An Improved Spectral Subtraction Algorithm for Speech Enhancement System

Shun Na, Weixing Li, Yang Liu*

College of Electronic Information Engineering, Inner Mongolia University, Hohhot, 010021, China
* Corresponding author:

E-mail address: yangliuimu@163.com

Keywords:speech enhancement, spectral subtraction, noise estimation.

Abstract

In this paper, we present an application of spectral subtraction (SS) algorithm in speech enhancement system to extract the pure speech signal as far as possible. In contrast to the existing research, the proposed algorithm improves the voice quality, which reduces speech distortion, eliminates background noise and improves the speech intelligibility. This paper first introduces the research significance of the speech enhancement, then introduces the relevant theories of speech signal processing, and expounds the basic spectral subtraction speech enhancement, through a lot of simulation experiments verify the effect of spectral subtraction. Based on the voice activation detection algorithm is studied and an improved spectrum subtraction（ISS）algorithm was presented. Our simulation results show that the proposed ISS Algorithm is effective with the lower computational complexity in speech enhancement system.

1. Introduction

There are many situations when speech has to be processed in the presence of undesirable background noise that degrades speech quality and intelligibility. A variety of speech enhancement methods capable to reduce background noise were studied in the literature. Environment noise inevitably influences our speech communication quality. Speech enhancement technology is an available approach to resolve the influence of noise, while engineers are in favour of single microphone speech enhancement technology based on short time spectral estimation, such as algorithms of spectral subtraction [1], Wiener filter [2], minimum mean square error(MMSE) estimation [3], vector Taylor Series[4] and cepstral histogram equalization [5] etc. These algorithms improve the enhancement effect on a certain extent, but the enhancement effect is still to be improved. That is because these algorithms have not concerned with the ultimate property of noise spectral.

Spectral subtraction method, as proposed by [6], is a commonly used noise reduction method that has high noise reduction performance. The basic principle of the spectral subtraction method is to subtract the shone-term spectral magnitude of noise from that of the noisy speech. The noise is assumed to be uncorrelated and additive to the speech signal. An estimate of the noise signal is measured during silence or non-speech activity in the signal.

Spectral subtraction from the voice of the short-term spectrum with noise value minus the noise in the short-term spectrum to achieve the purpose of the speech enhancement, it adopted by the algorithm is simple and easy to implement. The main idea of the spectral subtraction is assuming that noise and speech signal under the condition of independent of each other, from the band noise power spectrum minus the noise power spectrum, relatively pure speech spectrum is obtained, the following is the basic principle of spectrum subtraction.

2. Discussed problems

Hypothesis $y(t)$ is noise speech signal, $s(t)$ is pure speech signal, $n(t)$ denotes signal to noise. The voice is short time smoothly, so that it is in short-term spectral magnitude estimate stationary

random signal, assuming that noise is $n(t)$ and voice $s(t)$ additive noise. Get signal additive model [7],

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

Energy spectrum of speech signals with noise can be expressed as after dealing with the add window of the signal, FFT transform respectively is $y_w(t)$, $s_w(t)$, $n_w(t)$, there are

$$y_w(t) = s_w(t) + n_w(t) \tag{2}$$

After dealing with the add window Fourier transform (FFT) signal

$$Y_w(w) = S_w(w) + N_w(w) \tag{3}$$

Power spectrum is

$$\begin{aligned} |Y_w(w)|^2 &= |S_w(w)|^2 + |N_w(w)|^2 \\ &\quad + 2\mathrm{Re}\left[S_w(w)N_w^*(w)\right] \end{aligned} \tag{4}$$

Again on type available,

$$\begin{aligned} E(|Y_w(w)|^2) &= E(|S_w(w)|^{2)} + E(|N_w(w)|^2) \\ &\quad + 2E\left\{\mathrm{Re}\left[S_w(w)N_w^*(w)\right]\right\} \end{aligned} \tag{5}$$

