Speech Enhancement Based on Maximum Likelihood Adaptive Subspace Estimation

Qi Sun¹,²*, Xiaohui Zhao¹

¹College of Communication Engineering, Jilin University, Changchun 130012, China;
²School of Electronics and Information Engineering, Shenyang Aerospace University, Shenyang 110136, China

*Corresponding author E-mail: sq227@126.com

Abstract

Maximum likelihood adaptive subspace estimation (MALASE) is a new method to deal with fast adaptive eigen decomposition problem. We use it for our subspace based speech enhancement algorithm to realize subspace tracking. The tracking results are the eigenvalues and eigenvectors of noisy speech covariance matrix, and the estimated eigenvectors are strictly orthogonal. In our proposed algorithm, the noise eigenvalue estimation with speech presence probability in subspace domain can be calculated by recursive smoothing. The simulation results show that the proposed algorithm achieves a better speech enhancement under different noise environment and input SNR with lower speech distortion than the classical algorithm and MCRA algorithm.

Keywords: speech enhancement, subspace, MALASE, noise eigen value estimation.

1. INTRODUCTION

Subspace based speech enhancement algorithm has balancing mechanism for control speech distortion and residual noise, and has been deeply developed in recent years. The original signal subspace approach for speech enhancement is proposed only for the case of white noise (Ephraim and Van Trees, 1995). A subspace speech enhancement algorithm by optimal speech estimation for both white noise and color noise is provided (Hu and Loizou, 2003). In order to further improve performance of speech enhancement, this method can also be studied by introducing many other methods (Saadoune et al., 2014; Yahya et al., 2014; Saadoune et al., 2013; Jabloun and Champagne, 2003; Lu et al., 2013; Jensen et al., 2013), for example, a subspace method combined with human hearing properties (Jabloun and Champagne, 2003), a subspace method combined with kernel Hilbert space (Lu et al., 2013).

Unfortunately, the subspace method requires eigen decomposition of speech data matrix with $O(K^3)$ computational complexity, where $K$ is the frame length of the sampled speech data. In many practical environments, the speech data matrix is time varied. Therefore, to develop adaptive subspace algorithm for speech enhancement is very important.

Subspace tracking algorithms can be divided into two categories. The first one is modified eigen value problem (MEP). In this category, the algorithm usually adopts rank-1 or rank-2 method to modify covariance matrix of the speech signal so that the classical eigen decomposition is directly extended from the stationary process to non-stationary process. While the other is non modified eigen value problem (non-MEP) or adaptive method, in which the update calculation can finally result in the convergent estimation of eigen decomposition. Projection approximation subspace tracking is one of the non-MEP
methods widely used in speech enhancement (Yang1995; Rezayee and Gazor, 2001; Choi2013; Oyerinde and Mneney, 2012; Chan et al., 2012; Nguyen and Yamada, 2014), since this algorithm has low computational complexity. But it adopts various hypotheses and uses voice activity detection (VAD) to estimate the noise eigen values. In the non-stationary noise environment and low SNR, it is difficult to accurately track the changeing environment.

In order to solve the above problems, a new subspace speech enhancement algorithm using Maximum Likelihood Adaptive Subspace Estimation (MALASE) with noise eigenvalue estimation is proposed. The adaptive tracking of eigen values and eigenvectors of data covariance matrix realizes instantaneous eigen decomposition with $O(K^2)$ computational complexity (Chonavel et al., 2003). Since MALASE algorithm uses a similar Givens rotation, it ensures the orthogonality of the estimated eigenvectors. The proposed noise eigen value estimation is conducted in subspace domain through recursive smoothing to obtain noise eigen values (Wu et al., 2009; Cohen and Berdugo, 2002).

2. CLASSICAL SUBSPACE SPEECH ENHANCEMENT

Assuming the noisy speech signal at time $n$ is

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

(1)

Where $y(n), x(n)$ and $n(n)$ represent $K$-dimensional noisy speech vector, clean speech vector and additive noise vector uncorrelated with $x(n)$, respectively

$$R_y(n) = R_x(n) + R_n(n)$$

(2)

Where $R_y(n), R_x(n)$ and $R_n(n)$ are the $K \times K$ covariance matrixes of $y(n), x(n)$ and $n(n)$ respectively.

