Single Channel Speech Enhancement Based on Masking Properties of the Human Auditory System

an Article by Nathalie Virag[1]

Presents: Yoel Paciuk
EE048831- Spring 2003
Topics Covered

• Auditory Masking & its proposed usage
• Calculation of the Masking Threshold
• Generalized Spectral Subtraction
• Block Diagram & Parameter Adaptation
• Performance
• A different Approach based on Ephraim-Malah STSA estimator
Auditory Masking Phenomenon

- The auditory system is characterized by a minimum threshold of audibility.
- In the presence of a signal this threshold is elevated in its vicinity of time & frequency.
- Signals masked by this elevated threshold will not be audible, and thus doesn’t have to be removed or reduced any further.
Calculation of the Masking Threshold
(Based on [3])

- **Critical Band Analysis**
The spectrum is partitioned into critical bands, and spectral energy is summed for each band. There are 18 bands for a 4Khz signal.

\[
B_i = \sum_{\text{band low frequency}}^{\text{band high frequency}} P(w)
\]

- **Spreading Operation**
Convolve with a spreading function to account for masking effects across critical bands. Convolution is performed with a matrix multiplication.

\[
C_{\text{spread}} = S \ast B
\]
Calculation of the Masking Threshold

• Tonality Measure
Masking is different for noise like or tone like masking signal

\[
SF_{M_{dB}} = 10 \log_{10} \left( \frac{G_m}{A_m} \right)
\]

\[
\alpha = \min \left( \frac{SF_{M_{dB}}}{SF_{M_{dB_{max}}}} , 1 \right)
\]

\[
G_m – \text{geometric mean of spectrum}
\]

\[
A_m – \text{arithmetic mean of spectrum}
\]

• Threshold Calculation
Tone masking threshold in band i: \( C_i-(14.5+i) \) dB

\[
O_i = \alpha (14.5 + i) + (1 - \alpha)5.5
\]

\[
T_i = 10^{\log_{10}(C_i) - (O_i/10)}
\]

\[
O_i – \text{masking offset (dB)}
\]

\[
T_i – \text{masking threshold}
\]
Calculation of the Masking Threshold

• **Renormalization and comparison with absolute audibility thresholds**

  Instead of De-convolution of the spreading function energy normalization is performed
  Received threshold is checked against absolute threshold and overridden if smaller

• **The Output of this process is T(w) – the masking threshold per frequency bin**
Help Slide

• Show figure of Masking calculation from Ref [3]
Generalized Spectral Subtraction

- Advantages:
  Easy to implement
  Reduces Noise Considerably (dependent on Quality of noise estimation.)

- Disadvantages:
  Speech Distortion
  Musicality Effects
Generalized Spectral Subtraction

\[ y(n) = x(n) + d(n) \]

\[ X_{est}(w) = G(w) |Y(w)| \exp(j \arg(Y(w))) \]

\[ SNR_p = \frac{|Y(w)|^2}{|D_{est}(w)|^2} \]

\[ G(w) = \begin{cases} 
(1 - \frac{\alpha}{\gamma_1^2}) \frac{\gamma^2}{\gamma_1^2} & \text{if } SNR_p > (\alpha + \beta)^2 \\
SNR_p^2 & \text{else} \\
(\frac{\beta}{\gamma_1^2}) \frac{\gamma^2}{\gamma_1^2} & \end{cases} \]
Generalized Spectral Subtraction

\( \alpha \) – determines over/under subtraction
\( \beta \) – determines spectral floor
\( \gamma_1, \gamma_2 \) – method of subtraction
\( \gamma_1, \gamma_2 = (2,1) \) – wiener
\( \gamma_1, \gamma_2 = (1,1) \) – magnitude subtraction
\( \gamma_1, \gamma_2 = (2,0.5) \) – spectral subtraction
Speech Enhancement Reminder

**SPECTRAL SUBTRACTION** $\gamma_1 = 1; \gamma_2 = 0.5$

$$G(w) = \left(1 - \frac{1}{SNR_p^1}\right)^{\frac{1}{2}} \text{ if } SNR_p > 1$$

