



# 多媒体技术

## 2.2.1 声音心理学

---

- a. 音强

- 又叫响度，表示声音能量的强弱程度，主要取决于声波振幅的大小。一般用声压或声强来计量，单位是分贝（dB）。
- 常说的“音量”也是指响度。

## 2.2.1 声音心理学

- 声压级

— 将声压的有效值以对数的形式表示声音强弱的数值称为声压级。

$$SPL = 20 \lg ( P_{rms} / P_{ref} ) \quad (\text{单位用分贝dB})$$

$P_{rms}$  : 计量点的声压有效值

$P_{ref}$  : 零声级的参考声压值 ( $P_{ref}=2 \times 10^{-5}$ 帕)

## 2.2.1 声音心理学

- 声压级

— 将声压的有效值以对数的形式表示声音强弱的数值称为声压级。

$$SPL = 20 \lg ( P_{rms} / P_{ref} ) \quad (\text{单位用分贝dB})$$

1、 $P_{ref}$ 为具有正常听力的年轻人对1kHz的声音刚好能察觉的声压值。

2、声压级实际上是一种相对量，是某点的声压与零声压的比，是描述声音变化的动态范围的物理量。

## 2.2.1 声音心理学

---

- b. 音调

- 音乐也称音高，表示人耳对声调高低的主观感受。客观上音高大小主要取决于声波基频的高低。
- 基频越低，给人的感觉越低沉。基频频率增加一倍，音乐上称提高了一个八度。

## 2.2.1 声音心理学

---

- 在任一时刻  $t$ ，声波可以分解成一系列正弦波的线性叠加：

$$f(t) = \sum_{n=0}^{\infty} A_n \sin(n\omega t + \varphi_n)$$

- $A_n$  是振幅，表示声音的强弱
- $\omega$  是声波的基频，表示声音音调的高低
- $\varphi_n$  是  $n$  次谐波的初相位

## 2.2.1 声音心理学

- 基频

- 根据傅里叶变换的理论，声音可以分解成若干个不同频率纯音的叠加。
- 这些频率都是某一频率的倍数，这一频率就称作基频，也就决定了这个音的音高。

- 基音

- 基频所产生的音称为基音。（频率最低）

- 泛音

- 基频为 $f$ ，则频率为 $2f$ 的音称为第一泛音，频率为 $3f$ 的音称为第二泛音。
- 频率为基频整数倍的正弦，振荡为谐波。基波频率3倍的波称之为三次谐波。

## 2.2.1 声音心理学

---

- c. 音色

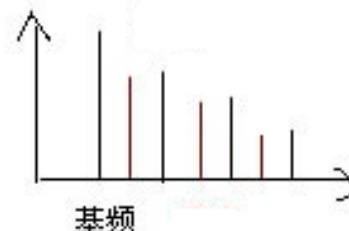
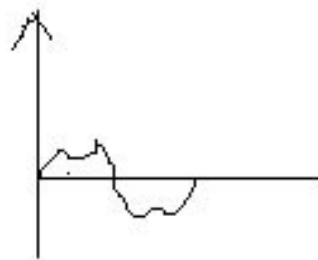
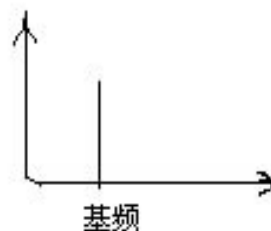
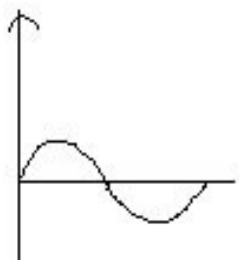
- 又称音品，由声音波形的谐波频谱分布决定。
- 在相同响度及音高情况下，辨别不同的乐器，或者不同的说话的人，就是因为这些声源产生的频谱的差异。



## 2.2.1 声音心理学

### • 频谱

- 对一个声源发出的声音的**频率**成份和**强度**的分析，叫频谱分析，并且可用**频谱图**表示，其**纵轴**为**幅度**，**横轴**为**频率**。
- 将声音的**基频**和各次**谐波**按**频率**分别用一竖线画在横轴上，它的长短表示幅度的大小。



## 2.2.1 声音心理学

---

- d. 音长

- 音长就是发声过程**延续时间的长短**。声音的长短取决于发声物体振动的持久程度。
- 不同长短的音相互结合起来就产生了音乐的节奏、节拍，从而构成了旋律的骨架。
- 在普通话里，**轻声音节**的音长要比非轻声的音节短，如“东西”中的“西”要比“西南”中的“西”短一些。再如英语中[u]和[u:]的不同就是长短的不同。

