



噪声抑制与 声学回声消除

主讲人 宋辉

清华大学电子工程系 博士
滴滴AI Labs 语音技术部



 5.1 子带分解

 5.2 噪声抑制

 5.3 声学回声消除

 5.4 实战



为什么需要声学回声消除和噪声抑制



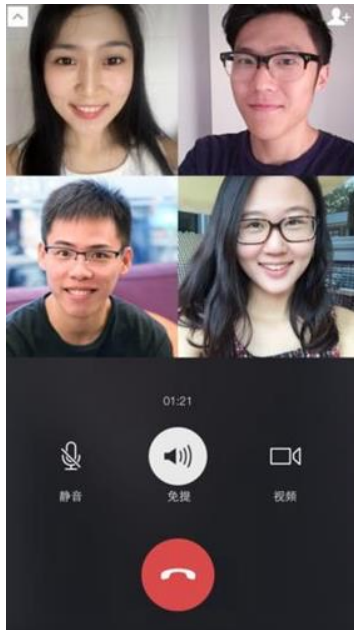
回声消除 (AEC)

消除设备自身的干扰



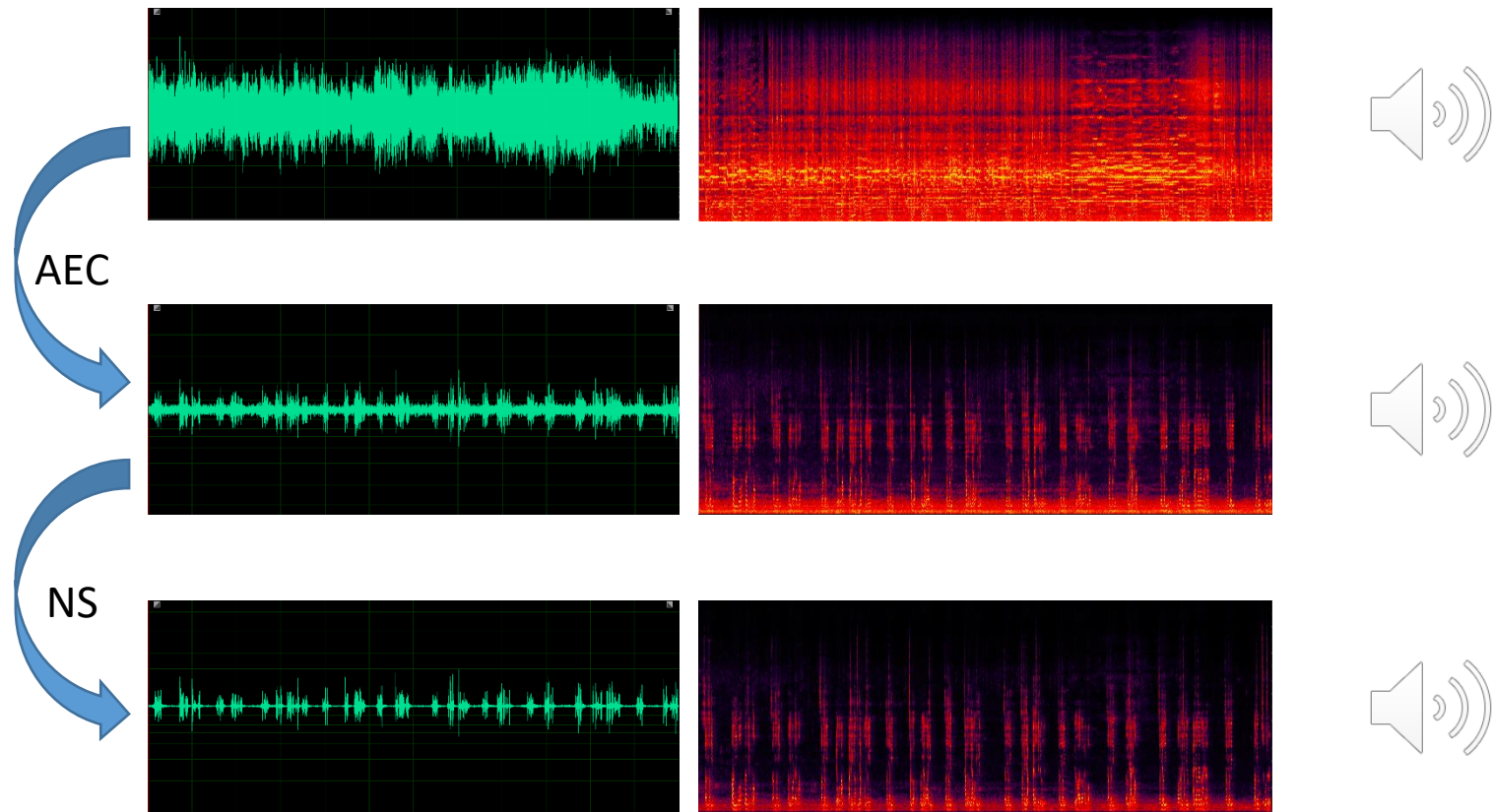
噪声抑制 (NS)

抑制环境噪声





案例





5.1 子带分解——DFT filterbank



信号的分频带处理，是语音信号处理中的常用手段。



DFT

$$X(k) = \sum_{n=0}^{N-1} x(n)W_N^{nk}$$



$$X(k, l) = \sum_{n=0}^{N-1} x(l+n)W_N^{nk}$$

k : frequency bin index
 l : frame index
 N : frame length





DFT filterbank



选择合适的抽取因子 M

$$X(k, l) = \sum_{n=0}^{N-1} x(lM + n) W_N^{nk}$$



选择窗函数 (长度为 N)