The linear estimator of $x(n)$ is given by

$$\hat{x}(n) = H(n) \cdot y(n)$$

(3)

Where $H(n)$ represents the $K \times K$ linear estimation matrix, the optimal estimation matrix $H_{opt}(n)$ can be obtained by the Lagrange multiplier method, then we have

$$H_{opt}(n) = R_x(n)(R_x(n) + \mu R_n(n))^{-1}$$

(4)

Where $\mu$ is the Lagrange multiplier.

Eq. (4) can be simplified by the eigen decomposition of $R_x(n) = U(n)\Lambda_x(n)U^T(n)$ as

$$H_{opt}(n) = U(n)\Lambda_x(n)(\Lambda_x(n) + \mu U^T(n)R_n(n)U(n))^{-1}U^T(n)$$

(5)

Where $U(n)$ is the eigenvector matrix of $R_x(n)$, $\Lambda_x(n)$ is the diagonal eigen value matrix of $R_x(n)$.

When $n(n)$ is white noise with variance $\lambda_n(n)$, substituting $R_n(n) = \lambda_n(n)I$ into Eq. (5), we have (Ephraim and Van Trees, 1995)
When $n(n)$ is color noise, simulation experiments show that $U^T(n)R_y(n)U(n)$ is not usually a diagonal matrix, $U^T(n)R_y(n)U(n)$ is approximately the diagonal matrix

$$\Delta_y(n) = \text{diag}(\lambda_{y1}(n), \lambda_{y2}(n), \cdots, \lambda_{yk}(n)) \quad (7)$$

Where $\lambda_{yn}(n)$ is the noise variance along the $i$th eigenvector. This approximation is less restrictive and covers the case of white noise, i.e., for all $i, \lambda_{yn}(n) = \lambda_{yn}$. When we substitute Eq. (7) into Eq. (5), the suboptimal estimation matrix can be obtained as

$$H(n) = U(n)A_y(n)(A_y(n) + \mu\Delta_y(n))^{-1}U^T(n) \quad (8)$$

Substituting $\Delta_y(n) \approx U^T(n)R_y(n)U(n)$ into Eq. (2), we get

$$R_y(n) \approx U(n)(A_y(n) + \Delta_y(n))U^T(n) \quad (9)$$

It is clear that the eigenvector of $R_y(n)$ is the same as that of $R_y(n)$. Since we have no access to the covariance matrix of the clean signal, the eigenvector matrix $U(n) = [u_1(n), u_2(n), \cdots, u_K(n)]$ and the eigenvalue diagonal matrix $\Delta_y(n) = \text{diag}(\lambda_{y1}(n), \lambda_{y2}(n), \cdots, \lambda_{yk}(n))$ can be calculated by the eigen decomposition of $R_y(n)$, and

$$A_y(n) = A_y - \Delta_y \quad (10)$$

Where $\Delta_y(n)$ is the noise eigen value diagonal matrix, its elements can be calculated by using noise samples obtained from silence intervals between speech with speech activity detector.

To solve the problem requires the eigen decomposition of $R_y(n)$ with high computational complexity $O(K^3)$. In order to reduce this complexity, a projection approximation subspace tracking algorithms is presented (Rezayee and Gazor, 2001), where the eigenvalues and the eigenvectors of $R_y(n)$ can be obtained in an iterative way with the complexity $O(K)$. This method can be considered as the classical adaptive subspace speech enhancement algorithm. It uses some hypotheses and VAD to estimate noise eigenvalues. In the cases of non-stationary noise and low SNR, the hypotheses of the classical algorithm may fail. To overcome these mentioned shortcomings, a new adaptive method for the estimation of the eigenvalues and the eigenvectors of $R_y(n)$ is developed by using MALASE.

3. PROPOSED SPEECH ENHANCEMENT BASED ON MALASE

3.1 Estimation of the Eigenvalues of $R_y(n)$ by MALASE

MALASE is proposed based on the log-likelihood criterion optimization (Chonavel et al., 2003), which can realize iterative estimation for the eigenvalues and eigenvectors of the sample data covariance matrix to quickly solve eigendecomposition.

