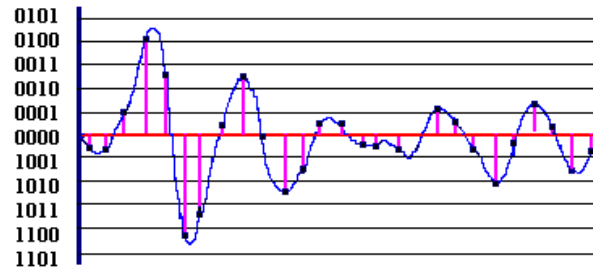


复 习

- 声音的数字化
 - 采样、量化



- 均匀量化、非均匀量化
 - 量化噪声 $SQNR = 20 \log_{10} \frac{V_s}{V_q} (dB)$
- 存储格式：wav文件
- MIDI：The Musical Instrument Digital Interface

本节课内容

- 语音信号编码
 - 1、语音编码概述
 - 2、波形编译码器
 - PCM、DM、ADM、ADPCM的概念
 - 3、混合编译码器
 - LPC、CELP的概念
 - 4、语音编码标准
 - G.721、G.722、G.723、G.726

林福宗： c3：语音编码

Ze-Nian Li： c6.3：音频的量化与传输

c13：音频压缩技术基础

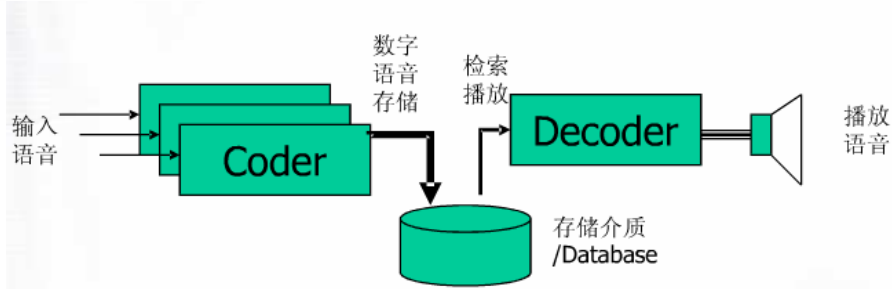
1、语音编码概述

- **语音编码**是数字化语音传输和存储的基础技术，用压缩语音信号的数字表示而使表达这些信号所需的比特数最小。压缩后的语音：
 - **用于传输**，可以降低每路语音所需带宽，在同样的带宽内传输更多路的语音；
 - **用于存储**，可以节约空间，提高存储语音长度，降低成本；
 - **用于网络通信**，可以减少通信成本，提高效率，有利于网络资源的管理。

未压缩音源的大致比特率

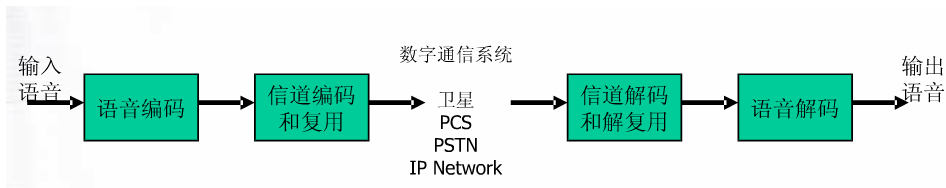
信源	数据流/码率
电话 (200Hz至3.4kHz)	$8000\text{样本/秒} \times 12\text{比特/样本} = 96\text{kbps}$
宽带语音 (50Hz至7kHz)	$16000\text{样本/秒} \times 14\text{比特/样本} = 224\text{kbps}$
宽带音频 (20Hz至20kHz)	$44100\text{样本/秒} \times 2\text{通道} \times 16\text{比特/样本} = 1.35\text{ Mbps}$

语音编码典型应用：数字语音录放系统



- 责任电话语音记录（如调度电话，重要热线电话）：多路实时录音功能和数据库检索功能，要求长时间海量存储
- 功能性语音记录（如数字录音电话，语音信箱，语音电子邮件，多媒体信息查询系统，发声字典等），要求实时录音，存储和实时播放，容量较小，通过适当编码算法减小存储容量以降低成本

语音编码典型应用：实时语音通信系统



- 编解码对称进行，需要实时处理
- 往往需要进行多路复用法和多种媒体的复用传输
- 编解码时延和传输时延需要控制在一定范围内（如小于300ms），否则将严重影响通话感知质量
- 当端到端延时大于几十毫秒，需要采用数字回波抵消器

