frac{1}{2}} \\ &= 20 \times N \times \log 2 = 6.02 N(\text{dB}) \end{aligned}$$

- 采样点中每个位增加**6dB**的分辨率，即采样精度每增加**1位**，信噪比增加**6dB**。

## 2.2.2 音频的数字化和符号化

---

- 数字音频文件格式

- WAV文件

- 也叫波形文件，是Microsoft公司的声音文件格式，存储文件扩展名为“.wav”。
    - 文件由三部分组成：文件头(标明是WAVE文件、文件结构和数据的总字节数)、数字化参数(如采样率、声道数、编码算法等)，最后是实际波形数据。
    - 数据量大，音质最好。大多数压缩格式的声音在此基础上经过数据的重新编码来实现，压缩前和回放时都是用WAV格式。

## 2.2.2 音频的数字化和符号化

---

- 数字音频文件格式

- AIF或AIFF文件

- AIF（Audio Interchange File Format，音频交换文件格式）是Apple公司的声音文件格式，被很多程序支持。
    - Windows的Convert工具可以把AIF格式的文件转换成Microsoft的WAV格式的文件。

## 2.2.2 音频的数字化和符号化

---

- 数字音频文件格式

- WMA文件

- Windows Media Audio，是Microsoft公司开发的**网上流式**数字音频压缩技术，可以一边下载一边播放，因此可以轻松的实现在线广播。
    - 同时兼顾了**保真度**和**网络传输**的要求，在64kbit/s的码率下，达到接近**CD**的音质。
    - 支持**防复制**功能，可以限制**播放时间**和**播放次数**甚至于**播放的机器**等。

## 2.2.2 音频的数字化和符号化

---

- 数字音频文件格式

- RA文件 (RM、RAM)

- 是Real Networks公司开发的一种流媒体音频文件。很流行，在低速率的广域网上实时传输音频信息。
    - 网络连接速率不同，客户端获得的音质也不相同
      - 14.4kbit/s的网络连接：获得调幅质量的音质；
      - 28.8kbit/s的网络连接：获得广播级的音质；
      - ISDN或更快的：CD音质的声音。
    - 不但支持边下载边播放，也同样支持使用特殊协议来隐匿文件的真实网络地址，实现只在线播放而不提供下载的方式。

## 2.2.2 音频的数字化和符号化

---

- 数字音频文件格式

- PCM格式

- 是模拟的音频信号经模/数转换直接形成二进制序列的文件。没有附加的文件头和文件介绍标志。
    - 可将PCM格式转换为VOC格式、WAV格式。

## 2.2.2 音频的数字化和符号化

---

- 数字音频文件格式

- MP1、MP2、MP3

- 是指MPEG运动图像专家组所制定的音频文件格式，根据压缩质量和编码复杂程度的不同分为3层。
    - 是一种有损压缩，具有很高的压缩率，MP1、MP2的压缩率是4:1-8:1，而MP3的压缩率高达12:1。
    - MP3体积小、有较好的声音质量。是目前最流行的音乐文件。

## 2.2.2 音频的数字化和符号化

---

- 数字音频文件格式

- MIDI文件

- 是**数字音乐电子合成乐器**的统一国际标准，规定计算机音乐程序、电子合成器和其他电子设备之间交换信息和控制信号的方法。
    - **MIDI文件**是“乐谱和指挥”。
    - **MIDI的数据量较小**，适合作为**背景音乐**音响效果，用来播**长时间、高质量**的音乐。
    - 扩展名为**.mid**和**.rmi**。



## 2.2.2 音频的数字化和符号化

---

- 数字音频文件格式

- CD-DA文件

- 是CD光盘采用的文件格式，是一个.cda文件，这只是个索引信息，并不是真正的包含声音信息。不论CD音乐的长度，文件都是44字节长。
    - 不能直接复制文件到硬盘上播放，需要使用软件把CD格式的文件转换成WAV格式。

## 2.2.2 音频的数字化和符号化

---

- 数字音频文件格式

- MP4

- 是对MP3的大众化、无版权的一种保护格式。
    - 使用MPEG-2 AAC (Advanced Audio Coding, a2b、AAC) 技术, MPEG-2是MPEG针对数码电视提出的。特点是音质更加完美而压缩比更加大(1:15)。

## 2.2.2 音频的数字化和符号化

---

### 2. 声音的符号化

波形声音可以把音乐、语音都进行数据化并且表示出来，但是并没有把它看成音乐和语音。

对于声音的符号化（也可以称为**抽象化**）表示包括两种类型：一种是**音乐**，一种是**语音**。

## 2.2.2 音频的数字化和符号化

### 音乐的符号化 - MIDI

- MIDI(Music Instrument Digital Interface)是指**乐器数字接口**的国际标准，是音乐与计算机结合的产物。
- MIDI标准规定了不同厂家的电子乐器与计算机连接的**电缆和硬件**，还指定了从一个装置传送数据到另一个装置的**通信协议**。
- 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**。