For simplicity, let $R_y(n) = R_y$, we have
\[
f(y(n); R_y) = \frac{1}{\pi^N |R_y|} \exp(-y^H(n)R_y^{-1}y(n))
\]  

(11)

Where \( f \) denotes the probability density function of \( y(n) \). After direct calculation for Eq. (11), we can define log-likelihood function as

\[
\psi(y(n); R_y) = \log |R_y| + y^H(n)R_y^{-1}y(n)
\]  

(12)

Since the eigen decomposition of \( R_y = U \Delta_y U^H \), where \( U \) is the eigenvector matrix, and \( \Delta_y \) is the diagonal eigen value matrix of \( R_y \), the maximization of log-likelihood function of \( y(n) \) is equivalent to the minimization of the following formula

\[
\Phi(y(n); U, \Delta_y) = \sum_{i=1}^{K} \log \lambda_{yi} + y^H(n)U \Delta_y^{-1}U^H y(n)
\]  

(13)

Based on iterative optimization of log-likelihood function by calculating the gradient, the iterative update equations for \( U(n) \) and \( \Delta_y(n) \) of \( R_y(n) \) are as follows (Chonavel et al., 2003)

\[
U(n) = U(n-1) \exp[\gamma(\Delta_y^{-1}(n-1)a(n)a^T(n) - a(n)a^T(n)\Delta_y^{-1}(n-1))]
\]  

(14)

\[
\Delta_y(n) = \Delta_y(n-1) + \gamma'(\Delta_y^{-2}(n-1)\text{diag}(a(n)a^T(n))) - \Delta_y^{-1}(n-1))
\]  

(15)

where

\[
a(n) = U^T(n-1)y(n)
\]  

(16)

\( \gamma \) and \( \gamma' \) are the step sizes. The convergence rate may be different for \( U(n) \) and \( \Delta_y(n) \) when \( \gamma = \gamma' \).

It has been proven that a square matrix \( A \) is a skew-symmetric matrix (i.e. \( A^T = -A \)) if and only if \( \exp(A) \) is a unitary matrix. According to this definition, matrix\( \Delta_y^{-2}(n-1)a(n)a^T(n) - a(n)a^T(n)\Delta_y^{-1}(n-1) \) is clearly skew-symmetric, so that the update factor \( \exp[\gamma(\Delta_y^{-1}(n-1)a(n)a^T(n) - a(n)a^T(n)\Delta_y^{-1}(n-1))] \) is a unitary matrix in Eq. (14). During initialization, if we take the initial matrix \( U(0) \) as a unitary matrix, we can guarantee that \( U(n) \) is unitary matrix after iteration, i.e. the updated eigenvectors are mutually orthogonal. Even \( U(0) \) is not a unitary matrix, due to the fast convergence rate of MALASE, \( U(n) \) is still close to the unitary matrix.

In MALASE algorithm, the matrix exponential in Eq. (14) is the principle part for the computational complexity. Its computational complexity is \( O(K^2) \) (Chonavel et al., 2003).

In practical applications, for \( K \)-dimensional noisy speech vector \( y(n) \), MALASE algorithm can be implemented as follows:

Initialize \( U(0), \Delta_y(0), i=1,2,\cdots,K \) and \( \gamma, \gamma' > 0 \)

For each step

input \( y(n) \)
\[ a(n) = U^T(n-1)y(n) \] (17)

\[ z_i(n) = \lambda_{yi}^{-1}(n-1)a_i(n) \] (18)

\[ c(n) = \sqrt{\|a(n)\|^2 - \|z(n)\|^2 - (a^T(n)z(n))^2} \] (19)

\[ Q(n) = \begin{bmatrix} \frac{z(n)}{\|z(n)\|} & \frac{z(n)(z^T(n)a(n)-a(n))\|z(n)\|^2}{c(n)\|z(n)\|^2} \end{bmatrix} \] (20)

\[ U(n) = U(n-1) + (U(n-1)Q(n)) \begin{bmatrix} \cos \gamma c(n) - 1 & -\sin \gamma c(n) \\ \sin \gamma c(n) & \cos \gamma c(n) - 1 \end{bmatrix} Q^T(n) \] (21)

\[ \lambda_{yi}(n) = \lambda_{yi}(n-1) + \gamma'(\lambda_{yi}^{r2}(n-1)|a_i(n)|^2 - \lambda_{yi}^{-1}(n-1)) \] (22)

Where \( \lambda_{yi}(0) \) can be chosen for an instance regularly spaced in the interval \([a,b]\), where \( a \) and \( b \) are rough estimates of the smallest eigenvalue and the largest eigenvalue, and \( U(0) \) can be set to a unit matrix.

