



## 主要内容

语音编码

■ 随着数字电话和数据通信容量日益增长的迫切要求，而又不希望明显降低传送语音信号的质量，除了提高通信带宽之外，对话音信号进行压缩是提高通信容量的重要措施。

- 重点介绍语音编码的**基本思想**
- 详细计算则留给那些开发和具体设计编解码器软硬件的读者去研究，并可从本章所列的参考文献和站点中找到你满意的文献资料。

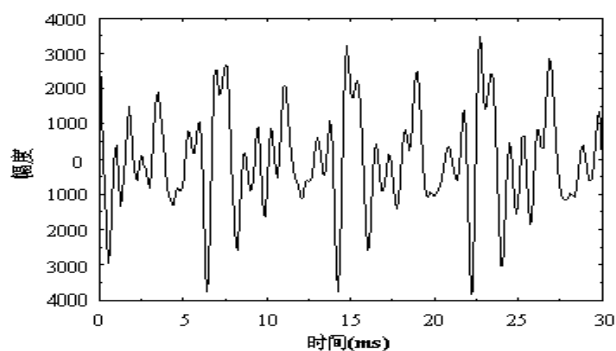


## 语音编码概要

语音编码

### ■ 三种类型

- 压缩空气通过声门激励声道滤波器，根据激励方式的不同，语音分成三种类型。
- 浊音
- 清音
- 爆破音



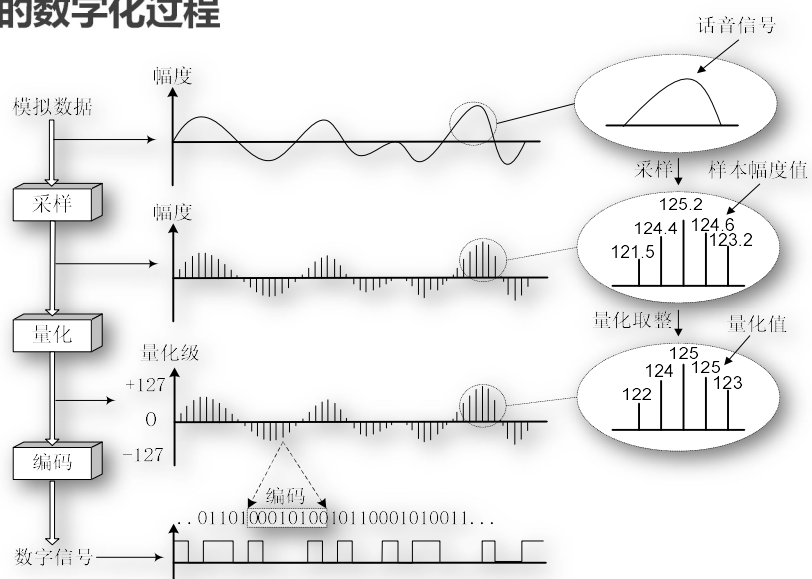
语音的特点



## 声音的波形与质量



## 声音的数字化过程





**Think**

**这么直接变成“二进制”？**



**减少存储空间，提高数据传输率？**

音频编码

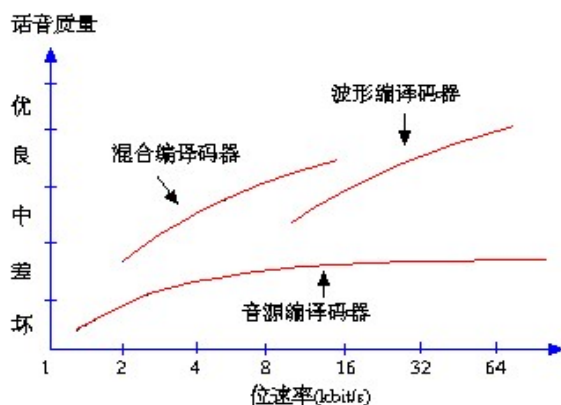


## 三大类

音频编码

- 波形编译码器(Waveform Codecs)
- 音源(参数)编译码器(Source Codecs)
- 混合编译码器(Hybrid Codecs)

### 普通编译码器的音质与数据率



| 编码技术 | 算法       | 编码标准  | 码率(kbit/s) | 质量  | 应用领域      |
|------|----------|-------|------------|-----|-----------|
| 波形编码 | PCM      | G.711 | 64         | 4.3 | PSTN、ISDN |
|      | ADPCM    | G.721 | 32         | 4.1 | -         |
|      | SB-ADPCM | G.722 | 64/56/48   | 4.5 | -         |
| 参数编码 | LPC      | -     | 2.4        | 2.5 | 保密语音      |
| 混合编码 | CELPC    | -     | 4.8        | 3.2 | -         |
|      | VSELP    | GIA   | 8          | 3.8 | 移动通信、语音信箱 |
|      | RPE-LTP  | GSM   | 13.2       | 3.8 | -         |
|      | LD-CELP  | G.728 | 16         | 4.1 | ISDN      |
|      | MPE      | MPE   | 128        | 5.0 | CD        |

