Speech Enhancement Algorithm Based on Combining Local Characteristic-Scale Decomposition and Difference Spectrum of Singular Values

Nan Li, Shu Deng and Quanbin Tang
College of Information Engineering, Northeast Dianli University, Jilin, P. R. China
whz101111@163.com

***Abstract*—De-noising of speech signal polluted by background noise, which is of great practical significance for effective transmission and accurate recognition of sound, is important. Local characteristic-scale decomposition is introduced, and it can divide the speech into low- and high-frequency parts without a loss in useful speech. The algorithm accelerates the speed of convergence, as well as pretreatment, and has less illusive component. In addition, two different criteria are proposed to select the reasonable noise reduction order based on the difference spectra of singular values. The proposed approach has improved the problem of noise reduction order selected by experience in singular value decomposition and enhanced de-noising effects. Simulation experiments verify the validity of the algorithm from subjective and objective evaluations.**

***Index Terms*—Local Characteristic-Scale Decomposition, Singular Value Decomposition, Difference Spectrum, Speech Enhancement**

I. INTRODUCTION

In the real communication environment, speech signals are inevitably subject to various interferences from transmission medium, communication equipment, and other speakers. The received speech influenced by interference becomes a noise-contaminated signal. Sometimes, in an extremely noisy harsh environment, speech information is covered, resulting in raw information that can hardly be recovered. Consequently, de-noising of noisy speech signals is important.

Speech enhancement algorithm is currently categorized into several transform domain methods, including time and frequency domains. The time domain methods, such as those based on parameters and model [1-3] and the subspace method [4, 5], can accurately estimate a speech model from noisy speech and obtain enhanced speech. However, estimating an accurate speech model with low signal-to-noise ratio (SNR) is difficult and requires a highly complex algorithm. The frequency domain methods, such as spectral subtraction [6] and minimum mean-squared error estimation [7], can estimate the original speech from noisy speech by using the strong relativity of speech short-time spectrum and weak relativity of noise. However, it mostly trends to produce musical noise so that perception results become poor. Therefore, researchers gradually concentrate on other transform domain methods. At present, the time-frequency domain method is the most frequently studied process.

The time-frequency domain methods in speech enhancement are mainly wavelet method [8] and empirical mode decomposition (EMD) [9-13]. However, a wavelet method lacks adaptability; a suitable wavelet basis and threshold are thus selected before using the wavelet method. Although EMD can overcome this drawback, this method must combine with the threshold at de-noising. Then, EMD produces a mass of illusive components and the problem of marginal effect when signals are decomposed by EMD. The local characteristic-scale decomposition (LCD) [14] is presented by Cheng Jun-sheng et al. to remedy these problems. This algorithm improves marginal effect and exhibits less illusive components in addition to fast calculation speed. The present study utilizes LCD to decompose speech signals and proposes two different criteria to reduce noise from noisy speech signals by analyzing the difference spectrum of singular values. This algorithm can remedy the selection of threshold and provide results with improved SNR and increased intelligibility.

In order to solve the problem, that it is difficult to select thresholds and unsatisfactory de-noising effect, a new algorithm is presented combining local characteristic scale decomposition algorithm and singular value decomposition (SVD). The singular value decomposition (SVD) algorithm is an efficient nonlinear filtering method, and it can decompose the signal into a series of singular value and its vector corresponding to time-frequency subspace. In the application of singular value decomposition, selection of effective singular value number is a difficulty. Selection criteria of effective singular values are proposed through analyzing the LCD characteristics of speech signal and difference spectrum characteristics of singular value. The speech signal is decomposed using LCD algorithm as different frequency signals, and the difference spectrum of singular value of these decomposition are analyzed. According to the ratio

doi:10.4304/jmm.8.3.233-240

of noise and useful speech signals after local characteristic scale decomposition, the effective singular values are determined by different criteria. When frequency signals contained mostly noise, the effective singular values numbers is determined by maximum criterion of difference spectrum and when frequency signals contained mainly speech signals, the effective singular values numbers is determined by terminal for zero criterion of difference spectra. This algorithm can remedy the selection of threshold and provide results with improved SNR and increased intelligibility.

II. BASIC THEORY

A. Local Characteristic-scale Decomposition

Based on the analysis of the frequency domain of a speech signal, the frequency of voice sound primarily focuses on the low band, whereas that of voiceless sound distributes in the high band. The noise signals mainly concentrate in the high frequency, and a small amount of noise information concentrates in the low band of speech. Hence, the low- and high-frequency components require different methods to reduce noise to achieve better sound effects.

Local characteristic-scale decomposition depends on the signal itself to divide signals. This process can obtain n-order intrinsic scale component (ISC) from high frequency to low frequency. An ISC must satisfy the following conditions. First, in the whole signal, the maximum value is positive, minimum value is negative, and any two adjacent maxima and minima must be monotonous. Second, a straight-line function can be obtained by any two adjacent maxima and minima. This condition can also determine the function value in the time between the two extreme points, so the ratio of function value and this extreme must be a constant.