## 2.2.2 音频的数字化和符号化

- **MIDI**不是把音乐的**波形**进行数字化采样和编码，而是将数字式电子乐器的弹奏过程以**命令符号**的形式记录下来，如按了哪一个键、力度多大、时间多长等。

例如：按下**MIDI**键盘中间的一个“**C**”键时，**MIDI**键盘会发送一个**三字节**组成的消息，用**16进制**表示为：**90 3C 40**

其中**90**是状态字节，表示字符**开始**，且向**0**号声道传送；

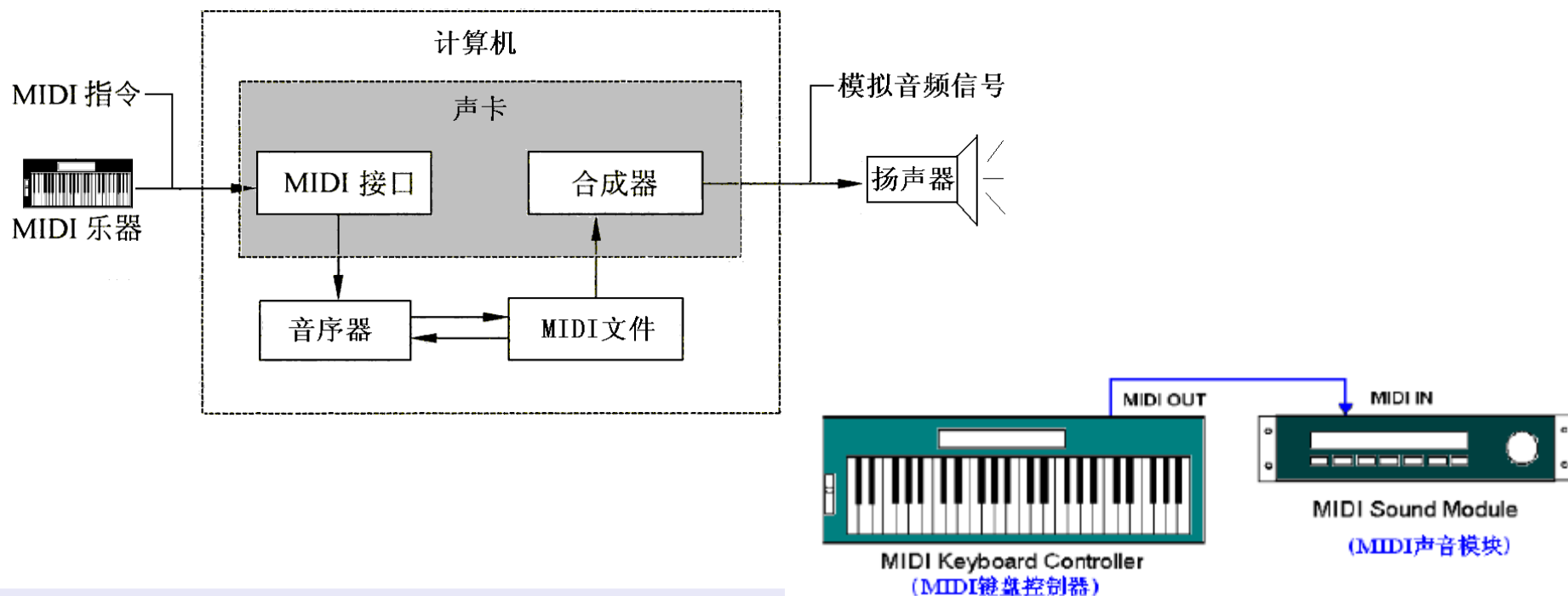
**3C**表示击键位置；**40**表示击键的速度

- 在播放这首乐曲时，根据记录的乐谱指令，通过**音乐合成器**生成音乐声波，放大后由**扬声器**播出。

## 2.2.2 音频的数字化和符号化

### • MIDI基本结构

- MIDI乐器、MIDI指令、MIDI接口、音序器、MIDI文件、合成器



## 2.2.2 音频的数字化和符号化

---

- **MIDI乐器（输入设备）**：制作MIDI音乐，演奏通过**MIDI接口**被**音序器**接收并存储为音序内容。能产生特定声音的合成器，如**MIDI键盘**、**MIDI吉他**等；它们相互间的数据传送符合MIDI的通信约定。
- **MIDI指令（消息）**：是指**乐谱的数字描述**。**MIDI软件通信协议**，用数字指令描述的音乐乐谱，其中包含音符、强度、定时及乐器的指派等（90 3C 40）。
- **MIDI接口**：**MIDI硬件**通信协议，可使电子乐器互连或与计算机硬件端口相连，可发送和接收MIDI消息。

## 2.2.2 音频的数字化和符号化

- **音序器（编曲机）**：实际上是记录了音乐的一般要素，拍子、音高、节奏等，以数字的形式记录下它们，让音源发音。**MIDI**文件的本质内容实际上就是音序内容。

