TELKOMNIKA, Vol. 11, No. 2, February 2013, pp. 726~735
ISSN: 2302-4046

A Signal Subspace Speech Enhancement Method for Various Noises

Wang Guang Yan*[1], Geng Yan Xiang[1], Zhao Xiao Qun[2]
[1]School of Information Engineering, Tianjin University of Commerce, Tianjin, China
[2]School of Electronics and Information Engineering, Tongji University, Shanghai, China Jinbagong Rd, Beichen Dis., Tianjin City, 300134, 86-22-26667577
*Corresponding author, e-mail: wanggy@tjcu.edu.cn

Abstract

In this paper, we propose a single-channel speech enhancement method in terms of subspace techniques to reduce the noises from speech signals in various noises environment. This subspace approach based on Karhunen-Loève transform and implemented via Principal component analysis. The optimal subspace selection is provided by a minimum description length criterion. An offset factor generated from the white noise was used to modify the variance to adapt to the specified colored noise. Several objective speech quality measures have been introduced to give an overall evaluation of the proposed method. A large amount of data and Figures, as well as the audio quality evaluation results, testify that the algorithm provides high performance for the input signal-to-noise ratio range from -5dB to 10dB. It is showed that the proposed approach have excellent characteristics for colored noises in strong background noise environment with lower signal-to-noise ratio.

Keywords: *Signal subspace, Speech enhancement, Principal component analysis, Minimum description length criterion, Colored noise*

Copyright © 2013 Universitas Ahmad Dahlan. All rights reserved.

1. Introduction

The background noises of our life have been influenced our speech communication and speech processing system for a long time. It degrades the speech quality and intelligibility, and also affects the listener's perception. In order to make voice communication feasible, natural and comforTable, it is desirable to develop speech enhancement method to 'clean' the noisy speech in real environment. Speech enhancement has been a subject of both theoretical interest and practical importance for about 40 years. There are variety speech enhancement methods have been presented in speech communication system, speech recognition system and other speech processing systems. Each method has its advantages and disadvantages. Many objective speech quality measures have been proposed in the past to predict the subjective quality of speech. A signal subspace speech model is a characterization of the speech signal in terms of its subspace information. It has been applied for speech enhancement techniques for the linear white noises. In this paper, we try to extend the subspace method based on PCA/KLT to speech enhancement field for colored noise.

In speech application, the signal subspace approach was originally introduced by Dendrinos [1] who proposed to use SVD method to remove the noise subspace for speech enhancement. The signal subspace approach (SSA) gained more popularity when Ephraim [2] proposed a different technique based on EVD by KLT (Karhunen-Loève transform). KLT is also called PCA (Principal component analysis). PCA subspace approach is one of the technologies of SSA (Signal subspace approach), which is to implement signal subspace decomposition for noisy speech signal by using of a signal/noise KLT. The speech enhancement method based on SSA mostly implifies the PCA/KLT method. In 1995, Ephraim [2] proposed the signal subspace speech enhancement method firstly. Since 2000, the signal subspace speech enhancemethod is more perfected with the development of the PCA/KLT algorithms, from the classical PCA to self-adapting PCA/KLT algorithm [3]. Gradually, the nonlinear PCA and Kernel PCA [4] have been introduced to eliminate the colored noise [5] or the nonstationary wideband noise signal [6-7]. In order to improve the performance of the specified method, the PCA has been incorparated with the human hearing properties [8] and Hilbert-huang transform [9], etc.

Received September 12, 2012; Revised December 27, 2012; Accepted January 13, 2013

In this paper, we propose a subspace approach for single channel speech enhancement in various noisy environment. This approach based on KLT and implemented via PCA. It is expected that the proposed method is more effective than spectral subtraction (SS) in 'musical noise' case. With respect to the colored noise or nolinear noise, some improvement measures will be introduced in this scheme. Several objective speech quality measures, such assegmental SNR (SegSNR), Weighted-slope spectral distance (WSS), Perceptual evaluation of speech quality (PESQ), Log-likelihood ratio (LLR), log spectral distance (LSD) have been used to evaluate the performance of the presented approach.

2. Signal Subspace Speech Enhancement

This scheme exploits the fact that the covariance matrix of a noisy speech signal frame can be decomposed into two mutually orthogonal vector spaces: a signal (+noise) subspace and a noise subspace. Noise reduction is obtained by discarding the noise subspace completely, while modifying the noisy speech components in the signal (+noise) subspace. The signal subspace is corresponding to the principal component, and the noise subspace is corresponding to the minor component. The dimension of the principal component can be determined by some criterions.