LCD is performed as follows [15]:

1) All the extreme X_k of the whole signal and all corresponding time τ_k, $k=1,2,\cdots M$, where M is the number of extreme, are identified;

2) Straight-line function is determined by two adjacent maxima (minima) (τ_k, X_k) and (τ_{k+2}, X_{k+2}). Following the straight-line function, the function value A_{k+1} is obtained in the time τ_{k+1}, $k=1,3,\cdots M$;

3) Based on the formula $L_k = aA_k + (1-a)X_k$ ($k=2,3,\cdots M-1$), L_k is calculated, and the edges extreme (τ_0, X_0) and (τ_{M+1}, X_{M+1}) can be obtained by mirror symmetry continuation method proposed by Grilling. Then, A_1 and A_M are calculated to finally obtain L_1 and L_M;

4) $L_1, L_2, \cdots L_M$ are connected with cubic spline lines as the mean value function m_1;

5) The first signal component can be calculated as $y(t) - m_1 = c_1$. Ideally, if c_1 satisfies the two conditions of ISCs, then it should be an ISC, as h_1. If c_1 does not satisfy the two conditions of ISCs, then c_1 is treated as the new data. This sifting procedure is repeated until the difference can satisfy the two conditions of ISCs, so that it can be treated as the first ISC component.

6) h_1 is separated from the rest of the signal by $y(t) - h_1 = r_1$. r_1 is treated as a new signal $y(t)$ and subjected to the same sifting process as described above, resulting in the second ISC h_2, the third ISC h_3,….

The termination standard of this algorithm is the standard deviation (SD) calculated as $SD = \sum_{t=0}^{T}\left[\frac{\left|h_{ik}(t) - h_{i(k-1)}(t)\right|^2}{h_{i(k-1)}^2(t)}\right]$, where T is the length of time. To ensure an ideal ISC, SD should be less than 0.5.

Using LCD, noisy speech signals can be decomposed as

$$y(t) = \sum_{i=1}^{n} h_i + r_n \tag{1}$$

where h_i is the i^{th}-order ISC, which represents the different frequency band signal components in noisy speech. The residue r_n is a monotonic function, which reflects the average tendency of original signals.

B. Difference Spectra of Singular Values

Singular value decomposition (SVD) is a traditional de-noising method. First, a Hankel matrix is created from a noisy speech signal. Second, singular values are obtained by SVD, and the significant singular values to reduce noise are determined. Finally, enhanced speech is reconstructed from the useful singular values.

For a real matrix $A \in R^{m\times n}$, its SVD is defined as follows [16]:

$$A = UDV^T \tag{2}$$

where U and V are the orthogonal matrices $U \in R^{m\times m}$ and $V \in R^{m\times m}$, D is the diagonal matrix. $D = [diag(\sigma_1, \sigma_2, \cdots \sigma_q), 0]$ $\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_q > 0, \sigma_i (i=1,2,\cdots q)$ are called the singular values of matrix A. q is the rank of matrix A, usually $q \le \min(m,n)$.

The properties of the singular values are given in Ref. [17]

$$\sigma_t(A+B) \le \sigma_t(A) + \sigma_t(B), 1 \le t \le q \tag{3}$$

Supposing that $y(t)$ is the noisy speech, $s(t)$ is the pure speech, $n(t)$ is the noise signal, then A_y, A_s, and A_n are the Hankel matrices created from $y(t)$, $s(t)$, and $n(t)$.

The signal is decomposed by SVD, and the following inequality is obtained:

$$\sigma(A_y) \le \sigma(A_s) + \sigma(A_n) \tag{4}$$

$$\sigma(A_y) \le (\sigma_{s1} + \sigma_n, \cdots \sigma_{sk} + \sigma_n, \sigma_n, \sigma_n \cdots \sigma_n) \qquad (5)$$

Based on expression (5), to extract the original speech signal $s(t)$, the key problem is to determine the sudden change point k in the singular values of the noisy speech signal. Then, the enhanced speech is reconstructed by the front k singular values. To describe its sudden change status, the concept of difference spectra is introduced. The forward difference of singular values is defined as follows: $b_i = \sigma_i - \sigma_{i-1}, i = 1, 2, \cdots, q-1$, then the sequence $B = (b_1, b_2, \cdots, b_{q-1})$ is called the difference spectrum of singular value, which reasonably describes the change status between adjacent singular values.

The simulation signal is $f(t) = \sin(2t) + \sin(6t)$, sampling 512 points in the interval [0, 2π], and the added noise follows Gaussian distribution $N(0,1)$. The waveform comparison charts of the input signal with noise and original signal are shown as Figure 1.

Difference spectrum sequence of the signal with noise shows in Figure 2(a) is analyzed. The maximum of sudden change point appeared in the four order, so pure signal is restored by the first four orders. The final de-noising signal is shown in Figure 2(b), which is restored by difference spectrum maximum.

