多媒体通信
Multimedia Communications

第2章　多媒体数据压缩国际标准

音频数据的压缩标准

2019年9月22日

第2章 多媒体数据压缩国际标准

- 2.1 多媒体数据压缩编码的重要性和分类
- 2.2 常见数据压缩方法分类与基本原理
- 2.3 音频压缩标准
  - 2.3.1 话音编码基础
  - 2.3.2 三种话音编码器
  - 2.3.3 MPEG Audio
  - 2.3.4 移动通信网中的音频编码
- 2.4 静态图像压缩编码的国际标准
- 2.5 视频压缩的国际标准
- 2.6 可伸缩性编码和分布式编码

{ynh,cxh}@ustc.edu.cn

History of Magnetic Tape
近半个世纪的辉煌

{ynh,cxh}@ustc.edu.cn

消费电子从磁存储步入半导体存储
索尼停产卡带Walkman随身听

告别：Sony停产卡带Walkman

Sony的卡带Walkman就是上个世纪的Ipod。在人们还傻里傻气的抱着纸箱子一样大的录音机在街上跳舞时，卡带Walkman的出现让这个世界街头音乐瞬间“旧貌换新颜”。（图1）

这种改变波及全球，从美国的街头沉迷其中的黑人妇女（图2）到中国那一代年轻人把刚刚看到的浪漫和流行浓缩进卡带（图3），那是数字时代来临前最让人兴奋的一次预言。

事实上，数字时代随身电子产品的一切主要特征，几乎都可以在当年的Walkman上面找到：小型化、电子化、时尚感、廉价……这一切再熟悉不过，却始于今天向我们道别的30年前的Walkman。

祝卡带Walkman走好，愿数字Walkman常青。我们博物馆再见！（图4）

2010年10月25日

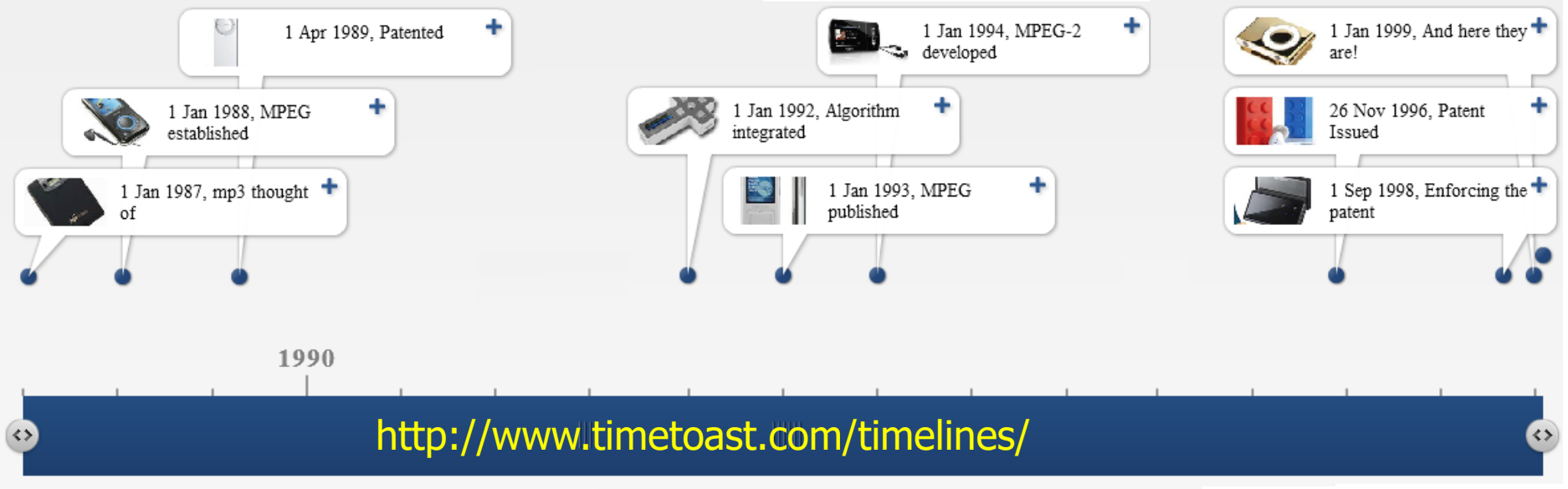
History of CD-DA
20世纪最后20年：在光盘上存放数字音乐

- 1969: The idea of a compact disc was born in the mind of Klass Compaan, a Ducth physicist.
- 1975: Research on laser and optical disc technology started by Sony.
- 1977: Philips began researching laser and optical disc technology.
- 1980: CD-DA format introduced by Philips and Sony, and standards were laid down.
- 1982: Manufacturing of CDs began on a large scale in a factory.
- 1982: First ever album on a CD released by Sony, which was Billy Joel's 52nd Street.
- 1983: CD players and discs hit the market in the US and the rest of the world.
- 1987: The first Video CD (VCD) format created for storing and playing video and audio.
- 1996: DVD technology hit the world.
- 1997: DVD released in the market, sidelining CDs.
- 1999: Super Audio CD (SACD) is released by Sony and Philips.
- 2003: The first consumer available Blu-Ray player is released in Japan by Sony.
- 2008: Sales for large label CDs drops 20% due to rising popularity of MP3 audio.

Development of the Compact Disc, CD-ROM, DVD and Blu-Ray
http://www.voices.com/resources/articles/audio-recording-technology/cd-history-timeline

{ynh,cxh}@ustc.edu.cn

History of MP3 players
21世纪，便携式MP3播放器



- 1987 - The **Fraunhofer** Institut in Germany began research (DAB).
- April 1989 - Fraunhofer received a German patent for MP3
- **1992** - Fraunhofer's and Dieter Seitzer's algorithm was integrated into MPEG-1.
- 1993 - MPEG-1 standard published
- 1994 - MPEG-2 developed and published a year later.
- 1996 - United States patent issued for MP3
- 1998 - Fraunhofer started to enforce their patent rights.
- **1999** - Portable MP3 players appear.

{ynh,cxh}@ustc.edu.cn

数字音频主要应用领域

- 应用范围
  - 无线电广播、电话、电视信号中的声音
  - 移动通信、卫星通信、音频文件
- Digital audio broadcasting
  - Digital Audio Broadcasting (DAB)
- Storage technologies
  - Digital audio player
  - Compact Disc (CD)
  - DVD-Audio
  - Super Audio CD
  - Blu-ray Disc (BD)

{ynh,cxh}@ustc.edu.cn

Technical Details of Lossy Audio Compression Formats

Audio compression format	Algorithm	Sample Rate	Bit rate	Latency
AAC	MDCT, Hybrid Subband (AAC-HE)	8 kHz to 192 kHz[36]	8 to 529 kbit/s (stereo)	20–405ms[37]
AC3	MDCT	32, 44.1, 48 kHz	32 to 640 kbit/s	40.6ms
AMR	ACELP	8 kHz	4.75, 5.15, 5.90, 6.70, 7.40, 7.95, 10.20, 12.20 kbit/s	25ms
GSM-HR	VSELP	8 kHz	5.6 kbit/s	25ms
GSM-FR	RPE-LTP	8 kHz	13 kbit/s	20-30ms
GSM-EFR	ACELP	8 kHz	12.2 kbit/s	20-30ms
HVXC	Speech	8 kHz	2 or 4 kbit/s	36ms
MP3 (MPEG-1, 2, 2.5 Audio Layer III)	MDCT, Hybrid Subband	8, 11.025, 12, 16, 22.05, 24, 32, 44.1, 48 kHz	8, 16, 24, 32, 40, 48, 56, 64, 80, 96, 112, 128, 144, 160, 192, 224, 256, 320 kbit/s	>100ms
Musepack	Subband	32, 37.8, 44.1, 48 kHz	3 to 1300 kbit/s	?
Vorbis (Ogg)	MDCT	1 Hz to 200 kHz	variable	>100ms
Windows Media Audio Pro	MDCT	8, 11.025, 16, 22.05, 32, 44.1, 48, 88.2, 96 kHz	4 to 768kbit/s	>100ms

http://en.wikipedia.org/wiki/Comparison_of_audio_codecs

{ynh,cxh}@ustc.edu.cn

Technical Details of Lossless Audio Compression Formats

Audio compression format	Algorithm	Sample Rate	Bits per sample	Latency
ALAC	Lossless	44.1 kHz to 192 kHz	16, 24[40]	?
FLAC	Lossless	1 Hz to 655350 Hz	8, 16, 20, 24, (32)	4.3ms - 92ms (46.4ms typical)
Monkey's Audio	Lossless	8, 11.025, 12, 16, 22.05, 24, 32, 44.1, 48 kHz	?	?
RealAudio Lossless	Lossless	Varies (see article)	Varies (see article)	Varies
True Audio	Lossless	0–4 GHz	1 to > 64	?
WavPack Lossless	Lossless, Hybrid	1 Hz to 16.777216 MHz	varies in lossless mode; 2.2 minimum in lossy mode	?
Windows Media Audio Lossless	Lossless	8, 11.025, 16, 22.05, 32, 44.1, 48, 88.2, 96 kHz	16, 24	>100ms

Lossless

指处理对象是数字音频 （采样、量化后得到的声音样本）

{ynh,cxh}@ustc.edu.cn

数字音频的研究对象

- 音频(Audio)信号
  - 频率范围为20 Hz～20 kHz的信号
  - 一般来说，人的听觉器官能感知的声音频率大约在20～20000 Hz之间，不同的人耳存在差异
- 话音(speech)信号
  - 人的发音器官发出的声音频率大约是80～3400 Hz；人说话的信号频率通常为300～3000 Hz，这种频率范围的信号称为话音，也称作**语音**
- 常见音频
  - 电话话音（窄带话音） 200～3400Hz / 13bits
  - 宽带话音 50～7000Hz / 16bits
  - 调频广播 20～15kHz / 16bits
  - 高质量音频 20～20k Hz / 16bits

{ynh,cxh}@ustc.edu.cn

第2章 多媒体数据压缩国际标准

- 2.1 多媒体数据压缩编码的重要性和分类
- 2.2 常见数据压缩方法分类与基本原理
- 2.3 音频压缩标准
  - 2.3.1 话音编码基础 ←话音信号的冗余
  - 2.3.2 三种话音编码器
  - 2.3.3 MPEG Audio
  - 2.3.4 移动通信网中的音频编码
- 2.4 静态图像压缩编码的国际标准
- 2.5 视频压缩的国际标准
- 2.6 可伸缩性编码和分布式编码

{ynh,cxh}@ustc.edu.cn

音频信号的冗余

- 时域信息的冗余度
  - 幅度的非均匀分布、样本间的相关 、周期之间的相关、基音之间的相关 、静音系数 、长时自相关函数
- 频域信息的冗余度
  - 非均匀的长时功率谱密度
  - 语音特有的短时功率谱密度
- 人的听觉感知机理
  - 在“MPEG Audio”小节展开

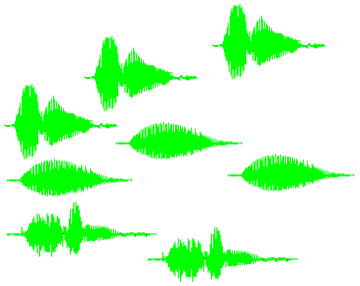
{ynh,cxh}@ustc.edu.cn

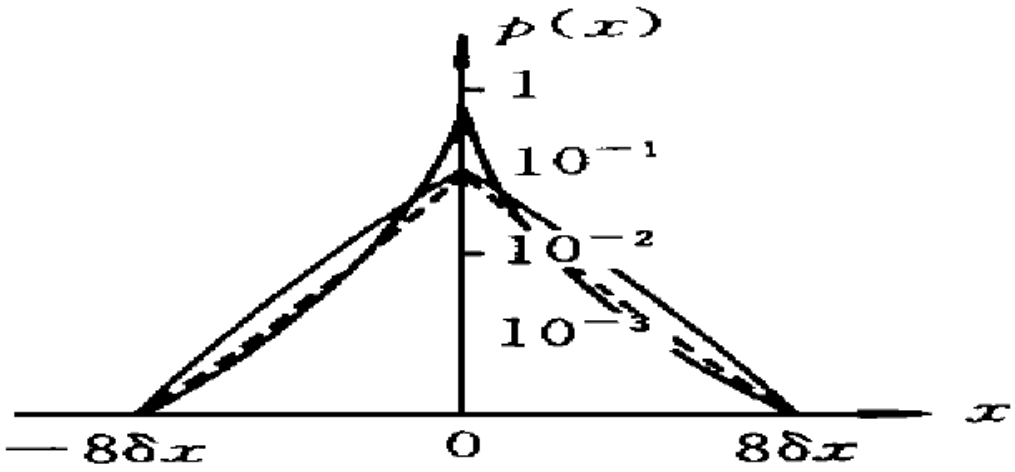
（1）幅度的非均匀分布

◆统计表明，语音中的小幅度样本比大幅度样本出现的概率要高。

◆由于通话中会有间隙，出现了大量的低电平样本。

◆实际讲话信号功率电平也趋向于出现在编码范围的较低电平端。





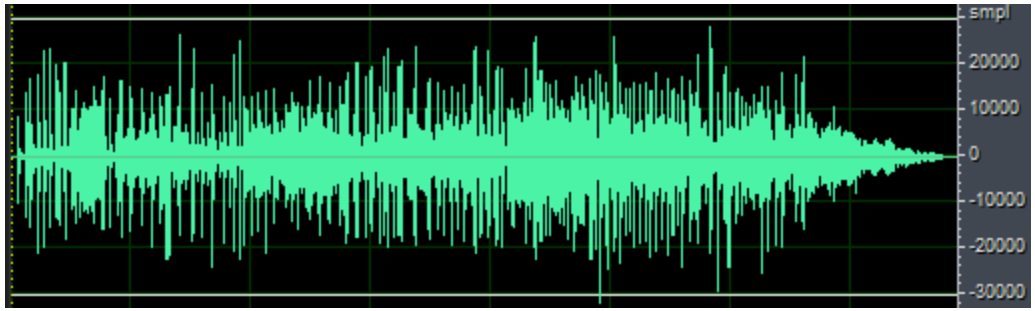
长时话音的振幅分布

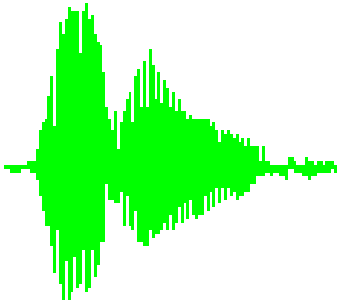
yinli,cxnj@ustc.edu.cn

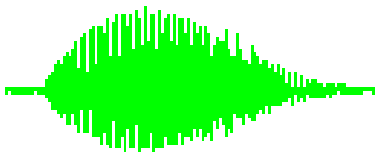
（2）样本间的相关

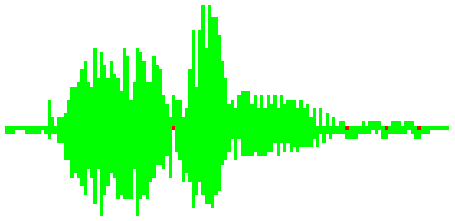
◆对语音波形的分析表明，取样数据的最大相关性存在于邻近样本之间。

◆当取样频率为8kHz时，相邻取样值间的相关系数大于0.85；甚至在相距10个样本之间，还可有0.3左右的数量级。如果取样速率提高，样本间的相关性将更强。







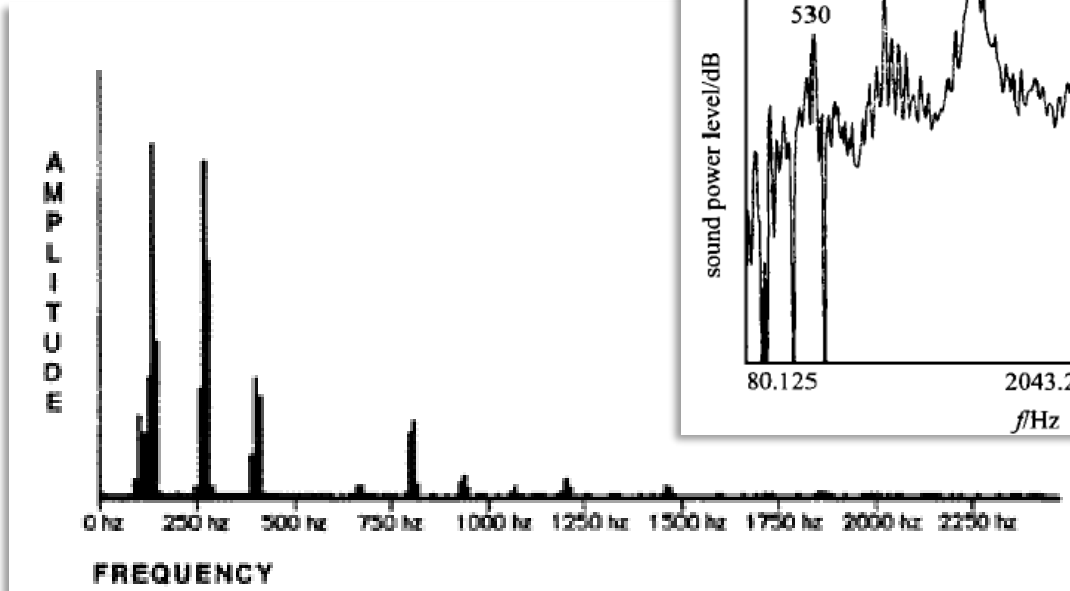


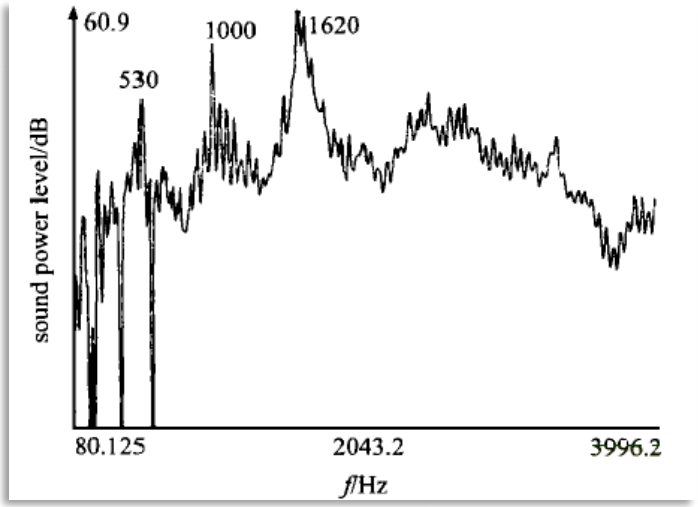
{ynh,cxh}@ustc.edu.cn

（3）周期之间的相关

◆当声音中只存在少数几个频率时，就会像某些振荡波形一样，在周期与周期之间，存在着一定的相关性。

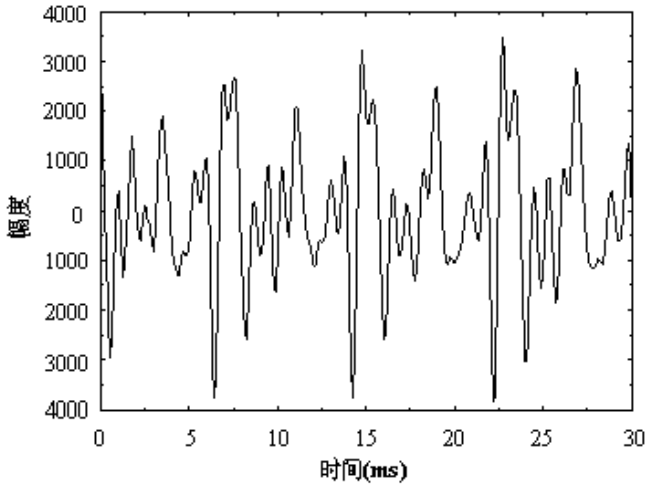
◆利用语音周期之间信息冗余度的编码器，比仅仅只利用邻近样本间的相关性的编码器效果要好，但要复杂得多。

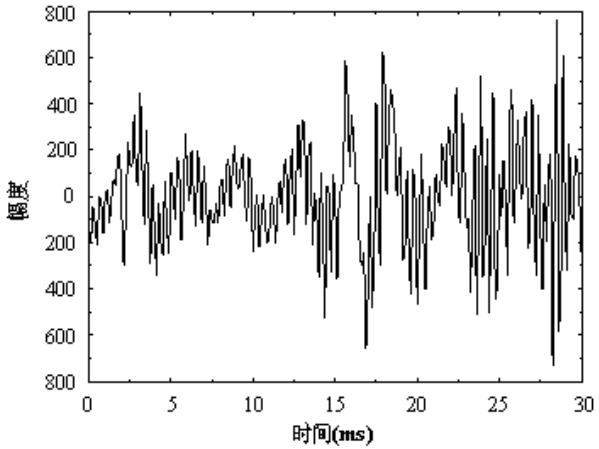


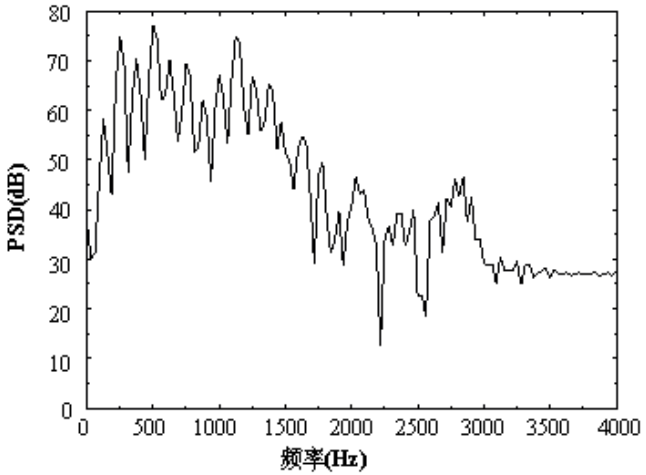


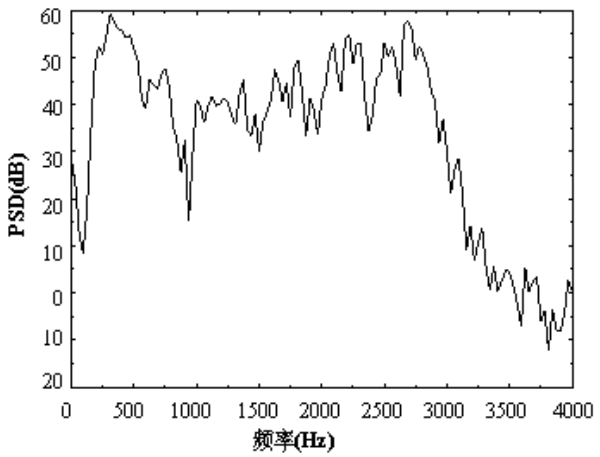
不光话音，很多音源发出的声音具有上述特性

（4）基音之间的相关









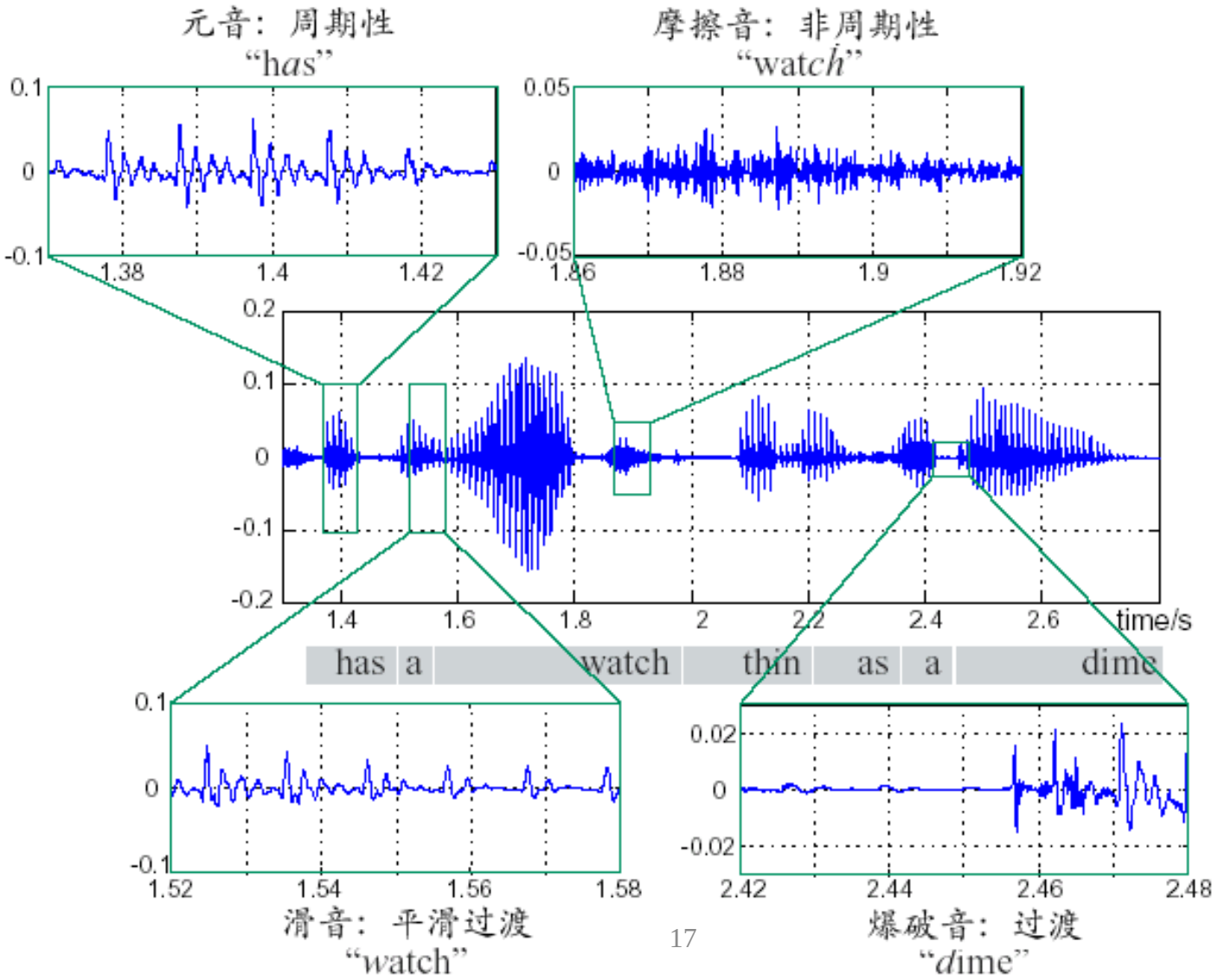
◆第一类称为浊音（voiced sound），一种准周期脉冲激励所发出的音。浊音表现在音节上有高度的周期性，其值在2-20ms之间，这个周期性称为长期周期性。

◆第二类称为清音（unvoiced sound），由不稳定气流激励所产生的，这种气流是在声门处打开状态下强制空气在声道里高速收缩产生的。

◆第三类称为爆破音（plosive sound）,它是在声道关闭之后产生的压缩空气然后打开声道所发出的音。

{ynh,cxh}@ustc.edu.cn

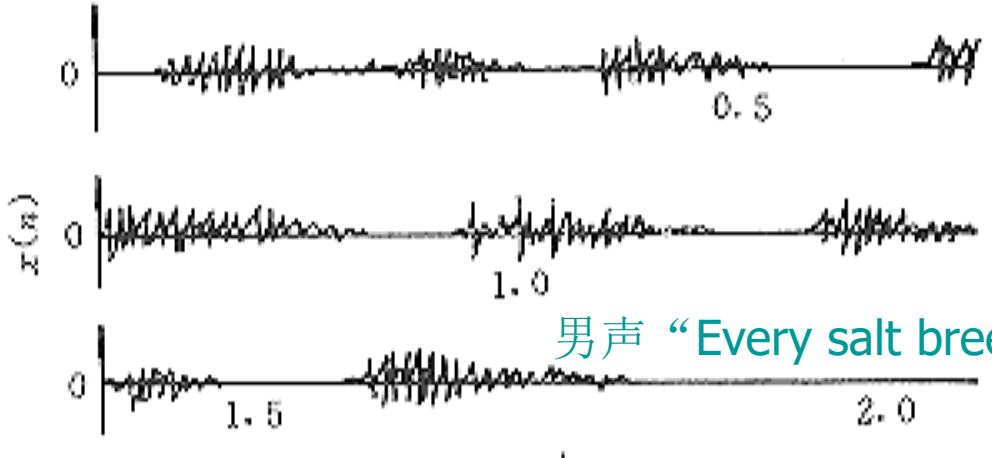
基音之间的相关举例



（5）静止系数

◆两个人之间打电话，平均每人的讲话时间为通话总时间的一半，另一半时间听对方讲。听的时候一般不讲话，而即使是在讲话的时候，也会出现字、词、句之间的停顿。

◆通过分析表明，话音间隙使得全双工话路的典型效率约为通话时间的40%（或静止系数为0.6）。显然，话音间隔本身就是一种冗余，若能正确检测出该静止段，便可“插空”传输更多的信息。