In order to get a better initialization for the proposed algorithm, \( U(0) \) and \( \lambda_{yi}(0) \) can be obtained by using the eigen decomposition of initial speech data covariance matrix. We take \( \gamma = 0.000065 \) and \( \gamma' = 300 \) for the proposed algorithm.

It needs to mention that the reciprocal of the eigenvalues is involved in Eq. (22). The computer simulations show that some eigenvalues of the speech signal are very small and close to zero, to avoid the happening of the numerical problem, Eq. (22) can be modified as follows

\[ \lambda_{yi}(n) = \lambda_{yi}(n-1) + \gamma'(\lambda_{yi}^{r2}(n-1)|a_i(n)|^2 - \lambda_{yi}^{-1}(n-1)) \] (23)

Where \( \beta > 0 \), in the proposed algorithm, we take \( \beta = 40 \).

### 3.2 Proposed Speech Enhancement Algorithm

According to MALASE algorithm, we can obtain the eigenvalues and the eigenvectors of \( R_y(n) \). Then we need to make the estimation for the eigenvalues of \( R_y(n) \). In the classical algorithm, the noisy speech is distinguished by voice activity detection (VAD) by updating the noise estimation in silent segment of speech. But the real-time change of noise, specifically, in the non-stationary noise environments and low SNR, it is hard to decide the silent segments of the noisy speech. As a result, some assumptions for the classical algorithm may fail, and the residual noise and the speech distortion would be unacceptable because of inaccurate eigenvalues estimation of the noise eigenvalues of \( R_y(n) \), which degrade the performance of the speech enhancement algorithm.

In order to overcome the above mentioned shortcomings, we propose an eigenvalue estimation method for \( R_y(n) \) with the speech presence probability developed in subspace domain, in which the corresponding noise eigenvalue \( \lambda_{yi}(n) \) can be obtained by recursive smoothing on the eigenvalue \( \lambda_{yi}(n) \) of \( R_y(n) \)(Wu et al., 2009; Cohen and Berdugo, 2002).

52
First-order recursive smoothing on $\lambda_i(n)$ is given by

$$\lambda_i(n) = \alpha_i \lambda_i(n-1) + (1 - \alpha_i) \lambda_i(n)$$  \hspace{1cm} (24)$$

where $0 < \alpha_i < 1$ is a parameter. We continuously search and update the local minimum of $\lambda_i(n)$ by

$$\lambda_{i(min)}(n) = \min \{ \lambda_{i(min)}(n-1), \lambda_i(n) \}$$  \hspace{1cm} (25)$$

At the same time, a temporary variable is defined as

$$\lambda_{i(tmp)}(n) = \min \{ \lambda_{i(tmp)}(n-1), \lambda_i(n) \}$$  \hspace{1cm} (26)$$

We also define a searching window of length $L$ in order to ensure that the searching window contains pure noise frame, $L$ is typically about $0.5 \sim 1.5$s in time. When $n$ is an integer multiple of $L$, we update the local minimum $\lambda_{i(min)}(n)$ and temporary variables $\lambda_{i(tmp)}(n)$, i.e.

$$\lambda_{i(min)}(n) = \min \{ \lambda_{i(min)}(n-1), \lambda_i(n) \}$$  \hspace{1cm} (27)$$

$$\lambda_{i(tmp)}(n) = \hat{\lambda}_i(n)$$  \hspace{1cm} (28)$$

The estimator for speech presence probability is

$$\hat{p}_i(n) = \alpha_p \hat{p}_i(n-1) + (1 - \alpha_p) I_i(n)$$  \hspace{1cm} (29)$$

where $0 < \alpha_p < 1$ is a smoothing parameter. $I_i(n)$ is an indicative function defined as

$$I_i(n) = \begin{cases} 1, & \eta_i(n) > \delta \\ 0, & \eta_i(n) < \delta \end{cases}$$  \hspace{1cm} (30)$$

Where $\eta_i(n) \equiv \lambda_i(n)/\lambda_{i(min)}(n)$, and $\delta$ is the threshold value. If $\eta_i(n)$ is greater than $\delta$, there is the presence of speech; otherwise, the absence of speech.