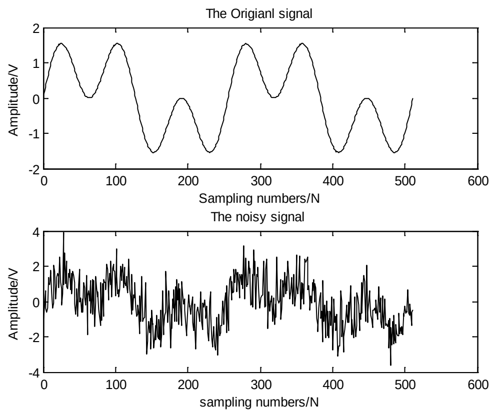


Figure 1. The original signal and noise signal time domain waveform

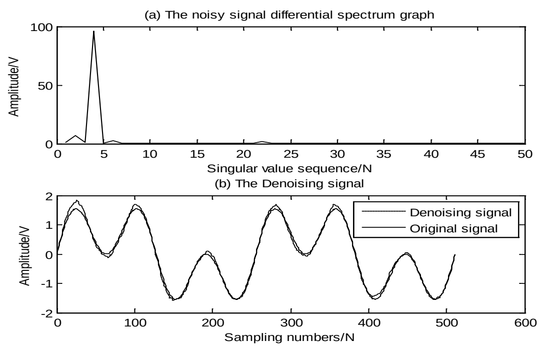


Figure 2. The noisy signal differential spectra and signal, and Time-domain waveform after singular value de-noising

C. LCD Characteristics of the Speech Signal with Noise

The noisy speech signal is decomposed by local characteristic-scale decomposition, and then it can obtain a lot of ISCs. Analyzing with the correlation coefficient of each ISC, the noise signal and the clean speech signal, it can determine the relationship between noise and speech signals in the ISCs.

The correlation coefficient is defined as

$$\rho_{XY} = \frac{\sum_{i=1}^{N}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{N}(X_i - \overline{X})^2}\sqrt{\sum_{i=1}^{N}(Y_i - \overline{Y})^2}} \qquad (6)$$

ISC values of voice “one” and “ten” with noise are showed in Table I and ISC values of pure signal are showed in Table II (The top eight ISCs are listed in tables) .Known from the Table I and Table II, the top two ISCs contain the large amount of noise components, because the similarity of ISC1, ISC2 and noise signal is the largest. However the similarity of component signal behind ISC3 and noise signal is diminishing, the similarity of component signal behind ISC3 and clean speech signal is increasing. The similarity is usually greater than 0.5.

TABLE I.
THE CORRELATION COEFFICIENT OF ISCS AND NOISE SIGNALS

SNR=0 dB	ISC1	ISC2	ISC3	ISC4	ISC5	ISC6	ISC7	ISC8
One	0.67	0.40	0.28	0.21	0.13	0.04	0.04	0.02
Two	0.67	0.40	0.31	0.13	0.09	0.08	0.04	0.05
Three	0.67	0.42	0.30	0.17	0.09	0.04	0.05	0.05
Four	0.67	0.43	0.27	0.13	0.07	0.07	0.05	0.05
Five	0.67	0.36	0.22	0.17	0.13	0.07	0.04	0.05
Six	0.69	0.42	0.28	0.15	0.09	0.02	0.04	0.04
Seven	0.67	0.43	0.26	0.17	0.09	0.05	0.04	0.05
Eight	0.69	0.42	0.24	0.11	0.10	0.07	0.05	0.05
Night	0.67	0.36	0.30	0.23	0.12	0.05	0.02	0.03
Ten	0.71	0.41	0.30	0.18	0.09	0.04	0.04	0.04

TABLE II.
THE CORRELATION COEFFICIENT OF ISCS AND CLEAN SPEECH SIGNAL

SNR=0 dB	ISC1	ISC2	ISC3	ISC4	ISC5	ISC6	ISC7	ISC8
One	0.01	0.12	0.13	0.23	0.33	0.66	0.51	0.10
Two	0.03	0.02	0.07	0.60	0.55	0.36	0.26	0.03
Three	0.01	0.04	0.10	0.33	0.56	0.52	0.39	0.10
Four	0.01	0.06	0.27	0.50	0.52	0.32	0.31	0.14
Five	0.03	0.19	0.34	0.25	0.29	0.44	0.46	0.23
Six	0.01	0.03	0.11	0.39	0.46	0.58	0.45	0.06
Seven	0.01	0.04	0.21	0.36	0.38	0.59	0.42	0.16
Eight	0.01	0.02	0.08	0.51	0.63	0.41	0.34	0.05
Night	0.02	0.16	0.14	0.13	0.28	0.64	0.50	0.16
Ten	0.02	0.03	0.07	0.33	0.55	0.63	0.28	0.05

D. Selection Criteria of Effective Singular Values

The results of SVD largely depend on the reasonable selection of noise reduction order. In signal processing, if too many singular values are selected and some noise will mix into the final signal, then it cannot desirably reduce noise. However, if too few singular values are selected, some useful signals will be lost and will result in waveform distortion of the signal. Thus, the selection of an effective singular value order is crucial. In Ref.[18], a method called singular entropy is presented to attempt to solve this concern. However, no obvious feature is found to determine the effective singular value, because the shape of a singular entropy sequence is similar to an inverse singular value sequence. The selection of an effective singular value based on user experience appears to be more frequently used than by singular entropy [19].