男声“Every salt breeze comes from the sea”

{ynh,cxh}@ustc.edu.cn

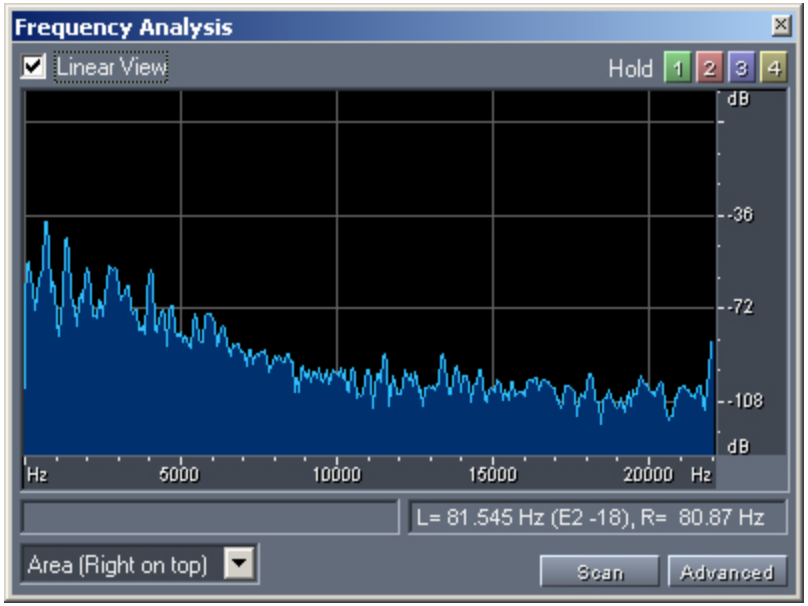
（6）长时自相关函数

- 上述样本、周期间的一些相关性，都是在20ms时间的间隔内进行统计的所谓短时自相关。如果在较长的时间间隔（比如几十秒）进行统计，便得到长时自相关函数。
- 长时统计表明，8kHz的取样语音的相邻样本间，平均相关系数高达0.9。
- SRD---short Range dependent
- LRD---long Range dependent

{ynh,cxh}@ustc.edu.cn

频域信息的冗余度：长时功率谱特性

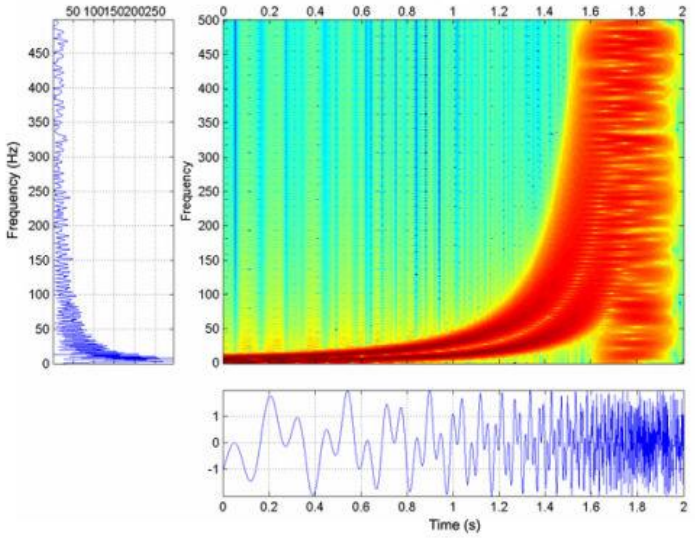
- 非均匀的长时功率谱密度
  - 在相当长的时间间隔内进行统计平均，可得到长时功率谱密度函数，其功率谱呈现强的**非平坦性**。
  - 从统计的观点看，这意味着没有充分利用给定的频段，或者说有着固有的冗余度。
  - 特别地，**功率谱的高频能量较低**，这恰好对应于时域上相邻样本间的相关性。



{ynh,cxh}@ustc.edu.cn

频域信息的冗余度：短时功率谱特性

- 语音特有的短时功率谱密度
  - 语音信号的短时功率谱，在某些频率上出现峰值，而在另一些频率上出现谷值。这些峰值频率，也就是能量较大的频率，通常称为**共振峰频率**。此频率不止一个，最主要的是第一和第二个，由它们决定了不同的语音特征。
  - 另外，整个谱也是随频率的增加而递减。更重要的是，整个功率谱的细节以基音频率为基础，形成了高次谐波结构。这都与电视信号类似，仅有的差异在于直流分量较小。



小结：话音信号的冗余

- 非均匀特性
  - 幅度的非均匀分布（时域）
  - 静音系数（时域）
  - 非均匀的长时功率谱密度（频域）
  - 语音特有的短时功率谱密度（频域）
- 相关性
  - 样本间的相关（时域）
  - 周期之间的相关（时域）
  - 基音之间的相关（时域）
  - 长时自相关函数（时域）

{ynh,cxh}@ustc.edu.cn

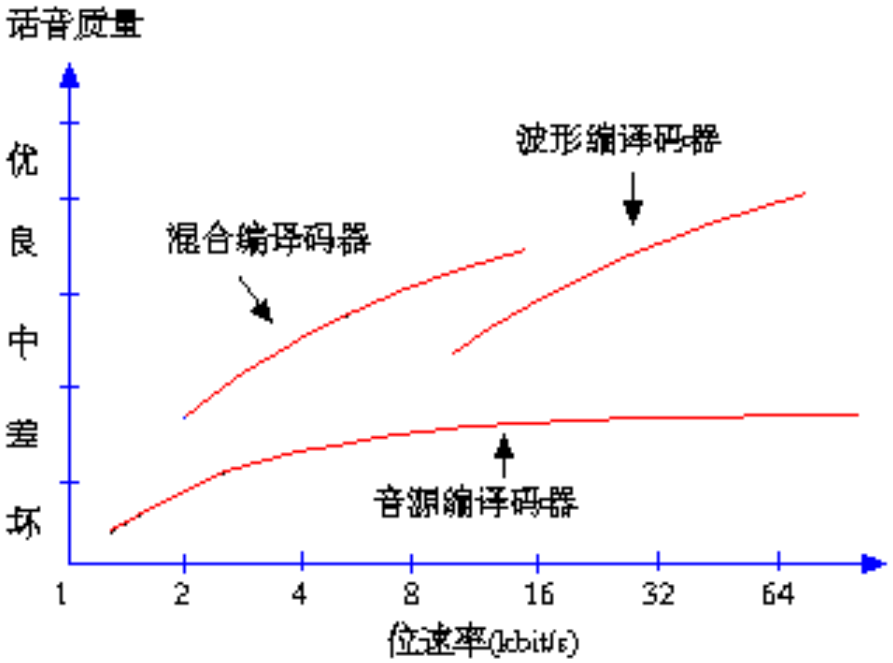
第2章 多媒体数据压缩国际标准

- 2.1 多媒体数据压缩编码的重要性和分类
- 2.2 常见数据压缩方法分类与基本原理
- 2.3 音频压缩标准
  - 2.3.1 话音编码基础
  - 2.3.2 三种话音编码器
    - 波形编译码器、音源编译码器、混合编译码器
  - 2.3.3 MPEG Audio
  - 2.3.4 移动通信网中的音频编码
- 2.4 静态图像压缩编码的国际标准
- 2.5 视频压缩的国际标准
- 2.6 可伸缩性编码和分布式编码

{ynh,cxh}@ustc.edu.cn

话音编译码器的分类

- **波形编码**：不利用声音的任何知识，数据率较高，实现简单
- **音源编码**：从声音的波形中提取生成话音的参数，数据率可以很低，实现复杂
- **混合编码**：以上两种思想的结合



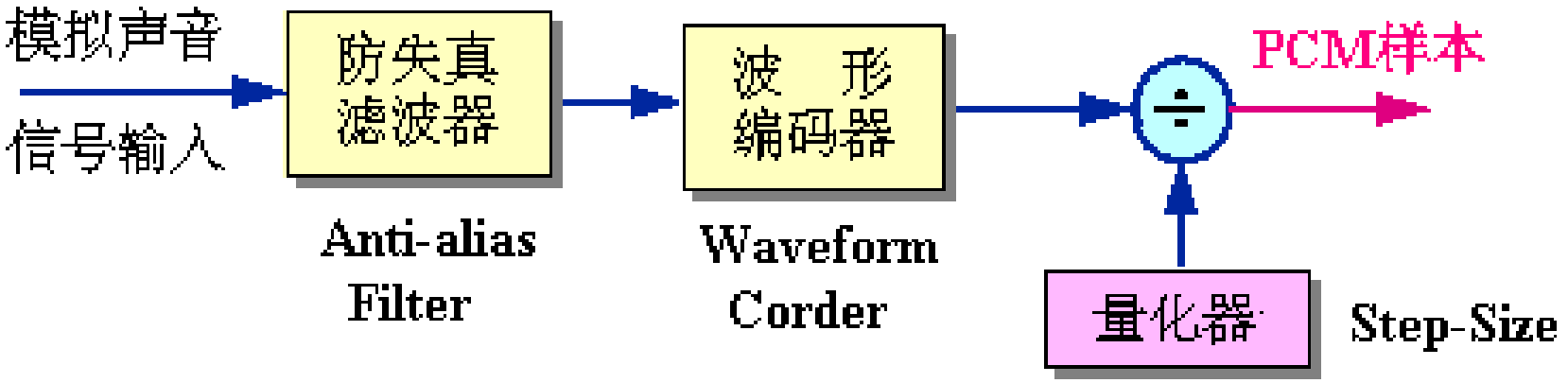
{ynh,cxh}@ustc.edu.cn

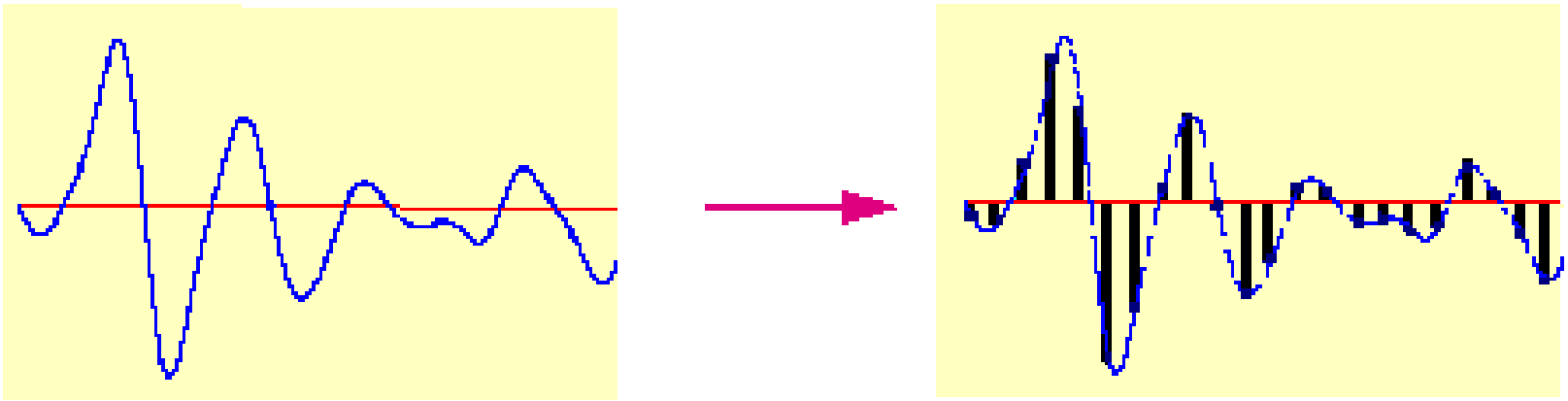
波形编译码器

- 波形编译码的想法
  - 不利用生成话音信号的知识产生而是产生一种重构信号，重构信号的波形和原始话音波形尽可能一致，这种编译码器的复杂程度低。
- 波形编码代表
  - PCM（脉冲编码调制）

{ynh,cxh}@ustc.edu.cn

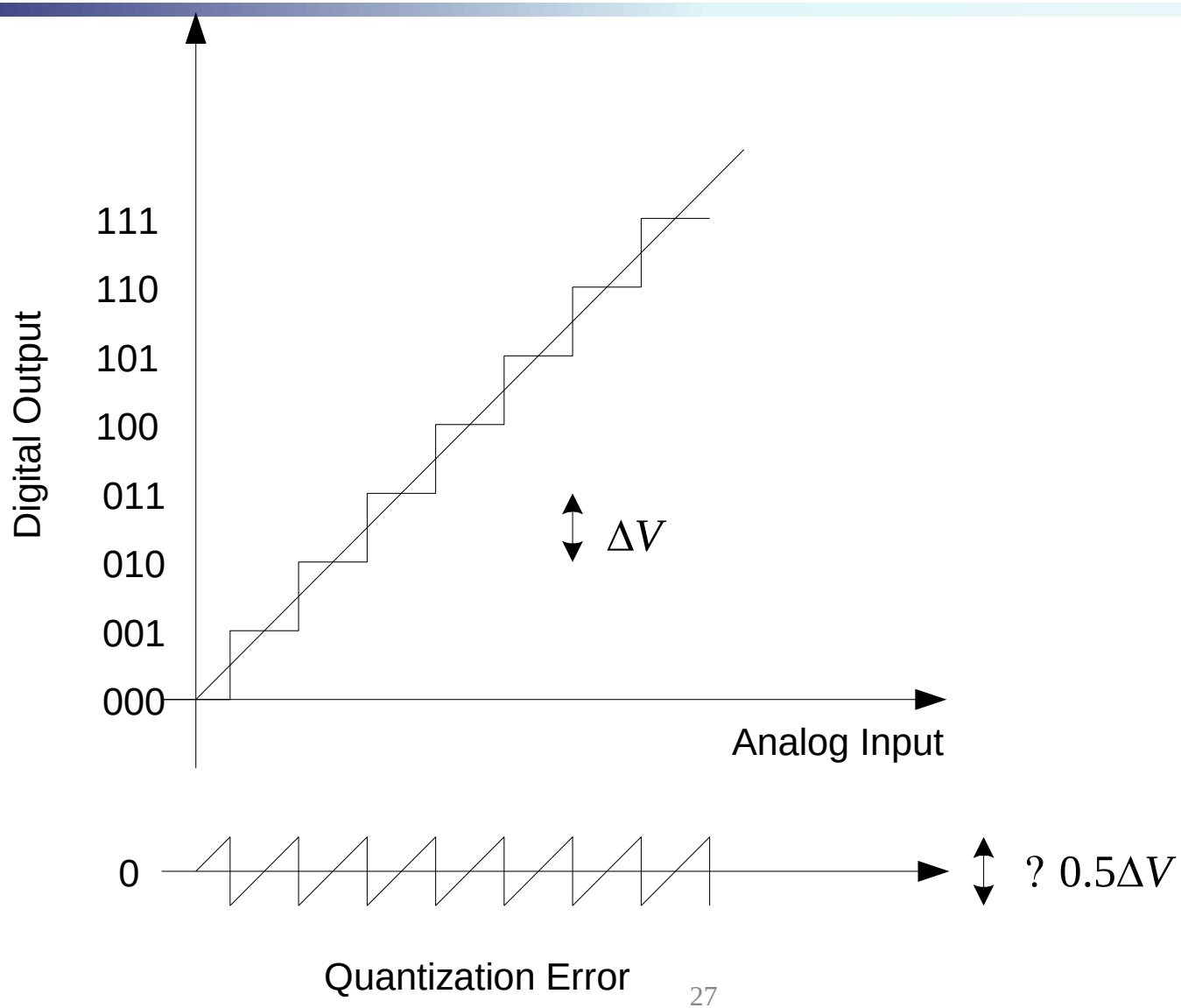
脉冲编码调制(PCM)





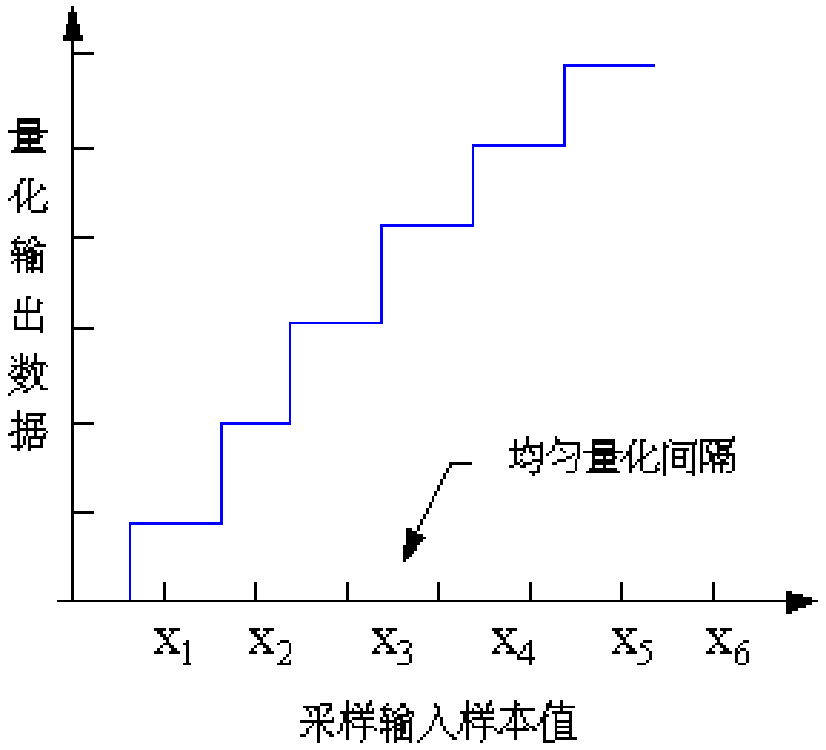
{ynh,cxh}@ustc.edu.cn

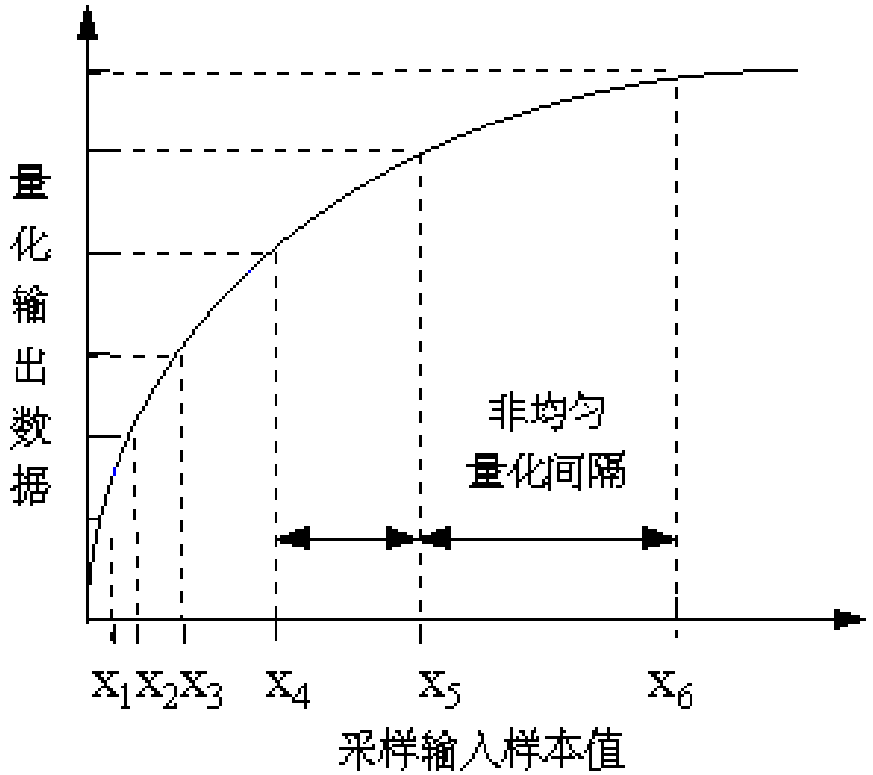
3bit量化过程中量化误差示意



{ynh,cxh}@ustc.edu.cn

脉冲编码调制(PCM)的量化方式





- μ 律(μ-Law)压扩(G.711)主要用在北美和日本等
- A律(A-Law)压扩(G.711)主要用在欧洲和中国大陆等

{ynh,cxh}@ustc.edu.cn

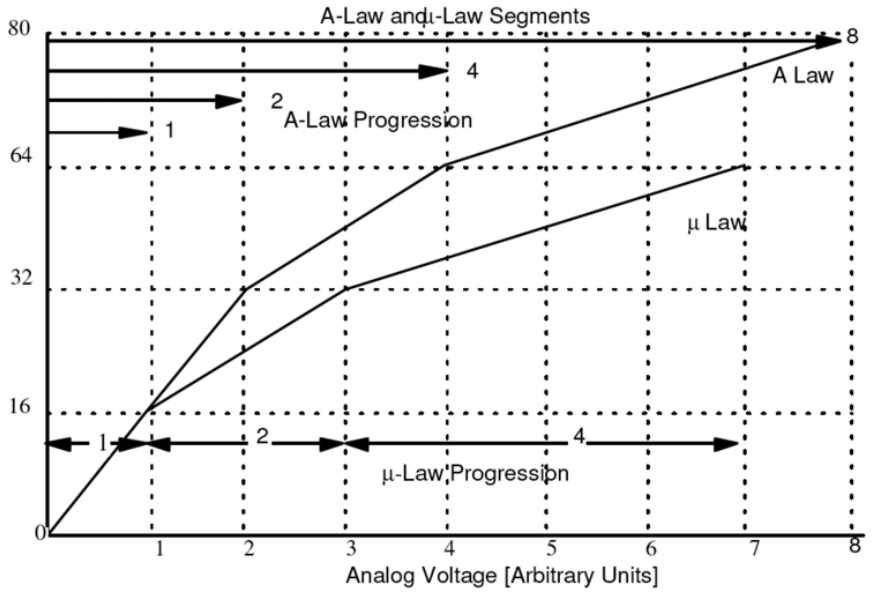
μ-Law & A-Law

- A-Law

$$F(x) = \operatorname{sgn}(x) \begin{cases} \frac{A|x|}{1+\ln(A)}, & |x| < \frac{1}{A} \\ \frac{1+\ln(A|x|)}{1+\ln(A)}, & \frac{1}{A} \le |x| \le 1, \end{cases}$$

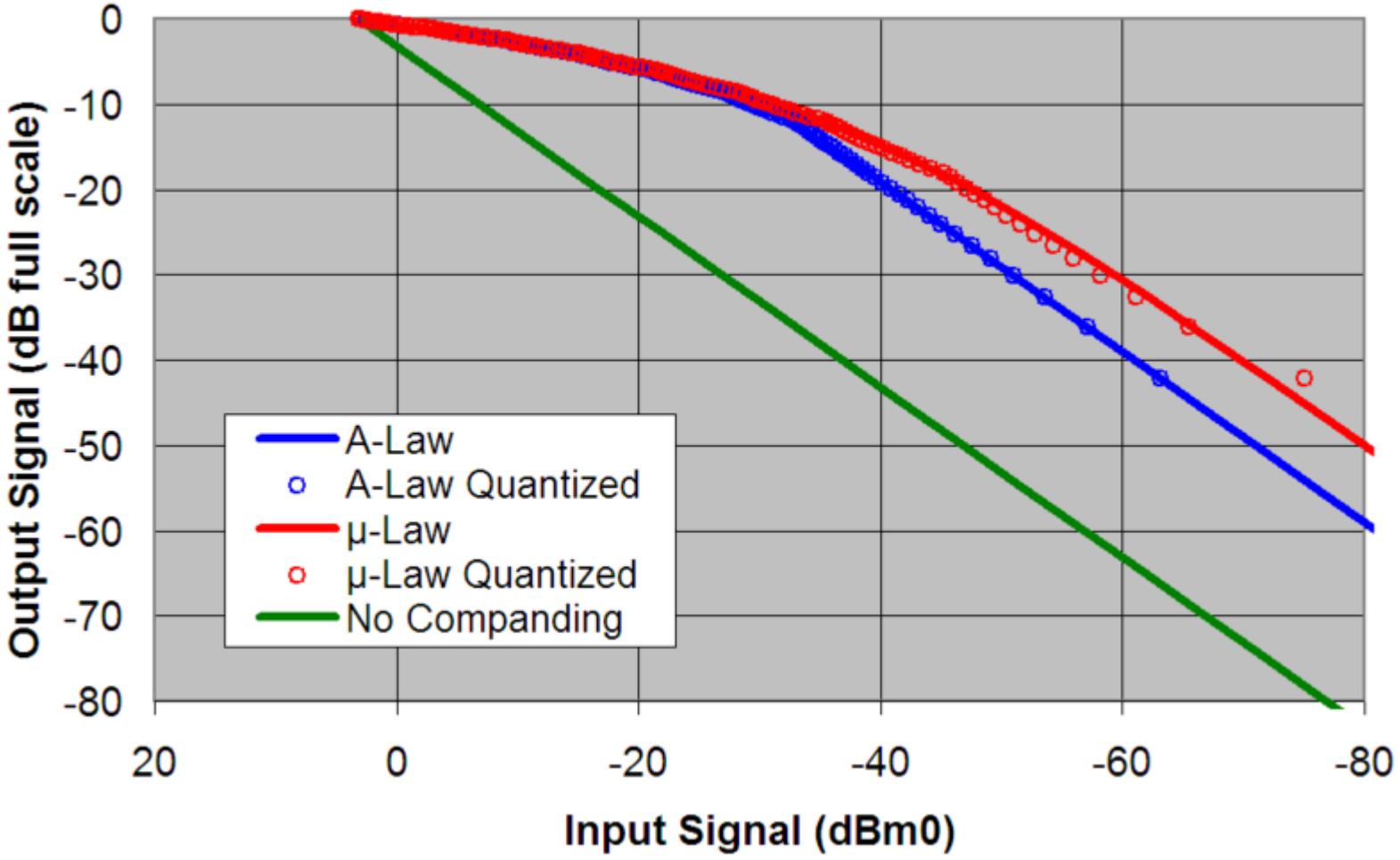
- μ-Law

$$F(x) = \operatorname{sgn}(x) \frac{\ln(1+\mu|x|)}{\ln(1+\mu)} \qquad -1 \le x \le 1$$



