Robust Speech Recognition Using a Cepstral Minimum-Mean-Square-Error-Motivated Noise Suppressor

Author: Dong Yu, Li Deng, Jasha Droppo, Jian Wu, Yifan Gong, Alex Acero

Professor: 陳嘉平
Reporter: 葉佳璋
Abstract—We present an efficient and effective nonlinear feature-domain noise suppression algorithm, motivated by the minimum-mean-square-error (MMSE) optimization criterion, for noise-robust speech recognition. Distinguishing from the log-MMSE spectral amplitude noise suppressor proposed by Ephraim and Malah (E&M), our new algorithm is aimed to minimize the error expressed explicitly for the Mel-frequency cepstra instead of discrete Fourier transform (DFT) spectra, and it operates on the Mel-frequency filter bank’s output. As a consequence, the statistics used to estimate the suppression factor become vastly different from those used in the E&M log-MMSE suppressor. Our algorithm is significantly more efficient than the E&M’s log-MMSE suppressor since the number of the channels in the Mel-frequency filter bank is much smaller (23 in our case) than the number of bins (256) in DFT. We have conducted extensive speech recognition experiments on the standard Aurora-3 task. The experimental results demonstrate a reduction of the recognition word error rate by 48% over the standard ICSSLP02 baseline, 26% over the cepstral mean normalization baseline, and 13% over the popular E&M’s log-MMSE noise suppressor. The experiments also show that our new algorithm performs slightly better than the ETSI advanced front end (AFE) on the well-matched and mid-mismatched settings, and has 8% and 10% fewer errors than our earlier SPLICE (stereo-based piecewise linear compensation for environments) system on these settings, respectively.
Introduction

• We proposed nonlinear feature-domain noise reduction algorithm motivated by the minimum-mean-square-error (MMSE) criterion on MFCC

• We derive the algorithm by
  - Assigning uniformly distributed random phase to the real-valued filter bank’s outputs
  - Assuming that the artificially generated complex filter bank’s outputs follow zero-mean complex normal distributions
Problem Formulation

• We assume that $x(t)$ is a corrupted with independent additive noise waveform $n(t)$ become the noisy speech waveform, i.e.

$$y(t) = x(t) + n(t)$$

• We get the relationship in the DFT domain

$$Y(f) = X(f) + N(f)$$

• The Mel-frequency filter bank’s output power for noisy feature

$$m_y(b) = \sum_f \omega_b(f) |Y(f)|^2$$

• The kth dimension of MFCC is calculated as

$$c_y(k) \approx \sum_b a_{k,b} m_y(b) \quad a_{k,b} = \cos \frac{\pi b}{B} (k - 0.5)$$
Problem Formulation

• Our goal is to find the MMSE estimate $\hat{c}_x(k)$ against to each separate and independent dimension $k$ of MFCC vector $c_x$

$$\hat{c}_x(k) = \hat{f}(c_y(k)) = \arg\min_f \{E\left(\left(f(c_y(k)) - c_x(k)\right)^2\right)\}$$

$$= \arg\min_f \int \left(\left(f(c_y(k)) - c_x(k)\right)^2 p(c_x(k)) dc_x(k)\right)$$

• Three reasons for choosing the dimension-wise instead of full-vector MMSE criterion
  - Each dimension of MFCC vector is known to be relatively independently with each others
  - The dynamic range of MFCC is vastly different across dimensions
  - The criterion decouples different dimensions, making the algorithm easier to develop and to implement.
Problem Formulation

- The solution is the conditional expectation

\[ \hat{c}_x(k) = E\{c_x(k) | m_y\} = E\left\{ \sum_b a_{k,b} \log m_x(b) | m_y \right\} \]

\[ = \sum_b a_{k,b}(f)E\{\log m_x(b) | m_y\} \]

- Can be further simplified to

\[ \hat{c}_x(k) \approx \sum_b a_{k,b}(f)E\{\log m_x(b) | m_y(b)\} \]

- The problem is reduced to finding the log-MMSE estimator of the Mel frequency filter bank's

\[ \hat{m}_x(b) \approx \exp\left( E\{\log m_x(b) | m_y(b)\} \right) \]
Noise Suppressor for MFCC

