视听觉信号处理

Visual and Auditory Signal Processing

哈尔滨工业大学 Harbin Institute of Technology

语音编码

哈尔滨工业大学 Harbin Institute of Technology

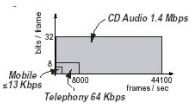
为什么要进行编码?

- 语音数据有多大?

--fs：每秒钟样本数（8k~44.1k）

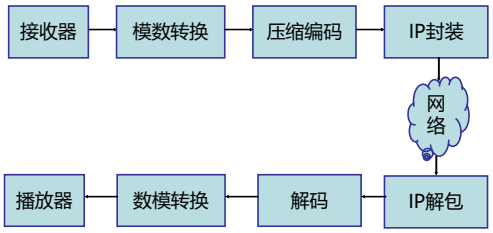
*c: 每个样本的通道数（1 or 2）

*b：每个样本点的位数（8 or 16）



- 比特率：fs*c*b，bit/s 或bps
- 压缩语音信号的传输带宽，降低信道的传输速率
- 语音编码就是使比特率（编码速率）尽可能小

语音编码应用实例（VoIP）



为什么语音是可以压缩的?

1. 存在冗余度：

（1）幅度非均匀分布

（2）语音信号样本间的相关性很强

（3）浊音具有准周期

（4）声道的形状及其变化缓慢

（5）语音间隙

为什么语音是可以压缩的?

2.人的听觉感知机理

（1）人类的听觉特性具有掩蔽效应

（2）人耳对不同频段声音的敏感程度不同

（3）人耳对语音相位不敏感

语音编码的类型：

---波形编码

---参数编码

---混合编码

波形编码：对采样值进行编码压缩，解码后的语音信号基本上与输入语音信号波形相同。

- 编码速率较高：9.6k~64kbit/s
- 包括：PCM、压扩PCM、ADPCM、DM、ADM、SBC等
- 适应能力强、语音质量好；编码速率高

参数编码：基于人类语音的产生机理建立数学模型，根据输入语音得出模型参数并传输，在收端恢复，重建的语音信号与原始信号样本之间没有一一对应关系，但内容相同。

- 编码速率较低：2.4k~4.8kbit/s
- 包括各线性预测编码（LPC）方法和余弦声码器等
- 编码速率低；语音质量差、自然度低、对环境噪声敏感

混合编码：波形编码+参数编码，在参数编码的框架下使波形尽可能相同。

- 编码速率较低：16k~2.4kbit/s
- 包括多脉冲激励线性预测编码(MPLPC)、规则脉冲激励线性预测编码(RPE-LPC)、码本激励线性预测编码(CELP)

在语音通信中，语音质量分为以下四等：

①广播质量：宽带，语音质量高，感觉不出噪声存在。

②长途电话质量：指通过电话网传输后得到的语音质量，信噪比大于30dB，谐波失真小于2%-3%。

③通信质量：可以听懂，但和长途电话质量相比，显有较大失真。

④合成质量：80%-90%可懂度，，听起来像机器说话，失去了讲话者的个人特征。

波形失真度

$$SNR = 10 * \log \left\{ \frac{\sum_{n=0}^{M} (s(n))^2}{\sum_{n=0}^{M} (s(n) - \hat{s}(n))^2} \right\}$$

已经标准化的语音编码

指定组织：国际电信联盟 ITU-T， （http://www.itu.int）

标准	编码类型	比特率 (kbps)	MOS	复杂性	时延 (ms)
G.711	PCM	64	4.3	1	0.125
G.726	ADPCM	32	4.0	10	0.125
G.728	LD-CELP	16	4.0	50	0.625
GSM	RPE_LPT	13	3.7	5	20
G.729	CSA-CELP	8	4.0	30	15
G.729A					15
G.723.1	ACELP	6.3	3.8	25	37.5
	MP-MLQ	5.3			
US Dod FS1015	LPC-10	2.4	合成语音	10	22.5

最具代表性的波形编码

(1)ADPCM：自适应差分脉冲编码 adaptive difference pulse code modulation

(2)CELP：码本激励线性预测 (code excited linear prediction)

(3)ACELP：代数码本激励线性预测 Algebraic-Code-Excited Linear-Prediction

(4)CS-ACELP：共轭结构的代数码本激励线性预测 Conjugate Structure Algebraic-Code-Excited Linear-Prediction

