```
% omlsa : Single Channel OM-LSA with IMCRA noise estimator
% Copyright (c) 2003. Prof Israel Cohen.
```

Note on algorithm implementation, list all formulas and parameters.

More details on the algorithm, please refer to the papers by Cohen.

- **state initialization at the first frame for all frequency bins**

$$\lambda_d(:,0) = |Y(:,1)|^2$$
$$G_{H1}^2(:,0) * \gamma(:,0) = 1$$
$$S(:,0) = S_f(:,1)$$
$$\tilde{S}(:,0) = S_f(:,1)$$

- **1，update with one subframe time domain data, STFT, power spectrum**

$$y0 = fin(one\ subframe)$$
$$y = \big[y\big((Mno+1){:}M\big); y0\big]$$
$$Y = fft(win.* y)$$
$$Y_{pow} = |Y(1{:}N)|^2$$

parameter: $Fs = 16000Hz$, $M = 512$, $Mno = 128$, $N = \frac{M}{2} + 1 = 257$

- **2，A priori SNR estimation $\boldsymbol{\xi(k,l)}$**

The following formulas are numbered as same as the paper of **Year2003** "*Noise Spectrum Estimation in Adverse Environments: Improved Minima Controlled Recursive Averaging*".

$$\gamma(k,l) = \frac{|Y(k,l)|^2}{\lambda_d(k,l-1)} \qquad (\mathbf{3})$$

$$\xi(k,l) = \alpha_\xi * G_{H1}^2(k,l-1) * \gamma(k,l-1) + \big(1-\alpha_\xi\big) * \max(\gamma(k,l)-1,0) \qquad (\mathbf{32})$$

$$\xi(k,l) = \min\{\xi(k,l),\ \xi_{min}\}$$

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

$\xi(k,l)$：the a priori SNR

$\gamma(k,l)$：the a posteriori SNR under speech presence uncertainty

k：frequency bin (subband) index

l：time frame index

parameter: $\alpha_\xi = 0.95$, $G\_dB_{min} = -18$, $\xi_{min} = 10^{\frac{G\_dB_{min}}{10}}$

- **3，Noise spectrum estimation with IMCRA**

(1) Compute the first iteration of smoothed power spectrum $S(k,l)$ and update its minimum $S_{min}(k,l)$

$$S_f(k,l) = \sum_{i=-w}^{w} b(i) * |Y(k-i,l)|^2 \qquad (\mathbf{14})$$

$$S(k,l) = \alpha_s * S(k,l-1) + (1-\alpha_s) * S_f(k,l) \qquad (\mathbf{15})$$

$$S_{min}(k,l) = \min\{S_{min}(k,l-1), S(k,l)\}$$

$$S_{mact}(k,l) = \min\{S_{mact}(k,l-1), S(k,l)\}$$

$$\gamma_{min}(k,l) = \frac{|Y(k,l)|^2}{B_{min} S_{min}(k,l)} \qquad (\mathbf{18})$$

$$\zeta(k,l) = \frac{S(k,l)}{B_{min} S_{min}(k,l)} \qquad (\mathbf{18})$$

S : local energy of the noisy signal

S_f : frequency average of the noisy signal's energy

S_{min} : local minimum of S

parameter: $w = 1$, $\alpha_s = 0.9$, $B_{min} = 1.66$

(2) Computer the indicator for the voice activity detection $I(k,l)$

$$I(k,l) = \begin{cases} 1, & \gamma_{min}(k,l) < \gamma_0 \ \ and\ \zeta(k,l) < \zeta_0 \ (speech\ absent) \\ 0, & otherwise \end{cases} \qquad (\mathbf{21})$$

parameter: $\gamma_0 = 4.6$, $\zeta_0 = 1.67$

(3) Compute the second iteration of smoothed power spectrum $\tilde{S}(k,l)$ and update its minimum $\tilde{S}_{min}(k,l)$

$$I_f(k,l) = \sum_{i=-w}^{w} b(i) * I(k-i,l)$$

$$\widetilde{S_f}(k,l) = \begin{cases} \dfrac{\sum_{i=-w}^{w} b(i) * I(k-i,l) * |Y(k-i,l)|^2}{I_f(k,l)}, & I_f(k,l) \neq 0 \\ \tilde{S}(k,l-1), & I_f(k,l) = 0 \end{cases} \qquad (\mathbf{26})$$

$$\tilde{S}(k,l) = \alpha_s * \tilde{S}(k,l-1) + (1-\alpha_s) * \widetilde{S_f}(k,l) \qquad (\mathbf{27})$$

$$\tilde{S}_{min}(k,l) = \min\{\tilde{S}_{min}(k,l-1), \tilde{S}(k,l)\}$$

$$\tilde{S}_{mact}(k,l) = \min\{\tilde{S}_{mact}(k,l-1), \tilde{S}(k,l)\}$$

$$\tilde{\gamma}_{min}(k,l) = \frac{|Y(k,l)|^2}{B_{min}\tilde{S}_{min}(k,l)} \qquad \mathbf{(28)}$$

$$\tilde{\zeta}(k,l) = \frac{S(k,l)}{B_{min}\tilde{S}_{min}(k,l)} \qquad \mathbf{(28)}$$

