An Engineering model of the masking for the noise-robust speech recognition

Ki-Young Park and Soo-Young Lee

Brain Science Research and
Department of Electrical Engineering and Computer Science
Korea Advanced Institute of Science and Technology
An Engineering model of the masking

Masking

✔ Definition

☒ Process or amount by which the threshold of audibility of a sound is raised by the presence of another sound

✔ Types of Masking

☒ Simultaneous masking
   – Function of masker signal
   – Frequency asymmetric

☒ Non-simultaneous (temporal) masking
   – Function of the intensities of the masker and the probe
   – Function of the time delay between masker and probe
An Engineering model of the masking

Lateral Inhibition Model

✔ Masking
  ❑ Suppression of weak signal adjacent to dominant signal
  ❑ Competition between signal components
  ❑ Suppression of the noise in signal

✔ Implementation of Competition in Neural Networks
  ❑ Winner-takes-all
  ❑ Lateral inhibition
    – Mexican hat
Simultaneous Masking using Lateral Inhibition

✔ Lateral Inhibition by Convolutional Filtering

✔ Filtering Domain

✔ Filter Shape

☐ Mexican hat from difference of Gaussian

\[
w(n) = \exp\left(\frac{-n^2}{\sigma_1^2}\right) - \gamma \exp\left(\frac{-n^2}{\sigma_2^2}\right)
\]

where, \(\sigma_1 < \sigma_2\), \(\gamma < 1\)
An Engineering model of the masking

**Lateral Inhibition with MFCC Model**

- ✔ Windowing the data with a hamming window
- ✔ Short-time Fourier transform
- ✔ Masking filtering
- ✔ critical-bank integration
- ✔ cepstral coefficients conversion

![Diagram of the process](image)
Recognition Experiments with MFCC Model

- Filters Used
  - Simple filter
  - Longer filter
  - Exact shape of the filter would not be critical for recognition performance
  - Downward spread
  - Upward spread

Computation and Neural Systems Lab.
Recognition Results with Various Filters

✔ Isolated Word Recognition Rates with MFCC Features

![Graph showing recognition rates with various filters at different SNRs](image_url)
Lateral Inhibition with ZCPA Model

✔ Model of ZCPA with Simultaneous Masking

Isolated Word Recognition Rates

<table>
<thead>
<tr>
<th>SNR</th>
<th>ZCPA</th>
<th>ZCPA with Masking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clean</td>
<td>95.6</td>
<td>90.8</td>
</tr>
<tr>
<td>30dB</td>
<td>94.3</td>
<td>89.6</td>
</tr>
<tr>
<td>25dB</td>
<td>91.9</td>
<td>87.2</td>
</tr>
<tr>
<td>20dB</td>
<td>87.1</td>
<td>81.4</td>
</tr>
<tr>
<td>15dB</td>
<td>77.2</td>
<td>72.6</td>
</tr>
</tbody>
</table>
An Engineering model of the masking

Temporal Masking

- ✔ Backward Masking
- ✔ Forward Masking
  - ❏ Suppression of weak signal following dominant signal
  - ❏ Adaptation model
  - ❏ Integration model
- ✔ Modeling of Forward Masking
  - ❏ Unilateral inhibition
### Temporal Masking using Integration Model

- **Implementation of Unilateral Integration**
  - Integration sum of preceding stimuli
  
  \[ y(n) = x(n) + A \sum_{k=1}^{\infty} \alpha^k x(n - k) - B \sum_{k=1}^{\infty} \beta^k x(n - k) \]

  - \( A, B \): constants reflecting the amount of integration
  - \( \alpha \): exponential decay of previous response
  - \( \beta \): exponential decay of masking term
  - \( \alpha \) (about 20ms) \(<\) \( \beta \) (about 200ms)

\[
\frac{Y(z)}{X(z)} = A \frac{1 + a_1 z^{-1} + a_2 z^{-2}}{(1 - \alpha z^{-1})(1 - \beta z^{-1})}
\]
An Engineering model of the masking

Analysis of the Model

✔ Impulse Response of the Model

A = .3, B = .03, α = 0.6 (20ms), β = 0.98 (200ms)
Other Feature Transform Methods

✔ RASTA Processing

- Band-pass temporal filtering
  \[ \frac{Y(z)}{X(z)} = 0.1 \frac{2 + 0.1 z^{-1} - 0.1 z^{-3} - 0.2 z^{-4}}{1 - 0.98 z^{-1}} \]

- Impulse response

✔ Hirsch Filter

- High-pass filter
  \[ \frac{Y(z)}{X(z)} = \frac{1 - z^{-1}}{1 - 0.7 z^{-1}} \]

- Impulse response

- Frequency response

Computation and Neural Systems Lab.
An Engineering model of the masking

Recognition Results with Temporal Masking

✔ Isolated Word Recognition Rates with ZCPA feature

![Graph showing misclassification rates with baseline and 16 bins at different SNRs (Clean, 30dB, 25dB, 20dB, 15dB).]
An Engineering model of the masking

**Discussions**

✔ Masking Suppresses Unwanted Noisy Components in Signal

✔ Simultaneous Masking by Lateral Inhibition
  - Recognition performance was enhanced with MFCC model
  - Proposed model can be used with any auditory model
  - ZCPA has spectral masking effect

✔ Temporal Masking by Unilateral Inhibition
  - Unilateral inhibition using the integration model
  - Model resembles other feature processing algorithms
  - Recognition performance was enhanced with RASTA parameters
An Engineering model of the masking

Adaptive Algorithms to Find Optimal Parameters

✔ EBP algorithms with MLP Classifier

❏ Lateral inhibition as the layered structure

❏ Learning at training phase

❏ Find ‘optimal’ values for the recognition

Auditory Model MLP Speech Recognition Result
Comparison with the Psychoacoustic data

✔ Auditory Models to Mimic the Human Performance
  ❏ Modeling and reproducing the psychoacoustic data will enhance the automatic speech recognition

✔ Observations from Psychoacoustics
  ❏ Asymmetric masking pattern
  ❏ Frequency dependent masking pattern

✔ Systematic Ways to Simulate Psychoacoustic Experiments