The eigenvalue estimation of $R(n)$ can be expressed as

$$\hat{\lambda}_w(n) = \tilde{a}_{di}(n) \hat{\lambda}_w(n-1) + [1 - \tilde{a}_{di}(n)] \hat{\lambda}_yi(n)$$  \hspace{1cm} (31)$$

where $\tilde{a}_{di}(n)$ is a smoothing factor, $\tilde{a}_{di}(n) \equiv a_d + (1 - a_d) \hat{p}_i(n), a_d$ is a constant, and $\tilde{a}_{di}(n)$ is in the range of $a_d < \tilde{a}_{di}(n) < 1$.

From Eq. (31), we know if the speech is present, $\tilde{a}_{di}(n) \approx 1$, $\hat{\lambda}_w(n) \approx \hat{\lambda}_w(n-1)$, which implies that there is no update for the estimation of the eigenvalues of the noise. Since the speech presence probability $\hat{p}_i(n)$ is independently along with each eigenvector, the estimation for the noise eigenvalues can be constantly adjusted and updated even in speech segment.
For $y(n)$, subspace speech enhancement algorithm can be realized by the combination of MALASE and noise eigenvalue estimation as follows:

Initialize $\hat{\lambda}_u(0), \mu=1$

For each step

For $i=1,2,\ldots,K$

$$\hat{\lambda}_u(n) = \max\{\hat{\lambda}_u(n) - \hat{\lambda}_n(n), 0\} \quad (32)$$

$$g_i(n) = \frac{\hat{\lambda}_u(n)}{\hat{\lambda}_u(n) + \mu \hat{\lambda}_n(n)} \quad (33)$$

Where $\lambda_u(n)$ is the $i$th eigenvalue of $R_u(n)$, and $\lambda_u(n) = \lambda_v(n) - \hat{\lambda}_n(n)$

END

$$\hat{x}(n) = H(n)y(n) \quad (34)$$

Where $H(n) = U(n)G(n)U^T(n)$, and $G(n) = \text{diag}(g_1(n), g_2(n), \ldots, g_K(n))$.

4. SIMULATION RESULTS AND ANALYSIS

In this section, the performance comparison of the proposed algorithm with that of the classical algorithm and the noise estimation by minima controlled recursive averaging (MCRA) is provided (Rezayee and Gazor, 2001; Cohen and Berdugo, 2002). We use 6 different types of noise: i.e. white, babble, hf channel, buccaneer, pink and factory from NOISEX-92 database. The speech is from TIMIT database, which consists of 3 male and 3 female test speeches at sampling frequency $f_s = 8$kHz with the noise under different SNR of -5dB, 0dB, 5dB, 10dB respectively. The speech signal and is divided into frames with the frame length $K=20$ and the frame overlap of 50%. We take $\alpha_s=0.8, \alpha_p=0.2, \delta=5$ and $\alpha_d=0.95$ for the proposed algorithm.
Figure 1. Eigenvalue corresponding to the twentieth eigenvector.

Under the babble noise and the factory noise with SNR=0dB, we provide the adaptive trackings of the eigenvalue corresponding to the twentieth eigenvector of the proposed algorithm in Figure 1. The noisy speech eigenvalue $\lambda_{y20}(n)$ can be obtained by MALASE, where $\lambda_{n20}(n)$ represents the true noise eigenvalue. The noise eigenvalue estimation $\hat{\lambda}_{n20}(n)$ can be obtained by recursive smoothing on $\lambda_{y20}(n)$. From Figure 1, we can see that the noise eigenvalue estimation and true noise eigenvalue is substantially the same, and the proposed algorithm has high estimation precision, which realizes continuous estimation and constantly update for the noise eigenvalues.

Figure 2. $C_{sig}$ of three algorithms.
We select signal distortion ($C_{\text{sig}}$), noise distortion ($C_{\text{bak}}$) and overall quality ($C_{\text{ovl}}$) to evaluate the performance of the speech enhancement algorithm (Hu and Loizou, 2008). The evaluation results are expressed in graded scores, high score means better enhanced speech quality.

