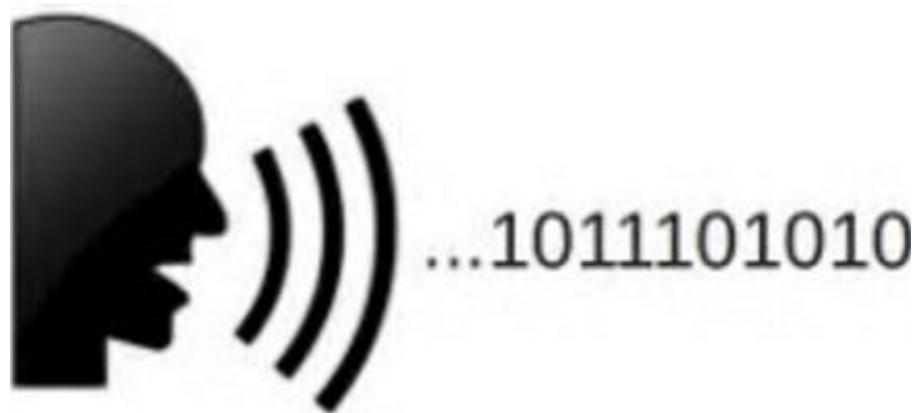


第3章 数字声音编码



目录

3.1 简介

3.1.1 声音简介

3.1.2 声音信号数字化

3.1.3 声音质量的MOS评分
标准

3.2.3 差分脉冲编码调制
(DPCM)与自适应差分脉冲
编码调制(ADPCM)

3.2.4 话音编码器
-子带编码(SBC)

3.2 声音的编码

3.2.1 脉冲编码调制 (PCM)
和自适应脉冲编码调制
(APCM)

-线性预测编码(LPC)

3.2.2 增量调制 (DM)与自适
应增量调制(ADM)

目录

3.1 简介

3.1.1 声音简介

3.1.2 声音信号数字化

3.1.3 声音质量的MOS评分
标准

3.2.3 差分脉冲编码调制
(DPCM)与自适应差分脉冲
编码调制(ADPCM)

3.2.4 话音编码器
-子带编码(SBC)

3.2 声音的编码

3.2.1 脉冲编码调制 (PCM)
和自适应脉冲编码调制
(APCM)

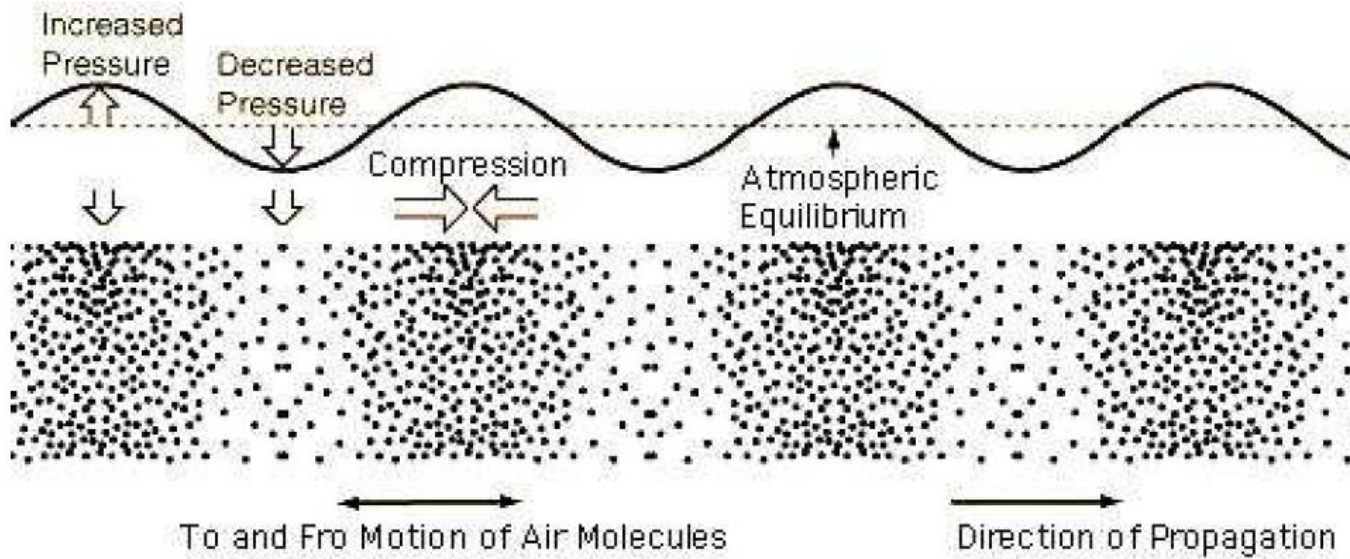
-线性预测编码(LPC)

3.2.2 增量调制 (DM)与自适
应增量调制(ADM)

3.1.1 声音简介

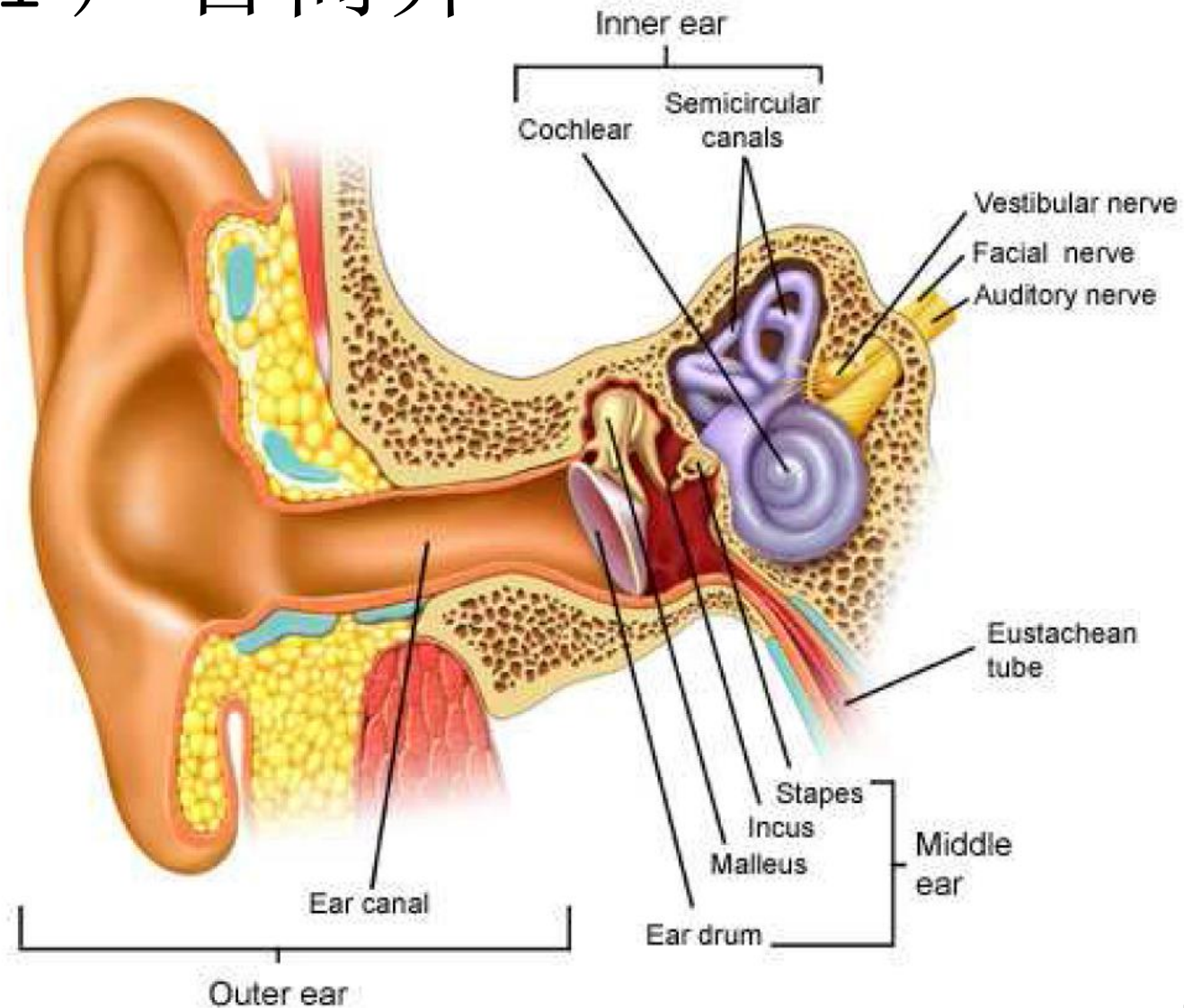
■ 什么是声音？

- 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



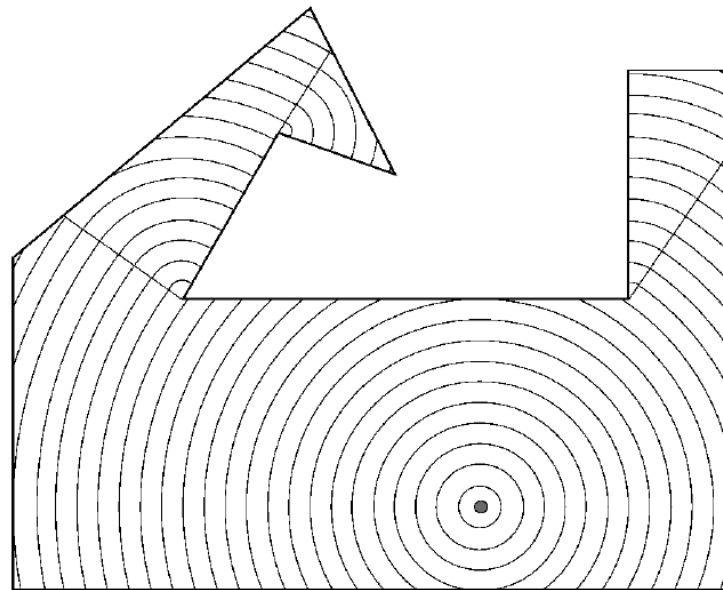
- 在20度的空气中，声波的传播速度是343.2m/s

3.1.1 声音简介



3.1.1 声音简介

- 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



3.1.1 声音简介

- 声音的强弱体现在声波压力的大小上，可用 decibels(dB)来衡量。

$$N_{db} = 20 * \log_{10}(p_e/p_r)$$

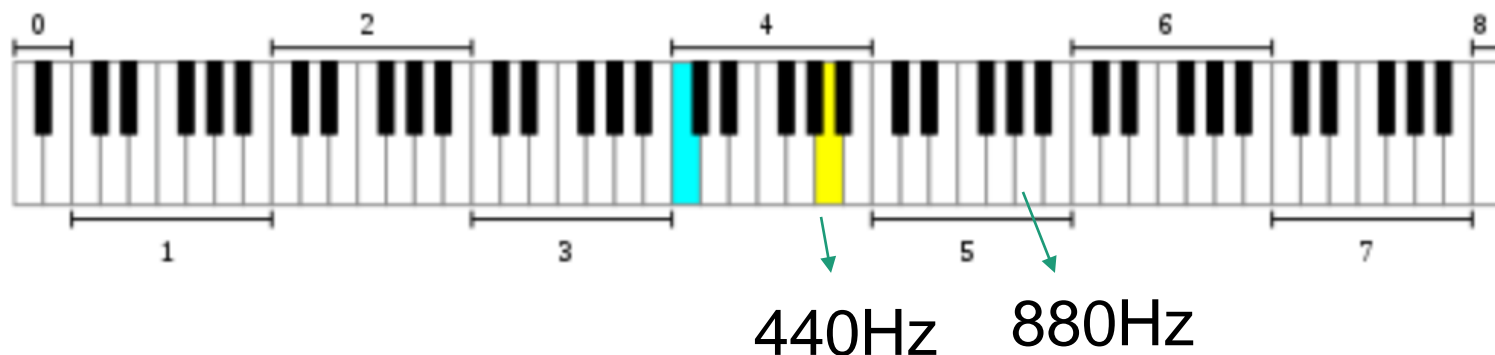
压力的峰值

$2 \times 10^{-5} \text{ pa}$

Threshold of hearing	0 dB
Rustle of leaves	10 dB
Very quiet room	20 dB
Conversation	60 dB
Busy street	70 dB
Loud radio	80 dB
Train through station	90 dB
Threshold of discomfort	120 dB
Threshold of pain	140 dB
Damage to ear drum	160 dB

