



多媒体技术

语音编码

数字语音编码

- 语音是携带信息的重要媒体，广泛用在多媒体作品和多媒体通信系统中
- 多个国际组织制定了一系列的语音编码标准，并继续开发质量更高、数据率更低的语音编码标准
- 数字语音编码的研究和开发方向
 - 在满足语音质量要求的前提下，尽量降低数字语音的数据率，以降低对存储容量和传输带宽的要求

数字语音编码

- 语音(speech): 频率为300~3400 Hz的信号
 - 声音(audio): 频率为20~ 20 000Hz的信号
- 语音是声音，但声音不一定是语音
- 本章介绍声音的基础知识和语音编码方法
 - 波形编码
 - 参数编码(LPC为主)
 - 混合编码(以CELP为主)
 - 这些算法主要针对语音编码，但其中的许多算法也适合声音编码

内容提纲

- 声音信号数字化
- 语音编码介绍
- 波形编码
- 参数编码
- 混合编码

声音信号的数字化

- 从模拟过渡到数字

- 大多数电信号是模拟信号，用模拟器件处理难度大，精度低，成本高
- 用数字表示模拟量：把模拟信号转变成数字信号，使用数字信号处理器(digital signal processor, DSP)
- 在数字域而不在模拟域中做信号处理的主要优点
 - 数字信号计算是精确运算，不受时间和环境变化影响
 - 复杂的数学运算容易实现
 - 可对数字运算部件进行编程

声音信号的数字化

- 模拟信号与数字信号

- 模拟信号：时间和幅度上都是连续的信号

- 数字信号：时间和幅度都用离散值表示的信号

- 在特定时刻对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

- 采样得到的幅值是无穷多个实数值中的一个。如果把信号幅度取值的数目加以限定，这种由有限数目的数值组成的信号称为离散幅度信号

声音信号的数字化

- 模拟信号与数字信号

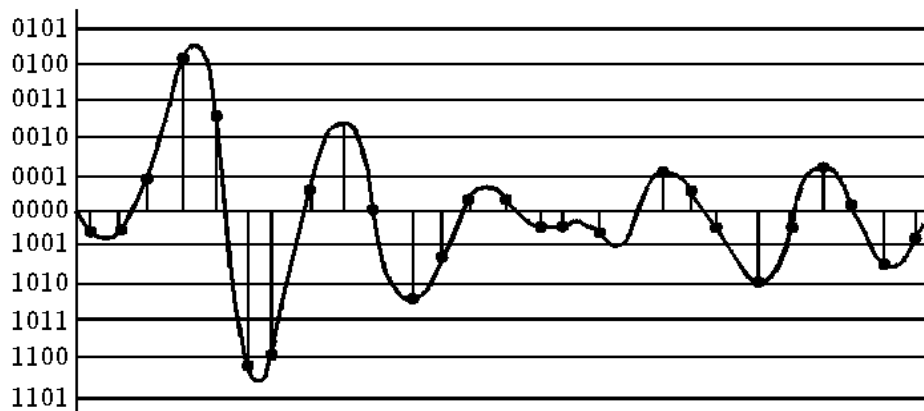
- 【例】假设输入电压范围为 $0.0 \sim 0.7 \text{ V}$ ，其取值只限定为 $0, 0.1, 0.2, \dots, 0.7$ 共8个值

- 如果采样得到的幅度值是 0.123 V ，它的取值就算作 0.1 V
 - 如果采样得到的幅度值是 0.26 V ，它的取值就算作 0.3 V

声音信号的数字化

- 数字化的概念

- 声音进入计算机的第一步就是数字化
- 模拟信号在时间和幅度都用离散值表示



声音信号的数字化

- 采样速率/采样频率

- 由信号本身包含的最高频率决定，信号的频率越高，需要的采样速率就越高
- 奈奎斯特理论(Nyquist theory)，采样速率不应低于声音信号最高频率的两倍
- 使用奈奎斯特理论采样，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(lossless digitization)

声音信号的数字化

- 采样精度

- 度量声音波形幅度的精确程度，通常用存储每个声音样本的位数(n)来表示
 - 例如，每个声音样本用16位(2字节)表示，测得的声音样本值是在 $[0\sim 65535]$ 范围里的数，采样精度是 $1/65536$
- 精度是度量模拟信号的最小单位，称为量化阶(quantization step size)
 - 如将 $0\sim 1$ V的电压用 $n=256$ 个数表示，其量化阶等于 $1/256$ V。

声音信号的数字化

- 声音质量与数据率

- 根据声音频带，声音质量分成5个等级，由低到高

- 电话(telephone)
- 调幅(amplitude modulation, AM)广播声音
- 调频(frequency modulation, FM)广播声音

表 4-1 声音质量和数据率

质量	采样速率 (kHz)	样本精度 (bit/s)	单道声/ 立体声	数据率(kb/s) (未压缩)	频率范围
电话*	8	8	单道声	64.0	200~3 400 Hz
AM	11.025	8	单道声	88.2	20~15 000Hz
FM	22.050	16	立体声	705.6	50~7 000Hz
CD	44.1	16	立体声	1411.2	20~20 000 Hz
DAT	48	16	立体声	1536.0	20~20 000 Hz
* 电话使用 μ 律编码，动态范围为 13 位而不是 8 位					

声音信号的数字化

● 声音质量的MOS评分标准

➤ 衡量声音质量方法

- 声音信号的带宽
- 客观质量度量：主要用信噪比(SNR)
- 主观质量度量：人的感觉更具决定意义

表 4-2 声音质量 MOS 评分标准

➤ 声音

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)