说明：质量评价共五个等级（1、2、3、4、5），其中5.0为最高分。

- 一般来说，波形编译码器的话音质量高，但数据率也很高；音源编译码器的数据率很低，产生的合成话音的音质有待提高；混合编译码器使用音源编译码技术和波形编译码技术，数据率和音质介于它们之间。



## 波形编码

PCM  
DPCM  
ADPCM  
SBC  
SB-ADPCM

## 参数编码

LPC  
CELPC



只涉及波形编码

## 混合编码

MPE  
RPE  
CELP  
MELP



## 脉冲编码调制PCM

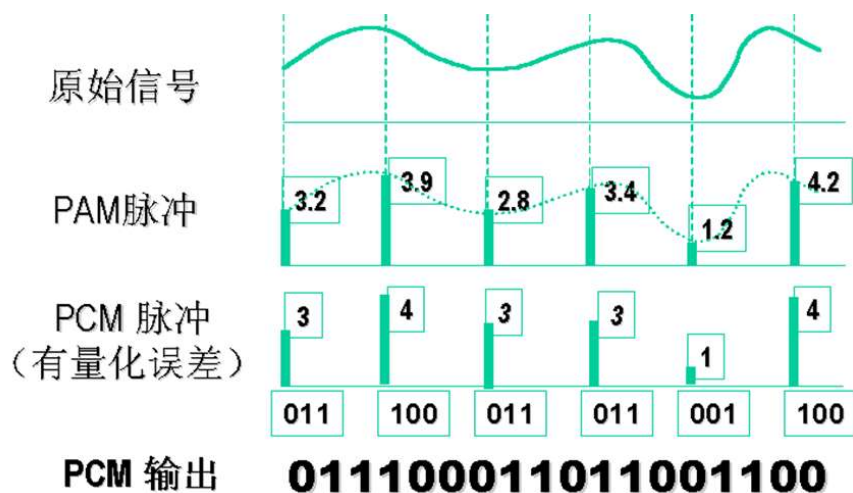
脉冲编码调制PCM(Pulse Code Modulation)



# PCM

脉冲编码调制PCM  
(Pulse Code Modulation)

- PCM是对音频信号采样、量化后不作其它处理直接以数字化格式存储。
- PCM可以选择不同的采样频率，但**并不对音频信号进行压缩**。
- 只要采样频率足够高，量化位数足够多，就能使解码后恢复的声音信号有很高的质量





## PCM采用的量化方法是均匀量化

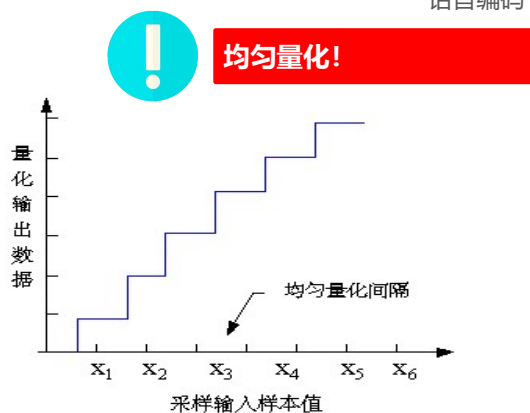
语音编码

### ■ 均匀量化

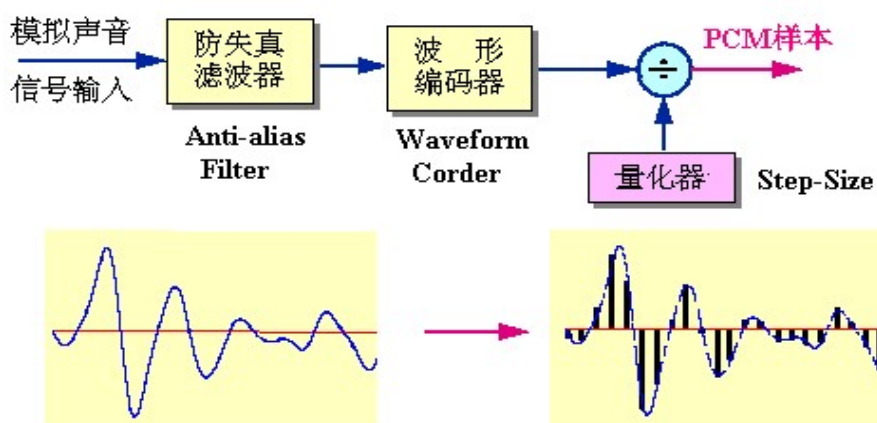
- PCM采用的量化方法是**均匀量化**。

### ■ 主要问题:

- 为了提高音质,就必须提高音频信号的采样精度,即增加量化位数,但数据量会明显增加。



- 采用相同的“等分尺”来度量采样得到的幅度。
- 量化后的样本值Y和原始值X的差  $E = Y - X$  称为**量化误差或量化噪声**。
- 特点: 大的输入信号和小的输入信号都采用相同的量化间隔。