**WEINER FILTER** $\gamma_1 = 2; \gamma_2 = 1$

$$\hat{x} = h(n) * x(n); \text{ minimize } E\{(x - \hat{x})^2\}$$

$$G(w) = \left(1 - \frac{1}{SNR_p^1}\right)^1$$
The General Idea

• Use $T(w)$ to determine & adapt alfa, beta such as to reduce noise up to the masking threshold and not more than that

• Used Parameters in this work

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alfa</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>Beta</td>
<td>0</td>
<td>0.02</td>
</tr>
<tr>
<td>Gamma</td>
<td>G1=2</td>
<td>G2=0.5</td>
</tr>
</tbody>
</table>
Block Diagram

* Reproduced from [2]
Parameter Adaptation

\[ \alpha(i) = (1 - \frac{T(i) - T(i)_{\text{min}}}{T(i)_{\text{max}} - T(i)_{\text{min}}})^x \alpha_{\text{max}} - \alpha_{\text{min}} + \alpha_{\text{max}} \]

\[ \alpha(i) = \alpha_{\text{min}} \text{ for } T(i) = T(i)_{\text{max}} \]

\[ \alpha(i) = \alpha_{\text{max}} \text{ for } T(i) = T(i)_{\text{min}} \]

\[ \beta \text{ is updated in a similar manner} \]

perform exponential averaging on frames
## Comparison & Performance

<table>
<thead>
<tr>
<th>Method</th>
<th>Alfa</th>
<th>Beta</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spectral subtraction</td>
<td>Fixed at 1</td>
<td>Fixed at 0</td>
<td>Reference</td>
</tr>
<tr>
<td>Nonlinear Subtraction</td>
<td>Function of SNR</td>
<td>Fixed at 0.01</td>
<td>Better than above</td>
</tr>
<tr>
<td>Proposed method</td>
<td>Function of T(w)</td>
<td>Function of T(w)</td>
<td>Better than above</td>
</tr>
</tbody>
</table>
Claimed Advantages

• The estimation of $T(w)$ is more accurate than the SNR estimation because it is estimated from the subtracted signal
• $T(w)$ has a smoother evolution than the SNR
• The adaptation based on $T(w)$ is better correlated with perception
Audio Samples [2]

- Wiener filter
- Suggested algorithm

Issues to be further explored

- Is the setting of minimum alfa to 1 not contrary to the general idea of under subtraction when possible?
- Compare against more sophisticated methods:
  - Ephraim&Malah optimum STSA estimator
  - McAulay&Malpass ML STSA estimator
A Different Approach

• P.J. Wolfe and S.J. Godsill

• Based on Ephraim-Malah STSA estimator
Ephraim-Malah STSA

\[ Y_k = X_k + D_k \]
\[ X_k = A_k \cdot \exp(j \cdot \alpha_k) \]
\[ X_k, D_k \text{ independent gaussian R.V} \]
\[ \hat{A}_k \text{ -- spectral amplitude estimator} \]

\[ C(\hat{A}_k, A_k) = (\hat{A}_k - A_k)^2 \]
\[ \hat{A}_k = \min \int \int C(\hat{A}_k, A_k) \cdot p(A_k, \alpha_k / Y_k) \, dA_k \, d\alpha_k \]
Modified Cost Function

\[ M_k \text{ – masking threshold in bin } k \]

\[ C(\hat{A}_k, A_k) = \begin{cases} 
(\hat{A}_k - A_k)^2 - M_k^2 & \text{if } |\hat{A}_k - A_k| > M_k \\
0 & \text{otherwise} 
\end{cases} \]

\[ \hat{A}_k = \min \iint C(\hat{A}_k, A_k) p(A_k, \alpha_k / Y_k) dA_k d\alpha_k \]
Audio Samples [4]

Noisy Signal (0 dB)
Ephraim-Malah STSA
Proposed Algorithm
References