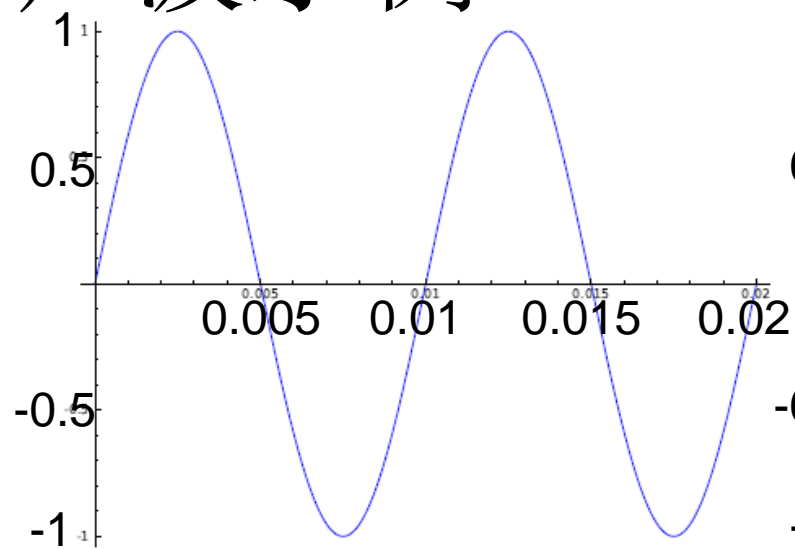
3.1.1 声音简介

- 音调的高低体现在声音的频率上。

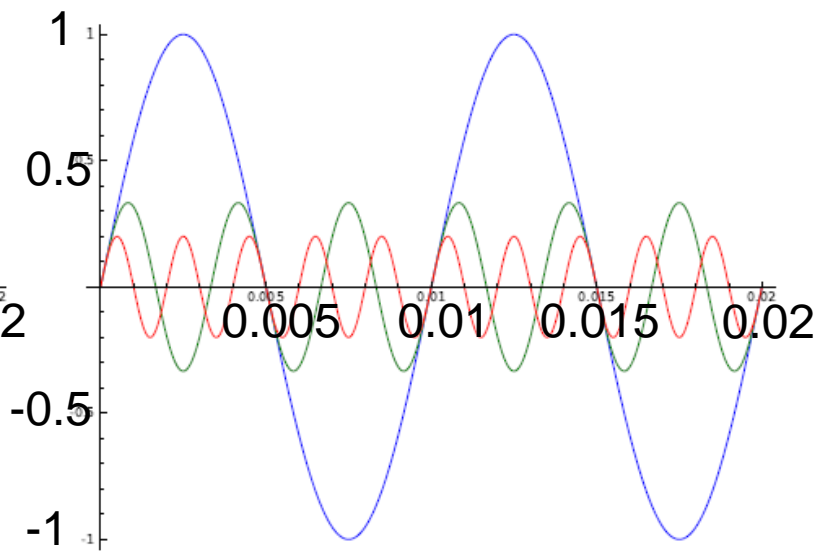


- 一个纯音可以用正弦波来表示，只有一个对应的频率；更复杂的声音可以由一系列纯音组成。

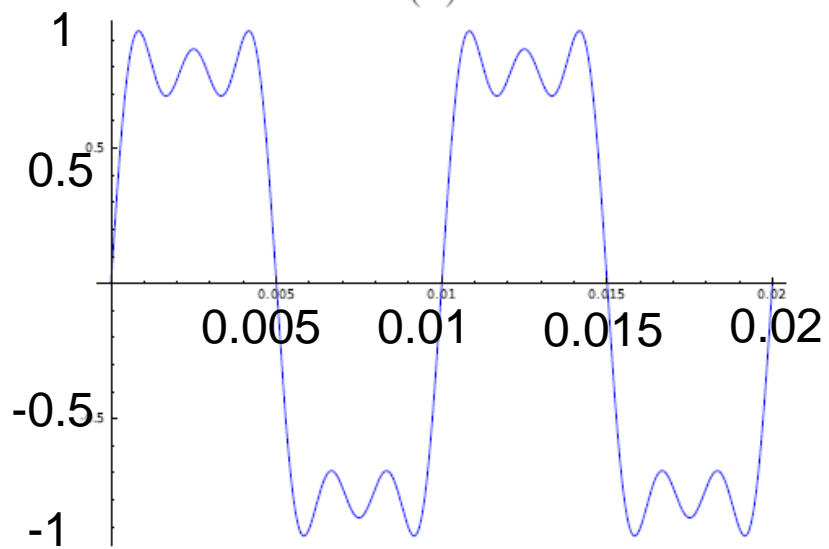
声波示例



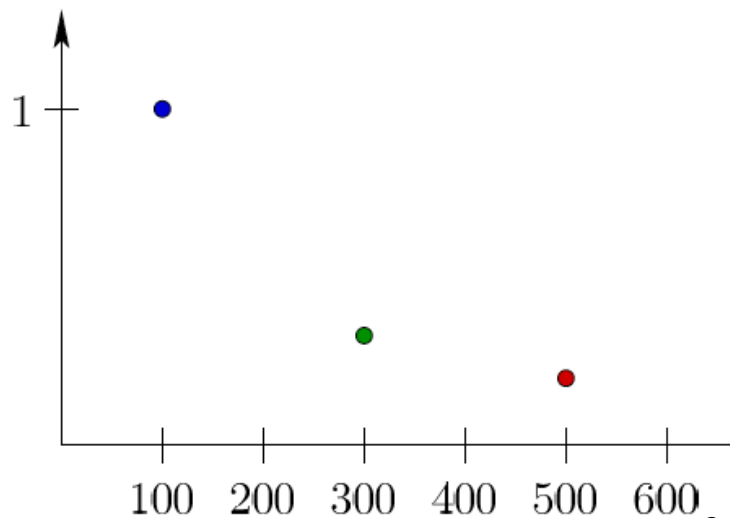
(a)



(b)

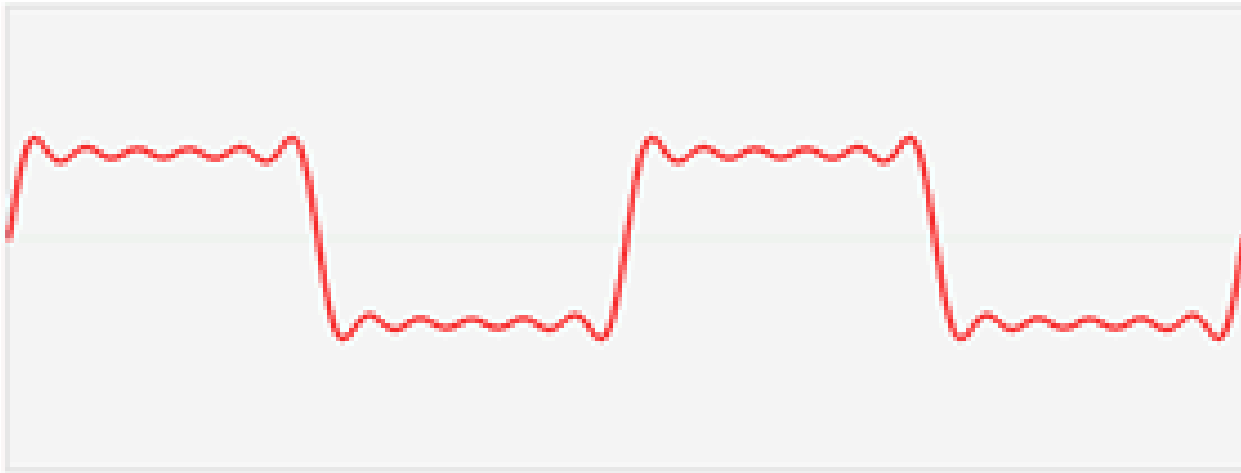


(c)

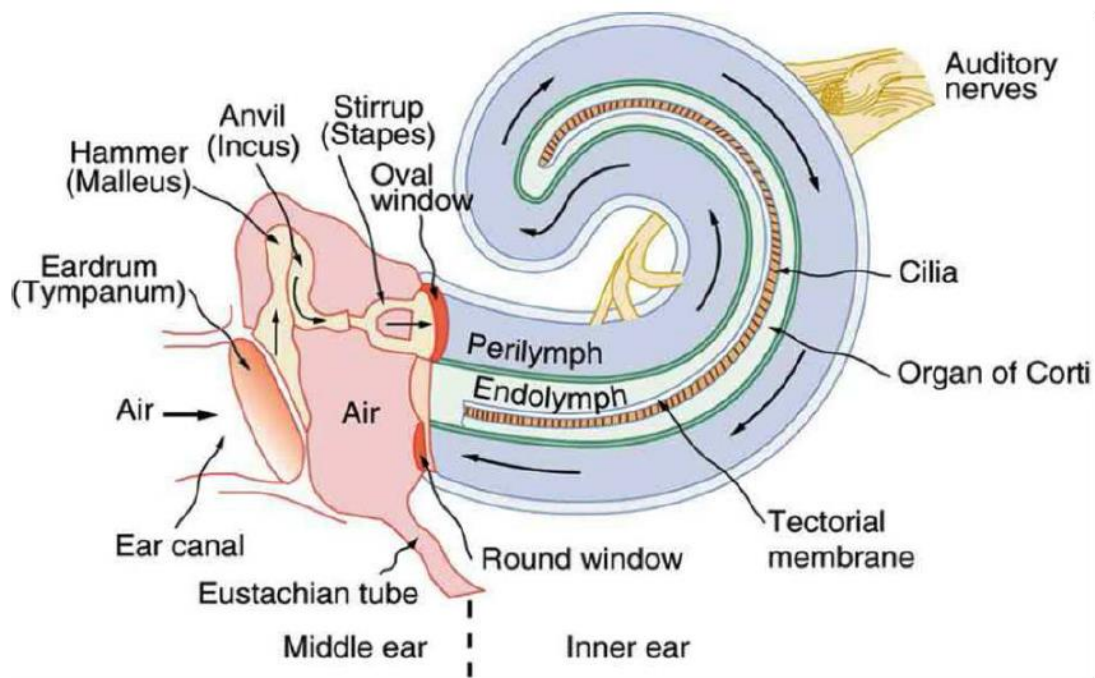


(d)

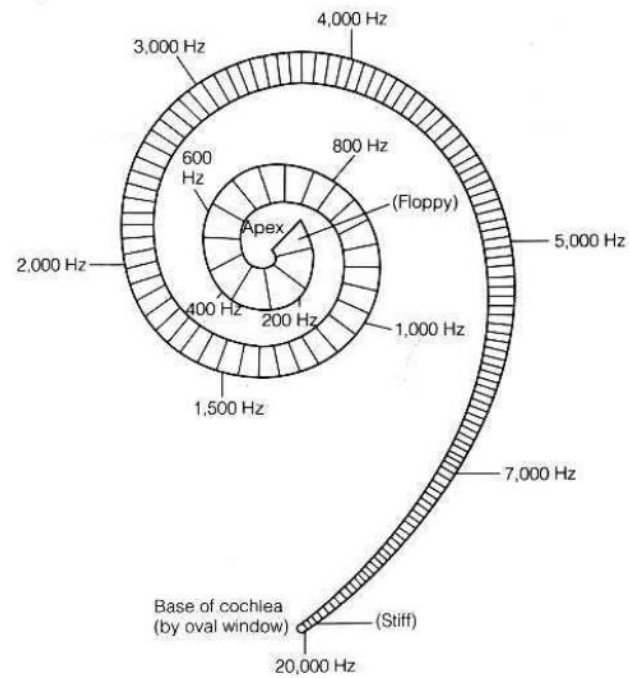
声波示例



3.1.1 声音简介



(a)