1、语音编码概述

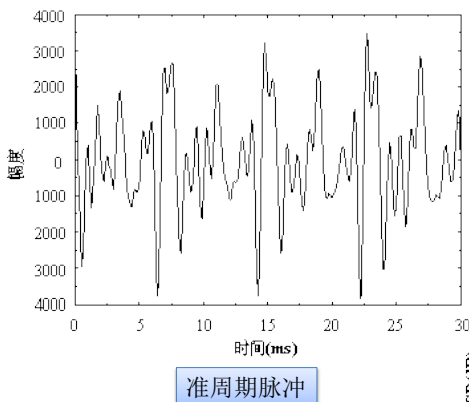
1.1 语音的形成原理

- 肺中的空气受到挤压形成气流，气流通过声门（声带）沿着声道（由咽、喉、口腔等组成）释放出去，就形成了语音。
- 气流、声门可以等效为一个**激励源**，声道可以等效为一个**时变滤波器**（共振峰）。
- 语音信号具有很强的相关性（长期相关、短期相关）。

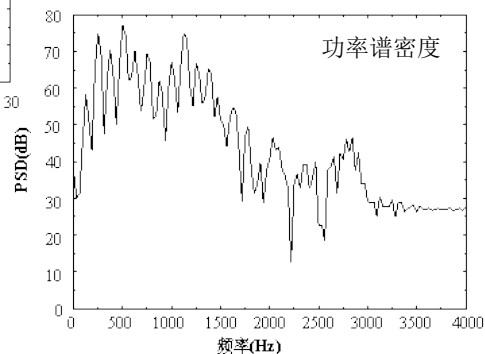


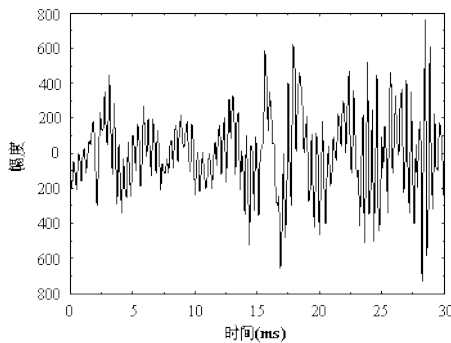
7

1、语音编码概述

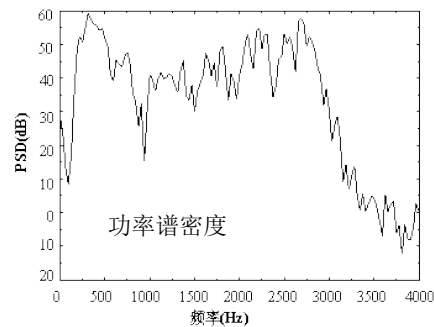


浊音 (voiced sounds) :
声道打开，声带先打开后关闭，在此期间声带要发生振动。浊音的激励源被等效为准周期的脉冲信号。





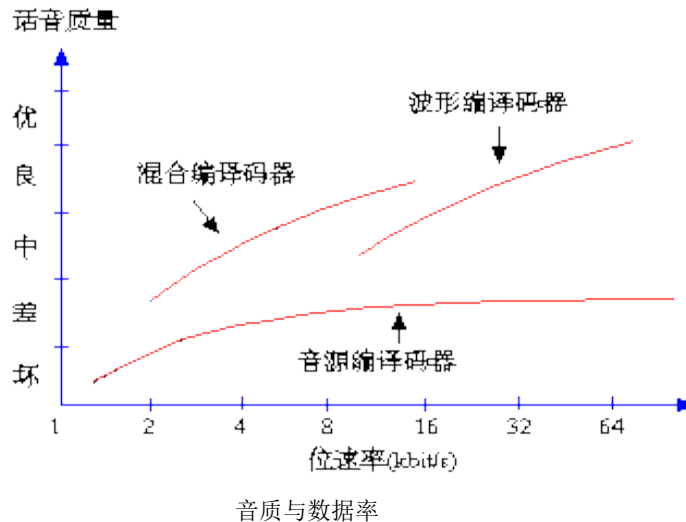
清音 (unvoiced sounds) :
 声门打开，声带不振动，声音靠空气在声道里高速收缩产生。清音的激励源被等效为一种白噪声信号。



- **爆破音 (plosive sounds) :** 声道关闭之后产生压缩空气然后突然打开声道所发出的声音。
- **混合音:** 某些音不能归属到上述三种音中的任何一种，例如在声门振动和声道收缩同时出现的情况下产生的摩擦音，这种音称为混合音。
- 虽然各种各样的话音都有可能产生，但声道的形状和激励方式的变化相对比较慢，因此话音在短时间周期 (20 ms的数量级) 里可以被认为是**准定态** (quasi-stationary) 的，也就是说基本不变的。

1、语音编码概述

1.2 不同类型语音编码器的性能比较



1、语音编码概述

① 波形编译码器

- 不利用生成语音信号的任何知识而企图产生一种重构信号，它的波形与原始语音波形尽可能地一致。
- 一般来说，这种编译码器的复杂程度比较低，数据速率在16 kb/s以上（24~64kbps），质量相当高。低于这个数据速率时，音质急剧下降。抗误码、丢帧、噪声等能力强。（鲁棒性，Robust，好）
- 方法：
 - 基于时域：PCM、ADPCM...
 - 基于频域：
 - 子带编码(sub-band coding, SBC)
 - 自适应变换编码(adaptive transform coding, ATC)。