• Set up a “straw man” by first rewriting

\[ \hat{m}_x(b) = \exp\left( E\{\log m_x(b) | m_y(b)\} \right) = \exp\left( 2E\{\log \sqrt{m_x(b)} | \sqrt{m_y(b)} \} \right) \]

• The same form in the objective function as the E&M log-MMSE amplitude spectral suppressor

• This naive approach has produced poor recognition results in our experiments.
Fig. 1. Feature extraction pipeline where the E&M log-MMSE magnitude suppressor is directly applied to the magnitude spectrum of the filter bank output.
Noise Suppressor for MFCC

• Note that the filter bank’ output $m_x(b)$, $m_n(b)$, and $m_y(b)$ take real value in the range of $(0, \infty]$, and thus it is inappropriate to model them with real-valued normal distributions.

• To develop appropriate models, we construct three artificial complex variable $M_x(b)$, $M_n(b)$, and $M_y(b)$ such that

$$|M_x(b)| = m_x(b) = \sum_f \omega_b (f)|X(f)|^2$$

$$|M_n(b)| = m_n(b) = \sum_f \omega_b (f)|N(f)|^2$$

$$|M_y(b)| = m_y(b) = \sum_f \omega_b (f)|Y(f)|^2$$

• We choose the ones with uniformly distributed random phases $\theta_x(b)$, $\theta_n(b)$, and $\theta_y(b)$. 
Noise Suppressor for MFCC

• Since $M_y(b)$ contains all information there is in $m_y(b)$, can be rewritten as

$$\hat{m}_x(b) \equiv \exp(E\{\log m_x(b) | M_y(b)\})$$

• We follow the approach adopted in E&M by first evaluating the moment generating function

$$\Phi_b(\mu) = E\{\exp(\mu \log m_x(b) | M_y(b))\}$$

$$= E\{m_x^\mu(b) | M_y(b)\}$$

$$\hat{m}_x(b) = \exp\left(\frac{d}{d\mu} \Phi_b(\mu) |_{\mu=0}\right) \quad \frac{d}{d\mu} m_x^\mu = m_x^\mu \log m_x$$
Noise Suppressor for MFCC

- We assume that $\theta_x(b)$, $\theta_n(b)$, and $\theta_y(b)$ are independent and uniformly distributed random variables.

\[
\Phi_b(\mu) = E\{m^\mu_x(b) | M_y(b)\} = \int_0^\infty \int_0^{2\pi} m^\mu_x(b) p(M_y(b), m_x(b), \theta_x(b)) dm_x(b) d\theta_x(b) / p(M_y(b))
\]

\[
= \int_0^\infty \int_0^{2\pi} m^\mu_x(b) p(M_y(b) | m_x(b), \theta_x(b)) p(m_x(b), \theta_x(b)) dm_x(b) d\theta_x(b)
\]

\[
= \int_0^\infty \int_0^{2\pi} (b) p(M_y(b) | m_x(b), \theta_x(b)) p(m_x(b), \theta_x(b)) dm_x(b) d\theta_x(b)
\]

- $M_x(b)$ is assumed to follow the zero-mean complex normal distribution

\[
p(m_x(b), \theta_x(b)) = \frac{m_x(b)}{\pi \sigma_x^2(b)} \exp\left\{-\frac{m^2_x(b)}{\sigma_x^2(b)}\right\}
\]

- Where

\[
\sigma_x^2(b) \overset{\text{def}}{=} E\{M_x(b)^2\} = E\{m_x^2(b)\}
\]
Noise Suppressor for MFCC

• Similarly, given that \( M_y(b) - M_x(b) \)

\[
p(M_y(b) | m_x(b), \theta_x(b)) = \frac{1}{\pi \sigma_d^2(b)} \exp \left\{ -\frac{|M_y(b) - m_x(b) e^{j\theta_x(b)}|^2}{\sigma_d^2(b)} \right\}
\]

\[
= \frac{1}{\pi \sigma_d^2(b)} \exp \left\{ -\frac{|m_y(b) e^{j\theta_y(b)} - m_x(b) e^{j\theta_x(b)}|^2}{\sigma_d^2(b)} \right\}
\]

\[
= \frac{1}{\pi \sigma_d^2(b)} \exp \left\{ -\frac{|m_y(b) \cos(\theta_y(b)) - m_x(b) \cos(\theta_x(b)) + j(m_y(b) \cos(\theta_y(b)) - m_x(b) \cos(\theta_x(b)))|^2}{\sigma_d^2(b)} \right\}
\]

