



中国科学院自动化研究所
INSTITUTE OF AUTOMATION
CHINESE ACADEMY OF SCIENCES

2019—2020学年(春)第二学期
中国科学院大学课程

语音交互技术 ——语音信号处理



中国科学院自动化研究所
模式识别国家重点实验室

陶建华

jhtao@nlpr.ia.ac.cn

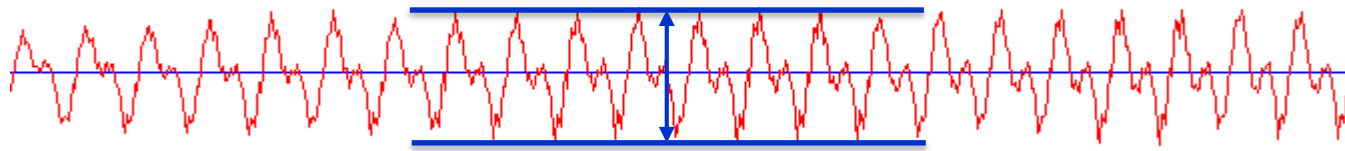
本节课提纲

- 时域分析
- 频域分析
- 倒谱分析
- 线性预测分析
- 基音周期分析

本节课提纲

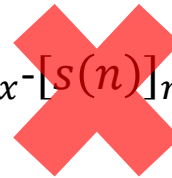
- 时域分析
- 频域分析
- 倒谱分析
- 线性预测分析
- 基音周期分析

语音的强度

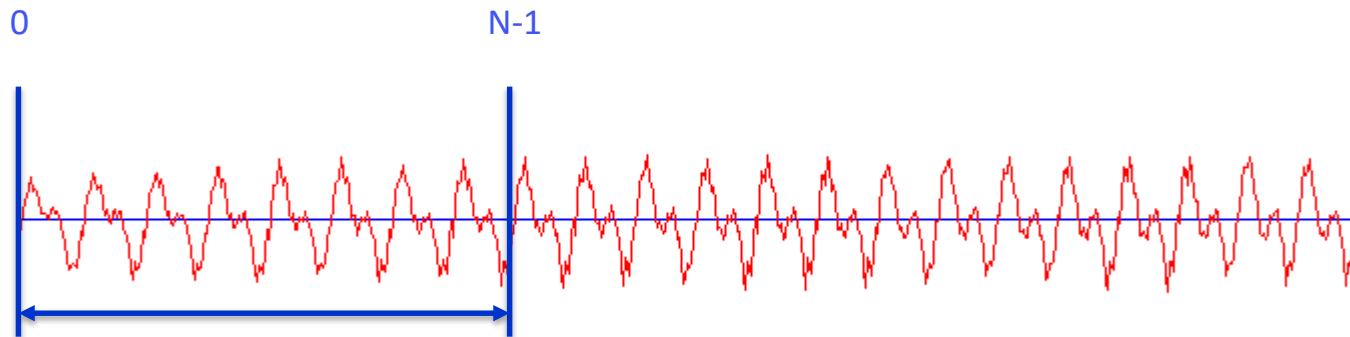


设语音信号为： $s(n)$

则语音的强度为： $[s(n)]_{max} - [s(n)]_{min}$



短时能量分析

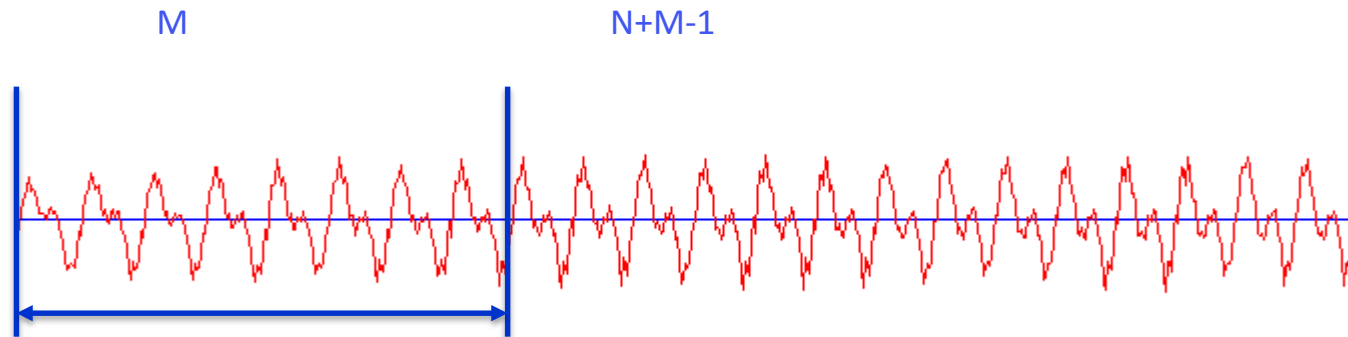


设语音信号为: $s(n)$

正确的语音的强度为:

$$E_0 = \sum_{m=0}^{N-1} [s(m)]^2$$

短时能量分析



$$E_M = \sum_{m=M}^{N+M-1} [s(m)]^2$$

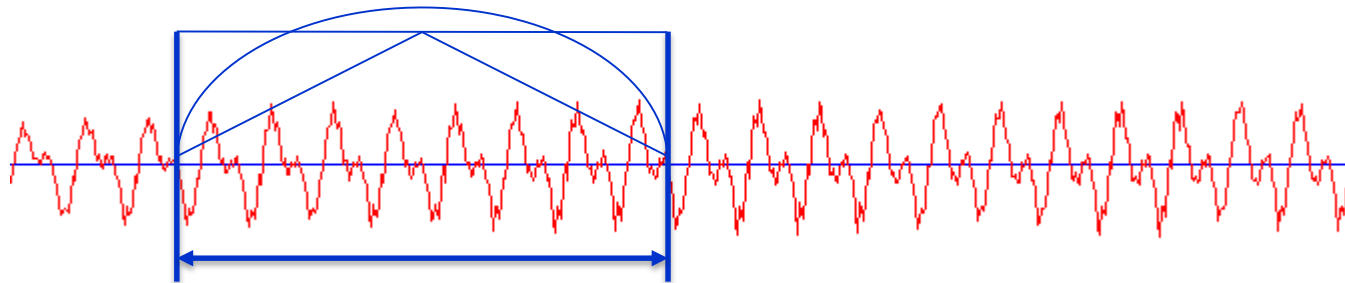
N称为语音短时分析的帧长 Frame Length

$$E_{n \times M} = \sum_{m=n \times M}^{N+n \times M-1} [s(m)]^2$$

M称为语音短时分析的步长 Step

窗函数

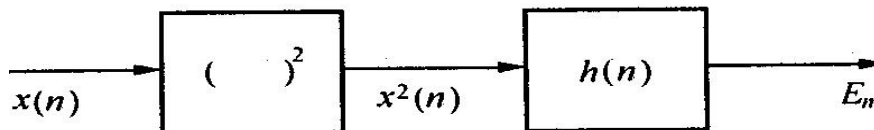
决定短时能量特性有两个条件：不同的窗口的**形状**和**长度**。



$$E_{n \times M} = \sum_{m=n \times M}^{N+n \times M-1} [s(m)w(m - n \times M)]^2$$

$$E_n = \sum_{m=-\infty}^{\infty} [s(m)w(n - m)]^2 = \sum_{m=-\infty}^{\infty} s^2(m)h(n - m) = s^2(n) * h(n)$$

$$h(n) = w^2(n)$$



典型的窗函数

矩形窗：

$$w(n) = \begin{cases} 1 & 0 \leq n \leq M - 1 \\ 0 & \text{其它} \end{cases}$$

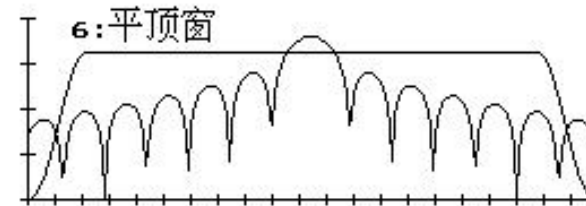
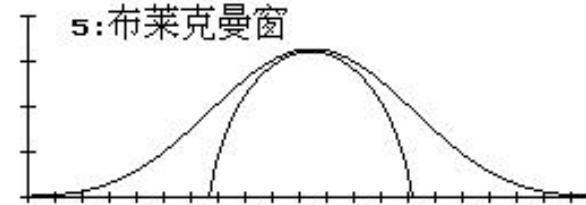
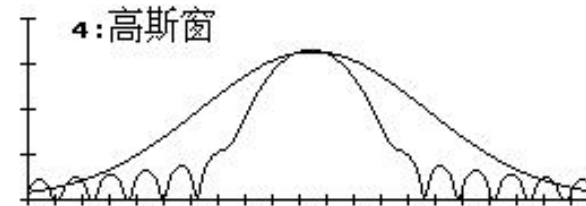
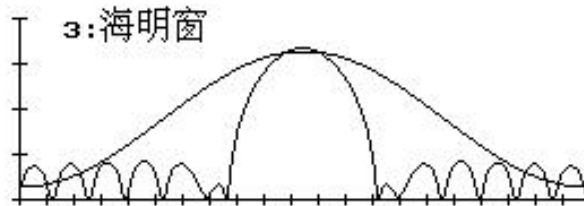
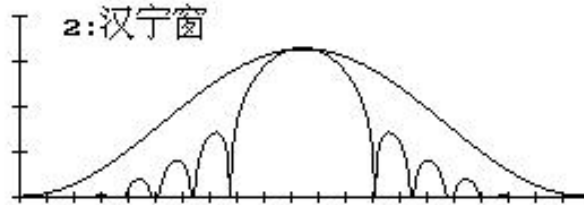
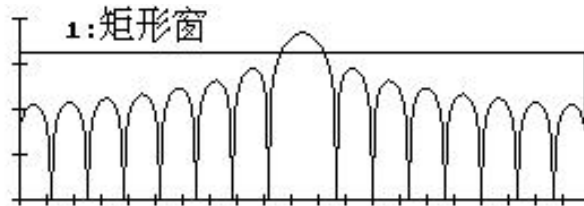
汉宁窗：

$$w(n) = \begin{cases} 0.5 - 0.5 \cos(2\pi n / (M - 1)) & 0 \leq n \leq M - 1 \\ 0 & \text{其它} \end{cases}$$