# 2.2.1 声音心理学

---

## 2. 听觉特性

### ✴ 等响曲线

由于响度与频率和强度有关，所以在不同频率上的强度是不同的。

例如1000Hz、40分贝、持续0.5秒；

再给一个音也持续0.5秒，但频率不同，通过调整使其响度听起来一样，得到的这样一组曲线称之为等响曲线。

## 2.2.1 声音心理学

---

### ☀ 等响曲线

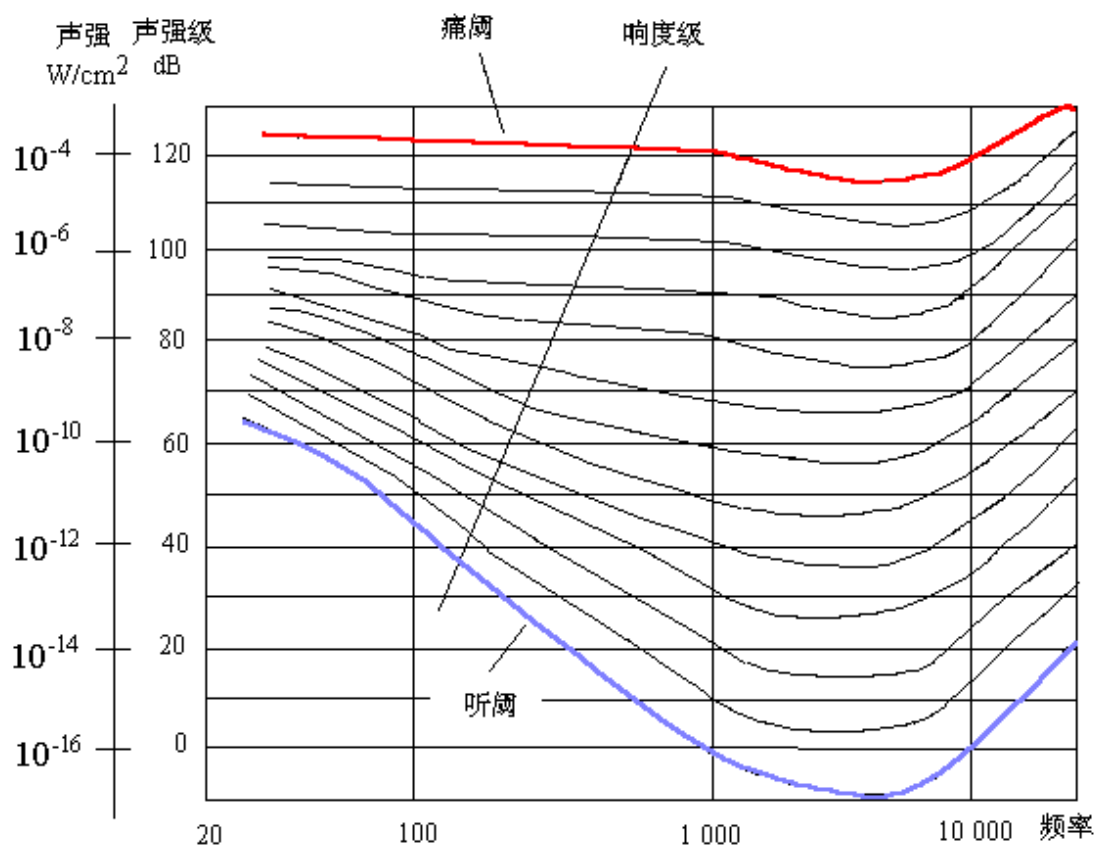
等响曲线描述的是响度与频率和强度的关系。

从声音心理学考虑，对**同一响度**的声音在**频率**上和**强度**上可以有很大的差别，这对声音表现有重要意义。

- **听阈**：当声音弱到人耳朵**刚可听见时**的声音强度。
- **痛阈**：声音强到使人耳**感到疼痛时**的声音强度。
- **听觉范围**：位于听阈和痛阈之间。

## 2.2.1 声音心理学

### ☀ 等响曲线



## 2.2.1 声音心理学

---

### ☀ 掩蔽（masking）

声音的响度不仅取决于自身的强度和频率，而且也依赖于同时出现的其它声音。

各种声音可以互相掩蔽，也就是说一种声音的出现可能使得另一种声音难于听清。由于声音的掩蔽效果，可以欺骗人的听觉。

例如，本来是多种频率的声音的复合，但听众以为是另一种声音。所以，声音的掩蔽特性常常用于声音的压缩。

## 2.2.1 声音心理学

---

### ☀ 掩蔽（masking）

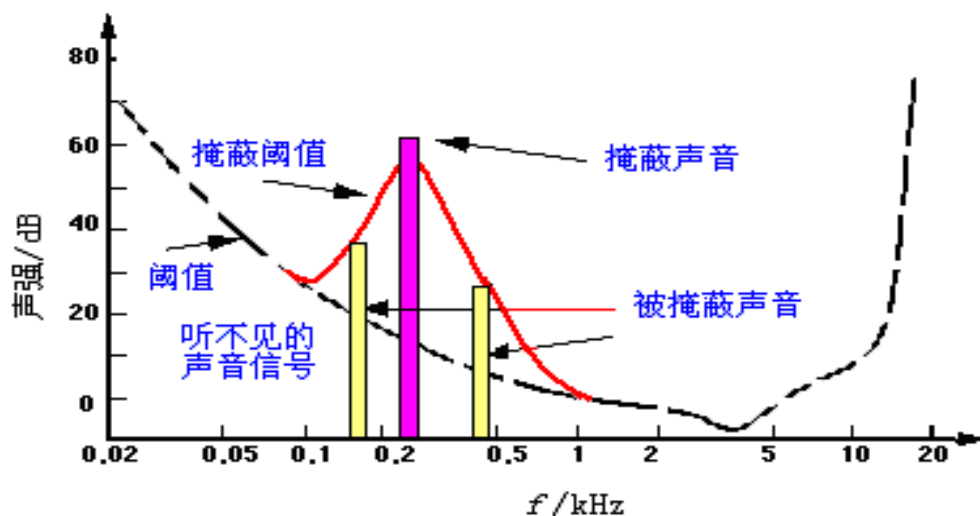
- 一种频率的声音阻碍听觉系统感受另一种频率的声音的现象。
  - 前者称为掩蔽声音
  - 后者称为被掩蔽声音
  - 掩蔽可分成
    - 频域掩蔽
    - 时域掩蔽