Because of the assumptions are $n(t)$ and voice $s(t)$ are independent of each other, $N(\omega)$ and $S(\omega)$ is also independent of each other, according to the characteristics of stationary random noise $N(\omega)$ is a zero mean Gaussian distribution. So the expression of $E\left\{\mathrm{Re}\left[S_w(w)N_w^*(w)\right]\right\}$ is 0, Therefore, (3-5) deformation for

$$E(|Y_w(w)|^2) = E(|S_w(w)|^{2)} + E(|N_w(w)|^2) \tag{6}$$

For the frame inside short time stationary process

$$|Y_w(w)|^2 = |S_w(w)|^2 + |N_w(w)|^2 \tag{7}$$

where $Y_w(w)$ is,

$$Y_w(w) = \sum_{n=0}^{N-1} y(n)e^{-j\frac{2\pi kn}{N}} = |Y(w)|e^{j\phi(w)} \tag{8}$$

where $\phi(w)$ is the phase of noise speech of $Y_w(w)$.

Because directly by noise speech noise energy spectrum cannot estimated in $|N_w(w)|^2$, generally several frames silent phase noise signal energy spectrum of the noise power spectrum estimate $|N_w(w)|^2$ is calculated. Due to the power spectrum of smooth voices in voice before and after can be thought of as basic did not change, so we can through the voice of so-called wave silence before to estimate the noise power spectrum

$$\{\ p_s(w) = p_y(w) - p_n(w) \tag{9}$$

Such reduction can be thought of as the power spectrum of the relatively pure speech, however, from the power can be restored after noise reduction of speech signal.

In the concrete operation, in order to prevent the negative power spectrum, spectral reduction $p_y(w) < p_n(w)$, To $p_s(w) = 0$, i.e., full spectrum reduction computation formula is:

$$p_s(w) = p_y(w) - p_n(w), p_y(w) \geq p_n(w) \tag{10}$$

For assumptions were not associated with the noise and speech signal, energy spectrum estimation for speech

$$|\hat{S}_w(w)|^2 = |\hat{Y}_w(w)|^2 - |\hat{N}_w(w)|^2 \tag{11}$$

where clean speech power spectrum estimate $|\hat{S}_w(w)|^2$ by the energy spectrum of speech signals with noise minus, which named the noise power spectrum estimation. Because of the noise power spectrum estimation and noise in the speech signals with noise exist differences, energy spectrum of (10) may be negative, in order to avoid the negative energy spectrum, the negative value is set to zero, this process is called half-wave rectifier (or half wave rectification). Through the half-wave rectifier, pure voice energy spectrum estimation can be expressed as $|\hat{S}_w(w)|^2$.

$$|\hat{S}_w(w)|^2 = \begin{cases} |\hat{S}_w(w)|^2 & |\hat{S}_w(w)|^2 > 0 \\ 0 & |\hat{S}_w(w)|^2 < 0 \end{cases} \tag{12}$$

Combining with noise speech phase information, through the inverse discrete Fourier transform (IDFT), get clean speech signal estimation of signal is $\hat{s}(n)$

$$\hat{s}(n) = \mathbf{IDFT}(\left|\hat{S}_w(w)\right| e^{j\phi(w)}) \tag{13}$$

Spectral subtraction in the frequency domain will take noise power spectrum minus the noise power spectrum, in order to get clean speech power spectrum estimation, prescribing after get speech spectral amplitude estimation, with the phase noise to approximate the phase of the pure voice. Again the inverse Fourier transform for USES to restore time domain signals, flow chart of spectrum subtraction speech enhancement algorithm is shown in Figure 1.