海明窗：

$$w(n) = \begin{cases} 0.54 - 0.46 \cos(2\pi n / (M - 1)) & 0 \leq n \leq M - 1 \\ 0 & \text{其它} \end{cases}$$

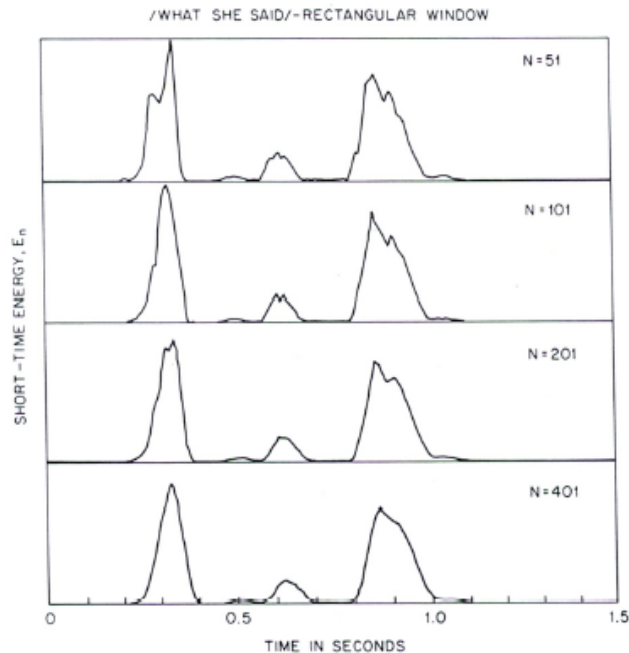
典型的窗函数的频谱



矩形窗谱平滑性能好，但损失高频成分，波形细节丢失，海明窗与之相反。

窗长度的选择

- 窗太长：等效于很窄的低通滤波器，此时随时间的变化很小，语音信号的变化细节就看不出来；
- 窗太短：滤波器的通带变宽，随时间有急剧的变化，不能得到平滑的信息。



标准：一帧内含有1~7个基音周期，
10-20ms。

语音短时过零分析

■ 定义：

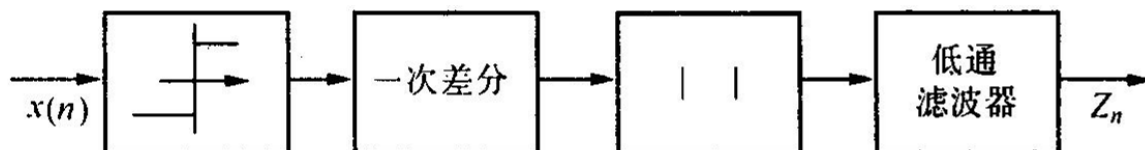
$$Z_n = \sum_{m=-\infty}^{\infty} |\text{sgn}[x(m)] - \text{sgn}[x(m-1)]| w(n-m)$$

$$= |\text{sgn}[x_w(m)] - \text{sgn}[x_w(m-1)]| * w(n)$$

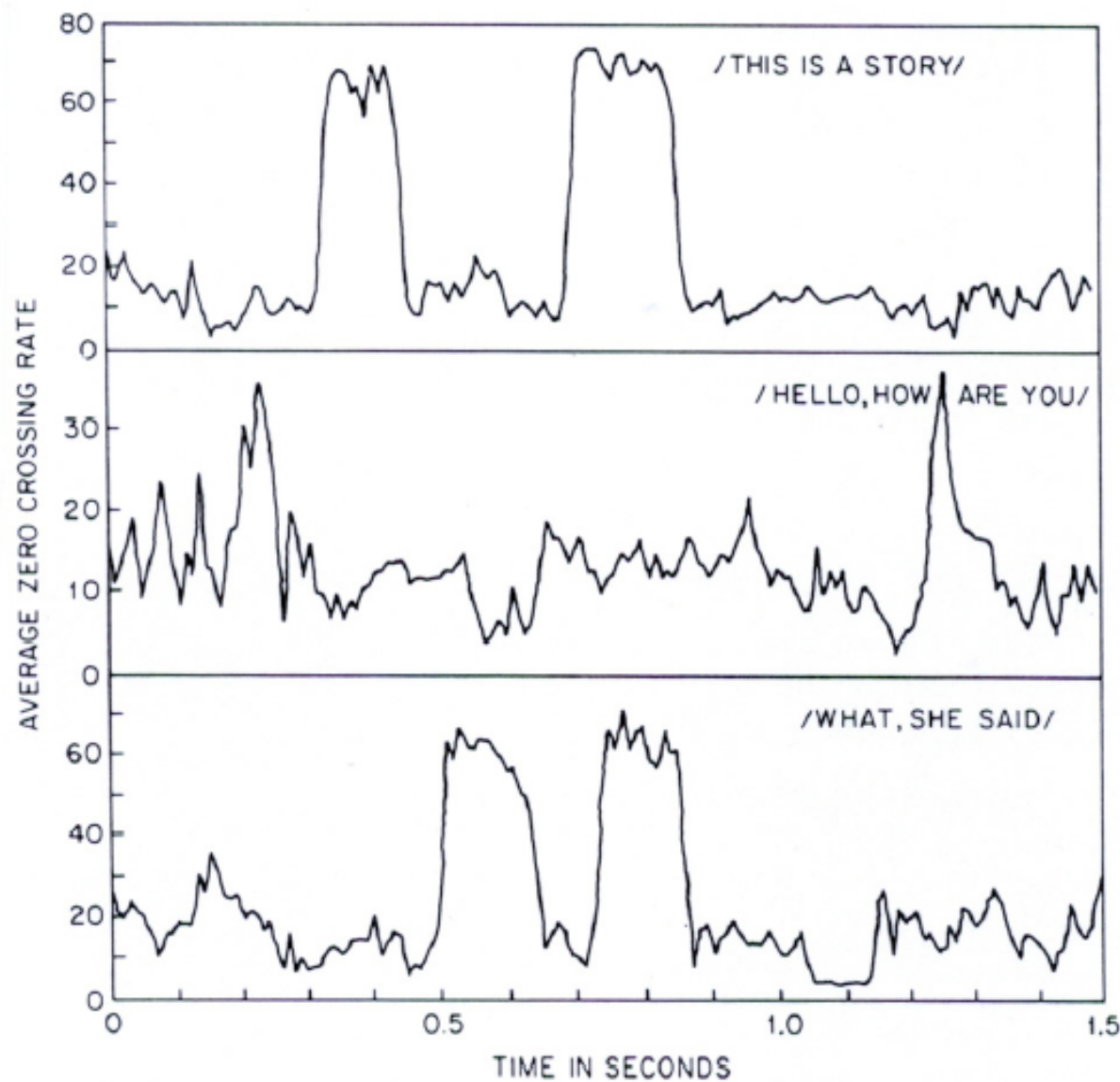
其中：

$$\text{sgn}[x(n)] = \begin{cases} 1 & x(n) \geq 0 \\ -1 & x(n) < 0 \end{cases} \quad w(n) = \begin{cases} 1/2N & 0 \leq n \leq N-1 \\ 0 & \text{其它} \end{cases}$$

● 框图：



语音短时过零分析

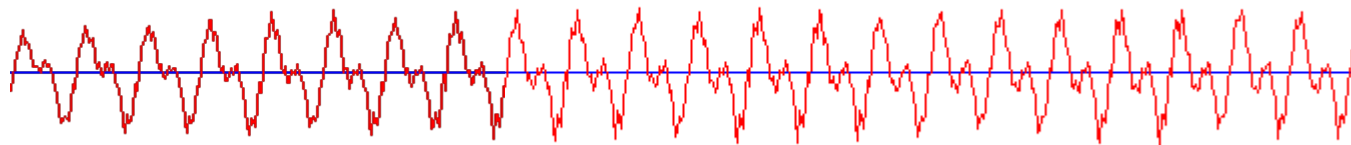


短时能量、短时过零率在语音上的特点

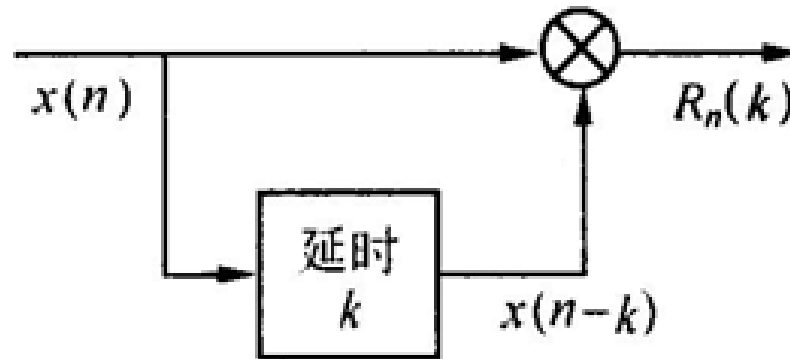
- 浊音的短时能量最大，过零率最低
- 清音短时能量居中，过零率最高
- 无声的短时能量最低，过零率居中
- 还可以帮助判断是否存在语音
 - 可用于判断寂静无语音和有语音的起点和终点位置。
 - 在背景噪声较小时用短时能量判断效果较好，而在背景噪声较大时用短时过零率判断较好。

