

实时语音通信的音频后处理技术

触宝科技
吴威麒

2018年4月21日
陆家嘴软件园

目录



1

概述



2

基本
概念



3

变速控制



4

3A处理



5

总结
展望

概述

- 1 回声消除
- 2 语音去噪
- 3 音量控制
- 4 消除混响
- 5 啸叫抑制
- 6 编解码
- 7 音效
- 8 混音
- 9 vad
- 9 其它

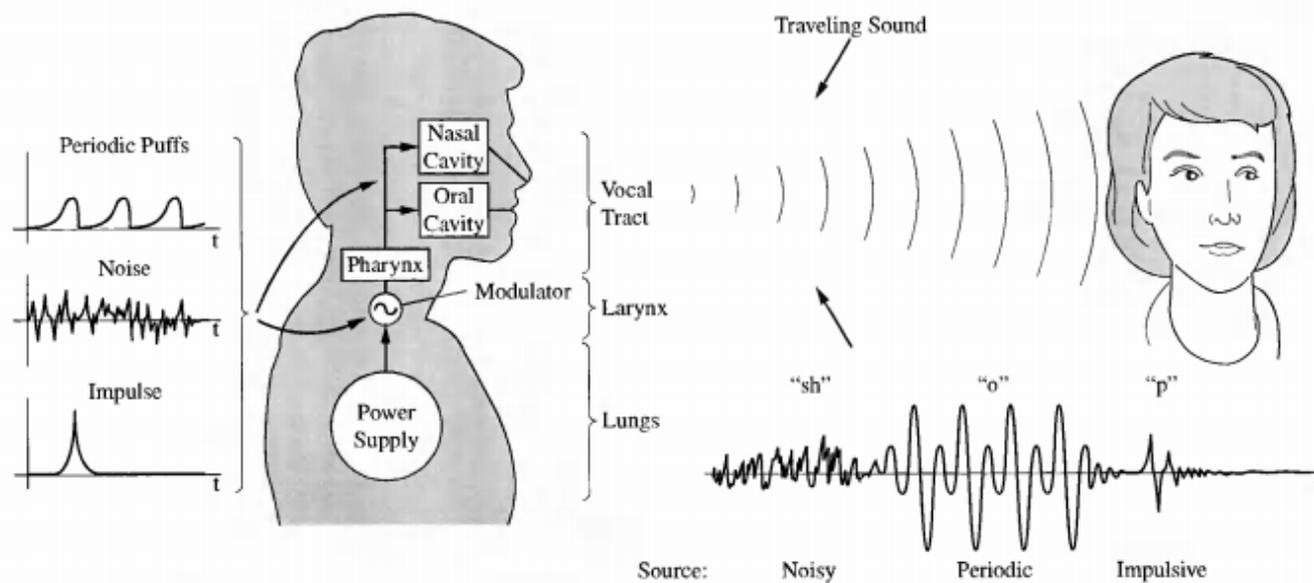
基本概念

语音的产生

时频分析

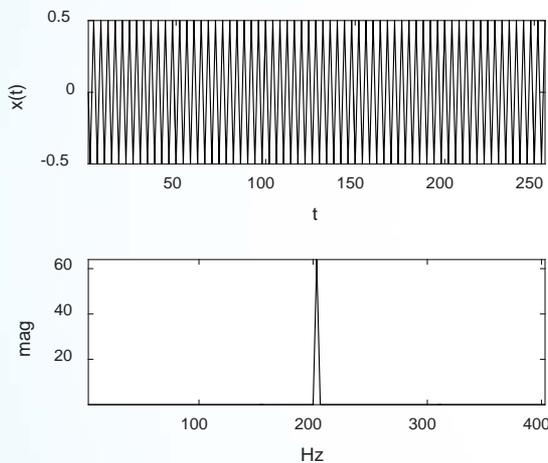
变速变调

语音的产生



时频分析

傅里叶变换的物理意义



$$x(t) = 0.5 \cdot \cos(200t)$$

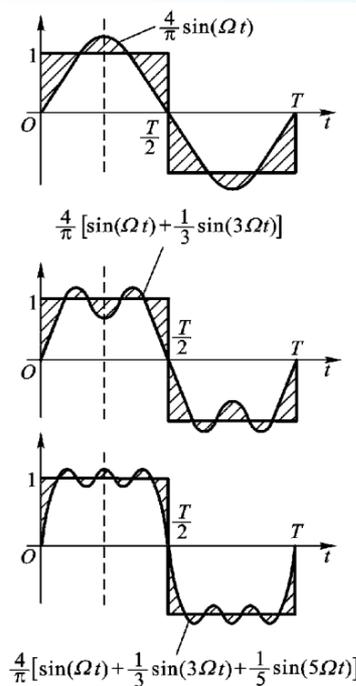
完备的正交函数集

$$f(t) = \frac{a_0}{2} + \sum_{k=1}^n [a_k \cos(k\Omega t) + b_k \sin(k\Omega t)] + \epsilon_n(t)$$

