



第五章 语音编码

- 概述
- 压缩编码原理
- 波形编码
- 参数编码
- 极低速率编码
- 编码标准



概述

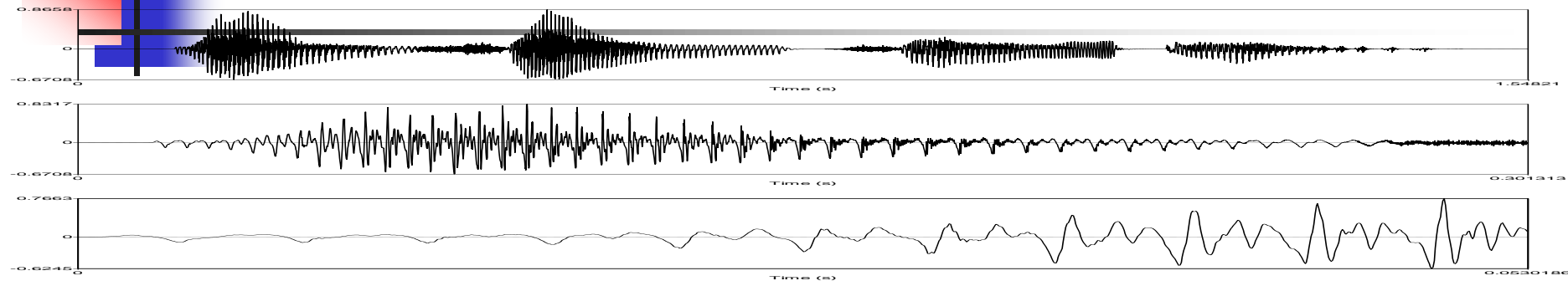
- 应用背景，数字网络及多媒体内容
 - 通信、存储、传播
- 性能要求，在可接受的失真下追求最小的数码率
 - 指标：编码速率、鲁棒性、时延、计算复杂度
 - 评价：语音可懂度及语音质量MOS



压缩编码原理

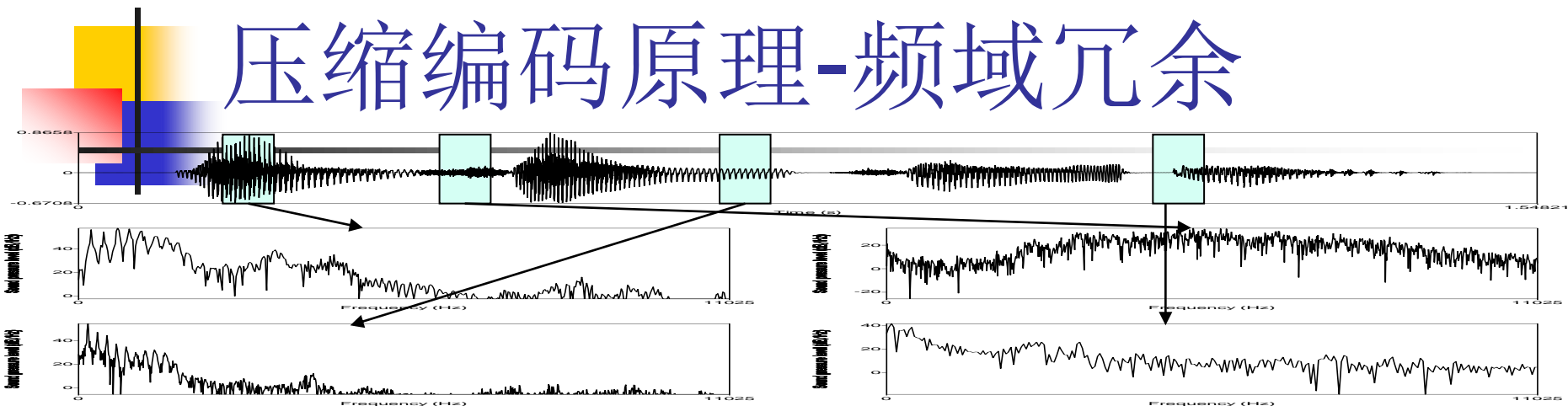
- 实例：普通电话8kHz采样率，12比特量化，则数码率为96Kpbs
- 基本思路：语音信号的冗余度和听觉感知机理出发，去除‘多余’成分
- 语音信号的冗余
 - 时域、频域
- 听觉感知
- 信息论——压缩的潜力