2.1. Subspace Decomposition

The model used in the subspace approach assumes that the noisy signal is additive and uncorrelated with the speech signal. The signal picked up by the microphone can be modeled as a superposition of the clean speech and noise. Suppose that the single-channel noisy speech signal $y(n)$ be constructed by the clean speech $s(n)$ and additive noise $d(n)$, that is:

$$y(n) = s(n) + d(n) \tag{1}$$

Let $\hat{x}(n) = H\cdot y(n)$ be a linear estimation of $y(n)$, where *H* is a $K \times K$ matrix. The error signal is given by

$$\varepsilon = \hat{x} - x = Hy - x = \underbrace{(H - I)x}_{\varepsilon_x} + \underbrace{Hd}_{\varepsilon_d} \tag{2}$$

where I is a $K \times K$ unit matrix. The estimation error ε consists of two parts: the speech signal distortion ε_x and the residual noise ε_d. Denoting the signal distortion energy by $\mathrm{tr}\ \mathrm{E}[\varepsilon\varepsilon^{\mathrm{T}}]$. According to the assumption that speech and noise are independent, the cross term of the mean square error is zero.

$$\begin{aligned} \mathrm{tr}\ \mathrm{E}[\varepsilon\varepsilon^{\mathrm{T}}] &= \mathrm{tr}\ \mathrm{E}[\varepsilon_x\varepsilon_x^{\mathrm{T}}] + \mathrm{tr}\ \mathrm{E}[\varepsilon_d\varepsilon_d^{\mathrm{T}}] \\ &= \mathrm{tr}\ (H - I)R_x(H - I)^{\mathrm{T}} + \mathrm{tr}\ HR_dH^{\mathrm{T}} \end{aligned} \tag{3}$$

where $R_x = \mathrm{E}[xx^{\mathrm{T}}]$ and $R_d = \mathrm{E}[dd^{\mathrm{T}}]$ are the covariance matrix of speech and noise signal respectively. Signal subspace approach for speech enhancement is a trade off between speech distortion and residual noise. Its purpose is to reduce the amount of residual noise $\mathrm{tr}\ \mathrm{E}[\varepsilon_d\varepsilon_d^{\mathrm{T}}]$ while reducing the amount of speech distortion $\mathrm{tr}\ \mathrm{E}[\varepsilon_x\varepsilon_x^{\mathrm{T}}]$. This optimization problem can be solved by Lagrange multiplication

$$Hs = R_x(R_x + \mu R_d)^{-1} \tag{4}$$

where $\mu > 0$ is the Lagrange multiplication operator. When $\mu = 1$, Hs includes the Weiner filter Hw. Otherwise, the difference between the subspace method and the Weiner filter is that the covariance matrix of the former is diagonalized by KLT. Under the condition that

$\operatorname{rank}(R_x) = M \le \operatorname{rank}(R_y)$. Using subspace method, the noisy speech signal can be decomposed into the speech subspace which contains a little noise signal and the pure noise subspace. Enhancement is then performed by nulling the components of noise subspace and modifying the components of speech subspace by the gain function $Hs = U\Lambda_x(\Lambda_x + \Lambda_d)^{-1}U^{\mathrm{T}}$. The estimation result is

$$H_{Sub} = U\begin{pmatrix} G_\mu & 0 \\ 0 & 0 \end{pmatrix}U^{\mathrm{T}} \quad (5)$$

where $G_\mu = \Lambda_x'(\Lambda_x' + \mu\sigma_d^2 I)^{-1}U^{\mathrm{T}}$, Λ_x' is a $M \times M$ diagonal matrix witch contains the former M eigenvalues of matrix R_x. I is a $M \times M$ unit matrix, and σ_d^2 is the variance of the white noise. Here, the noise was assumed to be white. For colored noise, it was suggested that the noise could be whitened, or to combine with other enhancement method, such as the human hearing properties [8], is the efficient choice.

2.2. Minimum Description Length Criterion

Provided by the minimum description length (MDL) criterion overcomes the limitations encountered with other selection criteria, like the overestimation of the signal–plus–noise subspace or the need for empirical parameters. MDL criterion is obtained by minimizing the following equations.

$$\mathrm{MDL} = -\log\left(p\left(x(1), x(2), \cdots, x(N) \middle| \hat{\Theta}\right)\right) + \frac{1}{2} n \log N \quad (6)$$

where $p\left(x(1), x(2), \cdots, x(N) \middle| \hat{\Theta}\right)$ is the parameterized probability density function, and $\hat{\Theta}$ is the maximum likelihood estimation of the reference vector Θ, N is the length of mixture signal $x(k)$. n is a variable which is the estimation of the number of the source signal who used to make the MDL value minimum. Although MDL criterion is the second-best estimation of n, it is the most value equation for determining the number of the signal.