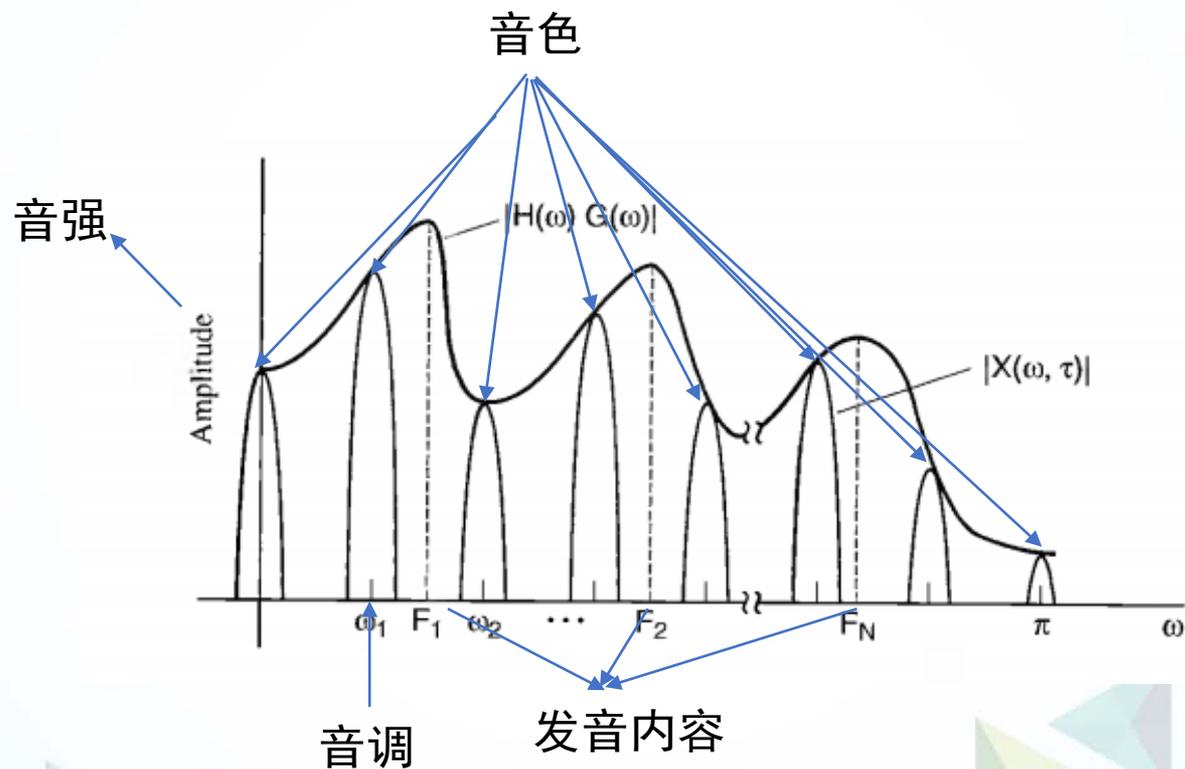
方均误差最小

$$a_n = \frac{2}{T} \int_0^T f(t) \cos(n\Omega t) dt$$

$$b_n = \frac{2}{T} \int_0^T f(t) \sin(n\Omega t) dt :$$



主观认识



变速变调

时域法	参量法
剪贴法	相位声码器
SOLA、SOLA-FS、WSOLA	正弦模型

变速不变调

时域法	参量法
频率调制	相位声码器
变速不变调+重采样	正弦模型

变调不变速

• 应用场景

- 缓冲区包过多
- 抗丢包
- 预防啸叫
- Tomy猫
- 男声<->女声
- 老人<->小孩
- 老鼠吱吱声

音频3A处理

回声消除
语音去噪
音量自动控制

回声消除