只要采样频率足够高,量化位数足够多,就能使解码后恢复的声音信号有很高的质量。



## 大部分信息集中在低幅度的区域

语音信号的**大部分信息集中在低幅度的区域**

## 大信号—大阶距；小信号—小阶距

对输入信号进行量化时，大的输入信号采用大的量化间隔，小的输入信号采用小的量化间隔。

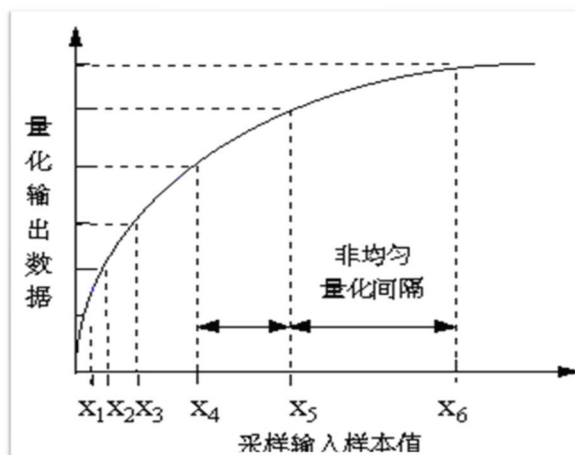




## 非均匀量化

话音编码

- 音频信号的统计特性是：
  - 语音信号的**大部分信息集中在低幅度的区域**。
- 非线性量化的**基本想法**：
  - 对输入信号进行量化时，**大的输入信号采用大的量化间隔，小的输入信号采用小的量化间隔**。