3. Performance Evaluation for Speech Enhancement Algorithms

Objective evaluation usually serves as the main protocol during the process of algorithm research and system design. In this paper, several objective speech quality measures were evaluated: SegSNR, WSS, PESQ, LLR, LSD. In order to obtain an overall objective evaluation, a composite evaluation measure which combines some measures with different weighted coefficients has been proposed.

3.1. Segmental SNR (SegSNR) Measure

The frame-based segmental SNR is a reasonable measure of speech quality. It is formed by averaging frame level SNR estimates as follows [10]

$$\mathrm{SegSNR} = \frac{10}{M}\sum_{i=1}^{M}\log_{10}\left[\frac{\sum_{n=0}^{L-1} s_c^2(i,n)}{\sum_{n=0}^{L-1}\left[s_c(i,n) - s_p(i,n)\right]^2}\right] \quad (7)$$

where $s_c(i,n)$ and $s_p(i,n)$ are the ith frame of the original and enhanced speech signal respectively. L is the length of each frame, and M is the number of frames. $s_c(i,n) - s_p(i,n)$ is the i-th noise frame. Frames with SNRs above 35dB do not reflect large perceptual difference. Consequently, we limited the range of SegSNR to -10~35dB.

3.2. Weighted Spectral Slope (WSS) Measure

The WSS measure is a frequency domain measure based on an auditory model. There are 36 overlapping filters of progressively larger bandwidth were used to estimate the smoothed short-time speech spectrum. The measure finds a weighted difference between the spectral slopes in each band. The jth frame measure in decibels is defined as [10]

$$d_{\mathrm{WSS}}(j) = K_{spl}(K - \hat{K}) + \sum_{k=1}^{36} w_a(k)(S(k) - \hat{S}(k))^2 \tag{8}$$

where K and $\hat{K}$ are related to overall sound pressure level of the original and enhanced utterances, and K_{spl} is a parameter which can be varied to increase overall performance. Generally, let $K = 25$, and length of frame is 256.

3.3. The Log-Likelihood Ratio (LLR)

LLR measure is one of the LPC-based objective measures, and it is defined as [11]

$$d_{LLR}(\vec{a}_p, \vec{a}_c) = \log\left(\frac{\vec{a}_p \Re_c \vec{a}_p^{\mathrm{T}}}{\vec{a}_c \Re_c \vec{a}_c^{\mathrm{T}}}\right) \tag{9}$$

where $\vec{a}_c$ and $\vec{a}_p$ are the LPC vector of the original speech frame and the enhanced speech frame respectively. $\Re_c$ is the autocorrelation matrix of the original speech signal. LLR measure mainly concern on the similarity of spectral envelope, and not consider of the displacement produced by the model gain.

3.4. Log Spectral Distance (LSD)

The LSD measure is defined as follows [12]

$$\mathrm{LSD} = \frac{1}{L}\sum_{l=1}^{L}\sqrt{\frac{1}{K/2+1}\sum_{k=0}^{K/2}(10\log\left|\hat{S}(k)\right| - 10\log\left|S(k)\right|)^2} \tag{10}$$

where $S_c(k)$ and $S_p(k)$ are the log-spectrum of the original speech frame and the enhanced speech frame respectively, which based on K-point DFT calculation. L is the number of frames. LSD measure is a measure of frequency domain. The smaller the LSD is, the closer the shape of log-spectrum of the clean speech.

3.5. Perceptual Evaluation of Speech Quality (PESQ)

PESQ was selected in May 2000 as draft ITU-T recommendation P.862, and are aggregated in frequency and time and mapped to a prediction of subjective mean opinion score (MOS). In PESQ the original and degraded signals are mapped onto an internal representation using a perceptual model. The difference in this representation is used by a cognitive model to predict the perceived speech quality of the degraded signal. This perceived listening quality is expressed in terms of Mean Opinion Score, an average quality score over a large set of subjects.Most of the subjective experiments used in the development of PESQ used the ACR (Absolute Category Rating) opinion scale of Table 1, and the block diagram and calculation procedure are as shown in references [13-14].