内容提纲

- 声音信号数字化
- 语音编码介绍
- 波形编码
- 参数编码
- 混合编码

语音编码介绍

- 语音编码 (speech coding) 方法

- 针对语音数字数据的压缩技术，主要用于语音通信

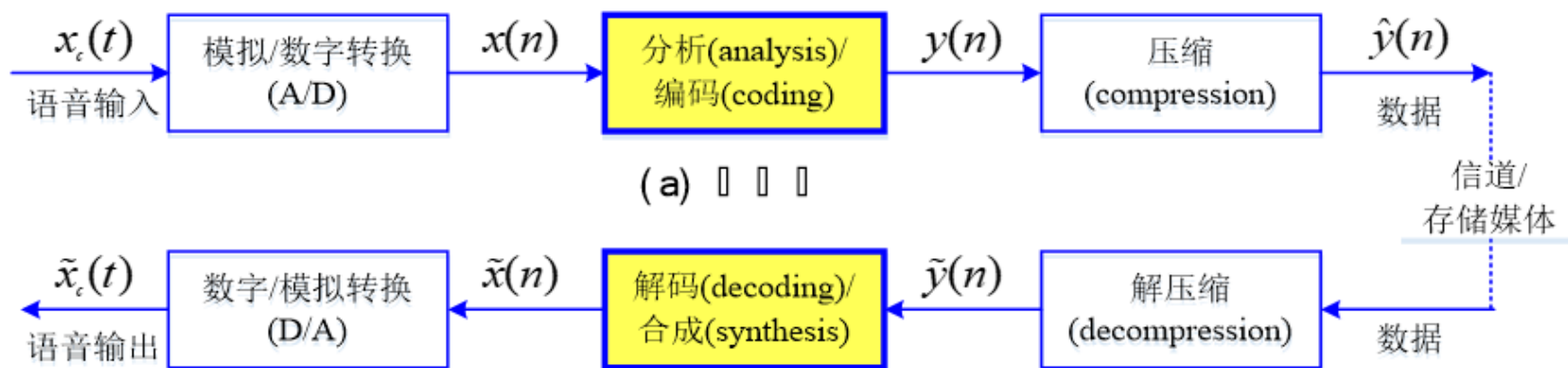


图4-3 语音编码概念

- 输入语音 经过A/D转换(采样和量化)后，通过分析和编码转换成代表语音的数字信号，解码过程与编码相反

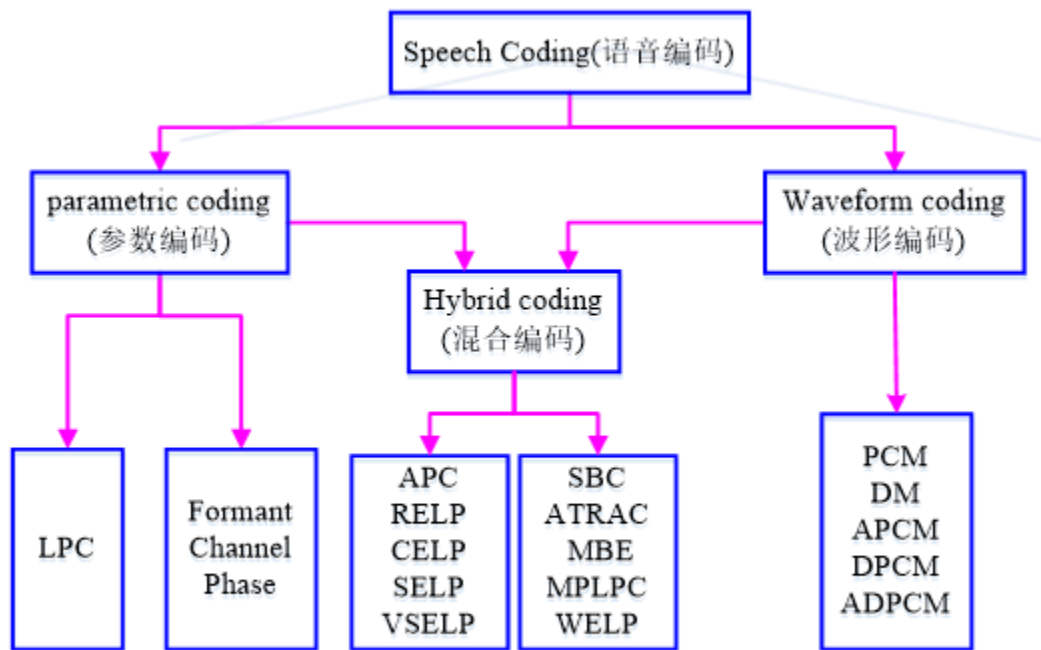
语音编码介绍

- 编码方法：三种类型

- 波形编码

- 参数编码

- 混合编码



语音编码介绍

- 波形编码

- 用数字形式精确地表示模拟信号波形的编码方法，不考虑语音产生和感知特性
- 编码器输出的数据速率为9.6~64 kbps或更高(取决于采样速率和量化精度)

语音编码介绍

- 波形编码

- 算法包括

- PCM(pulse code modulation): 脉冲编码调制
- APCM(adaptive pulse code modulation): 自适应脉冲编码调制
- DPCM(differential pulse code modulation): 差分脉冲编码调制
- ADPCM(adaptive difference pulse code modulation): 自适应差分脉冲编码调制

语音编码介绍

- 参数编码

- 利用发音器官生成语音信号的模型，对从语音信号中抽出的语音特征参数(如发音模型、有声/无声、音量大小、音调)进行编码的方法
- 解码器根据模型参数重构语音信号
- 编码器输出的数据速率约为2~4.8 kbps

语音编码介绍

- 参数编码

- 算法包括

- LPC(Linear Predictive Coding): 线性预测编码
 - Channel coding: 信道编码
 - Phase coding: 相位编码
 - Formant coding: 共振峰编码

语音编码介绍

- 混合编码

- 综合使用波形编码和参数编码技术，组合波形特性和语音特征参数的编码方法
- 既有波形编码语音质量高的优点，又有参数编码数据速率低的优点
- 主要用在语音质量要求高的移动通信系统
- 混合编码器输出的数据速率为4.0~16 kbps

语音编码介绍

- 混合编码

- 算法包括

- APC(adaptive predictive coding): 自适应预测编码
- RELP(residual-excited linear prediction): 残余激励线性预测
- CELP(code excited linear predictive): 码激励线性预测
- SBC(sub-band coding): 子带编码

内容提纲

- 声音信号数字化
- 语音编码介绍
- 波形编码
- 参数编码
- 混合编码

波形编码

- 基本思想

- 使重建语音信号的波形与原始信号的波形尽量接近

- 语音质量

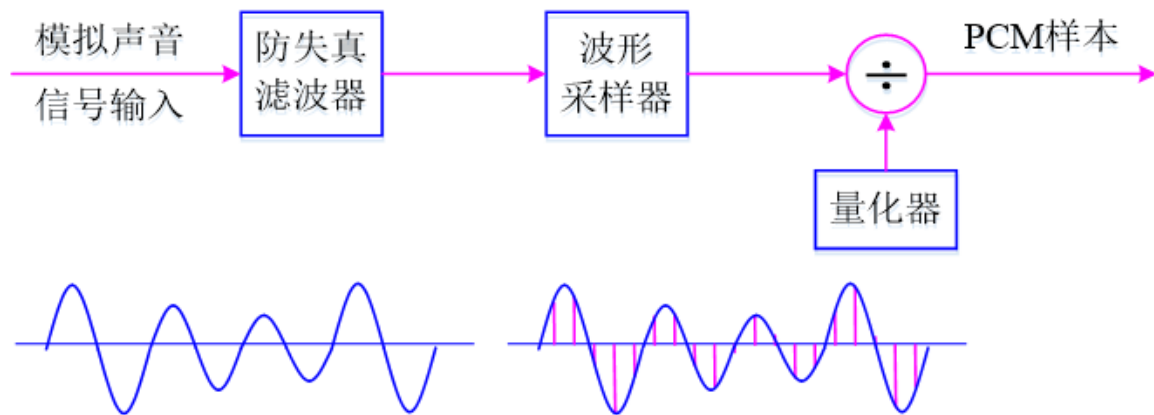
- 语音编码质量最好的编码方法

- 编码方法

- PCM, DM, ADM, APCM, DPCM, ADPCM和SB-ADPCM

PCM编码 (G.711)

- 脉冲编码调制(pulse code modulation, PCM)
 - 概念最简单、理论最完善的编码系统
 - 最早研制成功、使用最为广泛的编码系统
 - 数据量最大的编码系统
 - 1972年ITU将它作为G.711声音(audio)编码标准

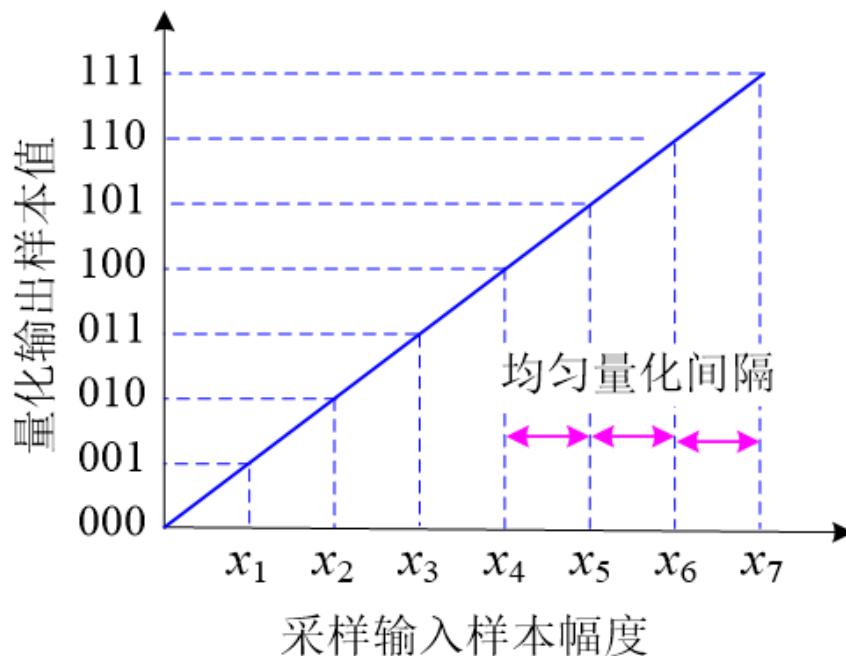


PCM编码 (G.711)

- 均匀量化(也称线性量化)

- 采用相等的量化间隔对采样信号进行量化

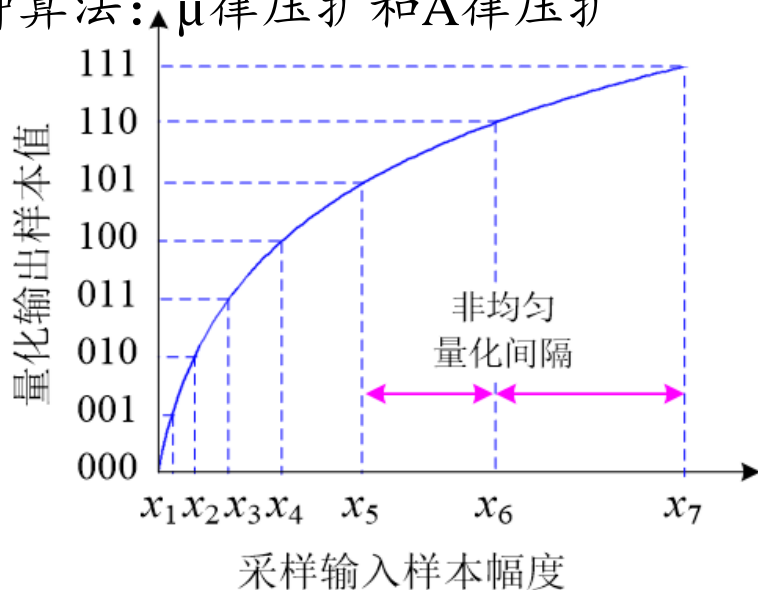
- 量化误差或量化噪声：样本值Y和原始值X之差 $E=Y-X$



PCM编码 (G.711)