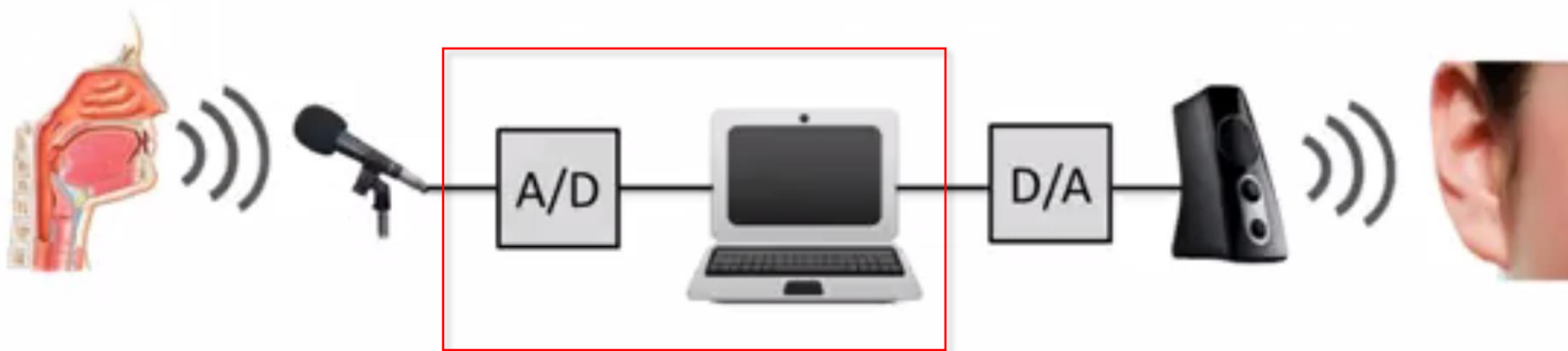
(b)

3.1.1 声音简介

■ 声音的频率

- 声音(audio): 20~ 20 000Hz
- 高保真声音(high-fidelity audio): 10 ~ 20 000 Hz
- 话音(speech): 300~3000/3400 Hz
- 亚音/次音(subsonic): < 20 Hz
- 超声(ultrasonic): > 20 000 Hz

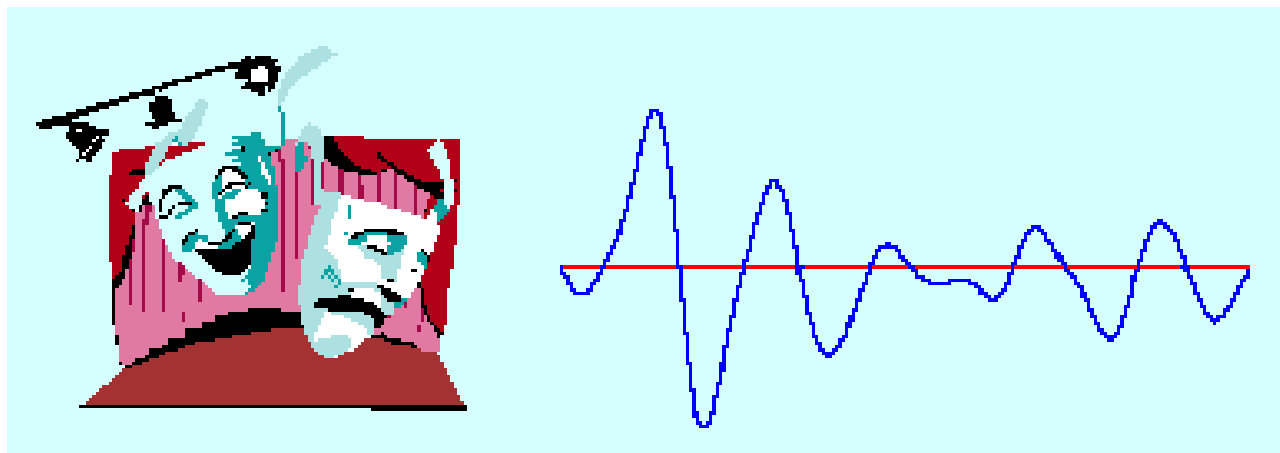
3.1.1 声音简介



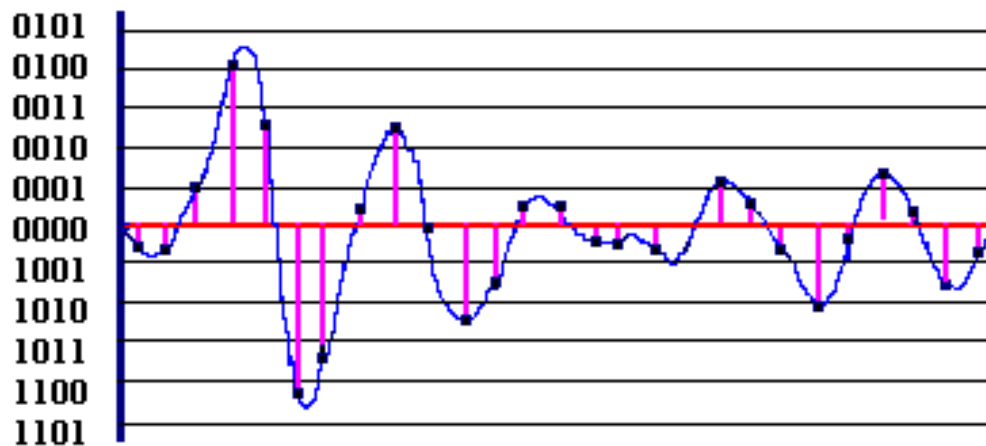
声音的产生、传播、处理、感知示例

3.1.1 声音简介

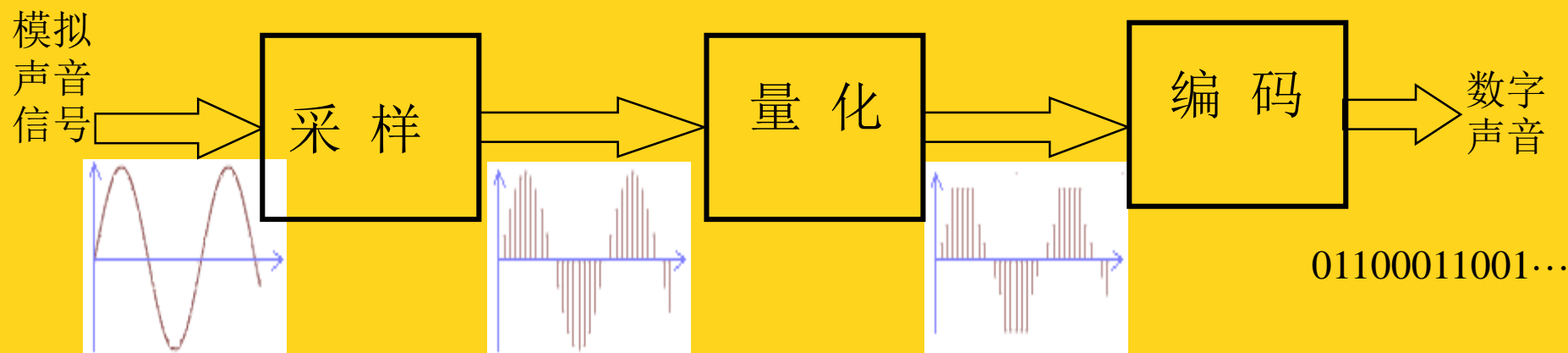
- 声音用电压信号表示时，声音信号在时间和幅度上都是连续的模拟信号



3.1.2 声音信号数字化



声音的采样和量化



3.1.2 声音信号数字化

- 采样(sampling):连续时间的离散化
 - 均匀采样: 每隔相等的一段时间采样一次
- 量化(quantization):连续幅度的离散化
 - 线性(均匀)量化: 幅值划分等间隔
 - 非线性(非均匀)量化: 幅值划分间隔不等

3.1.2 声音信号数字化

■ 采样频率

- 奈奎斯特采样定理:采样频率不低于声音信号最高频率 f_{max} 的两倍, 就能把以数字表达的声音还原成原来的声音

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

3.1.2 声音信号数字化

- 采样精度或量化阶(quantization step size)
 - 可用每个声音样本的位数(即bps)表示，大小影响到声音的质量
 - 度量声音波形幅度的精确程度，度量模拟信号的最小单位

3.1.2 声音信号数字化

- 采样精度也可用信号噪声比表示，简称为信噪比 (signal-to-noise ratio, SNR),

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

信号电压

单位为分贝 (dB)

$$= 20 \log_{10} \left(\frac{V_{signal}}{V_{signal} (1/2^n)} \right) \approx 6.02n$$

量化噪声电压
(模拟信号的采样值
和与它最接近的数字
数值间的差值)

3.1.2 声音信号数字化

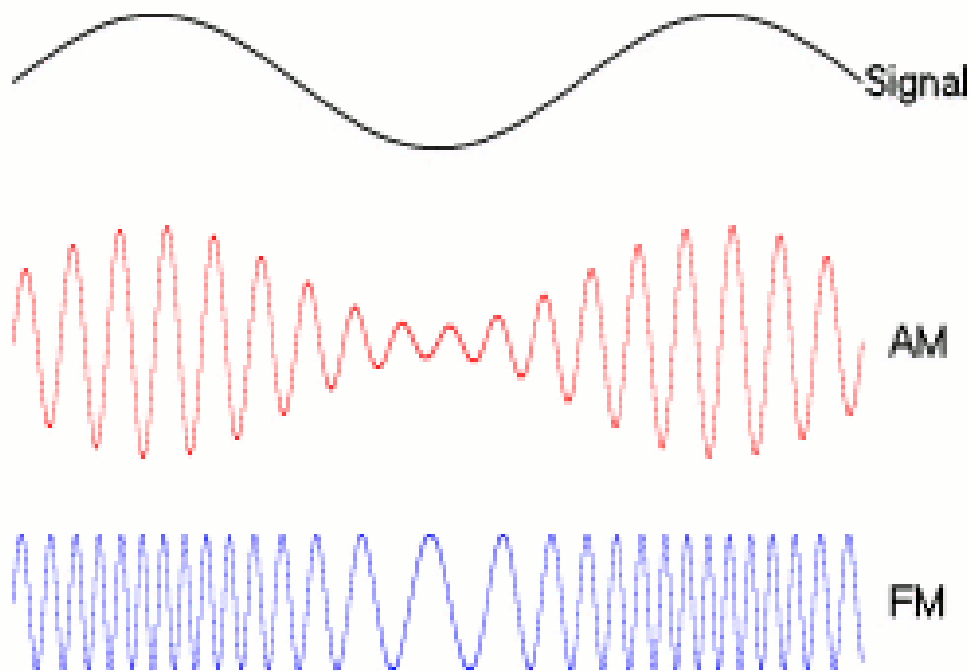
- 声音质量和数据率——质量度量

质量	采样频率 (kHz)	样本精度 (bit/s)	单道声/立 体声	(未压缩的)数 据率(kb/s)	频率范围 (Hz)
电话*	8	8	单道声		200~3400
AM	11.025	8	单道声		100~6600
FM	22.050	16	立体声		20~11000
CD	44.1	16	立体声		5~20000
DAT	48	16	立体声		20~20000

*AM: amplitude modulation
FM: frequency modulation

CD: compact disc audio
DAT: digital audio tape

3.1.2 声音信号数字化



中央人民广播电台节目频率分布表

2009年1月1日00:00起执行

语言	播音时间	频率{千赫}	服务区	备注
中一	00:00-24:00	106.1	北京	调频
中一	03:55-01:35	639	华北	中波
中一	03:55-01:35	6030	华北	

3.1.3 声音质量的衡量方法

■ 声音质量的衡量方法

- 客观质量度量
 - 声音带宽、信噪比（SNR）、采样精度
- 主观质量度量
 - 主观平均分（MOS）

3.1.3 声音质量的MOS评分标准

- 主观平均分(mean opinion score, MOS)
 - 对声音主观质量度量比较通用的标准是5分制

声音质量MOS评分标准

分数	质量等级	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)

目录

3.1 简介

3.1.1 声音简介

3.1.2 声音信号数字化

3.1.3 声音质量的MOS评分
标准

3.2.3 差分脉冲编码调制
(DPCM)与自适应差分脉冲
编码调制(ADPCM)