(5)RPE-LTP：长时预测的规则脉冲激励的线性预测 Regular-Pulse Excited LPC with a Long-Term Predictor

最具代表性的混合编码

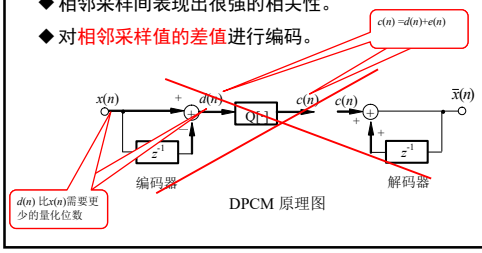
波形编码之ADPCM

DPCM

◆ 在PCM中，各采样值都独立编码，需要较多位数，比特率较高。

◆ 相邻采样间表现出很强的相关性。

◆ 对相邻采样值的差值进行编码。



DPCM 原理图

DPCM

分析存在的问题。用z变换考察各点信号的时域关系，有：

$$C(z)=X(z)(1-z^{-1})+E(z)$$

$$\bar{X}(z)=\frac{C(z)}{1-z^{-1}}=X(z)+\frac{E(z)}{1-z^{-1}}$$

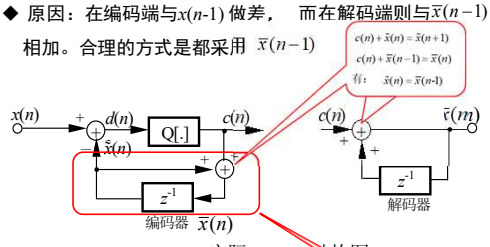
其中$E(z)$为量化器量化噪声$e(n)$的z变换。有：

$$\bar{x}(n)=x(n)+\sum_{m=1}^{n}e(m)$$

可以看出，量化器所产生的量化噪声被累积叠加到了输出信号中。

DPCM

◆ 原因：在编码端与$x(n\text{-}1)$做差， 而在解码端则与$\bar{x}(n-1)$相加。合理的方式是都采用 $\bar{x}(n-1)$



实际 DPCM 结构图

◆ 在编码器中包含解码器

解码器

DPCM

$$\tilde{X}(z)=\frac{C(z)z^{-1}}{1-z^{-1}}$$

$$C(z)=X(z)-\tilde{X}(z)+E(z)$$

$$C(z)=(X(z)+E(z))(1-z^{-1})$$

$$\bar{X}(z)=\frac{C(z)}{1-z^{-1}}=X(z)+E(z)$$

有： $\bar{x}(n)=x(n)+e(n)$

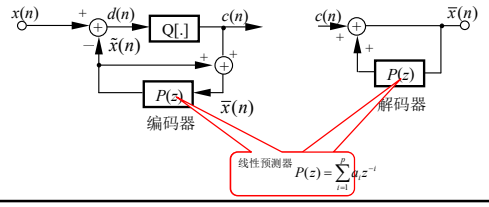
可以看出，已经消除了量化噪声的累积。

ADPCM

◆ 仅仅利用相邻的两个采样点之间的相关性，其差可能不够小，仍需要较多的量化bit数。

◆ 用线性预测来刻画更多采样间的相关性。

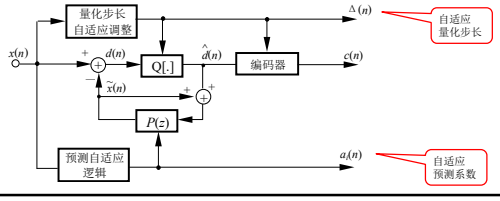
◆ 对采样值与预测值间的差值进行编码。



ADPCM

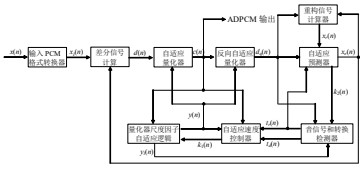
◆ 最佳线性预测器系数每帧(段)不同，需要每帧(段)传送一次。每帧(段)可以采用不同的量化步长(位数)，被称之边信息。

◆ 每帧(段)的边信息根据预测残差自适应确定。



ADPCM

- ADPCM已形成国际标准，ITU-T（原CCITT）在1988年制定了G.726标准，将1984年和1986年分别制定的ADPCM标准G.721和G.723进行了合并，同时也删除了上述两个标准。
- G.726能提供四种数码率：40kbit /s、32kbit/s、24kbit/s、16kbit/s。其语音质量相当于64kbit/s的PCM编码，并具有很好的抗误码性能 .