\[
= \frac{1}{\pi \sigma_d^2(b)} \exp \left\{ -\frac{|m_y^2(b) + m_x^2(b) + 2m_y(b)m_x(b) \cos(\theta_y(b) - \theta_x(b))|^2}{\sigma_d^2(b)} \right\}
\]

where \( \sigma_d^2(b) \stackrel{\text{def}}{=} E\left\{|M_y(b) - M_x(b)|^2\right\} \geq E\left\{(m_y(b) - m_x(b))^2\right\} \)
Noise Suppressor for MFCC

• Since

\[ m_y(b) = \sum_f \omega_b(f)|Y(f)|^2 \]

\[ = \sum_f \omega_b(f)\left(|X(f)|^2 + |N(f)|^2 + 2|X(f)||N(f)|\cos \phi(f)\right) \]

\[ = m_x(b) + m_n(b) + \sum 2\omega_b(f)|X(f)||N(f)|\cos \phi(f) \]

Where \( \phi(f) \) is the phase difference of \( X(f) \) and \( N(f) \)

\[ \sigma_d^2(b) \geq E\left\{ \left(m_n(b) + \sum 2\omega_b(f)|X(f)||N(f)|\cos \phi(f)\right)^2 \right\} \]

\[ = E\left\{ m_n^2(b) \right\} + E\left\{ \left(\sum_f 2\omega_b(f)|X(f)||N(f)|\cos \phi(f)\right)^2 \right\} \]

where \( E\left\{ 2m_n(b)\left(\sum_f 2\omega_b(f)|X(f)||N(f)|\cos \phi(f)\right) \right\} \approx 0 \)

\[ \sigma_d^2(b) \equiv \sigma_x^2(b) + \sigma_\phi^2(b) \]

• One of major different from E&M. In E&M

\[ \sigma_d^2(b) \overset{\text{def}}{=} E\left\{|Y(f) - X(f)|^2\right\} = E\left\{|N(f)|^2\right\} = \sigma_n^2(b) \]
Noise Suppressor for MFCC

• By substituting and replacing variable $\theta_y(b) - \theta_x(b)$ by $\beta(b)$

$$
\Phi_b(\mu) = E\{m_x^\mu(b) | M_y(b)\} = 
\frac{\int_0^\infty m_x^{\mu+1}(b) \exp\left\{-\frac{m_x^2(b)}{\sigma_x^2(b)} - \frac{m_x^2(b)}{\sigma_d^2(b)}\right\} g(m_x(b)) dm_x(b)}{
\int_0^\infty m_x(b) \exp\left\{-\frac{m_x^2(b)}{\sigma_x^2(b)} - \frac{m_x^2(b)}{\sigma_d^2(b)}\right\} g(m_x(b)) dm_x(b)}
$$

$$
g(m_x(b)) = \int_0^{2\pi} \frac{1}{\pi \sigma_x^2(b)} \exp\left\{-\frac{2m_x(b)m_x(b)\cos(\beta(b))}{\sigma_d^2(b)}\right\} d\beta(b)
$$

• This can be show simplified

$$
g(m_x(b)) = I_0\left(2m_x(b)\sqrt{\frac{v(b)}{\sigma^2(b)}}\right), \text{where } I_0(z) = \int_0^{2\pi} \exp(z \cos \beta) d\beta
$$

$$
\frac{1}{\sigma^2(b)} = \frac{1}{\sigma_x^2(b)} + \frac{1}{\sigma_d^2(b)}
$$

$$
v(b) = \frac{\xi(b)}{1 + \xi(b)} \gamma(b)
$$
Noise Suppressor for MFCC

• $v(b) = \frac{\xi(b)}{1 + \xi(b)} \gamma(b)$ is defined from a priori signal-to-noise ratio

\[
\xi(b) \overset{\text{def}}{=} \frac{\sigma_x^2(b)}{\sigma_d^2(b)} \approx \frac{\sigma_x^2(b)}{\sigma_n^2(b) + \sigma_\phi^2(b)}
\]