$$X(k, l) = \sum_{n=0}^{N-1} x(lM + n) w(n) W_N^{nk}$$



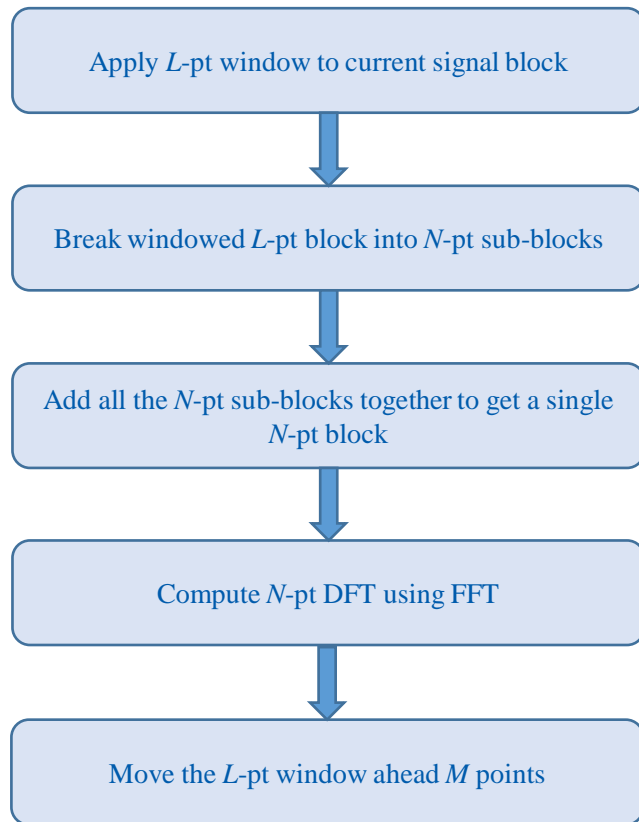
DFT filterbank



考虑使用更长的窗函数（长度为 L ）

$$X(k, l) = \sum_{n=0}^{L-1} x(lM + n)w(n)W_N^{nk}$$

NOT a DFT!!





5.2 噪声抑制 (Noise Suppression, NS)

顾名思义，噪声抑制技术用于消除背景噪音，改善语音信号的信噪比和可懂度，让人和机器听得更清楚。



常见的噪声



CAF



STR



BUS



PED



噪声抑制方法的种类

单通道降噪

平稳噪声

被动降噪

多通道降噪

非平稳噪声

主动降噪



5.2 噪声抑制 (Noise Suppression, NS)



评价指标

Signal-to-noise ratio (SNR)

Signal-to-distortion ratio (SDR)

Mean opinion score (MOS)

Perceptual evaluation of speech quality (PESQ)

Short-time objective intelligibility (STOI)

Word error rate (WER)

Minima Controlled Recursive Averaging (MCRA)

 噪声抑制的核心是对噪声的估计 (noise estimation / noise tracking)

 噪声估计的常用方法

递归平均
Recursive averaging

最小值追踪
Minima controlled

直方图统计
histogram

Minima Controlled Recursive Averaging (MCRA)

Recursive averaging

只要某个频带中语音存在的概率很低，则可以利用这个频带估计/更新噪声谱。

Minima controlled / tracking

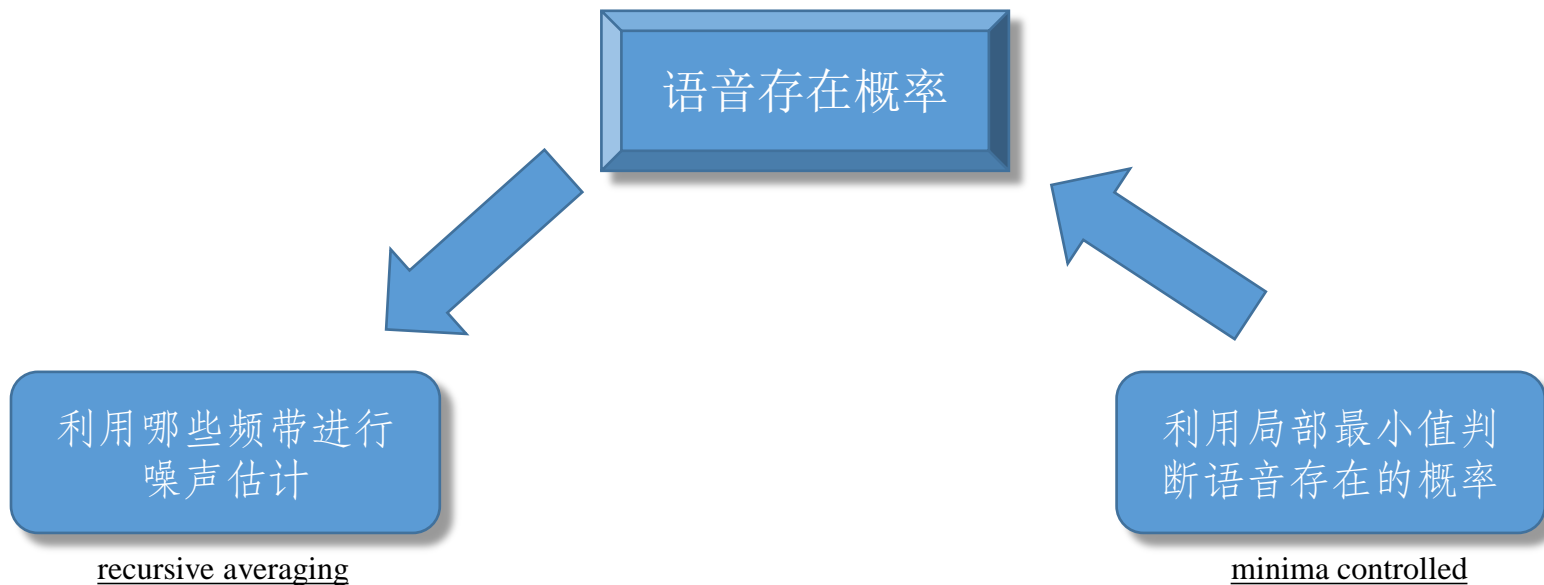
由于语音信号的稀疏性，即使语音存在，在一个短时窗（0.5s ~ 1.5s）内，每个频带的最小值也会大概率趋近于噪声功率。因此可以通过在短时窗内追踪最小值的方法获得每个频带的噪声估计。

Histogram-based method

带噪语音信号的每个频带在一个短时窗内做直方图统计，出现频次最高的值对应这个频带的噪声水平。

Minima Controlled Recursive Averaging (MCRA)

 speech presence probability



Minima Controlled Recursive Averaging (MCRA)