Table 1. ACR Listening Quality Opinion Scale Used in The Development of PESQ

Quality of the speech	Score
Excellent	5
Good	4
Fair	3
Poor	2
Bad	1

3.6. Composite Measure

Composite measures, named as Cov measure, obtained by combining a subset of the above measures were also evaluated. The Cov measure is the overall planning and combination of the evaluation measures in time domain, frequency domain and perceptual field. The Cov measure is defined as follows

$$\mathrm{Cov} = 1.594 + 0.805 \times \mathrm{PESQ} - 0.512 \times \mathrm{LLR} - 0.007 \times \mathrm{WSS} \tag{11}$$

4. Algorithm Implementation

The proposed principle component subspace scheme is a speech enhancement method in time domain. The input noisy speech signal is processed by voice activity detection (VAD), noise estimation and frame classification by window signal firstly. Considering the Gaussian White noise and Colored noise respectively, the signal-plus-noise and noise subspace can be divided by use of MDL criterion frame by frame. Because the residual noise in the signal-plus-noise frame is less than the threshold setting beforehand, it is suggested that the signal-plus-noise can be considerate as the signal subspace, followed by the KLT operation for the speech frame and noise frame respectively. Figure 1 shows a system block diagram for the proposed speech enhancement algorithm. There are some key points of this scheme.

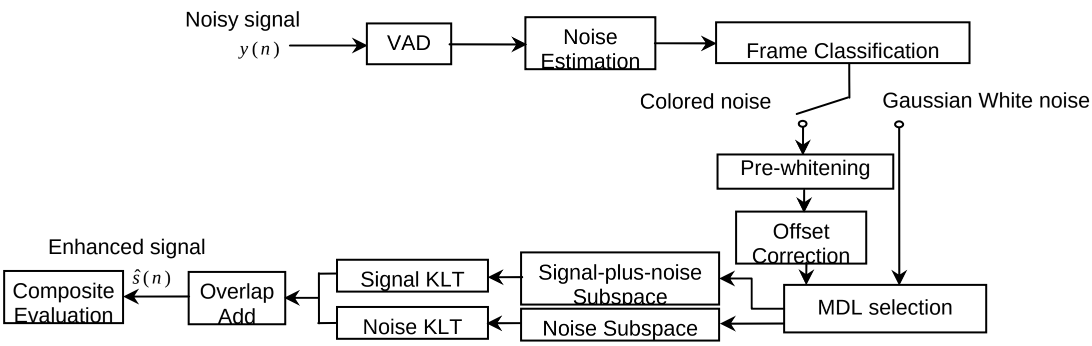


Figure 1. System Block Diagram for The Proposed Speech Enhancement Algorithm

4.1. Subspace dimensions selection

Theoretically, the dimension of subspace is defined by the order of the linear signal. Actually, contents of each phonetic segment are different frame to frame. Sometimes, the dimension can be determined by the number of positive eigenvalues of the covariance matrix. In this scheme, the sampling frequency of the original speech signal is 16kHz, the dimension of subspace is set as $p \approx 12$ according to the MDL criterion.

4.2. Frame length selection

In order to separate the noise signal as thorough as possible, it is necessary to make the frame length N is greater than the matrix dimension M , that is $N \geq M$. Limited by the short time stationarity of speech signal, the frame length N is generally about 20-30ms. At the same time, the longer of the frame, the lower computational complexity of the KLT algorithm. In this scheme, the sampling frequencies of speech signals from corpus data are generally 16kHz. Consequently, we select the frame length N for 320-480 sampling points.

4.3. Solving for the Colored Noise

Because the autocorrelation matrix of colored noise signal can not be diagonalized, the Singular Value Decomposition (SVD) of it would be invalid. According to this situation, an improved method is proposed. After pre-whitening, the SVD noise cancelling method for white noise can be applied to colored noise by correcting some factors. Supposing that the Cholesky factor R of the noise signal is known, it is pre-whitened by multiplied the factor R^{-1} on both sides of the equation:

$$H_y R^{-1} = H_s R^{-1} + H_d R^{-1} \quad (12)$$

During the process of signal and noise KLT, the autocorrelation matrix of the estimated clean speech signal is obtained by subtracting the estimated noise signal ($\hat{R}_d$) from the noisy speech signal ($\hat{R}_y$)

$$R_s = R_y - R_d \quad (13)$$

$$\begin{aligned} R_s &= V \Lambda V^{\mathrm{T}} \\ R_d &= V(\sigma_w^2 I) V^{\mathrm{T}} \\ R_y &= V(\Lambda + \sigma_w^2 I) V^{\mathrm{T}} \end{aligned} \quad (14)$$