3.2.4 话音编码器

-子带编码(SBC)

-线性预测编码(LPC)

3.2 声音的编码

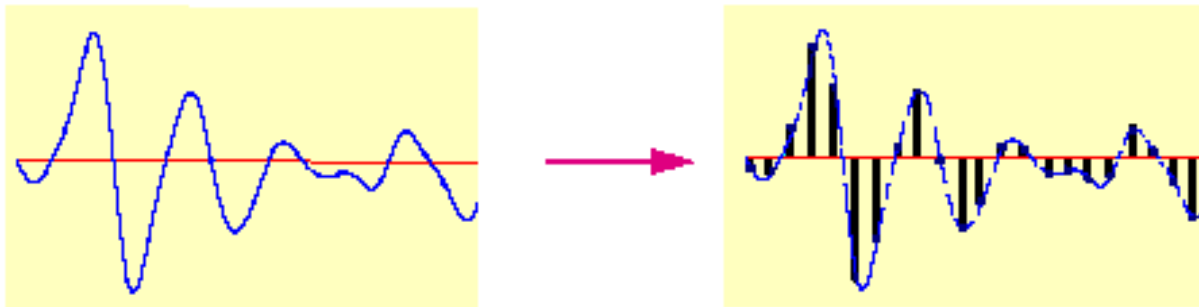
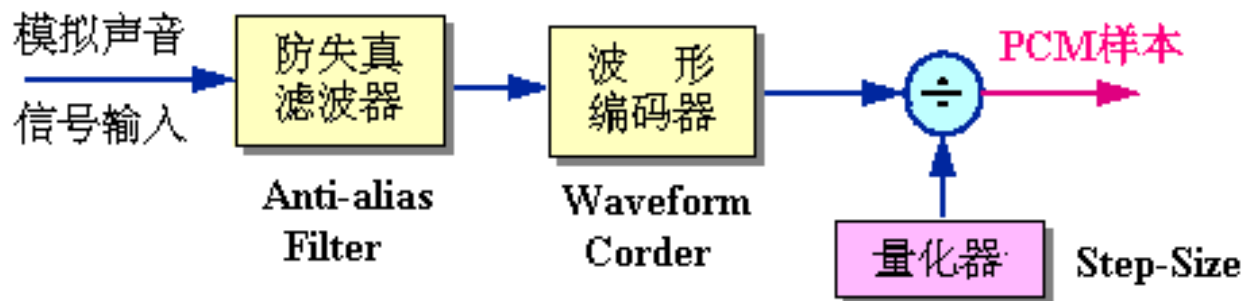
3.2.1 脉冲编码调制 (PCM)
和自适应脉冲编码调制
(APCM)

3.2.2 增量调制 (DM)与自适
应增量调制(ADM)

3.2.1 脉冲编码调制(PCM)

■ 脉冲编码调制(Pulse Code Modulation, PCM)

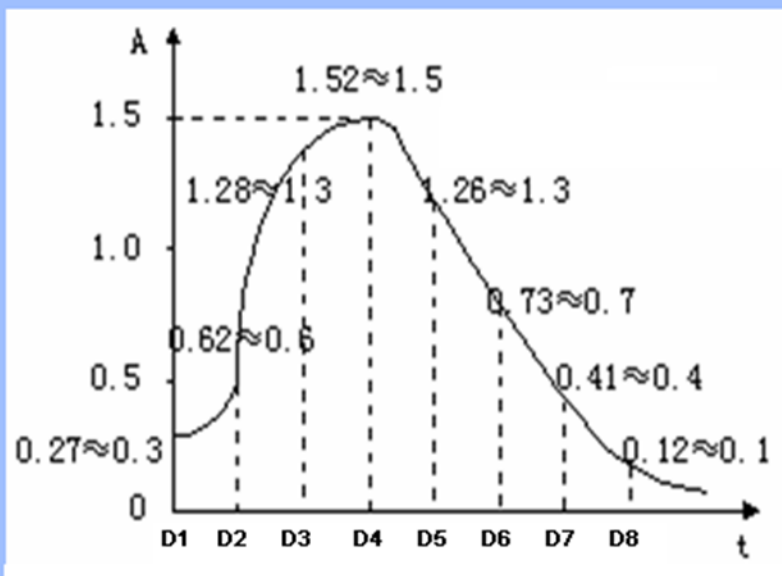
概念上最简单、理论上最完善、最早研制成功、使用最为广泛、数据量最大的编码系统



3.2.1 脉冲编码调制

- PCM实际上是模拟信号数字化。
 - 采样：每隔一段时间读声音的幅度
 - 量化：把采样得到的声音幅度转换成数字值
- 量化可分为均匀量化和非均匀量化两类。

3.2.1 脉冲编码调制



[0,1.5]分成15个区间，间隔长度0.1，形成16个量化级。

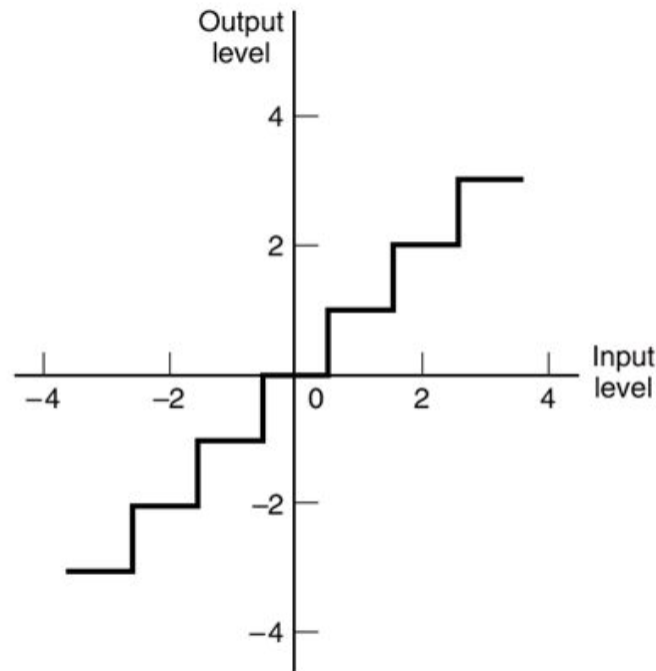
样本	量化级	二进制编码	编码信号
D ₁	3	0011	
D ₂	6	0110	
D ₃	13	1101	
D ₄	15	1111	
D ₅	13	1101	
D ₆	7	0111	
D ₇	4	0100	
D ₈	1	0001	

均匀量化？非均匀量化？

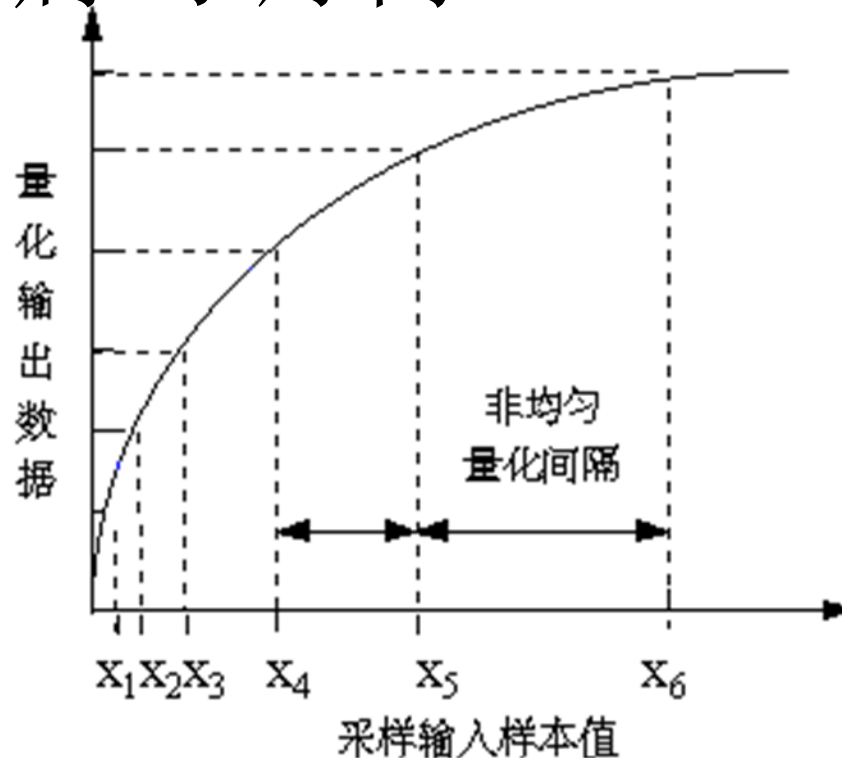
3.2.1 脉冲编码调制

■ 均匀量化

- 采用相等的量化间隔对采样得到的信号作量化。
- 量化误差(量化噪声): 量化后的样本值 Y 和原始值 X 的差 $e=y-x$ 。



3.2.1 脉冲编码调制



■ 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

3.2.1 脉冲编码调制

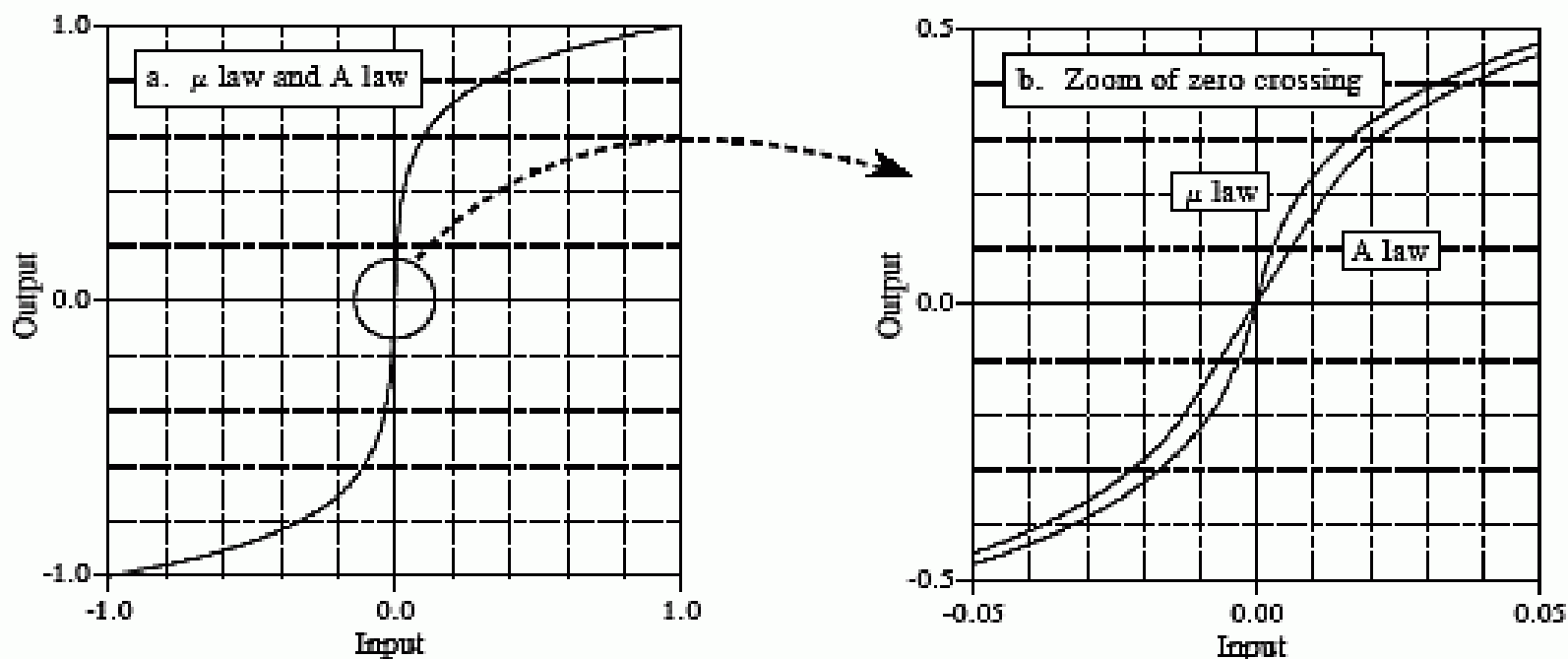


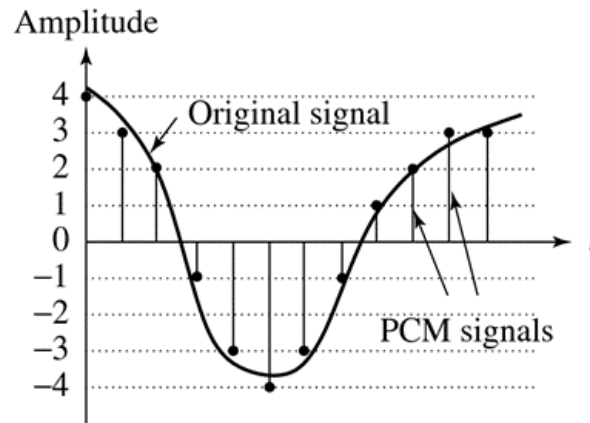
FIGURE 22-7

Companding curves. The μ 255 law and "A" law companding curves are nearly identical, differing only near the origin. Companding increases the amplitude when the signal is small, and decreases it when it is large.

