Application of Grey Filtering to Speech Enhancement

Cheng-Hsiung Hsieh
Department of Electronic Engineering
Chien Kuo Institute of Technology
Changhua, TAIWAN, R.O.C.
E-mail. hsieh@ckit.edu.tw

Abstract

In this paper, a speech enhancement approach based on grey filtering is proposed. In a previous paper [1], it is shown that additive noise can be estimated accurately by grey filtering approach with an appropriate scaling factor. Note that the spectral subtraction approach to speech enhancement is heavily dependent on the accuracy of statistics of additive noise and that the grey filtering is able to estimate additive noise appropriately. A magnitude spectral subtraction (MSS) approach for speech enhancement is proposed where the mechanism to determine the non-speech and speech portions is not required. Simulation results are provided to justify the proposed MSS approach based on grey filtering. It indicates that the objective of speech enhancement has been achieved by the proposed MSS approach.

Keywords: Grey filtering, GM(1,1) model, additive noise, estimation error, speech enhancement, spectral subtraction

1. Introduction

The purpose of filtering is to recover signal component from noisy observations [2]. Filtering is required in many engineering applications. One example is the speech enhancement. Assume that the signal model is the additive signal model, which is expressed as $x(k) = s(k) + n(k)$ where $x(k)$, $s(k)$, and $n(k)$ are noisy speech, clean speech, and the additive noise, respectively. The objective of speech enhancement is to recover $s(k)$ from noisy speech $x(k)$. Note that filtering $s(k)$ out of $x(k)$ is equivalent to the estimation of additive noise $n(k)$. Therefore better performance of speech enhancement results from appropriate noise estimation. Basically, the speech enhancement consists of two stages: noise estimation and noise removal. Up to present, several noise estimation approaches have been reported. Some of representative approaches are as follows. By a mechanism to determine non-speech and speech portions in $x(k)$, additive noise is estimated during non-speech period in [3]. Note that the spectrum above speech frequency component comes from $n(k)$ if it is white noise. In [4], the spectral component of white noise is estimated through linear prediction coefficients while higher sampling rate is used for spectral estimation of $n(k)$ in [5]. By signal insertion in the transmitted speech signal, in [6] the contaminated inserted signals are used to estimate noise. Since additive noise is random, it is appropriate to deal with $n(k)$ in a statistical way. Therefore, statistics of $n(k)$ is sufficient in many practical applications of speech enhancement. When statistics of noise are estimated, a noise removal technique is applied. The noise removal approach can be Weiner filtering as in [3], Kalman filtering as in [4], and a popular approach called spectral subtraction as in [5] and [6].

In this paper, we apply grey filtering, which is based on GM(1,1) model [7], to speech enhancement whose noise removal technique is based on magnitude spectral subtraction (MSS) [8]. This paper is motivated by the following observations. In [1], it is noted that the estimation error of GM(1,1) model is zero for a constant signal and approximately zero for random signal when additive noise is absent. When additive noise is present, both signals have non-zero estimation error. These observations imply that estimation error of GM(1,1) model can be related to additive noise. Furthermore, the speech signal generally consists of two parts: non-speech and speech. The non-speech portion can be considered as constant signal while speech portion as random signal. Consequently, there is a hope to estimate additive noise in noisy speech through estimation error of GM(1,1) model and therefore speech enhancement by spectral subtraction is possible.

This paper is organized as follows. In Section 2, a brief review of GM(1,1) model [7] is given and grey filtering and noise estimation based on GM(1,1) model is described. In Section 3, the application of grey filtering to MSS [8] for speech enhancement is proposed and described. Then simulation results are provided to justify the proposed MSS approach in Section 4. Finally, conclusive remarks are made in Section 5.

2. Grey Filtering Based on GM(1,1) Model

In this section, a brief review of GM(1,1) model is given first and then the grey filtering approach to noise estimation is described.