 考虑带有加性噪声的数据模型:

$$y(n) = x(n) + d(n)$$

输入信号经过overlapping分帧、加窗、傅里叶变换:

$$Y(k, \ell) = \sum_{n=0}^{N-1} y(n + \ell M) h(n) e^{-j \frac{2\pi}{N} nk} \quad (5.1)$$

k : frequency bin index
 ℓ : time frame index
 h : analysis window of size N (frame length)
 M : frame shift

Minima Controlled Recursive Averaging (MCRA)

 给定两个假设 $H_0(k, \ell)$ 和 $H_1(k, \ell)$, 分别表示 “语音不存在” 和 “语音存在” :

$$\begin{aligned} H_0(k, \ell) : Y(k, \ell) &= D(k, \ell) \\ H_1(k, \ell) : Y(k, \ell) &= X(k, \ell) + D(k, \ell) \end{aligned} \quad (5.2)$$

定义 $\lambda_d(k, \ell) = E[|D(k, \ell)|^2]$ 。

噪声估计

采用时间递归平滑的方式, 当 “语音不存在” 时, 迭代更新 $\lambda_d(k, \ell)$:

$$\begin{aligned} H'_0(k, \ell) : \hat{\lambda}_d(k, \ell + 1) &= \alpha_d \hat{\lambda}_d(k, \ell) + (1 - \alpha_d) |Y(k, \ell)|^2 \\ H'_1(k, \ell) : \hat{\lambda}_d(k, \ell + 1) &= \hat{\lambda}_d(k, \ell) \end{aligned} \quad (5.3)$$

其中 α_d ($0 < \alpha_d < 1$) 是平滑因子。

Minima Controlled Recursive Averaging (MCRA)

 speech presence probability:

$$p'(k, \ell) \triangleq P(H'_1(k, \ell) | Y(k, \ell)) \quad (5.4)$$

则噪声估计可以表示为:

$$\begin{aligned} \hat{\lambda}_d(k, \ell + 1) &= \hat{\lambda}_d(k, \ell) p'(k, \ell) \\ &+ \left[\alpha_d \hat{\lambda}_d(k, \ell) + (1 - \alpha_d) |Y(k, \ell)|^2 \right] (1 - p'(k, \ell)) \\ &= \tilde{\alpha}_d(k, \ell) \hat{\lambda}_d(k, \ell) + [1 - \tilde{\alpha}_d(k, \ell)] |Y(k, \ell)|^2 \end{aligned} \quad (5.5)$$

其中

$$\tilde{\alpha}_d(k, \ell) \triangleq \alpha_d + (1 - \alpha_d) p'(k, \ell)$$

现在的问题是：如何确定 $p'(k, \ell)$



语音存在概率




计算语音存在概率的思路：

在一个短时窗内，计算局部能量（local energy，通常用 $S(k, \ell)$ 表示）与最小能量 $S_{\min}(k, \ell)$ 的比值。

频域平滑：

$$S_f(k, \ell) = \sum_{i=-w}^w b(i) |Y(k-i, \ell)|^2. \quad (5.6)$$

 频域窗

时域平滑：

$$S(k, \ell) = \alpha_s S(k, \ell - 1) + (1 - \alpha_s) S_f(k, \ell), \quad (5.7)$$

其中 α_s ($0 < \alpha_s < 1$) 是local energy的时域平滑因子。



语音存在概率



计算最小能量: $S_{\min}(k, \ell)$

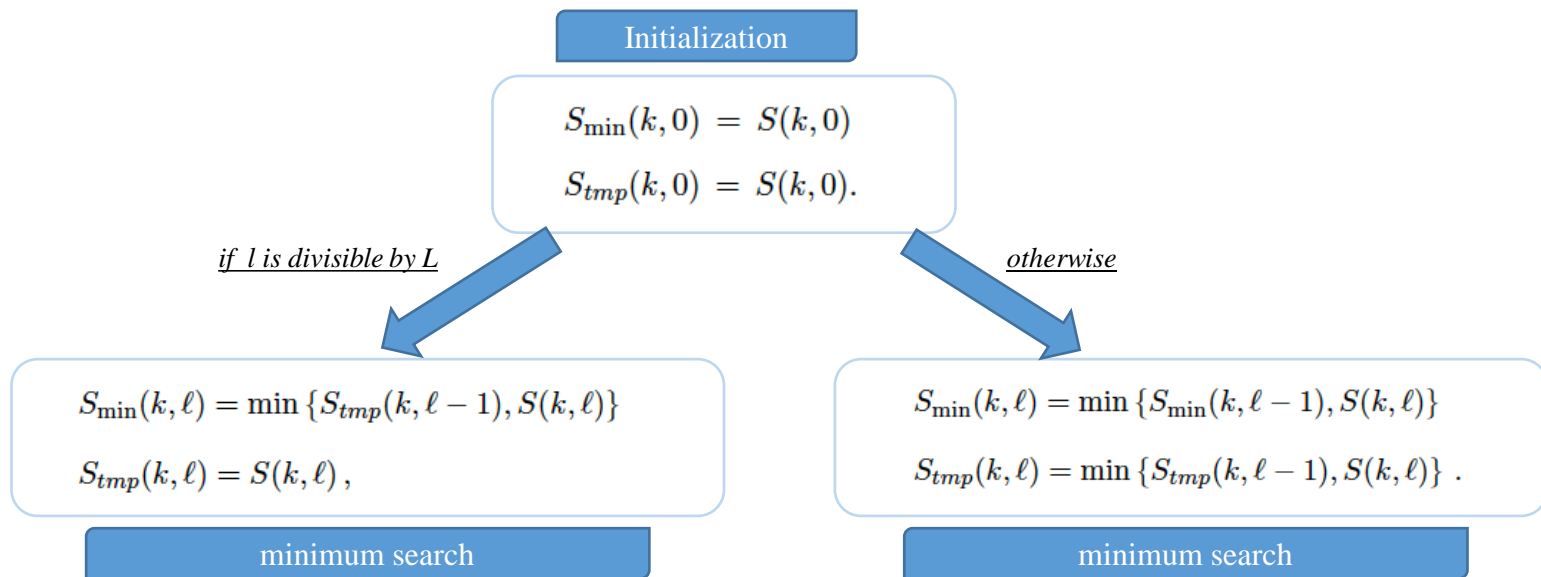
采用局部最小值追踪的方法, 设定一个时间窗 L , 搜索局部最小值。



语音存在概率



计算 $S_{\min}(k, \ell)$ 的另一种简化方法:





语音存在概率



定义: $S_r(k, \ell) \triangleq S(k, \ell) / S_{\min}(k, \ell)$

则语音是否存在的判决式为:

$$S_r(k, \ell) \underset{H'_0}{\overset{H'_1}{\geq}} \delta. \quad (5.8)$$

根据上式, 可以得到语音存在概率的迭代估计:

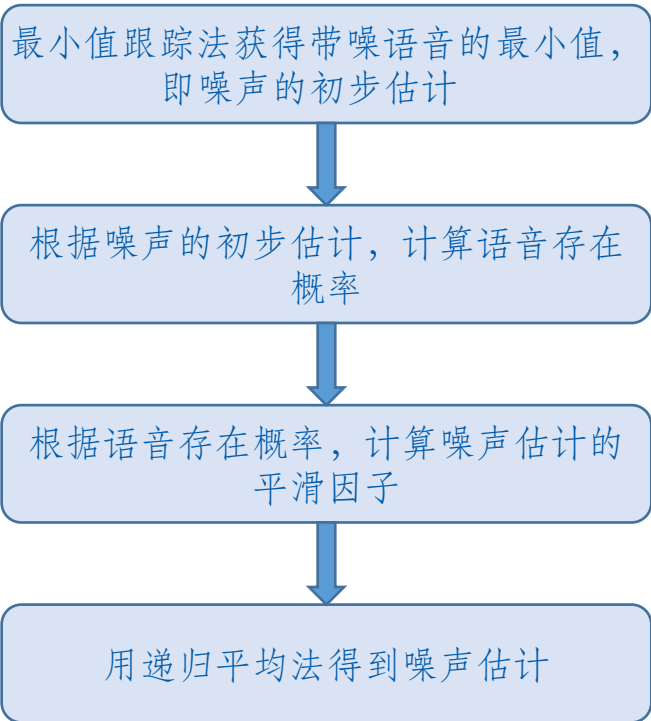
$$\hat{p}'(k, \ell) = \alpha_p \hat{p}'(k, \ell - 1) + (1 - \alpha_p) I(k, \ell) \quad (5.9)$$

$$I(k, \ell) = 1 \text{ if } S_r(k, \ell) > \delta \text{ and } I(k, \ell) = 0 \text{ otherwise.}$$

其中 α_p ($0 < \alpha_p < 1$) 是语音存在概率的平滑因子。



MCRA噪声估计流程



$$S_{\min}(k, \ell)$$

$$\hat{p}'(k, \ell) = \alpha_p \hat{p}'(k, \ell - 1) + (1 - \alpha_p) I(k, \ell)$$

$$\tilde{\alpha}_d(k, \ell) \triangleq \alpha_d + (1 - \alpha_d) p'(k, \ell)$$

$$\hat{\lambda}_d(k, \ell + 1) = \tilde{\alpha}_d(k, \ell) \hat{\lambda}_d(k, \ell) + [1 - \tilde{\alpha}_d(k, \ell)] |Y(k, \ell)|^2$$



MCRA的参考参数列表

$$N = 512 \text{ (32ms)}$$

$$M = 128 \text{ (75\% overlap)}$$

$$\alpha_d = 0.95$$

$$w = 1$$

$$\alpha_s = 0.8$$

$$L = 125 \text{ (1s minima search window)}$$

$$\delta = 5$$

$$\alpha_p = 0.2$$



增益因子 G 的确定



当噪声估计已知后，如何进行噪声抑制？

“增益因子与带噪语音相乘” 即可。

$$\hat{X}(k, l) = G(k, l)Y(k, l)$$

或者写成：

$$|\hat{X}(k)|^2 = G^2(k)|Y(k)|^2 \quad (5.10)$$



确定增益因子的方法：

谱减法、维纳滤波、MMSE、……



增益因子 G 的确定



谱减法——假设噪声平稳或缓慢变化，从带噪语音谱中，将噪声谱减掉

根据等式：

$$|\hat{X}(k)|^2 = G^2(k)|Y(k)|^2$$

可以得到增益因子的表达式：

$$\begin{aligned} G(k) &= \sqrt{\frac{|\hat{X}(k)|^2}{|Y(k)|^2}} = \sqrt{\frac{|Y(k)|^2 - |\hat{D}(k)|^2}{|Y(k)|^2}} = \sqrt{\frac{|Y(k)|^2 - \lambda_d(k)}{|Y(k)|^2}} \\ &= \sqrt{1 - \frac{\lambda_d(k)}{|Y(k)|^2}} = \sqrt{1 - \frac{1}{\gamma(k)}} \end{aligned} \quad (5.11)$$

后验信噪比



增益因子 G 的确定



频域维纳滤波——估计出的纯净语音幅度谱与真实幅度谱的均方误差最小

频域估计误差：

$$E(k) = X(k) - \hat{X}(k) = X(k) - G(k)Y(k) \quad (5.12)$$

目标函数：

$$J = E[|E(k)|^2]$$

最小化目标函数，可以得到频域维纳滤波器的增益因子：

$$G(k) = \frac{\lambda_x(k)}{\lambda_x(k) + \lambda_d(k)} = \frac{\xi(k)}{\xi(k) + 1} \quad (5.13)$$

先验信噪比

等式(5.3.13)的证明，请大家自行推导。

$$\lambda_x(k) = E[|X(k)|^2]$$



后验信噪比 & 先验信噪比



后验信噪比

$$\gamma(k) = \frac{|Y(k)|^2}{\lambda_d(k)} \quad (5.14)$$

先验信噪比

$$\xi(k) = \frac{\lambda_x(k)}{\lambda_d(k)} \quad (5.15)$$

根据后验信噪比估计先验信噪比 —— 判决引导法 (Decision Direct)

$$\xi(k, l) = \alpha_{DD} \frac{|\hat{X}(k, l-1)|^2}{\lambda_d(k, l-1)} + (1 - \alpha_{DD}) \max\{\gamma(k, l) - 1, 0\} \quad (5.16)$$