基本框架

线性自适应

非线性处理

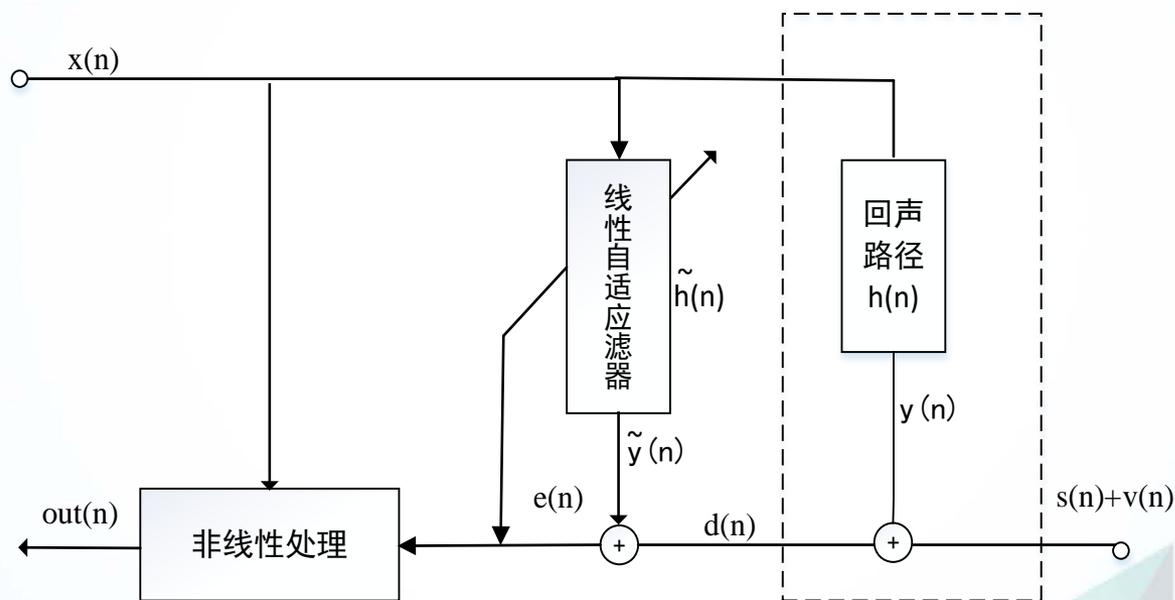
近端干扰

信号对齐

PBFDAF

效果对比

基本框架



线性自适应

$$\mathbf{x}^* \tilde{\mathbf{h}} \rightarrow d(n) = y(n)$$

均方误差意义最小:

$$e(n) = y(n) - \mathbf{x}^* \tilde{\mathbf{h}}$$

$$E[e(n)^2] = 0$$

$$E[y(n)^2] - 2\tilde{\mathbf{h}}^T E(\mathbf{x} * y(n)) + \tilde{\mathbf{h}}^T E(\mathbf{x}\mathbf{x}^T) \tilde{\mathbf{h}} = 0$$