短时相关分析

- 自相关用于研究信号本身，如信号波形的同步性、周期性等。



自相关函数



自相关函数
$$R(k) = \sum_{m=-\infty}^{+\infty} x(m)x(m-k)$$

短时自相关函数

$$R_n(k) = \sum_{m=-\infty}^{\infty} x(m)w(n-m)x(m-k)w(n-(m-k))$$

k是最大延时点数。


由于自相关函数是偶函数，所以上式可写成：

$$R_n(k) = R_n(-k) = \sum_{m=-\infty}^{\infty} x(m)x(m-k)[w(n-m)w(n-m+k)]$$


短时自相关函数

如果定义：

$$h_k(n) = w(n)w(n+k)$$

$$R_n(k) = \sum_{m=-\infty}^{\infty} x(m)x(m-k) \underline{[w(n-m)w(n-m+k)]}$$


则上式可写为：

$$R_n(k) = \sum_{m=-\infty}^{\infty} [x(m)x(m-k)] h_k(n-m) = [x(n)x(n-k)] * h_k(n)$$


所以，短时自相关函数可看作序列 $[x(n)x(n-k)]$ 通过单位样值响应为 $h_k(n)$ 的数字滤波器的输出。

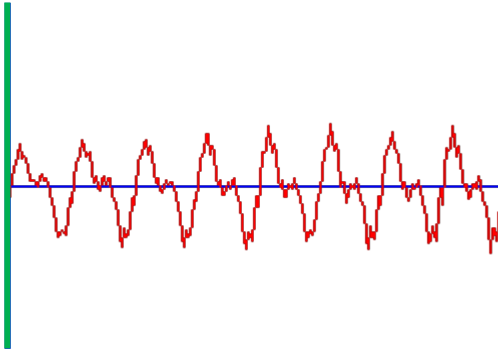
短时自相关函数

短时自相关分析在语音中可有下面两个方面的应用：

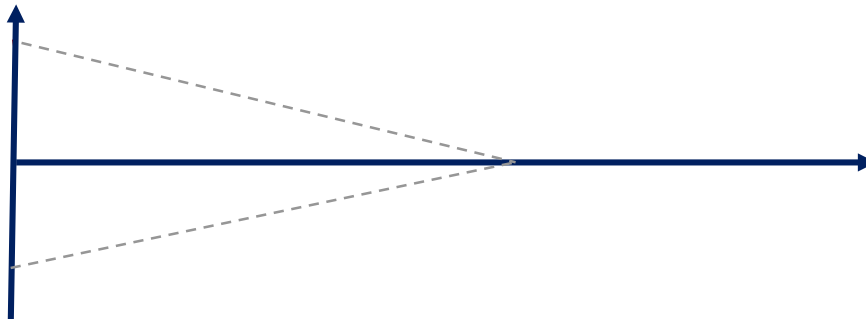
- 用来区分清音和浊音，因为浊音信号是准周期性的，对浊音语音可以用自相关函数求出语音波形序列的基音周期；
- 另外在进行语音信号的线性预测分析时，也要用到短时自相关函数。