Thus, following the characteristic of difference spectra, two different criteria are proposed to select the reasonable noise reduction order:

1) Maximum criterion

If the position of the maximum k exists in the sequence of the difference spectra, then this point is the sudden change point. The maximum sudden change point represents the border between the singular value of the useful signal and noise. To determine the sudden change point, the singular values of noise can be reasonably detected, and then it can reduce noise.

2) Terminal for zero criterion

Given the equality of the singular values of noise, zeros exist in the sequence of the difference spectra. The noise reduction order is determined by searching for the first zero in the terminal of the difference spectrum [20].

III. Algorithm Implementation

As introduced in the basic theory of algorithms in section B, the implementation of the proposed algorithm will be discussed in this section. It mainly includes the signal pretreatment, selection of noise reduction order, and reconstruction of speech signal.

A. Signal Pretreatment

Signal pretreatment is an important step prior to implementation, and it contains the signal frame process, frequency division, and elimination of false components.

Sub-frame processing divides the signal, in which the long speech signal is truncated, improving the speed of operation. The unvoiced and voiced frames are obtained after this process, so that it effectively extracts the noise, unvoiced, and voiced data during frequency division. Then, LCD is used to divide the frame signal to produce a small amount of false ISCs. If the correlation coefficient is greater than 0.2, then it is considered as an effective ISC. Otherwise, the value is discarded. The time-domain waves of the pure and noisy speech frames after LCD are illustrated in Figure 3.

B. Selection of Noise Reduction Order

The effective ISCs remove noise by using SVD. The first and second order ISCs select the maximum criterion of the difference spectrum to determine the noise reduction order, while the other ISCs select the terminal for zero criterion.

The frequency spectrum of the ISC shown in Figure 4 is analyzed. Noise component is mainly found in the low-level ISCs, and a few useful speech signals are found. The high-level ISCs are mainly speech signals. However, a small amount of noise exists in this level. Therefore, different frequencies of ISCs must select different order of noise reduction using SVD.

Simulation experiment shows that the low-input SNR speech frame signal can remove more noise components and preserve most of the speech components in the maximum criterion of the difference spectrum when the input SNR of the speech frame is higher. These phenomena indicate that the content is mostly speech signal, and the maxima of the difference spectra cannot represent the border between the singular values of the useful signal and noise, which can influence the selection of order. Therefore, these signals should effectively remove noise by choosing the terminal for zero criterion.

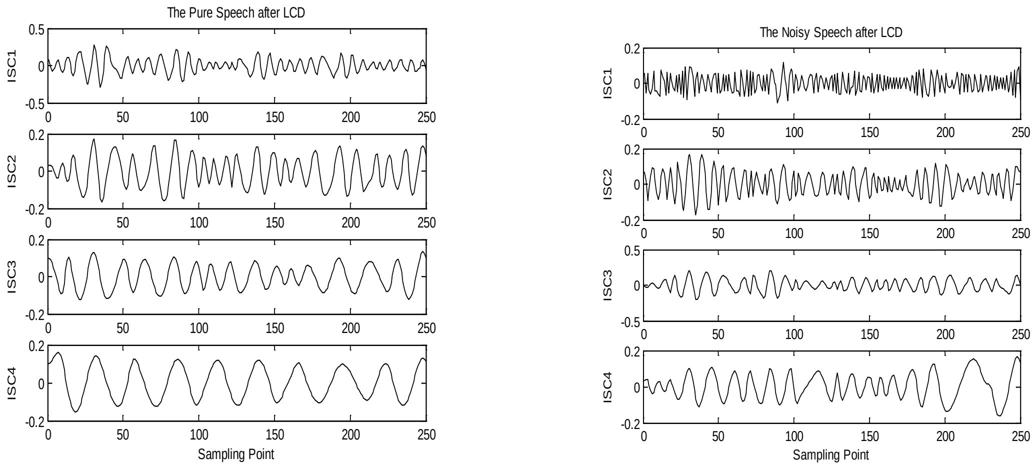
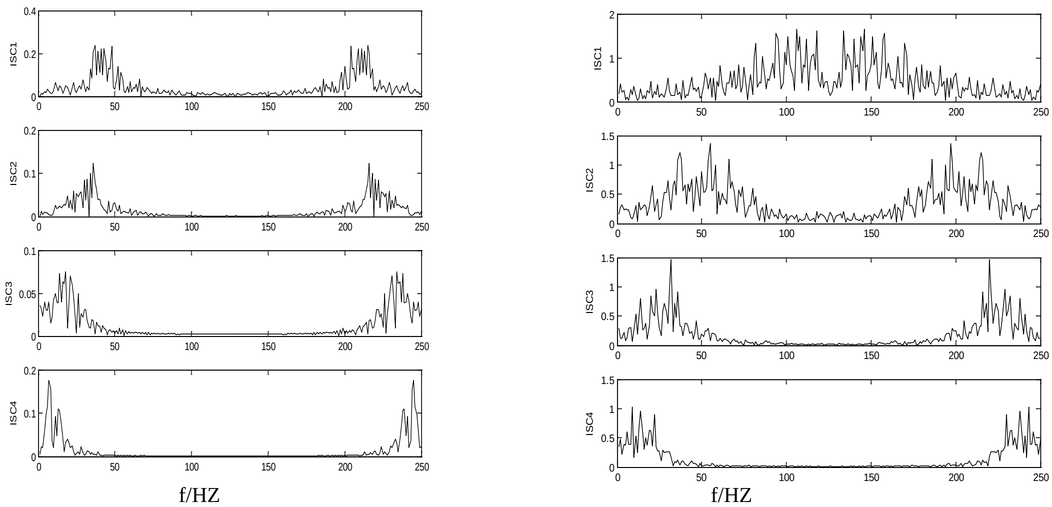