最初**专用硬件**，现多指计算机上用于编辑音乐的软件，如：**FL Studio (Fruity Loops Studio)**



- **合成器/音源**：是一个装了很多音色的东西，所以它是用来发声音的；声卡。

**波表合成**：把真实乐器发出的声音以数字的形式记录下来，播放时再加以调整、修饰和放大，生成各种音阶的音符。

不同机器音色库不一样，同样**MIDI**文件可能效果差别。



## 2.2.2 音频的数字化和符号化

---

- MIDI

- 音轨

- 在音序器中，音乐按照音轨来组织；
    - 在录音和回放时，可以打开或关闭一个音轨。

- 通道

- 某种乐器和一个专门的MIDI通道相对应；
    - 用来分隔消息，一共16个通道（0-15），消息的后4位用来存储通道编号。
      - 例如，通道1代表钢琴，通道10代表鼓

- MIDI消息：通道消息、系统消息

## 2.2.2 音频的数字化和符号化

### 1、通道消息

- 由1个状态字节和随后的2个数据字节组成;
  - 状态字节: 通道编号等
- 状态字节: 最高位是1; 数据字节: 最高位是0。

十六进制	二进制	数据	描述
8x	1000xxxx	nn vv nn=音符号 vv=速度	音符关闭 (释放键盘)
9x	1001xxxx	nn vv nn=音符号 vv=速度	音符打开 (按下键盘)
Ax ~ Ex			

## 2.2.2 音频的数字化和符号化

---

### 1、通道消息

- 要在**13号通道**上以**最高速度**来播放**80号音符**，MIDI设备会发送3个16进制值：

9C

10011100

50

01010000

7F

01111111

## 2.2.2 音频的数字化和符号化

---

### 2、系统消息

- 不带通道编号，表示不是针对某一个具体的通道，如同步的时钟信号等。
- **种类 + 参数**：如以**FF**开头的命令 字节数 数据

**FF 51 03 xx xx xx**：四分音符的绝对时间长度

**FF 51 03 07 A1 20**：07A120 (500000) 微秒  
**00 78**：120 tick

**1个tick**： $500000 / 120 = 4166$  微秒

## 2.2.2 音频的数字化和符号化

---

### 2、系统消息

- 不带通道编号，表示不是针对某一个具体的通道，如同步的时钟信号等。
- **种类 + 参数**：如以**FF**开头的命令 字节数 数据

**FF 59 02 sf mi**：音调符号

**sf**：升调/降调    **0**：基准C调

**mi**：大调/小调    **0**：大调； **1**：小调

**FF 59 02 00 00**

## 2.2.2 音频的数字化和符号化

---

### MIDI文件结构:

- ❖ 一个MIDI文件基本上由两部分组成，一个**头块**和紧接的一个或者多个**轨道块**
  - ❖ 头块：MThd，描述文件的格式、轨道块数量等内容
  - ❖ 轨道块：MTrk

## 2.2.2 音频的数字化和符号化

### MIDI文件结构:

Chunk 块结构					
MIDI 文件	类型	长度	数据		
	MThd	6	<格式>	<tracks>	<division>
	MTrk	<长度>	<delta-time><event>		
	....				
	MTrk	<长度>	<delta-time><event>		

## 2.2.2 音频的数字化和符号化

---

### MIDI头块结构:

- ❖ 头块出现在文件的开头，当头块数据为 **4D 54 68 64 00 00 00 06 ff ff nn nn dd dd** 时
  - ❖ 前4个字节等同于ASCII码MThd;
  - ❖ 接着的4个字节是头的大小，一直是6，因为现行的头信息是6个字节;
  - ❖ ff ff是文件的格式，有3种格式
    - ❖ 0: 单轨; 1: 多轨，同步; 2: 多轨，异步



## 2.2.2 音频的数字化和符号化

---

### MIDI头块结构:

- ❖ 头块出现在文件的开头，当头块数据为 **4D 54 68 64 00 00 00 06 ff ff nn nn dd dd** 时
  - ❖ **nn nn** 是MIDI文件中的轨道数;
  - ❖ **dd dd** 是每个4分音符delta-time节奏数。

## 2.2.2 音频的数字化和符号化

### MIDI头块结构:

Header Chunk				
Chunk 类型	长度	数据		
4 字节  (ASCII)	4 字节  (32 位二进制)  Chunk 数据部分的长度。这个是一个 32 位二进制数。	<长度>		
		16 位	16 位	16 位
		<格式>  MIDI 文件的格式。这是一个 16 位二进制数，有效格式是：0、1 和 2。	<tracks>  MIDI 文件中 track chunk 的数量。这是一个 16 位二进制数。	<division>  这个定义在 MIDI 文件中（一个）单位的 delta-time 数。这是一个 16 位二进制数。
MThd	<长度>	<delta-time><event>		