<u>parameter:</u> $\alpha_s = 0.9$, $B_{min} = 1.66$

(4) Computer the a priori speech absence probability $\tilde{q}(k,l)$ and the speech presence probability $\hat{p}(k,l)$ **for noise estimation**

$$\hat{q}(k,l) = \begin{cases} 1, & \tilde{\gamma}_{min}(k,l) \leq 1 \quad and\ \tilde{\zeta}(k,l) < \zeta_0 \\ \dfrac{\gamma_1 - \tilde{\gamma}_{min}(k,l)}{\gamma_1 - 1}, & 1 < \tilde{\gamma}_{min}(k,l) < \gamma_1 \quad and\ \tilde{\zeta}(k,l) < \zeta_0 \\ 0, & \tilde{\gamma}_{min}(k,l) \geq \gamma_1 \quad and\ \tilde{\zeta}(k,l) \geq \zeta_0 \end{cases} \quad \mathbf{(29)}$$

$$\hat{p}(k,l) = \begin{cases} 0, & \tilde{\gamma}_{min}(k,l) \leq 1 \quad and\ \tilde{\zeta}(k,l) < \zeta_0 \\ \widehat{p_0}(k,l), & 1 < \tilde{\gamma}_{min}(k,l) < \gamma_1 \quad and\ \tilde{\zeta}(k,l) < \zeta_0 \\ 1, & \tilde{\gamma}_{min}(k,l) \geq \gamma_1 \quad and\ \tilde{\zeta}(k,l) \geq \zeta_0 \end{cases} \qquad \mathbf{(7)}$$

$$\widehat{p_0}(k,l) = \frac{1}{1 + \dfrac{\hat{q}(k,l)}{1-\hat{q}(k,l)} * \big(1+\xi(k,l)\big) * \exp\big(-v(k,l)\big)}$$

<u>parameter:</u> $\gamma_1 = 3$, $\zeta_0 = 1.67$

(5) Update the noise spectrum estimate $\lambda_d(k,l)$

$$\widetilde{\alpha_d}(k,l) = \alpha_d + (1-\alpha_d) * \hat{p}(k,l) \qquad \mathbf{(11)}$$

$$S_f(k,l) = \sum_{i=-w}^{w} b(i) * |Y(k-i,l)|^2 \qquad \mathbf{(14)}$$

$$\bar{\lambda}_d(k,l) = \widetilde{\alpha_d}(k,l) * \bar{\lambda}_d(k,l-1) + \big(1-\widetilde{\alpha_d}(k,l)\big) * |Y(k,l)|^2 \qquad \mathbf{(10)}$$

$$\lambda_d(k,l) = \beta * \bar{\lambda}_d(k,l) \qquad \mathbf{(12)}$$

<u>parameter:</u> $\alpha_d = 0.85$, $\beta = 1.4685$

(6) The minimum tracking of $S_{min}(k,l)$ and $\tilde{S}_{min}(k,l)$

Description from paper: we divides the window of D samples into N_{win} sub-windows of V samples($D = N_{win}V$). Whenever V samples are read, the minimum of the current subwindow is determined and stored for later use. The overall minimum is obtained as the minimum of past samples within the current subwindow and the N_{win} previous subwindow minima.

V frames is one sub-window. every V frames, we update $S_{min}(k,l)$ and $\tilde{S}_{min}(k,l)$:

(1) use the latest sub-window's minimum $S_{mact}, \tilde{S}_{mact}$ to update to $S_W, \tilde{S}_W$ buffer of length N_{win};

(2) use the minimum of buffer $S_W, \tilde{S}_W$, and update to $S_{min}(k,l)$ and $\tilde{S}_{min}(k,l)$;

(3) use the latest frame's local energy $S, \tilde{S}$ to reset next sub-window's minimum's initial value $S_{mact}, \tilde{S}_{mact}$.

```
if l_mod_lswitch==V
    l_mod_lswitch=0;
    SW=[SW(:,2:Nwin) SMact];
    Smin=min(SW,[],2);
    SMact=S;
    SWt=[SWt(:,2:Nwin) SMactt];
    Smint=min(SWt,[],2);
    SMactt=St;
```

parameter: $V = 15$, $N_{win} = 8$

- **4，A priori speech absence probability estimation $\boldsymbol{q(k,l)}$**

The following formulas are numbered as same as the paper of **Year2001** "*Speech enhancement for non-stationary noise environments*".