自相关的问题

语音

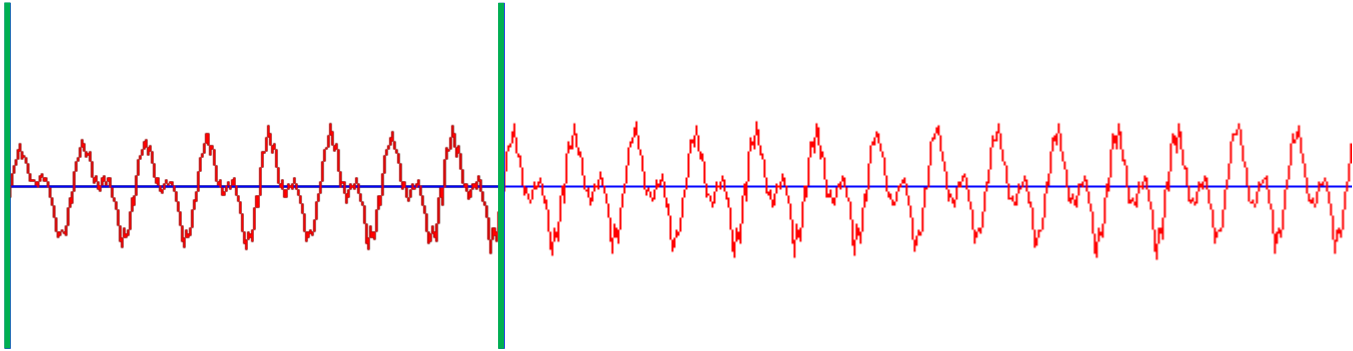


自相关



修正的自相关函数

语音



自相关



本节课提纲

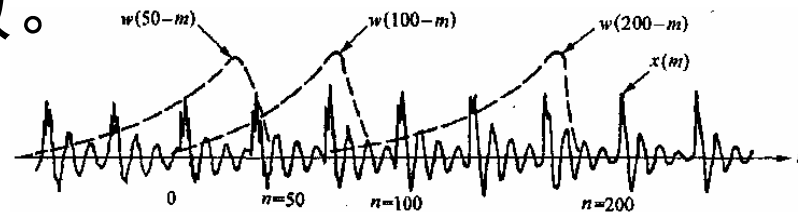
- 时域分析
- 频域分析
- 倒谱分析
- 线性预测分析
- 基音周期分析

短时傅里叶变换

短时傅里叶变换的定义：

$$X_n(e^{j\omega}) = \sum_{m=-\infty}^{\infty} x(m)w(n-m)e^{-j\omega m}$$

短时傅里叶变换有两个自变量： n 和 ω ；所以它既是关于时间 n 的离散函数，又是关于角频 ω 率的连续函数。



在几个 n 值上 $x(m)$ 与 $w(n-m)$ 的示意图

短时功率谱

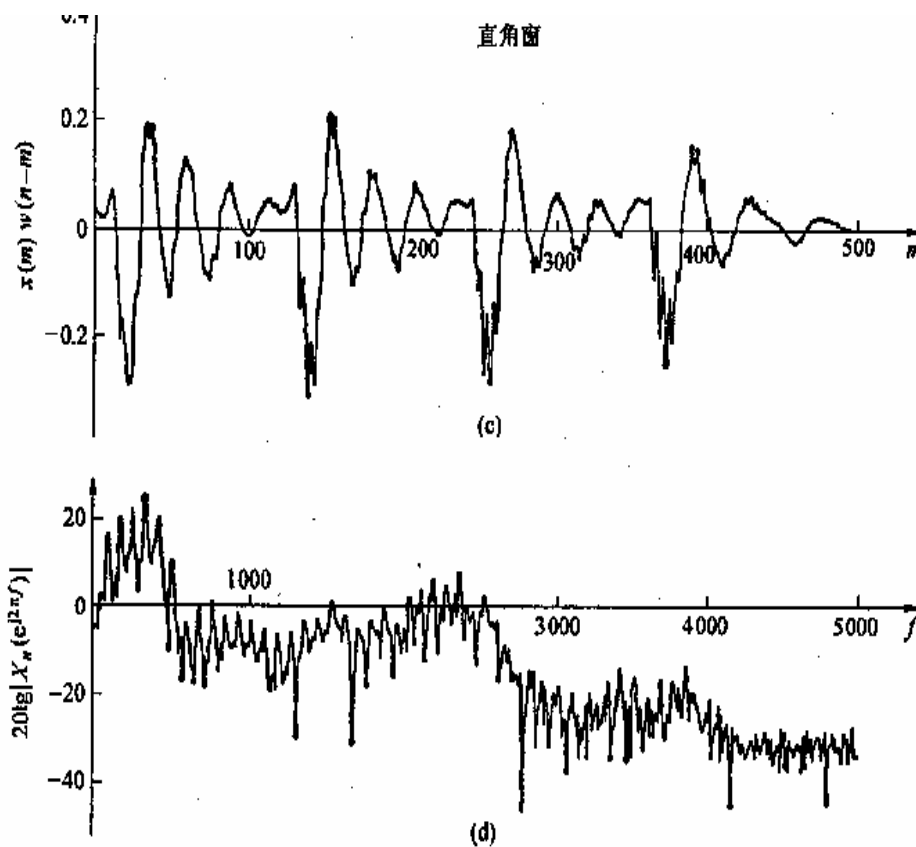
根据功率谱的定义，短时功率谱和短时傅里叶变换之间的关系为：

$$S_n(e^{j\omega}) = X_n(e^{j\omega})X_n^*(e^{j\omega}) = |X_n(e^{j\omega})|^2$$

短时功率谱是短时自相关函数的傅里叶变换：

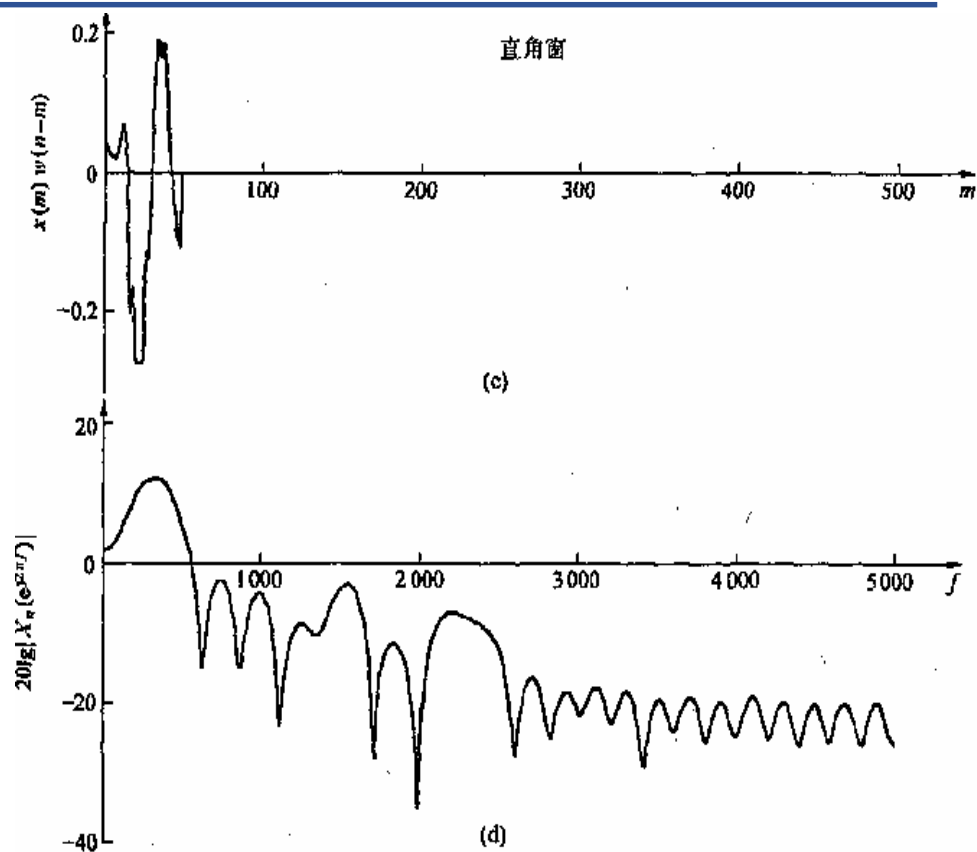
$$R_n(k) = \sum_{m=-\infty}^{\infty} x(m)w(n-m)x(m-k)w(n-m+k)$$

不同窗函数下的功率谱



N=500时海明窗与直角窗的功率谱分析

不同窗函数下的功率谱



N=50时海明窗与直角窗的功率谱分析

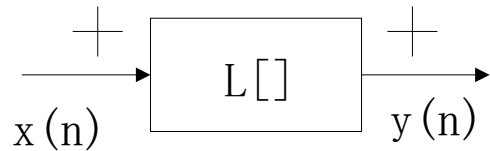
本节课提纲

- 时域分析
- 频域分析
- 倒谱分析
- 线性预测分析
- 基音周期分析

为什么要倒谱分析

■ 问题的提出

● 线性系统



$$y(n) = L[x(n)]$$

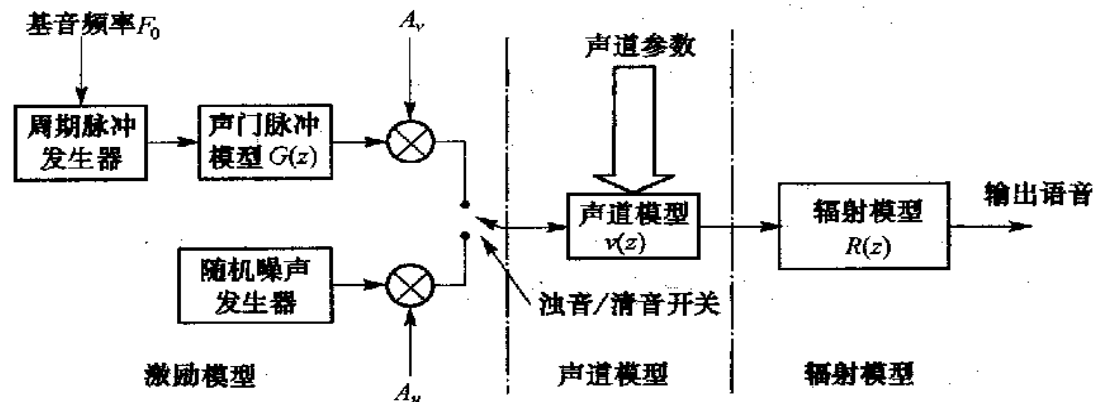
$$x_1(n) + x_2(n) \rightarrow L[x_1(n) + x_2(n)] = L[x_1(n)] + L[x_2(n)] = y_1(n) + y_2(n) = y(n)$$

$$ax(n) \rightarrow L[ax(n)] = aL[x(n)] = ay(n)$$

为什么要倒谱分析

■ 卷积同态系统

- 语音信号可视为声门激励信号与声道冲激响应的卷积
- 如何将二者分开



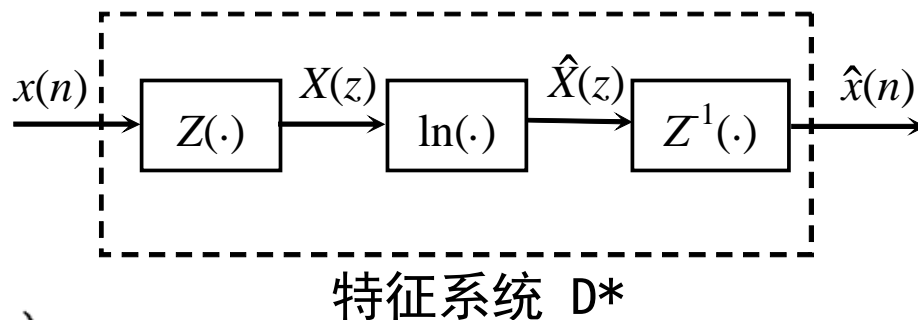
语音信号产生的离散时域模型

特征系统

$$x(n) = x_1(n) * x_2(n)$$

■ 1. Z变换

$$Z[x(n)] = X(z) = X_1(z) \cdot X_2(z)$$



■ 2. 对数运算

$$\hat{X}(z) = \ln X(z) = \ln X_1(z) + \ln X_2(z) = \hat{X}_1(z) + \hat{X}_2(z)$$

■ 3. 逆Z变换

$$\hat{x}(n) = Z^{-1}[\hat{X}(z)] = Z^{-1}[\hat{X}_1(z) + \hat{X}_2(z)] = \hat{x}_1(n) + \hat{x}_2(n)$$