压缩编码原理-时域冗余



- 幅度非均匀，小幅度频度高
- 相关性强
- 浊音段准周期性
- 声道变化缓慢
- 对话中的停顿

压缩编码原理-频域冗余

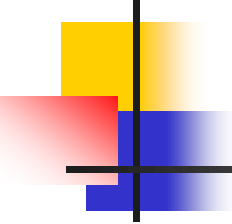


- 非均匀的功率谱密度分布
- 功率谱的结构
谱包络与谐波结构 对应于 共振峰+基频



压缩编码原理-听觉感知

- 掩蔽效应（Masking Effect）。
 - 去除被掩蔽信号成分
 - 抑制/掩盖量化噪声
- 不同频率段的感知灵敏度不同。不同频段分配不同的比特数
- 相位变化不敏感。对于相位不分/少分比特数



压缩编码原理-压缩的潜力

- 信息论角度
 - 256个音素，10个音素/秒，则有信息率 $I = [\log_2(256)^{10}] = 80\text{bps}$
- 标准编码速率64Kbps， 10^3 量级的压缩潜力！



压缩编码原理-主观质量评价

- 清晰度（Intelligibility）
 - 可懂程度
 - 方法：MRT（Modified Rhyme Test，押韵测试）
- 语音质量（Quality）
 - 自然程度
 - 方法：MOS（Mean Opinion Score）
- 测试条件
 - 发音人人数
 - 发音数据



压缩编码原理-客观质量评价

- 客观测度，与主观评价结果的相关性

- 时域测度

- 信噪比SNR (Signal Noise Rate)

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

- 频域测度

- 基于听觉特性失真测度：感知谱、对数谱距离、**Itakura**测距

- 近期出现的客观测度指标

- PESQ (Perceptual Evaluation of Speech Quality)语音质量感知评价

- STOI (Short-Time Objective Intelligibility)短时客观可懂度



波形编码

- 均匀量化PCM(Pulse-Code Modulation)
- 非均匀量化PCM
- 自适应量化PCM



波形编码-均匀量化PCM

- A/D转换直接得到
 - 数码率高
- 量化信噪比 $\text{SNR}(\text{dB})=6.02B-7.2$



波形编码-非均匀量化PCM

- 问题：对于PCM，信号幅度小时，SNR变坏
- 解决方法：通过非线性变换，使语音信号的幅度变化范围变小，然后再量化
 - 依据：掩蔽效应
 - 方案：u律和A律



u律算法

■ u律压缩

$$F_{\mu}[x(n)] = X_{\max} \frac{\ln \left[1 + \mu \frac{|x(n)|}{X_{\max}} \right]}{\ln(1 + \mu)} \operatorname{sgn}[x(n)]$$

- u越大，压缩率越大
- u=255，7位与12位相当



u律算法

- 预处理：原编码加33偏移（0x10001）
 - 方便硬件实现
 - 数值范围：(0 - 8158) 偏移后 (33 - 8191)
- 查表
 - 保留4位有效数字，7位完成了对12~13位的编码

| * Biased Linear Input Code | Compressed Code |
|----------------------------|-----------------|
| * ----- | ----- |
| * 00000001wxyza | 000wxyz |
| * 0000001wxyzab | 001wxyz |
| * 000001wxyzabc | 010wxyz |
| * 00001wxyzabcd | 011wxyz |
| * 0001wxyzabcde | 100wxyz |
| * 001wxyzabcdef | 101wxyz |
| * 01wxyzabcdefg | 110wxyz |
| * 1wxyzabcdefgh | 111wxyz |