## 2.2.1 声音心理学

### ☀ 掩蔽 (masking)

#### • 频域掩蔽

- 一个**强纯音**掩蔽在其附近**同时发声**的**弱纯音**的特性，也称**同时掩蔽**。



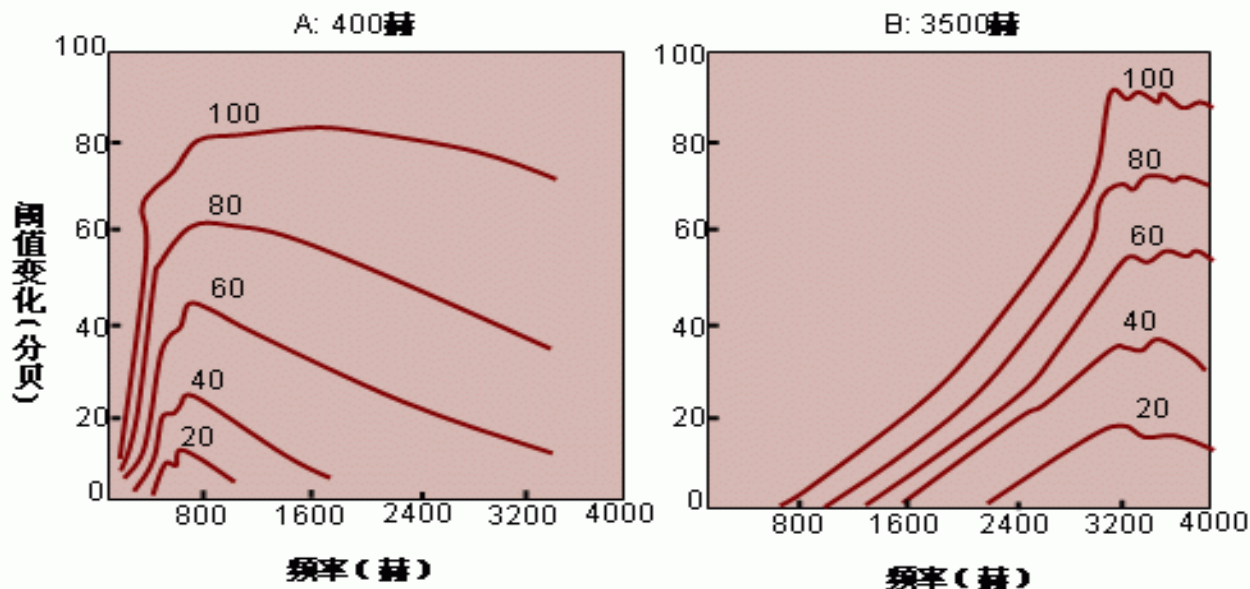


## 2.2.1 声音心理学

### ☀ 掩蔽（masking）

#### • 频域掩蔽

- A、B为掩蔽音；横坐标为各种频率的被掩蔽音，纵坐标为掩蔽阈值。

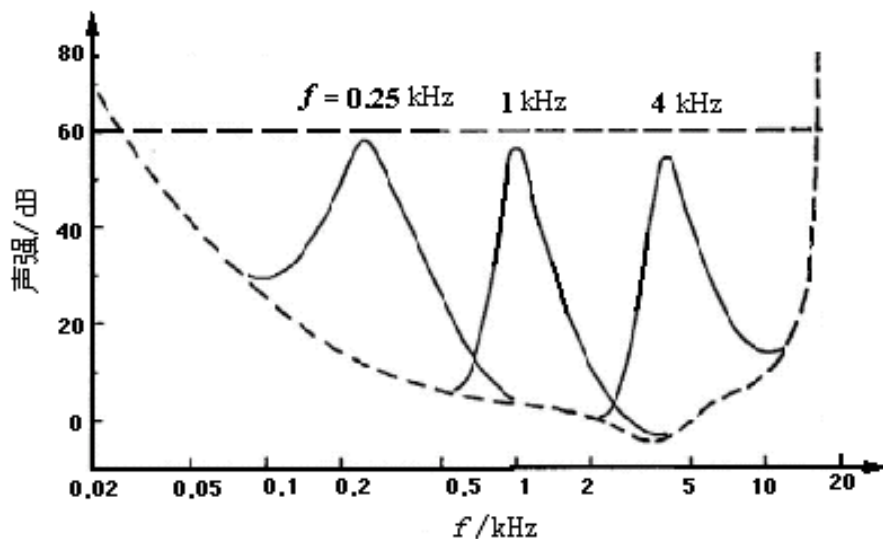


## 2.2.1 声音心理学

### ☀ 掩蔽 (masking)

#### • 频域掩蔽

- 曲线表示为250 Hz, 1 kHz和4 kHz纯音的掩蔽效应, 它们的声强均为60 dB。
- 低频纯音可有效地掩蔽高频纯音, 相反则不明显。



## 2.2.1 声音心理学

---

### ☀ 掩蔽（masking）

#### • 时域掩蔽

– 在时间上相邻的声音之间的掩蔽现象。

- 滞后掩蔽：信号出现在掩蔽音消失后出现的现象，可以持续**50 ~ 200 ms**。
- 超前掩蔽：信号出现在掩蔽音出现之前产生的现象。某个信号遮掩刚好在它产生之前的信号。

虽然对超前掩蔽有许多研究报告，但这种现象依然令人费解。超前掩蔽很短，通常只有大约**2 ~ 20 ms**。

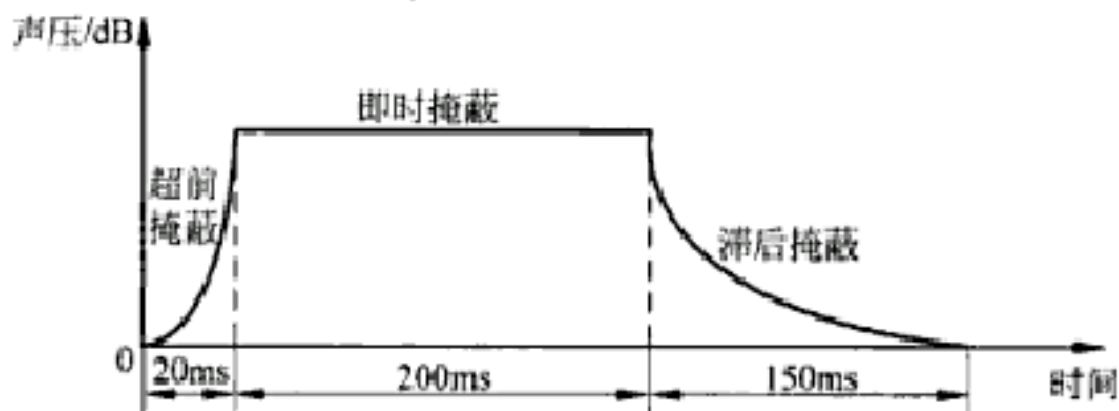
## 2.2.1 声音心理学

### ☀ 掩蔽（masking）

#### • 时域掩蔽

— 产生时域掩蔽的主要原因

- 人的大脑处理信息需要花费一定的时间。



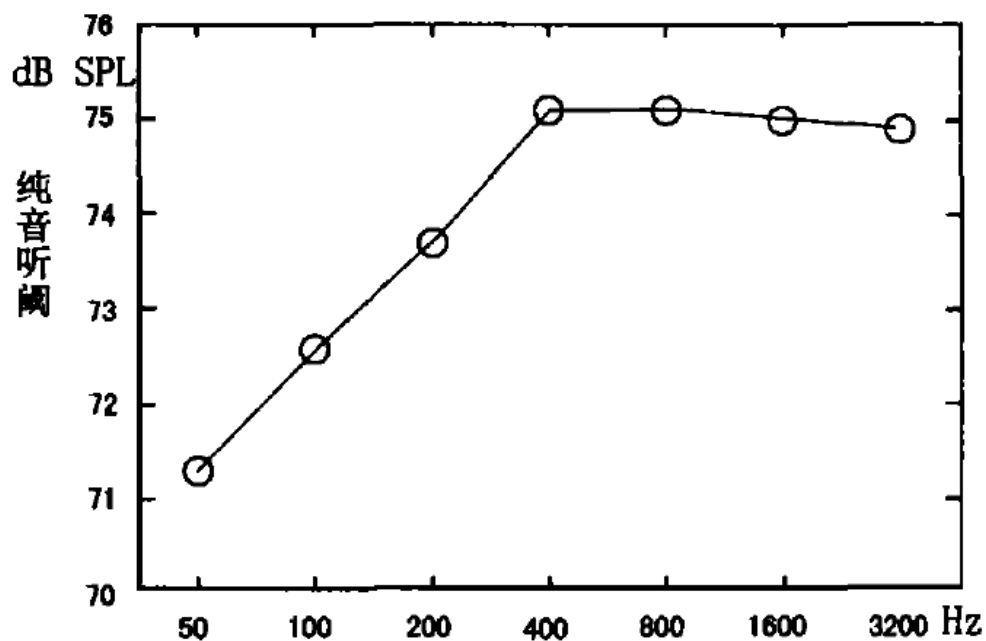
## 2.2.1 声音心理学

### ✪ 临界频带

- 通过带宽可变的**噪声**对**纯音信号**的掩蔽实验发现：
  - 当以**1000Hz**为中心频率的噪声**增加带宽**时,它对**1000Hz**纯音信号的掩蔽效应也随着增加。
  - 但带宽增至一定程度,再增加就对掩蔽的改变不起作用。
  - 就是说,这个噪声的掩蔽作用只限制在这个频带内,以外的声音无作用。这个频带称为**临界频带**,它随频率的提高而加宽。
  - 临界频带的掩蔽作用最明显。