$C_{\text{sig}}$ and $C_{\text{bak}}$ of three different algorithms are shown in Figure 2 and Figure 3 respectively. We can see that the proposed algorithm shows an improvement in both $C_{\text{sig}}$ and $C_{\text{bak}}$ score for all noise signals, preserves more speech components, and restrains more background noise. The performance of the proposed algorithm is obviously superior to that of the classical algorithm and MCRA algorithm. In addition, as the SNR decreases, even the performance of the proposed algorithm decreases gradually, but in low SNR (-5dB), the proposed algorithm still performs well.

Table 1 gives the $C_{\text{ovl}}$ score of three different algorithms. From this table, we can see that the overall speech quality of the proposed algorithm is obviously better than that of the classical algorithm and MCRA algorithm. Especially in low SNR (-5dB), the proposed algorithm has more than 0.4~0.6 increased $C_{\text{ovl}}$ score than the classical algorithm, than 0.25~0.4 increased $C_{\text{ovl}}$ score than MCRA algorithm. This means that the subjective listening for the enhanced speech by the proposed algorithm and the clean speech is quite similar in auditory perception.
Table 1 $C_{out}$ of three algorithms

<table>
<thead>
<tr>
<th>Noise type</th>
<th>SNR(dB)</th>
<th>classical</th>
<th>MCRA</th>
<th>proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>-5</td>
<td>1.0627</td>
<td>1.2975</td>
<td>1.5389</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1.6784</td>
<td>1.8898</td>
<td>2.0316</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>2.1471</td>
<td>2.5054</td>
<td>2.5815</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>2.0083</td>
<td>3.0418</td>
<td>3.1324</td>
</tr>
<tr>
<td>Babble</td>
<td>-5</td>
<td>1.2960</td>
<td>1.4469</td>
<td>1.8255</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1.7514</td>
<td>2.0611</td>
<td>2.3565</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>2.1443</td>
<td>2.5060</td>
<td>2.8137</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>2.4241</td>
<td>3.0479</td>
<td>3.2503</td>
</tr>
<tr>
<td>Hfchannel</td>
<td>-5</td>
<td>1.0713</td>
<td>1.3539</td>
<td>1.6112</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1.6795</td>
<td>1.8763</td>
<td>2.0761</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>2.1588</td>
<td>2.5050</td>
<td>2.6187</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>2.3877</td>
<td>3.0804</td>
<td>3.0830</td>
</tr>
<tr>
<td>Buccaneer</td>
<td>-5</td>
<td>1.2757</td>
<td>1.4269</td>
<td>1.6861</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1.7692</td>
<td>1.9357</td>
<td>2.1086</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>2.1647</td>
<td>2.4225</td>
<td>2.7052</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>2.4530</td>
<td>2.9398</td>
<td>3.1247</td>
</tr>
<tr>
<td>Pink</td>
<td>-5</td>
<td>1.2917</td>
<td>1.5584</td>
<td>1.8303</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1.8327</td>
<td>2.0966</td>
<td>2.3826</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>2.2025</td>
<td>2.6613</td>
<td>2.8624</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>2.5097</td>
<td>3.1383</td>
<td>3.2699</td>
</tr>
<tr>
<td>Factory</td>
<td>-5</td>
<td>1.3016</td>
<td>1.3623</td>
<td>1.7632</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1.7150</td>
<td>1.9865</td>
<td>2.2554</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>2.1086</td>
<td>2.5776</td>
<td>2.7258</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>2.4182</td>
<td>3.0500</td>
<td>3.1472</td>
</tr>
</tbody>
</table>

In order to illustrate the effectiveness and the differences between the enhanced speech and the clean speech, time-domain waveforms and spectrograms of three different algorithms under babble noise with SNR=5dB are given in Figure 4. We can see that the proposed algorithm is better than the classical algorithm both in the elimination of background noise and in the retention of speech components. It is more effective in background noise suppression compared with the MCRA algorithm, since it has almost no residual noise.

5. CONCLUSION

A subspace based speech enhancement algorithm using MALASE and noise eigenvalue estimation is proposed without eigen decomposition and voice activity detection. The proposed algorithm can be used for different kinds of noise environment compared with the classical algorithm and MCRA algorithm, and has the advantages of high estimation precision, easy implementation, low voice distortion, small residual noise and good overall quality. It is also suitable for the application of low SNR and non-stationary noise environment.
Figure 4. Time-domain waveforms and spectrograms of three algorithms under babble noise with SNR=5dB.

6. REFERENCES