• And the adjusted a posteriori SNR

\[
\gamma(b) \overset{\text{def}}{=} \frac{\sigma_y^2(b)}{\sigma_d^2(b)} \approx \frac{m_y^2(b)}{\sigma_n^2(b) + \sigma_\phi^2(b)}
\]

• Rewritten as

\[
\Phi_b(\mu) = E\{m_x^\mu(b) | M_y(b)\} = \frac{\int_0^\infty m_x^{\mu+1}(b) \exp\left\{ -\frac{m_x^2(b)}{\sigma_x^2(b)} - \frac{m_x^2(b)}{\sigma_d^2(b)} \right\} I_0(2m_x(b)) \sqrt{\frac{v(b)}{\sigma^2(b)}} dm_x(b)}{\int_0^\infty m_x(b) \exp\left\{ -\frac{m_x^2(b)}{\sigma_x^2(b)} - \frac{m_x^2(b)}{\sigma_d^2(b)} \right\} I_0(2m_x(b)) \sqrt{\frac{v(b)}{\sigma^2(b)}} dm_x(b)}
\]
Noise Suppressor for MFCC

\[ \Phi_b(\mu) = \sigma^{\mu/2} \Gamma(\mu/2 + 1) M(\mu/2; 1; -v(b)) \]

, where \( \Gamma(\bullet) \) gamma function \( M(a;c;x) \) confluent hypergeometric function

\[ \left. \frac{\partial}{\partial \mu} M(\mu/2; 1; -v(b)) \right|_{\mu=0} = -\frac{1}{2} \sum_{r=1}^{\infty} \frac{(-v)^r}{r!} \frac{1}{r} \quad \left. \frac{\partial}{\partial \mu} \Gamma\left(\frac{\mu}{2} + 1\right) \right|_{\mu=0} = -\frac{c}{2} \]

\[ \left. \frac{d}{d\mu} \Phi_b(\mu) \right|_{\mu=0} = \frac{1}{2} \ln \sigma + \frac{1}{2} \left( \ln v(b) + \int_{v(b)}^{\infty} \frac{(e^{-t})}{t} dt \right) \]

\[ \hat{m}_x(b) = \exp\left( E\{\log m_x(b) | m_y(b)\}\right) = G(\xi(b), v(b))m_y(b) \]

where

\[ G(\xi(b), v(b)) = \frac{\xi(b)}{1 + \xi(b)} \exp\left\{ \frac{1}{2} \int_{v(b)}^{\infty} \frac{e^{-t}}{t} dt \right\} \]

- The MMSE estimate for MFCC is thus

\[ \hat{c}(k) = \sum_b a_{k,b} E\{\log m_x(b) | m_y(b)\} = \sum_b a_{k,b} \log\left( G(\xi(b), v(b))m_y(b)\right) \]
Estimation of Parameters

- To apply the noise reduction algorithm, we need to estimate the noise variance $\sigma_n^2(b)$, the variance $\sigma_\phi^2(b)$ and clean speech variance $\sigma_x^2(b)$

- Estimate of $\sigma_n^2(b)$
  - Using a minimum-controlled recursive movie-average noise tracker
  - A decision on whether a frame contains speech is made based on energy ratio test
    \[
    \frac{|\bar{m}_y(b)|_t^2}{|\bar{m}_n(b)|^2_{\text{min}}} > \theta
    \]
    Where $\theta$ is threshold, $|\bar{m}_n(b)|^2_{\text{min}}$ is the smoothed minimum noise power, $|\bar{m}_y(b)|_t^2$ is the smoothed power of the bth filter’s output at the tth frame.

  - If the energy ratio is true the frame is assumed to contain speech the new noise estimate of the noise variance becomes
    \[
    \sigma_n^2(b)_t = \sigma_n^2(b)_{t-1}, \quad \text{otherwise} \quad \sigma_n^2(b)_t = \alpha \sigma_n^2(b)_{t-1} + (1-\alpha)|m_y(b)|_t^2
    \] using smoothing factor $\alpha$
Estimation of Parameters

• Estimation of $\sigma_x^2(b)$

- Using decision-directed approach.
- $\sigma_x^2(b)$ for the current frame is estimated using the estimated clean speech from the previous frame and smoothed over the past frames.
Estimation of Parameters