α_{DD} 经验值: 0.95—0.98

有时用 $\lambda_d(k, l)$ 代替



增益因子 G 的确定



MMSE的增益因子:

$$G(k) = \Gamma(1.5) \frac{\sqrt{v(k)}}{\gamma(k)} \exp\left(-\frac{v(k)}{2}\right) \left[\left(1 + v(k)\right) I_0\left(\frac{v(k)}{2}\right) + v(k) I_1\left(\frac{v(k)}{2}\right) \right] \quad (5.17)$$

$$v(k) = \frac{\xi(k)}{1 + \xi(k)} \gamma(k)$$

“Speech Enhancement Using a Minimum Mean-square Error Short-Time Spectral Amplitude Estimator. Ephraim & Malah, 1984”



5.2 噪声抑制 (Noise Suppression, NS)



推荐阅读

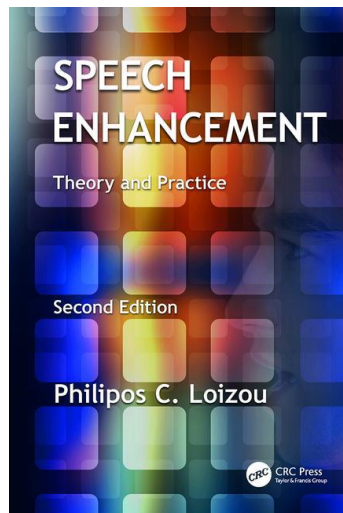


Table of Contents

- Chapter 1 Introduction
- Chapter 2 Discrete-Time Signal Processing and Short-Time [Fourier Analysis](#)
- Chapter 3 Speech Production and Perception
- Chapter 4 Noise Compensation by Human Listeners
- Chapter 5 Spectral-Subtractive Algorithms
- Chapter 6 Wiener Filtering
- Chapter 7 Statistical-Model-Based Methods
- Chapter 8 Subspace Algorithms
- Chapter 9 Noise-Estimation Algorithms
- Chapter 10 Evaluating Performance of Speech Enhancement Algorithms
- Chapter 11 Objective Quality and Intelligibility Measures
- Chapter 12 Comparison of Speech Enhancement Algorithms
- Chapter 13 Algorithms That Can Improve Speech Intelligibility
- Appendix A: Special Functions and Integrals
- Appendix B: Derivation of the MMSE Estimator
- Appendix C: MATLAB ® Code and Speech/Noise Databases

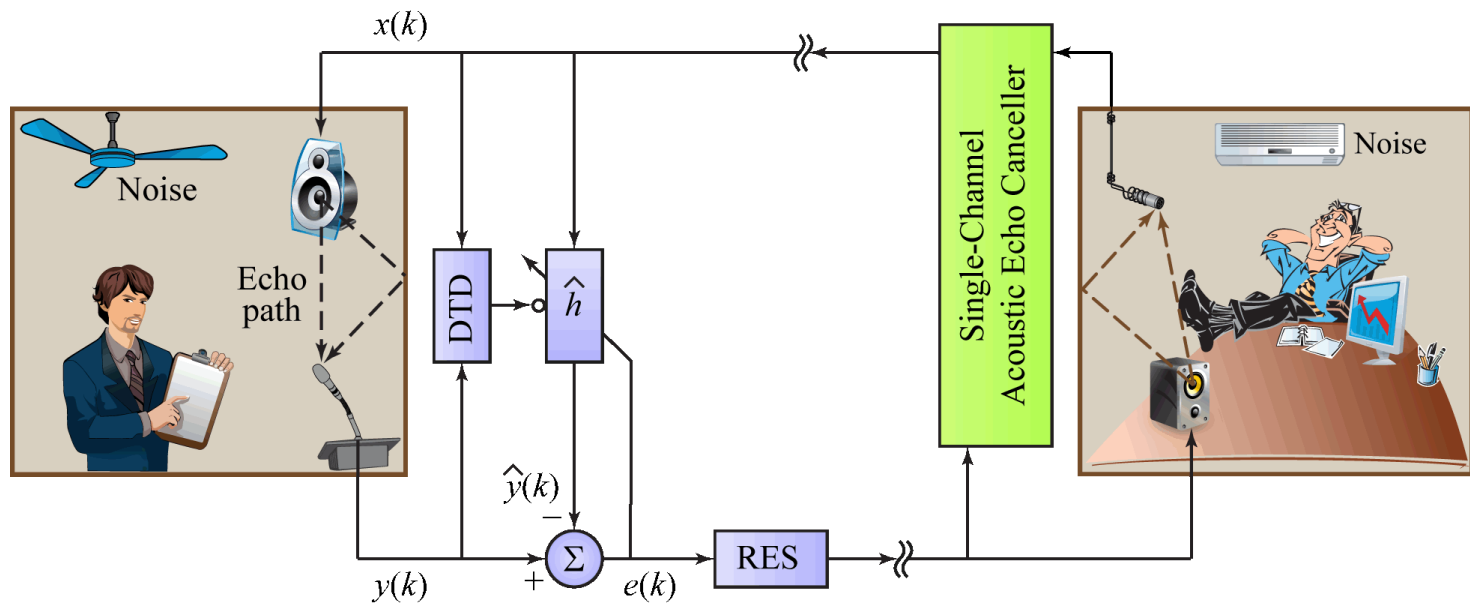




5.3 声学回声消除 (Acoustic Echo Cancellation, AEC)