# $\mu$ 律

采样频率8kHz  
8位编码

# A律

- **13位**PCM码
- 主要用在北美和日本等地区的数字电话通信中。

- **14位**PCM码
- 欧洲和中国大陆等地区的数字电话通信中。



**13位PCM码**

$\mu$ 律

**14位PCM码**

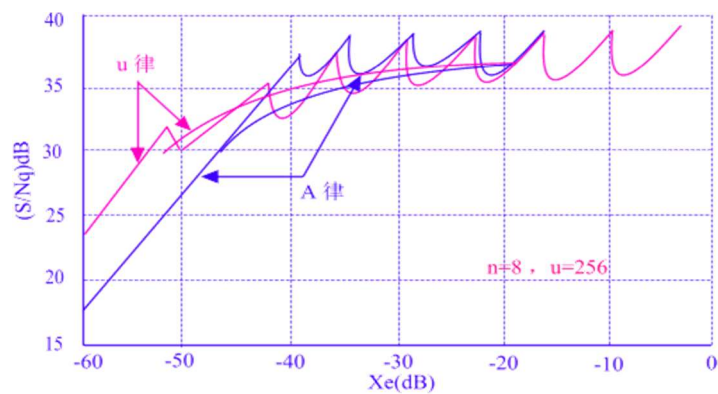
A律

**8位编码**

采样频率8kHz

## $\mu$ 律和A律编码标准(G.711)

- 非线性量化的方法



$\mu$ 律15折线和A律13折线量化信噪比的比较



## $\mu$ 律编码

主要用在北美和日本等地区的数字电话通信中。

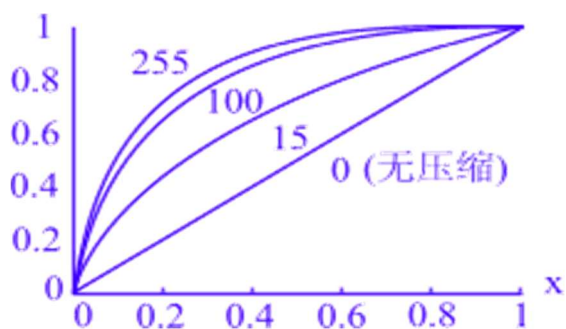
### ■ 量化中采用了**对数压缩法**：

- 量化前用对数函数压缩样值幅度，使信噪比对信号幅度不敏感；
- 解码后用指数函数进行幅度扩张。

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

## $\mu$ 律编码

主要用在北美和日本等地区的数字电话通信中。



$\mu$ 律压缩特性曲线

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



## 自适应脉冲编码调制(APCM)

### 自适应脉冲编码调制(APCM)

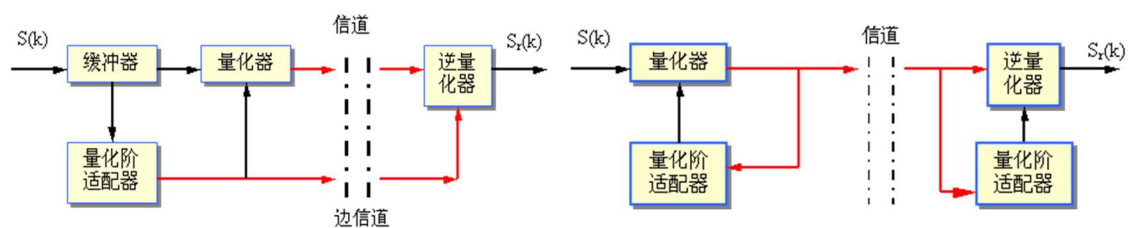
语音编码



自动改变量化阶距

#### 基本思想

- 根据输入信号幅度大小来**改变量化阶距**大小。
- 量化阶大小的方法：
  - 1、前向自适应(forward adaptation)：根据未量化的样本值的均方根值，边信息
  - 2、后向自适应(backward adaptation)：收发两端自动生成量化阶





## 差分脉冲编码调制(DPCM)

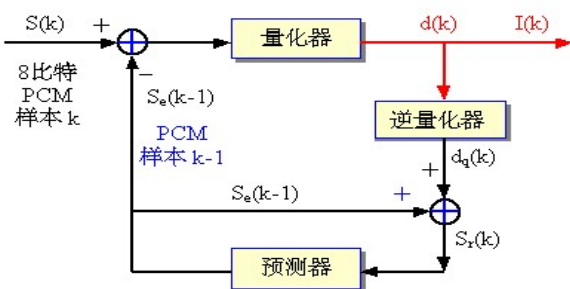
语音编码



传输差值!

### 基本思想

- 根据过去的样本值估算出预测值，然后对预测值和实际样本值的**差值编码**，从而降低音频数据的编码率，达到压缩目的。
- 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。
  - (身高, 110/65100110)



存储或传送的是差值而不是幅度绝对值，降低了传送和存储的数据量。

## 自适应差分脉冲编码调制(ADPCM)

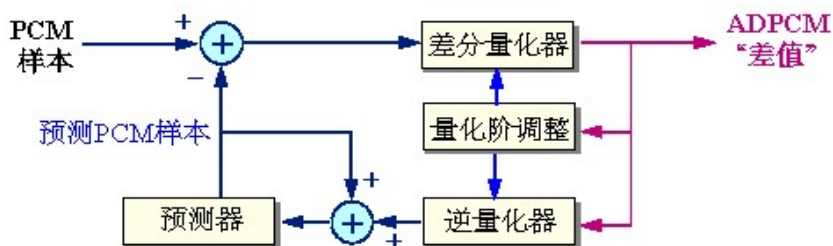
语音编码



自动阶距+差值

### 基本思想

- 差分量化：利用相邻样值之间的差值进行量化。
- 自适应量化：
  - 使量化器自动适应输入电平的变化，随时估计输入信号的时变幅度，以修正量化阶距，当输入信号与预测信号的差值小时，使用小的量化阶距，差值大时，使用大的量化阶距。

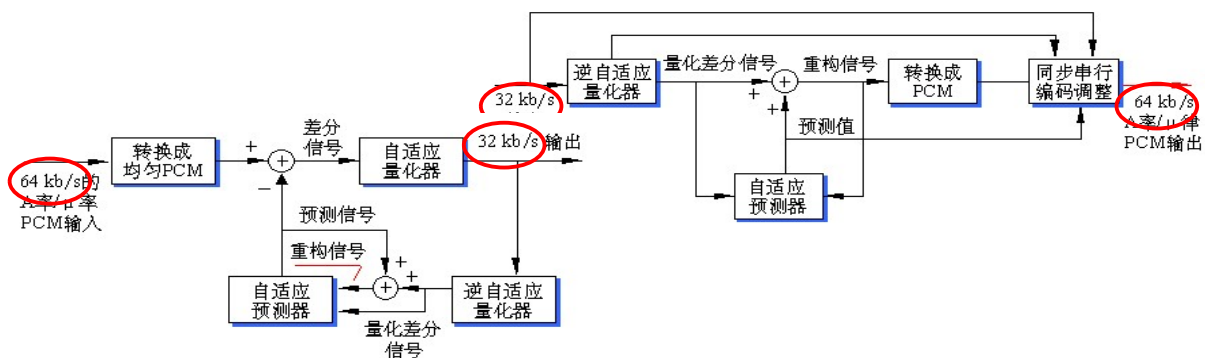




## 四、G.721 ADPCM 标准

### ■ 具体作用

- 实现64kb/s A律/μ律PCM与32kb/s ADPCM之间的转换。
- 压缩率 2:1 。



## 子带编码

子带编码(SBC)