Figure 3. Time-domain waveforms of pure and noisy speech after LCD decomposition



(a)The spectrum of pure Speech frames after LCD (b)The spectrum of noise Speech frames after LCD

Figure 4. Frequency spectrum of pure and noisy speech frames after LCD

C. Reconstruction of Speech Signal

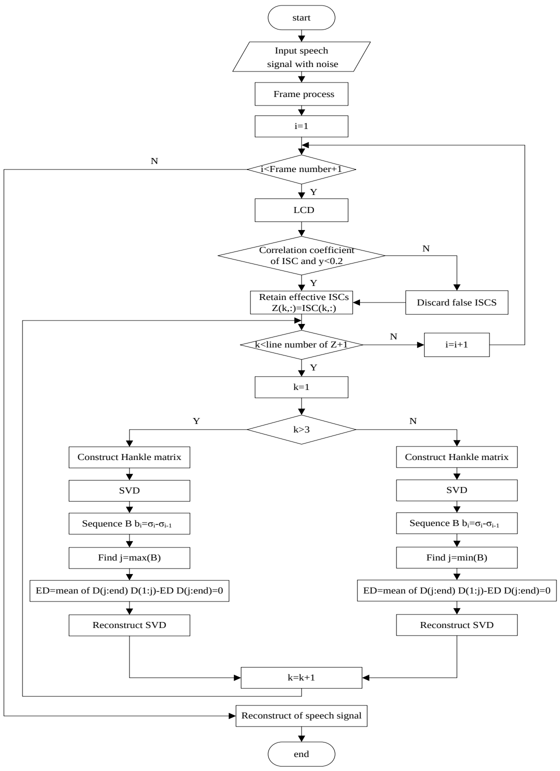


Figure 5. The algorithm flow chart

After the noise reduction order k is determined, $\sigma_k, \sigma_{k+1}, \cdots, \sigma_q$ are singular values of noise. Then, the average singular values of noise can be calculated as $\sigma_{mean(n)} = \frac{1}{(q-k-1)} \sum_{i=k+1}^{q} \sigma_i$. The sequence of singular values subtracts the average singular values of noise. Then, the k-point after the singular values are set to zero, is $\sigma_k{}' = \sigma_k - \sigma_{mean(n)}$, $\sigma_{k+1} = \sigma_{k+2} = \cdots = \sigma_q = 0$.

Finally, the de-noising speech is recovered from a new singular value sequence $Y' = (\sigma_1', \sigma_2', \cdots \sigma_k', 0, 0 \cdots, 0)$.

According to the above procedure, the algorithm flow chart is shown in Figure 5.

IV. Experimental Results

The time-domain waveform for an enhanced speech, input and output SNRs, the speech distortion measurement value, and the mean opinion score are selected as the evaluation performance indices to assess the performance of the proposed speech enhancement algorithm. Therefore, this performance can be analyzed objectively and subjectively.

A. Simulation

In this experiment, a female sound "seven" is chosen as a pure speech signal, with sampling frequency of $10kHz$ with 16 bits. The frame length is $W = 0.025 \times f_s$, and the frame shift is $50\% \times W$. In the case of input SNR ranging from -8 dB to 8 dB, the comparative results of the output SNRs in the different speech enhancement methods are shown in Figure 6. The different speech enhancements contain the traditional wavelet method, spectral subtraction, soft threshold in EMD, and the proposed algorithm.