语音通信

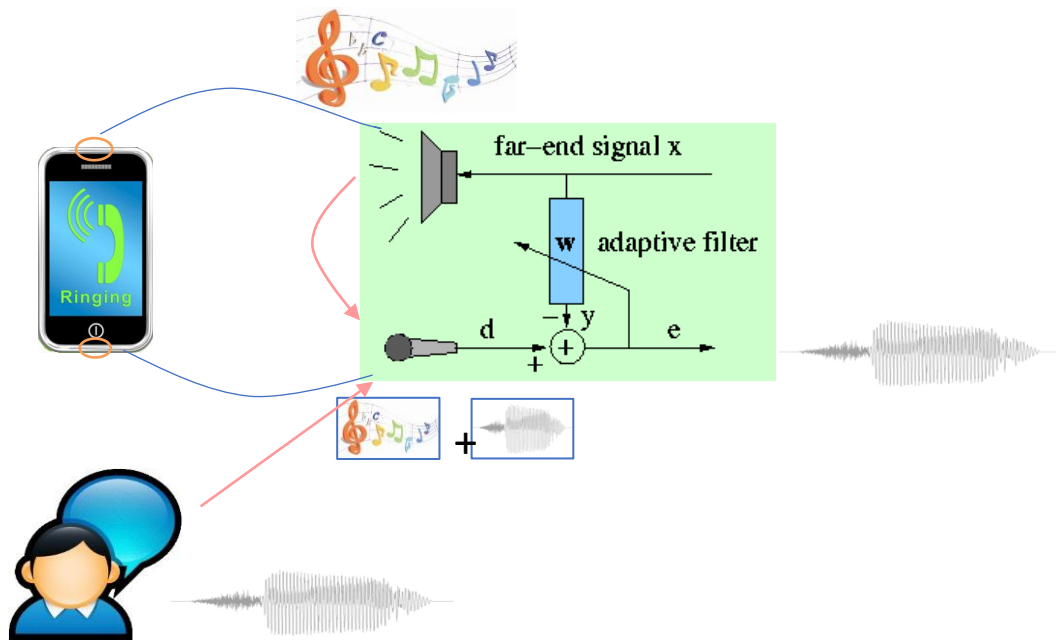




5.3 声学回声消除 (Acoustic Echo Cancellation, AEC)

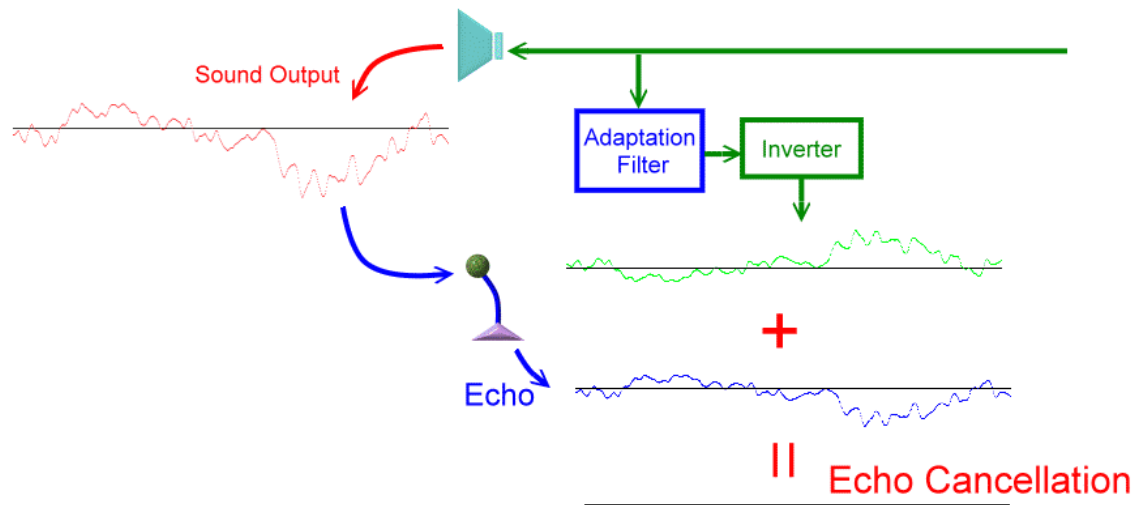


Bargeln



5.3 声学回声消除 (Acoustic Echo Cancellation, AEC)