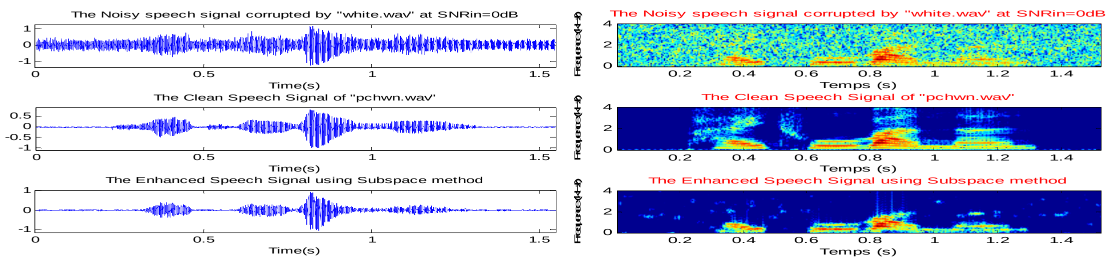
Hankel matrix of the noise signal H_d and principle eigenvalue of signal subspace R_s can be represented by the following diagonal matrix elements

$$\sum\nolimits_{c,\mathrm{proj}}^{2} = \mathrm{diag}\left\{ V^{\mathrm{T}} R_d V \right\} \quad (15)$$

Substitute the noise variance σ_w^2 of Equation 14 by elements of $\sum\nolimits_{c,\mathrm{proj}}^{2}$, σ_w^2 is no more than a constant, but a offset on the projection of each signal subspace. By use of this offset to modify the variance σ_w^2 of the white noise, we can get the noise variance of the specified colored noise.

5. Simulation Results and Analysis

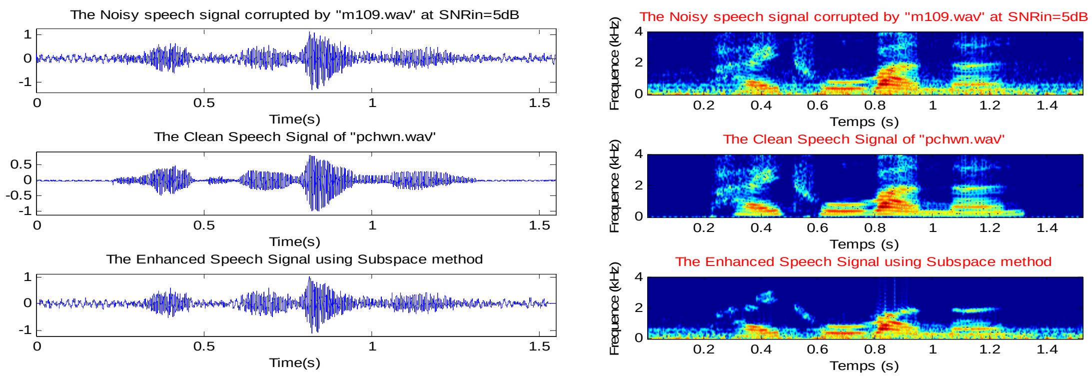
An utterance of Chinese female speech which comes from the CASIA corpus data called “pai chu wan nan” is selected as the experimental pure speech. Different noise types taken from the NOISEX-92 database are used to generate the experimental noisy speech signals at different signal-to-noise ratio, such as -5dB, 5dB, 0dB, 10dB, 15dB. Figure 3 shows the waveform comparison and spectrogram comparison among the noisy speech, the clean speech and enhance speech signal under the Gaussian white noise at SNR=0dB, and Figure. 2 shows the same case under the m109 noise at SNR=5dB. Under the Gaussian white noise environment, we executed the objective speech quality evaluation for the noisy speech before enhancing and the estimated speech signal after enhancing respectively. In order to evaluate the performance of the proposed method effectively, we compared the performance of the proposed PCA subspace speech enhancement method with the classical SS method at the concrete evaluation index. The results is as shown in Table 2~4. Table 5~8 show the evaluation results of the estimated speech signal based on the proposed subspace enhancement method under the hfChannel noise, pink noise, buccaneer2 noise and m109 noise respectively.



(a) Waveforms comparison (b) Spectrogram comparison

Figure 2. Comparison of the Waveforms and Spectrogram f The Noisy Speech Signal Contaminated by White noise of TIMIT Noise Group at SNRin=0dB

It can be seen from Figure 2 that the proposed method is more effective for white noise even at lower SNR ratio. Otherwise, for the colored noise, such as m109 noise, factory1 and factory2 noise etc., the enhancement performance is not as effective as the expected results. Comparing the first and third subplot of Figure 3. (a), we can find that the waveforms have little changed after enhancing. According to the third subplot of Figure3. (b), it can be found that the high-frequency information of the original speech signal has been destroyed, but the most frequency information of noise signal has been reserved during the enhancement process.