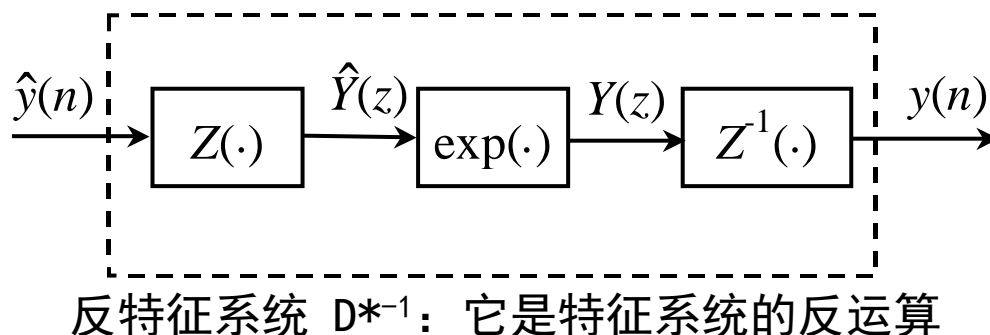
复倒谱

线性系统处理

$$\hat{y}(n) = \hat{y}_1(n) + \hat{y}_2(n) = L[\hat{x}_1(n)] + L[\hat{x}_2(n)]$$

可以用线性系统处理

逆特征系统



■ Z变换

$$\hat{Y}(z) = Z[\hat{y}(n)] = \hat{Y}_1(z) + \hat{Y}_2(z)$$

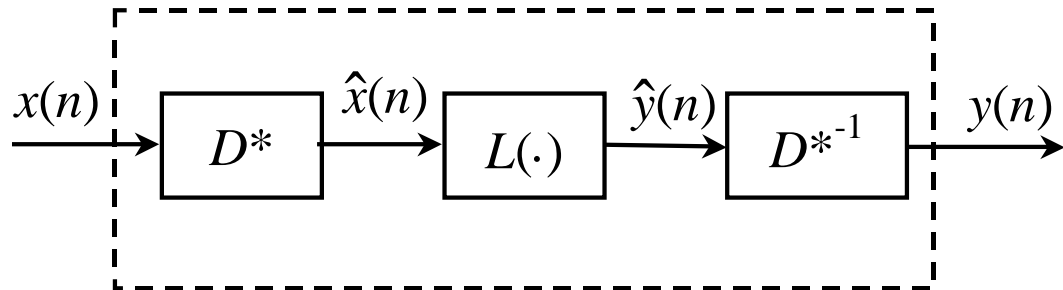
■ 指数运算

$$Y(z) = \exp[\hat{Y}(z)] = Y_1(z) \cdot Y_2(z)$$

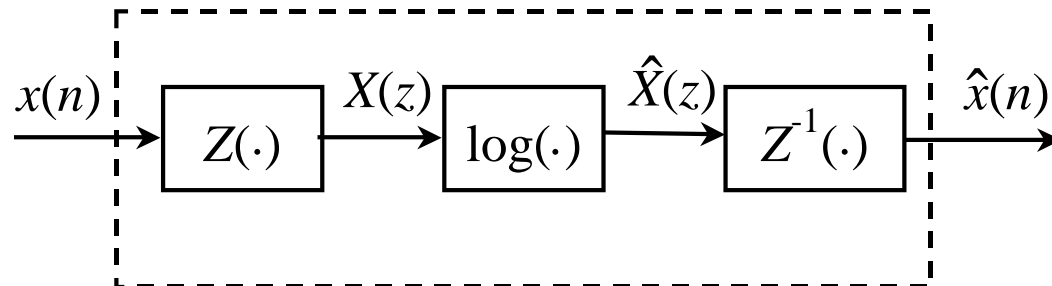
■ 逆Z变换

$$y(n) = Z^{-1}[Y_1(z) \cdot Y_2(z)] = y_1(n) * y_2(n)$$

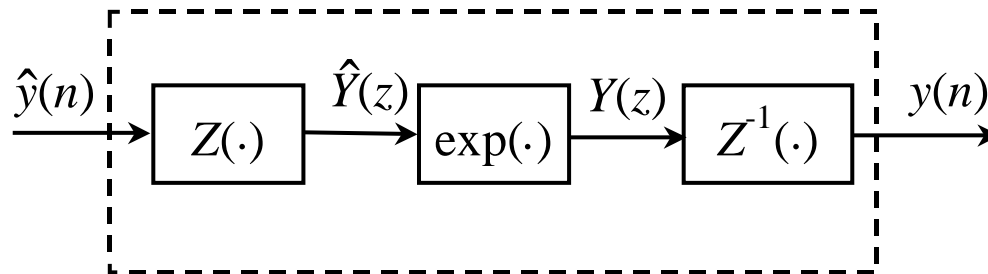
卷积同态系统



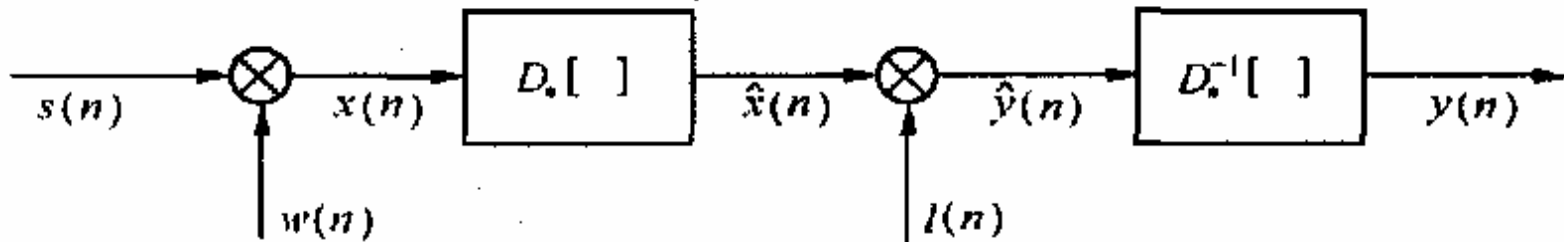
特征系统 D^*



反特征系统 D^{*-1} : 它是特征系统的反运算



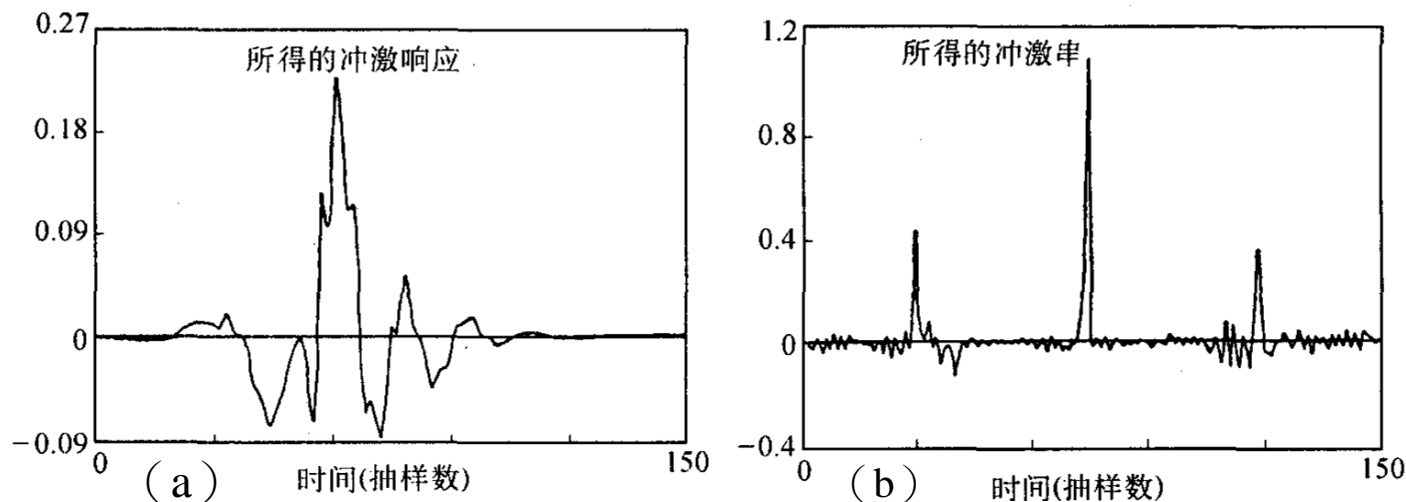
语音同态滤波系统



语音同态滤波系统的构成

先用窗 $w(n)$ 选择一个语音段，再计算倒谱，然后将得到的倒谱分量用一个“倒谱窗”分离出来。所得到的窗选倒谱用逆特征系统进行处理以恢复所需的卷积分量。

同态滤波分离出的语音声门激励和声道响应



浊音语音用同态滤波分离出声门激励和声道响应的示例

上图给出了经过滤波和逆特征系统处理后的结果。图(a)为经过低复倒谱窗 $l(n)$ 和 $D_*^{-1}[\]$ 之后的输出波形即声道冲击响应，图(b)给出了声门激励信号。可以看出声门激励波形近视于一个冲击串，其幅度随时间变化保持了用来加权输入信号所用的海明窗形状。

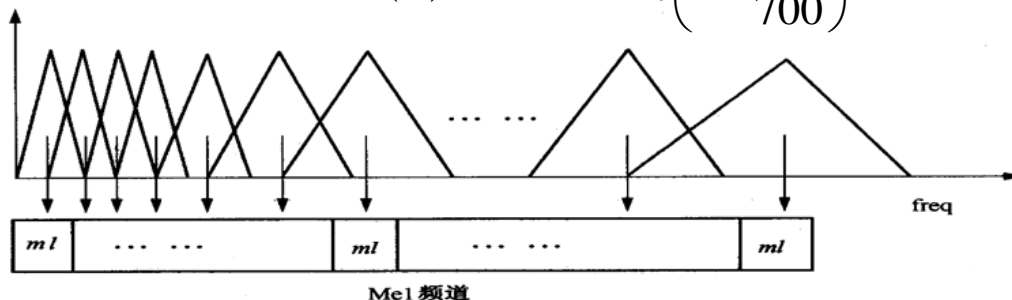
MeI 频率倒谱特征参数

MeI 频率倒谱参数利用了听觉原理和倒谱的解相关特性。另外，MeI 倒谱也具有对卷积性信道失真进行补偿的能力。由于这些原因，MeI 参数被证明是在语音识别任务中应用最成功的特征描述之一。

MFCC参数

借鉴人耳的听觉特性

$$Mel(f) = 2595 \log_{10} \left(1 + \frac{f}{700} \right)$$



Mel 频率尺度滤波器组

在 Mel 频率轴上配置 L 个通道的三角形滤波器组, L 的个数由信号的截止频率决定。每一个三角形滤波器的中心频率 $c(l)$ 在 Mel 频率轴上等间隔分配。设 $o(l)$ 、 $c(l)$ 和 $h(l)$ 分别是第 l 个三角形滤波器的下限、中心和上限频率, 则相邻三角形滤波器之间的下限、中心和上限频率有如图所示的如下关系成立:

$$c(l) = h(l-1) = o(l+1)$$

根据语音信号幅度谱 $|X_n(k)|$ 求每一个三角形滤波器的输出:

$$m(l) = \sum_{k=o(l)}^{h(l)} W_l(k) |X_n(k)| \quad l = 1, 2, \dots, L$$

$$W_l(k) = \begin{cases} \frac{k - o(l)}{c(l) - o(l)} & o(l) \leq k \leq c(l) \\ \frac{h(l) - k}{h(l) - c(l)} & c(l) \leq k \leq h(l) \end{cases}$$

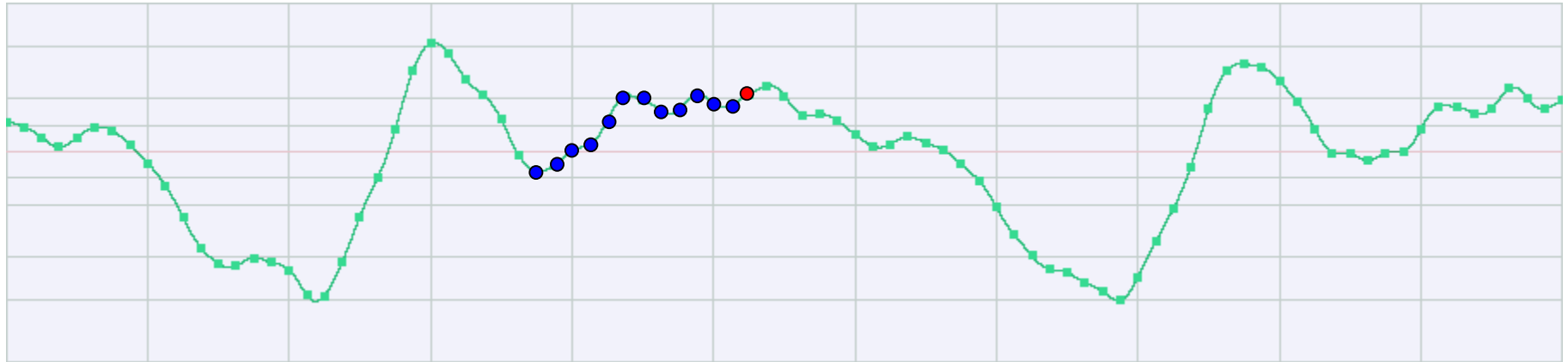
对所有滤波器输出做对数运算, 再进一步做离散余弦变换(DCT)即可得到 MFCC:

$$c_{mfcc}(i) = \sqrt{\frac{2}{N}} \sum_{l=1}^L \log m(l) \cos \left\{ \left(l - \frac{1}{2} \right) \frac{i\pi}{L} \right\}$$