$$\zeta(k,l) = \alpha_\zeta * \zeta(k,l-1) + (1-\alpha_\zeta) * \xi(k,l) \qquad \mathbf{(23)}$$

$$\zeta_{local}(k,l) = \sum_{i=-w_{local}}^{w_{local}} b_{local}(i) * \zeta(k-i,l) \qquad \mathbf{(24)}$$

$$\zeta_{global}(k,l) = \sum_{i=-w_{global}}^{w_{global}} b_{global}(i) * \zeta(k-i,l) \qquad \mathbf{(24)}$$

$$P_{local}(k,l) = \begin{cases} \zeta_{min}, & \zeta_{local}(k,l) \le \zeta_{local\_min} \\ \log\left(\dfrac{\zeta_{local}(k,l)/\zeta_{local\_min}}{\zeta_{local\_max}/\zeta_{local\_min}}\right), & \zeta_{local\_min} < \zeta_{local}(k,l) < \zeta_{local\_max} \\ 1, & \zeta_{local}(k,l) \ge \zeta_{local\_max} \end{cases} \qquad \mathbf{(25)}$$

$$P_{global}(k,l) = \begin{cases} \zeta_{min}, & \zeta_{global}(k,l) \le \zeta_{global\_min} \\ \log\left(\frac{\zeta_{global}(k,l)/\zeta_{global\_min}}{\zeta_{global\_max}/\zeta_{global\_min}}\right), & \zeta_{global\_min} < \zeta_{global}(k,l) < \zeta_{global\_max} \\ 1, & \zeta_{global}(k,l) \ge \zeta_{global\_max} \end{cases} \quad \mathbf{(25)}$$

$$\zeta_{frame}(l) = mean(\zeta(k,l)) \qquad \mathbf{(26)}$$

To prevent clipping of speech startings or weak components, speech is assumed whenever $\zeta_{frame}()$ increases.

$$\zeta_{peak}(l) = \max\left(\zeta_{peak}(l), \quad \zeta_{frame}(l)\right)$$

$$P_{frame}(l) = \begin{cases} \zeta_{min}, & \zeta_{frame}(l) \le \zeta_{peak}(l)\zeta_{frame\_min} \\ \log\left(\frac{\zeta_{frame}(l)/\zeta_{peak}(l)/\zeta_{frame\_min}}{\zeta_{frame\_max}/\zeta_{frame\_min}}\right), & otherwise \\ 1, & \zeta_{frame}(l) \ge \zeta_{peak}(l)\zeta_{frame\_max} \end{cases} \quad \mathbf{(25)}$$

$$q(k,l) = 1 - P_{local}(k,l) * P_{global}(k,l) * P_{frame}(l) \qquad \mathbf{(28)}$$
$$q(k,l) = \max(q(k,l), \quad q_{max})$$
$$q_{max} < 1$$

$\zeta_{local}(k,l)$, $\zeta_{global}(k,l)$: local and global averages of the a priori SNR

$\zeta_{frame}(l)$: frame average of the a priori SNR

$P_{local}(k,l)$, $P_{global}(k,l)$: local and global likelihood of speech

$P_{frame}(k,l)$: frame likelihood of speech

parameter: $\alpha_\zeta = 0.7$, $q_{max} = 0.998$

$w_{local} = 1$, $\zeta_{local\_\min(\mathrm{dB})} = -10$, $\zeta_{local\_\max(dB)} = -5$

$w_{global} = 15$, $\zeta_{global\_\min(\mathrm{dB})} = -10$, $\zeta_{global\_\max(dB)} = -5$

$\zeta_{frame\_\min(\mathrm{dB})} = -10$, $\zeta_{frame\_\max(dB)} = -5$

- **5, Calculate the spectral gain with new noise estimation $\lambda_d(k,l)$ and $q(k,l)$ for the OM-LSA**

$$\gamma(k,l) = \frac{|Y(k,l)|^2}{\max(\lambda_d(k,l), 1e-10)} \qquad \mathbf{(3)}$$

$$\xi(k,l) = \alpha_\xi * G_{H1}^2(k,l-1) * \gamma(k,l-1) + \left(1-\alpha_\xi\right) * \max(\gamma(k,l)-1,0) \qquad \mathbf{(32)}$$

$$\xi(k,l) = \min\{\xi(k,l),\ \xi_{min}\}$$

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

$$p(k,l) = \frac{1}{1+\frac{q(k,l)}{1-q(k,l)} * \left(1+\xi(k,l)\right) * \exp\left(-v(k,l)\right)} \qquad \mathbf{(9)}$$

$$if\ q(k,l) \geq 0.9, \qquad p(k,l) = 0$$

$$G_{H1}(k,l) = \frac{\xi(k,l)}{1+\xi(k,l)} \exp\left(\frac{1}{2}\int_{v(k,l)}^{\infty} \frac{e^{-t}}{t} dt\right) \qquad \mathbf{(33)}$$

$$G_{H0}(k,l) = 10^{\frac{G\_dB_{min}}{20}}$$

$$G(k,l) = G_{H1}(k,l)^{p(k,l)} * G_{H0}(k,l)^{1-p(k,l)} \qquad \mathbf{(16)}$$

$p(k,l)$：speech presence probability **for speech estimation**

$G(k,l)$：spectral gain function

$G_{H1}(k,l)$： conditional gain function

$G_{H0}(k,l)$： spectral gain floor

parameter: $\alpha_\xi = 0.95$， $G\_dB_{min} = -18$

- **6，reverse STFT, overlap add to time domain**

$$X = G * Y$$