Figure 6 shows that the highest output SNR of the enhanced speech was found in the proposed algorithm. The wavelet method and soft threshold in EMD results are nearly equal, because they essentially use the threshold to reduce noise.

The spectral subtraction results in the least effective noise reduction, because it depends on the estimation of the noise model to eliminate noise. If noise elimination is inaccurate, the final de-noising results are undesirable. Sometimes spectral subtraction introduces musical noise, which can influence the reduction of auditory effects.

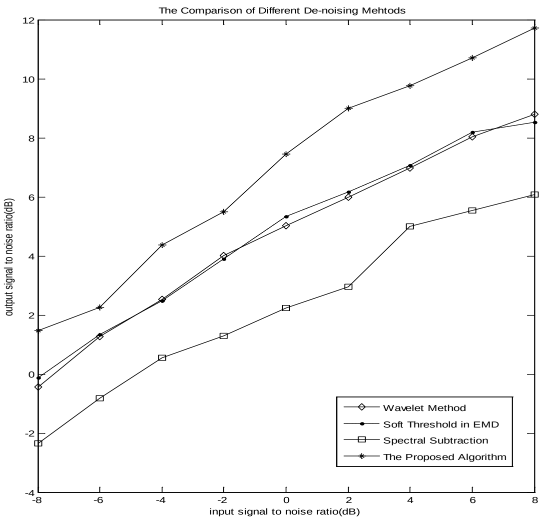


Figure 6. Comparison chart of output SNR using different algorithm

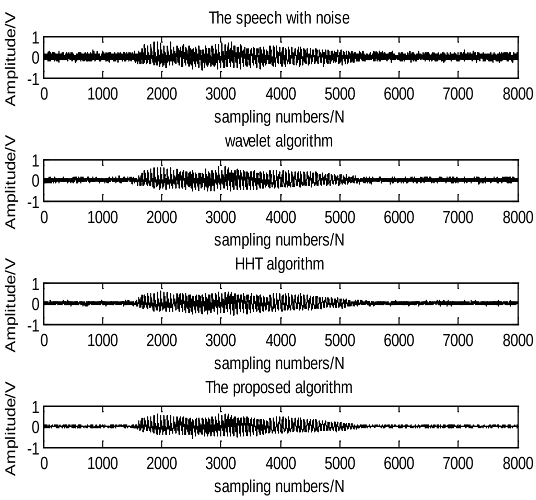


Figure 7. Three algorithms simulation comparison chart

In this paper, the subjective mean opinion score (MOS) acts as an objective evaluation criterion to assess the intelligibility of an enhanced speech algorithm. The noise of voice "seven" is reduced by four algorithms, and the results of MOS are the average of the scores achieved by 10 students to assess the enhanced speech. These results are tabulated in Table III, which shows that the intelligibility of enhanced speech by the proposed algorithm is better than the other algorithms.

When input SNR is -2dB, the three algorithms is used for the voice "seven". They are wavelet algorithm, HHT algorithm and improved algorithm. It is obvious that the improved algorithm can remove a lot of noise from Figure 7.

In this experiment, sound database of one to ten are chosen as speech signals. Mean square error (MSE) is acted as a criterion of measuring the enhancement speech distortion degree. Firstly, one to ten sounds in the input SNR ranging from $-8dB$ to $8dB$ are calculated their MSE. Secondly, each sound average its MSE is acted as final voice distortion measurement values. These results are tabulated in Table IV, which compares with the different speech enhancement methods.

TABLE III.
COMPARISON OF SUBJECTIVE MOS VALUES

The Score of MOS				
Input SNR(dB)	Wavelet Method	Soft Threshold in EMD	Spectral Subtraction	Proposed Algorithm
-8	1.8	2	1.2	3
-6	1.9	2.1	1.4	3.1
-4	1.9	2.1	1.8	3.4
-2	2	2.2	2.2	3.5
0	2.2	2.4	2.4	3.9
2	2.4	2.5	2.6	4.1
4	2.6	2.8	2.9	4.4
6	2.7	2.9	3	4.5
8	2.9	3	3.2	4.6

TABLE IV.
DISTORTION MEASUREMENT VALUES IN INPUT SNR FROM -8DB TO 8DB

DMV / voice	traditional wavelet method	soft threshold in EMD	spectral subtraction	proposed algorithm
one	0.0085	0.0084	0.0168	0.0081
two	0.0048	0.0050	0.0076	0.0043
Three	0.0050	0.0049	0.0092	0.0041
four	0.0084	0.0090	0.0100	0.0050
five	0.0027	0.0031	0.0032	0.0015
six	0.0011	0.0010	0.0014	0.0006
seven	0.0029	0.0025	0.0043	0.0018
eight	0.0025	0.0023	0.0033	0.0011
nigh	0.0030	0.0034	0.0050	0.0028
ten	0.0020	0.0017	0.0031	0.0020

The table IV shows that the least distortion measurement values (DMV) was appeared in the proposed algorithm, it explained the enhancement speech signal used by the proposed algorithm is the most similar to original signal .Therefore, the effect of the proposed algorithm is better than the traditional wavelet method, spectral subtraction and soft threshold in EMD.