- 两种非均匀量化方法： μ 律压扩算法和A律压扩算法

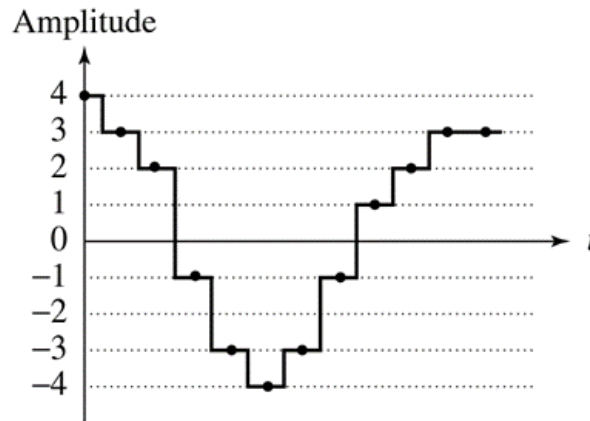
3.2.1 脉冲编码调制

Original analog signal and its corresponding PCM signals.

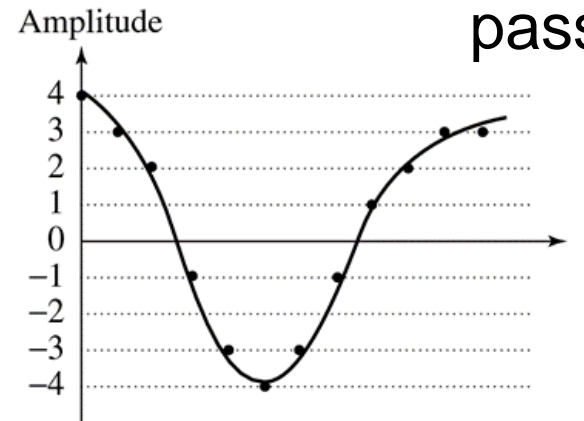


(a)

Reconstructed signal after low-pass filtering



(b)



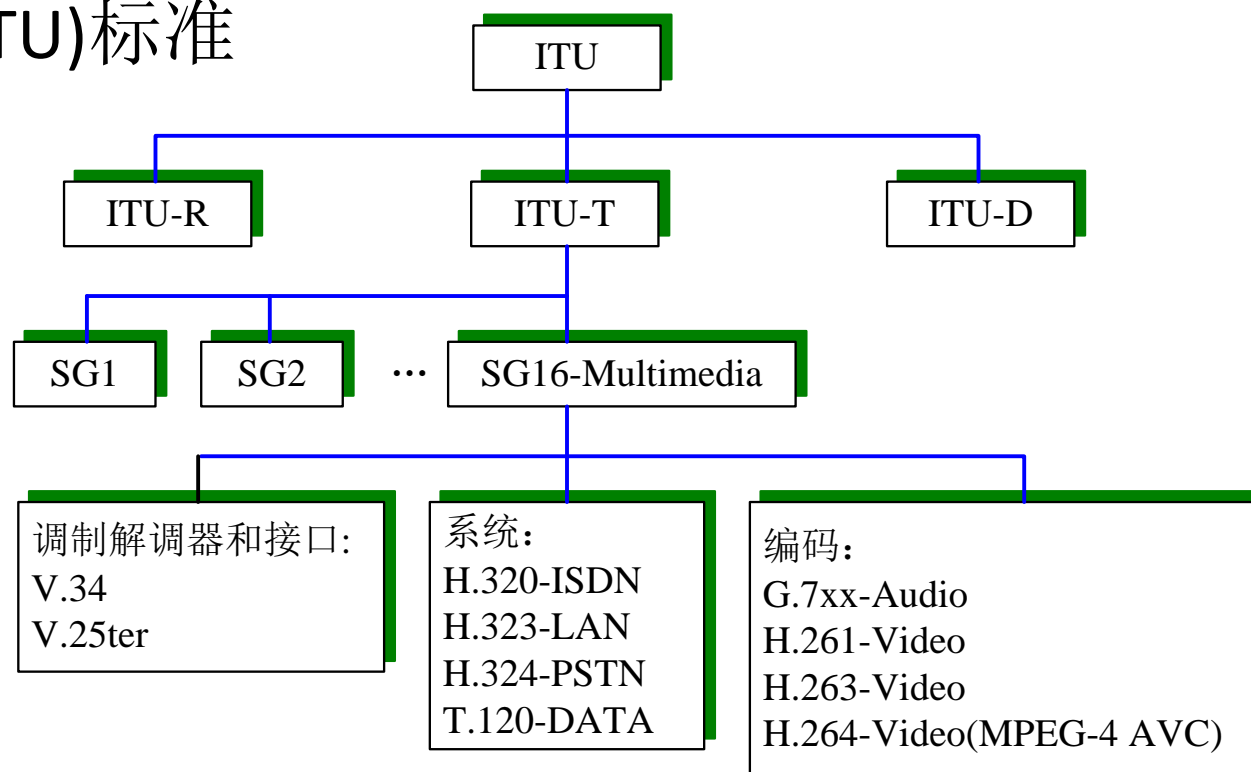
(c)

Decoded staircase signal.

ITU标准



■ 国际电信联盟(International Telecommunication Union, ITU)标准

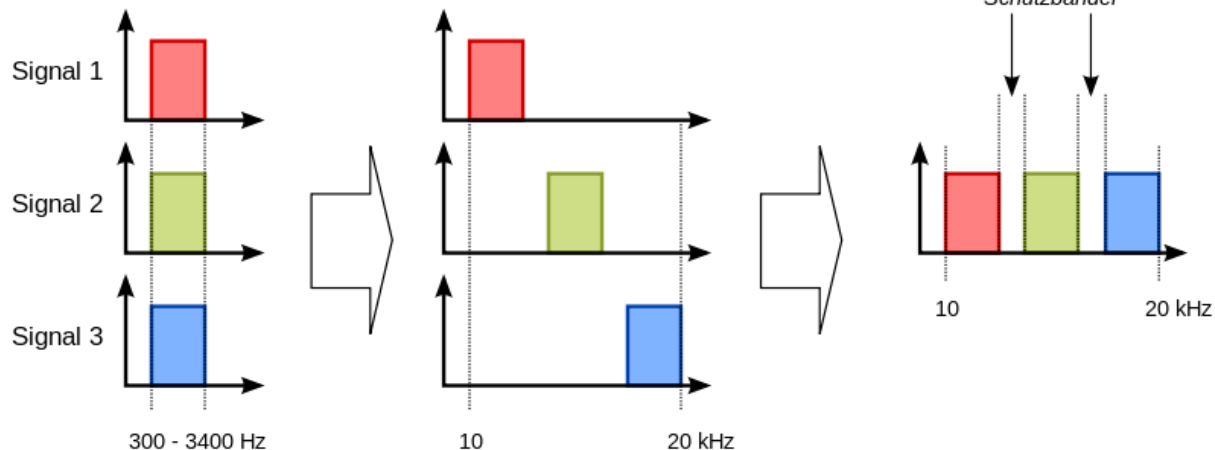


VCEG开发的标准

3.2.1 PCM在通信中的应用

■ 频分多路复用

- 在一条通信线路上使用不同频段同时传送多个独立信号的通信方法
- 模拟载波通信的主要方法



每一个信号的频率平移到一个特定的频率范围内

3.2.1 PCM在通信中的应用

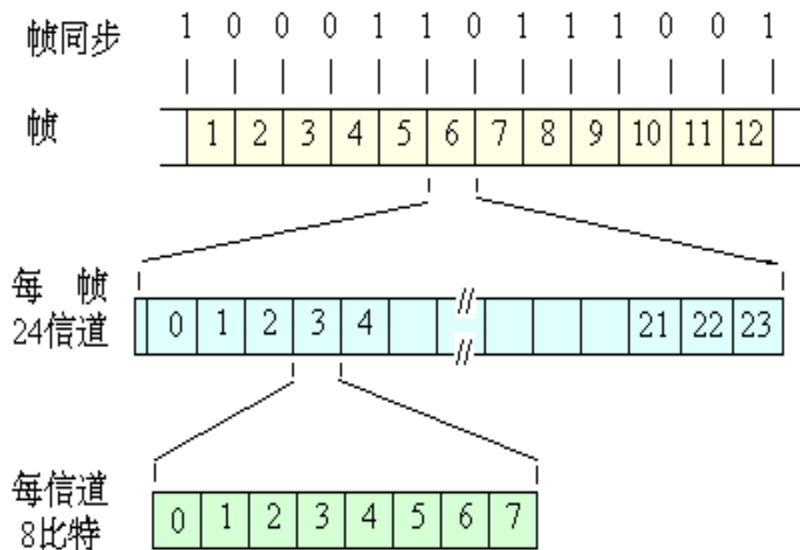
■ 时分多路复用

- 在一条通信线路上使用不同时段“同时”传送多个独立信号的通信方法
- 核心思想：将时间分成等间隔的时段，为每对用户指定一个时间间隔，每个间隔传输信号的一部分

3.2.1 PCM在通信中的应用

- 24路制:

- 每秒钟传送8000帧，每帧125 μ s
- 12帧组成1复帧(用于同步)。
- 每帧由24个时间片(信道)和1位同步位组成
- 每个信道每次传送8位代码，1帧有 $24 \times 8 + 1 = 193$ 位(位)
- 数据传输率R=
- 每一个话路的数据传输率=