非线性量化
时域自适应差分编码
自适应量化

如标准：G.711 G.726 G.722

1、语音编码概述

② 音源编译器 (参数编码器)

- 从话音波形信号中提取生成话音的参数，使用这些参数通过话音生成模型重构出话音。
- 数据率在2.4 kb/s左右 (<4.8kbps)，产生的语音虽然可以听懂，但其质量远远低于自然语音。
- 增加数据率对提高合成语音的质量无济于事，这是因为受到话音生成模型的限制。
- 编解码器复杂且不对称，对噪声和误差敏感 (即鲁棒性差)，编码速率低；
- 尽管它的音质比较低，但它的保密性能好，因此这种编译器一直用在军事上。

如标准：MPEG-4中HVXC声码器

LPC: Linear predictive coding

1、语音编码概述

③ 混合编译器

- 填补波形编译码和音源编译码之间的间隔
- 使用不同的激励信号形成不同的编译码器
 - 两个状态 (有声、无声) → LPC linear predictive coding
 - 接近原始话音波形的激励信号- → MPE
 - 使用固定间隔的脉冲，脉冲数目增加→ RPE
 - 使用量化矢量作为激励→ CELP
- 在中低速率上达到64k波形编码音质，鲁棒性较好，算法复杂度高。编码速率一般在4~16kbps。

如标准：G.723.1 G.729 (常用于互联网上)

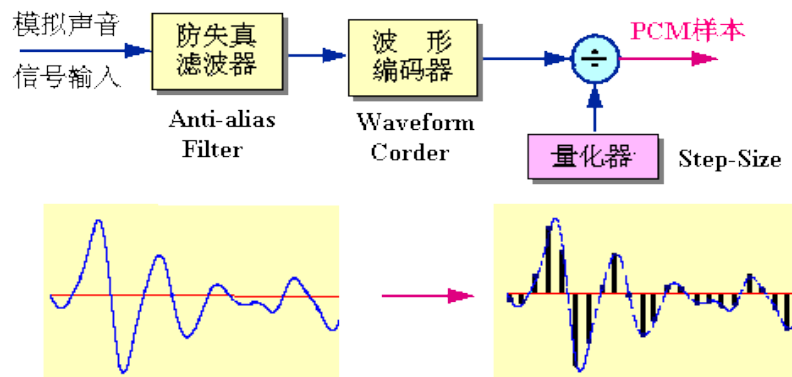
本节课内容

• 语音信号编码

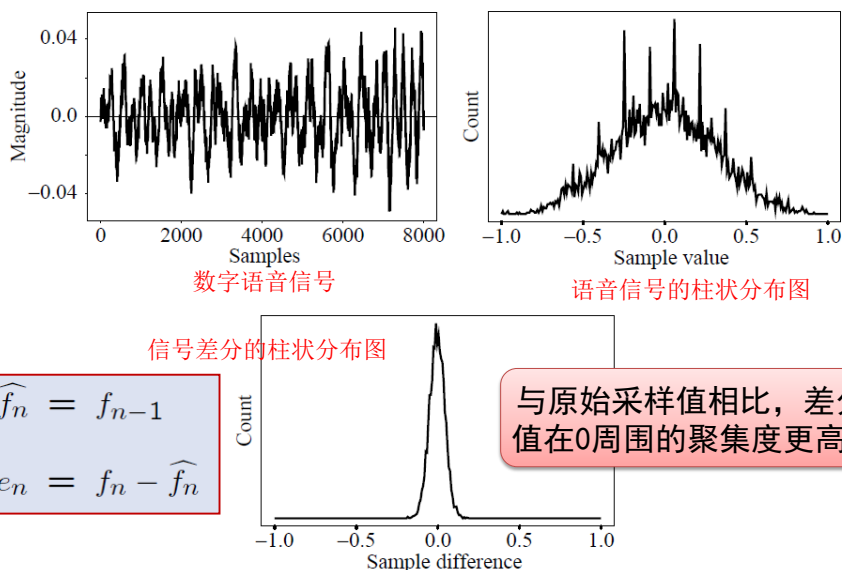
- 1、语音编码概述
- 2、波形编译码器
 - PCM、DM、ADM、ADPCM的概念
- 3、混合编译码器
 - LPC、CELP的概念
- 4、语音编码标准
 - G.721、G.722、 G.723 、 G.726

2、波形编译码器

- 2.1 PCM
- 脉冲编码调制(pulse code modulation , PCM)