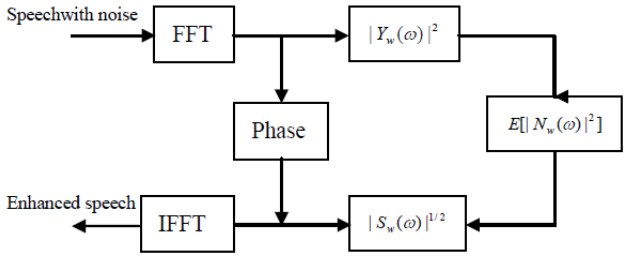

Fig.1: Spectral subtraction speech enhancement algorithm

3. Improved spectrum subtraction algorithm

The basic problems existing about the traditional spectral subtraction algorithm [8], [9], is that the inevitable music noise is introduced. Accurately estimate of noise in the speech signals is need with noise spectrum, in order to effectively filter out the noise of the voice signals. Noise spectrum estimation is more accurate, enhancing the smaller music noise in the signal spectrum. However, because noise spectrum can be obtained directly in the vast majority of spectral subtraction algorithm, which is obtained by the weighted average of silent phase noise spectrum estimation.

A result of noise spectrum estimation error is the negative energy value. The negative with half-wave rectifier or full wave rectifier (was set to the absolute value). It is not correct treat such mistake, resulting further distortion in the time domain. In the time domain, signals are generated to enhance, the phase of speech signals with noise did not make any changes. This is based on the fact that the phase distortion of the effects on the voice quality decline. When the signal-to-noise ratio (SNR) than high (> 5 db), phase distortion is little impact on the quality of voice, however, when the low SNR (< 0 db) the voice quality decline due to phase distortion can be felt.

The traditional noise estimation is based on the optimal smoothing and least statistical noise estimation, which the algorithm with the improved noise estimation based on voice activity detection [10].Voice activation detection refers to the voice from the speech signal contains certain starting point and end point, also known as the endpoint detection. The purpose of speech endpoint detection is from the continuous recording of the speech signal with noise isolated useful speech signal. Voice activation detection is need in the various speeches processing, and is an important link. Accurately determine the beginning and end of input speech will ensure the performance of the voice processing system [11].

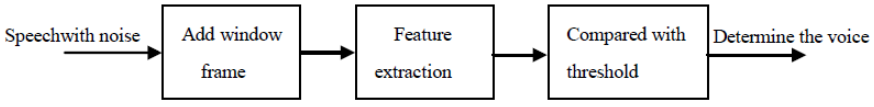

Fig.2: Voice activation detection block diagram

Specific short time energy detection algorithm based on signal is follows:

i) Calculate each frame of speech energy:

$$E_n = \sum_{m=0}^{N-1} x_n^2(m) \tag{14}$$

where N is the frame length, n is the serial number of the frame, m is each point in each frame, $1 \le n \le L$, L is the number of frames. It has a defect, however, that is sensitive to high level (signal of quadratic calculation). For this purpose, the definition of short time average magnitude functions to characterize a frame of speech signal energy size, definition.

$$M_n = \sum_{m=0}^{N-1} |x_n(m)| \tag{15}$$

ii) Average noise energy calculated before 20 frames
iii) Energy maximum and minimum value denote EAX and EMI
iv) According to the (15), determine the threshold.

$$T = \min[0.03(EAX - EMIN) + EMN, 4EMN] \tag{16}$$

4. Simulation results

Simulation respectively the draw the original clean speech waveform, spectrum and after high-pass filter the speech signal waveform and spectrum. The original speech signal waveform after sampling in time domain and spectrum graph, after adding noise signal plus noise speech signal waveform and spectrum diagram. The basic spectral subtraction is simulation as waveform diagram.

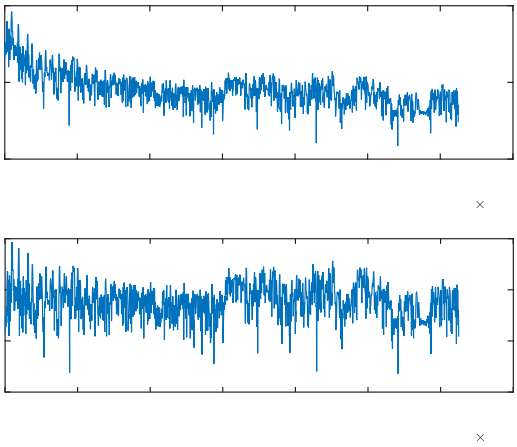

Fig.3: Clean speech signal

It can be seen from the Fig.3, this article MATLAB has read the selected voice signal sound clearer at the high SNR. In fact, this signal experimental contrast effect is not obvious. So before you eliminate noise experiments, we artificially add random white Gaussian noise to the original signal and reduce the SNR of speech signal.