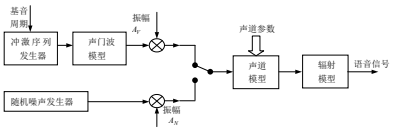


G.726 编码器方框图

参数编码之LPC-10

LPC编码

- 完全基于语音信号的产生模型。
- 在编码端计算模型参数，作为编码传输，在解码端基于该模型参数合成语音。
- 解码后语音波形一般都会发生改变。

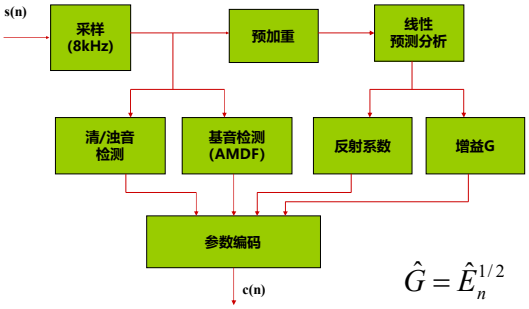


语音信号产生系统模型

LPC编码

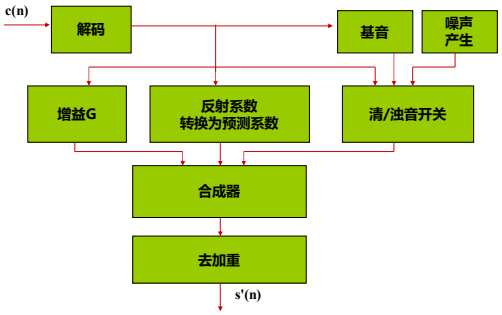
- 美国确定LPC-10作为2.4kb/s速率上的推荐编码形式，用于第三代保密电话中。
- 在其发送端，原始语音信号采用8kHz采样，然后每180个采样值分为一帧（22.5ms），提取语音特征参数并加以编码传送。每帧总共编码为54bits，每秒传输44.4帧，因此总传输速率为2.4kb/s。
- 其增强版本为LPC-10e

LPC-10编码器发送端



$$\hat{G} = \hat{E}_n^{1/2}$$

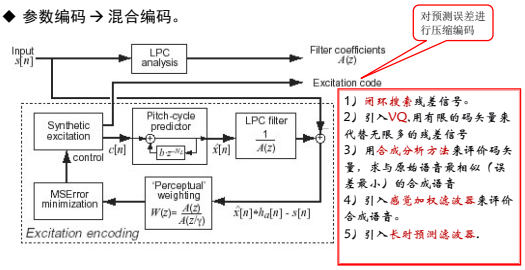
LPC-10编码器接收端



混合编码之CELP

混合编码

- LPC-10语音质量是“合成语音”级别。
- 对声门波的假定并不符合实际。（声门波=预测误差？）
- 参数编码→混合编码。



混合编码

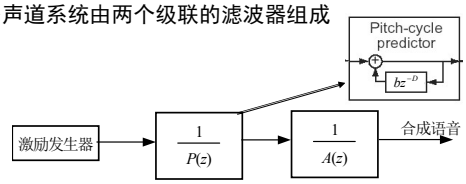
- 预测误差信号仍有较强的相关性，可以用长时预测去相关，使其更平坦。
- 对应声门波要经过一个长时预测综合滤波器，表示语音信号长时相关性的模型。它的一般形式为：

$$1/P(z)=1/[1-\sum_{i=-q}^{r}b_i z^{-(D+i)}]$$

- 其中延时参数D等于基音周期，$\{b_i\}$是语音信号的长时预测系数
- 预测系数的个数取1（q=r=0）或3(q=r=1)

混合编码

◆声道系统由两个级联的滤波器组成



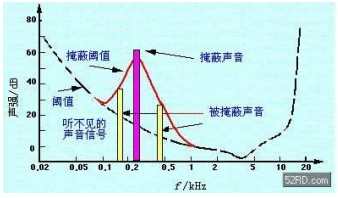
- 两种相关性：样本点之间的短时相关性和相邻基音周期之间的长时相关性。
- 对这两种相关性进行去相关后，可以得到更加平坦的预测残差信号，因而更加有利于进行量化编码。