24路PCM的帧结构

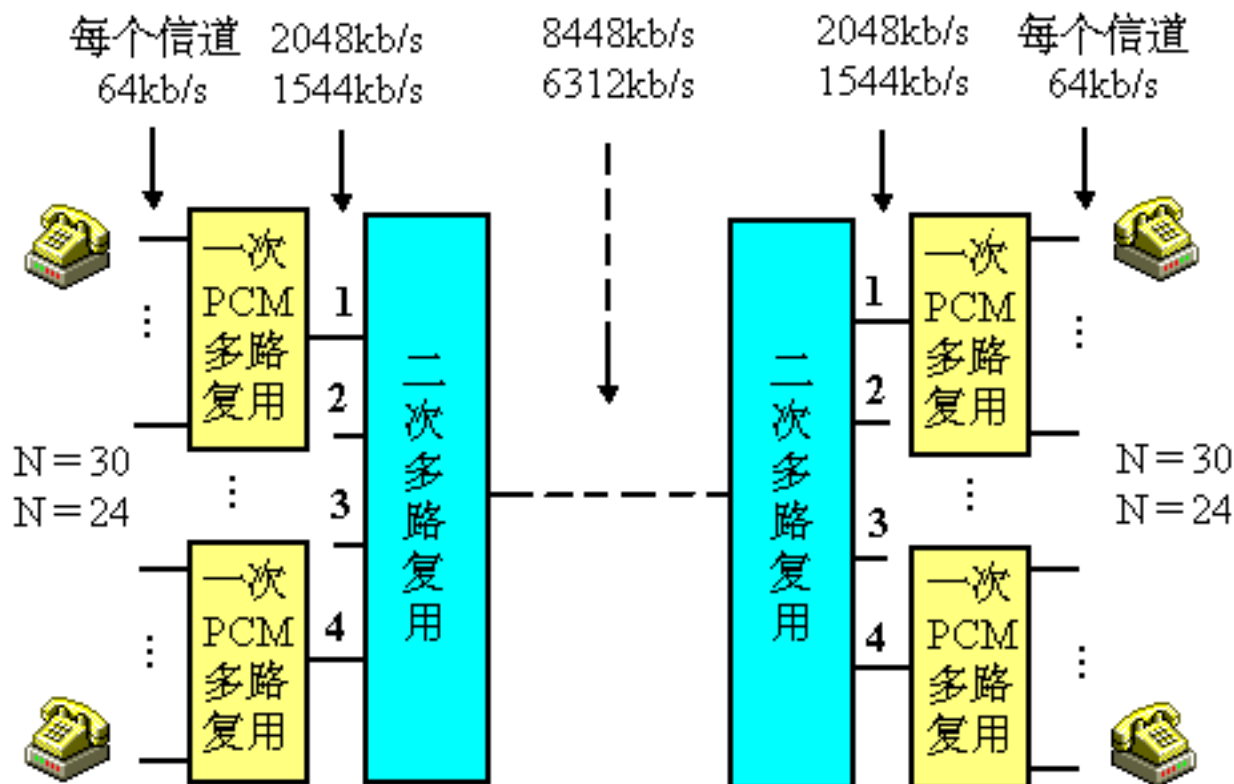
3.2.1 PCM在通信中的应用

■ 30路制:

- 每秒钟传送8000帧，每帧125 μs
- 16帧组成1复帧(用于同步)
- 每帧由32个时间片(信道)组成
- 每个信道每次传送8位代码
- 数据传输率: $R =$
- 每一个话路的数据传输率 =

3.2.1 PCM在通信中的应用

数字通信线路的数据传输率



二次复用示意图

3.2.1 PCM在通信中的应用

- T1/E1, T2/E2, T3/E3, T4/E4和T5/E5的数据传输率

表 3-3 多次复用的数据传输率(注)

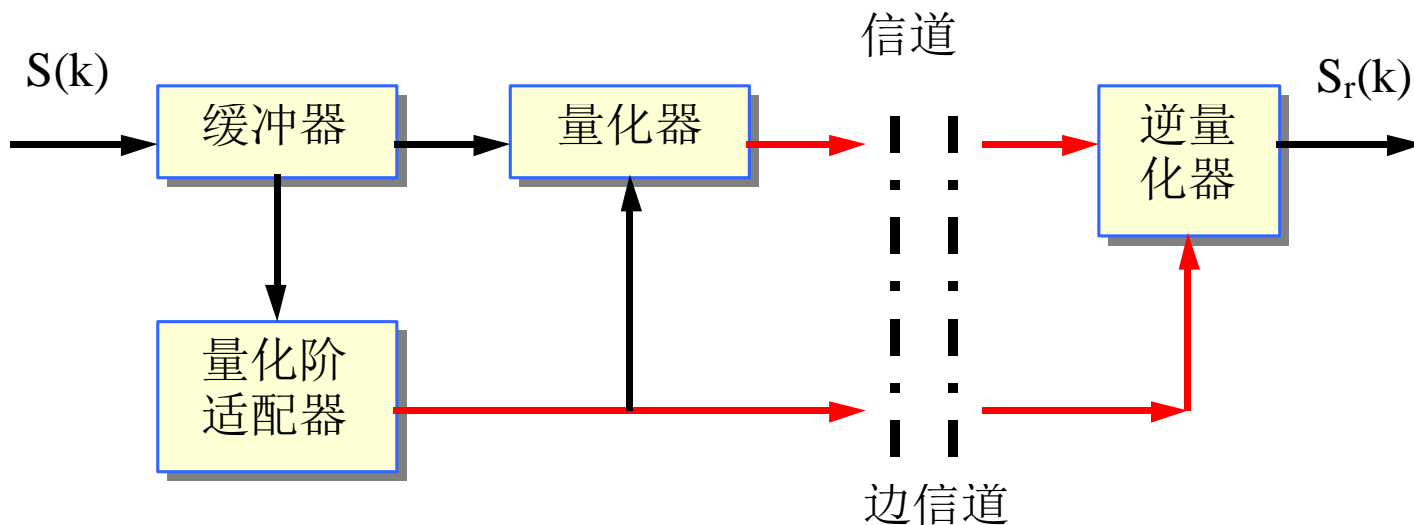
国别	数字网络等级	T1/E1	T2/E2	T3/E3	T4/E4	T5/E5
美国	64 kb/s话路数	24	96	672	4032	-
	总传输率(Mb/s)	1.544	6.312	44.736	274.176	-
	数字网络等级	1	2	3	4	5
欧洲	64 kb/s话路数	30	120	480	1920	7680
	总传输率(Mb/s)	2.048	8.448	34.368	139.264	560.000
日本	64 kb/s话路数	24	96	480	1440	-
	总传输率(Mb/s)	1.544	6.312	32.064	97.728	-

3.2.1 自适应脉冲编码调制

- 自适应脉冲编码调制(Adaptive Pulse Code Modulation, APCM)
 - 根据输入信号幅度大小来改变量化阶大小的一种波形编码技术
 - 自适应
 - 瞬时自适应：量化阶大小每隔几个样本就改变
 - 音节自适应：量化阶大小在较长时间里发生变化

3.2.1 自适应脉冲编码调制

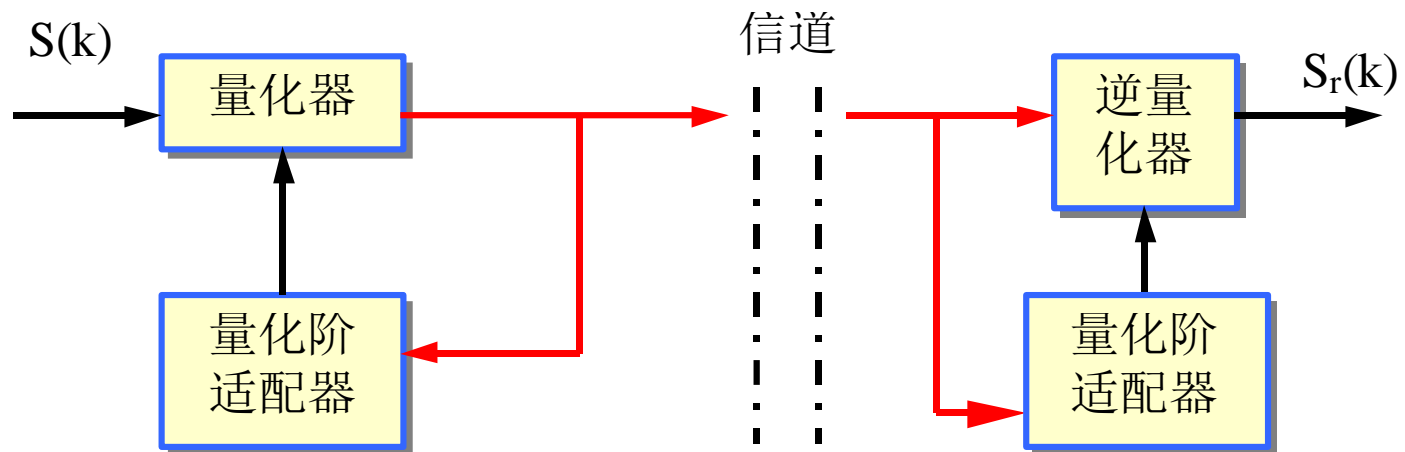
- 前向自适应(forward adaptation)
 - 根据未量化的样本值来确定量化阶的大小，并对其进行编码作为边信息(side information)传送到接收端



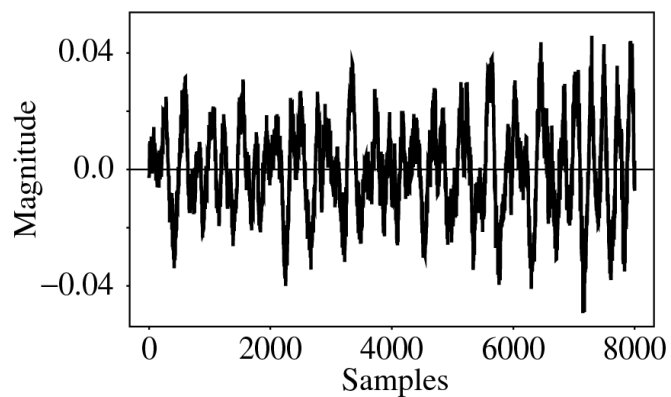
(a)前向自适应APCM

3.2.1 自适应脉冲编码调制

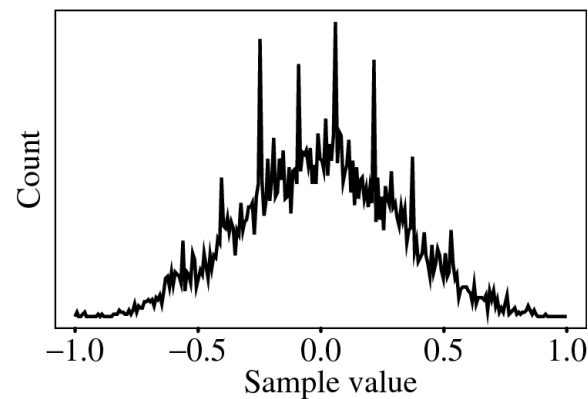
- 后向自适应(backward adaptation)
 - 从量化器刚输出的过去样本中提取量化阶信息。
 - 不需要传送边信息。



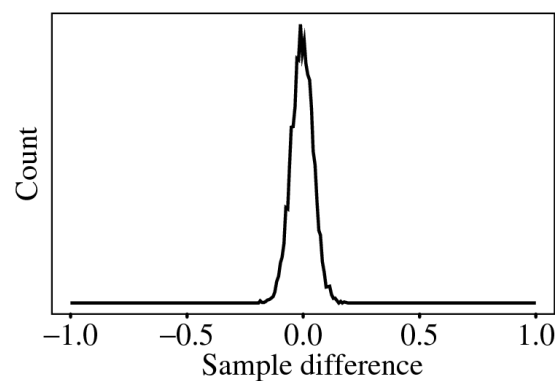
(b)后向自适应APCM



数字语音信号 (a)



(b) 数字语音信号
幅值直方图

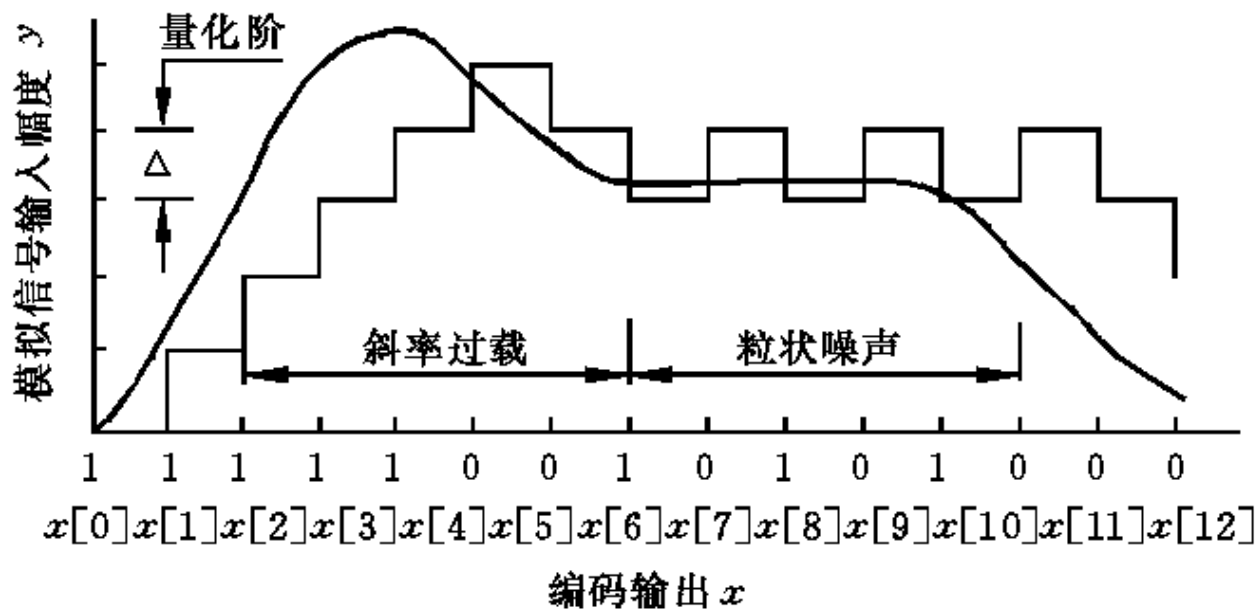


(c)