• Estimation of $\sigma_\varphi^2(b)$

$$
\sigma_\varphi^2(b) = E\left\{\left(\sum_f 2\omega_b(f)|X(f)||N(f)|\cos\varphi(f)\right)^2\right\}
$$

$$
= 4\sum_f E\left\{(\omega_b(f)|X(f)||N(f)|\cos\varphi(f))^2\right\}
$$

$$
= 4\sum_f E\left\{(\omega_b(f)|X(f)||N(f)|)^2\right\} \times \int_0^{2\pi} \frac{1}{2\pi} \cos\varphi(f) d\varphi(f)
$$

$$
= 2\sum_f E\left\{(\omega_b(f)|X(f)||N(f)|)^2\right\} = 2\sum_f \omega_b^2(f) E\{|N(f)|\}^2 E\{|X(f)|\}^2
$$

- Since we only estimate and keep track of statistics at the real-valued filter bank’s output, we approximate $\sigma_\varphi^2(b)$ as

$$
\sigma_\varphi^2(b) = 2\sum_f \omega_b^2(f) E\{|N(f)|\}^2 E\{|X(f)|\}^2
$$

$$
\approx 2E\{m_x(b)\} E\{m_n(b)\} \frac{\sum_f \omega_b^2(f)}{\sum_f \sum_{f_1} \omega_b(f_1)\omega_b(f_2)}
$$

$$
\approx 2 \frac{E\{m_x(b)\}}{E\{m_n(b)\}} E\{m_n^2(b)\} \frac{\sum_f \omega_b^2(f)}{\left(\sum_f \omega_b(f)\right)^2}
$$

$$
\approx 2 \frac{\sum_f \omega_b^2(f)}{\left(\sum_f \omega_b(f)\right)^2} \sqrt{\sigma_x^2(b)\sigma_n^2(b)}
$$
Experiment setup

- Aurora3 corpus
- Close-talking or a hand-free microphone
- 39-dimension features used in our experiment
  - 13-dimension(with energy and without c0) static MFCC
  - Their delta and delta-delta feature
- The threshold $g$ was set to 0.9, and the parameter $\alpha$ set to 5.
Fig. 5. Feature extraction pipeline for the ICSLP02 baseline system.

Fig. 6. Feature extraction pipeline for the CMN baseline system.
Fig. 1. Feature extraction pipeline where the E&M log-MMSE magnitude suppressor is directly applied to the magnitude spectrum of the filter bank output.

Fig. 7. Feature extraction pipeline for the E&M log-MMSE system [8], where the suppressor is applied to the DFT bins.
Fig. 8. Feature extraction pipeline for the MFCC-MMSE system.

Fig. 9. Feature extraction pipeline for the SPLICE systems.
### TABLE I
**Summary of Absolute WER on the Standard Test Sets in the Aurora-3 Task Under Different Experimental Settings**

<table>
<thead>
<tr>
<th>Summary of Aurora 3 Absolute Word Error Rate (Standard Set)</th>
<th>Well</th>
<th>Mid</th>
<th>High</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICSLP02 Baseline</td>
<td>8.96%</td>
<td>21.96%</td>
<td>48.85%</td>
<td><strong>23.48%</strong></td>
</tr>
<tr>
<td>CMN</td>
<td>6.87%</td>
<td>16.52%</td>
<td>31.11%</td>
<td>16.31%</td>
</tr>
<tr>
<td>FB Output Magnitude</td>
<td>6.87%</td>
<td>15.21%</td>
<td>31.29%</td>
<td>15.89%</td>
</tr>
<tr>
<td>E&amp;M log-MMSE</td>
<td>5.57%</td>
<td>12.79%</td>
<td>29.23%</td>
<td>14.01%</td>
</tr>
<tr>
<td>MFCC-MMSE</td>
<td>5.08%</td>
<td>12.26%</td>
<td>23.26%</td>
<td>12.13%</td>
</tr>
</tbody>
</table>

### TABLE II
**Summary of Relative WER Reduction on the Standard Test Sets in the Aurora-3 Task Under Different Experimental Settings**