## 2.2.1 声音心理学

### ☀ 临界频带



2kHz 纯音听阈随掩蔽噪声带宽变化的特性

## 2.2.1 声音心理学

---

### ☀ 临界频带

- 通常认为，在听力的音频频率范围内有**25**个临界频带。
- 临界频带的单位为**Bark**(巴克)
  - 1 **Bark**等于一个临界频带的宽度
- 临界频带的宽度随声音频率的变化而变化
  - 在**低频**端，宽度小于**100 Hz**，可认为接近于常数
  - 在**高频**端，宽度近似线性增加，宽度可大到**4 kHz**

## 2.2.1 声音心理学

---

### ☀ 临界频带

$$\text{临界频段值(Bark)} = \begin{cases} f / 100 & f < 500 \\ 9 + 4\log_2(f / 1000) & f \geq 500 \end{cases}$$

$$df = 25 + 75 * [1 + 1.4 f^2]^{0.69}$$

中心频率 $f$ 的单位是kHz，临界带宽 $df$ 的单位为Hz



## 2.2.1 声音心理学

### ☀ 临界频带

表 11-1 理想的临界频带[6]

频带号	临界频率 (Hz)				频带号	临界频率 (Hz)			
(Bark)	低端	中心频率	高端	宽度	(Bark)	低端	中心频率	高端	宽度
0	0	50	100	100	13	2000	2150	2320	320
1	100	150	200	100	14	2320	2500	2700	380
2	200	250	300	100	15	2700	2900	3150	450
3	300	350	400	100	16	3150	3400	3700	550
4	400	450	510	110	17	3700	4000	4400	700
5	510	570	630	120	18	4400	4800	5300	900
6	630	700	770	140	19	5300	5800	6400	1100
7	770	840	920	150	20	6400	7000	7700	1300
8	920	1000	1080	160	21	7700	8500	9500	1800
9	1080	1170	1270	190	22	9500	10500	12000	2500
10	1270	1370	1480	210	23	12000	13500	15500	3500
11	1480	1600	1720	240	24	15500	19500	22050	6550
12	1720	1850	2000	280					

## 2.2.1 声音心理学

---

### ☀ 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。

## 2.2.1 声音心理学

---

### ☀ 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。

主要应用于虚拟现实中。

方位的线索是各种声音到达两耳的**精确时间**和**强度**。

## 2.2.1 声音心理学

---

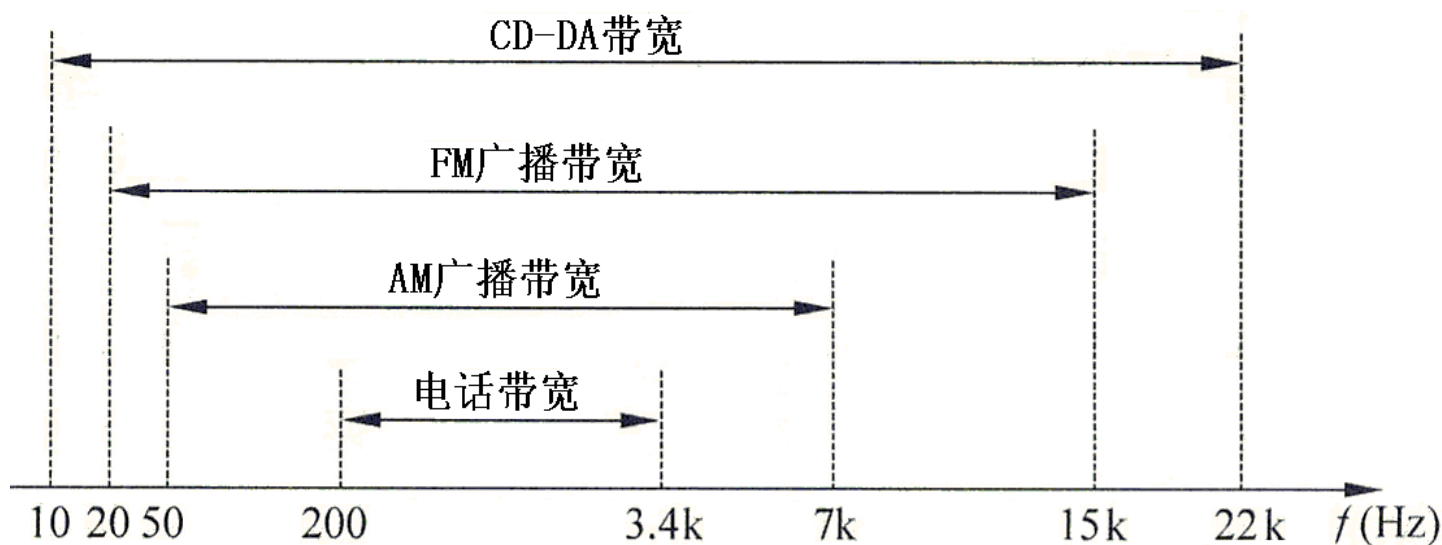
### 3. 声音音质-质量标准

#### ✦ 频带宽度

音频信号所包含的**谐波**分量越丰富，**音色**越好。在广播通信和数字音响系统中，以声音信号所包含的谐波分量的**频率范围**来衡量声音的质量，即**带宽**。

## 2.2.1 声音心理学

### ☀ 频带宽度



不同质量的声音的频带对比示意图

## 2.2.1 声音心理学

---

### ★ 动态范围

- 指声音**最大声压级**和**最小声压级**之间的差值。
- 动态范围**越大**，说明音频信号强度的相对变化范围越大，**音响效果越好**。
- 每种声源的动态范围依据各自的特性有所不同。
  - 女声的动态范围为**25~50dB**,
  - 男声为**30~50dB**,
  - 交响乐队的动态范围大于**100dB**。

## 2.2.1 声音心理学

---

### ✴ 动态范围

– 动态范围可以用信号的相对强度表示：

**信号的动态范围 =**

**$20 \times \lg(\text{信号最大强度} / \text{信号最小强度})(\text{dB})$**

– 其中信号可以用电压或功率衡量。

– 因为是一种比例关系，故只要采用相同的度量单位，其结果都是一致的。

## 2.2.1 声音心理学

---

### ★ 动态范围

- 动态范围一般用dB为单位来计量。
  - FM(调频)广播的动态范围约60dB,
  - AM(调幅)广播的动态范围约40dB。
  - CD-DA的动态范围约100dB,
  - 数字电话约50dB。