## 子带编码(SBC)

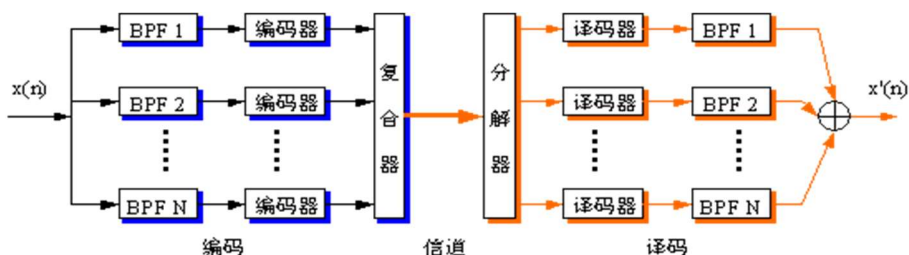
语音编码



每个子带信号进行编码

### 基本思想

- 子带编码将整个音频频带分成若干个子带，分别对每个子带信号进行编码，然后由多路编码器将各子带编码复合后输出。
  - 分解频带为多个子带
  - 对每个子带编码(ADPCM,APCM,PCM)
  - 复合子带
  - 接收端，每个子带单独译码，然后组合，还原成原来的音频信号。



低频

高频

小-量化阶  
多-量化级数

大-量化阶  
少-量化级数

低频子带中，为了保护音调和共振峰的结构，就要求用较小的量化阶、较多的量化级数

音中的摩擦音和类似噪声的声音，通常出现在高频子带中，对它分配较少的位数。



## SB+ADPCM?

结合?

### 3.6 G.722 SB-ADPCM编译码器

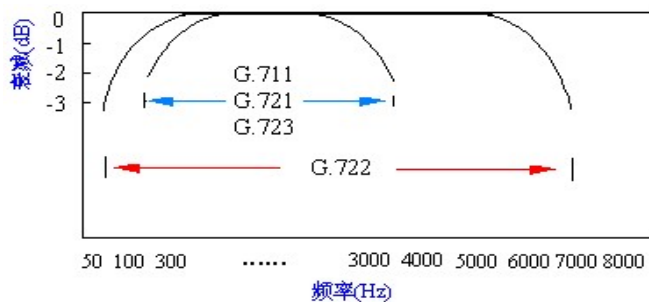
语音编码

■ 主要目标: 保持G.711标准的64kb/s数据率, 但音频质量明显提高。

- G.711: 采样频率**8kHz**、量化位数8位、数据率64kb/s
- 窄带语音: 信号频带: 300 ~ 3400Hz, 语音质量
- G.722: 采样频率**16kHz**、量化位数8位、数据率64kb/s
- 宽带语音: 信号频带: 50 ~ 7000Hz, 调幅广播(AM)质量



8kHz→16kHz







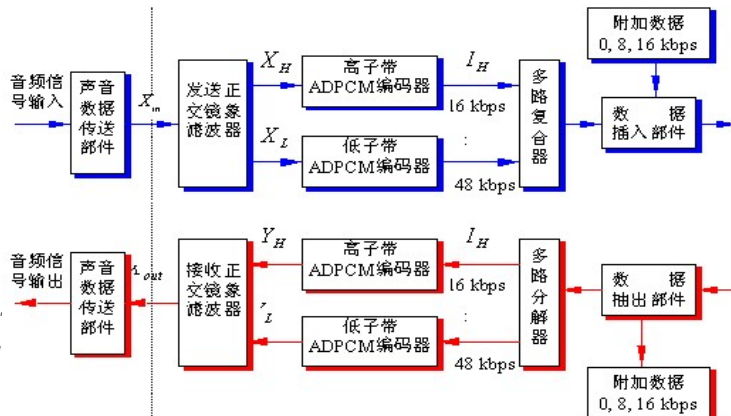
### 3.6 G.722 SB-ADPCM编译码器

语音编码

#### G.722标准

- 用正交镜像滤波器把音频频带分割成高频子带和低频子带，均采用DPCM编码：

- $X_H$ : 高频子带输入信号, 4k ~ 8kHz
- $X_L$ : 低频子带输入信号, 0 ~ 4kHz
- $I_H$ : 高频子带编码信号, 4阶量化, 2位二进制编码,
- $I_L$ : 低频子带编码信号, 6阶量化, 6位二进制编码,