2.2 无损预测编码



2.2 无损预测编码举例

- 假设预测公式如下：

$$\hat{f}_n = \lfloor \frac{1}{2}(f_{n-1} + f_{n-2}) \rfloor$$

$$e_n = f_n - \hat{f}_n$$

实际传输的是误差值 e_n

注意：假设原信号范围0~255，差分值的范围则是-255~255。我们可以定义SU/SD代表增加/减少32。那么就可以为一组有限的信号差值-15~16定义码字。如果差值不在该范围内，则用SU/SD加上一个-15~16范围内的值来代替。

$$f_1, f_2, f_3, f_4, f_5 = 21, 22, 27, 25, 22.$$

- 假设 $f_0 = f_1$, 并且第一个值不编码, $e_1 = 0$ 。那么

$$\hat{f}_2 = 21, e_2 = 22 - 21 = 1;$$

$$\hat{f}_3 = \lfloor \frac{1}{2}(f_2 + f_1) \rfloor = \lfloor \frac{1}{2}(22 + 21) \rfloor = 21, \\ e_3 = 27 - 21 = 6;$$

$$\hat{f}_4 = \lfloor \frac{1}{2}(f_3 + f_2) \rfloor = \lfloor \frac{1}{2}(27 + 22) \rfloor = 24, \\ e_4 = 25 - 24 = 1;$$

$$\hat{f}_5 = \lfloor \frac{1}{2}(f_4 + f_3) \rfloor = \lfloor \frac{1}{2}(25 + 27) \rfloor = 26,$$

$$e_5 = 22 - 26 = -4$$

误差值集中在0的附近,
编码会有更好的效率。

2.3 DPCM (差分脉冲编码调制)

- DPCM与无损预测压缩类似, 只是它含有一个**量化步骤**。
 - 量化步骤和PCM中的量化类似, 可以是均匀量化, 也可以是非均匀量化。有一种Lloyd-Max量化, 它是基于误差最小方差的。

- 形式化描述:

$$\hat{f}_n = \text{function_of}(\tilde{f}_{n-1}, \tilde{f}_{n-2}, \tilde{f}_{n-3}, \dots),$$

预测信号

$$e_n = f_n - \hat{f}_n,$$

$$\tilde{e}_n = Q[e_n],$$

原始信号

$$\text{transmit } \text{codeword}(\tilde{e}_n),$$

$$\text{reconstruct: } \tilde{f}_n = \hat{f}_n + \tilde{e}_n.$$

量化的重构信号

量化的重构信号

DPCM举例

- 预测器:

$$\hat{f}_n = \text{trunc}((\tilde{f}_{n-1} + \tilde{f}_{n-2})/2)$$

- 量化器:

$$\tilde{e}_n = Q[e_n] = 16 * \text{trunc}((255 + e_n)/16) - 256 + 8$$

$$\tilde{f}_n = \hat{f}_n + \tilde{e}_n$$

量化器把误差值平均划分为32块，每一块在原始信号中跨越16个取值范围

量化器重构级别

e_n in range	Quantized to value
-255 .. -240	-248
-239 .. -224	-232
⋮	⋮
⋮	⋮
-31 .. -16	-24
-15 .. 0	-8
1 .. 16	8
17 .. 32	24
⋮	⋮
⋮	⋮
225 .. 240	232
241 .. 255	248

$$\tilde{e}_n = Q[e_n] = 16 * \text{trunc}((255 + e_n)/16) - 256 + 8$$

- 举例原始信号

f_1 f_2 f_3 f_4 f_5
 130 150 140 200 230 .

- 假设 f_0 取值和 f_1 一样。初始误差为0。

预测值 \hat{f} =	130	,	130,	142,	144,	167
误差 e =	0	,	20,	-2,	56,	63
量化后误差 \tilde{e} =	0	,	24,	-8,	56,	56
重构值 \tilde{f} =	130	,	154,	134,	200,	223

实际误差为 0 -4 6 0 7

$$\tilde{e}_n = Q[e_n] = 16 * \text{trunc}((255 + 8 + e_n)/16) - 256$$

e_n in range	Quantized to value		e	eq
-255 .. -240	-248	
-239 .. -224	-232		-55~-40	-40
.	.		-39~-24	-32
.	.		-23~-8	-16
-31 .. -16	-24		-7~8	0
-15 .. 0	-8		9~24	16
1 .. 16	8		25~40	32
17 .. 32	24		41~56	40
.
.	.			
.	.			
225 .. 240	232			
241 .. 255	248			