{ynh,cxh}@ustc.edu.cn

μ-Law vs. A-Law

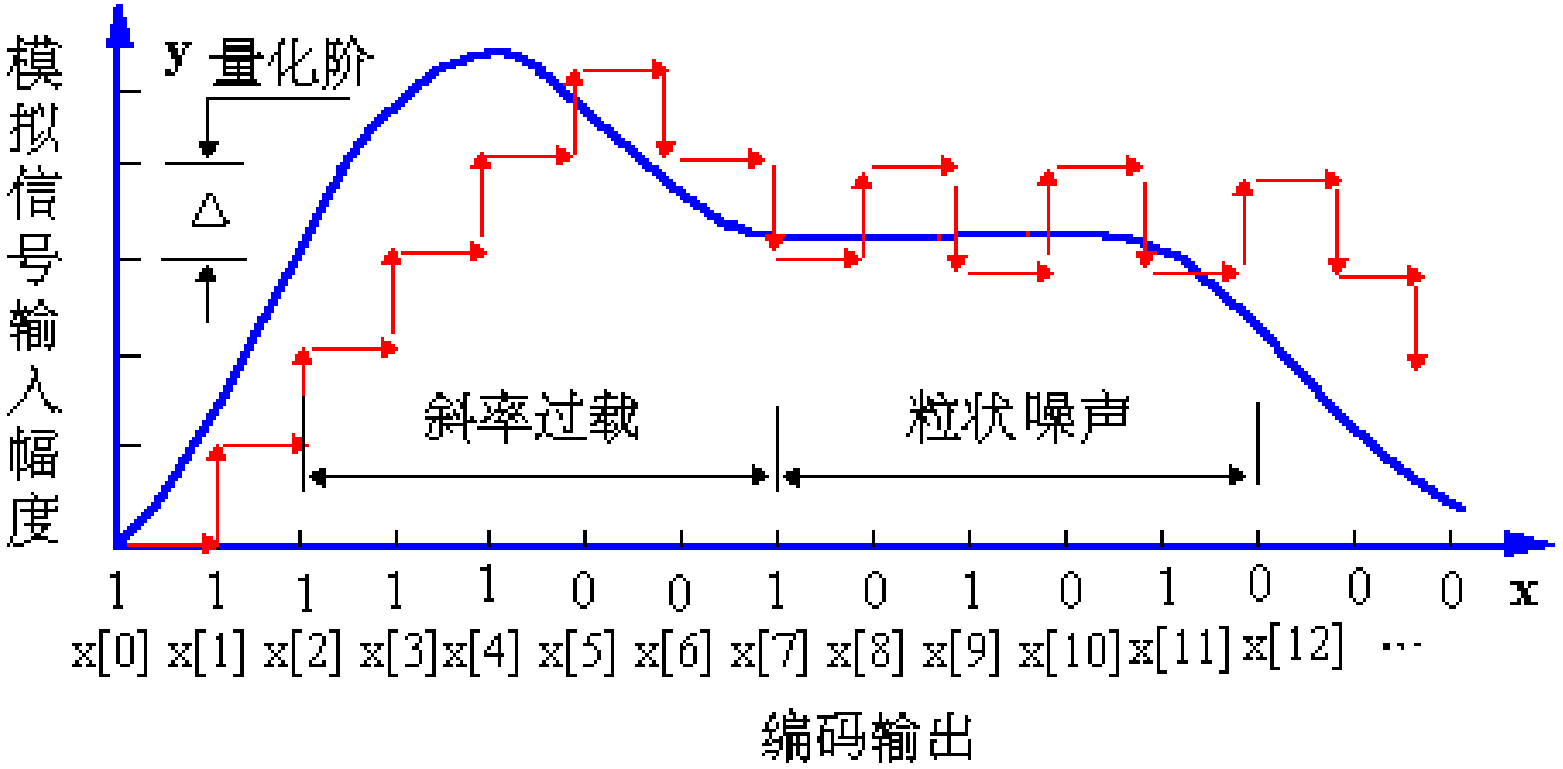


Source: http://en.wikipedia.org/wiki/Image:Ulaw_alaw.png

{ynh,cxh}@ustc.edu.cn

增量调制(△M)

两个固有的问题



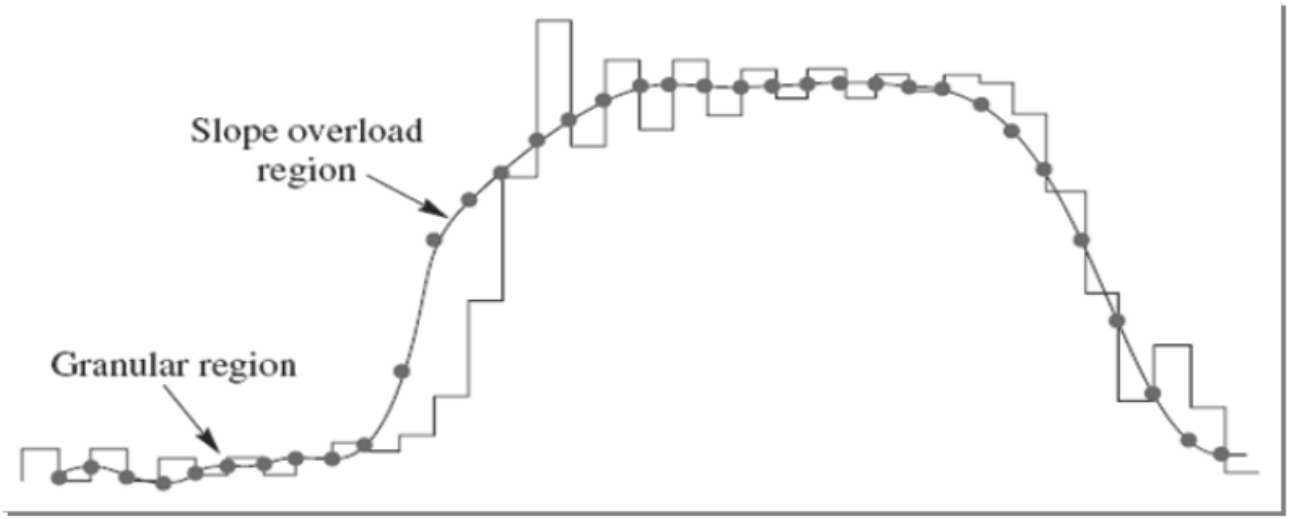
{ynh,cxh}@ustc.edu.cn

自适应增量调制

- 思路：自动调整量化阶△的大小；在检测到斜率过载的时候增大△，在输入信号斜率减小时降低△
- CFDM(Constant Factor Adaptive DM)
  - 根据量化器符号的判断当前区域是斜率过载还是颗粒噪声，进而改变△
- CVSD(Continuously Variable Slope DM)
  - 如果连续出现三个相同值△加大，反之减小

{ynh,cxh}@ustc.edu.cn

CFDM调整量化阶的过程



$$s_n = \begin{cases} 1 & \text{if } \hat{d}_n > 0 \\ -1 & \text{if } \hat{d}_n < 0 \end{cases} \qquad \Delta_n = \begin{cases} M_1 \Delta_{n-1} & \text{if } s_n = s_{n-1} \\ M_2 \Delta_{n-1} & \text{if } s_n \neq s_{n-1} \end{cases}$$

$$1 < M = M_1 = 1/M_2 < 2$$

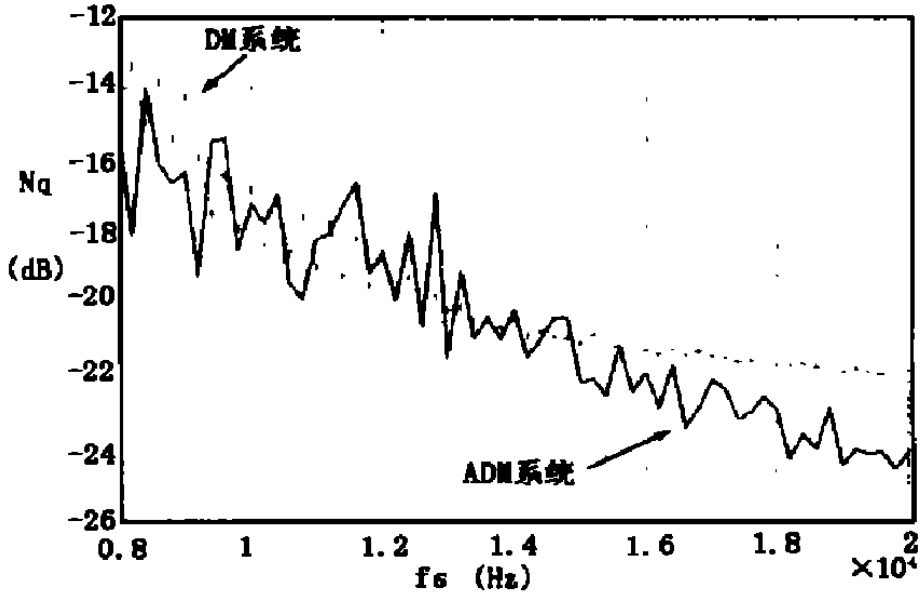
{ynh,cxh}@ustc.edu.cn

例：DM与ADM的量化噪声

◆输入的模拟信号是两个正弦信号的叠加，频率分别为：100 Hz和300 Hz。ADM采用两次交叠法，k=1.1，△初值=0. 125

$$\beta = \tilde{e}_n \times \tilde{e}_{n-1}$$

$$\Delta_n = \Delta_{n-1} \times k^{\beta}$$



{ynh,cxh}@ustc.edu.cn

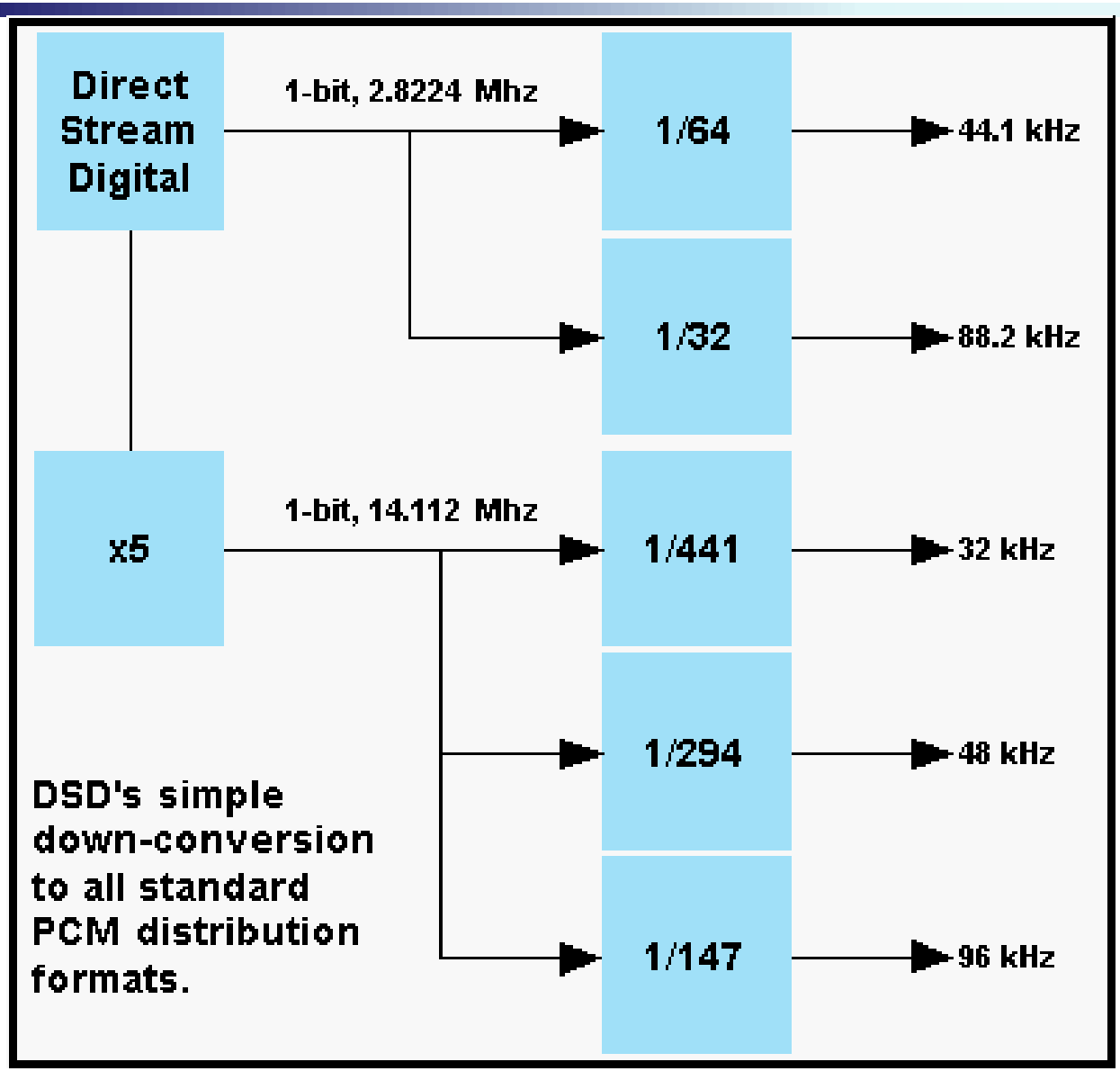
PCM vs. △M

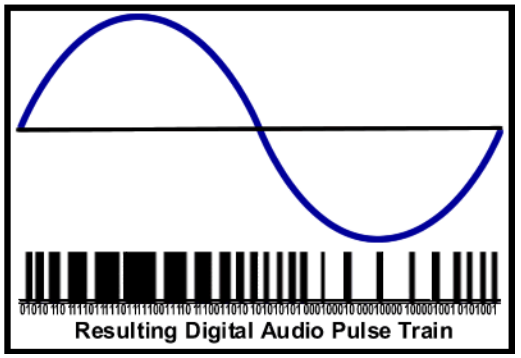
- 对于音频信号哪种更好？
- 各自怎样保证失真较小？

- CD
  - PCM（16bit/44.1kHz）
- SACD（Super Audio CD）
  - △M（1bit/2.8224MHz）

{ynh,cxh}@ustc.edu.cn

DSD (Direct Stream Digital)



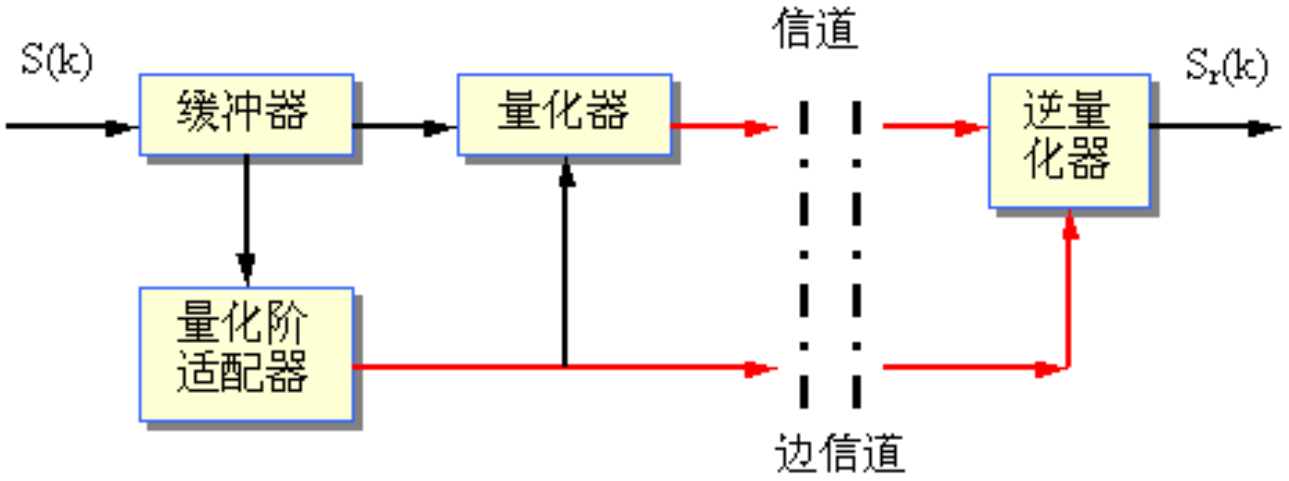


SUPER AUDIO CD

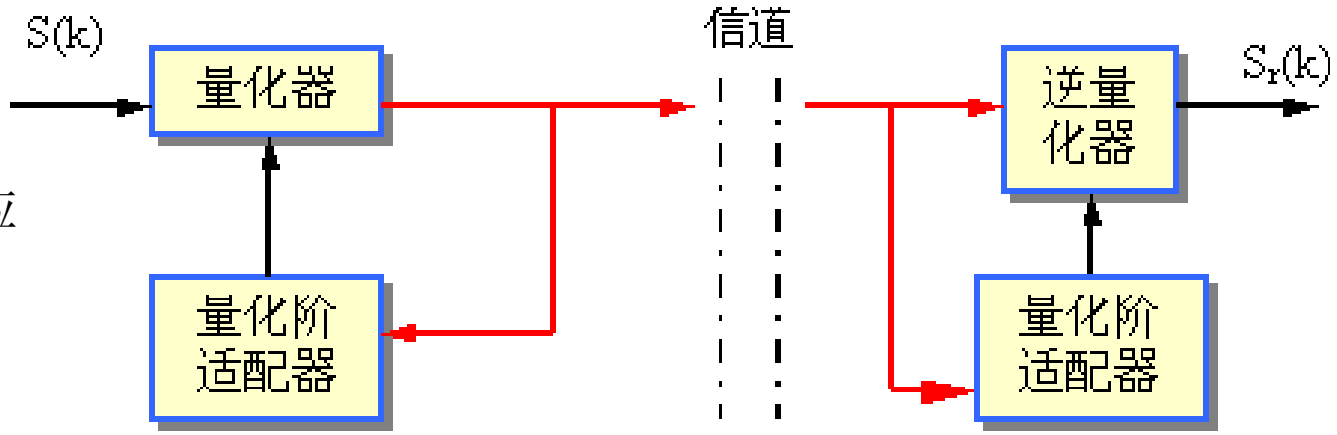
{ynh,cxh}@ustc.edu.cn

自适应脉冲编码调制 (APCM)

前向自适应

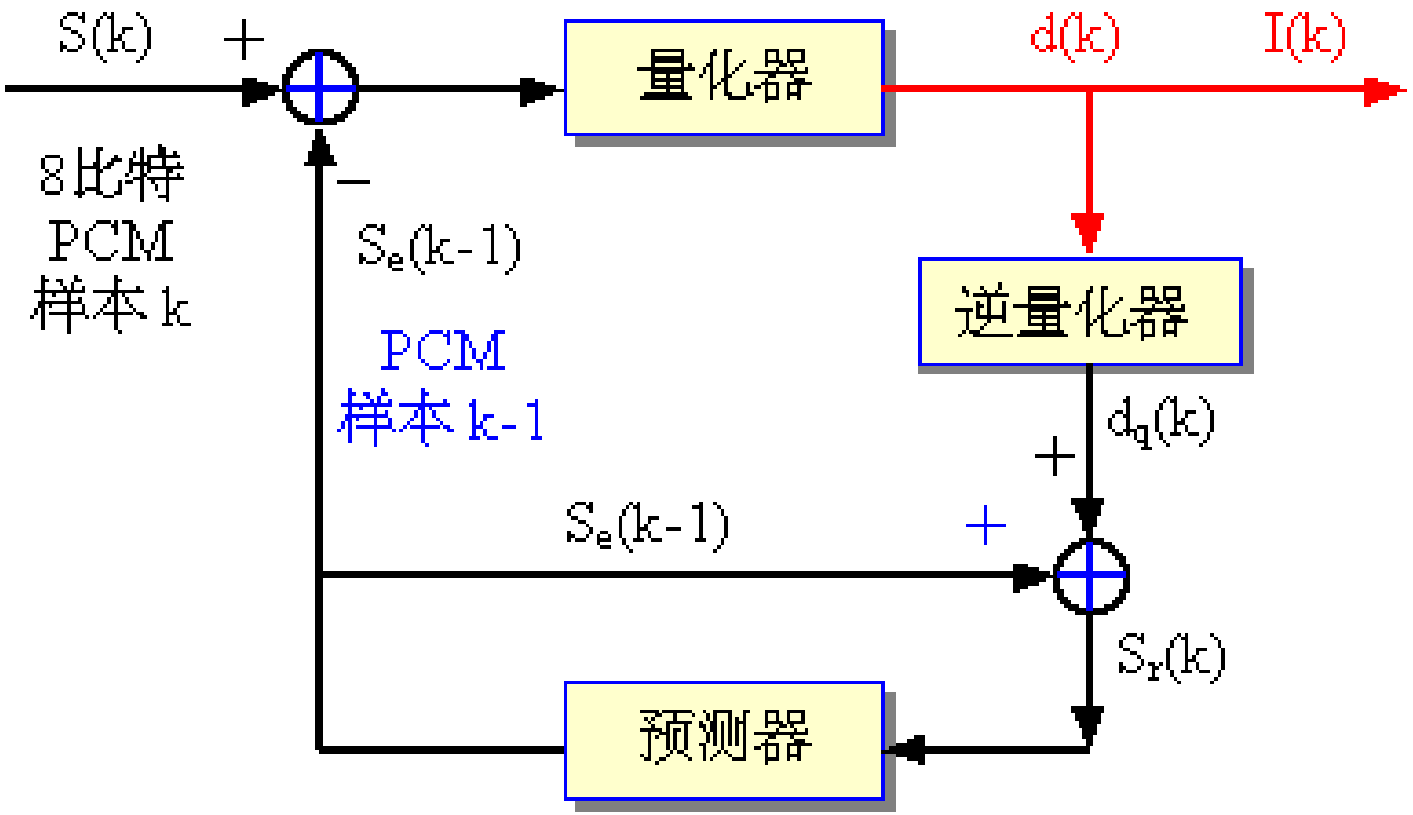


后向自适应



{ynh,cxh}@ustc.edu.cn

差分脉冲编码调制 (DPCM)



{ynh,cxh}@ustc.edu.cn

线性预测器

$$p_n = \sum_{i=1}^{N} a_i \hat{x}_{n-i}$$

◆N为线性预测器的阶数，使

$$\sigma_d^2 = E\left[\left(x_n - \sum_{i=1}^{N} a_i x_{n-i}\right)^2\right]$$

◆最小，求解采用Durbin算法【6系数字信号处理II内容】

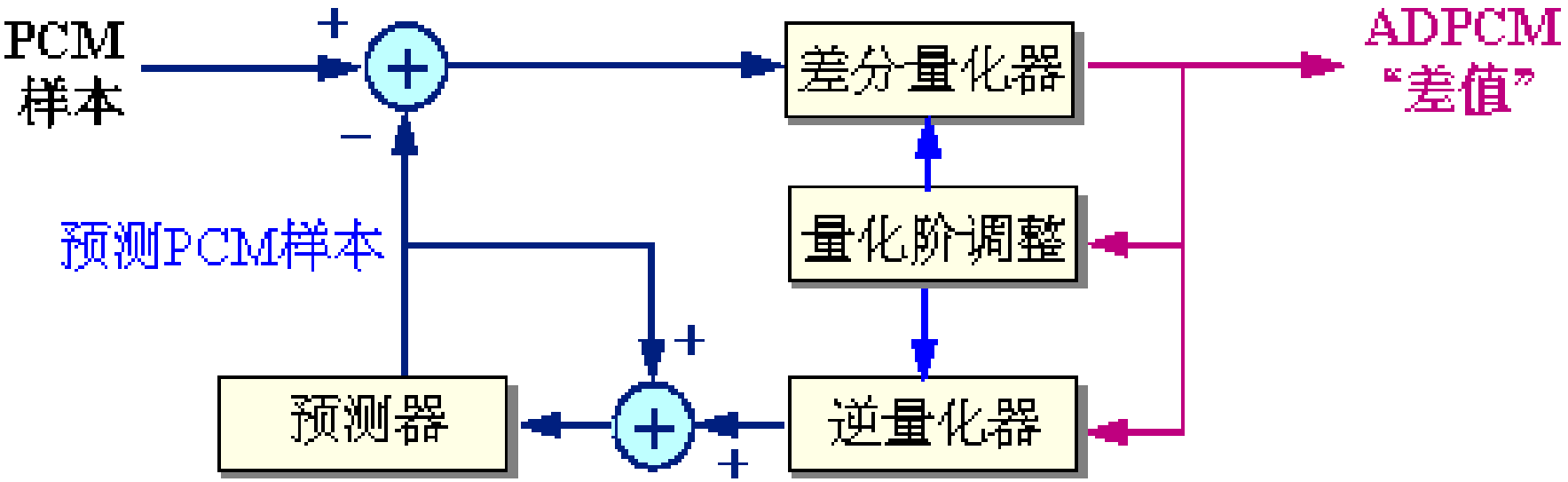
{ynh,cxh}@ustc.edu.cn

预测结果评判

- 利用误差最小的原则来确定预测模型的各项参数
  - 误差 – 实际值与预测值之间的偏差
  - 绝对平均偏差(MAD:mean absolute deviation)
    - 误差的绝对值的平均值
  - 均方差(MSE:mean square error)
    - 误差的平方的平均值
  - 累积误差(RSFE: running sum of forecast error)
- Such as：在企业效益不变的前提下，老板可以按照总体工薪成本最低的原则确定每年的人力资源计划

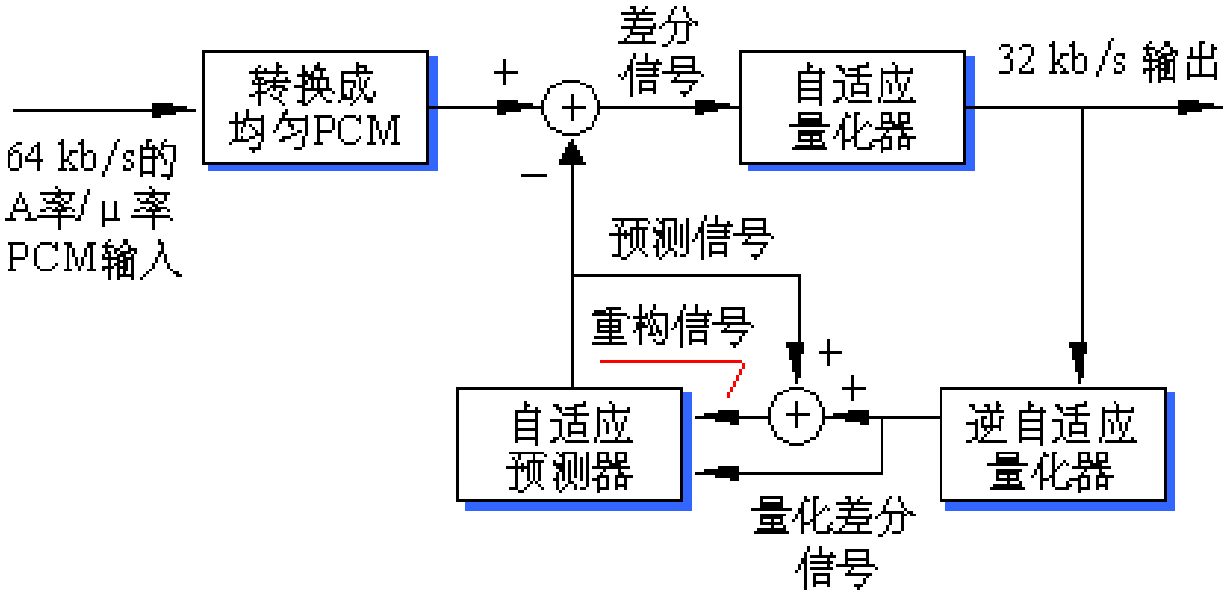
{ynh,cxh}@ustc.edu.cn

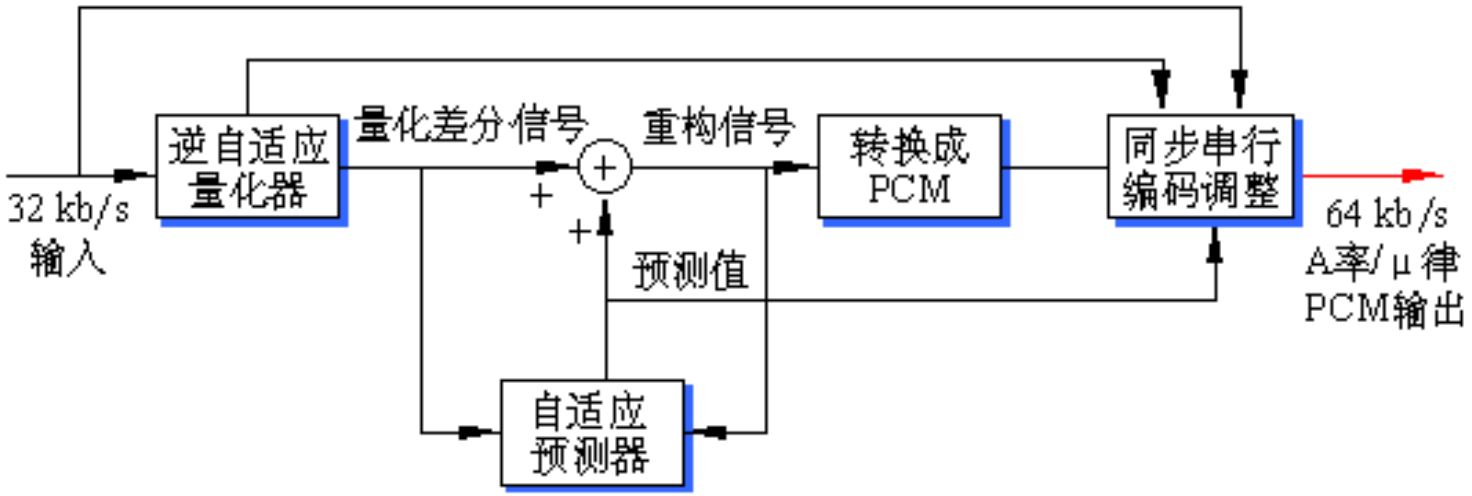
自适应差分脉冲编码调制(ADPCM)