## 2.2.1 声音心理学

### ★ 信噪比SNR(Signal Noise Ratio)

- 是**有用信号**与**噪声**之**比**的简称。
- 信噪比大，在一定程度上能够掩蔽噪声，从而获得较好的声音效果。

$$SNR = 10 \log_{10} \left[ \frac{(V_{signal})^2}{(V_{noise})^2} \right] = 20 \log_{10} \left( \frac{V_{signal}}{V_{noise}} \right)$$

$V_{signal}$ 表示信号电压，

$V_{noise}$ 表示噪声电压，

**SNR**的单位为分贝(db)

## 2.2.1 声音心理学

---

### ☀ 信噪比SNR(Signal Noise Ratio)

- 信噪比不仅是声音设备的性能指标，在声音的录制和播放时，也要注意环境噪声。**录制**时应尽可能减小环境噪音。**输出**时应使音量适当大，以减少环境噪音对听音的影响。
- 一般**话筒**和**音箱**的信噪比在75dB以上。**声卡**的信噪比在85 ~ 95dB。

## 2.2.1 声音心理学

---

### ☀ 数据量

- 音质与数据采样频率和数据位数有关：
  - 采样频率越低，位数越少，音质越差；
  - 采样频率越高，位数越多，音质越好。

## 2.2.2 音频的数字化和符号化

---

### 1. 音频的数字化与再现

在计算机中，所有的信息都以**数字**来表示。声音信号也是由一系列的**数字**来表示的，称为**数字音频**。数字音频的特点就是保真度好，动态范围大。

数字声音是一个数据序列。它是由外界声音经过**采样**、**量化**和**编码**后得到的。

## 2.2.2 音频的数字化和符号化

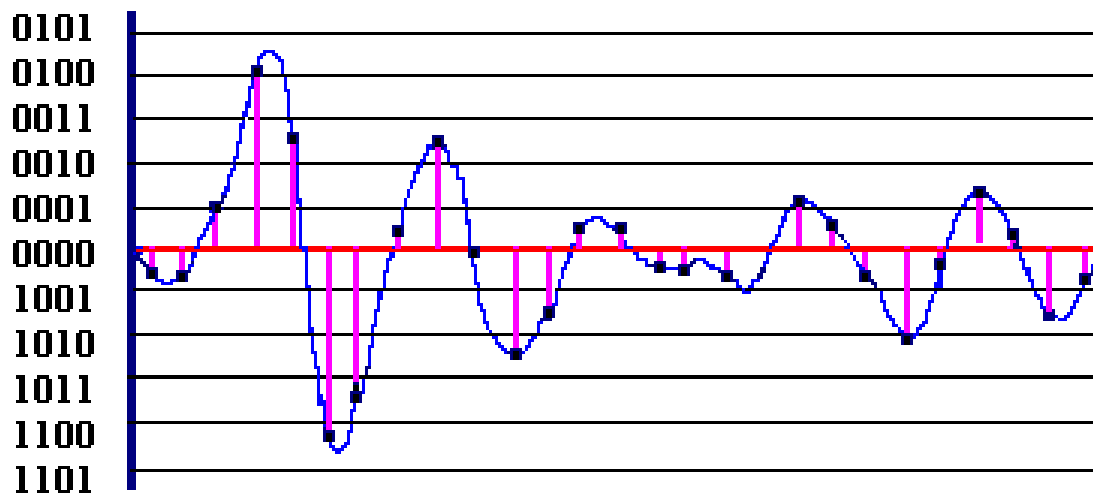
---

- 采样

- 模拟信号：连续时间函数  $X(T)$ ;
- 采样：按照规定的时间间隔  $T$  采集一段时间的模拟信号，以获得采集时刻模拟信号的振幅值即离散信号  $X(nT)$ 
  - 其时间间隔  $T$  称为采样周期， $1/T$  称为采样频率。

## 2.2.2 音频的数字化和符号化

- 采样



## 2.2.2 音频的数字化和符号化

- 采样

- 数字音频在播放声音时必须重新还原为模拟音频。
- 为了保证还原的音频信号不失真，数字化时采样频率必须满足采样定理的要求：
  - 采样频率至少是信号最高频率的两倍。又称奈奎斯特(Nyquist)定理。
- 例如，电话语音的信号频率约为3.4 kHz，若采样频率选为8kHz，就能无失真地重放原始声音。

## 2.2.2 音频的数字化和符号化

---

- 采样

- 采样率有多种，最常用的3个标准：

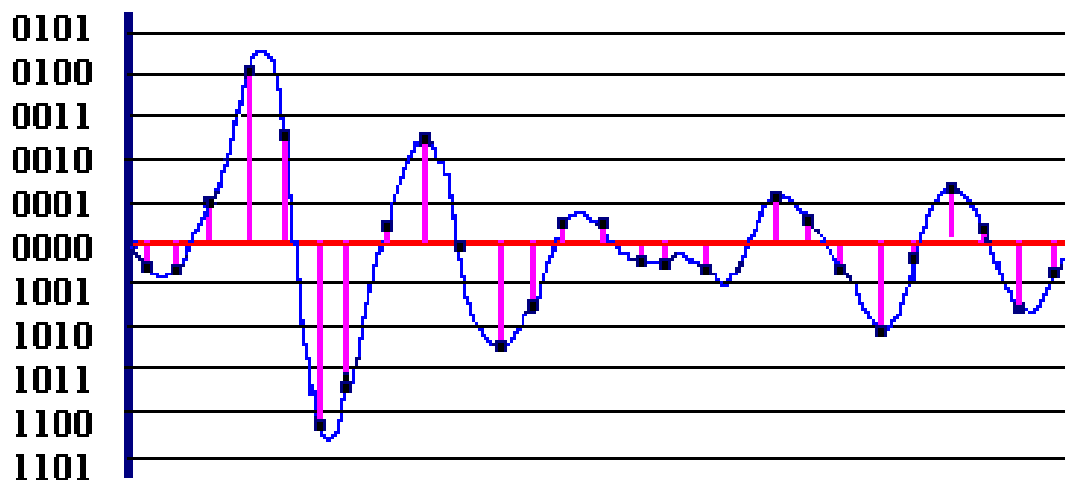
- 44.1kHz (CD质量)
    - 22.05kHz (盒式磁带质量)
    - 11.025kHz (普通声音)