A律算法

- A律压缩

$$F_A[x(n)] = \begin{cases} \frac{A |x(n)| / X_{\max}}{1 + \ln A} \operatorname{sgn}[x(n)] & (0 \leq \frac{|x(n)|}{X_{\max}} < \frac{1}{A}) \\ X_{\max} \frac{1 + \ln[A |x(n)| / X_{\max}]}{1 + \ln A} \operatorname{sgn}[x(n)] & (\frac{1}{A} \leq \frac{|x(n)|}{X_{\max}} \leq 1) \end{cases}$$

- 低于门限，线性编码
- 高于门限，对数压缩



A律算法

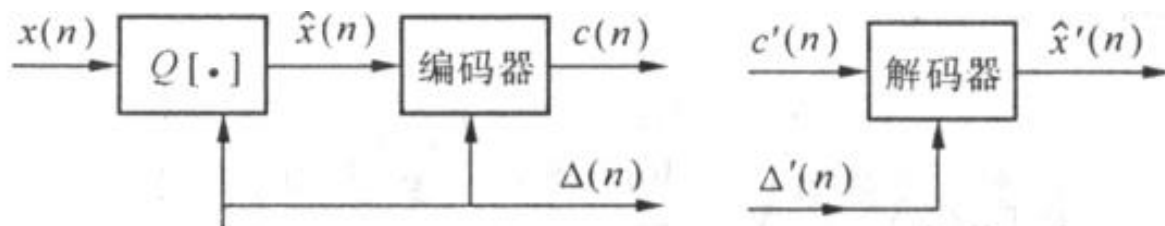
- 查表

- 保留4位有效数字，用7位完成了对12位的编码

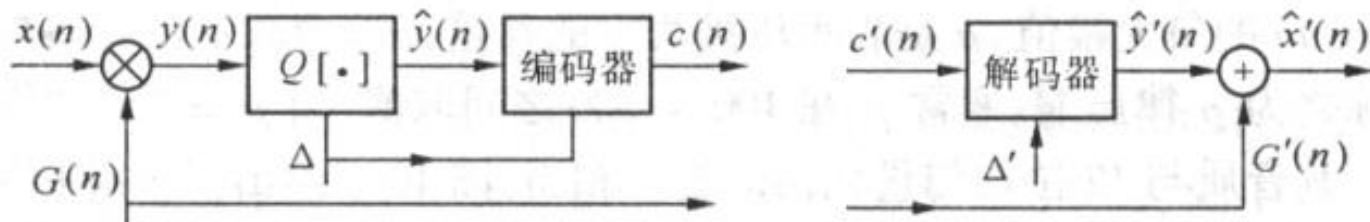
| * Linear Input Code | Compressed Code |
|---------------------|-----------------|
| * ----- | ----- |
| * 0000000wxyza | 000wxyz |
| * 0000001wxyza | 001wxyz |
| * 000001wxyzab | 010wxyz |
| * 00001wxyzabc | 011wxyz |
| * 0001wxyzabcd | 100wxyz |
| * 001wxyzabcde | 101wxyz |
| * 01wxyzabcdef | 110wxyz |
| * 1wxyzabcdefg | 111wxyz |