● 非均匀量化

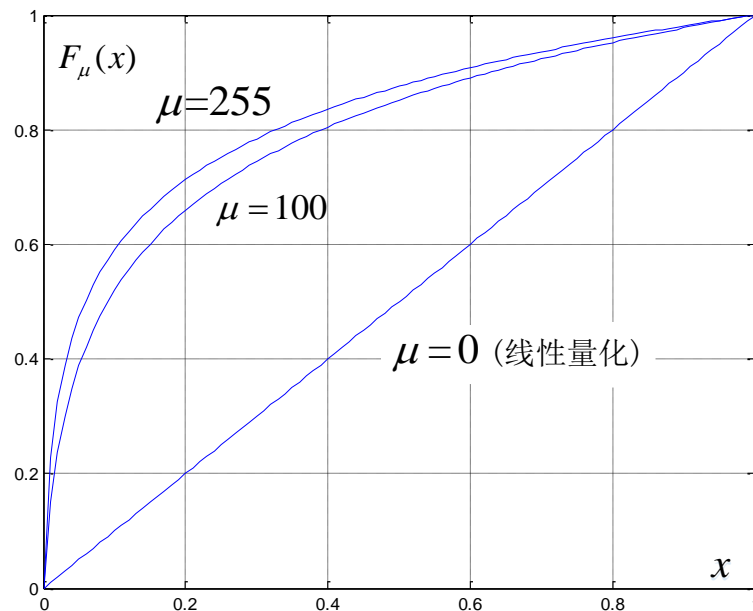
- 基本想法：对幅度大的信号采用大的量化间隔，幅度小的信号采用小的量化间隔
- 可在满足精度要求的情况下用较少的位数来表示
- 定义了两种算法： μ 律压扩和A律压扩



PCM编码 (G.711)

- 非均匀量化

- μ 律压扩



(1) μ 律压扩按下面的式子确定量化输入和输出的关系,

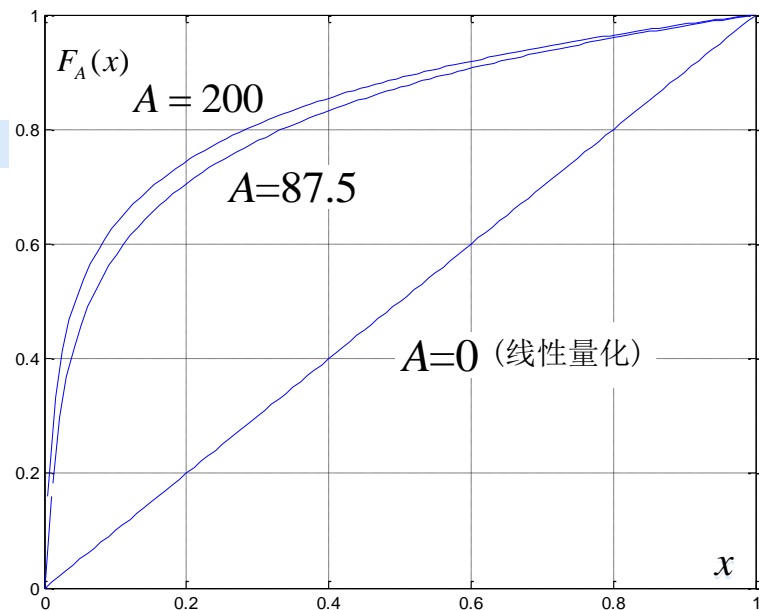
$$F_{\mu}(x) = \text{sgn}(x) \frac{\ln(1 + \mu |x|)}{\ln(1 + \mu)} \quad (4-2)$$

式中, x 为输入信号幅度, 规格化成 $-1 \leq x \leq 1$; 函数 $\text{sgn}(x)$ 的值为 1, 符号同 x 的正负号; μ 为确定压缩量的参数, 它反映最大量化间隔和最小量化间隔之比, 取 $100 \leq \mu \leq 500$ 。

PCM编码 (G.711)

- 非均匀量化

- A律压扩



(2) A 律压扩按下面的式子确定量化输入和输出的关系：

$$\begin{aligned} F_A(x) &= \text{sgn}(x) \frac{A|x|}{1 + \ln A}, & 0 \leq |x| \leq 1/A \\ F_A(x) &= \text{sgn}(x) \frac{1 + \ln(A|x|)}{1 + \ln A}, & 1/A < |x| \leq 1 \end{aligned} \quad (4-3)$$

式中， x 为输入信号幅度，规格化成 $-1 \leq x \leq 1$ ； $\text{sgn}(x)$ 为 x 的极性； A 为确定压缩量的参数，反映最大量化间隔和最小量化间隔之比。

APCM编码

- APCM(自适应脉冲编码调制)的概念
 - APCM=adaptive pulse code modulation
 - 根据输入信号幅度或音节来改变量化阶的大小
 - 自适应：量化阶的大小每隔几个样本改变一次
 - 音节自适应(量化阶的大小在较长时间周期里发生变化)

APCM编码

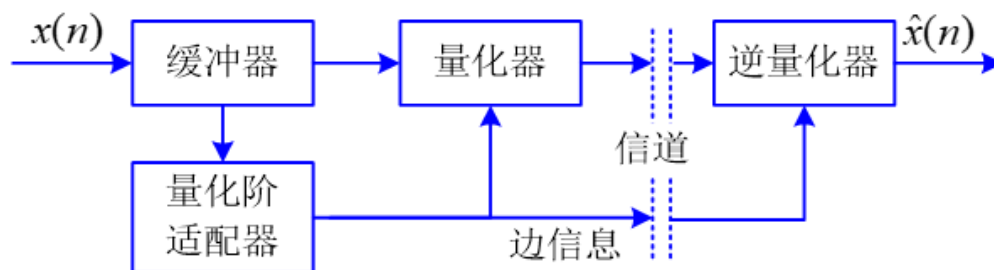
- APCM(自适应脉冲编码调制)的概念

- 改变量化阶大小的两种方法

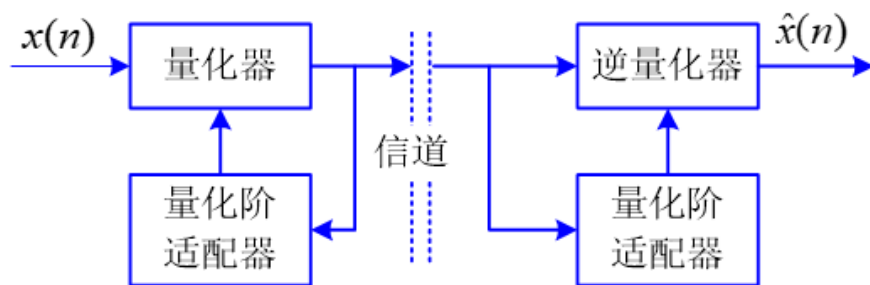
- 前向自适应(forward adaptation): 根据未量化的样本值的均方根估算输入信号的幅度, 以此来确定量化阶大小, 并对其幅度进行编码作为边信息 (side information)
- 后向自适应(backward adaptation): 从量化器刚输出的过去样本中来提取量化阶信息, 不需要传送边信息

APCM编码

- APCM(自适应脉冲编码调制)的概念



(a) □ □ □ □ □

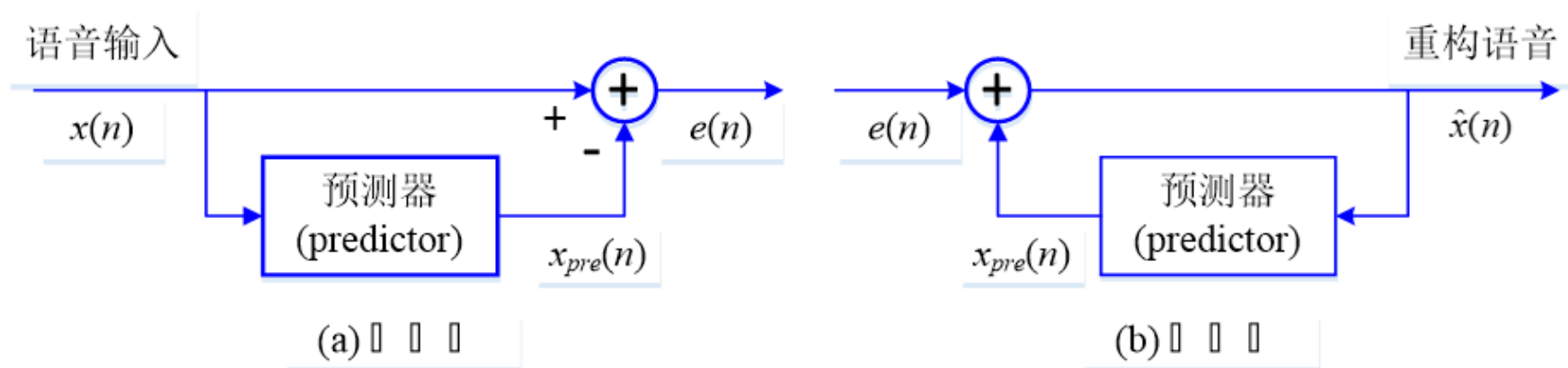


(b) □ □ □ □ □

DPCM编码

● 预测编码概念

- 预测编码(predictive coding)是对实际样本值与预测值之差进行编码的方法
- 在信道上传送的是相邻样本之间的差值，解码时使用差值和预测值进行重构