AEC的基本原理：对消

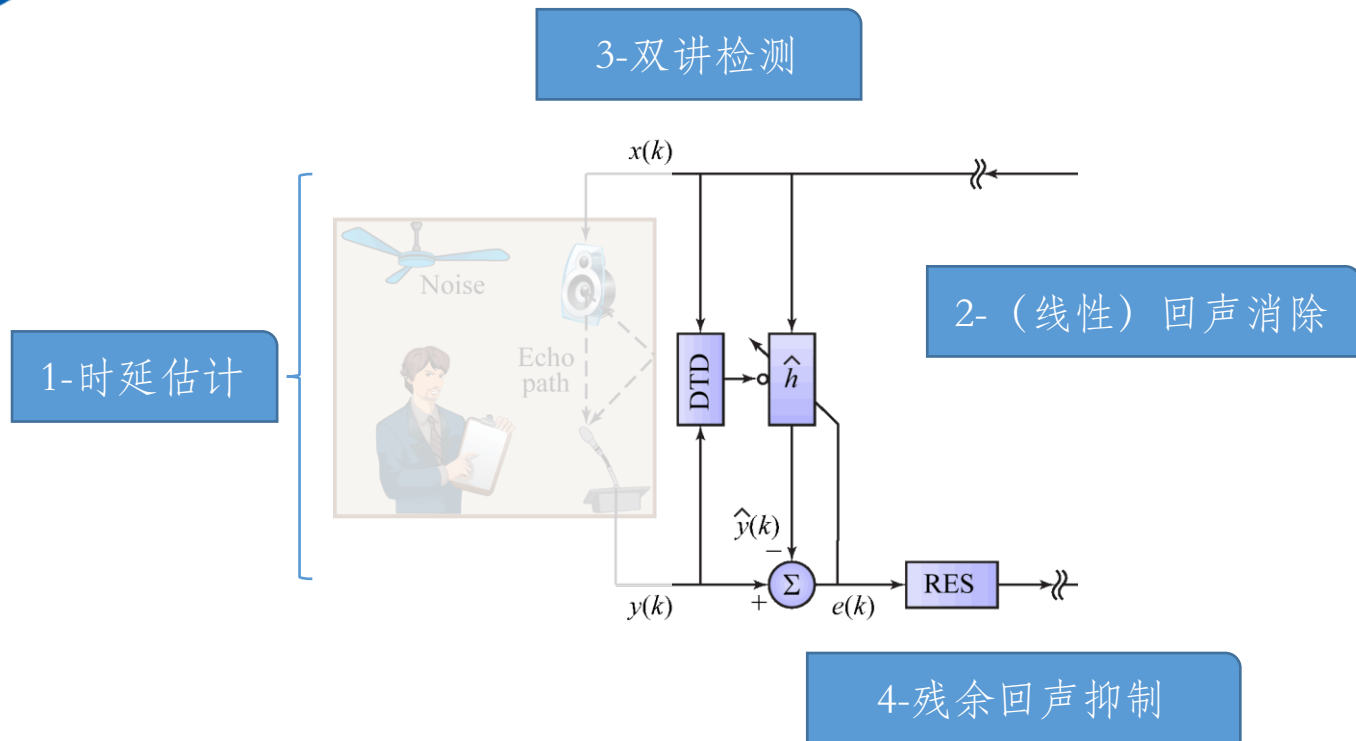




5.3 声学回声消除 (Acoustic Echo Cancellation, AEC)



AEC的基本模块组成:





时延估计 (Time Delay Estimation, TDE)



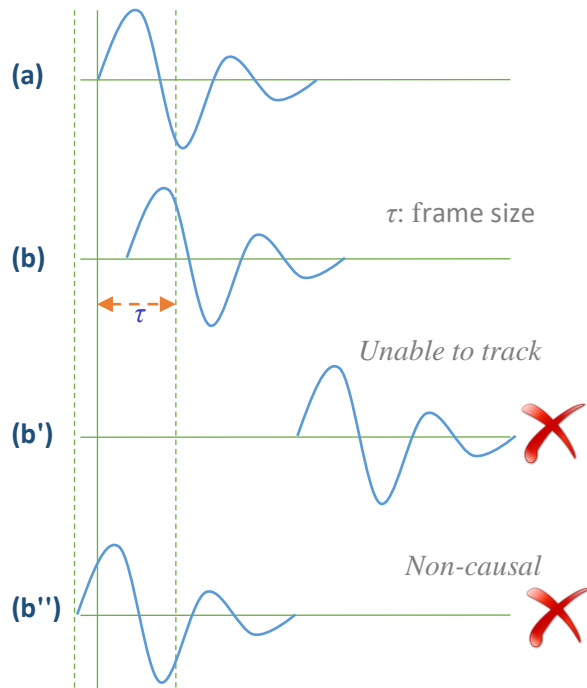
将麦克风信号 $y(n)$ 和参考信号 $x(n)$ 的时延调整到合适的范围。

时延估计的出发点是信号的相关性。

$$r_{yx}(m) = \mathbf{F}^{-1}\{Y(k)X^*(k)\} = \mathbf{F}^{-1}\{P_{yx}(k)\}$$

互功率谱

如果麦克风信号和参考信号的真实时延为 \tilde{m} ，那么理想情况下，互相关函数在 $r_{yx}(\tilde{m})$ 处取最大值。





时延估计 (Time Delay Estimation, TDE)



广义互相关法 (Generalized Cross Correlation)

互相关法：

$$r_{yx}(m) = \mathbf{F}^{-1}\{Y(k)X^*(k)\} = \mathbf{F}^{-1}\{P_{yx}(k)\}$$

广义互相关法：延时信息体现在互功率谱的相位上，而与幅度无关。

$$r_{yx}(m) = \mathbf{F}^{-1}\left\{\frac{P_{yx}(k)}{|P_{yx}(k)|}\right\} \quad (5.18)$$

这种互功率谱加权方式，称为GCC-PHAT (PHase Transformation)，或CSP (Cross Spectral Phase)。

白化



线性回声消除



典型的自适应滤波场景（子带）

FILTER

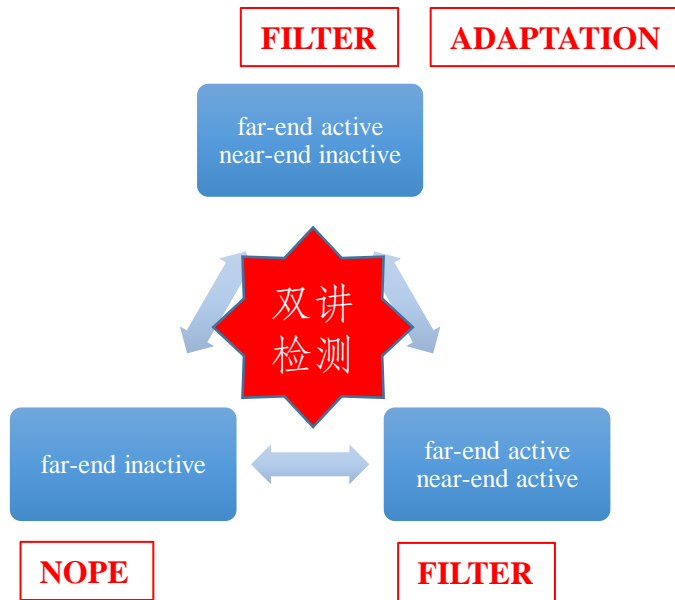
$$y(n) = \mathbf{x}^T(n) \mathbf{w}^*(n)$$

$$e(n) = d(n) - y(n)$$

ADAPTATION

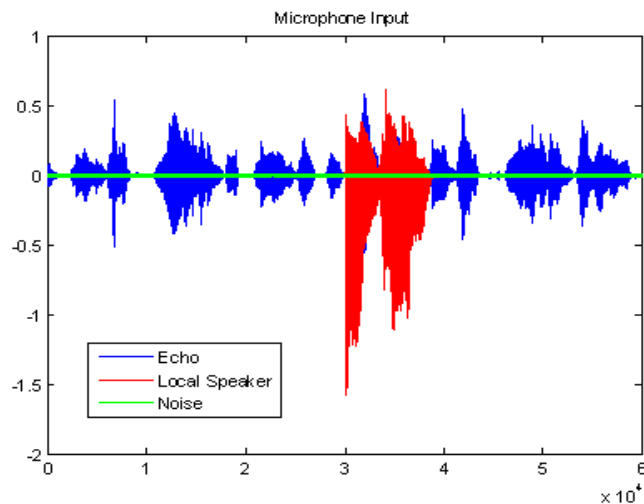
$$\mathbf{w}(n+1) = \mathbf{w}(n) + 2\mu \frac{\mathbf{x}(n)}{\mathbf{x}^H(n)\mathbf{x}(n)} e^*(n)$$