波形编码-自适应量化PCM

- 问题：对于PCM的量化台阶 Δ ，信号动态幅度与SNR是一对矛盾
- 解决方法：量化台阶 Δ 根据输入信号的幅值的变化而变化
 - 自适应参数



(a) Δ 匹配自适应



(b) G 匹配自适应



波形编码-自适应量化PCM

- 问题：对于PCM的量化台阶 Δ ，信号动态幅度与SNR是一对矛盾
- 解决方法：量化台阶 Δ 根据输入信号的幅值的变化而变化

- 自适应参数

- $\Delta(n)$ 、 $G(n)$

- 自适应参数产生，

$$\Delta(n) = \Delta_0 \bullet \sigma(n)$$

$$G(n) = G_0 / \sigma(n)$$

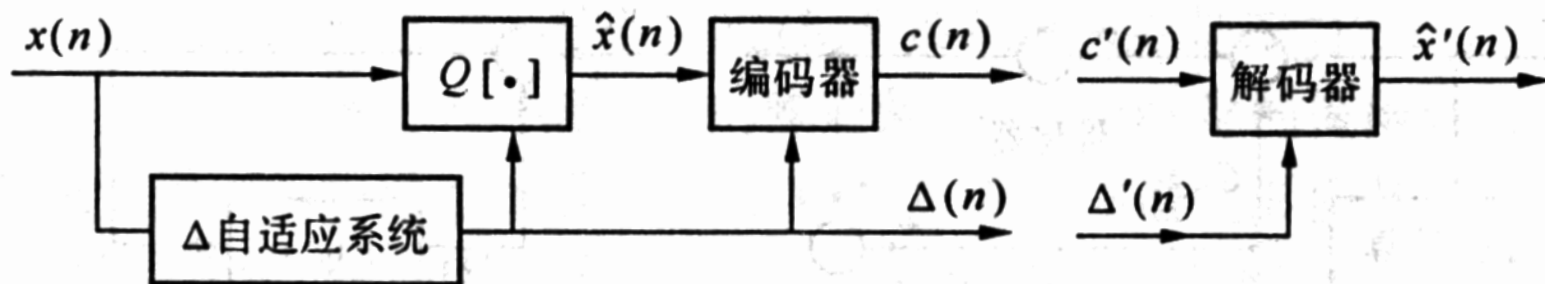
$$\sigma^2(n) = \sum_{m=-\infty}^{\infty} x^2(n)h(n-m)$$