信号幅度差值直方图

3.2.2 增量调制与自适应增量调制

- 增量调制 (Δ 调制, Delta Modulation, DM)
 - 预测编码技术: 对实际信号与预测信号之差的极性进行编码。如果极性为“正”, 用“1”表示; 反之用“0”表示, 或相反
 - 只用1位对话音信号编码, 称为“1位系统”



3.2.2 增量调制与自适应增量调制

- 自适应增量调制(Adaptive Delta Modulation, ADM)
 - 根据输入信号斜率的变化自动调整量化阶 Δ 的大小，以减少斜率过载和粒状噪声。
 - 检测到斜率过载时增大量化阶 Δ ，斜率减小时降低量化阶 Δ
 - 宋：每当输出不变时量化阶增大50%；每当输出值改变时，量化阶减小50%
 - 格林弗基斯：如果连续可变斜率增量调制器的输出连续出现三个相同值时，量化阶加一个大的增量，反之，就加一个小的增量。

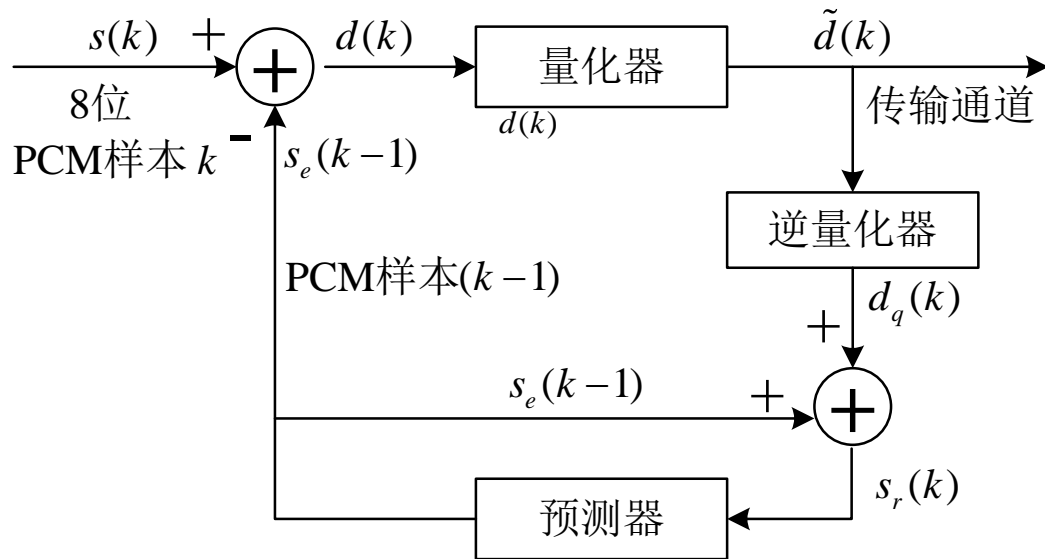
3.2.3 差分脉冲编码调制

- 差分脉冲编码调制(Differential Pulse Code Modulation, DPCM)
 - 利用样本间存在的信息冗余来进行数据压缩
 - 基本思想：根据过去的样本去预测下一个样本信号的幅度大小，对实际值与预测值之差进行量化编码，从而减少表示样本信号的比特数

3.2.3 差分脉冲编码调制

- DPCM原理

- 差分信号 $d(k)$: 离散输入信号 $s(k)$ 和预测器输出的估算值 $s_e(k-1)$ 之差
- 对 $d(k)$ 进行量化编码, 得到



问题

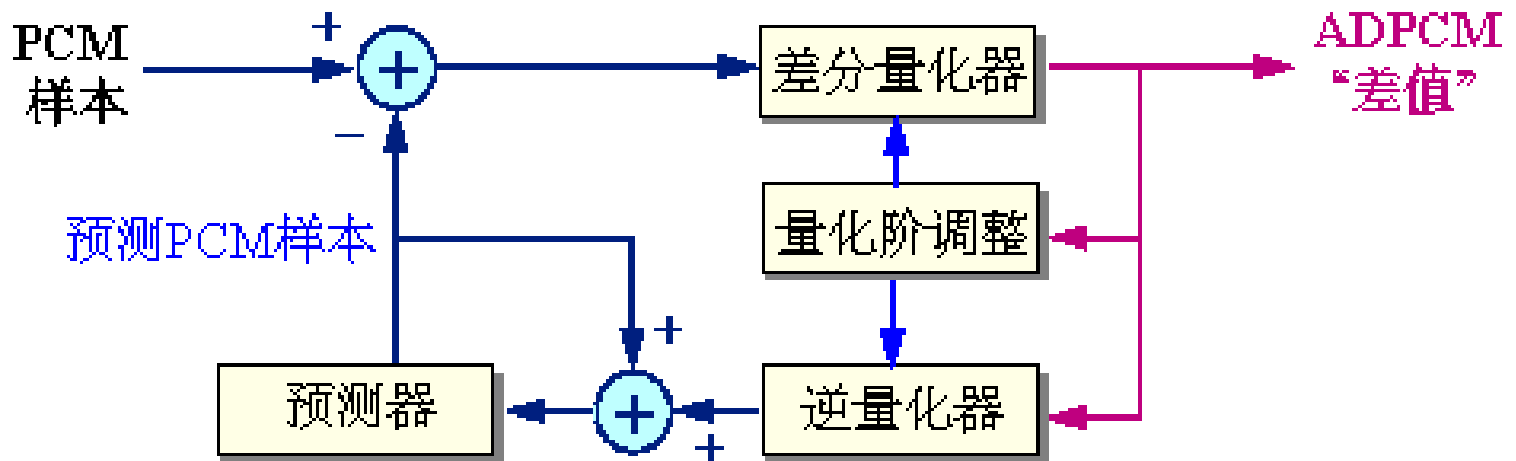
- PCM与DPCM的区别是什么？
- DM与DPCM的区别是什么？

3.2.3 自适应差分脉冲编码调制

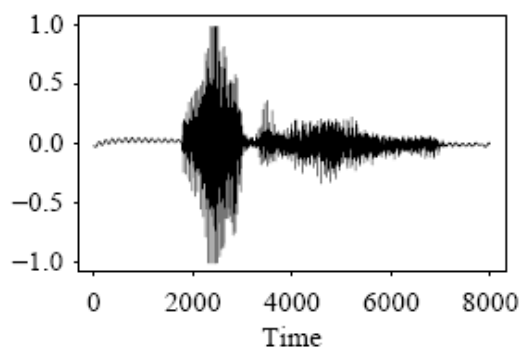
- 自适应差分脉冲编码调制（Adaptive Differential Pulse Code Modulation, ADPCM）
 - 综合APCM的自适应特性和DPCM系统的差分特性，性能较好
 - 利用自适应的思想改变量化阶的大小
 - 使用过去的样本值预测下一个输入样本的预测值

3.2.3 自适应差分脉冲编码调制

- 接收端的译码器使用与发送端相同的算法，利用传送来的信号来确定量化器和逆量化器中的量化阶大小，并且用它来预测下一个接收信号的预测值



3.2.3 ADPCM

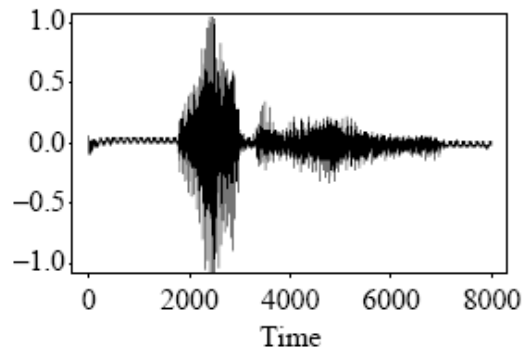


(a)

线性PCM

8kHz采样频率

16 bits per sample

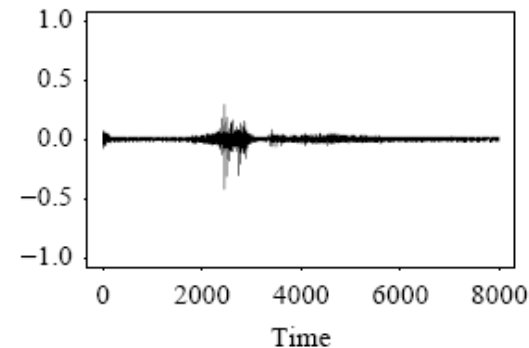


(b)

G.721(ADPCM)

8kHz采样频率

4 bits per sample



(c)

两者差值

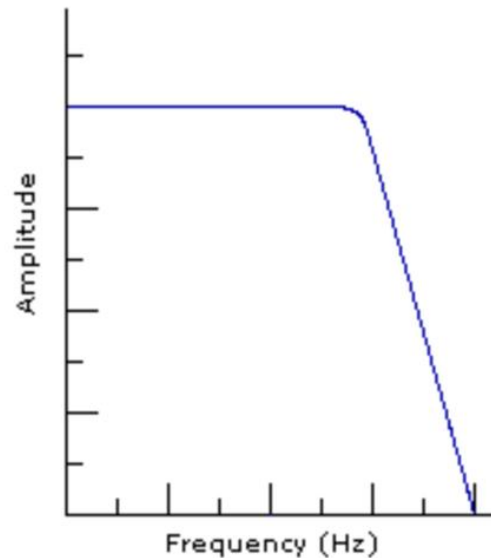
“Audio”波形图

3.2.4 语音编码器(Vocoders)

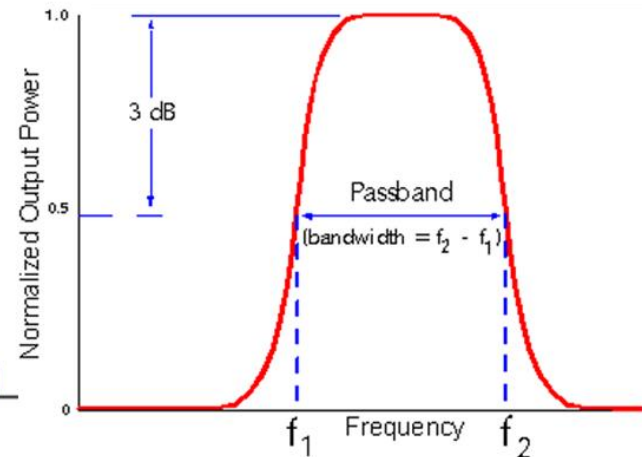
- 对语音进行建模，以便以尽可能少的比特数表示语音显著特征
 - 将信号分解为频率分量并对它们进行建模（通道声码器和共振峰声码器）
 - 使用时序语音波形模型（线性预测编码, LPC）

3.2.4 子带编码(sub-band coding, SBC)

- 用一组带通滤波器(band-pass filter, BPF)把输入声音信号的频带分成若干个连续的频段，每个频段称为子带。对每个子带中的声音信号采用单独的编码方案去编码