$$\tilde{e}_n = Q[e_n] = 16 * \text{trunc}((255 + e_n)/16) - 256 + 8$$

- 举例原始信号

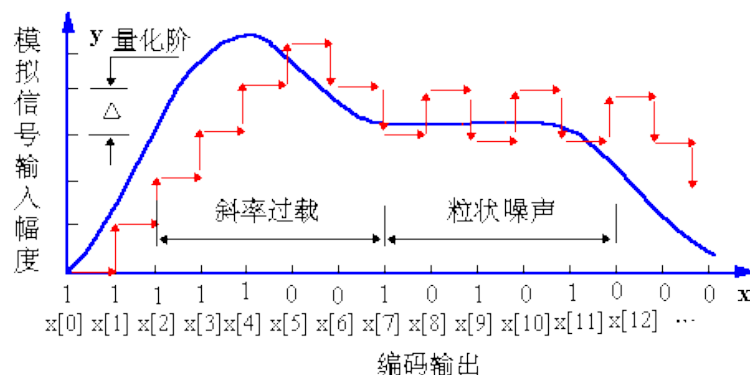
f_1	f_2	f_3	f_4	f_5
130	150	140	200	230

- 假设 f_0 取值和 f_1 一样。初始误差为0。

预测值 \hat{f}	=	130	, 130,	138,	142,	223
误差 e	=	0	, 20,	2,	58,	7
量化后误差 \tilde{e}	=	0	, 16,	0,	64,	0
重构值 \tilde{f}	=	130	, 146,	138,	208,	223
实际误差		0,	4,	2,	-8,	7

2.4 增量调制DM

- 增量调制也称 Δ 调制(delta modulation, DM), 它是一种预测编码技术, 是PCM编码的一种变形。



2.4 DM增量调制

- DM是DPCM的简化版本，常常作为一个快速模数转换器。
- 描述如下：

$$\begin{aligned} \text{预测值: } \hat{f}_n &= \tilde{f}_{n-1}, \\ \text{误差: } e_n &= f_n - \hat{f}_n = f_n - \tilde{f}_{n-1}, \end{aligned}$$

$$\text{量化后误差: } \tilde{e}_n = \begin{cases} +k & \text{if } e_n > 0, \text{ where } k \text{ is a constant} \\ -k & \text{otherwise} \end{cases}$$

$$\text{重构值: } \tilde{f}_n = \hat{f}_n + \tilde{e}_n.$$

DM举例

如果信号波形非常陡峭或者平缓，自动调整k的取值，即为自适应增量调制(ADM)

- 考虑时间信号：

$$\begin{array}{cccc} f_1 & f_2 & f_3 & f_4 \\ 10 & 11 & 13 & 15 \end{array}.$$

假设重建信号 $\tilde{f}_1 = f_1 = 10$.

如果步长k=4，那么

$$\begin{array}{llll} \hat{f}_2 = 10, & e_2 = 11 - 10 = 1, & \tilde{e}_2 = 4, & \tilde{f}_2 = 10 + 4 = 14 \\ \hat{f}_3 = 14, & e_3 = 13 - 14 = -1, & \tilde{e}_3 = -4, & \tilde{f}_3 = 14 - 4 = 10 \\ \hat{f}_4 = 10, & e_4 = 15 - 10 = 5, & \tilde{e}_4 = 4, & \tilde{f}_4 = 10 + 4 = 14. \end{array}$$