{ynh,cxh}@ustc.edu.cn

G.721





{ynh,cxh}@ustc.edu.cn

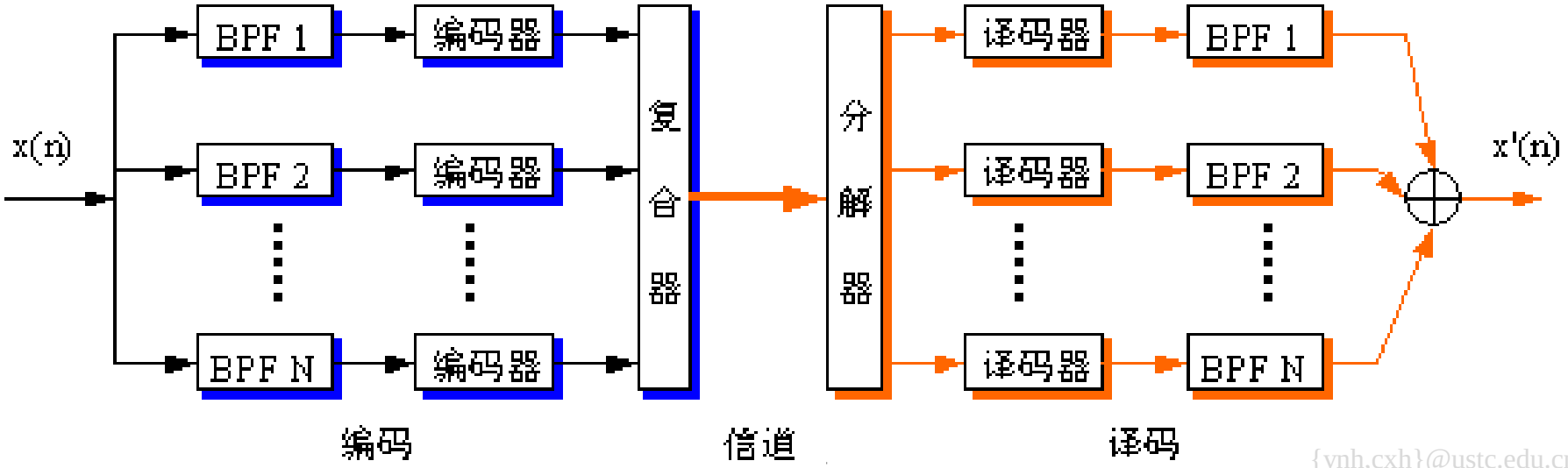
时域和频域的波形编码

- 时域法(time domain approach)
- 频域法(frequency domain approach)
  - **子带编码**(sub-band coding，SBC)，输入的话音信号被分成好几个频带(即子带)，变换到每个子带中的话音信号都进行独立编码。
  - **自适应变换编码**(adaptive transform coding，ATC)。用快速变换(例如离散余弦变换)把话音信号分成许许多多的频带，用来表示每个变换系数的位数取决于话音谱的性质，获得的数据率可低到16 kb/s。

{ynh,cxh}@ustc.edu.cn

子带编码SBC(subband coding)

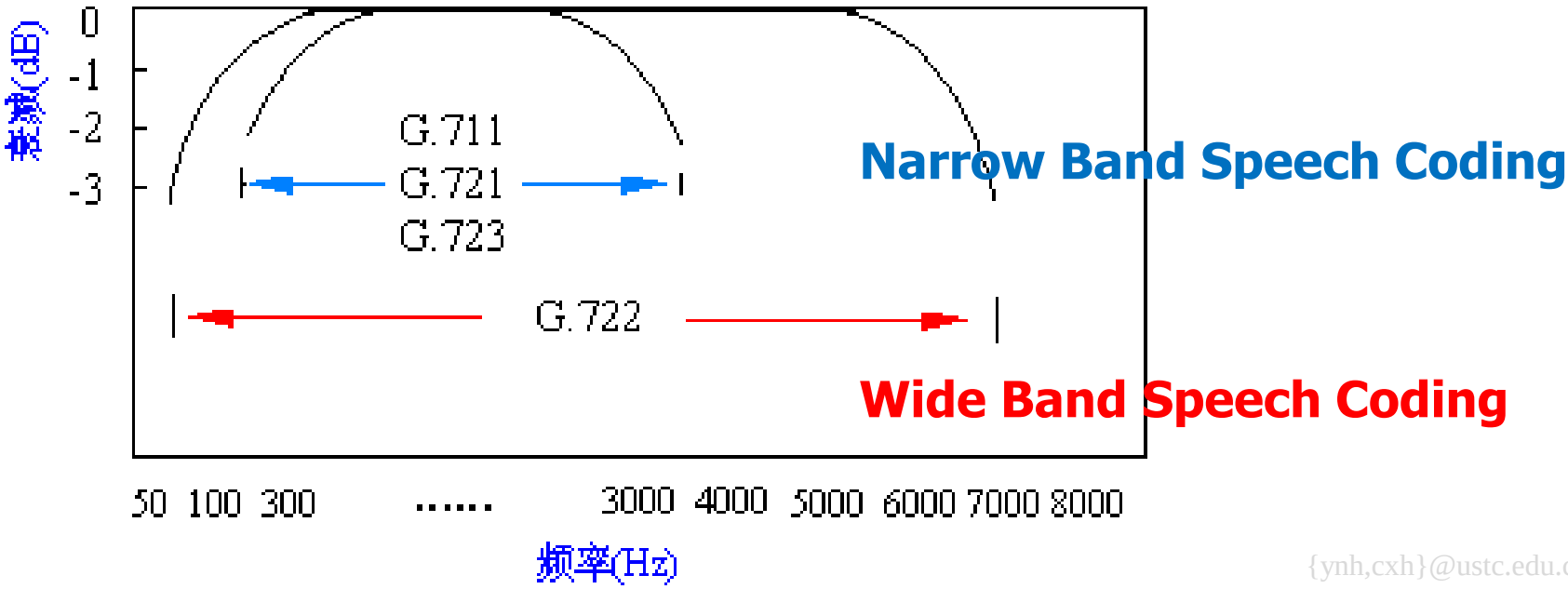
- 用一组带通滤波器BPF(band-pass filter)把输入音频信号的频带分成若干个子带。每个子带中的音频信号采用单独的编码方案编码。
- 传送时，将每个子带的代码复合起来。在接收端译码时，将每个子带的编码单独译码，然后把它们组合起来。



{ynh,cxh}@ustc.edu.cn

子带-自适应差分脉冲编码调制(SB-ADPCM)

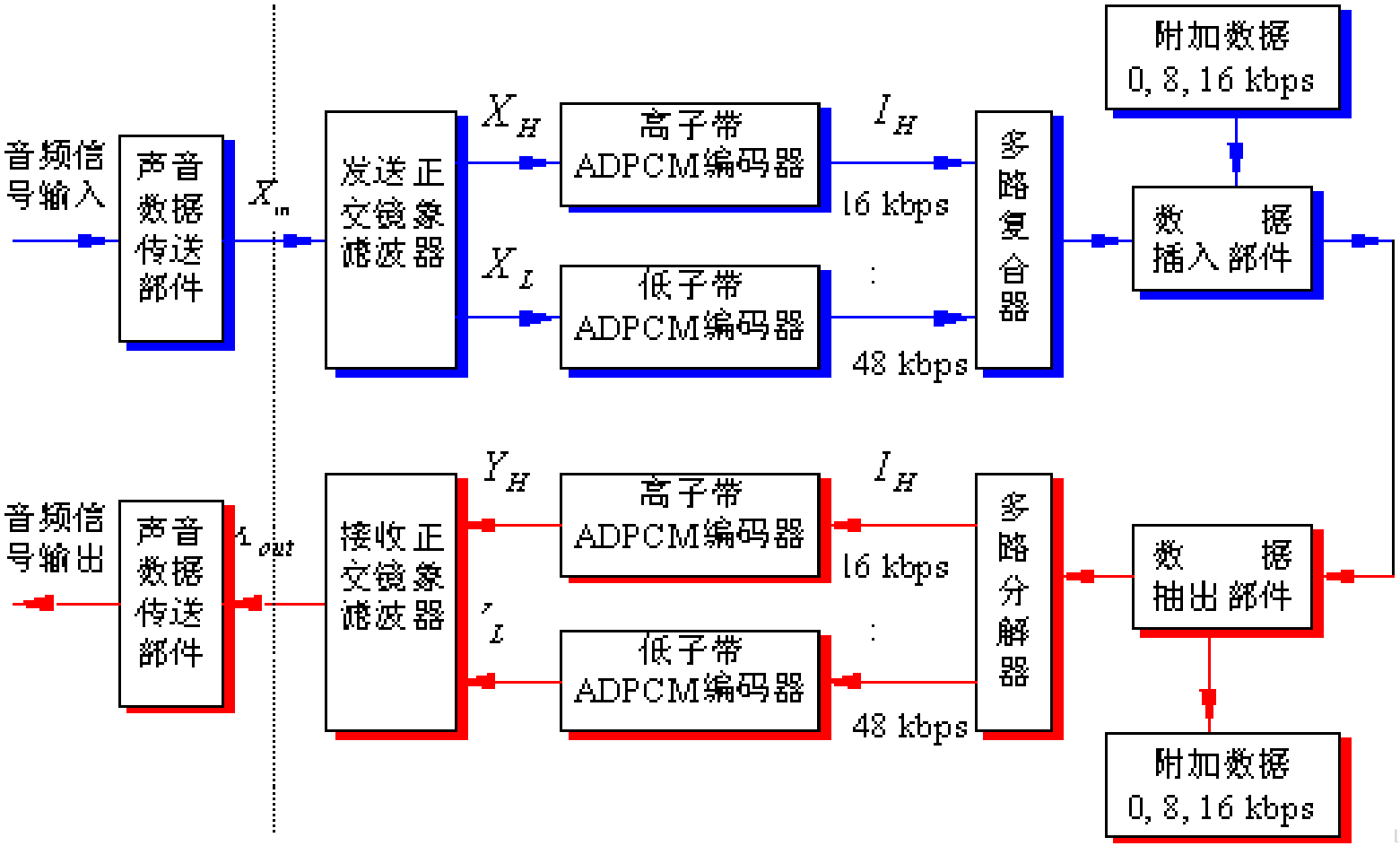
- G.722标准把采样频率由8kHz提高到16kHz
- 信号频率由原来的3.4 kHz扩展到7 kHz
- 低频端把截止频率扩展到50 Hz



{ynh,cxh}@ustc.edu.cn

G.722 SB-ADPCM

- 低频段6bit/sample；高频段2bit/sample

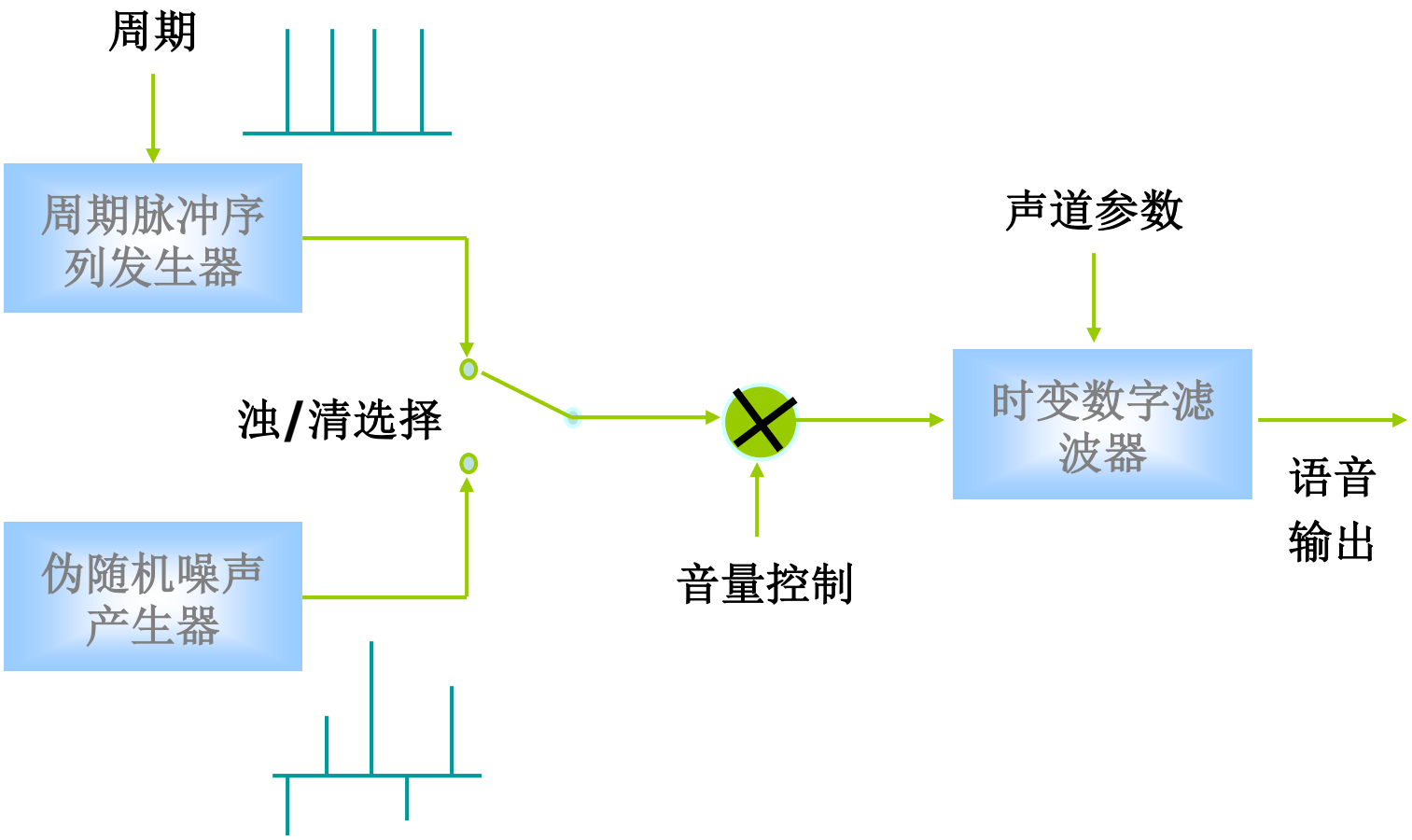


第2章 多媒体数据压缩国际标准

- 2.1 多媒体数据压缩编码的重要性和分类
- 2.2 常见数据压缩方法分类与基本原理
- 2.3 音频压缩标准
  - 2.3.1 话音编码基础
  - 2.3.2 三种话音编码器
    - 波形编译码器、音源编译码器、混合编译码器
  - 2.3.3 MPEG Audio
  - 2.3.4 移动通信网中的音频编码
- 2.4 静态图像压缩编码的国际标准
- 2.5 视频压缩的国际标准
- 2.6 可伸缩性编码和分布式编码

{ynh,cxh}@ustc.edu.cn

话音产生的数字模型



{ynh,cxh}@ustc.edu.cn

音源编译码器

- 通过话音波形的信号中**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。
- 在模型中声道被等效成一个随时间变化的滤波器，叫时变滤波器，激励函数是由白噪声，无声话音段激励或者由有声话音段激励。
- 数据率2.4bps，产生的语音质量很低，可以听懂而已。增加数据率对于话音质量没有用，因为这是由模型限制的，但保密性好。

{ynh,cxh}@ustc.edu.cn

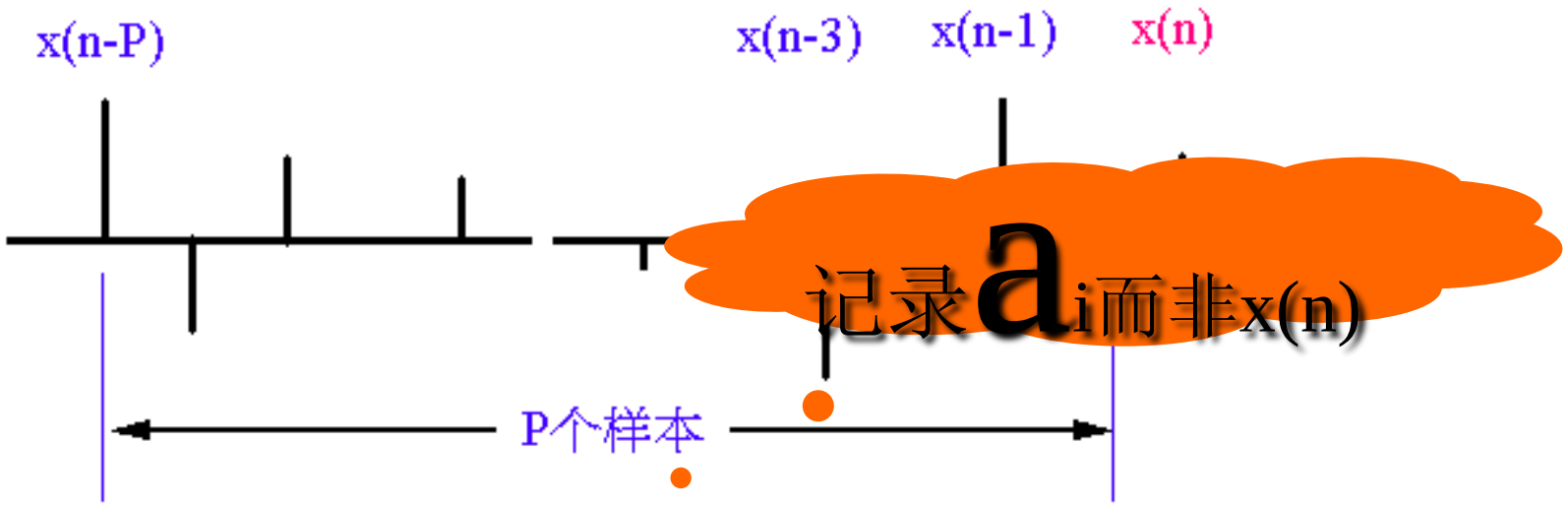
音源编译码起因

- 一般的语音传输每隔20ms传输一次
  - 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。
- 波形编码的数据量大
  - 20ms的CD音乐的存储量
    - 20ms/1000ms*44.1k*2byte*2 = 3.528kB
  - 20ms的G.721的存储量
    - 20ms/1000ms*64kbps = 1.28kb
- 用声道参数表示声音
  - LPC速率2.4kbps(平均20ms传输48bit)

{ynh,cxh}@ustc.edu.cn

线性预测编码(LPC)

- 发送端产生声道激励和转移函数的参数
- 接收端通过话音合成器重构话音
- 随着话音波形的变化，周期性地使模型的参数和激励条件适合新的要求。



$$x_{pre}(n) = -[a_1 x(n-1) + a_2 x(n-2) + \cdots\cdots + a_p (n-p)]$$

{ynh,cxh}@ustc.edu.cn

第2章 多媒体数据压缩国际标准

- 2.1 多媒体数据压缩编码的重要性和分类
- 2.2 常见数据压缩方法分类与基本原理
- 2.3 音频压缩标准
  - 2.3.1 话音编码基础
  - 2.3.2 三种话音编码器
    - 波形编译码器、音源编译码器、混合编译码器
  - 2.3.3 MPEG Audio
  - 2.3.4 移动通信网中的音频编码
- 2.4 静态图像压缩编码的国际标准
- 2.5 视频压缩的国际标准
- 2.6 可伸缩性编码和分布式编码

{ynh,cxh}@ustc.edu.cn

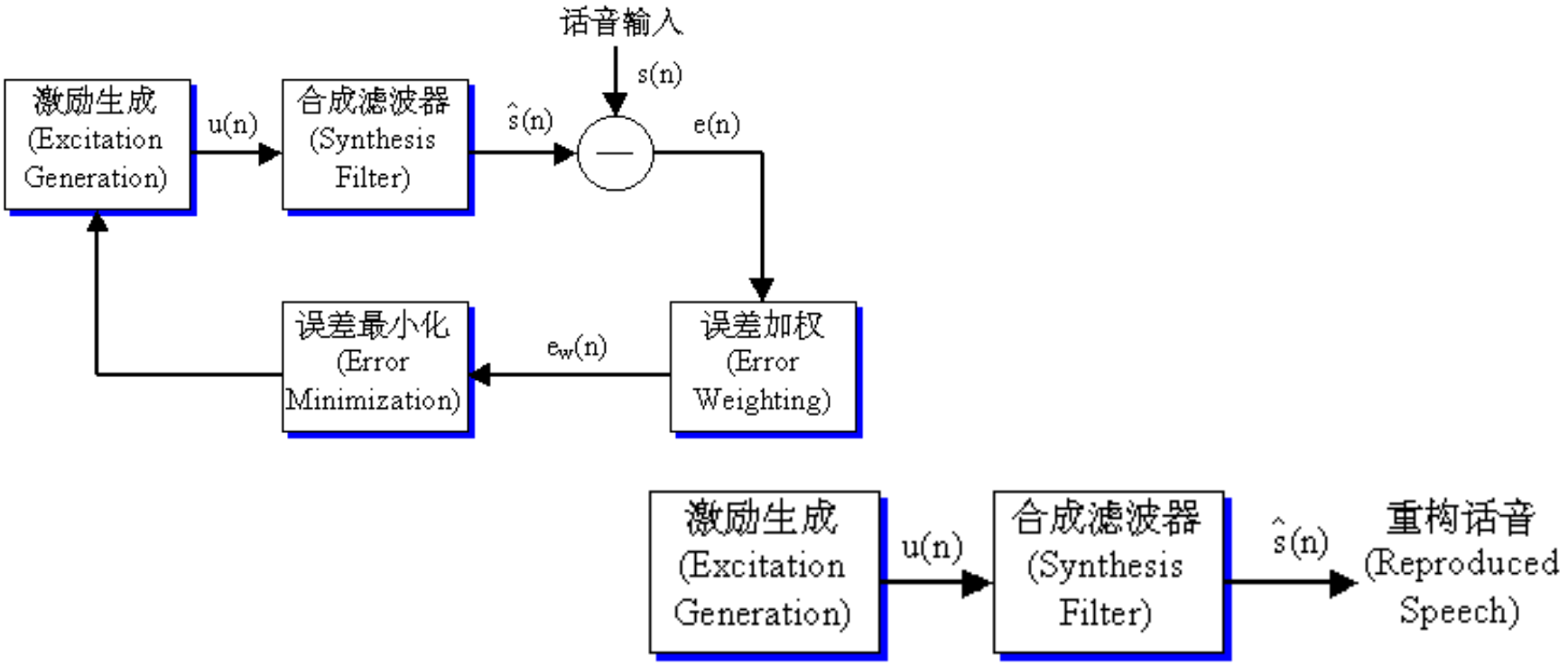
混合编译码

- 波形编译码器虽然可提供高话音的质量，但数据率低于16 kb/s的情况下，在技术上还没有解决音质的问题；
- 音源编码器的数据率虽然可降到2.4 kb/s甚至更低，但它的音质根本不能与自然话音相提并论。
- 混合编译码的想法是企图填补波形编译码和音源编译码之间的间隔。
- 历史上出现过很多形式的混合编译码器，但最成功并且普遍使用的编译码器是时域合成-分析(analysis-by-synthesis，AbS)编译码器。

{ynh,cxh}@ustc.edu.cn

时域合成-分析编码 (AbS)

◆使用的声道线性预测滤波器模型与LPC使用的模型相同，不使用两个状态(有声/无声)的模型来寻找滤波器的输入激励信号，而是企图寻找这样一种激励信号，使用这种信号激励产生的波形尽可能接近于原始话音的波形。



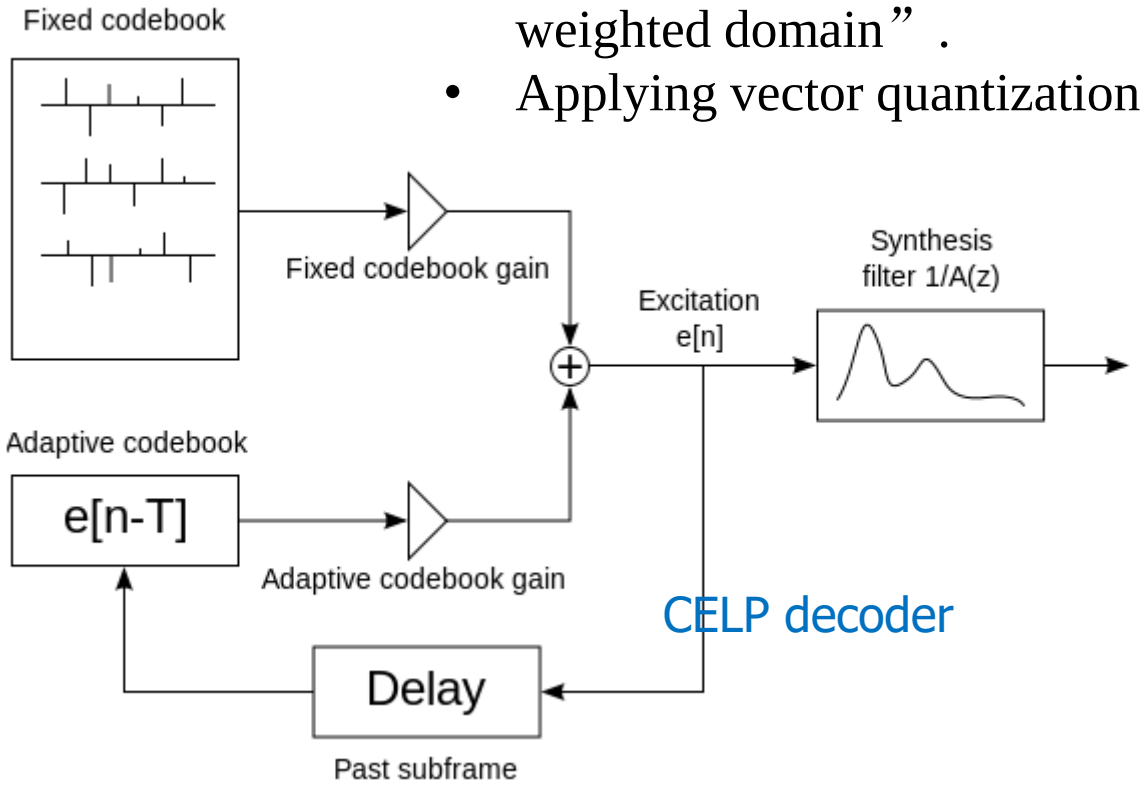
AbS的激励

- 多脉冲激励MPE (multi-pulse excited)
  - 每5ms使用4个脉冲，每个脉冲的位置和幅度由编码器确定
- 等间隔脉冲激励RPE (regular-pulse excited)
  - 每5ms用10个脉冲，记录第一个脉冲的位置和所有脉冲的幅度
- 码激励线性预测CELP(code excited linear predictive)
  - 激励信号由一个矢量量化大码簿的表项给出

{ynh,cxh}@ustc.edu.cn

Code-excited linear prediction (CELP)

- The CELP algorithm is based on four main ideas:
  - Using the source-filter model of speech production through linear prediction (LP);
  - Using an adaptive and a fixed codebook as the input (excitation);
  - Performing a search in closed-loop in a “perceptually weighted domain”.
  - Applying vector quantization (VQ)