Two image contrast can be seen from the Fig.4, add Gaussian random noise in pure speech signal, the signal waveform becomes relatively vague and spectral change is also very obviously. The introduction of random noise signal, greatly reduces the signal-to-noise ratio of speech signals, this step is to prepare for further study on the back.

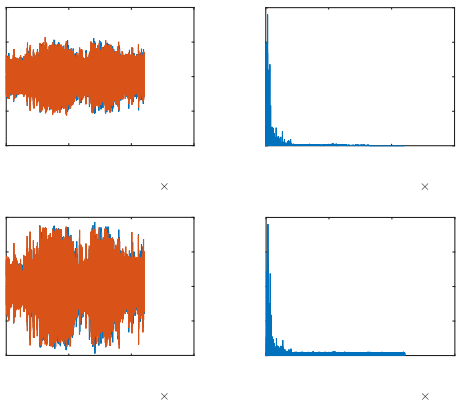

Fig.4: Speech signal with noise

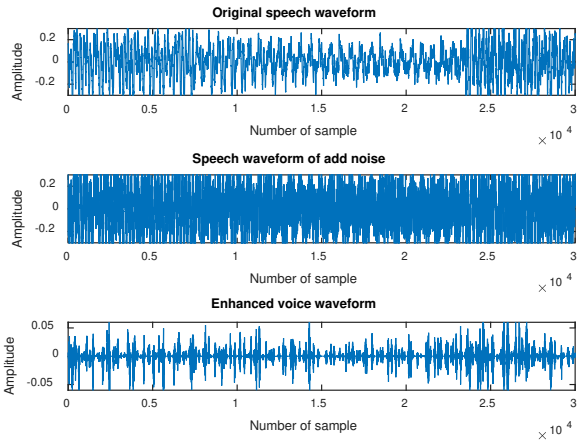

Fig.5: The realization of the basic spectral subtraction speech waveform

The basic spectral subtraction can be seen from the Fig. 5, which before and after implementation. The original speech waveform is relatively smooth and clear, after joining the noise speech waveform is blurry, obviously after the classic spectral subtraction. After deducting the spectrum of noise reduction of waveform is compared with the original pure voice. The basic back to the original voice clarity, so it is proves that the spectral subtraction is to realize the voice strengthen the use of a very good tool through the basic spectral subtraction practice [12].

Based on the improved original speech signals waveform figure, add noise speech signal waveform and speech to strengthen after the waveform diagram shown in Fig.6. According to the above analysis, we can conclude that the noise estimation algorithm based on voice activity detection after the simulation waveform is better than the classic basic spectral subtraction resulting from the simulation waveform.