三种工作模式





Double-Talk Detect



1、能量法

$$\frac{|Y(k, l)|}{|X(k, l)|} > \alpha(k) \quad \text{"double talk"}$$

$$\frac{|Y(k, l)|}{|X(k, l)|} \leq \alpha(k) \quad \text{"single talk"}$$

Echo return loss

2、利用残余回声抑制的结果

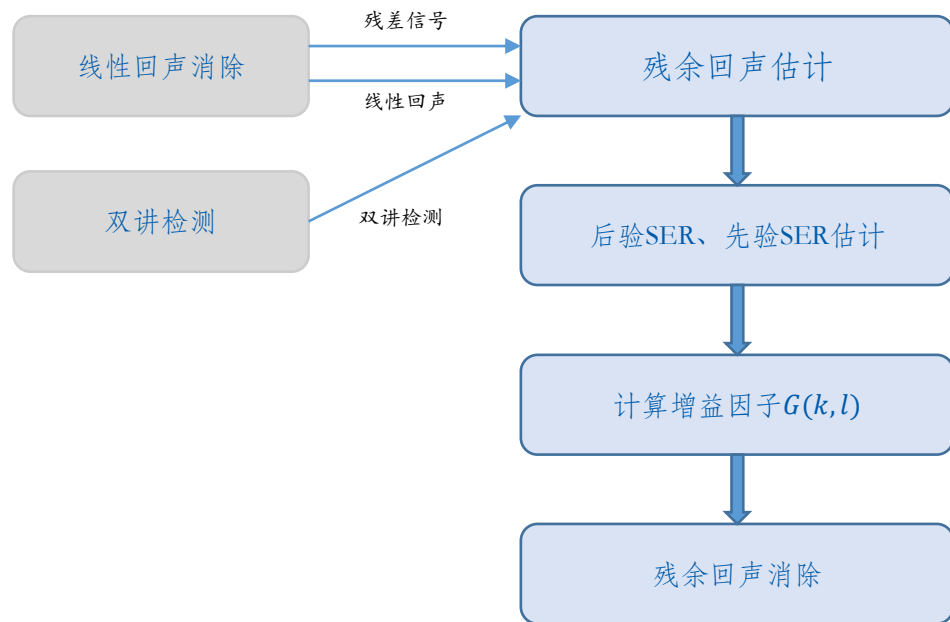


残余回声抑制 (Residual Echo Suppression)



RES的作用之一：抑制残余回声

方法：采用NS的思路。



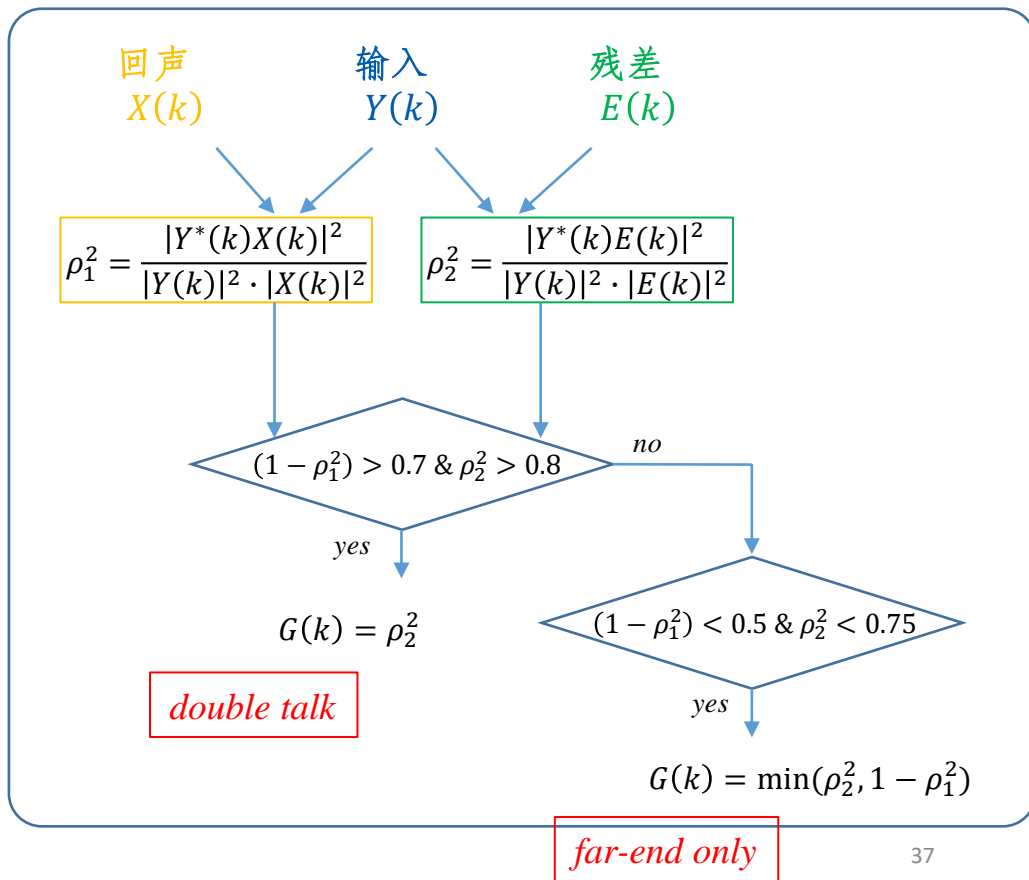


残余回声抑制 (Residual Echo Suppression)



RES的作用之二：修正DTD

利用“相关性”





本章回顾



5.1 子带分解



5.2 噪声抑制



5.3 声学回声消除



5.4 实战



TDE和NLMS算法的实现。

请大家编程实现

时延估计TDE算法；

NLMS算法用于线性回声消除中的滤波器更新。



结语

感谢各位聆听!
Thanks for Listening