### 增量调制( DM ) 自适应增量调制(ADM)

语音编码



### 3.4 增量调制(DM)与自适应增量调制(ADM)

语音编码

#### ■ 增量调制( $\Delta$ 调制)

- 也称 $\Delta$ (delta modulation, DM), 调制, 是一种预测编码技术。

#### ■ 基本思想:

- 根据当前信号预测下一个信号, 产生一个预测值, 对实际值与预测值之差的极性进行编码。

- 输入信号 **实际值**  $\geq$  **预测值**:  $x[i]=1$
- 输入信号 **实际值**  $<$  **预测值**:  $x[i]=0$

- 编码长度: **1位**

$$y[i+1] = y[i] \pm \Delta$$

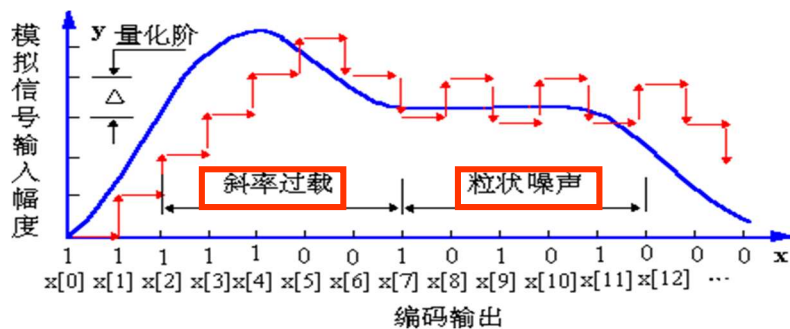
+: 实际值  $\geq$  预测值  
-: 实际值  $<$  预测值

### 缺陷

语音编码

#### ■ DM的主要问题:

- 1) **斜率过载**: 输出代码不能保持跟踪输入信号的快速变化。
- 2) **粒状噪声**: 输入信号与预测值的差值接近0的区域, 输出出现随机交变的“0”和“1”的现象。





## 问题的提出与解决思路

话音编码

### 思路1:

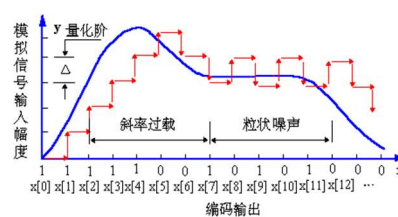
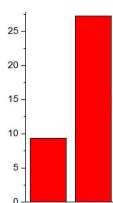
- 加大量化阶距可避免斜率过载，但会加大粒状噪声；
- 减少量化阶距可减少粒状噪声，但又使避免斜率过载更严重。

### 思路2: 斜率的变化自动调整量化阶距的大小?

- 提出了自适应增量调制的编码思想。

## 编码方式小结

话音编码



| 缩写           | 中文名称           | 缩写  | 中文名称    |
|--------------|----------------|-----|---------|
| PCM          | 脉冲编码调制         | DM  | Δ增量调制   |
| μ-law, A-law | μ-律, A-律       | ADM | 自适应增量调制 |
| APCM         | 自适应脉冲编码调制      |     |         |
| DPCM         | 差分脉冲编码调制       |     |         |
| ADPCM        | 自适应差分脉冲编码调制    |     |         |
| SB-ADPCM     | 子带-自适应差分脉冲编码调制 |     |         |



|          | 算法                | 名称             | 数据率                          | 标准             | 应用           | 质量        |
|----------|-------------------|----------------|------------------------------|----------------|--------------|-----------|
| 波形编码     | PCM               | 脉冲编码调制         |                              |                | 公网<br>ISDN配音 | 4.0~4.5   |
|          | $\mu$ -law, A-law | $\mu$ -律, A-律  | 64kbps                       | G.711          |              |           |
|          | APCM              | 自适应脉冲编码调制      |                              |                |              |           |
|          | DPCM              | 差分脉冲编码调制       |                              |                |              |           |
|          | ADPCM             | 自适应差分脉冲编码调制    | 32kbps                       | G.721          |              |           |
|          | SB-ADPCM          | 子带-自适应差分脉冲编码调制 | 64kbps<br>5.3kbps<br>6.3kbps | G.722<br>G.723 |              |           |
| 音源(参数)编码 | LPC               | 线性预测编码         | 2.4kbps                      |                | 保密话声         | 2.5 ~ 3.5 |
| 混合编码     | CELP              | 码激励LPC         | 4.6kbps                      |                | 移动通信         | 4.0~3.7   |
|          | VSELP             | 矢量和激励LPC       | 8kbps                        |                | 语音邮件         |           |
|          | RPE-LTP           | 规则码激励长时预测      | 13.2kbps                     |                | ISDN         |           |
|          | LD-CELP           | 低延时码激励LPC      | 16kbps                       | G.728<br>G.729 |              |           |
|          | MPEG              | 多子带, 感知编码      | 128kbps                      |                | CD           |           |
|          | Dolby AC-3        | 感知编码           |                              |                | 音响           | 5.0       |