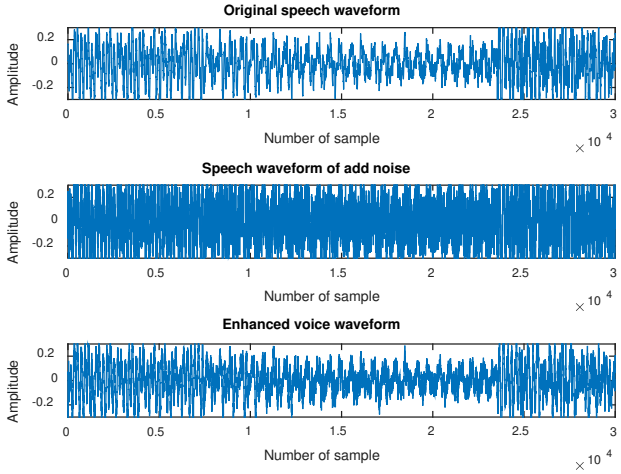

Fig.6: Noise estimation based on voice activity detection algorithm simulation

5. Conclusions

This paper generalizes the classic spectral subtraction for speech enhancement system. The goal of this article is try to remove the interference of noise to restore a pure voice, such as speech signal acquisition, extract the noise from the speech signal and implement to remove noise from the speech signal. We learned some relevant characteristics of signal and noise of estimation in time domain and frequency domain. We first research speech to strengthen from the classic spectral subtraction. On the basis of the found some shortages existing in the traditional spectral subtraction, we put forward and research a new modified spectral subtraction algorithms. Our analysis and simulation indicate that the performance of this improved algorithm is much better than that of standard spectral subtraction.

Acknowledgements

The authors are grateful to the National Science Foundation of China for its support of this research. This work is partly supported by the National Science Foundation of China under Grant 61362027, 61461036 and 61561037, and the Natural Science Foundation of Inner Mongolia Autonomous Region of China under Grant 2016MS0604.

References

[1] Sui, L, Zhang, X, Huang, J, Zhou, B, An improved spectral subtraction speech enhancement algorithm under no stationary noise. 2011 International Conference on Wireless Communications and Signal Processing (WCSP), IEEE, 2011.

[1] Zhang, B, The algorithm research of speech enhancement based on wavelet. Harbin: Harbin Engineering University, 1992.

[3] Gupta, V. K., et al., Speech Enhancement Using MMSE Estimation and Spectral Subtraction Methods. 2011 International Conference on Devices and Communications (ICDeCom), IEEE, 2011.

[4] Yong, L, Zhen, W, Robust Speech Recognition Based on Vector Taylor Series. Journal of Tianjin University, 22(3), PP.103-108, 2011.

[5] Zhao, Y, Jiang, B, Stranded Gaussian mixture hidden Markov models for robust speech recognition. 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2012.

[6] Boll, S. F, Suppression of acoustic noise in speech using spectral subtraction, IEEE Trans. on Acoustics, Speech and Signal Processing, 27(7), pp. 113-120, 1979.

[7] Boll, S. F, Suppression of acoustic noise in speech using spectral subtraction, IEEE Trans. on Acoustics, Speech and Signal Processing, 27(7), pp. 113-120, 1979. SOON, I. Y, KOH S. N, Speech enhancement using 2-D Fourier transform. IEEE Trans Speech Audio Process, 1l(6), pp.717—724, 2003.

[8] Kamath, S, Loizou, P, A multi-band spectral subtraction method for enhancing speech corrupted by colored noise, Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Processing, 2002.

[9] Zhang, C, Hu, X, Zhou, Y, Spectral subtraction based on the structure of noise power spectral density, ACTA ACUSTICA, 35(2), pp.216-222, 2010.

[10] ZCheng, G, Guo, L, Zhao, T, He, S, A More Effective Speech Enhancement Algorithm under Non-Stationary Noise Environment, Journal of Northwestern Polytechnical University, 28(5), pp. 664-668, 2010.

[11] Kamath, S, Loizou, P, A multi-band spectral subtraction method for enhancing speech corrupted by colored noise, Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Processing, 2002. Inoue, T, Saruwatari, H, Takahashi, Y, Shikano, K, Theoretical analysis of iterative weak spectral subtraction via higher-order statistics, Proc. MLSP'10, pp. 220–225, 2010.

[12] Zhang, C, Hu, X, Zhou, Y, Spectral subtraction based on the structure of noise power spectral density, ACTA ACUSTICA, 35(2), pp.216-222, 2010. Takahashi, Y, Saruwatari, H, Shikano, K, Musicalnoiseanalysisinmethodsofintegratingmicrophonearrayandspectral subtraction based on higher-order statistics, EURASIP J. Adv. Signal Process, 11, 2010.