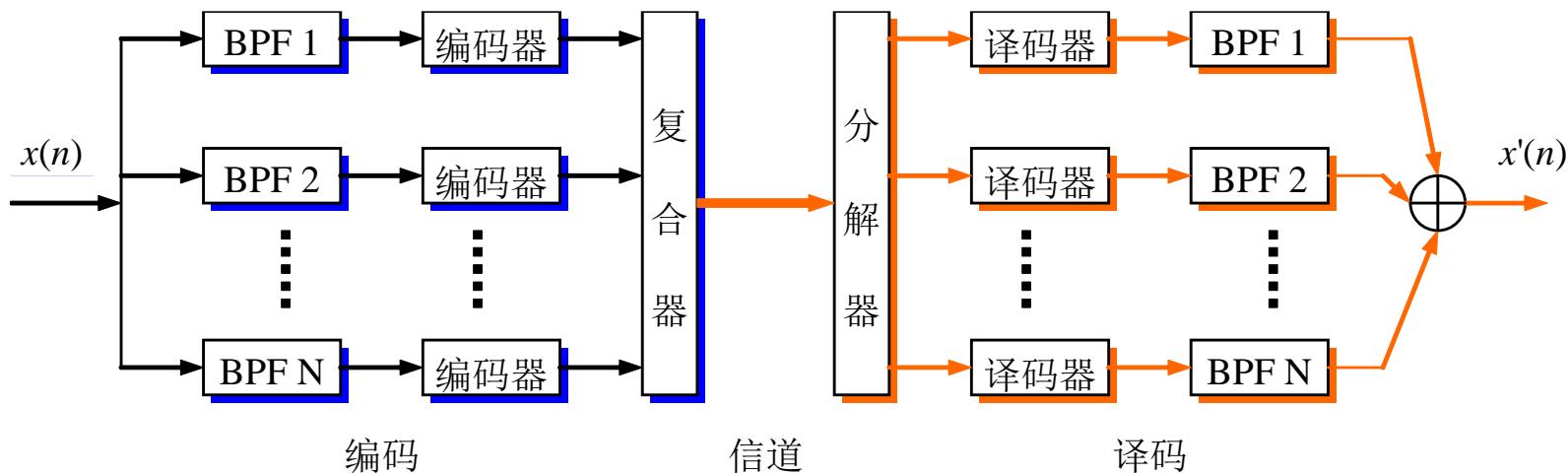
低通滤波器



带通滤波器

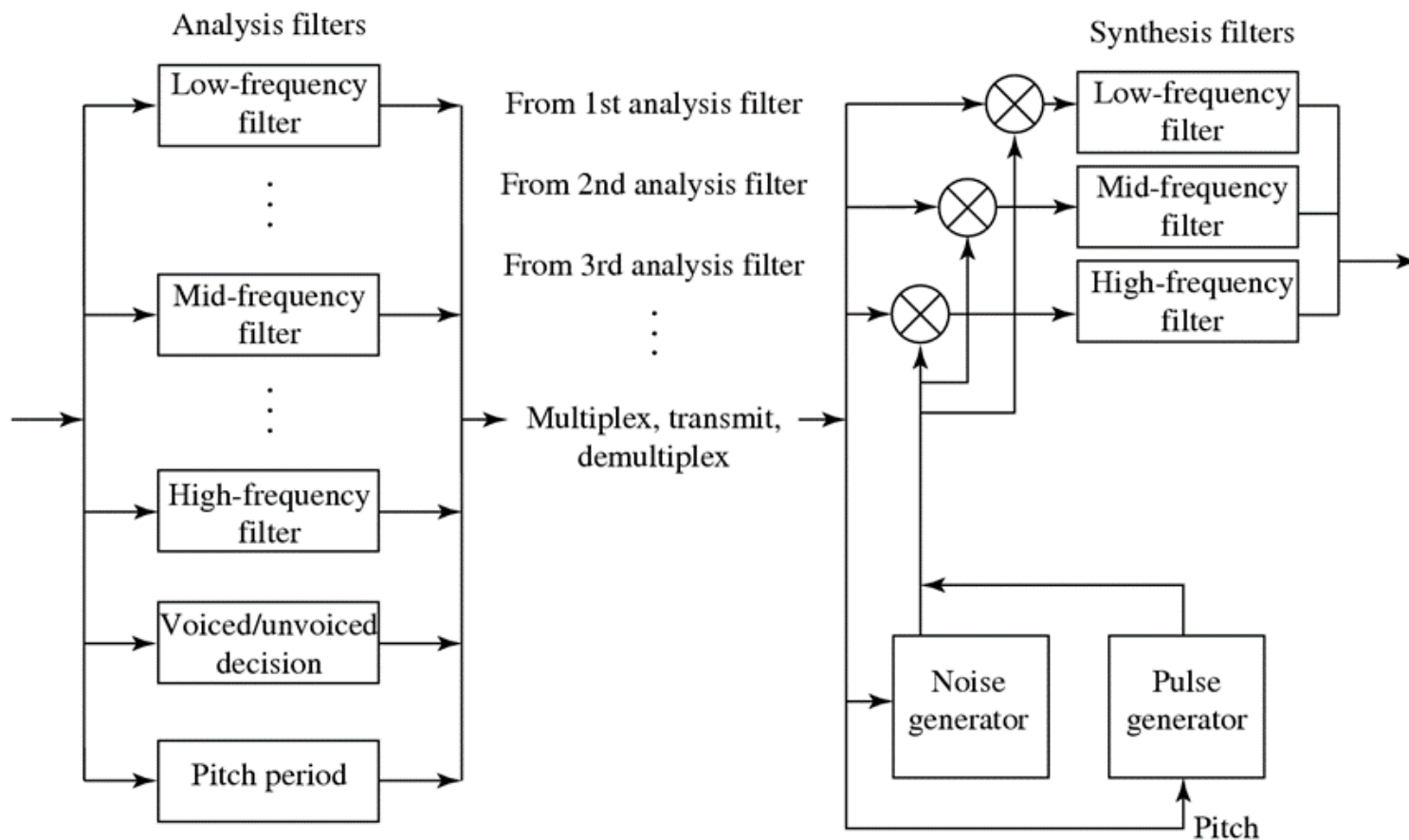
3.2.4 G.722 SB-ADPCM编译码器

- 编码/译码器采用ADPCM



子带编码方块图

3.2.4 子带语音编码




3.2.4 子带语音编码

- 通道声码器分析信号以确定语音的总体音调（低音或高音），以及语音的激励。
- 通道声码器应用声道传递模型来生成描述声音模型的激励参数矢量，并且还猜测声音是发声还是发声。

3.2.4 线性预测编码(LPC)

- 线性预测编码(Linear Predictive Coding, LPC)
 - LPC声码器直接从波形中提取语音的显著特征，而不是将信号转换到频域
 - 将话音生成机理模型化为一个离散的、时变的、线性的递归滤波器
 - 编码时使用线性预测分析话音波形产生声道激励和转移函数的参数，对声音波形的编码实际就转化为对这些参数的编码，这就可减少声音的数据量

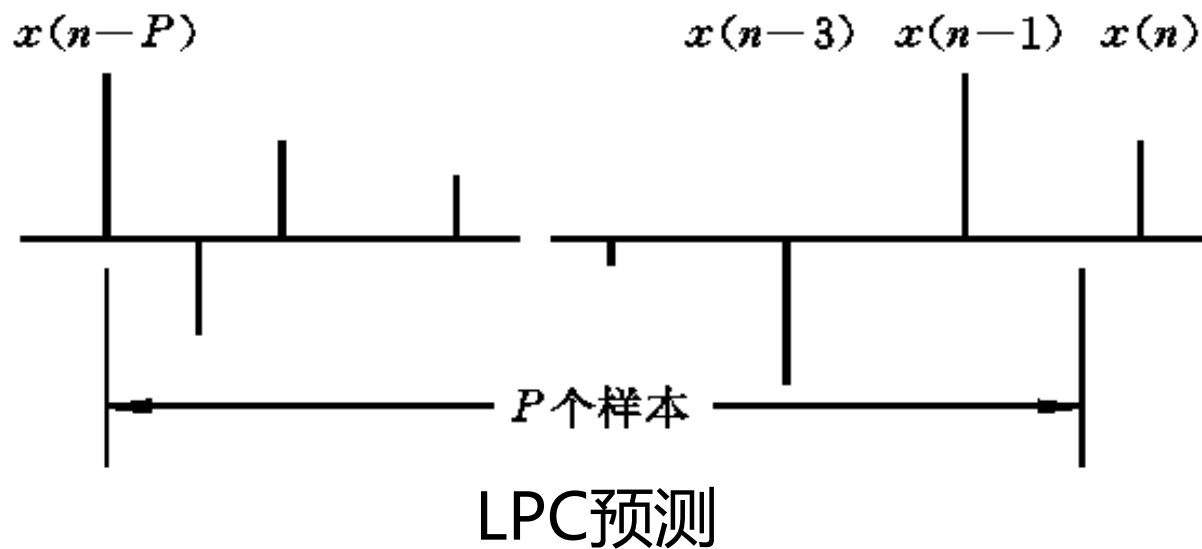
3.2.4 线性预测编码(Linear Predictive Coding, LPC)

- LPC声码器直接从波形中提取语音的显着特征，而不是将信号转换到频率域
- LPC特点：
 - 使用从给定激励产生的时变声道声音模型
 - 仅传输模拟声道形状和激励的一组参数，而不传输实际信号或差异  数据率低

3.2.4 线性预测编码(LPC)

■ 线性预测器

- 使用过去的 P 个样本值的线性组合来预测当前采样值 $x(n)$



3.2.4 线性预测编码(LPC)

- 译码时使用线性预测分析得到的参数，通过话音合成器重构话音。合成器实际上是一个离散的随时间变化的时变线性滤波器，它代表人的话音生成系统模型
- 时变线性滤波器
 - 分析话音波形时，当作预测器使用
 - 合成话音波形时，当作生成模型使用

3.2.4 线性预测编码(LPC)

- 预测值

$$\begin{aligned}x_{pre}(n) &= -[a_1x(n-1) + a_2x(n-2) + \cdots + a_px(n-p)] \\ &= -\sum_{i=1}^p a_i x(n-i)\end{aligned}$$

预测误差

$$e(n) = x(n) - x_{pre}(n) = \sum_{i=0}^p a_i x(n-i)$$

在给定的时间范围里, 如 $[n_0, n_1]$, 使 $e(n)$ 的平方和即

$$\beta = \sum_{n=n_0}^{n_1} [e(n)]^2$$

为最小, 这样可使预测得到的样本值更精确。

3.2.4 LPC的应用

- 全球数字移动通信系统（Global System for Mobile communications, GSM）



- GSM算法使用了LPC码。
- GSM算法是1992年柏林技术大学根据GSM协议开发的，这个协议是欧洲最流行的数字蜂窝电话通信协议。

3.2.4 LPC的应用

■ GSM的性能

- GSM的输入：数据分成帧(frame)，一帧(20毫秒)由带符号的160个样本组成，每个样本为13位或16位的线性PCM(linear PCM)码
 - 使用的采样频率为8 kHz时，如果每个样本为16位，那么未压缩的话音数据率为128 kb/s
- GSM的输出：一帧(160×16 位)的数据压缩成260位的GSM帧，相当于13 kb/s。由于260位不是8位的整数倍，因此编码器输出的GSM帧为264位的线性PCM码
 - 使用GSM压缩后的数据率为：
- GSM的压缩比：

编码算法的性能比较

部分编码器的MOS分

编码器	MOS分
64 kb/s脉冲编码调制(PCM)	4.3
32 kb/s自适应差分脉冲编码调制(ADPCM)	4.1
16 kb/s低时延码激励线性预测编码(LD-CELP)	4.0
8 kb/s码激励线性预测编码(CELP)	3.7
3.8 kb/s码激励线性预测编码(CELP)	3.0
2.4 kb/s线性预测编码(LPC)	2.5

常见音频编码格式

■ 常见音频编码格式

 MP3 (MPEG-1 and/or MPEG-2 Audio Layer III) 有损压缩



AAC(Advanced Audio Coding):包含在MPEG2, MPEG4中；有损压缩；相同速率下音质比MP3更好。



- APE(Monkey's Audio):无损压缩，压缩比比FLAC大，解码更复杂



- FLAC(Free Lossless Audio Codec)：无损压缩，50%-60%

总结

- 声音、声音数字化

- 声音的编码

- PCM,APCM
- DM, ADM
- DPCM, ADPCM
- SBC, LPC

作业

- 课后题3.7

END