## 不同音频文件的容量和视听效果

语音编码

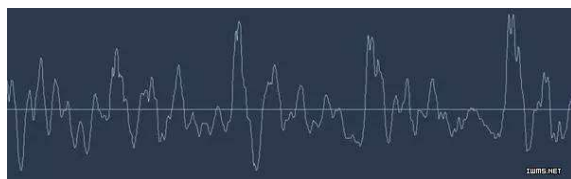
| 歌曲长度 | 采样率 (kHz) | 量化位 (bit) | 声道数 | 编码算法           | 文件格式 | 压缩比 | 文件容量 (KB) | 试听效果 |
|------|-----------|-----------|-----|----------------|------|-----|-----------|------|
| 30秒  | 44.1      | 16        | 双   | PCM            | WAVE |     | 5167      |      |
| 30秒  | 22.05     | 16        | 单   | PCM            | WAVE |     | 1292      |      |
| 30秒  | 11.025    | 8         | 单   | PCM            | WAVE |     | 323       |      |
| 30秒  |           |           |     | CCITT A Law    | WAVE | 1:2 | 2584      |      |
| 30秒  |           |           |     | Creative ADPCM | WAVE | 1:4 | 1292      |      |
| 30秒  |           |           |     | mp3            | mp3  |     | 471       |      |



Think

## 怎样将声波转换成特征？

声音以波的形式传播。



音频「Hello」的波形



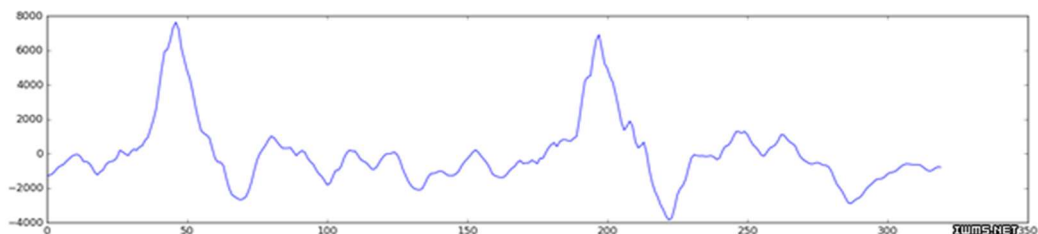
## 周期性

话音信号有**高度**的周期性。

- 话音在短时间周期里(20ms数量级) 可以认为**基本不变准定态**(Quasi-Stationary) ；
- 话音编码器正是利用这种周期性来压缩数据。话音编码器就是企图揭示这种周期性，目的是为了减少数据率而又尽可能不牺牲声音的质量。