(a) Waveforms comparison (b) Spectrogram comparison

Figure 3. Comparison of The Waveforms and Spectrogram of The Noisy Speech Signal Contaminated by m109 Noise of TIMIT Noise Group at SNRin=5dB

On the basis of calculation of tested data from Table 2~4, the comparison of the output SegSNR according to different methods at different SNR is as shown in Figure 4. It is indicated that both of the two methods have elevated the SegSNR index to a great extent. When the input SNR is less than 10dB, the proposed subspace method is more effective than the classical spectral subtraction method. However, when the input SNR is greater than 10dB, the spectral subtraction method is more effective than the proposed subspace method. The same situation is appeared for the pink noise, m109 noise and other colored noise according to Table 5~8 and Figure 5. The mainly reason for this case is that the proposed method will destroy the waveform or frequency information of the original speech signal. If the degree of the damage is greater than the degree of the de-noising, the total performance of the proposed method will be degrading. In a specific range of input SNR as -5~10dB, de-noising is the main function which lead to the excellent performance of the proposed method. A large number of experimental data, as well as the waveforms and spectrograms, show that the proposed speech enhancement is effective for the stationary noise as white noise, hfChannel noise and pink noise under low input SNR environment, as is shown in Figure 5. According to Table 4~8, assigned that the input SNR as 10dB, we can get the output subjective measures for performance evaluation with different measures for different noises, as shown in Figure 6. It is indicated that the proposed subspace enhancement method has better performance in time domain (SegSNR) for White noise. Meanwhile, for pink noise, it has a higher MOS scores in perceptual field.

Table 2. Evaluation Results of The Noisy Speech Signal (White Noise)

SNRin	SegSNR	LSD	LLR	WSS	PESQ	Cov
-5	-6.48331	2.54506	1.65619	71.67568	1.51276	1.46207
0	-4.03184	2.37255	1.42028	55.35715	1.72244	1.86588
5	-0.90719	2.08372	1.12739	41.56679	2.0282	2.35851
10	2.58676	1.73696	0.82709	30.11072	2.3999	2.89167
15	6.31889	1.38333	0.55251	21.25395	2.80822	3.42296

Table 3. Evaluation Results of The Estimated Signal Using SS Method (White Noise)

SNRin	SegSNR	LSD	LLR	WSS	PESQ	Cov
-5	-1.87136	1.99476	1.51715	88.17905	1.8327	1.67528
0	-0.05099	1.70415	1.23656	73.29963	2.09662	2.13556
5	2.20482	1.45449	0.93207	60.47747	2.48065	2.69036
10	4.47241	1.3073	0.83469	49.6791	2.79069	3.06539
15	10.07635	1.10117	0.50315	40.18064	3.54519	3.909

Table 4. Evaluation Results of The Estimated Signal Using Subspace Method (White Noise)

SNRin	SegSNR	LSD	LLR	WSS	PESQ	Cov
-5	-1.09731	1.61808	1.20763	93.78725	1.70001	1.68769
0	1.59715	1.50144	1.04735	72.91486	2.09143	2.23095
5	4.24163	1.45879	0.97516	54.96773	2.50776	2.72869
10	6.38955	1.43406	0.95688	41.3377	2.82947	3.09243
15	8.10153	1.42962	0.72075	32.4766	3.12116	3.51017
20	9.62428	1.3936	0.65708	26.16233	3.21584	3.66319

Table 5. Evaluation Results of The Estimated Signal Using Subspace Method (hfchannel Noise)

SNRin	SegSNR	LSD	LLR	WSS	PESQ	Cov
-5	-3.06807	2.04017	1.87502	91.38266	1.74005	1.39505
0	-0.53592	1.79436	1.59591	70.11109	2.02732	1.91811
5	1.83065	1.58216	1.39221	53.43882	2.41089	2.44788
10	4.45355	1.43162	1.29309	42.23599	2.75979	2.85792
15	7.12278	1.32665	1.01963	32.97625	3.0208	3.27286
20	9.27166	1.29535	0.91973	28.13602	3.16784	3.47626

Table 6. Evaluation Results of The Estimated Signal Using Subspace Method (Pink Noise)