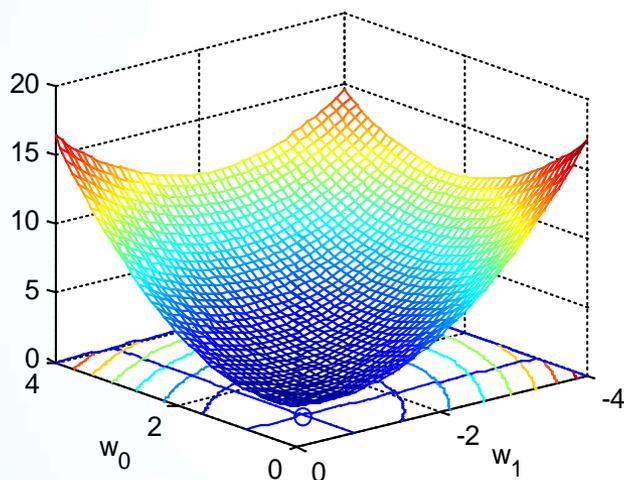
最优滤波器系数

$$\mathbf{h}^* = R_{xx}^{-1} * E(\mathbf{x} * y(n))$$

梯度下降法

$$\mathbf{h}(n+1) = \alpha * \mathbf{h}(n) + 2\mu * \mathbf{x} * e(n)$$

学习率与误差



步长太大，容易发散
步长太小，收敛慢

一般来说
梯度越大，离极值点越远，步长应该较大
梯度越小，离极值点越近，步长应该较小

由于回声路径的稀疏性，很多滤波器权重系数非常小。
对于权重大的系数，步长等比例的大。

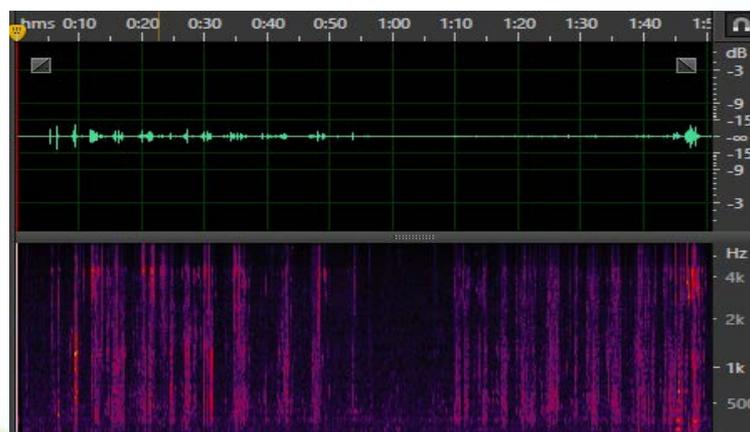
非线性处理

中心削波

相关性抑制



近端语音



线性处理后的残差

近端干扰

能量法

互相关

双滤波器

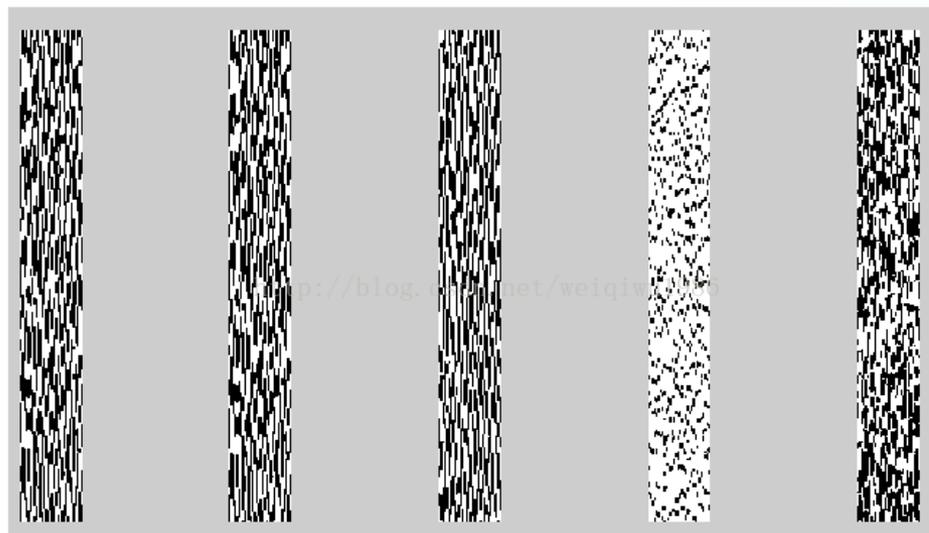
分类识别

信号对齐

互相关

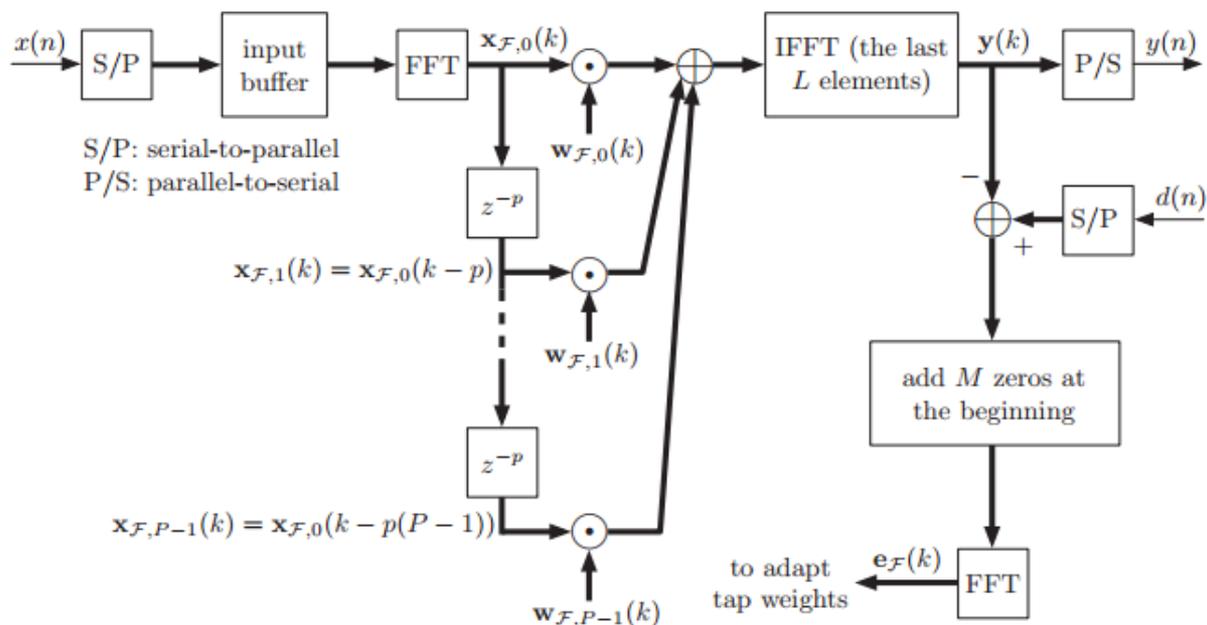