本节课提纲

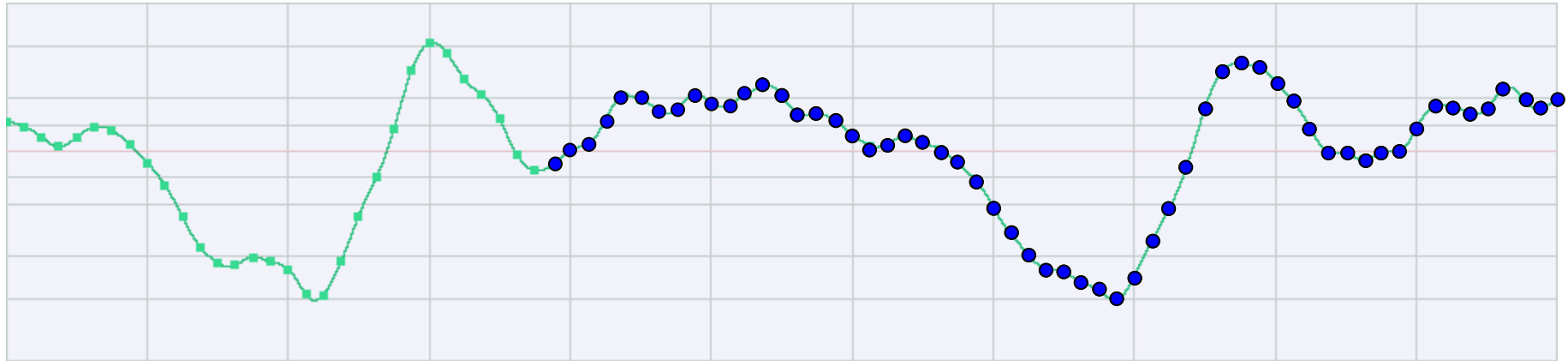
- 时域分析
- 频域分析
- 倒谱分析
- 线性预测分析
- 基音周期分析

线性预测



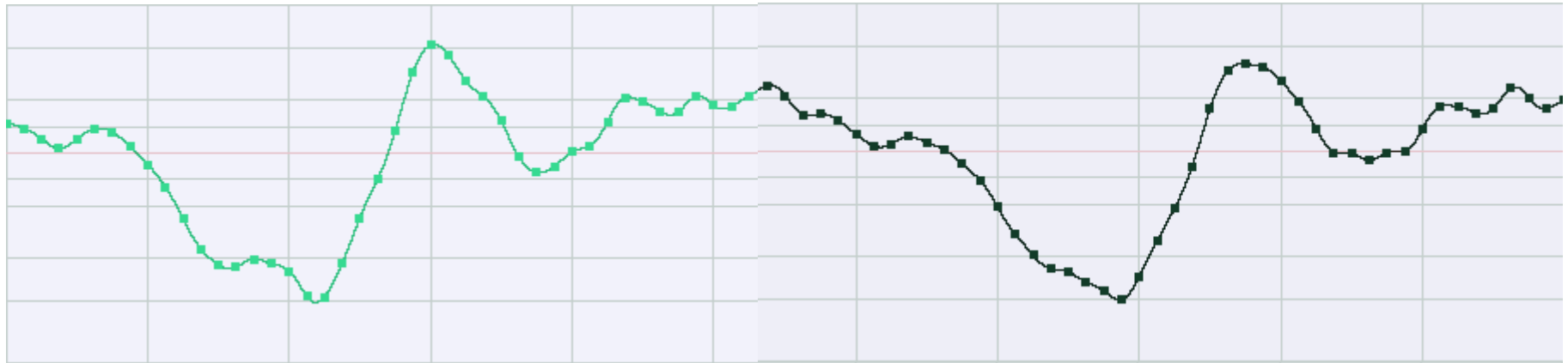
$$\tilde{s}(n) = \sum_{k=1}^p \alpha_k s(n-k)$$

线性预测



$$\tilde{s}(n) = \sum_{k=1}^p \alpha_k s(n-k)$$

线性预测



语音信号的线性预测分析

■ 最有效的语音分析技术之一

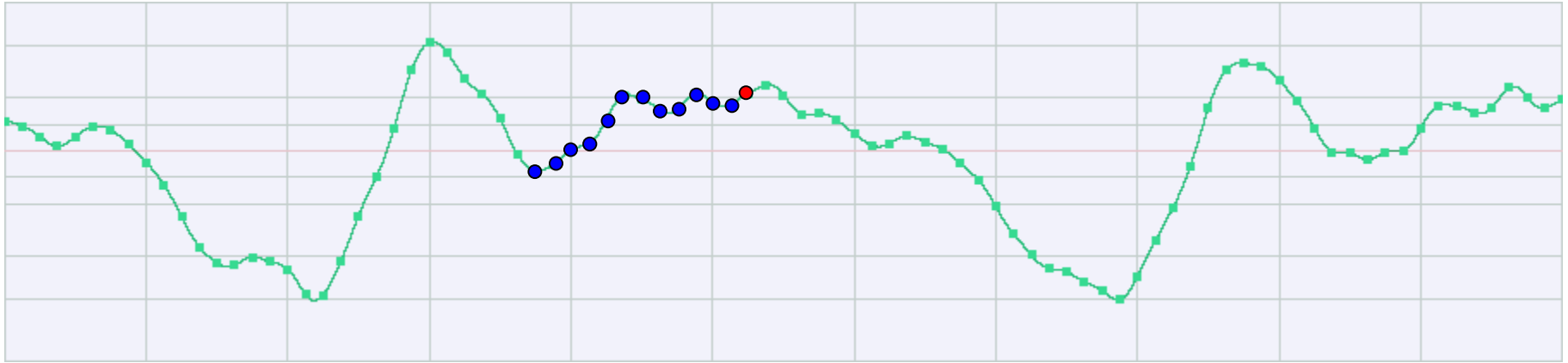
- 1947年维纳提出
- 1967年板仓等人应用于语音分析与合成

■ 语音信号处理与分析的核心技术

■ 用以估计语音的基本参数

- 基音、共振峰
- 频谱、声道截面积函数
- 特征参数
-

线性预测系数



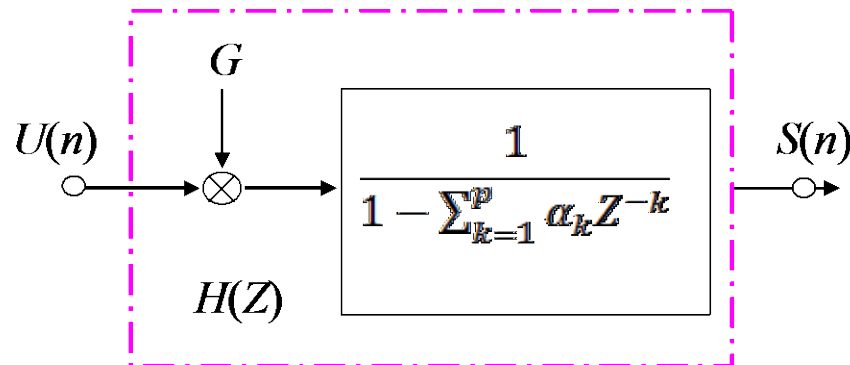
语音可以简化成：第 n 个语音采样 $S(n)$ 是前面 p 个采样的线性组合

$$s(n) = \sum_{k=1}^p \alpha_k s(n-k) + G \cdot U(n)$$

$\{\alpha_k\}$ $k = 1, 2, \dots, p$ 是线性预测系数



$$H(z) = \frac{G}{1 - \sum_{k=1}^p \alpha_k z^{-k}}$$



线性预测系数求解

短时平均预测误差：

$$E_n = \sum_m e_n^2(m) = \sum_m \left\{ s_n(m) - \sum_{k=1}^p \alpha_k s_n(m-k) \right\}^2$$

其中 $s_n(m)$ 是在 n 点附近选 m 语音： $s_n(m) = s(n+m)$

$$\text{令： } \frac{\partial E_n}{\partial \alpha_k} = 0, \quad k = 1, 2, \dots, p$$

得到

$$\sum_m s_n(m-i)s_n(m) = \sum_{k=1}^p \alpha_k \left\{ \sum_m s_n(m-i)s_n(m-k) \right\} \quad \text{其中 } 1 \leq i \leq p$$