SNRin	SegSNR	LSD	LLR	WSS	PESQ	Cov
-5	-6.03205	1.75542	0.92152	86.63887	1.74249	1.91841
0	-3.22656	1.64371	0.84061	69.45695	2.09642	2.36503
5	-0.33538	1.4802	0.79172	52.78236	2.4561	2.79632
10	2.19048	1.35415	0.79319	39.51864	2.8531	3.208
15	4.73762	1.29616	0.79521	31.3369	3.1643	3.51476
20	7.17693	1.31097	0.75469	26.50992	3.23864	3.62913

Table 7. Evaluation Results of The Estimated Signal Using Subspace Method (buccaneer2)

SNRin	SegSNR	LSD	LLR	WSS	PESQ	Cov
-5	-4.25267	1.81267	0.98239	97.72877	1.6027	1.69709
0	-1.88987	1.68208	0.84436	78.60153	1.91064	2.14954
5	0.51209	1.50695	0.7486	61.82256	2.28054	2.61379
10	2.64165	1.36817	0.71244	47.04776	2.72802	3.09595
15	5.11809	1.2969	0.7009	37.28226	3.05802	3.43587
20	7.40704	1.2954	0.69846	29.5876	3.05533	3.48882

Table 8. Evaluation Results of The Estimated Signal Using Subspace Method (m109 Noise)

SNRin	SegSNR	LSD	LLR	WSS	PESQ	Cov
-5	-5.87967	1.73678	1.06396	86.47256	1.88629	1.9624
0	-3.53906	1.61875	1.02014	70.02595	2.17827	2.33502
5	-0.948	1.50393	0.97184	54.92194	2.50209	2.72615
10	1.47779	1.40015	0.91576	40.34525	2.82459	3.11651
15	3.86937	1.35873	0.86293	30.51063	3.15808	3.48086
20	6.28076	1.36718	0.80818	25.66788	3.20281	3.5788

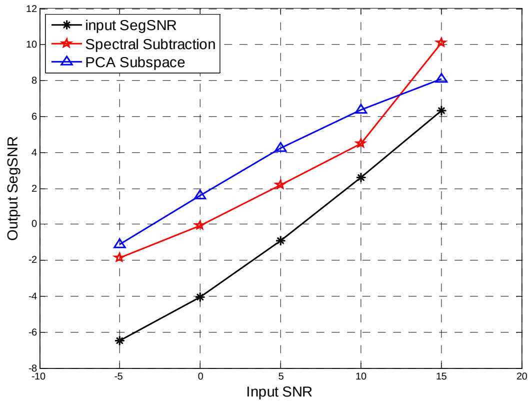


Figure 4. Comparison of The Output Segsnr According to Different Methods

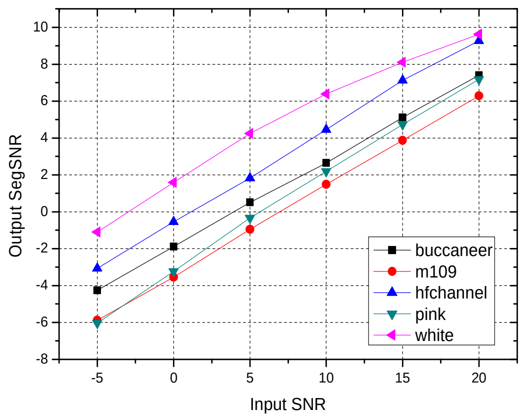


Figure 5. Comparison of The Output Segsnr According to The Background Noise

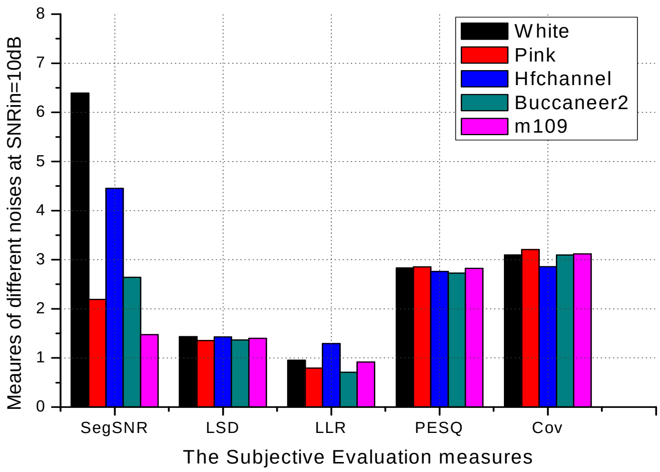


Figure 6. Comparison of Subjective Evaluation Measures of the Proposed Method for Various Noises at 10dB input SNR