DPCM编码

● 预测编码举例

假设当前样本 $x(n)$ 的预测值 $x_{pre}(n)$ 为前一个样本值，预测误差 $e(n)$ 可写成，

$$x_{pre}(n) = x(n-1)$$

$$e(n) = x(n) - x(n-1)$$

对差值 $e(n)$ 编码比对原始语言样本编码，可用比较少的位数来表达一个样本。

对于具体的语音序列样本，如 $\dots, x(n-2), x(n-1), x(n), \dots$ ，如果使用过去几个样本值来预测当前的样本值，合理选择预测系数，产生的预测误差会更小。预测函数可写成，

$$x_{pre}(n) = \sum_{k=1}^N a_{n-k} x(n-k)$$

其中， a_{n-k} 为预测系数； N 为参加预测的样本数，通常 $N=2, 3$ 或 4

DPCM编码

● 预测编码举例

【例】 假设 $x(n-3)=21$, $x(n-2)=24$, $x(n-1)=27$, $x(n)=23$, 分别求 $N=1$ 和 $N=2$ 的预测误差 $e(n)$ 。

(1) 当 $N=1$ 时, $e(n)=23-27=-4$ 。

(2) 当 $N=2$ 时, 为简单起见, 假设预测系数 $a_{n-1}=a_{n-2}=1/2$, 那么

$$x_{pre}(n) = a_{n-1}x(n-1) + a_{n-2}x(n-2) = (27 + 24) / 2 \Rightarrow 25$$

$$e(n) = x(n) - x_{pre}(n) = 23 - 25 = -2$$

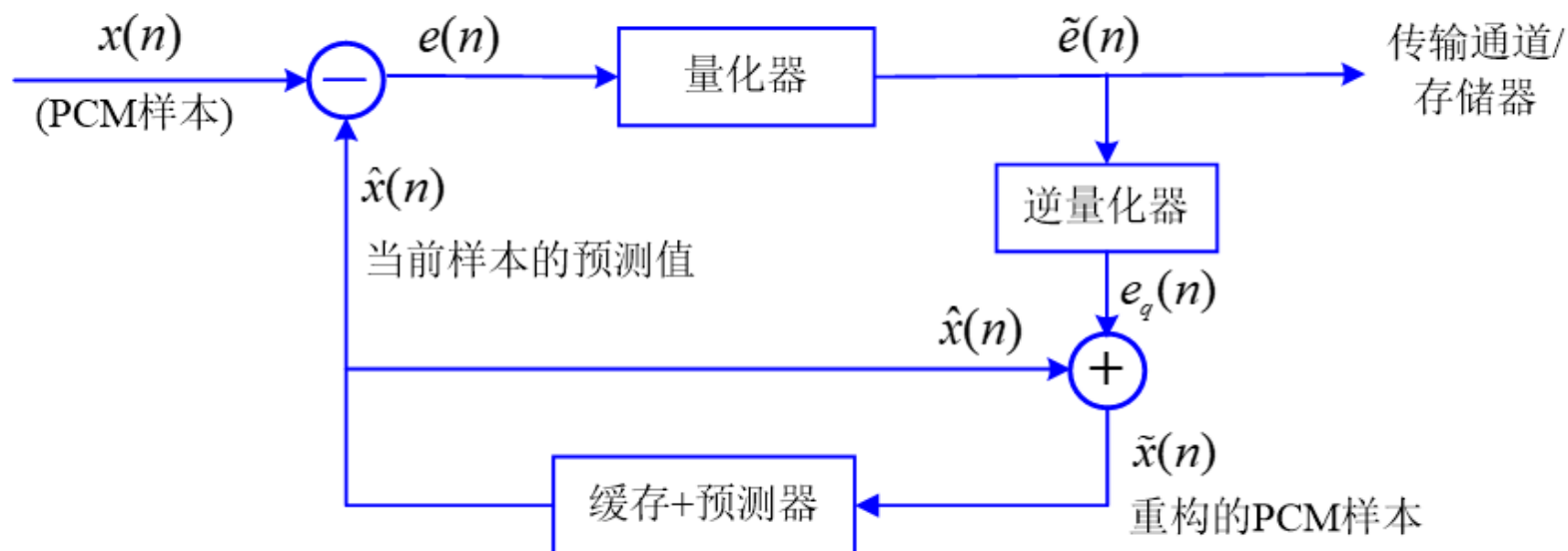
DPCM编码

- DPCM（差分脉冲编码调制）概念

- DCPM=differential pulse code modulation：利用样本与样本之间存在冗余进行编码的数据压缩技术
- 基本思想：根据过去的样本估算下一个样本信号幅度，称为预测值，然后对实际信号值与预测值之差进行量化编码

DPCM编码

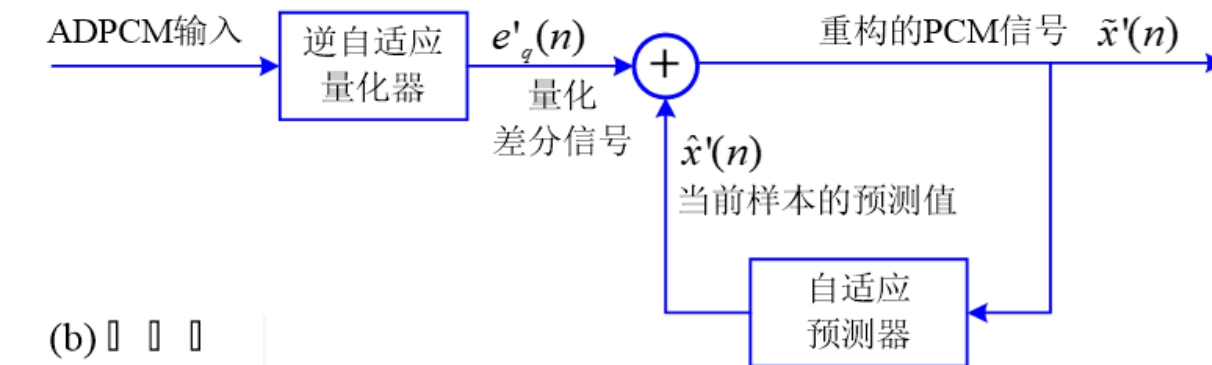
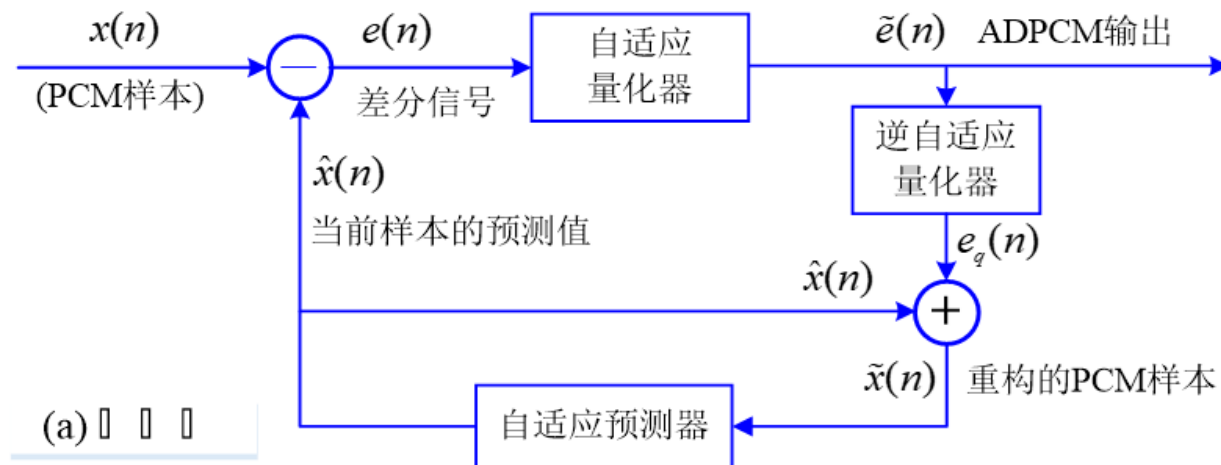
- DPCM（差分脉冲编码调制）概念



ADPCM编码 (G.726)