预测值

误差值

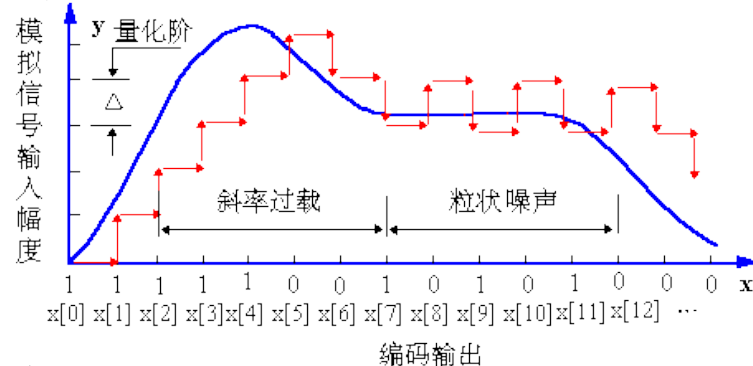
量化后误差值

重构信号

实际误差为：0 3 -3 1

2.5 自适应增量调制ADM

- 增量调制

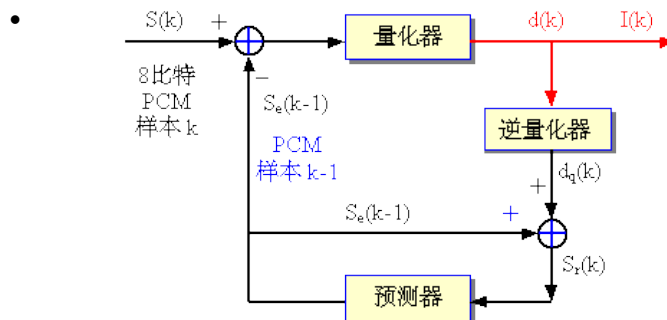


- **自适应增量调制(ADM) :**

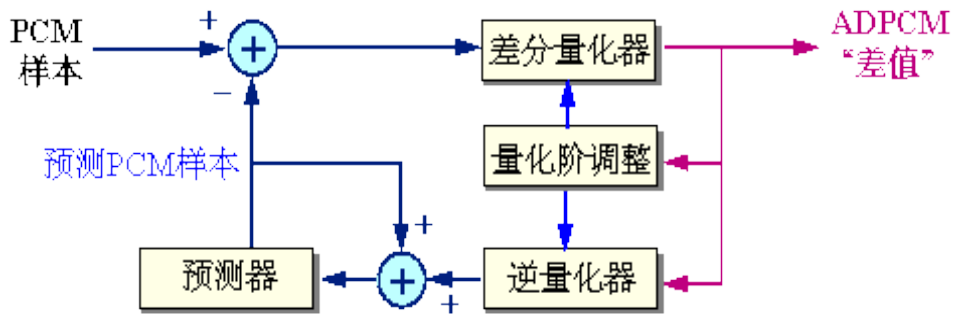
- 根据输入信号斜率的变化自动调整量化阶 Δ 的大小，以使斜率过载和粒状噪声都减到最小：检测到斜率过载时开始增大量化阶 Δ ，而在输入信号的斜率减小时降低量化阶 Δ

2.6 自适应差分脉冲编码调制 (ADPCM)

- **差分脉冲编码调制DPCM**(differential pulse code modulation)是利用样本与样本之间存在的信息冗余度来进行编码的一种数据压缩技术。
- **自适应脉冲编码调制**(adaptive pulse code modulation , APCM)是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。



- ADPCM(adaptive difference pulse code modulation)
- 综合了APCM的自适应特性和DPCM系统的差分特性，是一种性能比较好的波形编码。



本节课内容

- 语音信号编码
 - 1、语音编码概述
 - 2、波形编译码器
 - PCM、DM、ADM、ADPCM的概念
 - 3、混合编译码器
 - LPC、CELP的概念
 - 4、语音编码标准
 - G.721、G.722、G.723、G.726

3、混合编译码器

• 3.1 LPC (线性预测编码)

- 使用过去的P个样本值来预测现时刻的采样值 $x(n)$:

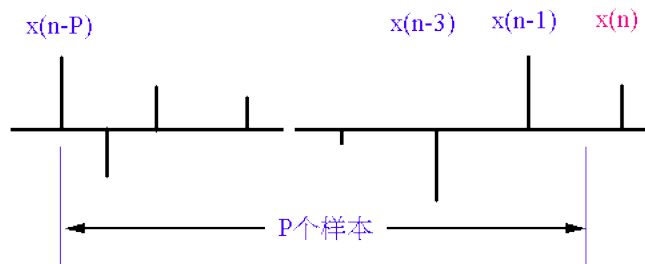
$$x_{pre}(n) = -[a_1x(n-1)+a_2x(n-2)+\dots+a_px(n-p)] = -\sum_{i=1}^p a_i x(n-i)$$