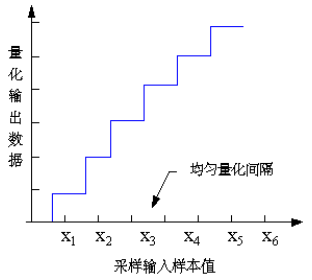
{ynh,cxh}@ustc.edu.cn

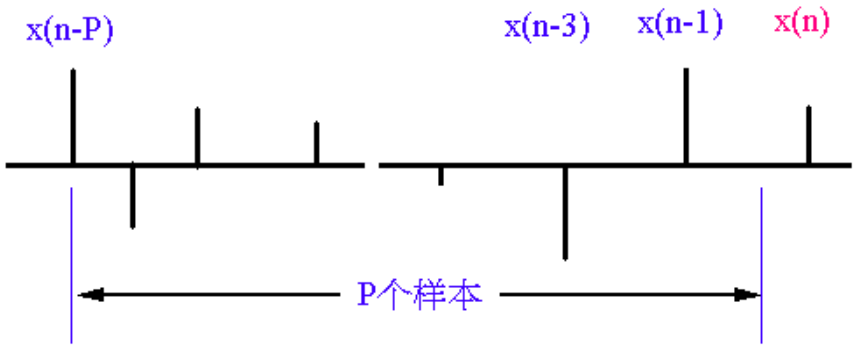
小结：两种思路

Waveform Model

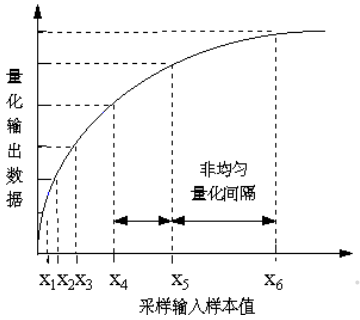
Source-Filter Model

- 用一个波形逼近原始信号
  - 产生一种重构信号，其波形和原始波形尽可能一致
  - PCM
- 用数学模型模拟
  - 利用生成话音信号的知识产生
  - LPC





$$x_{pre}(n) = -[a_1 x(n-1) + a_2 x(n-2) + \cdots\cdots + a_p (n-p)]$$



第2章 多媒体数据压缩国际标准

- 2.1 多媒体数据压缩编码的重要性和分类
- 2.2 常见数据压缩方法分类与基本原理
- 2.3 音频压缩标准
  - 2.3.1 话音编码基础
  - 2.3.2 三种话音编码器
  - 2.3.3 MPEG Audio
  - 2.3.4 移动通信网中的音频编码
- 2.4 静态图像压缩编码的国际标准
- 2.5 视频压缩的国际标准
- 2.6 可伸缩性编码和分布式编码

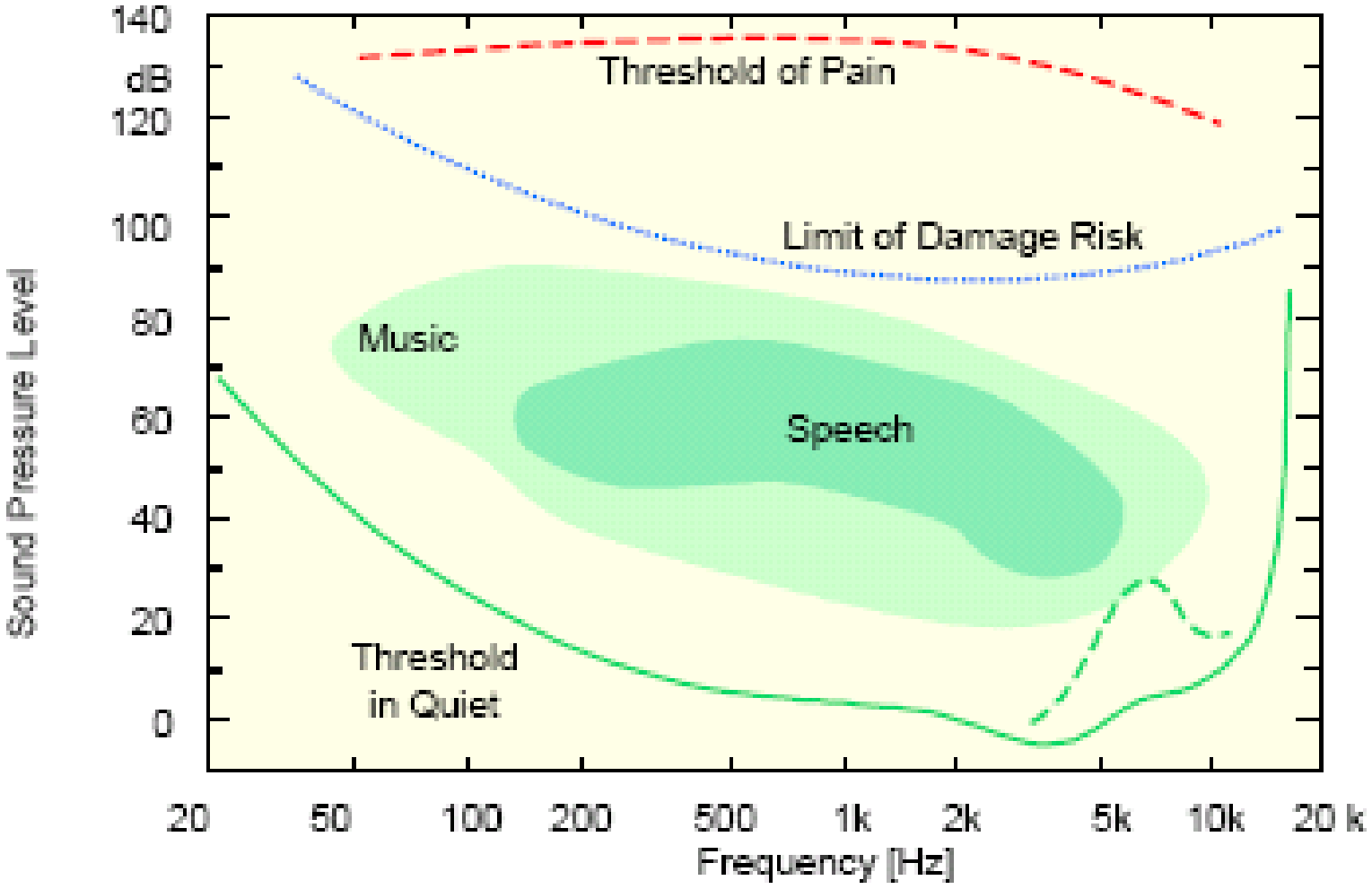
{ynh,cxh}@ustc.edu.cn

人的听觉感知机理

- 人的听觉具有掩蔽效应
  - 当几个强弱不同的声音同时存在时，强声使弱声难以听见的现象称为**同时掩蔽**，它受掩蔽声音和被掩蔽声音之间的相对频率关系影响很大；
  - 声音在不同时间先后发生时，强声使其周围的弱声难以听见的现象称为**异时掩蔽**。
- 人耳对不同频段的声音的敏感程度不同
  - 通常对低频端较之对高频端更敏感。既使是对同样声压级的声音，人耳的实际感觉到的音量也是随频率而变化的。
- 人耳对语音信号的相位变化不敏感
  - 人耳听不到或感知极不灵敏的声音分量都视为冗余

{ynh,cxh}@ustc.edu.cn

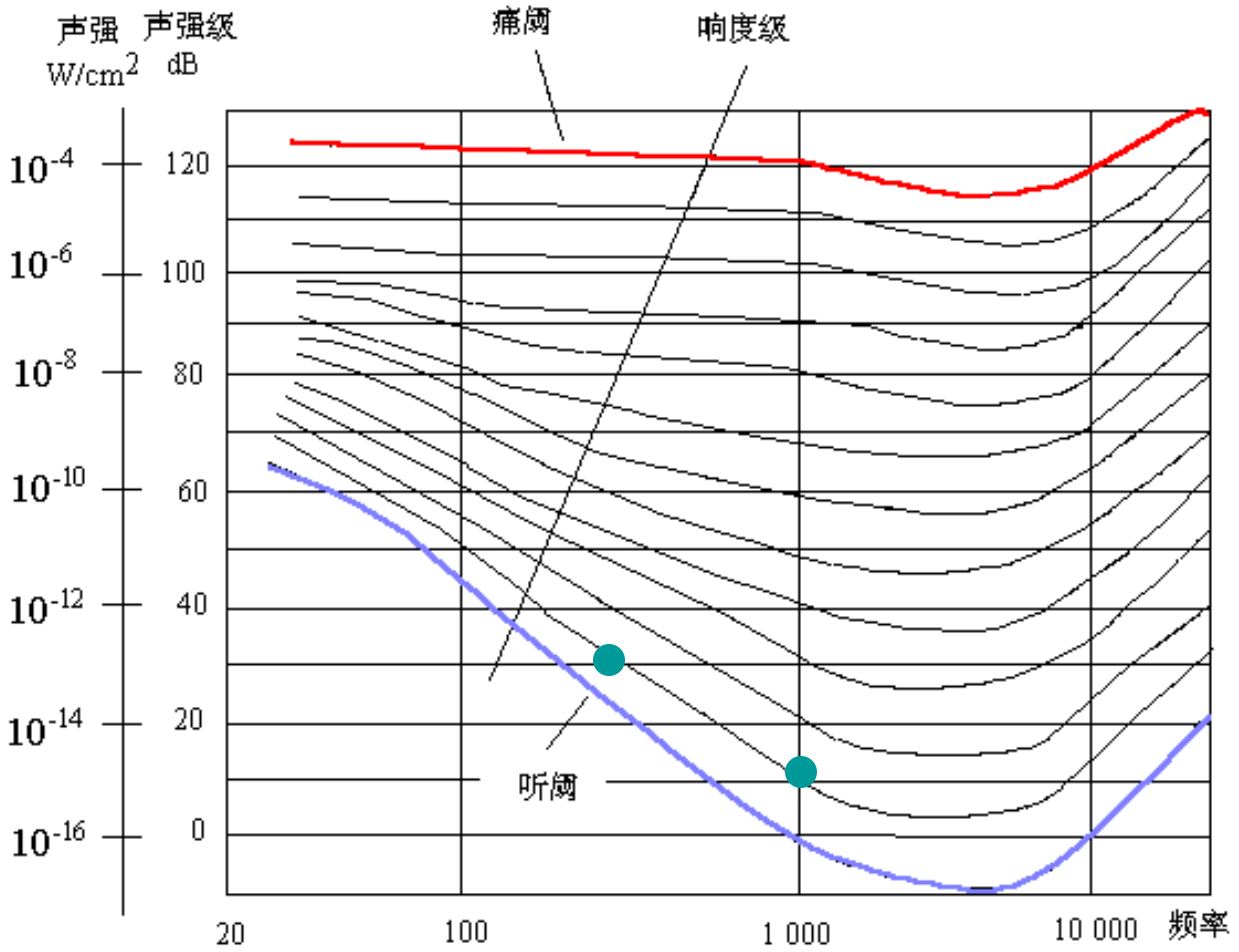
Sound Pressure Level



http://www.transport07.co.nz/uploads/File/ppt/C%20Roberts.ppt

{ynh,cxh}@ustc.edu.cn

“听阈—频率”和“痛阈—频率”曲线



◆1 kHz的10 dB的声音和200 Hz的30 dB的声音，在人耳听起来具有相同的响度

{ynh,cxh}@ustc.edu.cn

听觉掩蔽效应

- 一种频率的声音阻碍听觉系统感受另一种频率的声音的现象称为掩蔽效应。
- 频域掩蔽：同时发出的频率接近的两个纯音，声强低的纯音会被声强高的纯音淹没
- 时域掩蔽：在时间上相邻的声音之间也有掩蔽现象，称为时域掩蔽。产生的主要原因是人的大脑处理信息需要花费一定的时间。

{ynh,cxh}@ustc.edu.cn

时域掩蔽

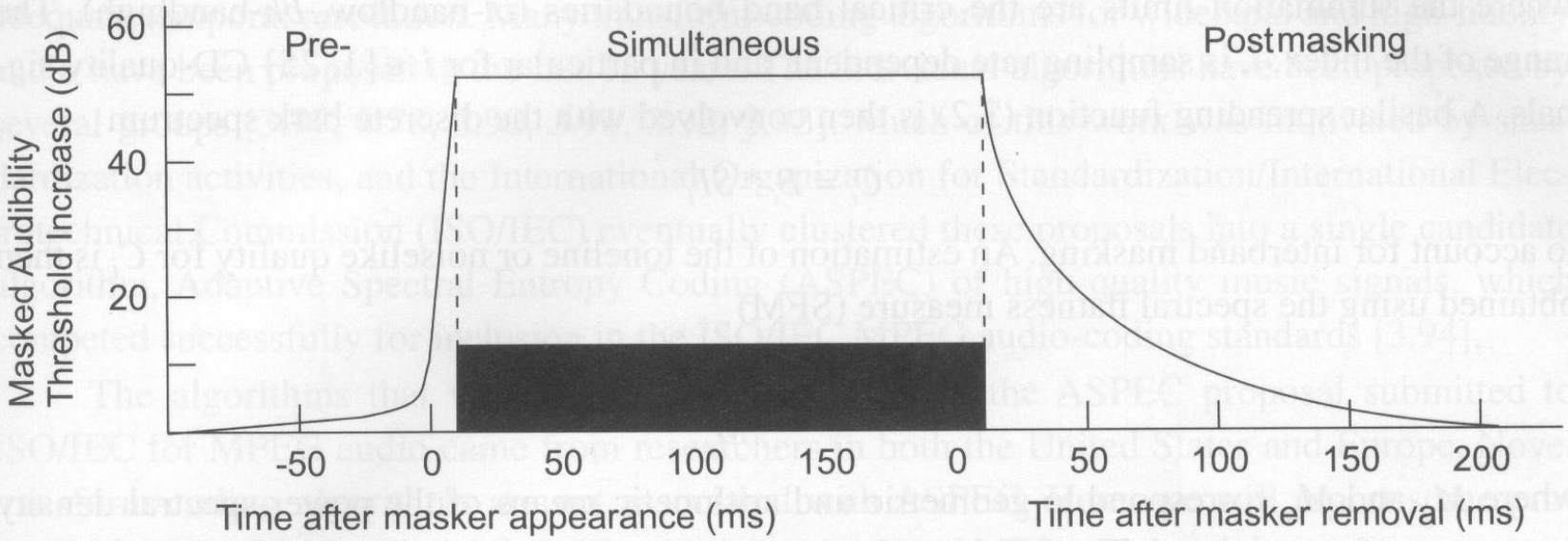
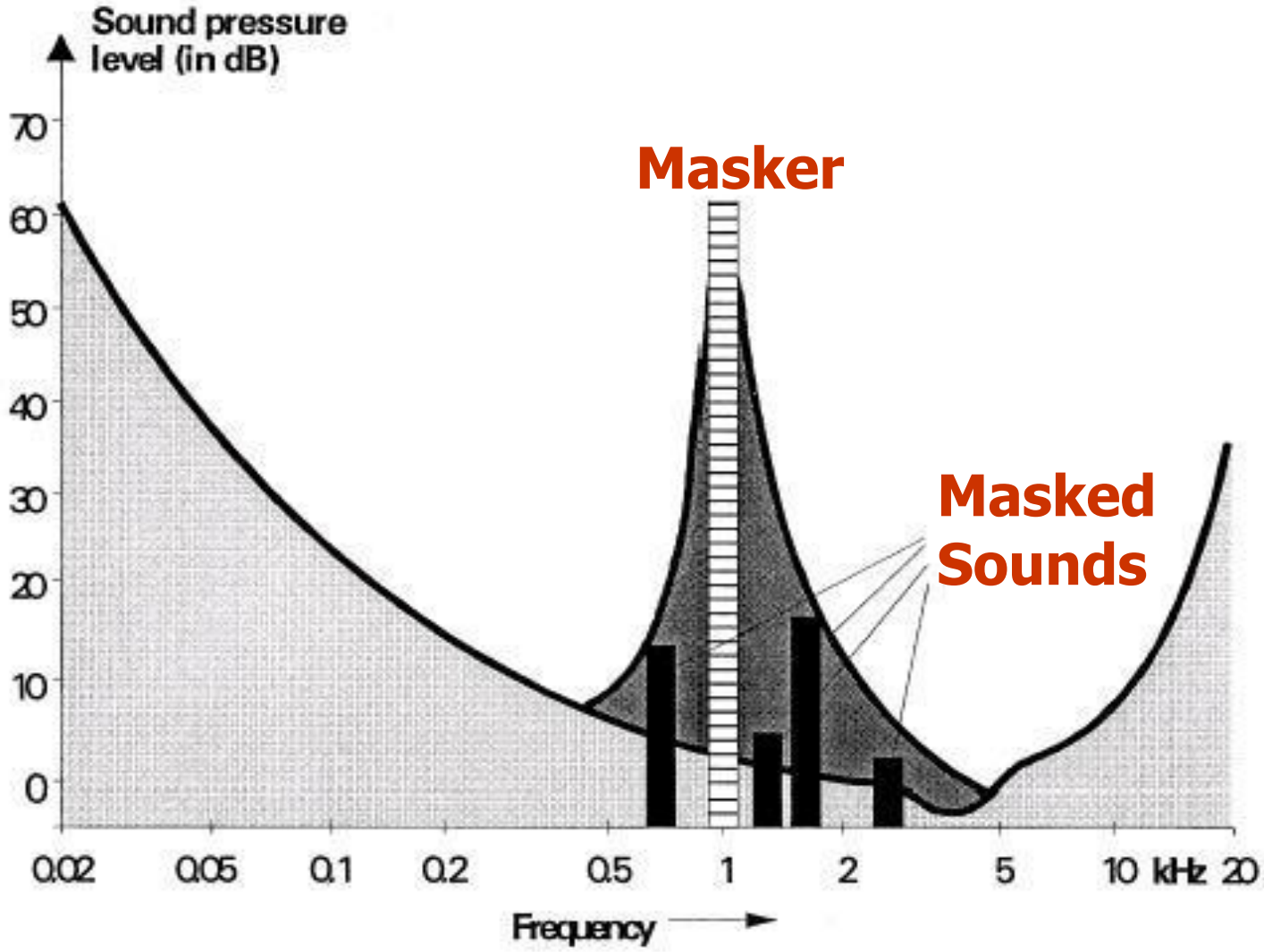


Figure 3.7 Schematic representation of temporal masking properties of the human ear [3.82]. ©1997 IEEE.

{ynh,cxh}@ustc.edu.cn

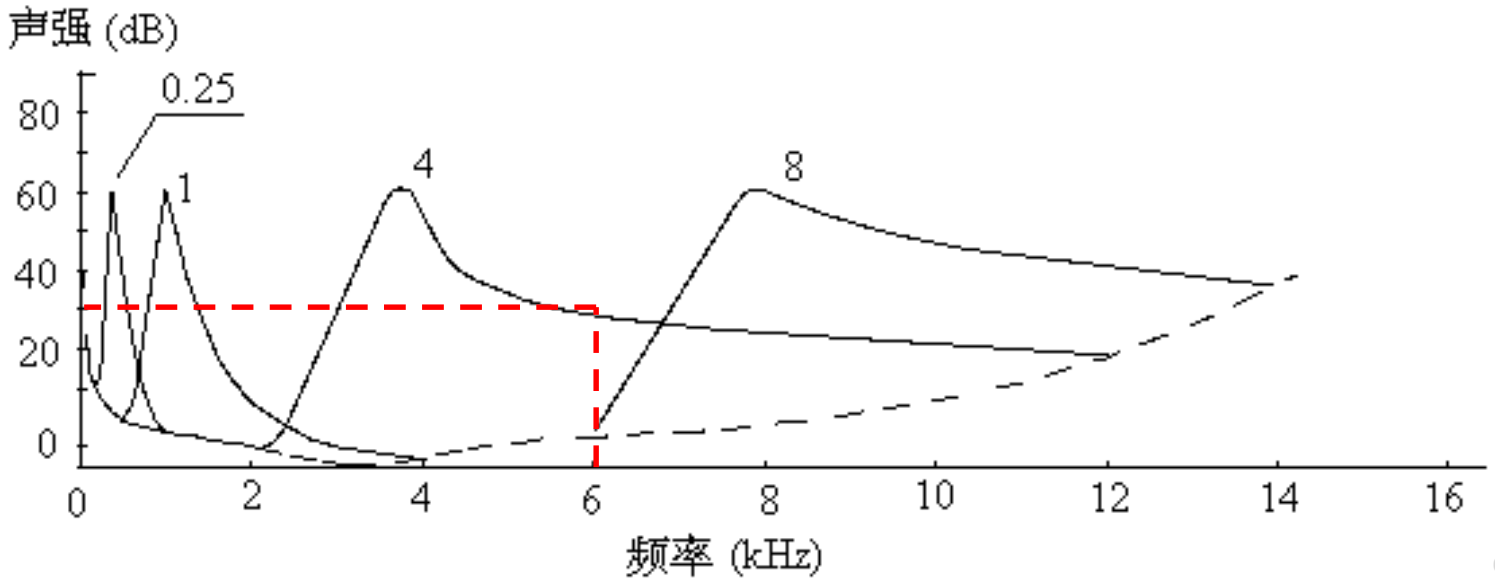
频域掩蔽



不同纯音的掩蔽效应曲线

◆图中的一组曲线分别表示频率为250 Hz、1 kHz、4 kHz和8 kHz纯音的掩蔽效应。从图中可以看到：

- ①在纯音附近，对其他纯音的掩蔽效果最明显，
- ②低频纯音可以有效地掩蔽高频纯音，但高频纯音对低频纯音的掩蔽作用则不明显。



临界频带

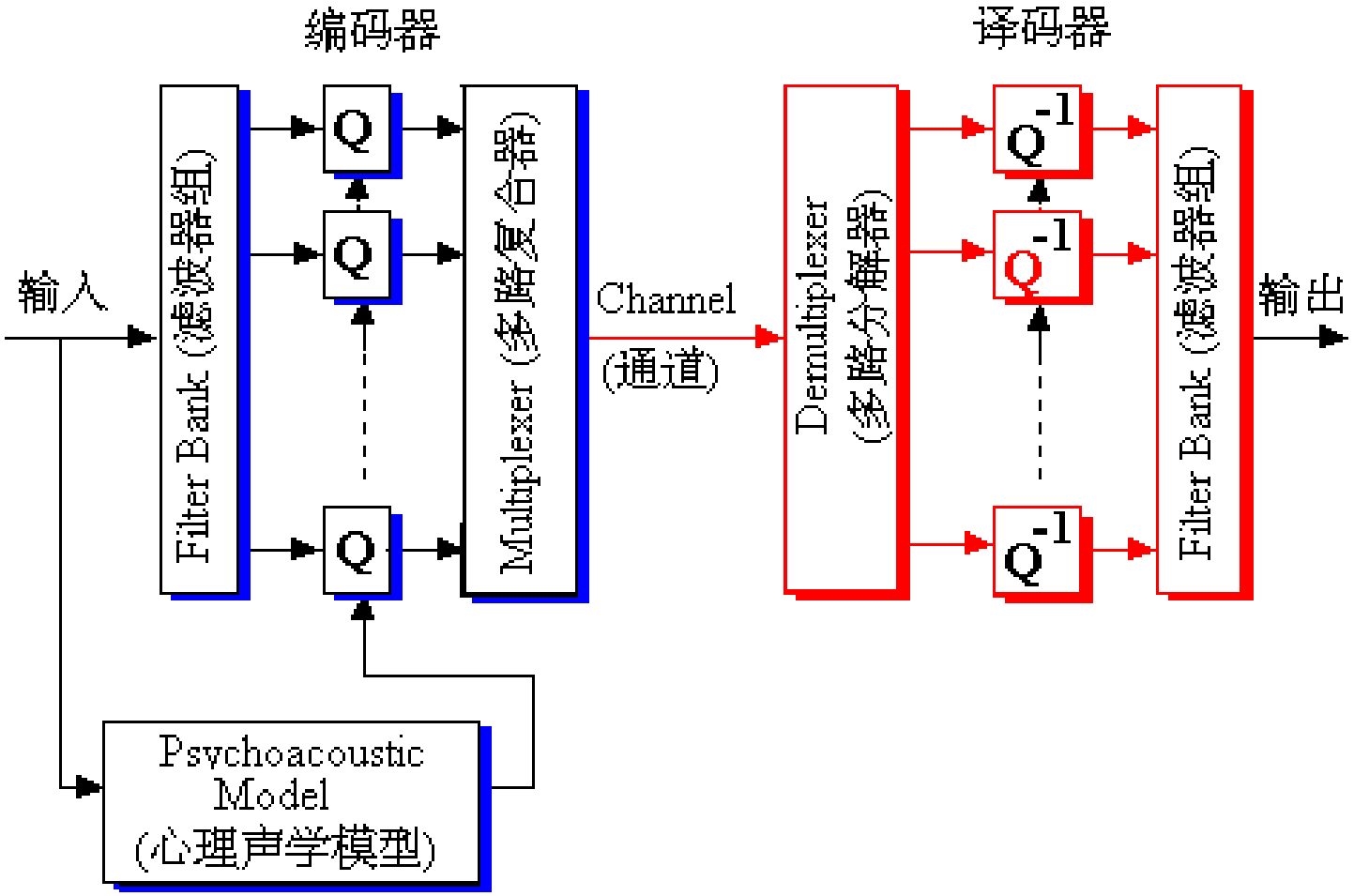
临界	频率（Hz）			临界	频率（Hz）		
频带	低端	高端	宽度	频带	低端	高端	宽度
0	0	100	100	13	2000	2320	320
1	100	200	100	14	2320	2700	380
2	200	300	100	15	2700	3150	450
3	300	400	100	16	3150	3700	550
4	400	510	110	17	3700	4400	700
5	510	630	120	18	4400	5300	900
6	630	770	140	19	5300	6400	1100
7	770	920	150	20	6400	7700	1300
8	920	1080	160	21	7700	9500	1800
9	1080	1270	190	22	9500	12000	2500
10	1270	1480	210	23	12000	15500	3500
11	1480	1720	240	24	15500	22050	6550
12	1720	2000	280				

利用感知特性进行压缩编码

- 人耳对不同临界频带感知不同，可以将临界频带分成子带采用不同的量化阶
- 听觉系统中存在一个听阈电平，低于这个电平的声音信号就听不到，因此就可以把这部分信号去掉。
- 利用掩蔽效应将人耳无法感知的频率分量消除