<table>
<thead>
<tr>
<th>Summary of Aurora 3 Relative Improvement (Standard Set)</th>
<th>Relative to CMN</th>
<th>ICSLP02 Baseline</th>
<th>CMN</th>
<th>E&amp;M log-MMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative to CMN</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CMN</td>
<td></td>
<td>30.55%</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>E&amp;M log-MMSE</td>
<td></td>
<td>40.33%</td>
<td>14.08%</td>
<td>--</td>
</tr>
<tr>
<td>MFCC-MMSE</td>
<td></td>
<td>48.33%</td>
<td>25.59%</td>
<td>13.41%</td>
</tr>
</tbody>
</table>
TABLE III  
**DETAILED AURORA-3 ABSOLUTE WER RESULTS ON THE STANDARD TEST SETS UNDER THE MFCC-MMSE EXPERIMENTAL SETTING**

<table>
<thead>
<tr>
<th>Aurora-3 Word Error Rate with MFCC-MMSE (Standard Set)</th>
<th>Finnish</th>
<th>Spanish</th>
<th>German</th>
<th>Danish</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Well (x40%)</td>
<td>3.54%</td>
<td>5.90%</td>
<td>5.20%</td>
<td>5.66%</td>
<td>5.08%</td>
</tr>
<tr>
<td>Mid (x35%)</td>
<td>15.12%</td>
<td>5.39%</td>
<td>10.67%</td>
<td>17.84%</td>
<td>12.26%</td>
</tr>
<tr>
<td>High (x25%)</td>
<td>17.99%</td>
<td>34.77%</td>
<td>10.78%</td>
<td>29.49%</td>
<td>23.26%</td>
</tr>
<tr>
<td>Overall</td>
<td>11.21%</td>
<td>12.94%</td>
<td>8.51%</td>
<td>15.88%</td>
<td>12.13%</td>
</tr>
</tbody>
</table>

TABLE IV  
**DETAILED AURORA-3 WER REDUCTION RESULTS ON THE STANDARD TEST SETS AGAINST THE ICSLP02 BASELINE UNDER THE MFCC-MMSE**

<table>
<thead>
<tr>
<th>Aurora-3 Relative Improvement with MFCC-MMSE (Standard Set)</th>
<th>Finnish</th>
<th>Spanish</th>
<th>German</th>
<th>Danish</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Well (x40%)</td>
<td>51.24%</td>
<td>16.43%</td>
<td>40.91%</td>
<td>55.50%</td>
<td>43.36%</td>
</tr>
<tr>
<td>Mid (x35%)</td>
<td>22.42%</td>
<td>67.71%</td>
<td>43.72%</td>
<td>45.41%</td>
<td>44.18%</td>
</tr>
<tr>
<td>High (x25%)</td>
<td>69.75%</td>
<td>28.24%</td>
<td>59.82%</td>
<td>51.36%</td>
<td>52.39%</td>
</tr>
<tr>
<td>Overall</td>
<td>54.44%</td>
<td>37.73%</td>
<td>49.54%</td>
<td>49.88%</td>
<td>48.32%</td>
</tr>
<tr>
<td>Relative to -&gt;</td>
<td>CMN</td>
<td>E&amp;M log-MMSE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>--------</td>
<td>--------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E&amp;M log-MMSE</td>
<td>20.33%</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MFCC-MMSE</td>
<td>21.72%</td>
<td>1.75%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**TABLE VII**

**Comparison Between the MFCC-MMSE System and the ETSI’s AFE on the Aurora-3 Task**

<table>
<thead>
<tr>
<th>Compare with AFE on Aurora 3 (Standard Set)</th>
<th>Well</th>
<th>Mid</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>ETSI AFE</td>
<td>4.70%</td>
<td>13.21%</td>
<td>12.75%</td>
</tr>
<tr>
<td>MFCC-MMSE</td>
<td>5.08%</td>
<td>12.26%</td>
<td>23.26%</td>
</tr>
</tbody>
</table>

**TABLE VIII**

**Comparison Between the MFCC-MMSE System and the SPLICE on Aurora-3 Where the SPLICE Code Book was Trained Using Additional Information to Make a Matching Condition**

<table>
<thead>
<tr>
<th>Comparisons with SPLICE on Aurora-3 (Standard Set)</th>
<th>Well</th>
<th>Mid</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPLICE</td>
<td>5.49%</td>
<td>13.55%</td>
<td>11.42%</td>
</tr>
<tr>
<td>MFCC-MMSE</td>
<td>5.08%</td>
<td>12.26%</td>
<td>23.26%</td>
</tr>
</tbody>
</table>