- ADPCM(自适应差分脉冲编码调制)的概念
 - ADPCM=adaptive difference pulse code modulation
 - 综合APCM的自适应特性和DPCM的差分特性
 - 核心理念
 - 利用自适应改变量化阶大小，用小的量化阶(step-size)编码小的差值，用大的量化阶编码大的差值
 - 使用过去的样本值估算当前输入样本的预测值，使实际样本值和预测值之间的差值总是最小

ADPCM编码 (G.726)



SB-ADPCM编码

- SB-ADPCM(G.722)

- 1988年CCITT制定G.722推荐标准

- 数据率为64 kb/s的7 kHz声音信号编码(7 kHz Audio-coding with 64 kb/s)

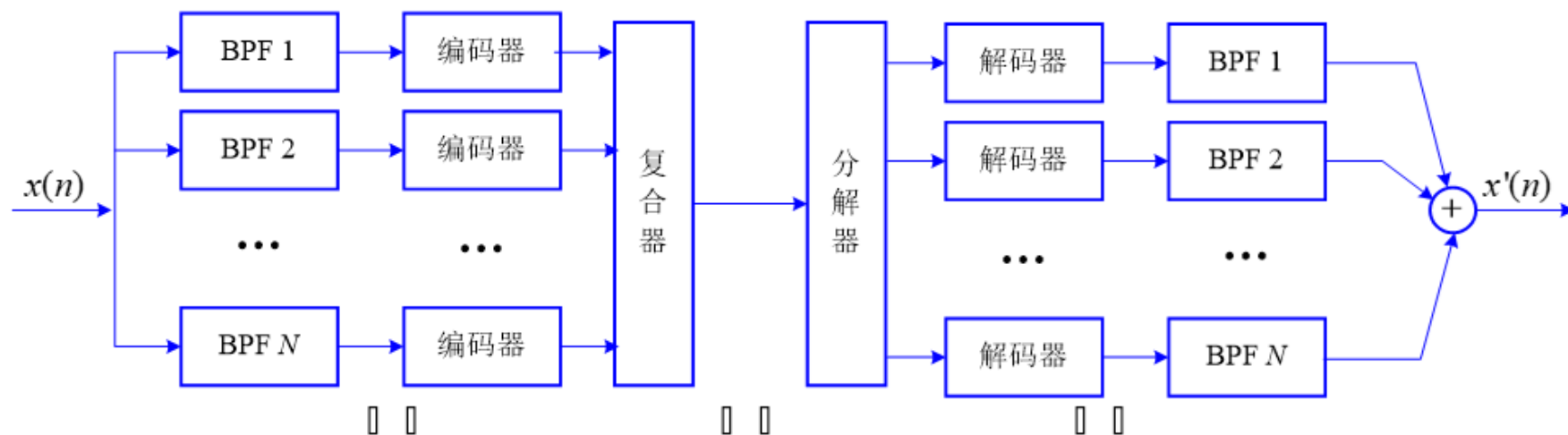
- 语音质量由电话质量提高到AM广播质量，数据传输率仍保持为64 kb/s

- 宽带语音是指带宽在50~7000 Hz的语音，在可懂度和自然度方面都比带宽为300~3400 Hz的语音有明显提高，也更容易识别对方的说话人

子带编码

- 子带编码(sub-band coding, SBC)的基本思想
 - 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
 - 对每个子带的声音信号采用单独(ADPCM, APCM或PCM)的编码方案编码
- 子带编码的两个特点
 - 可对每个子带信号分别进行自适应控制, 按能量调节量化精度
 - 可根据每个子带信号在感觉上的重要性, 采用不同的量化精度

子带编码



语音编码标准

- 国际上，对语音信号压缩编码的审议在CCITT下设的第15研究组进行，相应的建议为G系列。如：G.711、G.721、G.722、G.723等
 - **G.711**是CCITT为话音信号（300~3.4kHz）制定的编码标准。8kHz、8位/样本、64kb/s — A律和u律PCM
 - **G.726**是CCITT为话音信号制定的编码标准，可将64kb/s的PCM信号转换为40kb/s、32kb/s、24kb/s、16kb/s的ADPCM
 - **G.722**是CCITT推荐的音频信号编码标准。信号带宽为7kHz，采样频率16kHz，保持数据率仍为64kb/s — SB-ADPCM

内容提纲

- 声音信号数字化
- 语音编码介绍
- 波形编码
- 参数编码
- 混合编码

参数编码

- 参数编码（ parametric coding ） 的概念

- 参数编码：对语音的特征参数进行编码，也称声源编码(source coding)/
- 利用发音器官生成语音信号的模型(语音生成模型)，从语音信号中抽出表示语音信号的特征参数
 - 参数：有声/无声、音调周期、音量和模仿生成语音的声道参数
 - 对特征参数编码，使用特征参数重构语音信号
 - 不苛求重构语音波形与原始波形一致，而是让听者在主观感觉上逼近原始语音，追求高数据压缩率
 - 可将语音数据压缩到2~4.8 kbps，语音质量达到人能听懂的水平，但缺乏自然度

参数编码

- 参数编码（ parametric coding ） 的概念
 - 编码算法：线性预测的编码称为线性预测编码(linear predictive coding, LPC)
 - 参数编码涉及较多数学知识，计算量比较大，再介绍LPC之前，先简单介绍数字滤波器、Z变换和差分方程。

数字滤波器

- 数字滤波器

- 可对数字信号执行滤波功能的软硬件
- 在模拟系统中，时间 t 是以秒作单位的连续变量。在数字系统中，离散的时间是整数值 n ，它们之间的关系是 $t=nT$ ， T 是以秒作单位的时间间隔，常用 n 表示时间，即 $x(t) \Rightarrow x(nT) \Rightarrow x(n)$ 或 x_n
- 最简单的滤波器

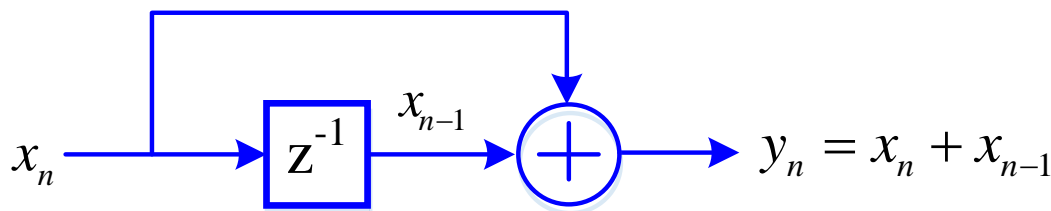
$$y_n = x_n + x_{n-1}$$

数字滤波器

- 最简单的滤波器

$$y_n = x_n + x_{n-1}$$

一阶零点低通滤波器可用如图 4-22 所示的信号流图表示。信号流图使用Z变换描述，图中的 z^{-1} 表示延时一个样本，于是 x_{n-1} 可写成 $x_{n-1} = z^{-1}x_n$ 。



Z变换

- Z变换的概念

- 滤波器对输入信号产生的影响可在时域或在频域中分析。在频域中分析时，常采用与傅立叶变换等价的Z变换(Z-transform)，对幅频特性和相频特性进行分析。
- Z变换是1947年由W.Hurewicz介绍的等效的拉普拉斯变换，将离散时间信号用复数频域中的信号来表达。实践表明，离散时间信号的Z变换是分析线性系统的重要工具

Z变换

Z变换定义为,

$$X(z) = \sum_{n=-\infty}^{\infty} x_n z^{-n}$$

其中, x_n 是已知的序列, z 是复数, 并限定 $z = e^{i\theta} = \cos \theta + i \sin \theta$ 为一个单位圆。

Z变换有许多定理, 其中一个称为位移定理: 若 $x(n)$ 的Z变换为 $X(z)$, 则 x_{n-k} 的Z变换为 $z^{-k}X(z)$,

$$Z[x_{n-k}] = z^{-k}X(z)$$

【例】 已知序列 $x_0=1, x_1=3, x_2=3, x_3=1$, 其他的 $x_n=0$, 这个系列的Z变换为,

$$\begin{aligned} X(z) &= x_0 z^0 + x_1 z^{-1} + x_2 z^{-2} + x_3 z^{-3} \\ &= 1 + 3z^{-1} + 3z^{-2} + z^{-3} \end{aligned}$$

差分方程

- 在数学上，差分方程是指任何类型的递推关系
 - 递推关系通常是指一个序列，在序列中的每一项定义为在该项之前几项的函数，序列的一个或多个初始项的值给定后，其后的项可用定义的函数重复计算。

【例】 斐波纳契数列(Fibonacci sequence)为[0, 1, 1, 2, 3, 5, 8, 13, 21, ...], 数列的递推关系是 $y_2 = y_1 + y_0, y_3 = y_2 + y_1, \dots$ 如图4-23所示，表示数列的差分方程是

$y_2 = y_1 + y_0$			$y_n = y_{n-1} + y_{n-2}$					
1	1	2	3	5	8	13	21	...
y_0	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_n