B. Simulation Experiment in Different Noisy Conditions

In this experiment, a female sound "seven" is selected as a pure speech signal, with sampling frequency of $10kHz$ with 16 bits. The frame length is $W = 0.025 \times f_s$, and the frame shift is $50\% \times W$. The four noise files from NOISEX-92 database consisting of pink, aircraft, factory, and babble noises are added to the speech files. The noises of the four noisy speeches are removed by the proposed algorithm, and the results of the output SNR and speech distortion values for the various SNRs are tabulated in Table V. The mean square error (MSE) is selected as the evaluation criterion of speech distortion. Table V shows that the output SNRs are improved significantly and the speech distortion values are smaller. Therefore, the proposed algorithm to reduce pink, aircraft, factory, and babble noises is practical.

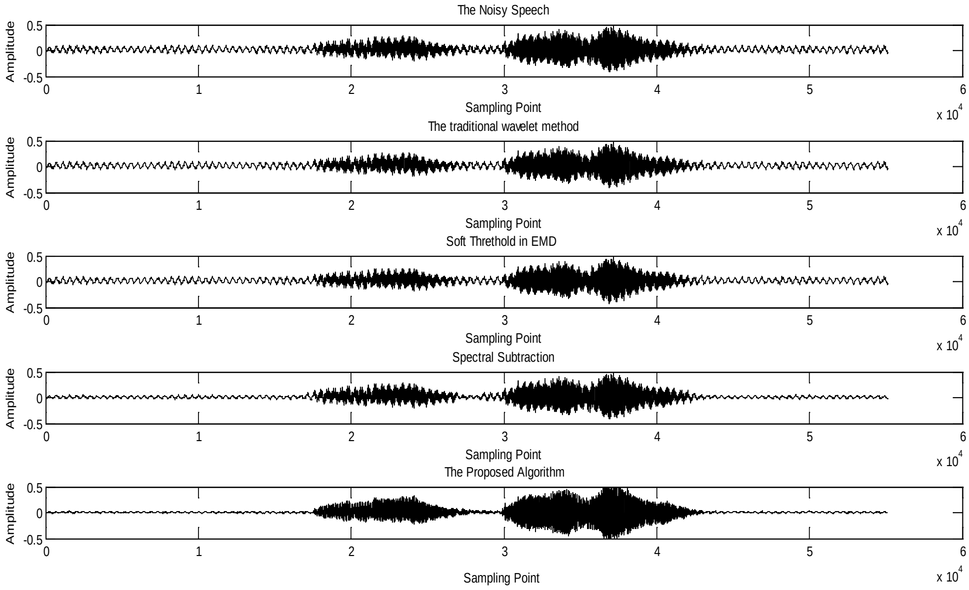


Figure 8. Waveforms of the enhanced speech in the different algorithms

TABLE V.
COMPARISON OF DE-NOISING RESULTS OF THE DIFFERENT NOISE CONDITIONS

Comparison	Pink Noise	Aircraft Noise	Factory Noise	Babble Noise
SNR_{input} (dB)	4.861	1.763	3.782	2.139
SNR_{output} (dB)	8.425	7.696	8.148	5.923
MSB	0.0009	0.0010	0.0009	0.0015

C. Simulation Experiment in De-noising Real Speech

In this experiment, a female sound “ni hao” is selected, with sampling frequency of $22.05kHz$. This voice is transcribed by Windows recording software in a real environment. The noise resources are derived from the sound of old fan blades and an old computer system. The waveforms for the enhanced speech using the traditional wavelet method, soft threshold in EMD, spectral subtraction, and the proposed algorithm are demonstrated in Figure 8. The proposed algorithm can considerably eliminate background noise from practically noisy speech, and the enhanced speech does not lose as much useful speech component as the other algorithms do.

V. CONCLUSIONS

A new speech enhancement algorithm based on LCD and difference spectrum of singular value is proposed. Based on the different frequency distributions of useful speech and noise signals after LCD, the two criteria of difference spectra can be used to determine the noise reduction orders so that a mass of noise can be removed. The proposed algorithm overcomes the drawbacks of the basis functions and threshold selection irregularity. However, this algorithm cannot introduce musical noise after noise reduction. The simulation experiments verify the effectiveness of the proposed algorithm.

REFERENCES

[1] Kim J B, Lee K Y and Lee C W.,” On the applications of the interacting multiple model algorithm for enhancing noisy speech”, *IEEE Transaction on Speech and Audio Processing*, vol. 8, no. 3, pp. 349-352, 2000.