$$\sigma^2(n) = \sum_{m=-\infty}^{\infty} \hat{x}^2(n)h(n-m)$$

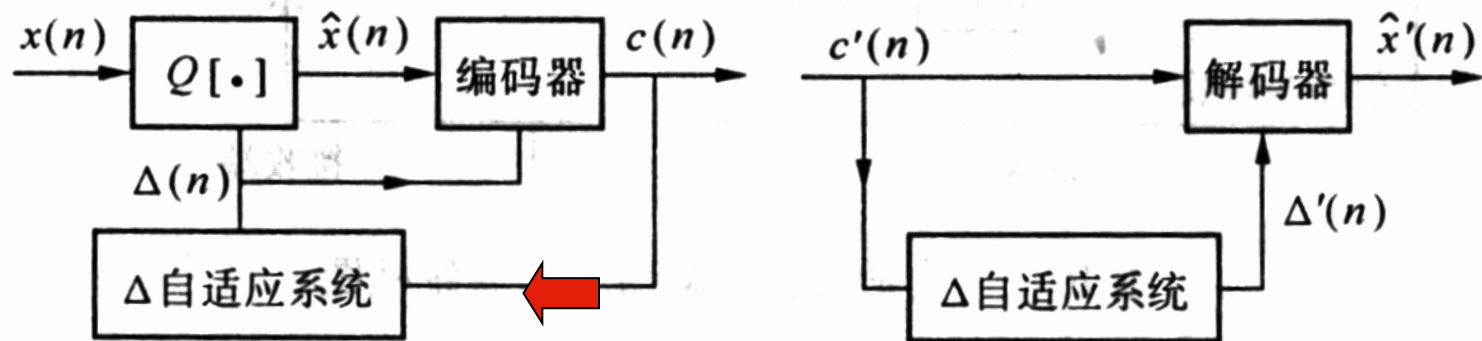
- 自适应参数来源

- 前馈式、反馈式

自适应量化PCM系统框图



(a) 前馈自适应



(b) 反馈自适应



波形编码-其它方案

■ 其它编码方案

■ 线性预测

- 对LPC及残差进行编码。LPC反映声道特性为慢变化；残差动态范围较语音信号动态小。
- 对语音信号的LP随量化残差自适应变化，自适应预测编码（APC）

■ 子带编码

- 频域编码
- 对应于听觉特性，子带可独立选择量化参数，满足主观听觉要求。
- 各带宽的宽度以对听觉的贡献为原则选择

■ 变换域编码

- 变换域中消除了语音信号样本间的冗余，数值分布范围变小

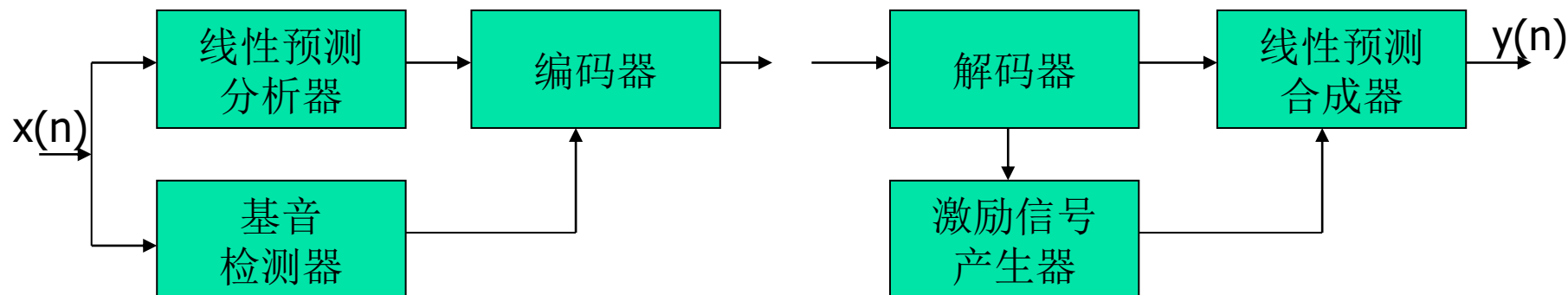


参数编码

- 波形编码特点
 - 逼近时域波形
 - 压缩效率低，16Kbps以下数码率语音质量下降
- 参数编码
 - 模型化参数
 - 压缩效率高，但语音质量较低
- 解决方案：混合编码
 - 模型参数编码
 - 激励源波形编码

参数编码-LPC声码器

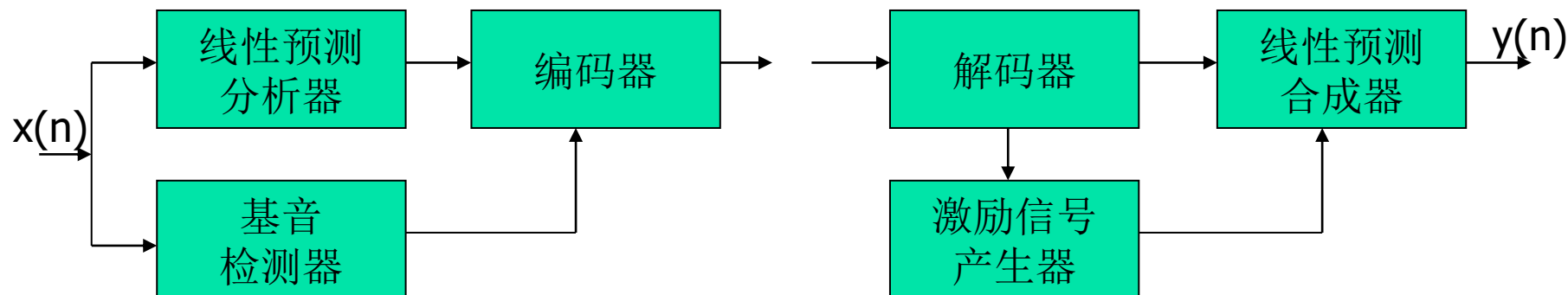
- 假定声道模型为全极点模型，LPC参数表示
- 激励源：基音周期脉冲/白噪声
- 编码信息：LPC参数、清浊及基音周期、增益



LPC声码器原理图

参数编码-LPC声码器

- 假定声道模型为全极点模型，LPC参数表示
- 激励源：基音周期脉冲/白噪声
- 编码信息：LPC参数、清浊及基音周期、增益



LPC声码器原理图

- LPC-10的问题分析：清浊判决误差、LPC参数敏感
- LPC-10e改进
 - 激励源：混合激励、脉冲抖动、码本与脉冲混合
 - 基音提取
 - LSP参数