图4-23 递推与差分方程的概念

线性预测编码

- 线性预测编码(LPC)

- 通过分析语音波形获得语音生成模型的参数，将对声音波形的编码转化为对这些参数的编码，达到压缩语音数据的目的
- 解码器使用LPC分析得到的参数，通过语音合成器重构语音。语音合成器代表人的语音生成模型，是一个离散的时变线性滤波器
- 线性预测编码技术不仅用于声音的数据压缩，而且也用于自然图像的数据压缩

线性预测编码

- 线性预测的概念

线性预测是使用过去 p 个样本值 $s_{n-1}, s_{n-2}, \dots, s_{n-p}$ 的线性组合来预测当前的样本值 \hat{s}_n ，如图 4-26 (a) 所示，用下面的差分方程表示，

$$\hat{s}_n = a_1 s_{n-1} + a_2 s_{n-2} + \dots + a_p s_{n-p} = \sum_{k=1}^p a_k s_{n-k} \quad (4-9)$$

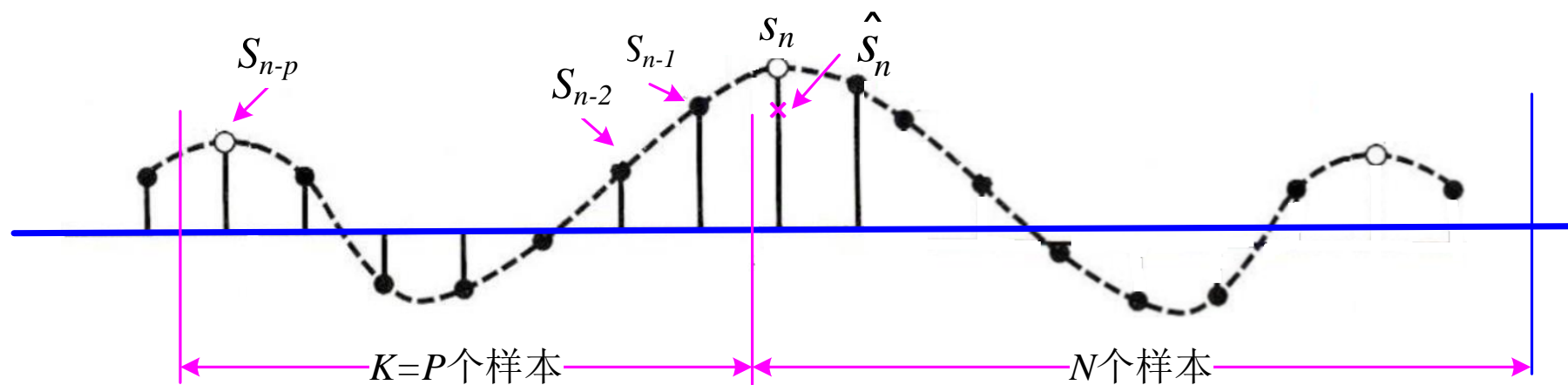
其中， a_k 称为预测系数。用它构造的预测器如图(b)所示， Z^{-1} 表示延时 1 个样本。

实际样本值与预测值的误差称为预测误差 e_n ，

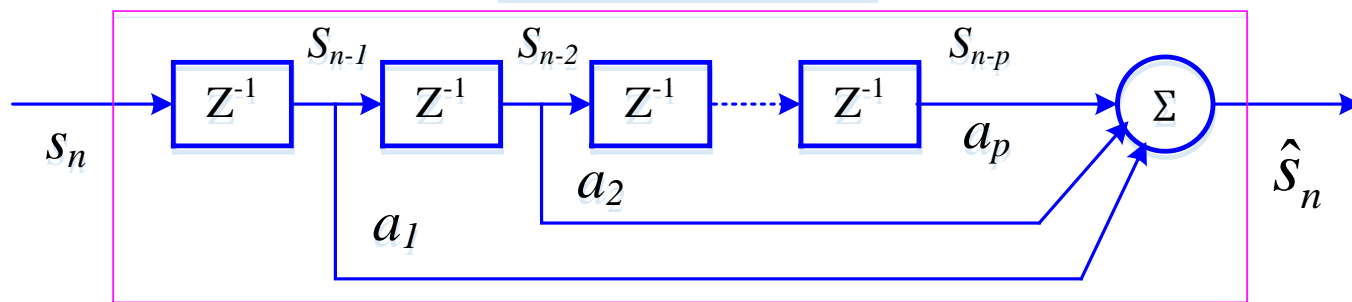
$$e_n = s_n - \hat{s}_n = s_n - \sum_{k=1}^p a_k s_{n-k} \quad (4-10)$$

预测误差 e_n 也称残差(residual)或语音合成的激励(excitation)信号。

线性预测编码



(a) 线性预测



(b) 线性预测器

线性预测编码

● 预测系数的求解

预测系数的求解问题表述为，对 N 个样本的空间，找出 p 个预测系数 a_1, a_2, \dots, a_p ，使预测误差为最小。预测误差 E 用每个样本的预测误差 e_n 的平方之总和来度量，

$$E = \sum_{n=0}^{N-1} e_n^2 = \sum_{n=0}^{N-1} (s_n - \sum_{k=1}^p a_k s_{n-k})^2 \quad (4-13)$$

这个函数的变量为 a_k ，当它取不同的值时，产生的预测误差 E 不同，而且有一个最小值。因为预测误差的最小值出现在 $\partial E / \partial a_k = 0$ 处，因此可分别对 a_1, a_2, \dots, a_p 求导，得到 p 个方程，求解这个方程组就可找到使预测误差为最小的 p 个预测系数 a_1, a_2, \dots, a_p 。

线性预测编码

- 预测系数的求解

假设一帧的语音数据 $N=160$ 个, 用 s_0, s_1, \dots, s_{159} 表示, 预测系数的数目 $p=10$, 将这两个数值代入式(4-13)得到,

$$E = \sum_{n=0}^{159} e_n^2 = \sum_{n=0}^{159} [s_n - (a_1 s_{n-1} + a_2 s_{n-2} + \dots + a_{10} s_{n-10})]^2$$

分别对 a_1, a_2, \dots, a_{10} 求导并设置为 0, 经过整理后得到,

$$\begin{aligned} \sum_{n=0}^{159} [s_n s_{n-1} - (a_1 s_{n-1} s_{n-1} + a_2 s_{n-2} s_{n-1} + \dots + a_{10} s_{n-10} s_{n-1})] &= 0 \\ \sum_{n=0}^{159} [s_n s_{n-2} - (a_1 s_{n-1} s_{n-2} + a_2 s_{n-2} s_{n-2} + \dots + a_{10} s_{n-10} s_{n-2})] &= 0 \\ \dots & \\ \sum_{n=0}^{159} [s_n s_{n-10} - (a_1 s_{n-1} s_{n-10} + a_2 s_{n-2} s_{n-10} + \dots + a_{10} s_{n-10} s_{n-10})] &= 0 \end{aligned}$$

线性预测编码

● 预测系数的求解

这个方程组可写成矩阵形式。假设在 N 个样本空间之外的样本值均为 0，矩阵中的元素可用 s_n 的自相关值 $R(k) = \sum_{n=0}^{159-k} s_n s_{n+k}$ 表示，于是这个方程组用矩阵表示为，

$$\begin{bmatrix} R(0) & R(1) & R(2) & R(3) & \cdots & R(9) \\ R(1) & R(0) & R(1) & R(2) & \cdots & R(8) \\ R(2) & R(1) & R(0) & R(1) & \cdots & R(7) \\ R(3) & R(2) & R(1) & R(0) & \cdots & R(6) \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ R(9) & R(8) & R(7) & R(6) & \cdots & R(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ \cdots \\ a_{10} \end{bmatrix} = \begin{bmatrix} R(1) \\ R(2) \\ R(3) \\ R(4) \\ \cdots \\ R(10) \end{bmatrix}$$