共振峰匹配

二值谱匹配

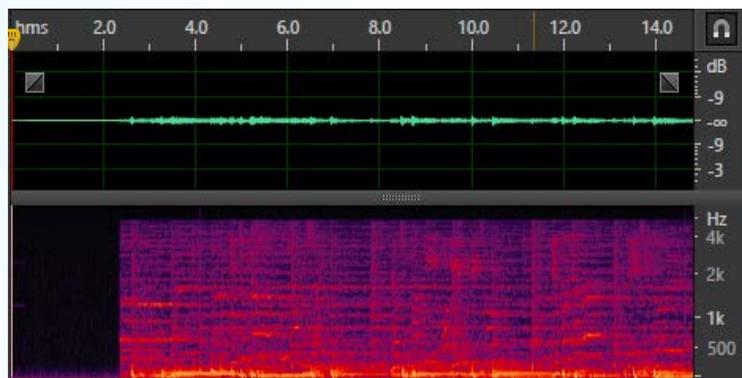


指纹匹配

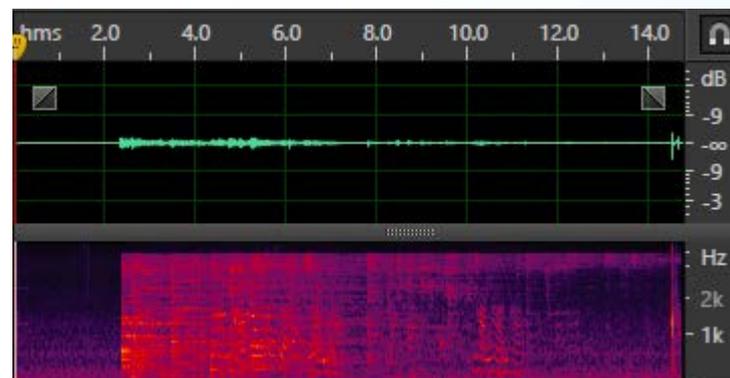
PBFDAF



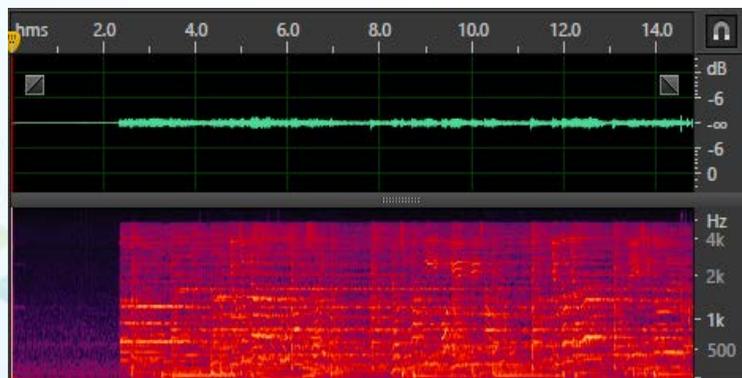
效果对比



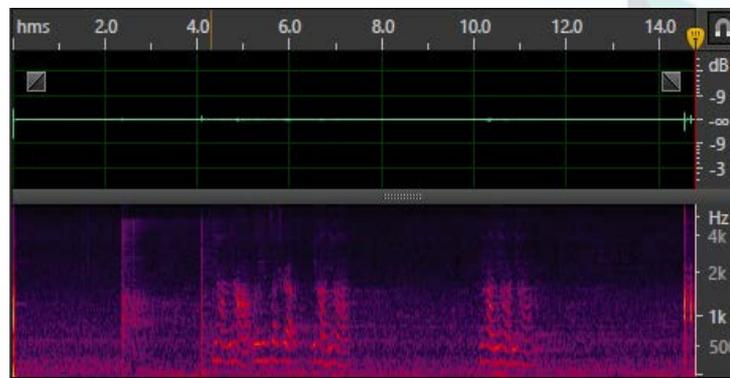
ref



webrtc



mic



ours

去噪

维纳滤波

webrtc的实现方法

语音质量客观评估

维纳滤波

$$H(\omega_k) = \frac{P_{xx}(\omega_k)}{P_{xx}(\omega_k) + P_{nn}(\omega_k)}$$



$$\xi_k \triangleq \frac{P_{xx}(\omega_k)}{P_{nn}(\omega_k)}$$



$$H(\omega_k) = \frac{\xi_k}{\xi_k + 1}$$



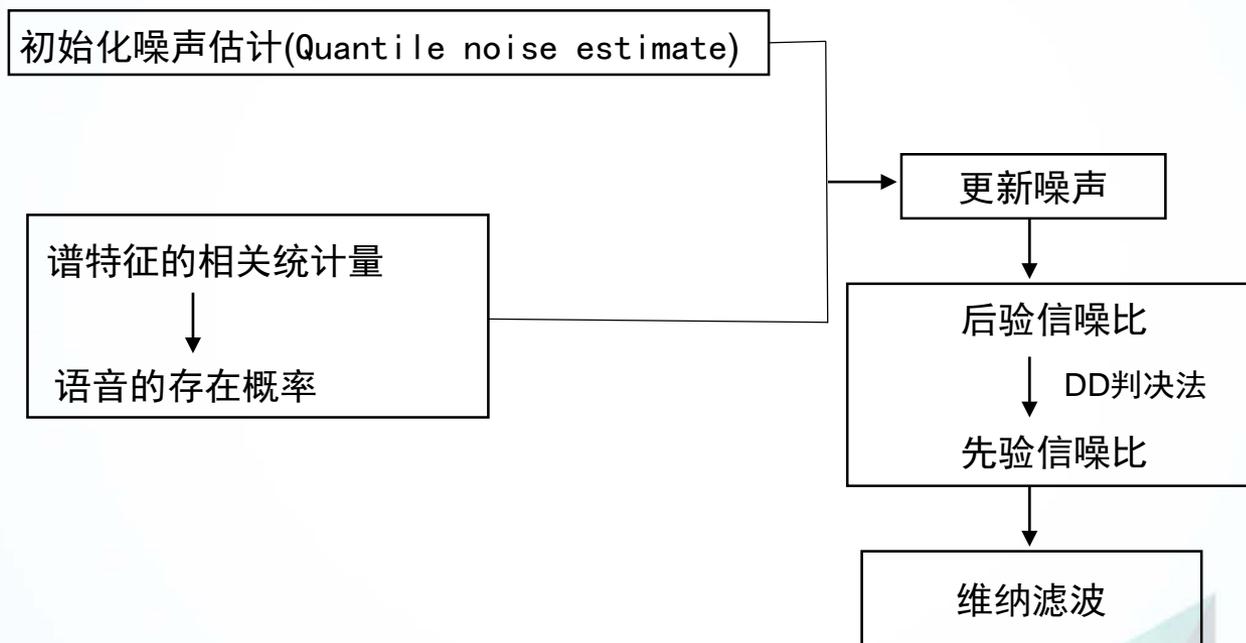
$$\xi_k(m) = E \left\{ \frac{1}{2} \frac{X_k^2(m)}{\lambda_d(k, m)} + \frac{1}{2} [\gamma_k(m) - 1] \right\}$$



噪声估计

相似的：
MMSE
LSA
OM-LSA

webrtc去噪



语音质量评估

语音质量主观评估

MOS	主观意见	听觉感受
4-5	优秀	很好
3-4	良好	清楚
2-3	一般	不太清
1-2	差	勉强
0-1	很差	极差

语音质量客观评估

- 1 SNR
- 2 PESQ
- 3 LLR
- 4 LSD

音量自动控制

三种工作模式：

kAdaptiveAnalog

kAdaptiveDigital

kFixedDigital

总结与展望

AEC

- ◆ 高采样率
- ◆ 音乐信号
- ◆ 双声道立体声AEC
- ◆ 语音识别前端

ANS

- ◆ DNN降噪

AGC

- ◆ 人声检测



IT大咖说
知识共享平台

livevideoStack
— 音视频技术社区 —

Thank You