## 2.2.2 音频的数字化和符号化

- 量化

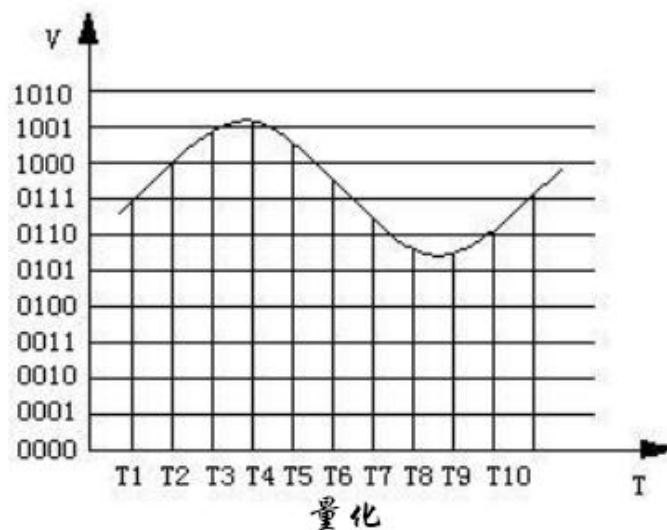
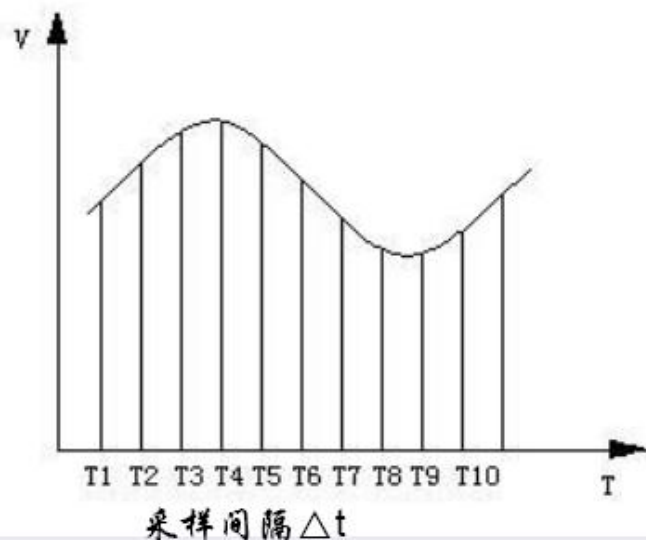
- 采样后的离散信号  $X(nT)$ ，其幅值仍然是连续变化的数值，为了便于在计算机处理，将取样值量化成一个有限个幅度值的集合。



## 2.2.2 音频的数字化和符号化

### • 量化

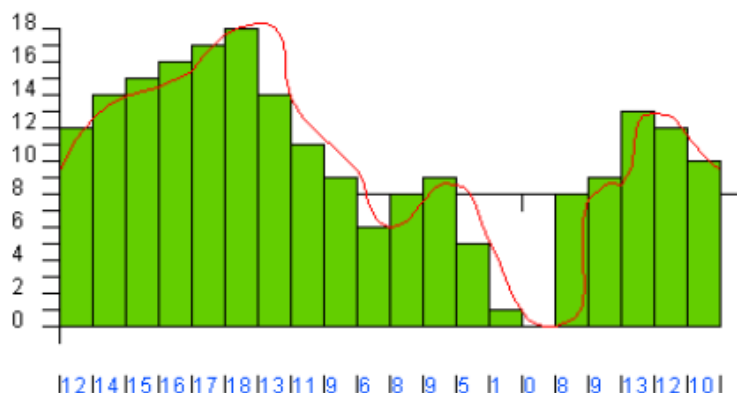
- 将整个幅度划分成为有限个小幅度（量化间隔）的集合，把落入某个间隔内的样值归为一类，并赋予相同的量化值。
- 量化间隔的数目，称为量化级。



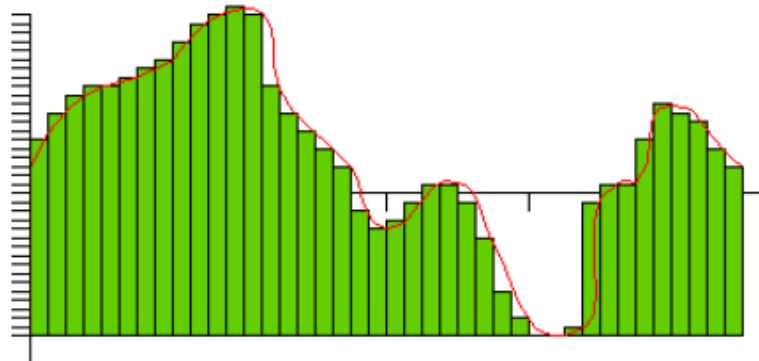
## 2.2.2 音频的数字化和符号化

### • 量化

- 量化过程存在**量化误差**，在还原信号的D/A转换后，误差作为**噪声再生**，称为**量化噪声**。
- 量化级数越多，量化误差越小，质量越好。



左图为采样率**2000Hz**，量化等级为**20**的采样量化过程



右图为采样率**4000Hz**，量化等级为**40**的采样量化过程

当采样率和量化等级提高一倍，当用**D/A**转换器重构原来信号时（图中的轮廓线），信号的失真明显减少，信号质量得到了提高

## 2.2.2 音频的数字化和符号化

- 例子

- 设一个连续信号的波形可以表示为：

$$f(t) = 8\sin(10t + \pi / 2) + 2\sin(5t + \pi / 4)$$

- 设采样频率为21Hz， $[-10, 10]$ 内的量化间隔取为1，试计算出该信号0到1秒内的量化数据。
    - 在0-1秒内，取21个采样点。在0、1/20、2/20、.....19/20、1秒时刻采样，将采样得到的数值取整，即得到量化数据。

## 2.2.2 音频的数字化和符号化

- 例子

序号	采样点	采样数据	量化值	序号	采样点	采样数据	量化值
0	0.00	9.41421	9	11	0.55	-9.37474	-9
1	0.05	-7.66754	-8	12	0.60	4.73547	5
2	0.10	1.30867	1	13	0.65	-2.56445	-3
3	0.15	1.07930	1	14	0.70	0.40759	0
4	0.20	-3.46729	-3	15	0.75	6.34915	6
5	0.25	8.93433	9	16	0.80	-9.36672	-9
.....	.....	.....	.....	.....	.....	.....	.....
9	0.45	-1.63831	-2	20	1.00	4.40090	4
10	0.50	7.89216	8				

## 2.2.2 音频的数字化和符号化

---

- 例子

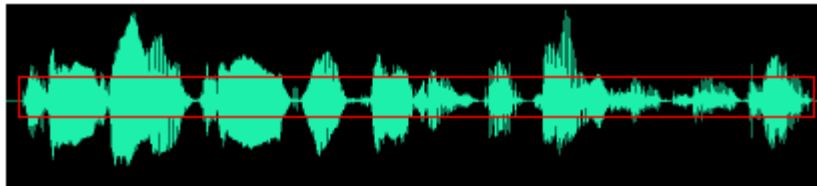
- 对于CD-DA，采样频率为**44.1kHz**，即每秒取**44,100**个点。幅度的取值范围是限制在 **$2^{16}=65,536$** 以内，量化间隔为**1**，即量化幅度可以取**65,536**个不同的值，计算机中用**16**位的存储空间就可以表示一个量化后的数值

