

本文是音频处理的朋友icoolmedia (QQ:314138065) 的投稿。对音频处理有兴趣的朋友可以通过下面的方式与他交流：

作者：icoolmedia

QQ：314138065

音视频算法讨论QQ群：374737122

这里假设读者具有自适应滤波器的基础知识。Speex 的 AEC 是以 NLMS 为基础，用 MDF 频域实现，最终推导出最优步长估计：残余回声与误差之比。最优步长等于残余回声方差与误差信号方差之比，这个结论可以记下，下面会用到的。

对于长度为 N 的 NLMS 滤波器，定义如下，这里  $d(n)$  为期望信号、 $\hat{y}(n)$  为估计信号。

$$e(n) = d(n) - \hat{y}(n) = d(n) - \sum_{k=0}^{N-1} \hat{w}_k(n)x(n-k)$$

则，滤波器的系数更新方程为：

$$\begin{aligned} \hat{w}_k(n+1) &= \hat{w}_k(n) + \mu \frac{e(n)x^*(n-k)}{\sum_{i=0}^{N-1} |x(n-i)|^2} \\ &= \hat{w}_k(n) + \mu \frac{(d(n) - \sum_i \hat{w}_i(n)x(n-i))x^*(n-k)}{\sum_{i=0}^{N-1} |x(n-i)|^2} \end{aligned}$$

则滤波器的系数误差： $\delta_k(n) = \hat{w}_k(n) - w_k(n)$ ，且  $d(n) = v(n) + \sum_k \hat{w}_k(n)x(n-k)$

这里  $v(n)$  为近端语音， $\sum_k \hat{w}_k(n)x(n-k)$  为残余回声，于是，滤波器系数更新方程可以重写为：

$$\delta_k(n+1) = \delta_k(n) + \mu \frac{(v(n) - \sum_i \delta_i(n)x(n-i))x^*(n-k)}{\sum_{i=0}^{N-1} |x(n-i)|^2}$$

如果每个时刻的失调定义为： $\Lambda(n) = \sum_k \delta_k^*(n)\delta_k(n)$

那么，在每一步的迭代中，滤波器的失调可表示如下：

$$\Lambda(n+1) = \sum_{k=0}^{N-1} \left| \delta_k(n) + \mu \frac{(v(n) - \sum_i \delta_i(n)x(n-i))x^*(n-k)}{\sum_{i=0}^{N-1} |x(n-i)|^2} \right|^2$$

假设远端信号与近端信号为白噪声，且不相干，则失调的更新方程为

$$E\{\Lambda(n+1) | \Lambda(n), x(n)\} = \Lambda(n) \left[ 1 - \frac{2\mu}{N} + \frac{\mu^2}{N} + \frac{2\mu^2\sigma_v^2}{\Lambda(n)\sum_{i=0}^{N-1} |x(n-i)|^2} \right]$$

其中：

$\sigma_v^2 = E\{|v(n)|^2\}$  为近端语音信号的方差、 $E\{\Lambda(n+1) | \Lambda(n), x(n)\}$  为凸函数，对它关于步长  $\mu$  求导，并导数为 0，可得：

$$\frac{\partial E\{\Lambda(n+1)\}}{\partial \mu} = \frac{-2}{N} + \frac{2\mu}{N} + \frac{2\mu\sigma_v^2}{\Lambda(n)\sum_{i=0}^{N-1} |x(n-i)|^2} = 0$$

最终推出最优步长为：

$$\mu_{opt}(n) = \frac{1}{1 + \frac{\sigma_v^2}{\Lambda(n)/N\sum_{i=0}^{N-1} |x(n-i)|^2}}$$

别看  $\Lambda(n)/N\sum_{i=0}^{N-1} |x(n-i)|^2$  式子挺长，其实意义很明确，可以近似理解为残余回声

$r(n) = y(n) - \hat{y}(n)$  的方差，表示为： $\Lambda(n)/N\sum_{i=0}^{N-1} |x(n-i)|^2 = \sigma_r^2(n)$

于是输出信号的方差为：近端语音的方差+残余回声的方差，用式子表示如下：

$$\sigma_e^2(n) = \sigma_v^2(n) + \sigma_r^2(n)$$

最终，导出最优步长：

$$\mu_{opt}(n) = \frac{1}{1 + \frac{\sigma_v^2}{\sigma_r^2(n)}} = \frac{1}{\frac{\sigma_r^2(n) + \sigma_v^2}{\sigma_r^2(n)}} \approx \frac{\sigma_r^2(n)}{\sigma_e^2(n)}$$
$$\mu_{opt}(n) = \min\left(\frac{\hat{\sigma}_r^2(n)}{\hat{\sigma}_e^2(n)}, 1\right)$$

上面的分析是在时域，基于 NLMS，可以看到：最优步长等于残余回声方差与误差信号方差之比。其中误差的方差比较好求，残余回声的方差比较难求。下面我们看下上面的结论在频域中如何解决，Speex 中在频域的自适应算法为：MDF (multidelay block frequency domain) 自适应滤波。

在频域中，设  $k$  为频率索引， $\ell$  为帧索引，上面的结论转换到频域，结果如下：

$$\mu_{opt}(k, \ell) \approx \frac{\sigma_r^2(k, \ell)}{\sigma_e^2(k, \ell)}$$

那么，在频域如何求残余回声的方差呢，我们可以定义一个泄露系数，表示回声相对于远端信号的泄露程度，这里， $\hat{\eta}(\ell)$  为远端信号的泄露系数。

$$\sigma_r^2(k, \ell) = \hat{\eta}(\ell) \hat{\sigma}_x^2(k, \ell)$$

因此，最优步长可以表示为：

$$\mu_{opt}(n) = \min\left(\hat{\eta}(\ell) \frac{|\hat{Y}(k, \ell)|^2}{|E(k, \ell)|^2}, \mu_{max}\right)$$

也就是说，根据泄露系数，可以估计出远端信号的残余回声，进而可以得到最优步长，那么，带来另一个问题，这里的泄露系数如何估计呢？如果读者熟悉下正交原理相关的内容的话，这个问题很容易解决，下面给出答案：

$$\hat{\eta}(\ell) = \frac{\sum_k R_{EY}(k, \ell)}{\sum_k R_{YY}(k, \ell)}$$
$$R_{EY}(k, \ell) = (1 - \beta(\ell)) R_{EY}(k, \ell) + \beta(\ell) P_Y(k) P_E(k)$$
$$R_{YY}(k, \ell) = (1 - \beta(\ell)) R_{YY}(k, \ell) + \beta(\ell) P_Y(k) (P_Y(k))^2$$
$$\beta(\ell) = \beta_0 \min\left(\frac{\hat{\sigma}_Y^2(\ell)}{\hat{\sigma}_e^2(\ell)}, 1\right)$$

这里，是通过递归平均处理方法得到每个频点的自相关、输入信号与误差信号的互相关。最终得到泄露系数，具体实现可以参考 speex 的代码实现，相关参数可以参考后面给出来参考论文。

Speex 的回声消除原理已经分析完了，最终得出结论是：只有改与泄露系数相关部分的代码，才是对效果影响最大的地方，因为根据泄露系数，最终会估计出滤波器的最优步长。

本文来自于 On Adjusting the Learning Rate in Frequency Domain Echo Cancellation With Double-Talk 这篇论文，由 icoolmedia 用自己的语言做了些整理，对算法感兴趣的朋友可以加入音视频算法讨论群 (374737122)，大家共同讨论！

文章标签： speex 回声消除

个人分类： 音频编码

此PDF由spyyg生成,请尊重原作者版权!!!

我的邮箱:liushidc@163.com