为方便起见，式中采用了负号。

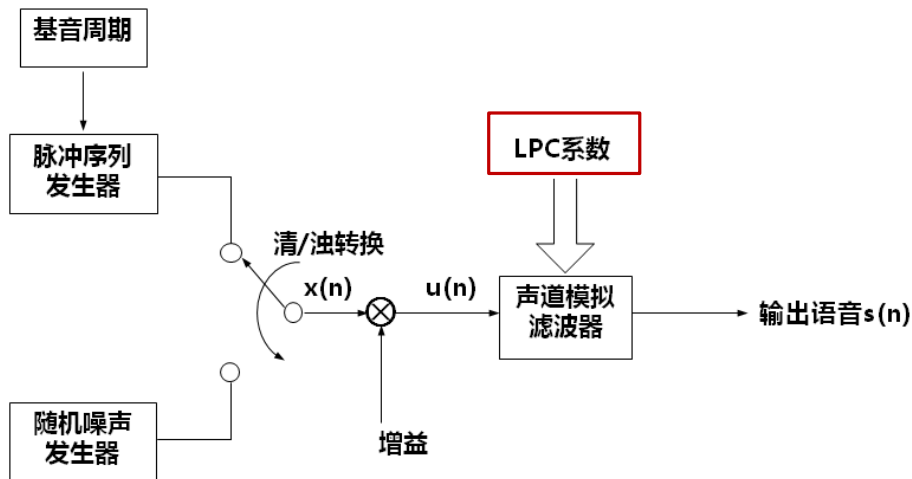
- 残差误差(residual error)即线性预测误差为

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

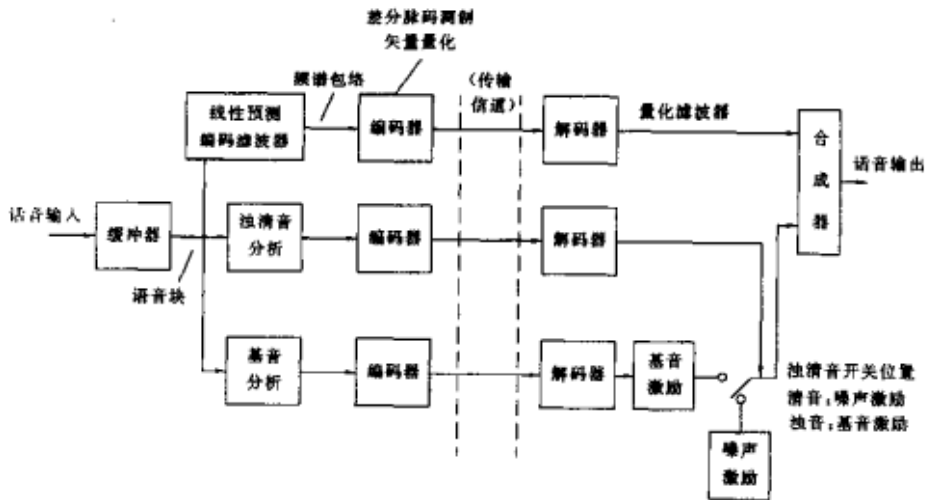
这是一个线性差分方程。



LPC 语音合成图



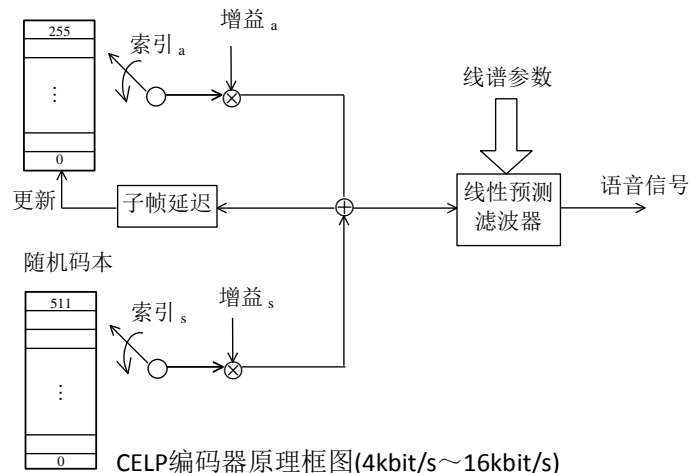
LPC完整原理图



3.2 CELP的概念

- CELP(Code Excited Linear Prediction , 码激励线性预测编码)

自适应码本



CELP编码器原理框图(4kbit/s~16kbit/s)

本节课内容

• 语音信号编码

- 1、语音编码概述
- 2、波形编译码器
 - PCM、DM、ADM、ADPCM的概念
- 3、混合编译码器
 - LPC、CELP的概念
- 4、语音编码标准
 - G.721、G.722、 G.723 、G.726

4、语音编码标准

ITU-T Telephony Speech Coding (F.700's A0 Audio Quality Level)

电话语音编码

- G.711 *PCM coding* (64 kbit/s) *late 60's*
- G.726 *ADPCM coding* (32; 40, 24 & 16 kbit/s) *1988*
- G.728 *LD-CELP coding* (16; 40, 11.8 & 9.6 kbit/s) *1992*
- G.723.1 *Dual-rate coding* (5.3 & 6.3 kbit/s) *1995*
- G.729 *CS-ACELP coding* (8; 12.8 & 6.4 kbit/s) *1996-2000*
- G.xxx *4kbit/s* (Ongoing)
- G.VBR (Variable bit rate)(New)

4、话音编码标准

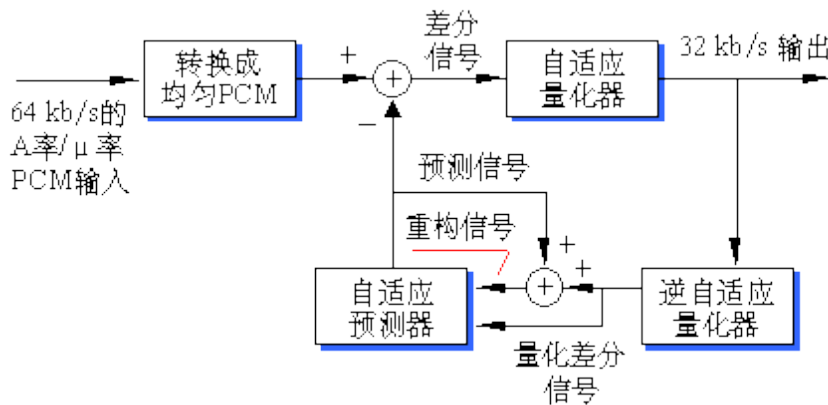
ITU-T Wideband Speech Coding (F.700's A1 Audio Quality Level)

宽带话音编码

- G.722
 - Coding of 7 kHz speech at 64, 56, and 48 kbit/s
 - Sub-band ADPCM
- G.722.1
 - Coding of 7 kHz speech at 32 and 24 kbit/s
 - Transform coding approach
- G.722.2 (Just completed)
 - Coding of 7 kHz speech at 16 kbit/s or lower
 - CELP-based; same as 3GPP AMR-WB
 - Optimized for speech, works well also with 7kHz music