动态范围为  $20 \times \lg(2^{16}) \approx 96\text{dB}$

## 2.2.2 音频的数字化和符号化

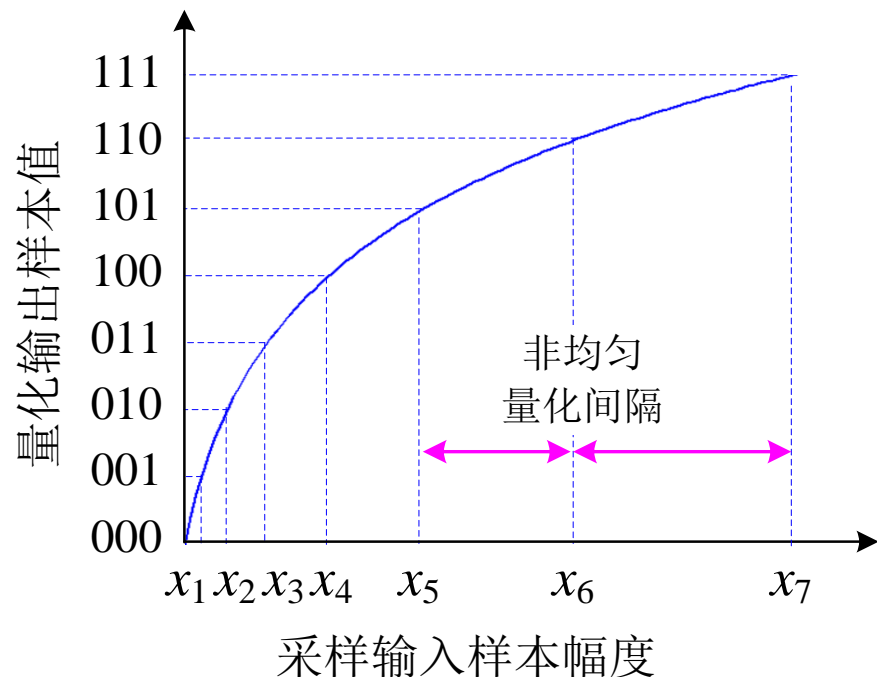
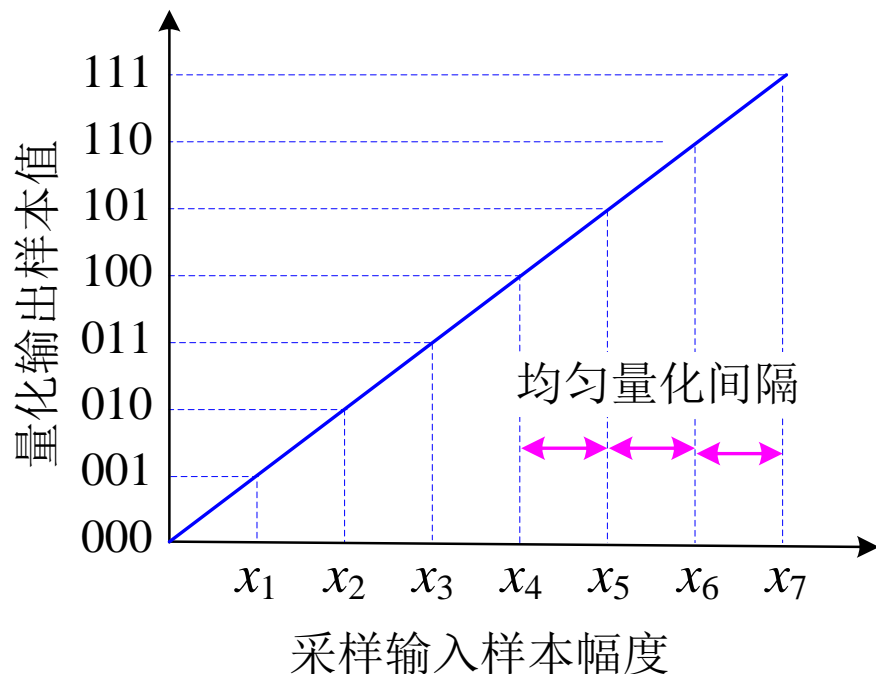
- 量化

- 如果采用相等的量化间隔对采样得到的信号作量化，称为均匀量化，也叫线性量化。
- 该方法将输入的声音信号的振幅范围分为 $2^B$ 个等份（ $B$ 为量化位数）。
- 为了满足听觉上的效果，均匀量化必须使用较多的量化位数，但会引起存储数据空间的增大。