## 采样声音数据的预处理



将这些数值绘制为简单的线图，给出了对于 20 毫秒时间段的原始声波的粗略近似。

## 采样声音数据的预处理

```
[ -1274, -1252, -1160, -986, -792, -692, -614, -429, -286, -134, -57, -41, -169, -456, -450, -541, -761, -1067, -1231, -1047, -952, -645, -489, -448, -397, -212, 193, 114, -17, -110, 128, 261, 198, 390, 461, 772, 948, 1451, 1974, 2624, 3793, 4968, 5939, 6857, 6581, 7382, 7640, 7223, 6119, 5461, 4820, 4353, 3611, 2740, 2004, 1349, 1178, 1085, 901, 301, -262, -499, -488, -707, -1406, -1997, -2377, -2494, -2605, -2675, -2627, -2500, -2148, -1648, -970, -364, 13, 260, 494, 788, 1011, 938, 717, 507, 323, 324, 325, 350, 183, -113, 64, 176, 93, -249, -461, -606, -909, -1159, -1307, -1544, -1815, -1725, -1341, -921, -959, -723, -261, 51, 210, 142, 152, -92, -345, -439, -529, -710, -907, -887, -693, -403, -180, -14, -12, 29, 89, -47, -398, -896, -1262, -1610, -1862, -2021, -2077, -2105, -2023, -1697, -1360, -1150, -1148, -1091, -1013, -1018, -1126, -1255, -1278, -1266, -1174, -1003, -707, -468, -300, -116, 92, 224, 72, -150, -336, -541, -820, -1178, -1289, -1345, -1385, -1365, -1223, -1004, -839, -734, -481, -396, -580, -527, -531, -376, -458, -581, -254, -277, 50, 331, 531, 641, 416, 697, 810, 812, 759, 739, 888, 1008, 1977, 3145, 4219, 4454, 4521, 5091, 6563, 6909, 6117, 5244, 4951, 4462, 4124, 3435, 2671, 1847, 1370, 1591, 1900, 1586, 713, 341, 462, 673, 60, -938, -1664, -2185, -2527, -2967, -3253, -3636, -3859, -3723, -3134, -2380, -2032, -1831, -1457, -804, -241, -51, -113, -136, -122, -158, -147, -114, -181, -338, -266, 131, 418, 471, 651, 994, 1295, 1267, 1197, 1291, 1110, 793, 514, 370, 174, -90, -139, 104, 334, 407, 524, 771, 1106, 1087, 878, 703, 591, 471, 91, -199, -357, -454, -561, -605, -552, -512, -575, -669, -672, -763, -1022, -1435, -1791, -1999, -2242, -2563, -2853, -2893, -2740, -2625, -2556, -2385, -2138, -1936, -1803, -1649, -1495, -1460, -1446, -1345, -1177, -1088, -1072, -1003, -856, -719, -621, -585, -613, -634, -638, -636, -683, -819, -946, -1012, -964, -826, -762, -788]
```

数组，每个数值代表声波在间隔为 **1 / 16,000** 秒的时刻时的高度值(振幅)。

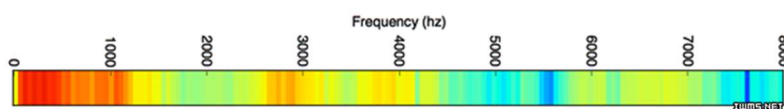


- **20 毫秒**的时长也包含不同频率的声音。
  - 有低音、中音，甚至高音混在一起。
  - 不同频率的声音混合在一起构成了人类复杂的语音。
- 分离不同频率得到**声音指纹**。
  - 为了使数据更容易进行神经网络处理，将这复杂的声波分成不同部分。
  - 将一步步分离低音部分，下一个最低音部分，以此类推。
  - 然后通过将每个频带(从低到高)中的能量相加，为各个类别(音调)的音频片段创建了一个指纹。
- 使用数学中的**傅里叶变换**等来完成。
  - 将复杂的声波分解成简单的声波，得到简单声波，将每一个声波包含的能量加在一起。

- 从低音(即低音音符)到高音, 每个频率范围的重要程度。
  - 下面的每个数字表示 20 毫秒音频剪辑中每个 50Hz 频带中的能量:



表中的每个数值表示每个 50Hz 频带中的能量



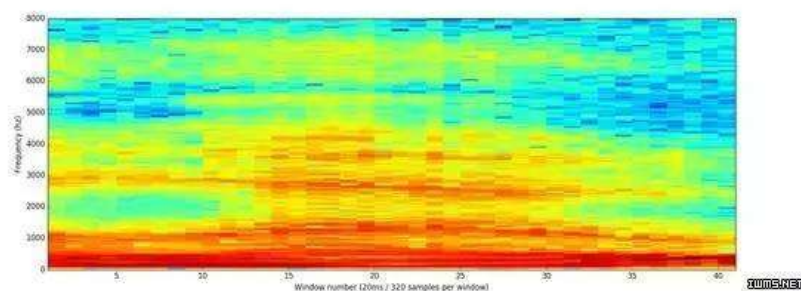
以图表形式画出

20 毫秒的声音片段中有很多低频能量，高频能量较少。这是典型的「男性」的声音。



## 频谱图

- 以每 20 毫秒的音频块重复这个过程，最终会得到一个频谱图
  - 每一列从左到右都是一个 20 毫秒的块



「hello」声音片段的完整频谱图

「频谱图很棒，因为可以从音频数据中看到音符和其他音高模式。相比于原始声波，神经网络可以更容易地从这种数据中找到规律。因此，这(频谱图)就是实际输入到神经网络的数据表征方式。」

## 语音编码

语音编码

### ■ 三种编码方式

- 波形编译码器(Waveform Codecs)
- 音源(参数)编译码器(Source Codecs)
- 混合编译码器(Hybrid Codecs)

### ■ 非均匀量化

- $\mu$ -律, A-律





## 话音编码

话音编码

### ■ 编码方式的思想

- PCM
- $\mu$ -law, A-law
- APCM
- DPCM
- ADPCM
- SB-ADPCM
- DM
- ADM

### ■ 2个标准

- G.721、G.722