6. Conclusion

Speech enhancement is a very difficult problem and still remains a challenge today even after 40 years of tremendous progress. There are many speech enhancement methods to eliminate the noise successfully from the noisy speech signal in relatively high input SNR. While for the lower input SNR cases or in strong background noise, the performance of some classical speech enhancement algorithms are not as effective as in high input SNR. There are many objective and subjective assessment indexes have been adopted to evaluate the performance of the specified algorithm. In this paper, we researched mainly on the speech enhancement algorithm based on PCA subspace technology. Comparing with the classical spectral subtraction method, the experimental results show that the proposed PCA method have potential application and effectiveness in strong background noise with input SNR lower than 10dB. Simulation results show that our proposed scheme outperforms the perfect spectral subtraction method in terms of segmental signal-to-noise ratio (SegSNR), Weighted-slope spectral distance (WSS), Perceptual evaluation of speech quality (PESQ), Log-likelihood ratio (LLR), log spectral distance (LSD). Otherwise, with respect to the high input SNR greater than 10dB, the proposed method will destroy the time-frequency characteristic of the original speech signal to make the performance degraded. The further research orientation will concentrate on the nonlinear revision and the combination with other methods.

References

[1] Dendrinos M, S Bakamidis, G Carayannis. Speech enhancement from noise: a regenerative approach. *Speech Communication*. 1991; 10(1): 45-57.
[2] Ephraim Y, HL Van Trees. A Signal Subspace Approach For Speech Enhancement. *Speech and Audio Processing, IEEE Transactions on*. 1995; 3(4): 251-266.
[3] Rezayee A, S Gazor. An adaptive KLT Approach For Speech Enhancement. *Speech and Audio Processing, IEEE Transactions on*. 2001; 9(2): 87-95.
[4] Takiguchi T, Y Ariki. PCA-based speech enhancement for distorted speech recognition. *Journal of Multimedia*. 2007; 2(5): 13-18.
[5] Lev-Ari H, Y Ephraim. Extension Of The Signal Subspace Speech Enhancement Approach To Colored Noise. *IEEE Signal Processing Letters*. 2003; 10(4): 104-106.
[6] Lou Hong Wei, Guang Rui Hu. An Approach Based On Simplified Klt And Wavelet Transform For Enhancing Speech Degraded By Non-Stationary Wideband Noise. *Journal of Sound and Vibration*. 2003; 268(4): 717-729.
[7] Chiung-Wen Li, Lei Sheau-Fang. *Signal Subspace Approach For Speech Enhancement In Nonstationary Noises*. Communications and Information Technologies ISCIT '07. International Symposium on. 2007: 1580-1585.
[8] Jabloun F, B Champagne. Incorporating the human hearing properties in the signal subspace approach for speech enhancement. *IEEE Transactions on Speech and Audio Processing*. 2003; 11(6): 700-708.
[9] Khademul Islam Molla, Hirose Keikichi. Single-Mixture Audio Source Separation by Subspace Decomposition of Hilbert Spectrum. *IEEE Transactions on Audio, Speech, and Language Processing*. 2007; 15(3): 893-900.
[10] Hansen Jhl, Bl Pellom. *An Effective Quality Evaluation Protocol For Speech Enhancement Algorithms*.ICSLP-98: International Conference on Spoken Language Processing. Sydney, Australia. 1998: 2819-2822.
[11] Yi Hu, PC Loizou. Evaluation of Objective Quality Measures for Speech Enhancement. *IEEE Transactions on Audio, Speech, and Language Processing*. 2008; 16(1): 229-238.
[12] Jounghoon Beh, RH Baran, Ko Hanseok. Dual channel based speech enhancement using novelty filter for robust speech recognition in automobile environment. *IEEE Transactions on Consumer Electronics.* 2006; 52(2): 583-589.
[13] Rix Aw, Mp Hollier, Ap Hekstra, Jg Beerends. Perceptual Evaluation of Speech Quality (PESQ): The New ITU Standard for End-to-End Speech Quality Assessment Part I-Time-Delay Compensation. *Journal of the Audio Engineering Society*. 2002; 50(10): 755-764.
[14] Beerends Jg, Ap Hekstra, Aw Rix, Mp Hollier. Perceptual evaluation of speech quality(PESQ) the new itu standard for end-to-end speech quality assessment. Part II: Psychoacoustic model. *Journal of the Audio Engineering Society*. 2002; 50(10): 765-778.