混合编码

感觉加权滤波器

- 语音质量与信噪比等价么？
- 掩蔽效应（频域）

'Perceptual' weighting
$W(z)=\frac{A(z)}{A(z/\gamma)}$



混合编码

感觉加权滤波器的传递函数

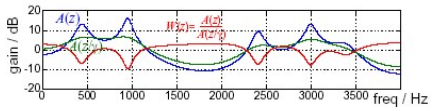
预测误差滤波器

$$W(z)=\frac{A(z)}{A(z/\gamma)}=\frac{1-\sum_{i=1}^{p}a_i z^{-i}}{1-\sum_{i=1}^{p}a_i\gamma^i z^{-i}}$$

加权因子γ取值在0~1之间

混合编码

感觉加权滤波器的频谱

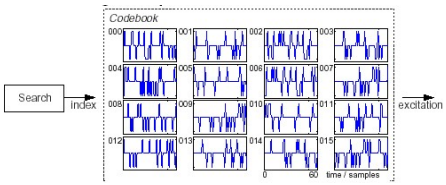


- 由于掩蔽效应，在语音频谱中，能量较高的频段（共振峰处）的噪声相对于能量较低的频段的噪声不易被感觉。在度量原始语音和合成语音之间的误差时，在高能量段应允许误差大。

CELP

- CELP是近10年来最成功的语音编码算法。
- CELP语音编码算法，用一个包含许多典型的激励矢量的码本作为激励参数，每次编码时都在这个码本中搜索一个最佳的激励矢量。
- 这个激励矢量的编码值就是这个序列的码本中的序号。
- 码本的获得：LBG算法，双重矢量量化

CELP



- 以码本激励线性预测（CELP）原理为基础的G.729、G.729可将经过采样的64kb/s话音以几乎不失真的质量压缩至8kb/s。
- G.723.1有两种编码速率：6.3kb/s和5.4kb/s

语音编码的评测方法

编码速率

降低编码速率往往是语音编码的首要目标。

分成两类：固定速率编码器和可变速率编码器。

❑ 固定编码速率

现有大部分编码标准都是固定速率编码，其范围为0.8 kbit/s ~64kbit/s。

语音编码的评测方法

❑ 可变编码速率

可变速率编码是近年来出现的新技术。两方通话大约只有40%的时间是真正有声音的，因此可采用通/断二状态编码。可变速率编码主要包括两个算法。一是有声检测，主要用于确定输入信号是语音还是背景噪声 。二是舒适噪声生成，主要用于接收端重建背景噪声，其设计必需保证发送端和接收端的同步。

语音编码的评测方法

❑ 顽健性

通过取多种不同来源的语音信号进行编码解码，并对输出语音质量进行比较测试得到的一种指标。

对存在部分数据丢失的情况下，语音编码器顽健性的研究也有重要的意义。

语音编码的评测方法

- 时延

（1）算法时延。等于帧长和前视长度之和，其值完全取决于算法，与具体的实现无关。

（2）计算时延。即编码器分析时间和解码器的重建时间，其值取决于硬件速度。

（3）复用时延。编码器发送之前和解码器解码之前，必需将整个数据块的所有比特装配好。

（4）传输时延。取决于是采用专用线还是共享信道。

语音编码的评测方法

- 时延

单向时延大于150ms就可感受到通话连续性受到影响，最大可容忍时延为400ms～500ms，超过此值只能进行半双工通信。

对于具有回声的情况，单向时延不能超过25ms，否则就需要装备回声抑制功能。

语音编码的评测方法

- 计算复杂度和算法的可扩展性

计算复杂度主要影响硬件实现的成本。算法的可扩展性是指一种编码算法不仅能解决当前的实际应用，而且可以兼顾将来的发展。

语音编码的评测方法

- 语音质量及其评价方法

用于评价输出语音质量的方法可分为主观和客观两种。语音主观评价方法种类很多，其中又可分为音质评价和可懂度评价两类。

可懂度评价方法有：

（1）判断韵字测试

（2）改进的韵字测试

音质的评价方法有：

（1）平均意见得分

（2）判断满意度测量

语音编码的评测方法

- 语音质量及其评价方法

目前所用的客观测度方法可以分为时域测度、频域测度和其它测度三类方法。

（1）时域测度：信噪比和分段信噪比等；

（2）频域测度：对数谱距离测度、LPC倒谱距离测度

还有在此二者的基础上发展起来的其它测度方法。