4、话音编码标准

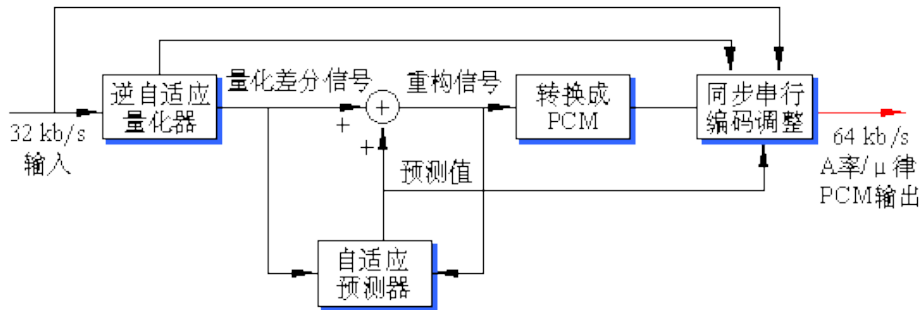
- G.721 : 32 kb/s自适应差分脉冲编码调制



G.721 ADPCM编码

4、话音编码标准

- G.721 : 32 kb/s自适应差分脉冲编码调制

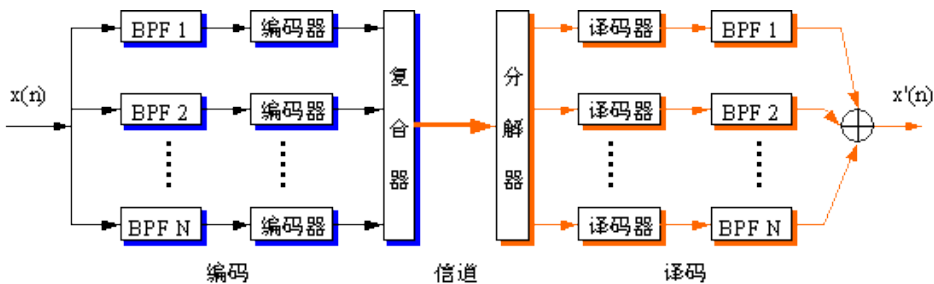


G.721 ADPCM译码器

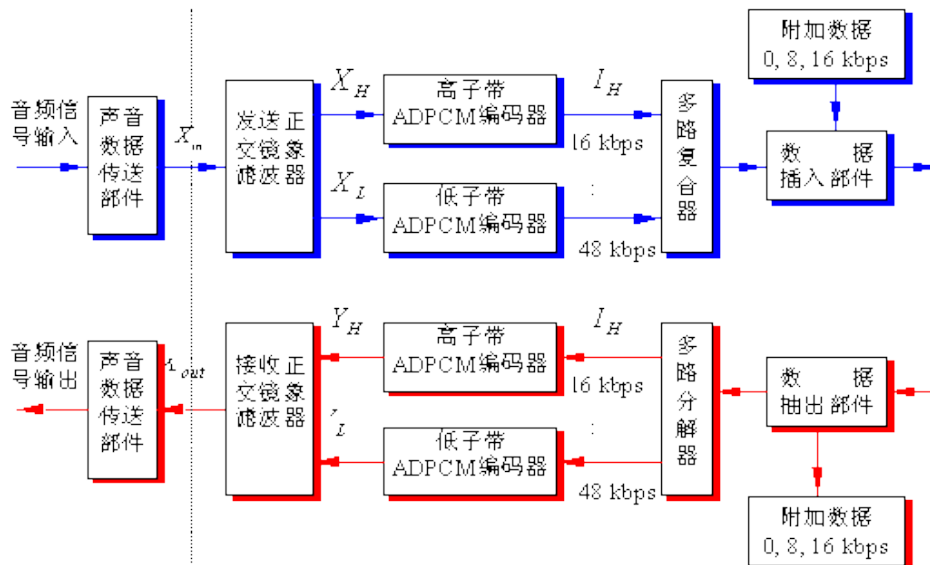
4、话音编码标准

- G.722 SB-ADPCM : 数据率为64 kb/s的7 kHz声音信号编码

宽带话音编码



4、话音编码标准



4、话音编码标准

- G.723语音编码器是一种用于多媒体通信，编码速率为5.3kb/s和6.3kb/s的双码率编码方案。可以应用于IP电话等系统中。其中，5.3kb/s码率编码器采用多脉冲最大似然量化技术（MP - MLQ），6.3kb/s码率编码器采用代数码激励线性预测技术。
- G.726是ITU前身CCITT于1990年在G.721和G.723标准的基础上提出的关于把64kbps非线性PCM信号转换为40kbps、32kbps、24kbps、16kbps的ADPCM信号的标准。