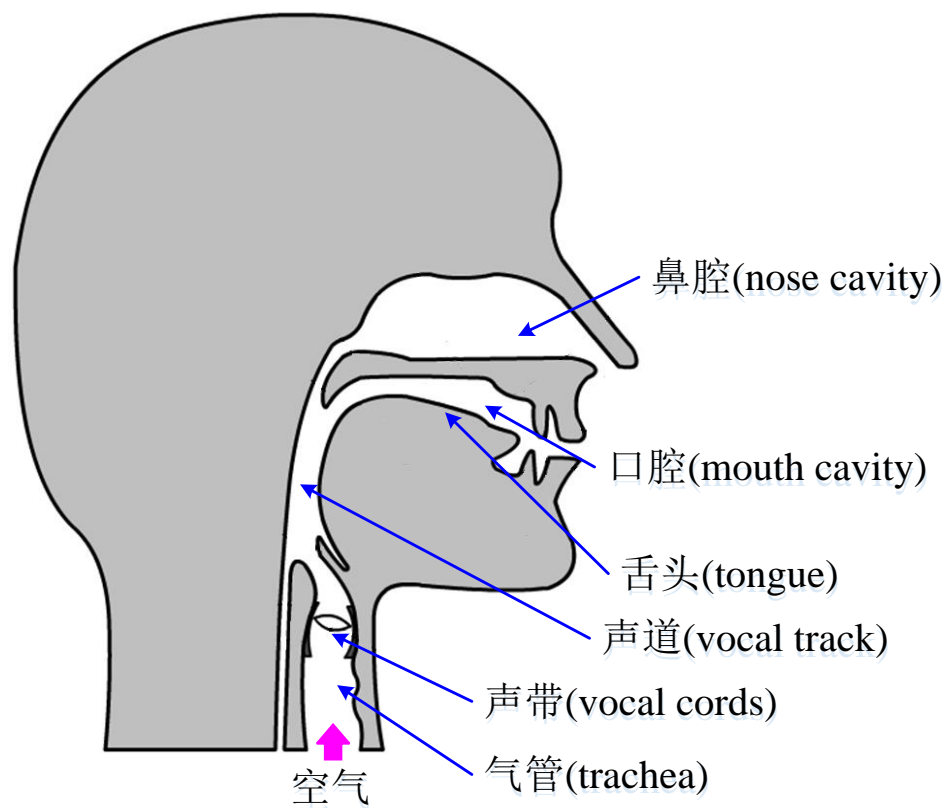
由于自相关值 $R(k)$ 可根据样本值求得，因此这个矩阵方程的预测系数可通过许多方法求解，如Levinson-Durbin递归法，矩阵运算法(可用MATLAB计算)或其他方法。介绍求解预测系数的书籍和文章都很多，网站上也容易找到通俗易懂的介绍和求解的源程序。

语音生成模型

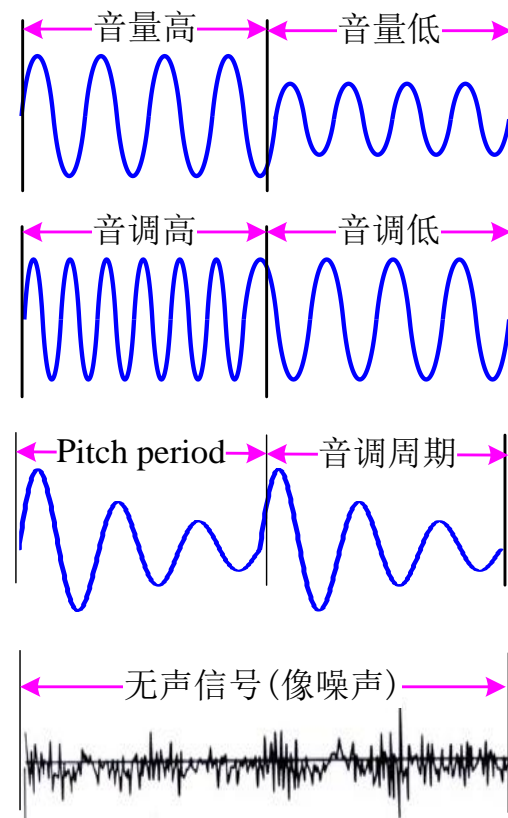
- 物理模型

- 人说话时的发音器官的物理模型如图(a)所示
- 语音音量、音调的高低和音调的含义，如图(b)所示
- 当我们说话时，来自肺部的空气通过声带、声道和口腔后从嘴巴发出就产生了语音

语音生成模型



(a) 发音器官的物理模型



(b) 语音的部分参数

语音生成模型

- 语音生成和发音器官的运动有如下特性：
 - 有声/无声：声带振动(开/闭)就有声音，声带不振动就无声音，但处于打开状态
 - 音调高低：声带的振动速率决定语音音调的高低，音调也称基频，反映音调周期(pitch)长短
 - 语音音量：肺部的空气量决定语音的音量大小

语音生成模型

- 语音生成和发音器官的运动有如下特性：

- 声道形状：形状决定语音，形状不同产生的语音不同。

声道被认为是滤波器，两个基本特性：

- 声道形状的变化相对比较慢(10~100 ms)，期间的声音被认为是平稳的，如在5~10 ms范围里可把语音看成是准周期性的脉冲串，但连续脉冲之间的间隔和脉冲幅度不完全相同
 - 声道的变化与空气量、有声/无声和声带振动频率密切相关，因此把它看成是一个语音合成滤波器。

语音生成模型

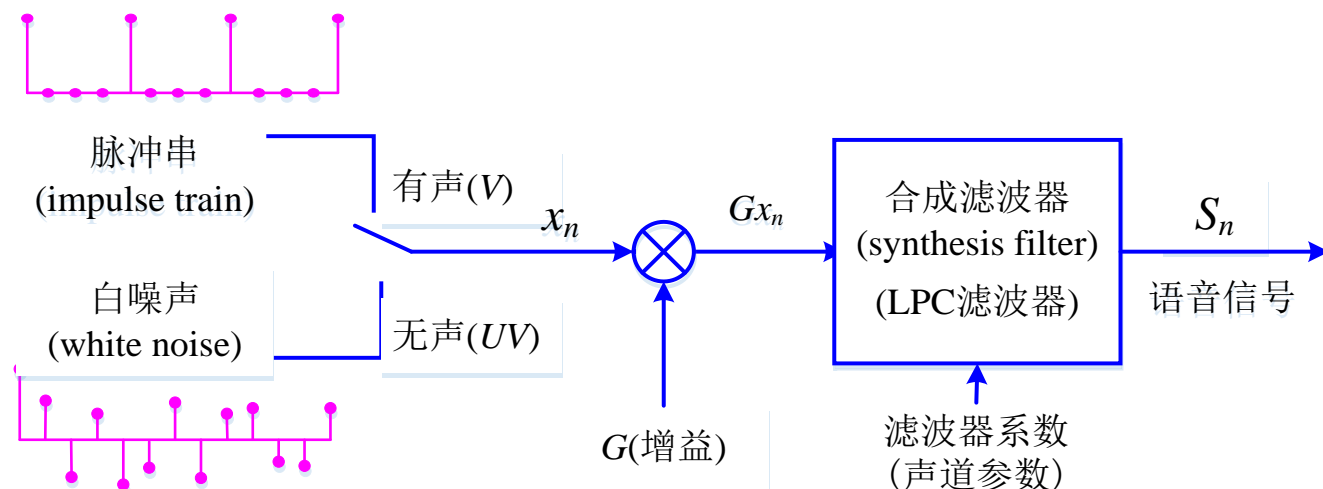
● 数学模型

➤ 发音器官用线性预测编码(LPC)模型描述

- 用在语音压缩、语音合成和语音识别技术中
- 经典数学模型，表示数字语音信号是合成滤波器的输出，合成滤波器的输入是表示有声的脉冲串或是表示无声的白噪声序列
- 物理模型与数学模型有如下对应关系，

物理模型	声道	声带振动	声带振动周期	擦音和爆破音	空气量
数学模型	合成滤波器 (LPC 滤波器)	V (有声)	T (音调周期)	UV (无声)	G (增益)

语音生成模型



- 开关：在有声(V)帧和无声(UV)帧之间切换
- 增益(gain)：一帧的能量大小
- 合成滤波器(PLC滤波器)：表示声道，滤波器系数反映合成滤波器的响应特性
- 音调周期T：有声的连续激励脉冲之间的时间间隔

语音生成模型

合成滤波器与下面的差分方程等效,

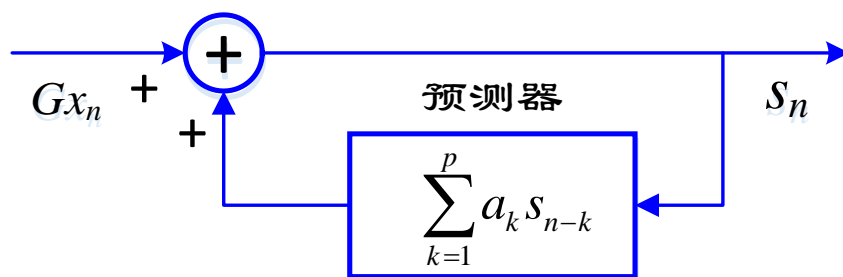
$$s_n = \sum_{k=1}^p a_k s_{n-k} + Gx_n \quad (4-14)$$

合成滤波器的传递函数为

$$H(z) = \frac{S(z)}{X(z)} = \frac{G}{1 - \sum_{k=1}^p a_k z^{-k}} \quad (4-15)$$

其中, x_n 是激励信号(excitation signal), s_n 是合成的语音信号, G 是语音帧的能量, a_k 是滤波器的系数。如果知道这 4 个参数, 就可产生合成的语音。