这是一个 p 元方程组，通过求解这个方程组就可以计算出线性预测系数

经典解法：自相关法、协方差法；

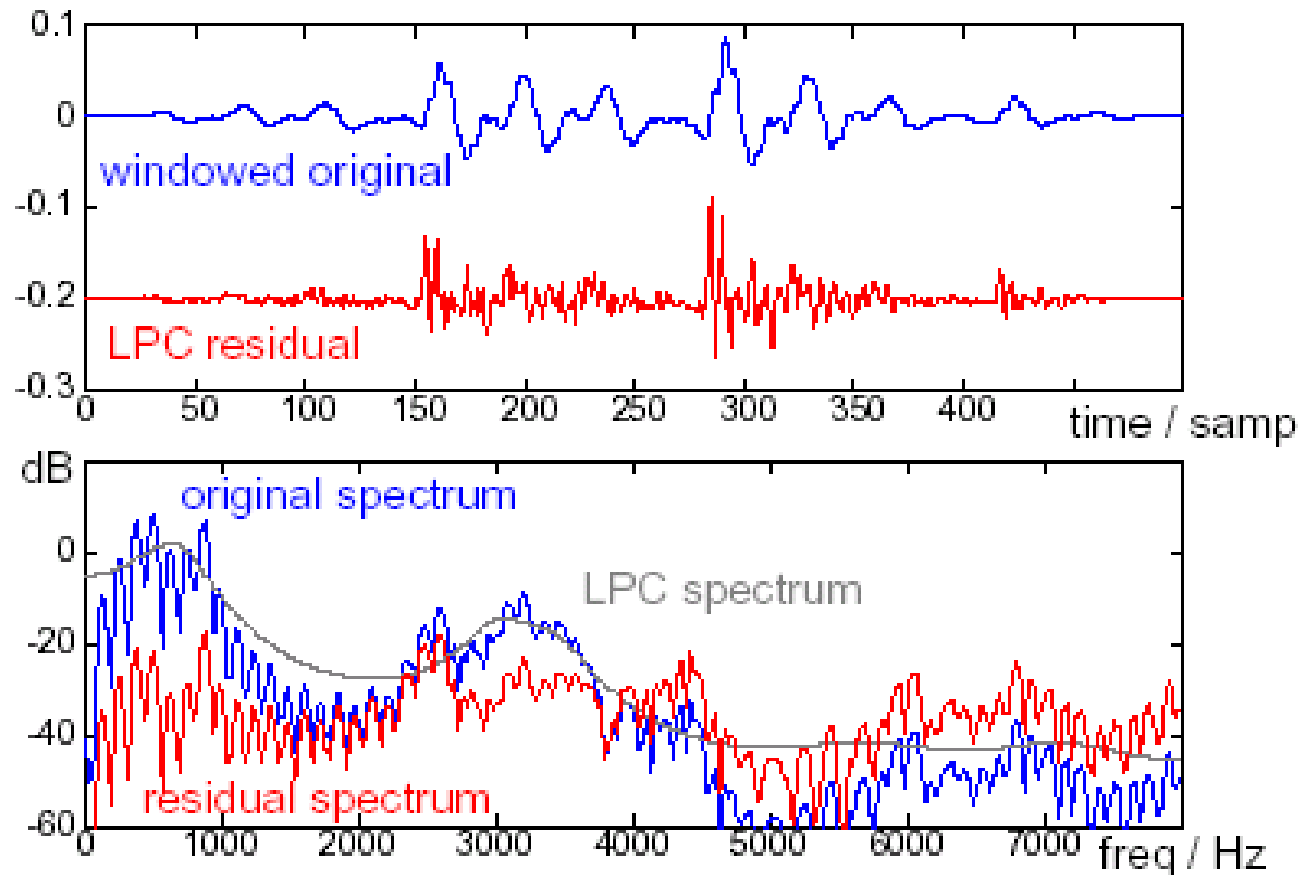
新方法和新技术：格型法、逆滤波器公式、谱估值公式、最大似然公式和内积公式。

LPC谱

当求出一组预测器系数后，就可以得到语音产生模型的频率响应，即：

可以预料在共振峰频率上其频率响应特性会出现峰值。所以线性预测分析法又可以看做是一种短时谱估计法。其频率响应 $H(e^{j\omega})$ 即称为LPC谱。也就是序列 $1, a_1, a_2, \dots, a_p$ 傅里叶变换的倒数。

LPC谱



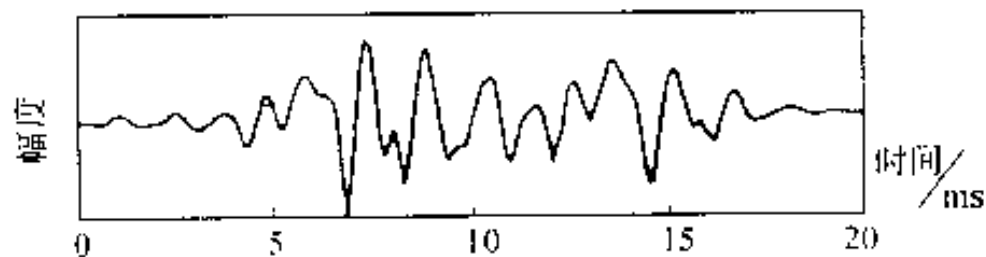
LPC参数阶数

P的选择要考虑频谱估计精度、计算量、存储量

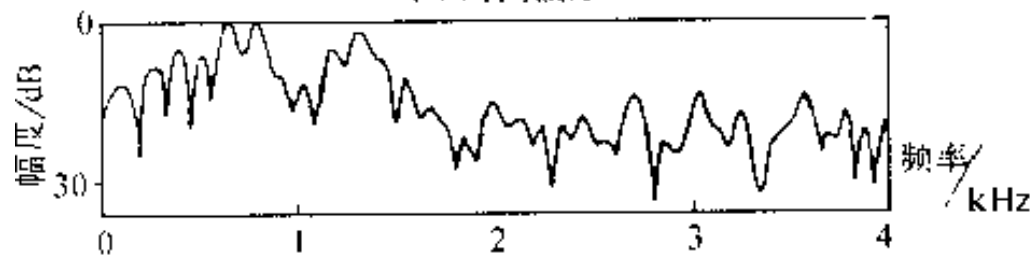
P值较大, 可使 $|H(e^{j\omega})|$ 精确匹配 $|S(e^{j\omega})|$ 但计算量、存储量较大。

原则: 要有足够多的极点来模型声道响应的谐振结构, 一般, 每1kHz需要一对共轭极点, 另外需要3~4个极点来逼近频谱中可能出现的零点及声门激励和辐射的组合效应, 16kHz取样下, p一般为12~14。

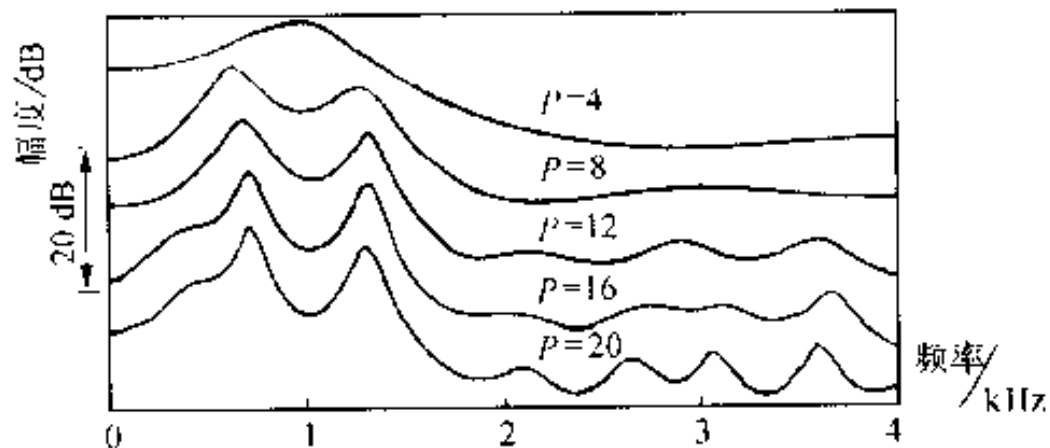
LPC参数阶数



(a) 时间波形



(b) 信号功率谱



(c) 不同阶数 p 的LPC谱

以 8 kHz 取样的一段元音[a] 的信号和功率谱

线谱频率

- ◆ 线谱频率 (Line Spectrum Frequency, LSF) 参数也称作线谱对 (Line Spectrum Pair, LSP) 参数, 它是线性预测系数的一种重要推演参数。
- ◆ LSF在数学角度上完全等价于其它的线性预测参数, 而且LSF参数还有一些特别的性质是其它的参数所不具备的, 例如, 一个LSF参数的误差仅仅影响全极点模型中邻近这个参数对应频率处的语音谱。
- ◆ 如果把声道等效为由声管级联而成, 则线谱频率参数表示声门完全开启或完全闭合状态下声管的谐振频率。

线谱频率

语音信号的线性预测模型可以表示为一个 P 阶的全极点数字滤波器 $H(z)$

通常 P 为偶数:

$$H(z) = \frac{1}{A(z)}$$

式中:

$$A(z) = 1 - \sum_{k=1}^p a_k z^{-k}$$

定义两个多项式:

$$P(z) = A(z) + z^{-(p+1)} A(z^{-1})$$

$$Q(z) = A(z) - z^{-(p+1)} A(z^{-1})$$

P+1阶

它们的零点是复数, 其相位表示的频率就是LSF参数。

$$A(z) = \frac{1}{2}[P(z) + Q(z)]$$

线谱频率

可以推出：

$$P(z) = 1 - (a_1 + a_p)z^{-1} - (a_2 + a_{p-1})z^{-2} \cdots - (a_p + a_1)z^{-p} + z^{-(p+1)}$$

$$Q(z) = 1 - (a_1 - a_p)z^{-1} - (a_2 - a_{p-1})z^{-2} \cdots - (a_p - a_1)z^{-p} - z^{-(p+1)}$$