极低速率编码

- 数码率1.2Kbps以下
- 特定场景：军事、安全、保密通讯，空间通讯
- 技术
 - LPC声码器为基础
 - 帧填充：去除帧间的相关性
 - VQ：对参数进行矢量量化编码



编码标准

- 语音编码-通讯网络
 - ITU (International Telecommunication Union)
 - 地区/行业标准

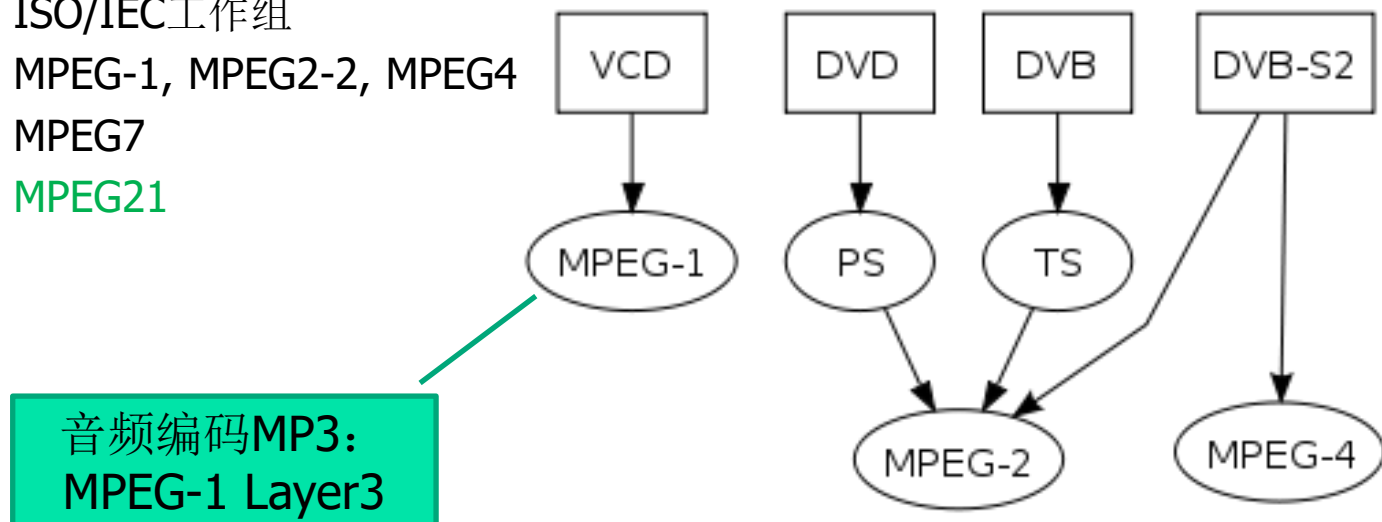


编码标准

- 语音编码-通讯网络
- 音频编码-数字媒体
 - 目标：高保真、声场景重构

编码标准

- 语音编码-通讯网络
- 音频编码-数字媒体
 - 目标：高保真、声场景重构
 - MPEG(Moving Picture Experts Group): MP3
 - ISO/IEC工作组
 - MPEG-1, MPEG2-2, MPEG4
 - MPEG7
 - MPEG21





编码标准

- 语音编码-通讯网络
- 音频编码-数字媒体
 - 目标：高保真、声场景重构
 - MPEG(Moving Picture Experts Group): MP3
 - AC-3: Dolby Surround Digital, 5.1声道
 - AAC: MPEG-2 Advanced Audio Coding
 - 版权控制
 - 多声道
 - 压缩率提高
 - MPEG-4 Advanced Audio Coding



作业4

课程中心平台-“作业4”

- 截止时间11月12日（周日）
- 格式：Word文档，命名规范：姓名_学号_作业4.doc

- 内容

1. 语音压缩的实质就是去除语音信号中冗余，请分析、归纳语音压缩编码的依据
2. 绘出 Δ 自适应PCM系统框图，说明前馈、反馈自适应的工作原理
3. 分析波形编码与参数编码的特点
4. 根据LPC声码器原理框图，分析其工作原理