[2] Zhao Huan, Zhao Kai, Liu He, et al. Improved MFCC feature extraction combining symmetric ICA algorithm for robust speech recognition. *Journal of Multimedia*, vol. 7, no.1, pp.74-81, 2012.

[3] Jindong Zhang, Guihe Qin, Ye Liu. Speech Separation in the Vehicle Environment Based on FastICA Algorithm. *Journal of Multimedia*, Vol. 7, no.1, pp.33-40,2012.

[4] Huan Zhao, He Liu, Kai Zhao, et al. Robust Speech Feature Extraction Using the Hilbert Transform Spectrum Estimation Method. *International Journal of Digital Content Technology and its Applications*, vol.5, no. 12,pp. 85-95, 2011.

[5] Huan Zhao, Yufeng Xiao. A Novel Robust MFCC Extraction Method Using Sample-ISOMAP for Speech Recognition. International Journal of Digital Content Technology and its Applications, vol.6, no. 19, pp. 393 - 400, 2012.

[6] Huan Zhao, Xiujuan Peng, Lian Hu, et al. “An Improved Speech Enhancement Method Based on Teager Energy Operator and Perceptual Wavelet Packet Decomposition”. *Journal of Multimedia*, vol. 6, no. 3, pp. 308-315,2011.

[7] Kuldip Paliwal, Belinda Schwerin, Kamil Wofcicki, “Speech enhancement using a minimum mean-square error short-time spectral modulation magnitude estimator”. Journal Speech Communication, vol.54, no.2,pp. 282-305, 2012.

[8] LI Ye, Wu Ya-feng, LIU Xue-fei.”Research on speech enhancement based on perceptual wavelet transform”,

Application Research of Computers, vol. 26, no. 4, pp. 1313-1315, 2009.

[9] LIU Li-wei, LI Jin-bao, ZHAO Kong-xin, DING Tie-fu, " The application of Hilbert-Huang Transform in speech Enhancement", *Microcomputer information*, vol.23, no.10, pp. 208-210,2007.

[10] Xiaojie Zou, Xueyao Li, and Rubo Zhang, "Speech Enhancement Based on Hilbert-huang Transform Theory", *Proceeding of the First International Multi-Symposiums on Computer and Computational Sciences*, vol.1, pp.208-213, 2006.

[11] LOU Li-xia, ZENG Yi-cheng, XIONG Min-quan , "An approach of speech enhancement combining EMD and adaptive filter", *Computer Engineering and Applications*, vol.48, no.9, pp. 104-107,2012.

[12] Zhao hua Wu,Norden E.Huang."A Study of the Characteristics of White Noise Using the Empirical Mode Decomposition Method". *Proceedings of the Royal Society of London*, vol. 460, no. 2046, pp. 1597-1611, 2004.

[13] Norden .E.Huang et al,"The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis", *Proc. R. Soc. Lond. A*, vol. 454, no.1971, pp. 903-995, 1998.

[14] Cheng Jun-sheng, ZHENG Jin-de, YANG Yu, "A nonstationary signal analysis approach-the local characteristic-scale decomposition method", *Journal of Vibration Engineering*, vol.25, no.2, pp. 215-220, 2012.

[15] HU Mou-fa, DONG Wen-juan, WANG Shu-hong, CHEN Zeng-ping, "Singular Value Decomposition Band-Pass-Filter for Image Background Suppression and Denoising", *Acta Electronica Sinica*,vol.36,no.1, pp. 111 -116, 2008.

[16] Xuezhi Zhao, Bangyan Ye, "Selection of effective singular values using difference spectrum and its application to fault diagnosis of headstock", *Mechanical Systems and Signal Processing,* Vol.25, no.l5, pp.1617-1631,2011.

[17] WANG Tai-yong, WANG Zheng-ying, XU Yong-gang, LI Rui-xin, "Empirical Mode Decomposition and Its Engineering Applications Based on SVD Denoising", *Journal of Vibration and Shock*, vol.24, no.4, pp. 96-98, 2005.

[18] Li Ruixin,Wang Dongfeng,Han Pu,etal."On the applications of SVD in fault diagnosis".*Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, vol.4, pp.3763-3768, 2003.

[19] N.Virag,Single Channel.Speech Enhancement Based on Masking Properties of Human Auditory System. *IEEE Transactions on Speech and Audio Processing*, vol.7, no.2, pp. 126-137, 1999.

[20] P. Lockwood, J. Boudy, M. Blanchet." Experiments with a Nonlinear spectral subtraction (NSS), Hidden Markov models and projection, for robust recognition in cars". *Journal Speech Communication*, vol. 1, no.2-3, pp.215-228, 1992.

Nan Li achieved the MS. in Power electronics and Power transmission in 2005. He is currently Assistant Professor at the Faculty of Information Engineering of Northeast Dianli University, P.R.China. Her research interests are mainly in weak signal detection, nonlinear systems application in signal processing. She published 10 papers on journals and 6 papers on international conferences.