## 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

- 非均匀量化：根据信号的不同区间来确定量化间隔。

- 对于信号取值小的区间，其量化间隔也小；
    - 反之，量化间隔就大。

- 优点

- 改善小信号时的量化信噪比。

## 2.2.2 音频的数字化和符号化

---

- 量化

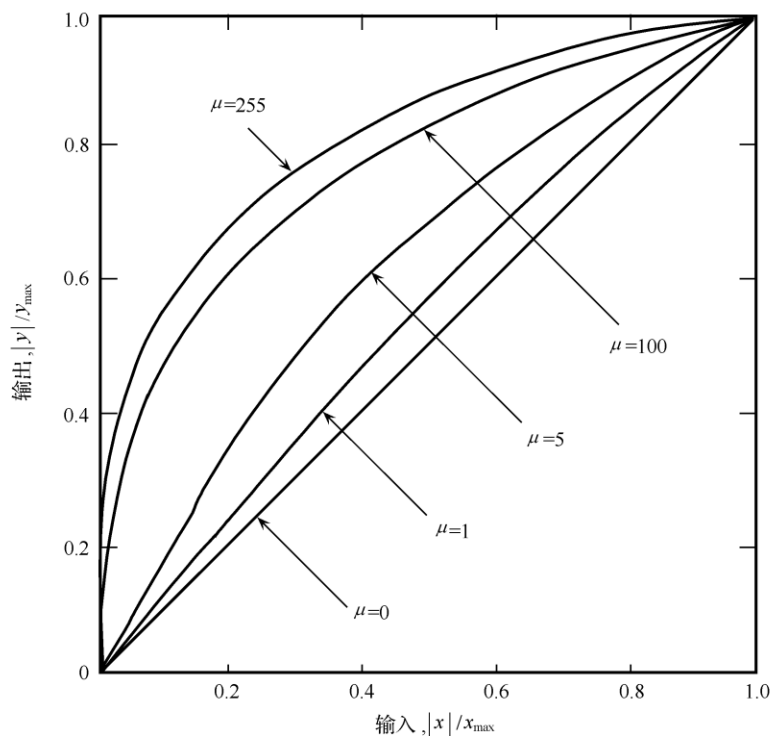
- 采样输入信号幅度和量化输出数据之间定义了两种对应关系：

- $\mu$ 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

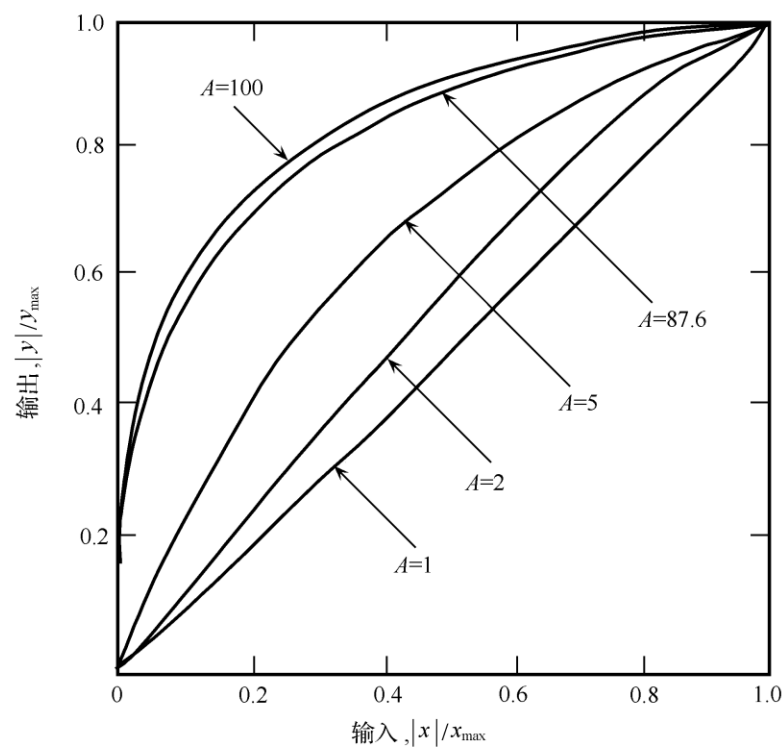
## 2.2.2 音频的数字化和符号化

- 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

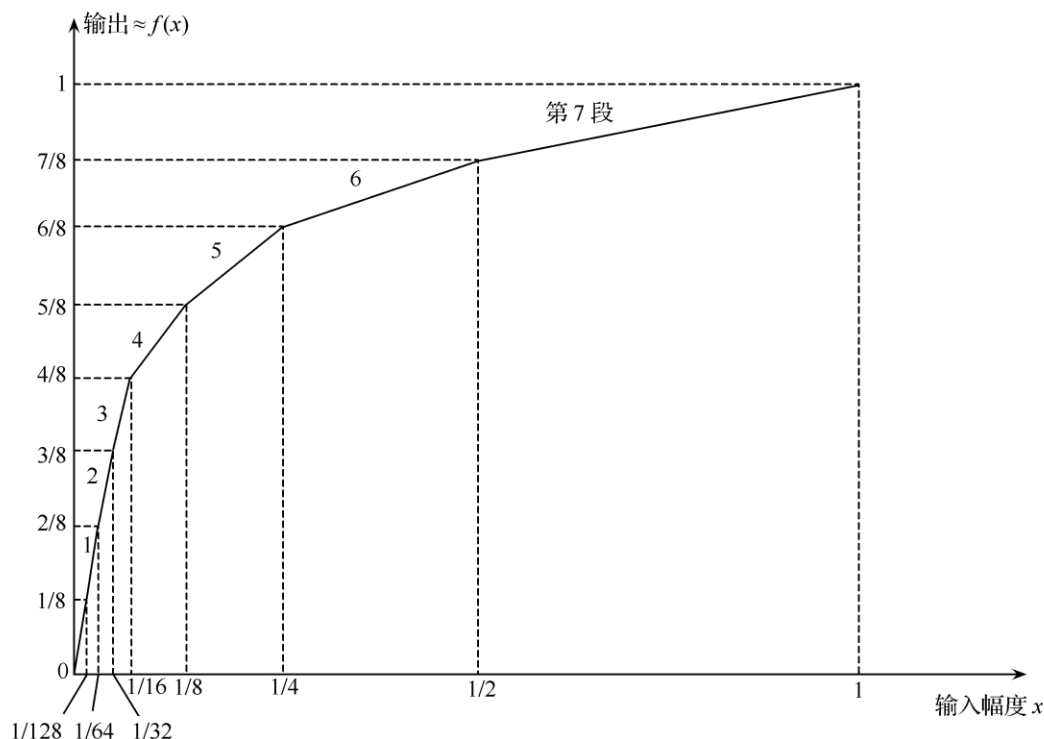
## 2.2.2 音频的数字化和符号化

### • 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:  $1/8 \div 1/32 = 4$

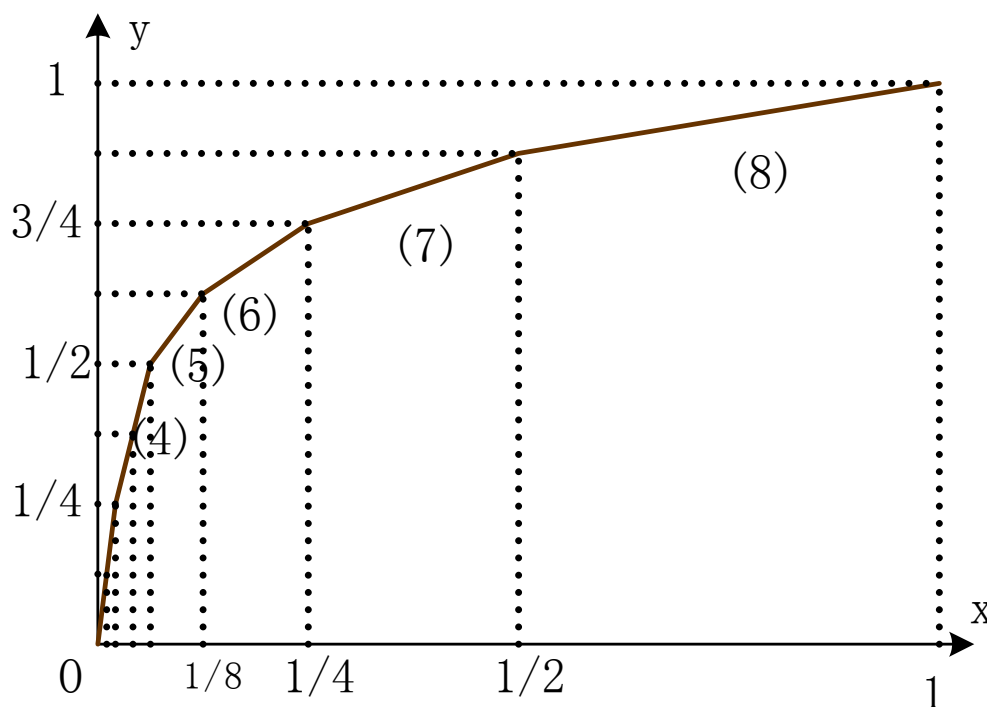
线段3斜率:  $1/8 \div 1/64 = 8$

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