{ynh,cxh}@ustc.edu.cn

MPEG Audio压缩算法框图



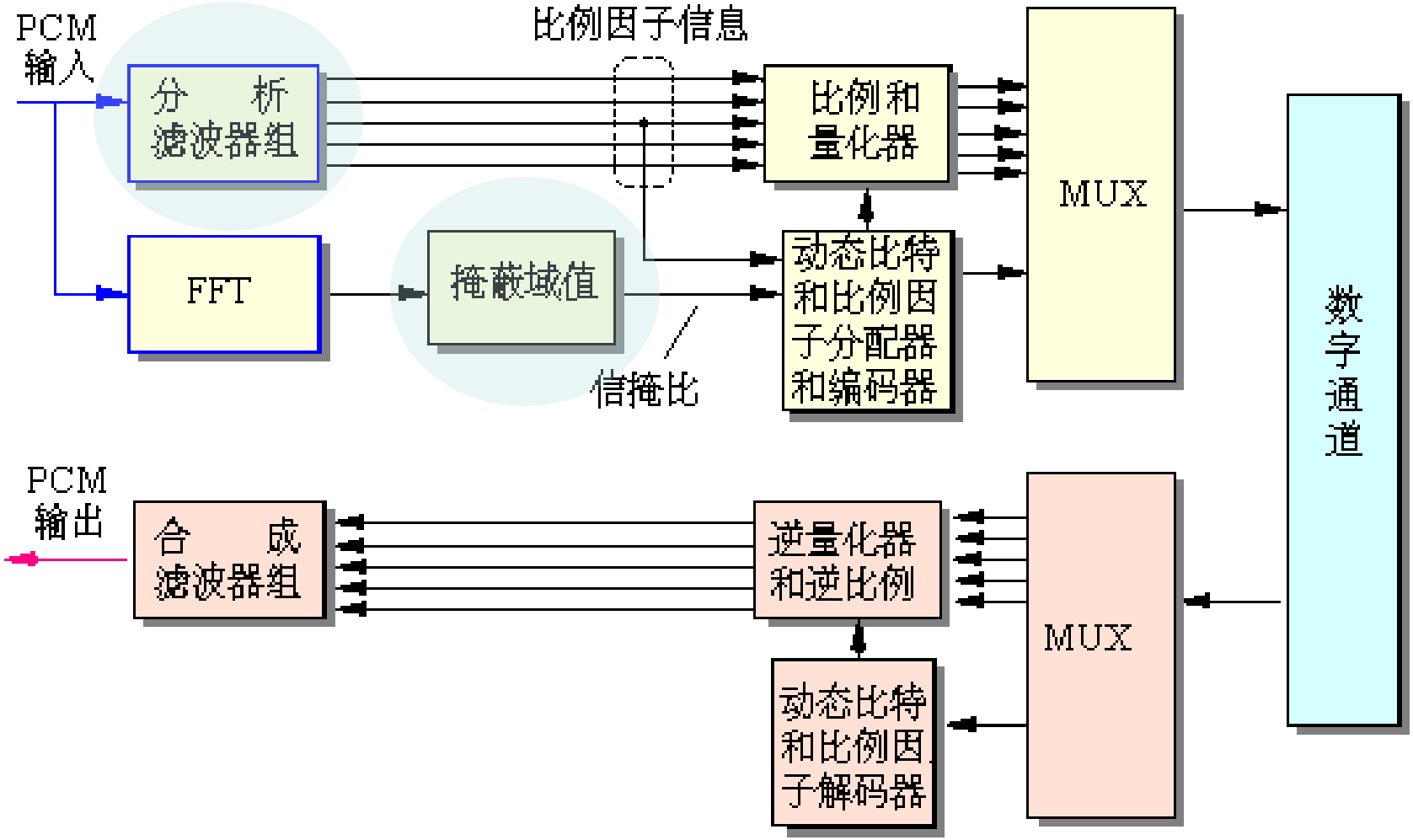
{ynh,cxh}@ustc.edu.cn

MPEG1 Audio

- Layer1
  - 频带相等的子带，使用频域掩蔽特性， 每个子带用6bit量化。
- Layer2
  - 频带相等的子带，除了使用频域掩蔽特性之外还利用了时间掩蔽特性，低频段的子带用4比特，中频段的子带用3比特，高频段的子带用2比特。
- Layer3
  - 使用比较好的临界频带滤波器，把声音频带分成非等带宽的子带，除了使用频域掩蔽特性和时间掩蔽特性之外，还考虑了立体声数据的冗余，并且使用了霍夫曼(Huffman)编码器。还使用了MDCT (modified discrete cosine transform) 把子带的输出在频域里进一步细分以达到更高的频域分辨率。

{ynh,cxh}@ustc.edu.cn

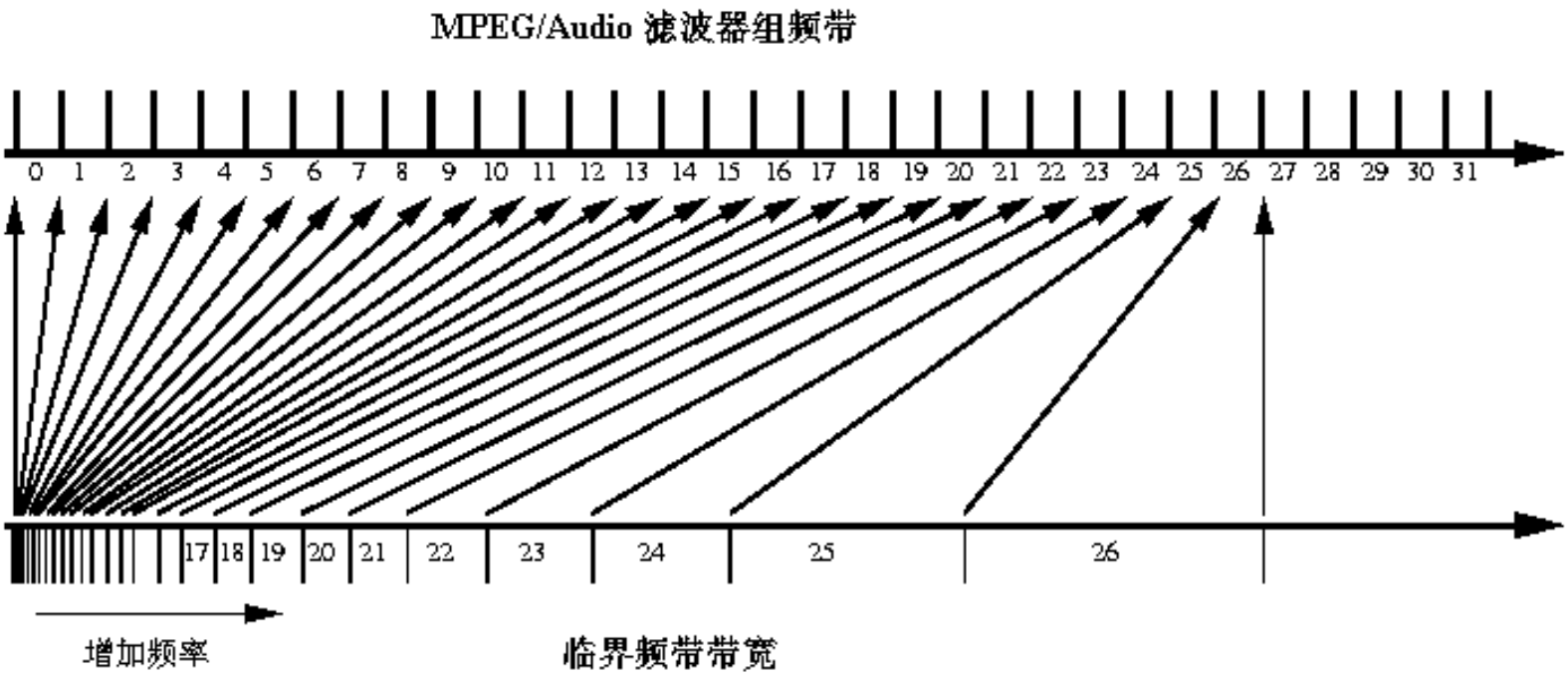
MPEG1 audio层1和层2编解码



{ynh,cxh}@ustc.edu.cn

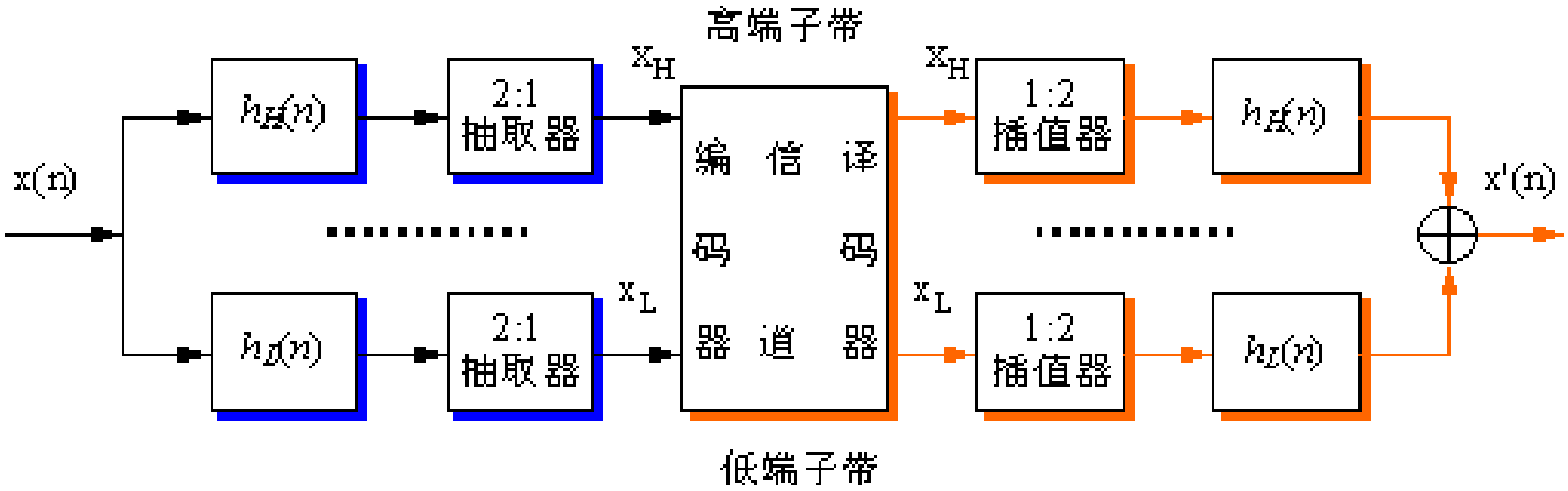
子带的划分

- 方法有两种，一种是线性划分，另一种是非线性划分。
- 如果把声音频带划分成带宽相等的子带，这种划分就不能精确地反映人耳的听觉特性，因为人耳的听觉特性是以“临界频带”来划分的，在一个临界频带之内，很多心理声学特性都是一样的。



子带分割

◆把音频信号分割成相邻的子带分量之后，用2倍于子带带宽的采样频率对子带信号进行采样，就可以用它的样本值重构出原来的子带信号。

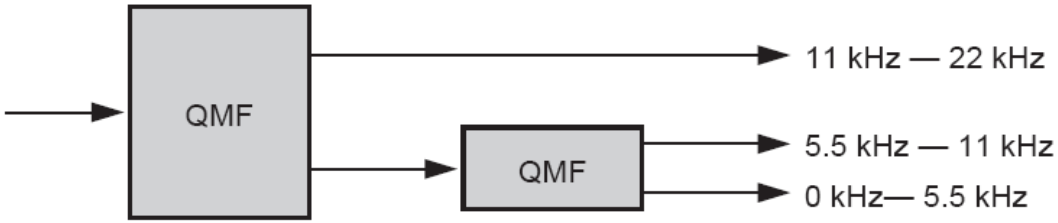


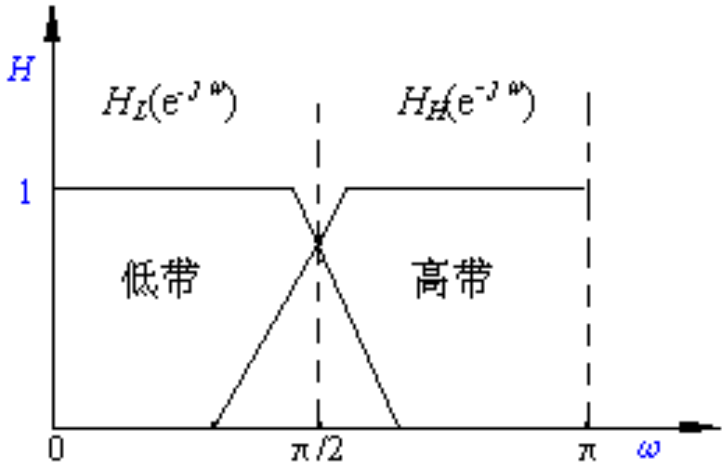
{ynh,cxh}@ustc.edu.cn

QMF

◆由于分割频带所用的滤波器不是理想的滤波器，经过分带、编码、译码后合成的输出音频信号会有**混迭效应**。

◆采用正交镜象滤波器(quandrature mirror filter，QMF)来划分频带，混迭效应在最后合成时可以抵消。





《信号与系统》，徐守时，p504-p507

https://en.wikipedia.org/wiki/Modified_discrete_cosine_transform

MDCT

In particular, it is a linear function $F: \mathbf{R}^{2N} \to \mathbf{R}^{N}$ (where **R** denotes the set of real numbers). The $2N$ real numbers $x_0, \ldots, x_{2N-1}$ are transformed into the N real numbers $X_0, \ldots, X_{N-1}$ according to the formula:

$$X_k = \sum_{n=0}^{2N-1} x_n \cos\left[\frac{\pi}{N}\left(n + \frac{1}{2} + \frac{N}{2}\right)\left(k + \frac{1}{2}\right)\right]$$

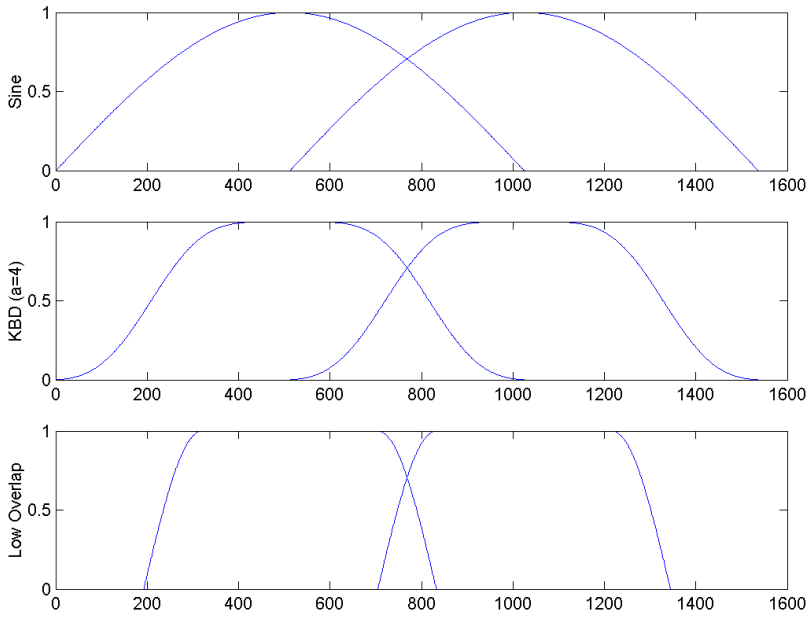
The inverse MDCT is known as the **IMDCT**. Because there are different numbers of inputs and outputs, at first glance it might seem that the MDCT should not be invertible. However, perfect invertibility is achieved by *adding* the overlapped IMDCTs of subsequent overlapping blocks, causing the errors to *cancel* and the original data to be retrieved; this technique is known as *time-domain aliasing cancellation* (**TDAC**).

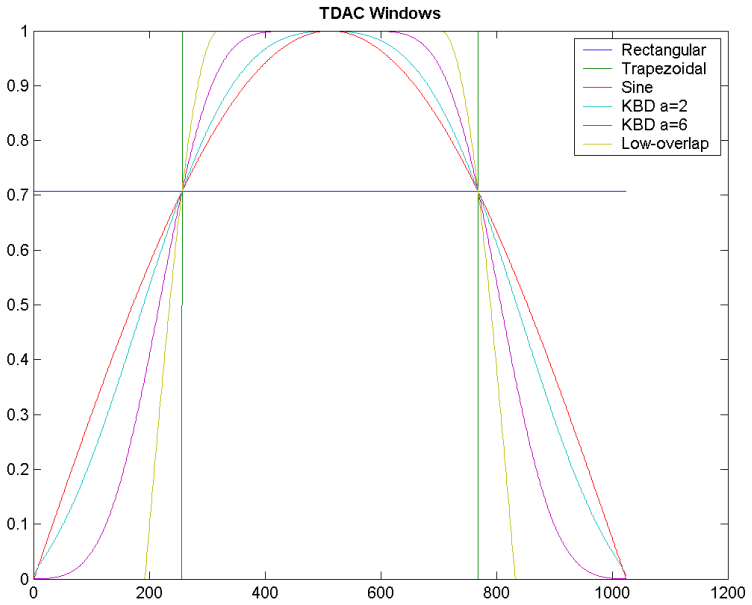
The IMDCT transforms N real numbers $X_0, \ldots, X_{N-1}$ into $2N$ real numbers $y_0, \ldots, y_{2N-1}$ according to the formula:

$$y_n = \frac{1}{N} \sum_{k=0}^{N-1} X_k \cos\left[\frac{\pi}{N}\left(n + \frac{1}{2} + \frac{N}{2}\right)\left(k + \frac{1}{2}\right)\right]$$

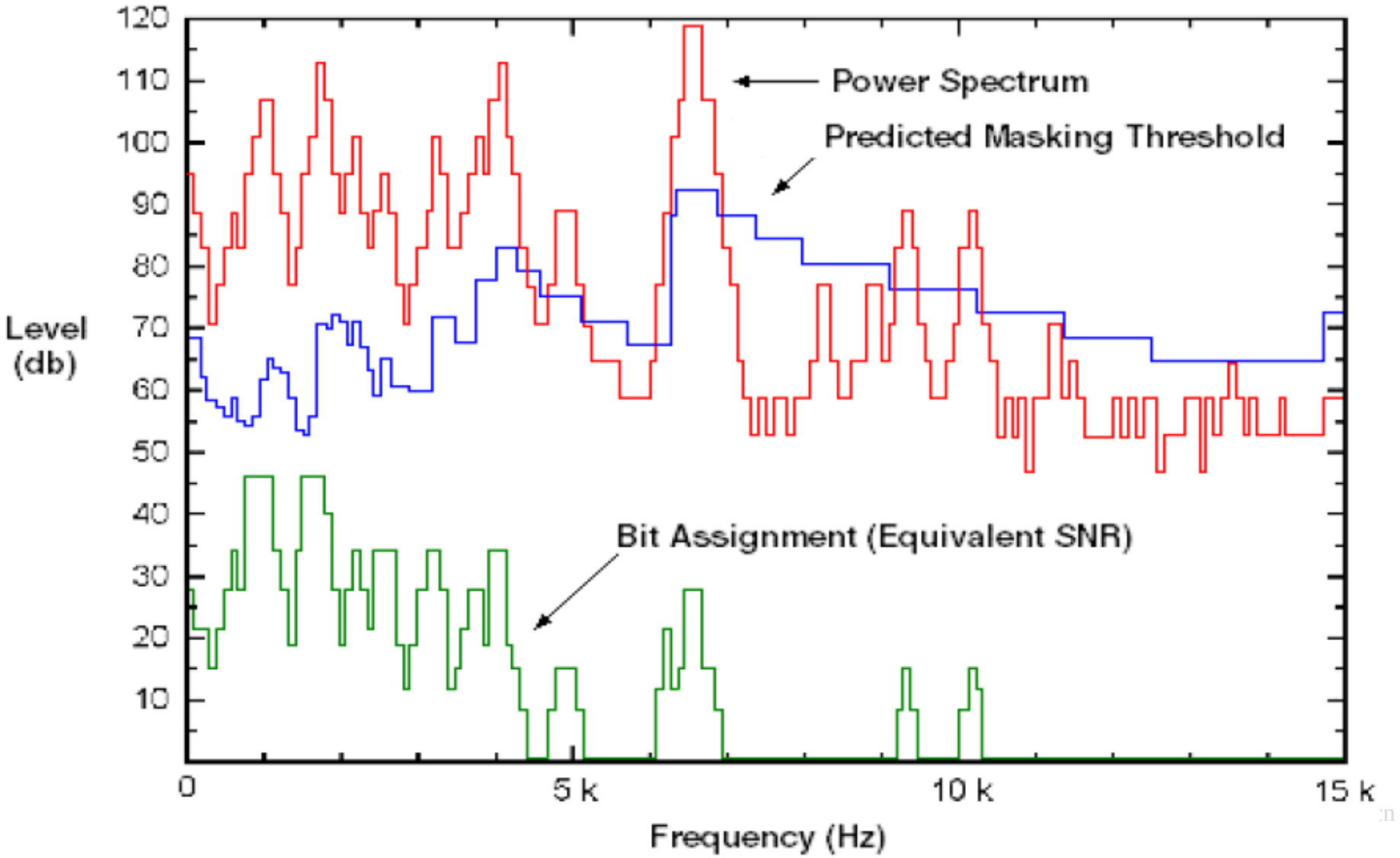
Kaiser-Bessel Derived (KBD) window

- A 50% overlap add (OLA) structure with certain pre and post, time domain aliasing cancellation (TDAC) windowing, the initial signal can be completely recovered.

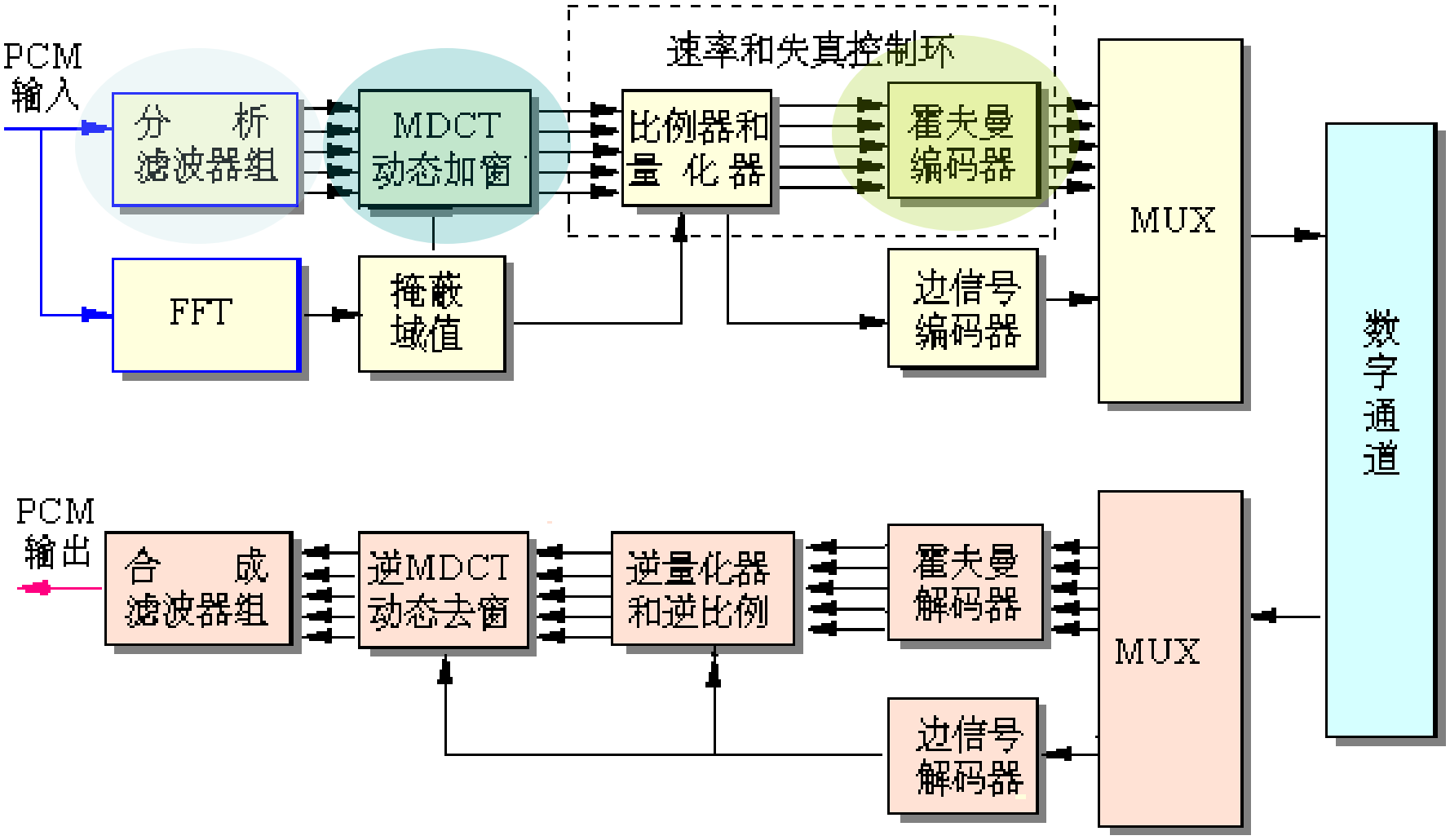




掩蔽特性应用举例


120
110
100
90
80
70
60
50
40
30
20
10
Level (db)
Power Spectrum
Predicted Masking Threshold
Bit Assignment (Equivalent SNR)
0
5 k
10 k
15 k
Frequency (Hz)

MP3编解码

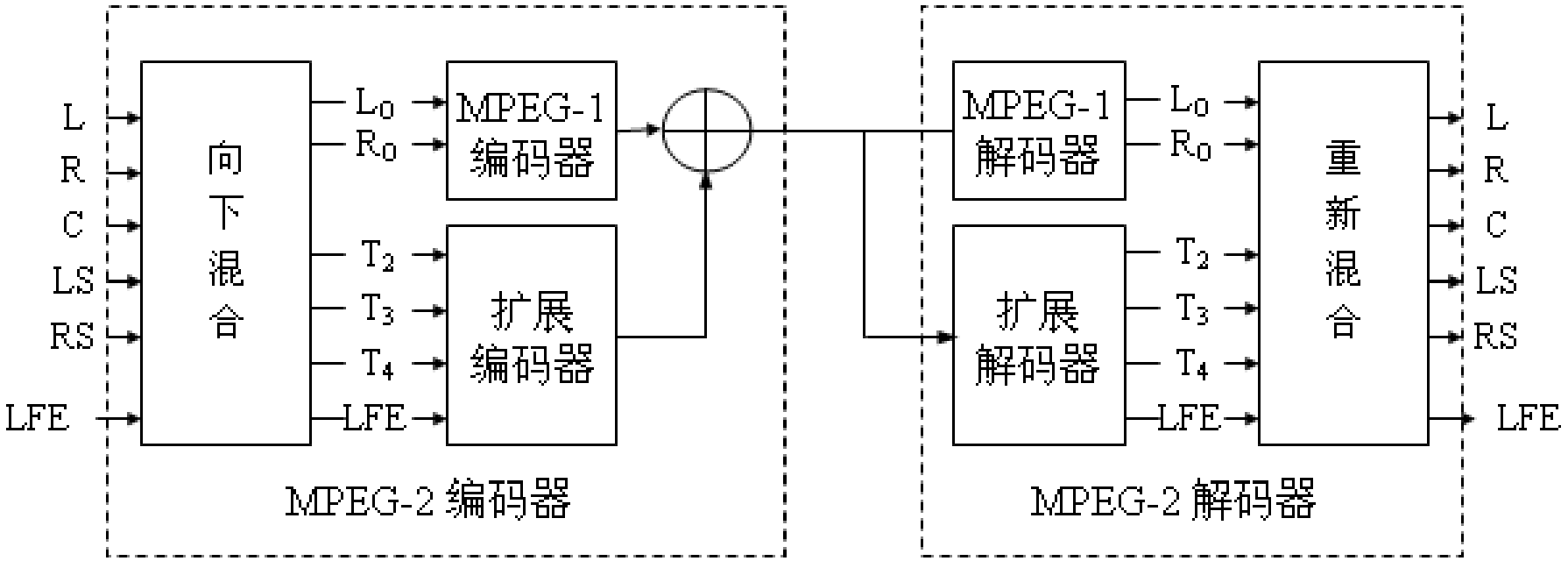

PCM
输入
分 析
滤波器组
MDCT
动态加窗
速率和失真控制环
比例器和
量 化 器
霍夫曼
编码器
MUX
数字通道
FFT
掩蔽
域值
边信号
编码器
PCM
输出
合 成
滤波器组
逆MDCT
动态去窗
逆量化器
和逆比例
霍夫曼
解码器
MUX
边信号
解码器

MPEG2 Audio

- MPEG-2 BC (Backward Compatible)
  - 增加了16 kHz, 22.05 kHz和24 kHz采样频率
  - 输出速率由32～384 kb/s扩展到8～640 kb/s
  - 支持5.1声道和7.1声道的环绕声
  - 支持Linear PCM(线性PCM)和Dolby AC-3(Audio Code Number 3)编码
- MPEG-2 AAC (Advanced Audio Coding)
  - 利用掩蔽特性减少数据量，并把量化噪声分散到各个子带中，用全局信号把噪声掩蔽掉。
  - 采用频率可从8 kHz到96 kHz，可支持声道数目极多