求解方法：

- 1) 代数方程求根
- 2) DFT法

它们都有共轭的复根。此外，它们分别有-1和+1的实根，即：

$$P(z)|_{z=-1} = 0 \quad Q(z)|_{z=1} = 0$$

$$0 < \omega_1 < \omega_2 < \omega_3 \cdots < \omega_p < \pi$$

即线谱频率为 $f_k = \omega_k / 2\pi, k = 1, \cdots, p$

本节课提纲

- 时域分析
- 频域分析
- 倒谱分析
- 线性预测分析
- 基音周期分析

基音周期提取的难点

- 语音并不是完全周期性的
- 噪声影响
- 发音的变化

自相关法

- 浊音信号的自相关函数在基音周期的整数倍位置上出现峰值，而清音的自相关函数没有明显的峰值出现。
- 峰—峰值之间对应的就是基音周期。
- 基音的周期性和共振峰的周期性混在一起时，被检测出来的峰值就可能会偏离原来峰值的真实位置。

语音信号先进行预处理

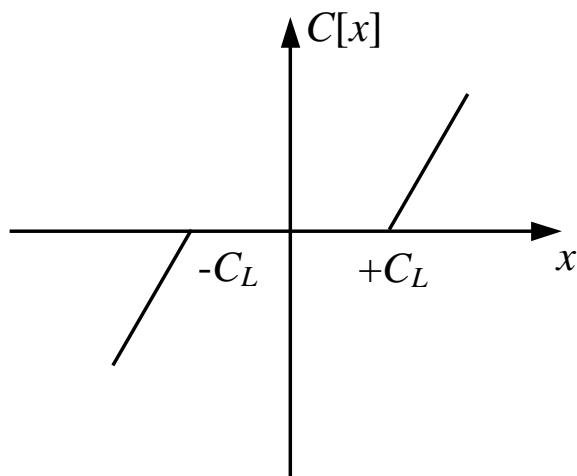
- 两个问题：
 - 1) 窗函数：选择矩形窗，且窗长大于两个基音周期。
 - 2) 去除声道的影响。

解决方法：

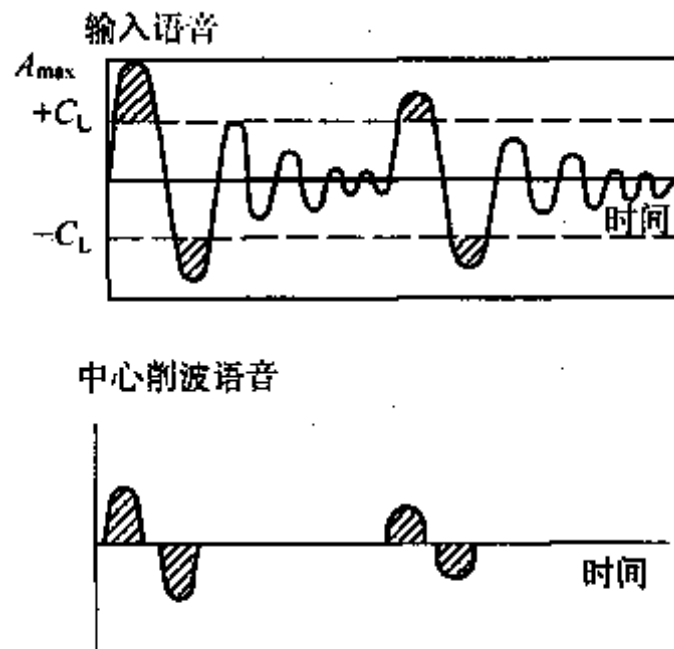
- 1) 减小共振峰影响，使用一个60~900Hz的带通滤波器。
- 2) 进行非线性变换，如中心削波。

中心削波

中心削波即是一种非线性处理，用以削除语音信号的低幅度部分，即 $y(n) = C[x(n)]$ ，其削波特性及工作过程如下图所示。

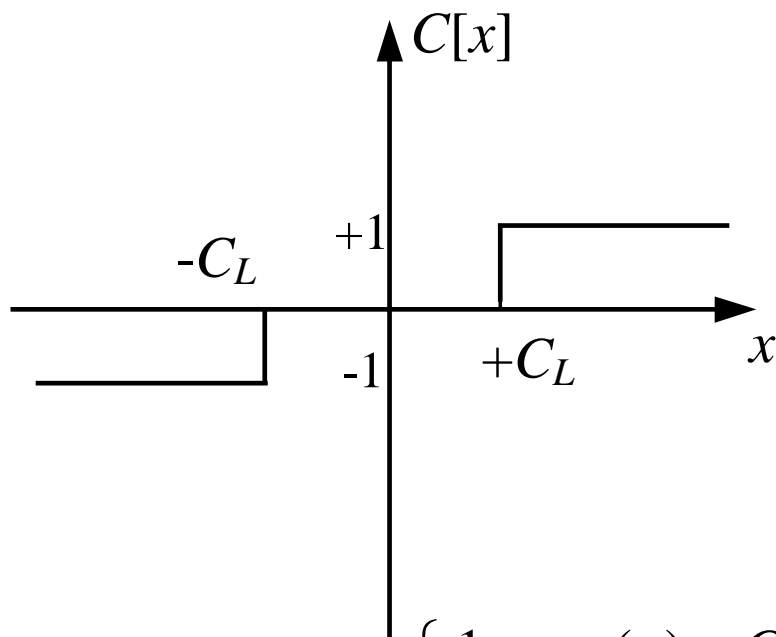


$$y(n) = C[x(n)] = \begin{cases} x(n) - C_L & x(n) > C_L \\ 0 & |x(n)| \leq C_L \\ x(n) + C_L & x(n) < -C_L \end{cases}$$



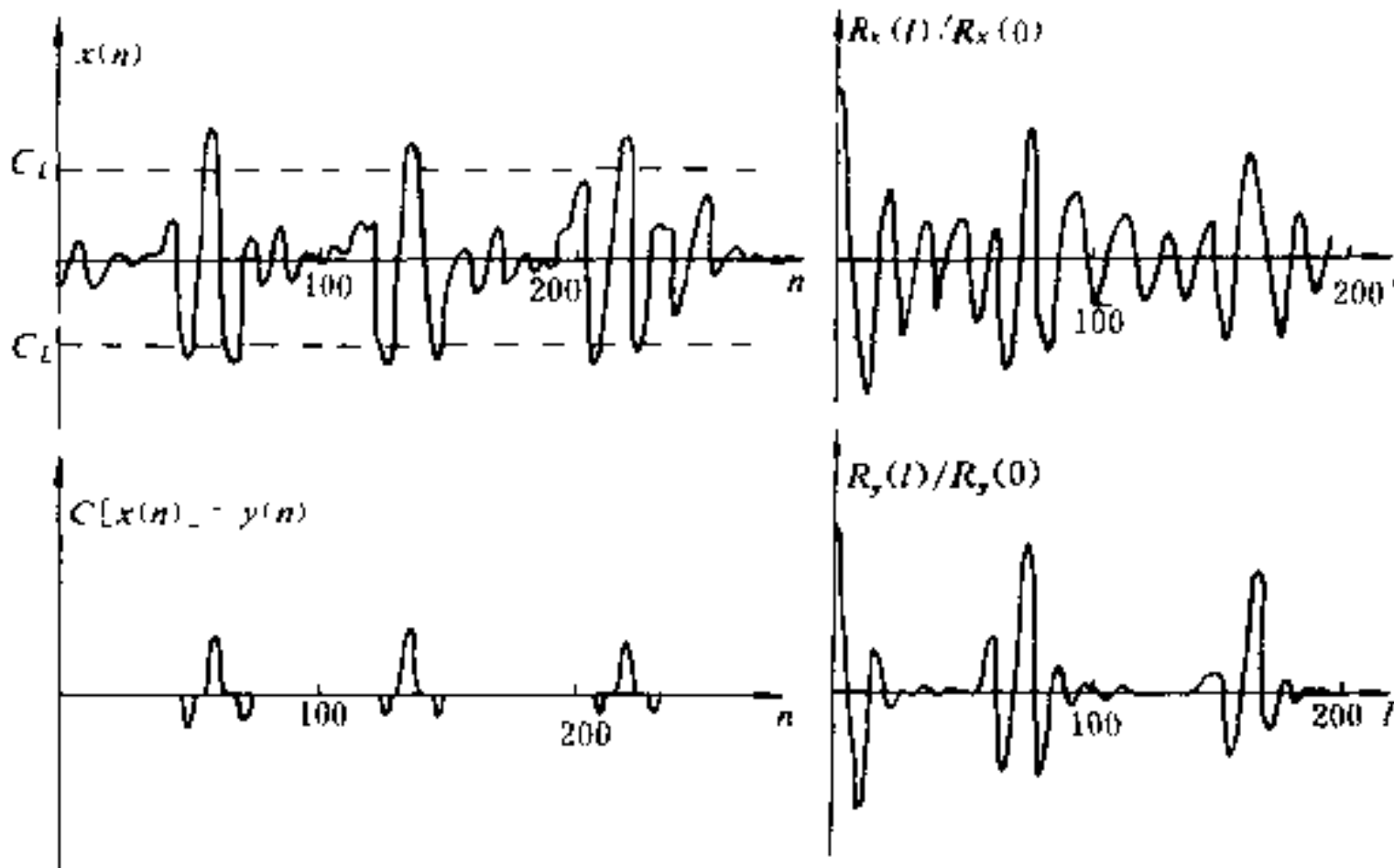
中心削波

为了减少运算量，可以采用三电平中心削波



$$y(n) = C[x(n)] = \begin{cases} 1 & x(n) > C_L \\ 0 & |x(n)| \leq C_L \\ -1 & x(n) < -C_L \end{cases}$$

实例：自相关+中心削波



语音信号经过中心削波后自关函数具有更尖锐峰起的示例

本节课提纲

- 时域分析
- 频域分析
- 倒谱分析
- 线性预测分析
- 基音周期分析

谢谢！