合成滤波器用图 4-30 表示。这个模型就是线性预测编码(LPC)中的解码器, 只是用激励信号 Gx_n 代替预测误差 e_n 。



LPC声码器

- 声码器(vocoder=voice coder): 语音编码+解码器
- 使用线性预测编码技术, 主要用于语音数据压缩
 - 如1984年美国国防部开发的安全电话语音标准FS-1015(LPC-10), 速率可低到2.4 kb/s

在线性预测编码中, 常把采样速率为 8 kHz的语音样本分成帧, 20 ms/帧, 50 帧/秒, 160 个语音样本/帧, $S = (s_0, s_1, \dots, s_{159})$, 用一个由 13 个值构成的矢量 A 代表 160 个样本值,

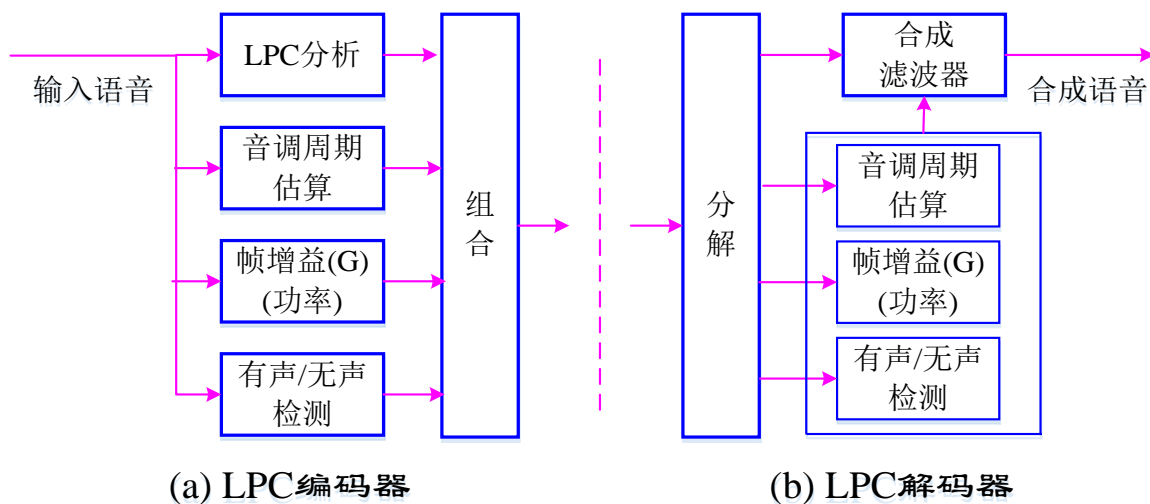
$$A = (a_1, a_2, \dots, a_{10}, G, V / UV, T)$$

- LPC 分析: 对给定的语音样本 S , 求解最适合表达 S 的矢量 A
- LPC 合成: 对给定的矢量 A , 使用合成滤波器产生 S

LPC声码器

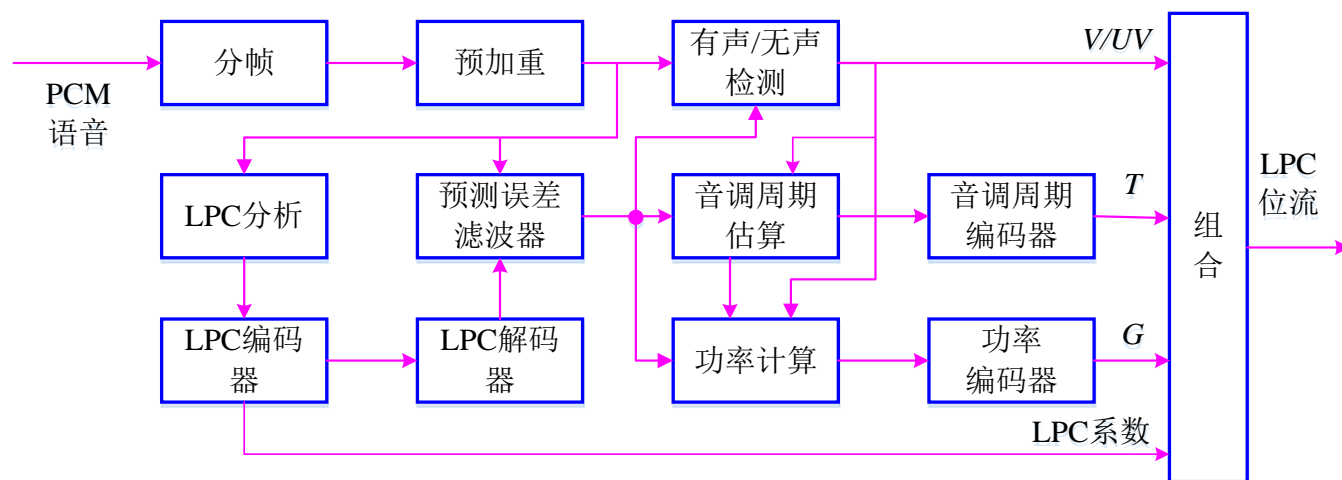
● LPC声码器的原理

- 核心思想：利用语音生成模型来压缩语音数据
- 编码器：从语音信号中抽出表达语音最重要的4个特征参数：声道参数、音调周期、有声/无声和功率(能量)参数
- 解码器：语音生成模型用4个特征参数重构语音



LPC声码器

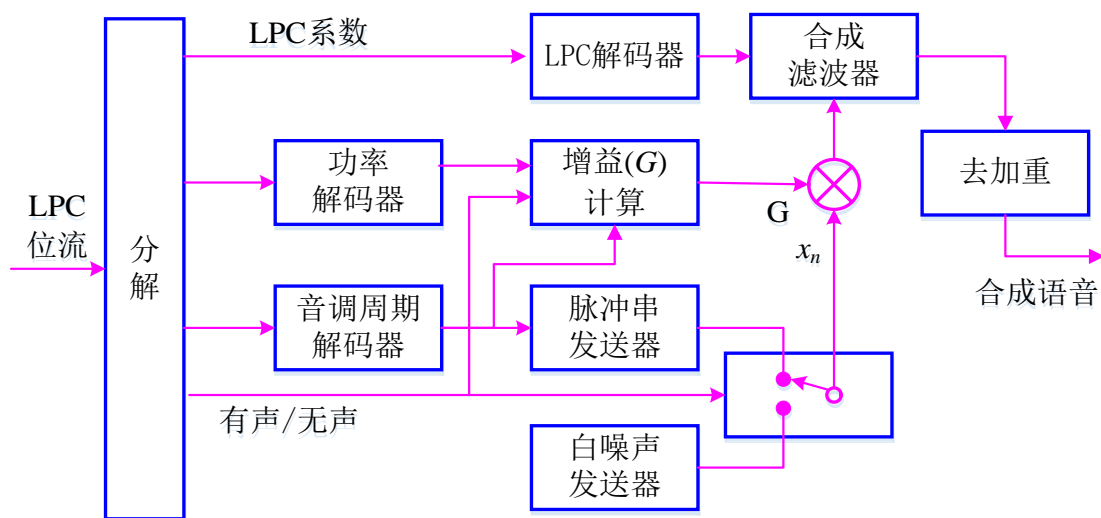
● LPC编码器的结构



(1) 将输入为PCM的语音数据分割成帧，如20 ms/帧；(2) 预加重用于调整信号的频谱特性；(3) 有声/无声检测用于将当前的帧分成有声/无声的帧；(4) 经过预加重的信号通过线性预测分析抽出LPC的系数，经过LPC编码器量化后作为输出；(5) 通过LPC解码器解码的LPC系数构造预测误差滤波器，它输出的预测误差用来估算音调周期和计算一帧的功率(能量/帧)

LPC声码器

● LPC解码器的结构



➤ 核心是语音生成模型

- LPC系数作为合成滤波器的系数，合成滤波器输入是增益出 G 和 x_n (有声/无声)的乘积，滤波器输出去加重后得到合成语音

LPC声码器

- LPC声码器的参数

- LPC声码器要计算的矢量为

$$A = (a_1, a_2, \dots, a_{10}, G, V / UV, T)$$

用于表达一帧的样本值，如160个

- 预测系数的求解见前面的“线性预测编码”

- 余下参数

- 有声/无声
 - 音调周期
 - 功率计算
 - 预测增益

内容提纲

- 声音信号数字化
- 语音编码介绍
- 波形编码
- 参数编码
- 混合编码

混合编码

- 综合使用波形编码和参数编码
 - 波形编码语音质量高 + 参数编码数据速率低
 - 数据率为4.0~16 kbps, 质量可达MOS 分4.0
- 核心思想：改变语音合成滤波器的激励信号，使重构语音波形接近原始语音波形
- 基础技术：AbS(analysis by synthesis)/分析合成法
 - 使用AbS去除语音样本之间和音调周期之间的相关性
 - 使用精心制作的码本生成语音合成器的激励信号
- 混合编码的复杂性高，计算量大