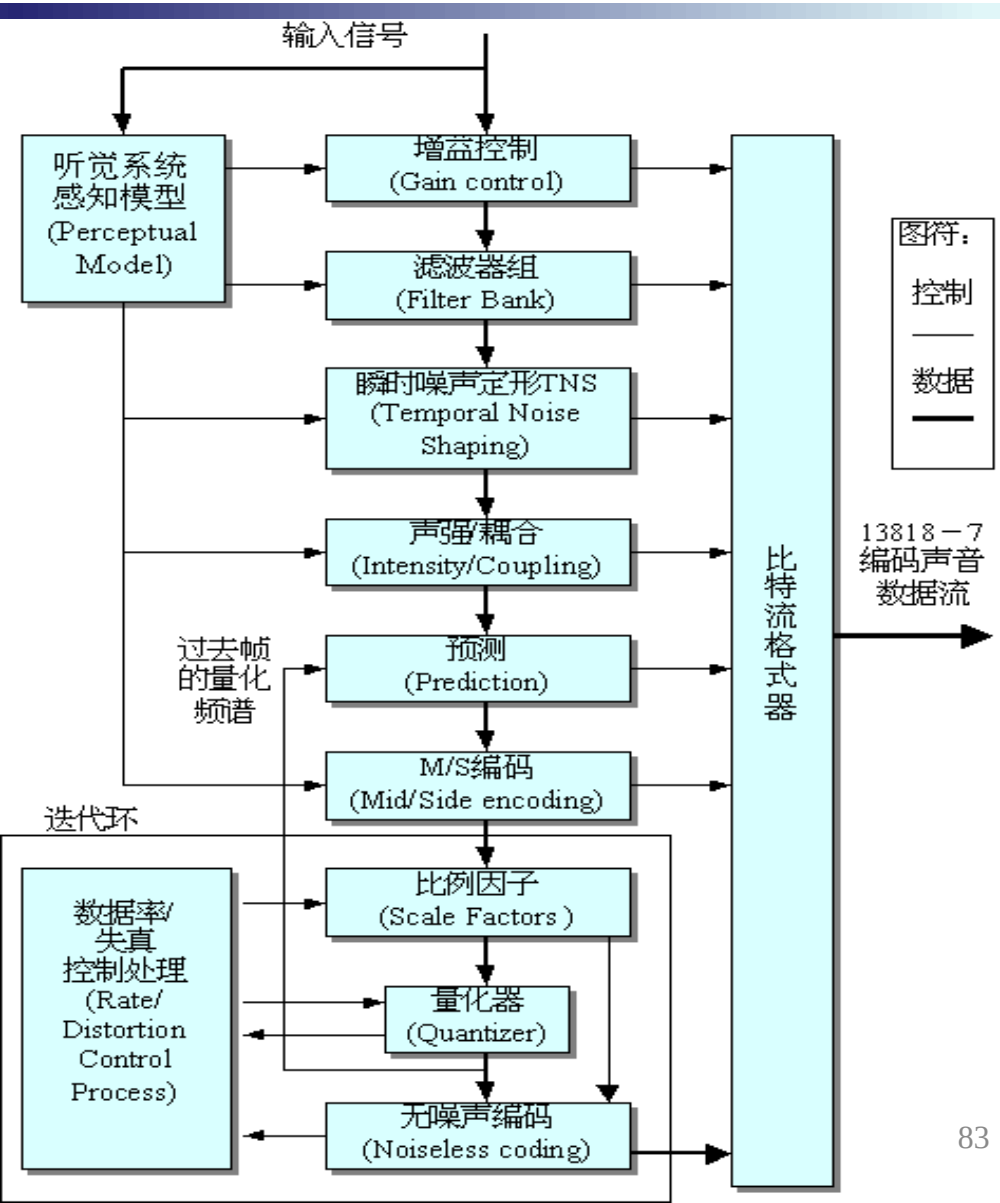
{ynh,cxh}@ustc.edu.cn

MPEG-2 BC编码器/解码器



{ynh,cxh}@ustc.edu.cn

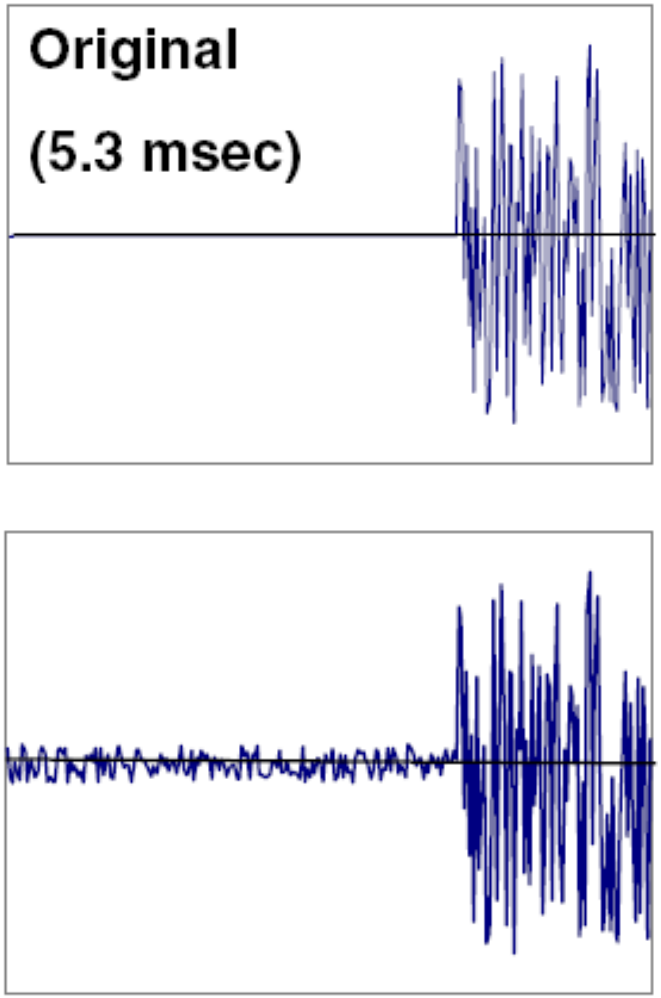
MPEG-2 AAC编码器



- 增益控制
- 滤波器组
  - MDCT/TDAC
- 瞬时噪声定形
- 声强/耦合
  - 单声道信号加上位置信息
- 预测
- Mid/Side
  - 左右声道转换为中央M(middle)和边S(side)声道
- 比例因子
- 量化器
- 无噪声编码
  - 霍夫曼编码

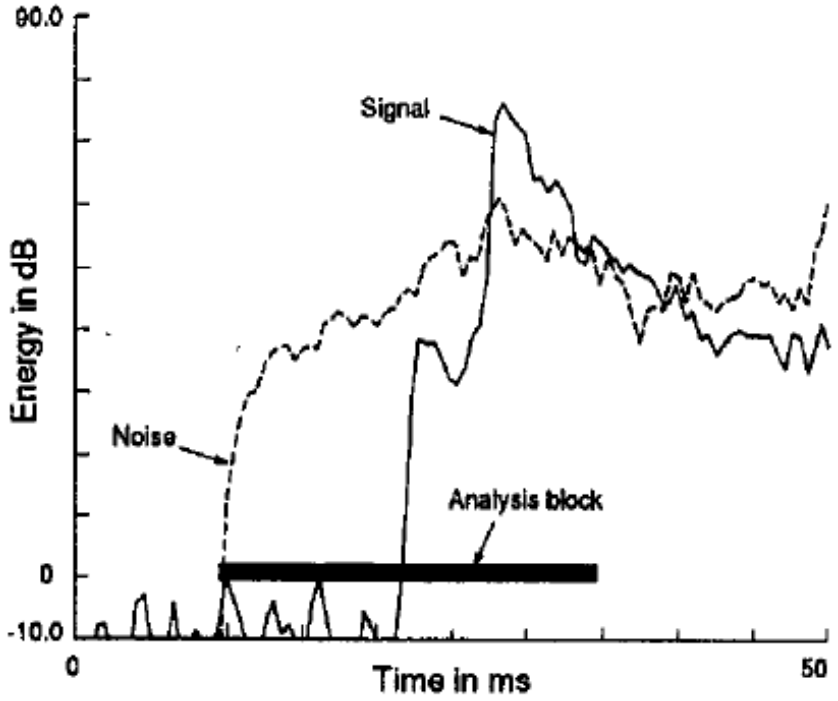
{ynh,cxh}@ustc.edu.cn

Pre-Echo



◆频域系数在编码过程中的量化产生的量化误差在时域被扩展了

◆当时域上出现能量突然变化的信号时（频域系数变化也比较大），量化噪声明显变大



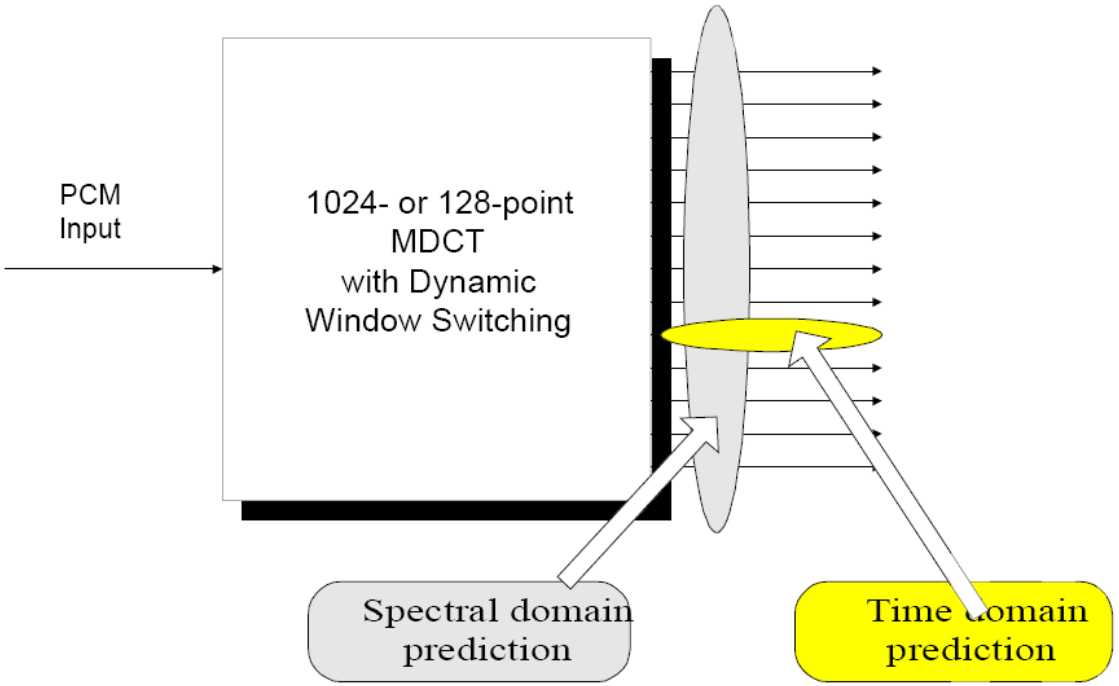
消除pre echo思路

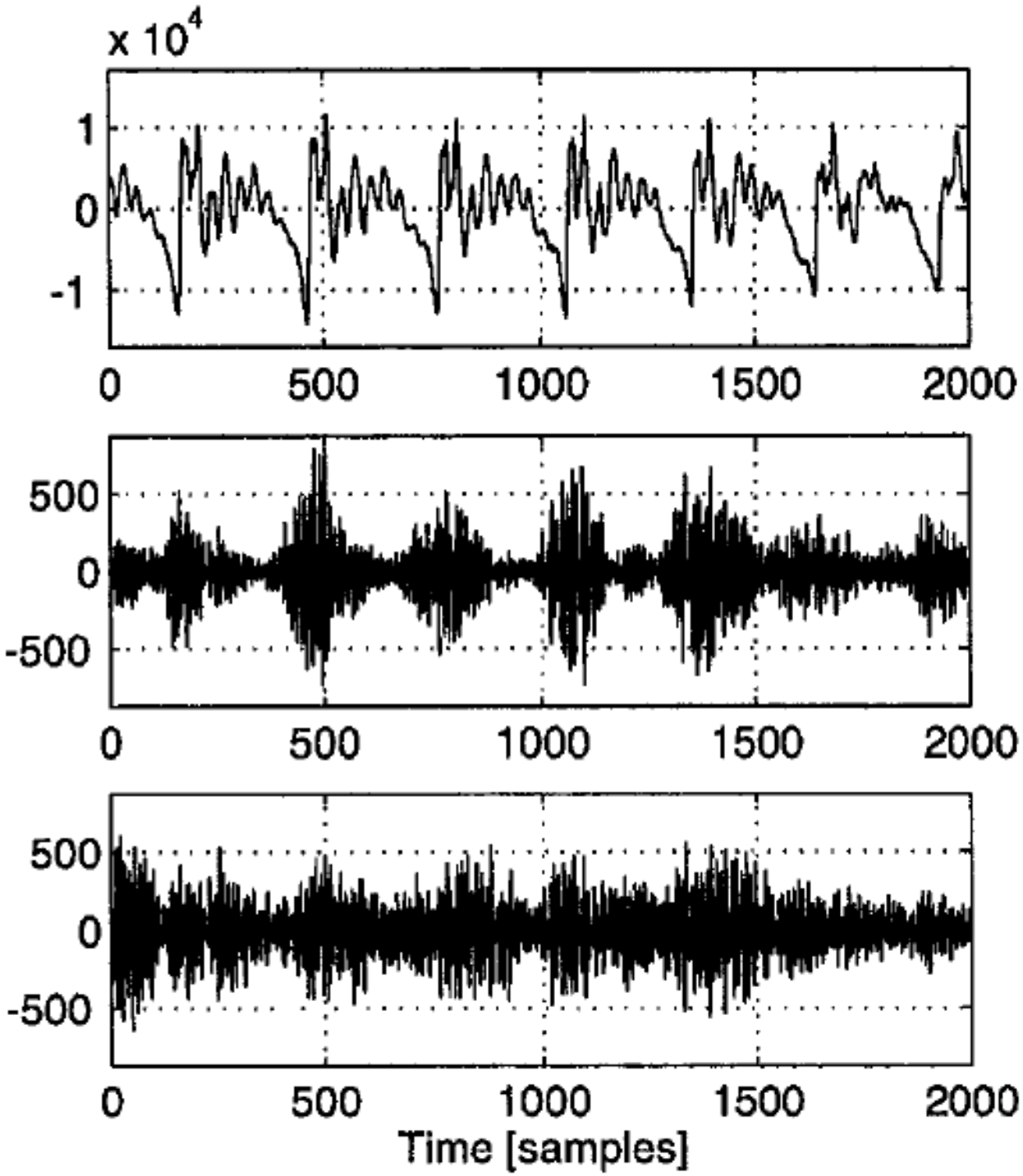
- 频域系数采用尽量小的量化阶（需要更多的比特）
- 自适应的窗口大小，对于stationary信号采用长的窗口，transient信号采用短的窗口（增加了复杂度）
- 采用增益控制，通过控制频域系数的动态范围使量化误差减小

{ynh,cxh}@ustc.edu.cn

TNS(Temporal Noise Shaping)

- 正常情况下，频域上的系数通过PCM进行编码；并随时对频率系数进行预测。当预测器发现频域系数变化超过一定阈值的时候，对频域系数采用DPCM编码
- 即通过对频域系数编码的调整降低频域上量化给时域带来的噪声





- 原始信号
- Coding with TNS
- Coding without TNS

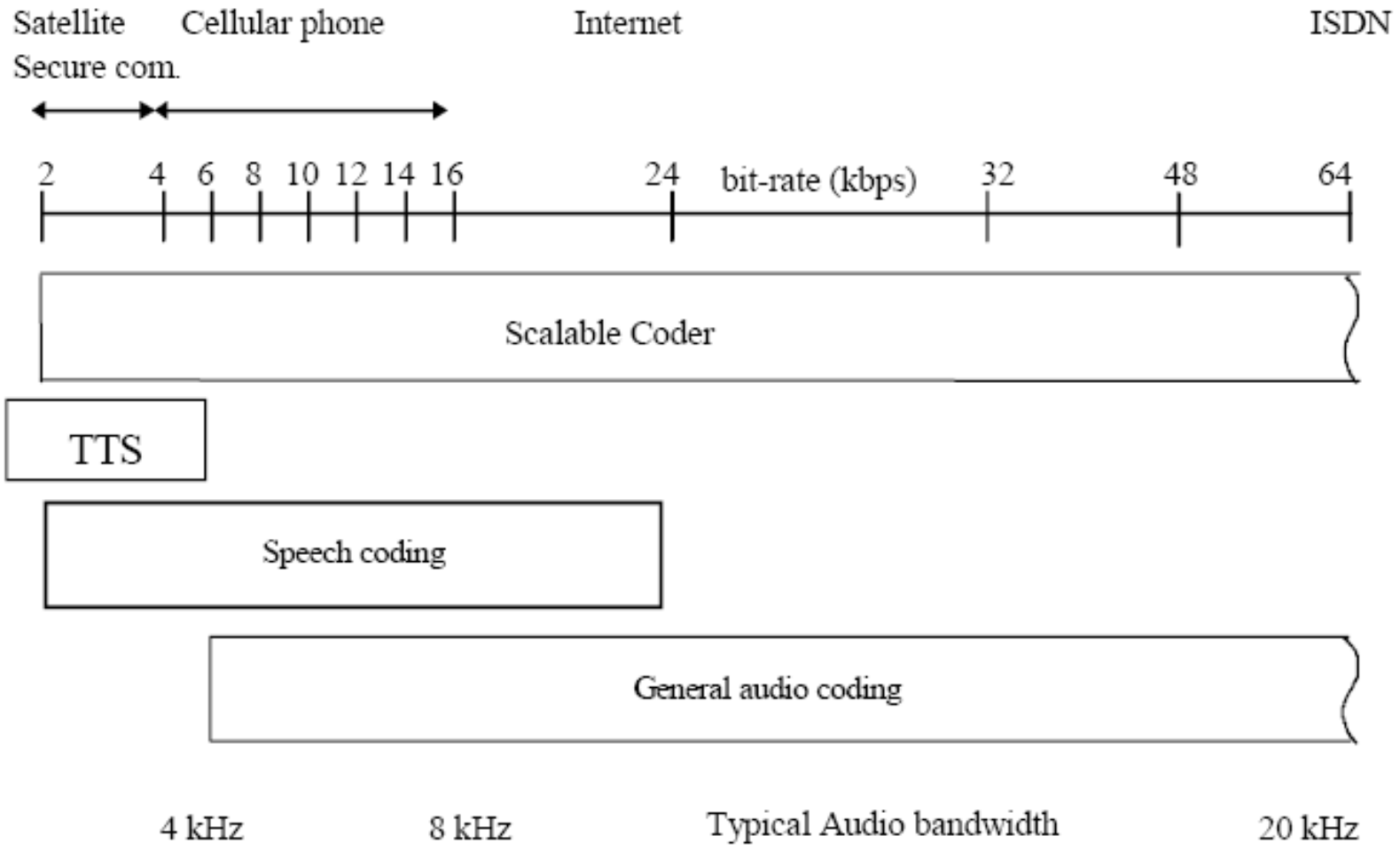
{ynh,cxh}@ustc.edu.cn

MPEG4 Audio

- MPEG-4 Audio标准可集成从话音到高质量的多通道声音，从自然声音到合成声音。
- 编码方法还包括
  - 参数编码(parametric coding)，
  - 码激励线性预测(code excited linear predictive，CELP)编码，
  - 时间/频率T/F(time/frequency)编码，
  - 结构化声音SA(structured audio)编码，
  - 文本-语音TTS(text-to-speech)系统的合成声音等

{ynh,cxh}@ustc.edu.cn

MPEG4的3种编码



{ynh,cxh}@ustc.edu.cn

Scalable Coding

URL http://www.ntt.co.jp/tr/0403/files/ntr200403053.pdf

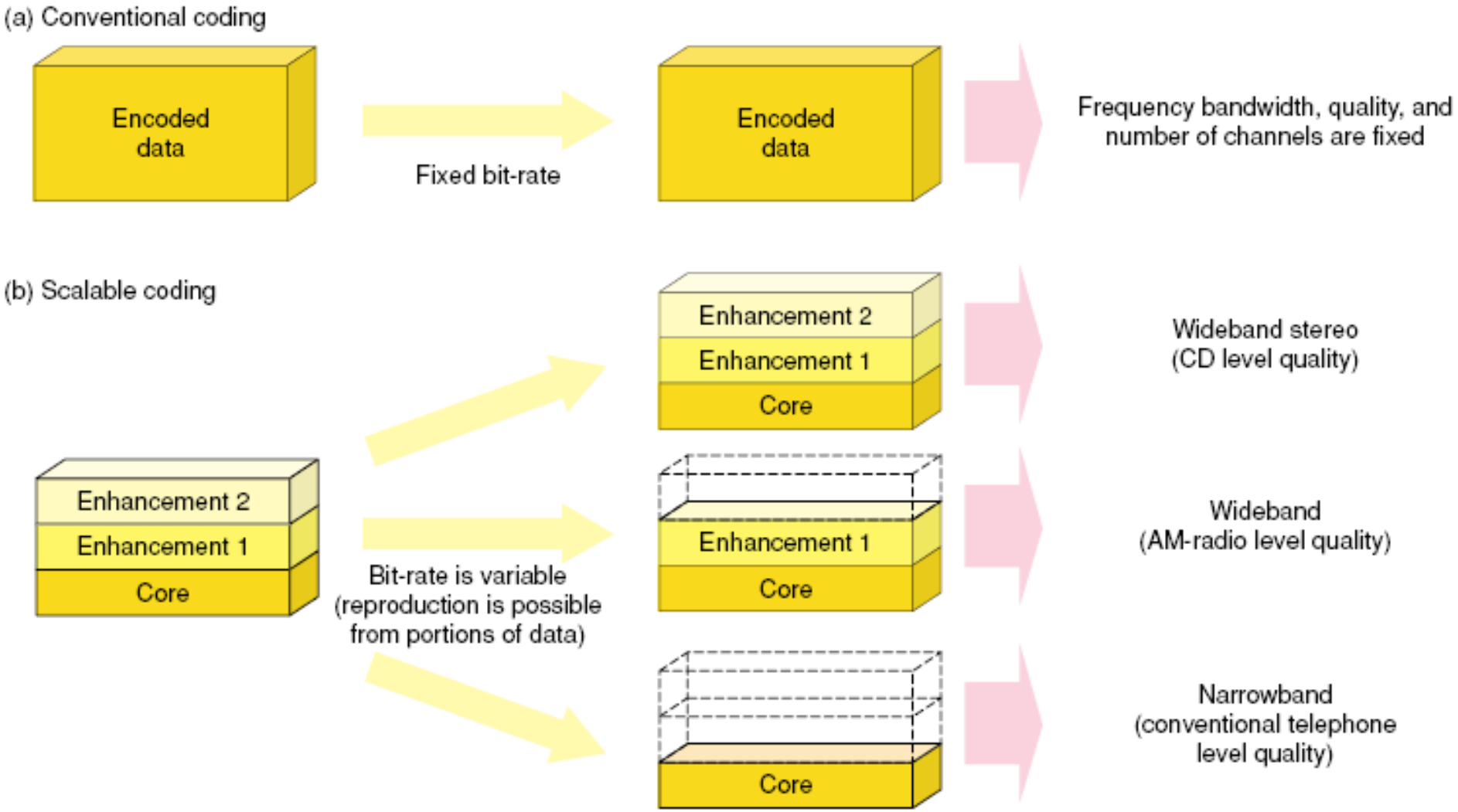
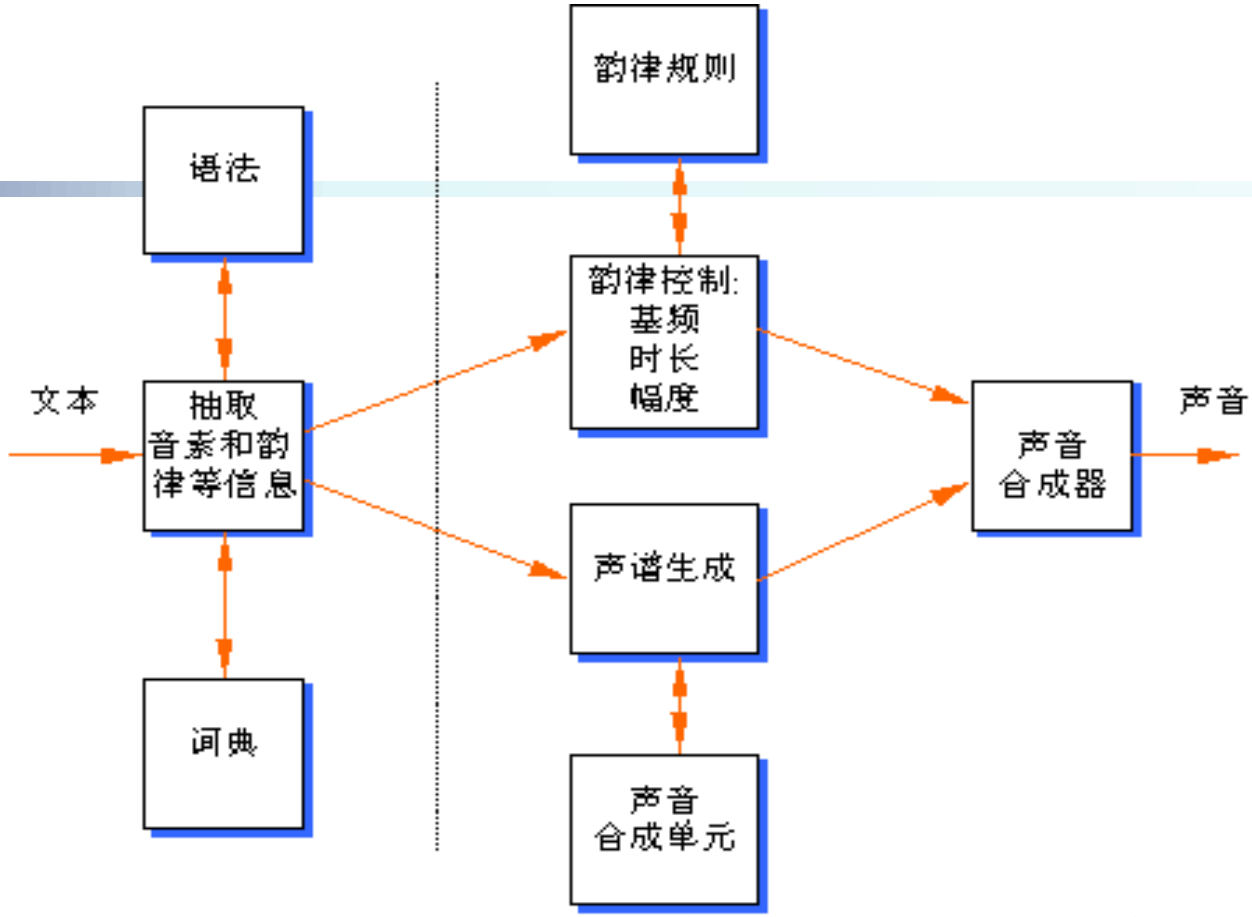


Fig. 2. Comparison of conventional coding and scalable coding.

TTS



◆虚线左边的部分是文本分析部分，通过对输入文本进行词法分析、语法分析，甚至语义分析，从文本中抽取音素和韵律等发音信息。

◆虚线右边的部分是语音合成部分，它使用从文本分析得到的发音信息去控制合成单元的谱特征(音色)和韵律特征(基频、时长和幅度)，送入声音合成器(软件或硬件)产生相应的语音输出。

{ynh,cxh}@ustc.edu.cn

音频压缩思路小结

- 基于音频数据的统计特性进行编码
  - μ 律(μ-Law)或A律(A-Law) 非均匀量化实现压缩
  - △M通过记录差值实现压缩
  - DPCM通过记录预测值与实际信号的差实现压缩
  - APCM通过调整量化阶实现压缩
  - ADPCM是DPCM和APCM思想的集合
  - SB-ADPCM通过改变不同子带样本的比特分配实现压缩（听觉特性）
- 基于音频的声学参数进行参数编码
  - LPC记录的是信道模型的参数
- 混合编码
  - MPE、RPE改变激励获取不同的效果，CELP通过建立码本进一步压缩
- 基于人的听觉特性进行编码
  - MPEG1 Layer1/2/3，基于听觉特性的变换域编码
  - MPEG2 BC & AAC ，基于听觉特性的变换域编码
  - MPEG4 Audio使用了参数编码和混合编码

{ynh,cxh}@ustc.edu.cn

音频编码算法和标准一览

	算法	名称	数据率	标准	应用	质量
波形编码	PCM	均匀量化			公用网 ISDN 配音	4.0-4.5
	μ(A)	μ(A)	64kb/S	G.711		
	APCM	自适应量化				
	DPCM	差值量化				
	ADPCM	自适应差值量化	32kb/S	G.721		
	SB-ADPCM	子带-自适应 差值量化	64kb/S	G.722		
			5.3kb/S 6.3kb/S	G.723		
参数编码	LPC	线性预测编码	2.4kb/S		保密话声	2.5-3.5
混合编码	CELPC	码激励LPC	4.8kb/S		移动通信	4.0-3.7
	VSELP	矢量和激励LPC	8kb/S		语音邮件	
	RPE-LTP	长时预测规则码激励	13.2kb/S		ISDN	
	LD-CELP	低延时码激励LPC	16kb/S	G.728 G.729		
	MPEG	多子带 感知编码	128kb/S		CD	5.0
	AC-3	感知编码			音响	5.0

第2章 多媒体数据压缩国际标准

- 2.1 多媒体数据压缩编码的重要性和分类
- 2.2 常见数据压缩方法分类与基本原理
- 2.3 音频压缩标准
  - 2.3.1 话音编码基础
  - 2.3.2 三种话音编码器
  - 2.3.3 MPEG Audio
  - 2.3.4 移动通信网中的音频编码 ←话音
- 2.4 静态图像压缩编码的国际标准
- 2.5 视频压缩的国际标准
- 2.6 可伸缩性编码和分布式编码

{ynh,cxh}@ustc.edu.cn

商用移动通信网中音频编码的要求
3GPP标准

Subject of specification series	3G and beyond / GSM (R99 and later)	GSM only (Rel-4 and later)	GSM only (before Rel-4)
General information (long defunct)			00 series
Requirements	21 series	41 series	01 series
Service aspects ("stage 1")	22 series	42 series	02 series
Technical realization ("stage 2")	23 series	43 series	03 series
Signalling protocols ("stage 3") - user equipment to network	24 series	44 series	04 series
Radio aspects	25 series	45 series	05 series
CODECs	26 series	46 series	06 series
Data	27 series	47 series (none exists)	07 series
Signalling protocols ("stage 3") -(RSS-CN) and OAM&P and Charging (overflow from 32.- range)	28 series	48 series	08 series
Signalling protocols ("stage 3") - intra-fixed-network	29 series	49 series	09 series
Programme management	30 series	50 series	10 series

http://www.3gpp.org/specifications/specification-numbering

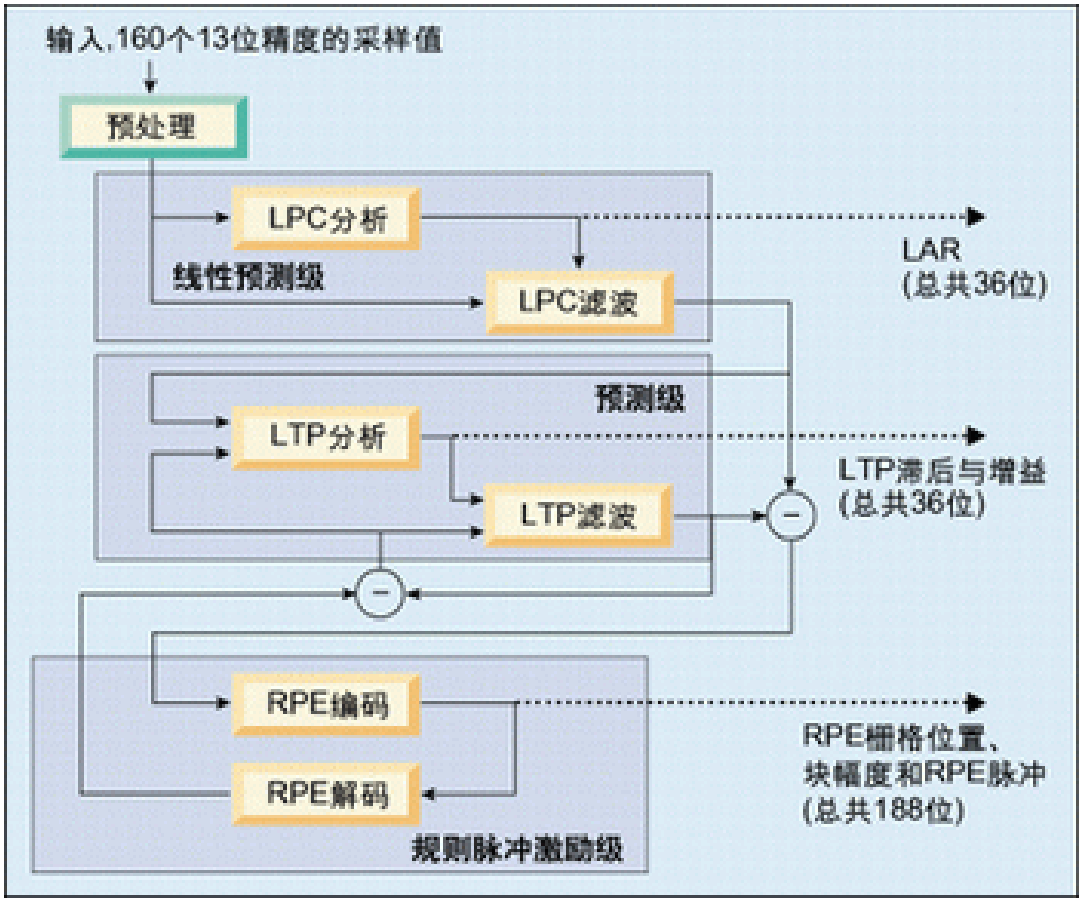
GSM网络中的音频编码

◆GSM系统中有四种编解码器：全速率(FR)、增强型全速率(EFR)、自适应多速率(AMR)及半速率语音压缩。

Standard	Description	Bit rate	Mos (Ber=0)
ETSI GSM 06.10	Full Rate (FR) speech transcoding (RPE-LTP:Regular Pulse Excitation-Long Term Prediction)	13 kbit/s	3.7
ETSI GSM 06.20	Half Rate (HR) speech transcoding (VSELP:Vector sum Excited Linear Prediction)	5.6 kbit/s	3.5
ETSI GSM 06.60	Enhanced Full Rate (EFR) speech transcoding (ACELP:Algebraic CELP)	12.2 kbit/s	3.9
ETSI GSM (AMR)	Used in UMTS	4.75 - 12.2 kbit/s	3.4 - 3.9
ETSI GSM (AMR-WB) = ITU-T G.722.2 WB	Used in UMTS	6.60 – 23.85 kbit/s	3.5 – 4.2

全速率

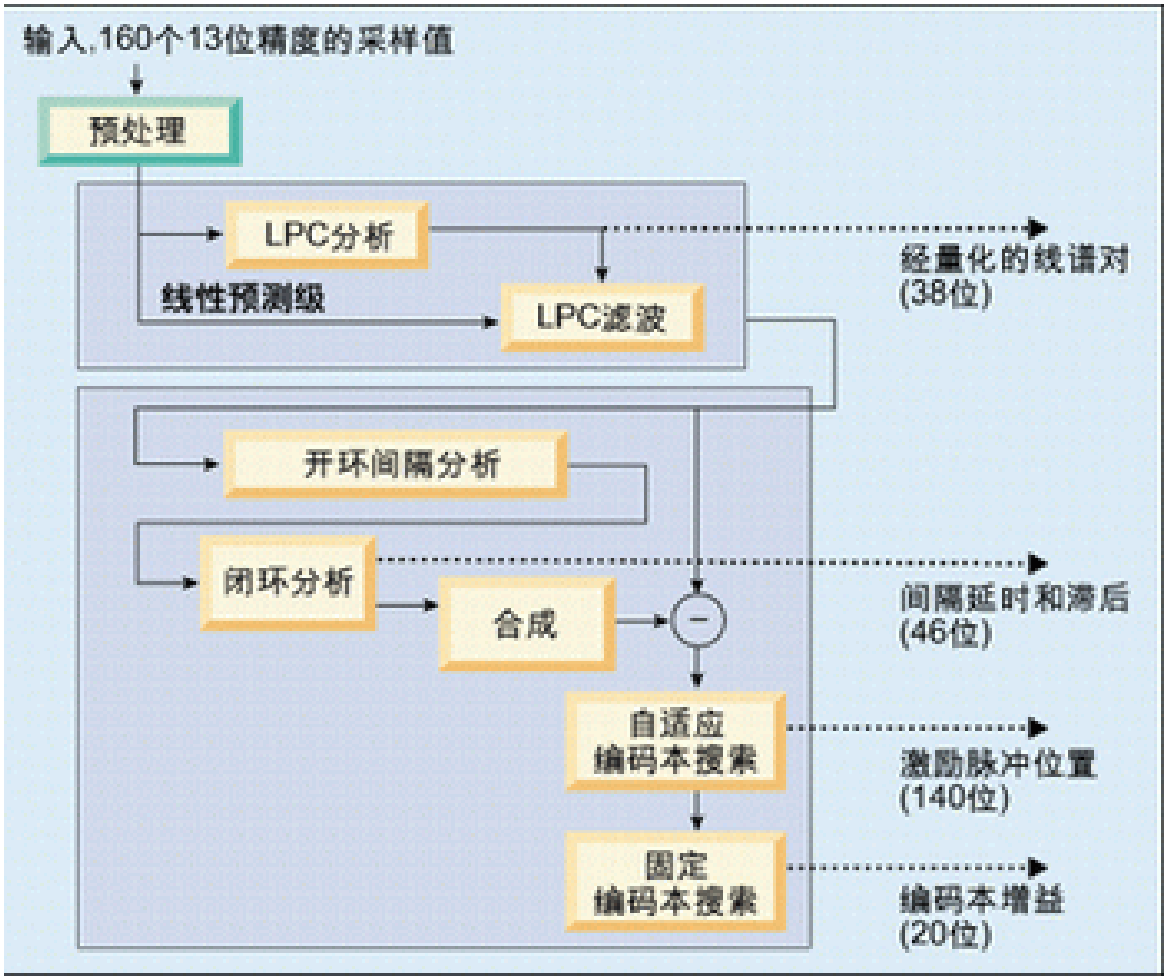
- 全速率编解码器就被称为RPE-LTP线性预测编码器
  - 长期预测(LTP)
  - 规则脉冲激励(RPE)
- 8个系数被变换成可以更少的位数来进行更佳量化的LAR(log-area ratio)。



The encoder accepts 13 bit linear PCM at a 8 kHz sample rate. This can be direct from an ADC in a phone or computer, or converted from G.711 8-bit nonlinear A-law or μ-law PCM from the PSTN with a lookup table.

增强型全速率(EFR)

◆EFR声码器是一种代数码激励线性预测(ACELP)编码器，EFR声码器的12.2kbps输出等于每帧244位。但编码语音是通过拥有260位容量的常规GSM全速率空中信道来传输，其余16位被填以CRC以及重复一些用于冗余的最重要编解码器参数。



{ynh,cxh}@ustc.edu.cn

代数码激励线性预测 ACELP

- Algebraic code-excited linear prediction (ACELP) is a patented speech coding algorithm by VoiceAge Corporation in which **a limited set of pulses** is distributed as excitation to **linear prediction filter**.
- The ACELP algorithm is based on that used in CELP, but ACELP codebooks have a specific algebraic structure imposed upon them.
- The ACELP method is widely employed in current speech coding standards such as AMR, EFR, AMR-WB (G.722.2), VMR-WB, EVRC, EVRC-B, SMV, TETRA, PCS 1900, MPEG-4 CELP and ITU-T G-series standards G.729, G.729.1 (first coding stage) and G.723.1.

With the patent ending on 9th February 2018, designers have the option of ACELP which the customer can optionally pay for now or for standard usage after the patent expires.

{xmh,cbl}@ustc.edu.cn

自适应多速率
Adaptive Multi-Rate (AMR)

◆当全部参数均能解码时，全速率及EFR编解码器可实现良好的语音再现。但当参数丢失或错误时，所接收信号的质量将迅速下降。

◆AMR编解码器组由速率从12.2kbps至4.75kbps的ACELP声码器组成，故可提供87%至480%的冗余。在一种很糟的情况下，即全速率及EFR帧丢失很久后，4.75kbps编解码器数据仍能恢复。

AMR utilizes Discontinuous Transmission (DTX), with Voice Activity Detection (VAD) and Comfort Noise Generation (CNG) to reduce bandwidth usage during silence periods.

{ynh,cxh}@ustc.edu.cn

3G中AMR

- 1999年初,3GPP采纳了由爱立信、诺基亚、西门子提出的自适应多速率(AMR)标准作为第三代移动通信中语音编解码器的标准。
- AMR标准针对不同的应用，分别提出了AMR－NB，AMR-WB和AMR-WB+三种不同的协议。AMR-NB应用于窄带，而AMR-WB和AMR-WB+则应用于宽带通信中。
- AMR声码器采用ACELP (Algebraic Code Excited Linear Prediction)编码方式，提供了8种编码速率(4.75~12.20kbit/s)，每种速率都有不同的容错率。

Adaptive Multi-Rate Wideband (AMR-WB)

{ynh,cxh}@ustc.edu.cn

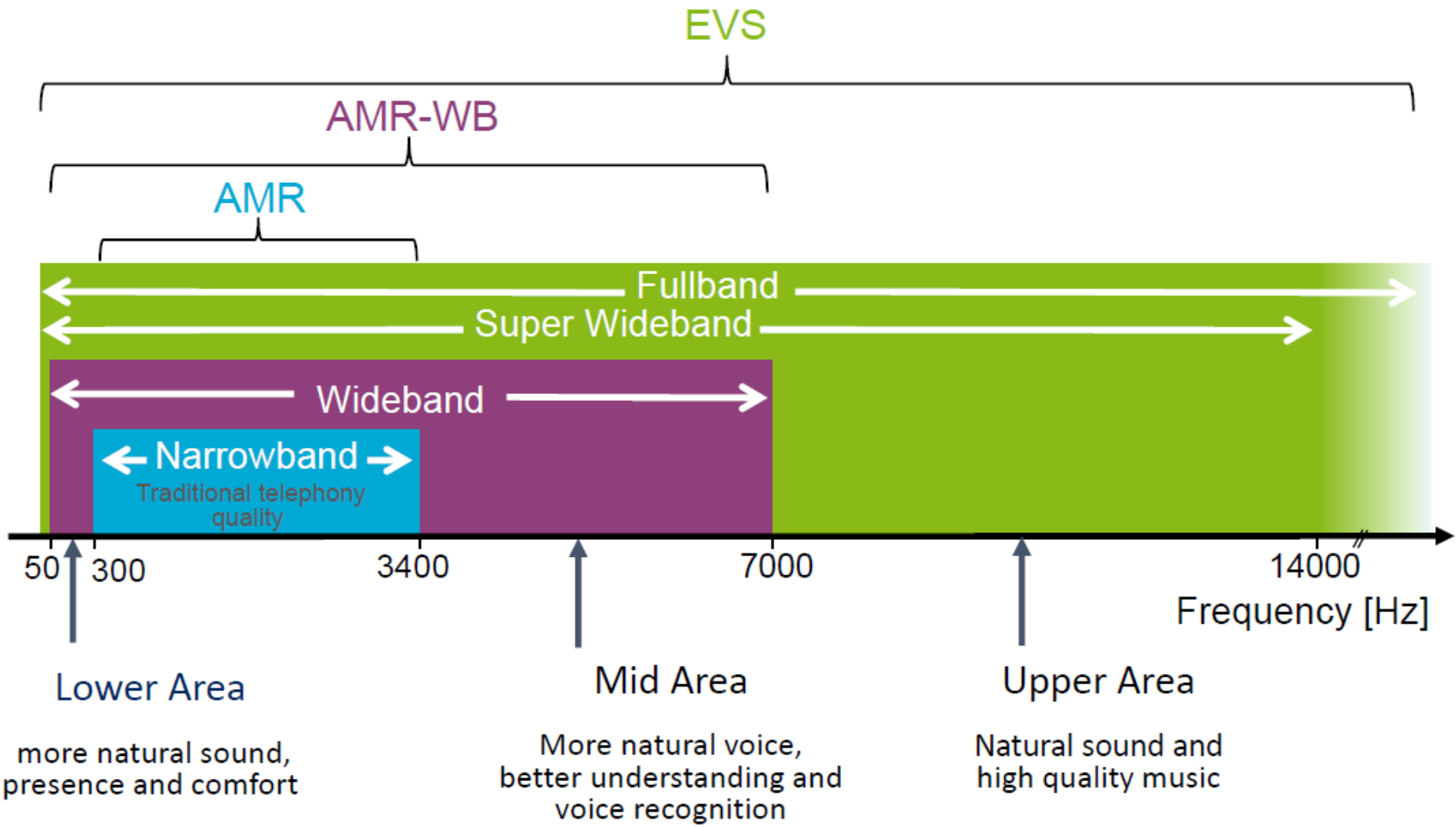
半速率

- GSM所采用的空中接口允许使用两个完全独立的半速率子信道，故能使蜂窝单元的语音容量加倍。
- 半速率声码器采用矢量和激励线性预测VSELP(Vector Sum Exited Linear Prediction)编码器，它以一种类似EFR及AMR编解码器的分析加合成方式工作，速率为5.7kbps。
- 人们对半速率语音的感觉普遍不佳，所以今天一般不采用此项技术。但以其自适应模式，AMR声码器的6种较低速率将适合半速率空中信道的可用容量，结果是采用带AMR的半速率信道将在高流量领域变得更为普遍。

{ynh,cxh}@ustc.edu.cn

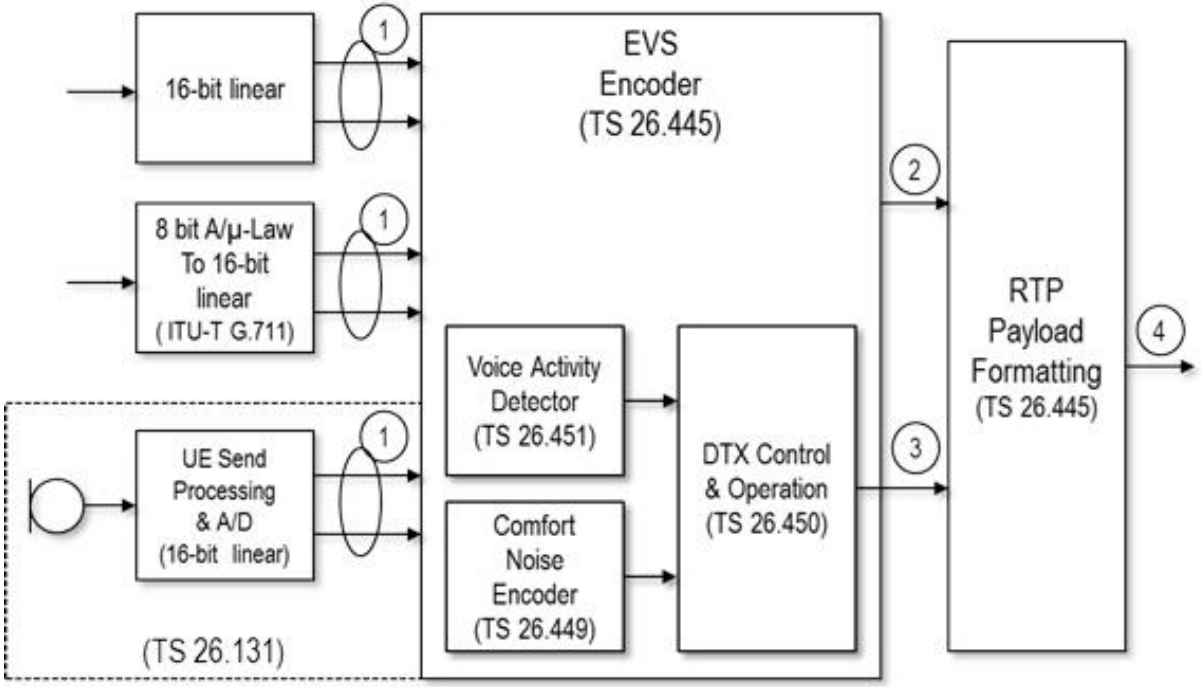
4G之后的话音编码EVS
Enhanced Voice Services (EVS) Codec

http://www.aes.org/technical/documentDownloads.cfm?docID=548



Codec for Enhanced Voice Services (EVS) 编码侧

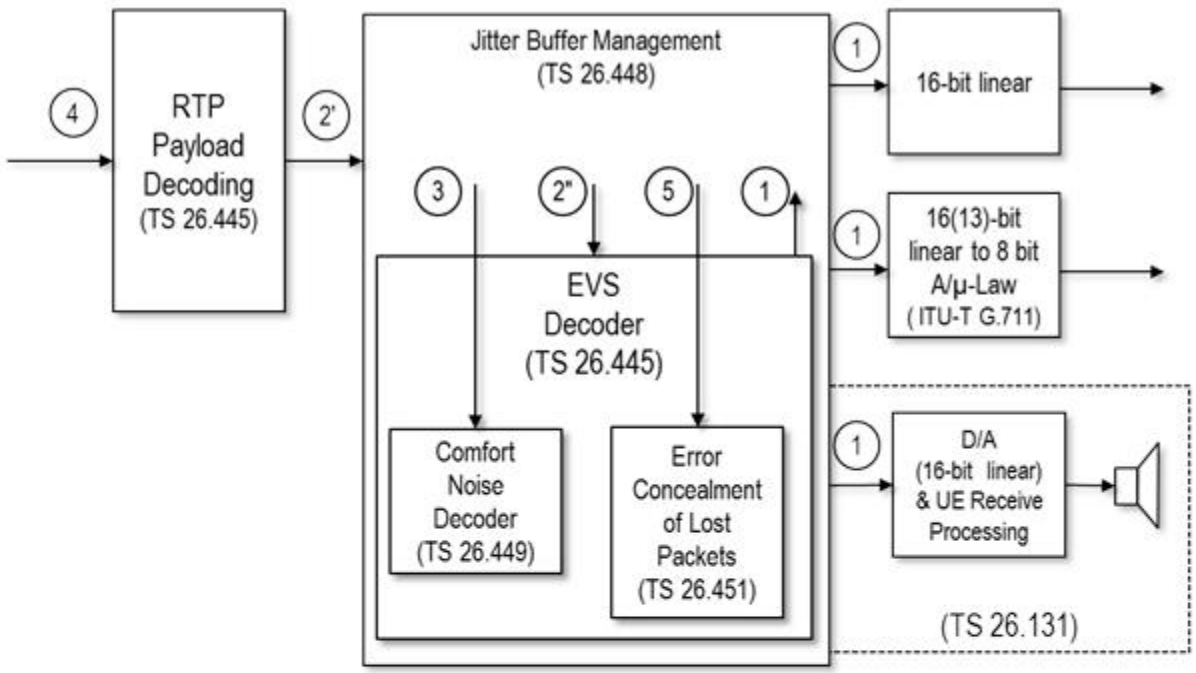
3GPP TS 26.441 V15.0.0 (2018-06)
Codec for Enhanced Voice Services (EVS); General Overview



1. 16-bit Linear PCM Samples and Sample Rate (8, 16, 32 or 48 kHz)
2. Encoded audio frame, 50 frames/s, number of bits/frame depending on the EVS codec mode
3. Encoded Silence Descriptor frames (variable frame rate)
4. RTP Payload Packets

Codec for Enhanced Voice Services (EVS) 解码侧

3GPP TS 26.441 V15.0.0 (2018-06)
Codec for Enhanced Voice Services (EVS); General Overview



1 16-bit Linear PCM Samples and Sample Rate (8, 16, 32 or 48 kHz)

2' 2" Impaired received audio frame, nominally 50 frames/s, number of bits/frame depending on the EVS codec mode

3 Encoded Silence Descriptor frames (variable frame rate)

4 RTP Payload Packets

5 Bad Frame Indication

Codec for Enhanced Voice Services (EVS) 3GPP TS

http://www.3gpp.org/news-events/3gpp-news/1639-evs_news
http://www.3gpp.org/dynareport/26-series.htm

TS 26.441	Codec for Enhanced Voice Services (EVS); General overview
TS 26.442	Codec for Enhanced Voice Services (EVS); ANSI C code (fixed-point)
TS 26.443	Codec for Enhanced Voice Services (EVS); ANSI C code (floating-point)
TS 26.444	Codec for Enhanced Voice Services (EVS); Test sequences
TS 26.445	Codec for Enhanced Voice Services (EVS); Detailed algorithmic description
TS 26.446	Codec for Enhanced Voice Services (EVS); Adaptive Multi-Rate - Wideband (AMR-WB) backward compatible functions
TS 26.447	Codec for Enhanced Voice Services (EVS); Error concealment of lost packets
TS 26.448	Codec for Enhanced Voice Services (EVS); Jitter Buffer Management
TS 26.449	Codec for Enhanced Voice Services (EVS); Comfort Noise Generation (CNG) aspects
TS 26.450	Codec for Enhanced Voice Services (EVS); Discontinuous Transmission (DTX)
TS 26.451	Codec for Enhanced Voice Services (EVS); Voice Activity Detection (VAD)
TS 26.452	Codec for Enhanced Voice Services (EVS); ANSI C code; Alternative fixed-point using updated basic operators
TS 26.453	Codec for Enhanced Voice Services (EVS); Speech codec frame structure
TS 26.454	Codec for Enhanced Voice Services (EVS); Interface to Iu, Uu, Nb and Mb

{ynh,cxh}@ustc.edu.cn

话音/音频自动检测

Range of Operating Points

Band-width	Bitrates [kbps]											
FB 20 kHz						16.4	24.4	32.0	48.0	64.0	96.0	128.0
SWB ≥ 14 kHz				9.6	13.2	16.4	24.4	32.0	48.0	64.0	96.0	128.0
WB 8 kHz	5.9 VBR	7.2	8.0	9.6	13.2	16.4	24.4	32.0	48.0	64.0	96.0	128.0
NB 4 kHz	5.9 VBR	7.2	8.0	9.6	13.2	16.4	24.4					

ACELP/MDCT ⟷ MDCT

- Bandwidth detector: automatically switches to effective bandwidth
- Seamless switching between any operating-points: adapt to transmission-channel

{ynh,cxh}@ustc.edu.cn

http://www.aes.org/technical/documentDownloads.cfm?docID=548 retrieved:201810

速率小结

波形编码→基于CELP发展起来混合编码

Codec	Rate (kHz)	Bitrate (kbps)	Delay (ms)
EVS	48	6.6-23.85	20
AMR-NB	8	4.75-12.2	20
AMR-WB (G.722.2)	16	6.6-23.85	20
G.729	8	8	15
GSM-FR	8	13	20
GSM-EFR	8	12.2	20
G.723.1	8	5.3 6.3	37.5
G.728	8	16	0.625
G.711 (μ/A-law)	8	64	
G.722	16	48 56 64	

{ynh,cxh}@ustc.edu.cn

小结：音频压缩标准

- 冗余
  - 幅度非均匀/样本、周期、基音相关静止系数长时自相关
  - 非均匀的长时功率谱密度/语音特有的短时功率谱密度
- 编译码器
  - 波形编译码器：PCM、DPCM、ADPCM、SB-ADPCM
  - 音源编译码器：LPC
  - 混合编译码器：MPE、RPE、CELP、MPEG4 Audio
  - 感知编码：mpeg1 Layer1/2/3、mpeg2 BC & AAC、Dolby AC-3
- 移动通信中音频编码
  - GSM-HR/GSM-FR/GSM-EFR/AMR-NB/AMR-WB/EVS

{ynh,cxh}@ustc.edu.cn