2.1 GM(1,1) model

The GM(1,1) modeling process is described in the following. For details, one may consult [6]. Given data sequence $\{x(k), \text{for}\, 1 \le k \le K\}$, a new data sequence $x^{(1)}(k)$ is found by 1-AGO (first-order accumulated generating operation) as

$$x^{(1)}(k) = \sum_{n=1}^{k} x(n) \qquad (1)$$

for $1 \le k \le K$, where $x^{(1)}(1) = x(1)$. To be effective in GM(1,1) modeling, $x(k)$ needs to meet two conditions: (i) data is of same sign, and (ii) the ratio between adjacent data in $x(k)$ should be within one order in magnitude. From (1), it is obvious that the original data $x(k)$ can be easily recovered from $x^{(1)}(k)$ as

$$x(k) = x^{(1)}(k) - x^{(1)}(k-1) \qquad (2)$$

for $2 \le k \le K$. This operation is called 1-IAGO (first-order inverse accumulated generating operation).

By sequences $x(k)$ and $x^{(1)}(k)$, a grey difference equation is formed as

$$x(k) + az^{(1)}(k) = b \tag{3}$$

where

$$z^{(1)}(k) = 0.5(x^{(1)}(k) + x^{(1)}(k-1)) \tag{4}$$

for $2 \le k \le K$, and parameters *a* and *b* are called developing coefficient and grey input, respectively. From (3), parameters *a* and *b* can be obtained as

$$\begin{bmatrix} a \\ b \end{bmatrix} = (\boldsymbol{B}^T \boldsymbol{B})^{-1} \boldsymbol{B}^T \boldsymbol{y} \tag{5}$$

where

$$\boldsymbol{B} = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(K) & 1 \end{bmatrix} \tag{6}$$

and

$$\boldsymbol{y} = \begin{bmatrix} x(2) \\ x(3) \\ \vdots \\ x(K) \end{bmatrix} \tag{7}$$

Next, to find the solution of $x^{(1)}(k)$ in (3), we utilize its associated differential equation which has the following form

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = b \tag{8}$$

It can be easily shown that the solution for $x^{(1)}(t)$ in (8) is

$$x^{(1)}(t) = ce^{-at} + \frac{b}{a} \tag{9}$$

where *c*, by the initial condition $x^{(1)}(t_0) = x(t_0)$, can be found as

$$c = (x(t_0) - \frac{b}{a})e^{at} \tag{10}$$

Therefore, the solution for $x^{(1)}(t)$ is given as

$$x^{(1)}(t) = (x(t_0) - \frac{b}{a})e^{-a(t-t_0)} + \frac{b}{a} \tag{11}$$

Letting $t_0 = 1$ and $t = k$, we have the solution of $x^{(1)}(k)$ as follows.

$$x^{(1)}(k) = (x(1) - \frac{b}{a})e^{-a(k-1)} + \frac{b}{a} \tag{12}$$

where parameters *a* and *b* are found in (5). By 1-IAGO, the estimate of $x(k)$, $\hat{x}(k)$, is obtained as

$$\hat{x}(k) = x^{(1)}(k) - x^{(1)}(k-1) \tag{13}$$

where $\hat{x}(1) = x^{(1)}(1) = x(1)$. The estimation error for $x(k)$ is given as

$$e(k) = x(k) - \hat{x}(k) \tag{14}$$

which will be used to estimate additive noise later in Subsection 2.2.

To sum up, the GM(1,1) modeling process consists of three steps. First, find parameter *a* and *b* by (5). Second, use (12) to estimate $x^{(1)}(k)$. Finally, find $\hat{x}(k)$ through (13). It should be noticed that the minimum number of data samples in GM(1,1) modeling is as few as four samples, i.e. $K = 4$.

2.2 Grey filtering and noise estimation

The grey filtering approach proposed in [1] is described here. Assume the available noisy signal $x(k)$ satisfies Conditions (i) and (ii) in Subsection 2.1 and has the additive signal model $x(k) = s(k) + n(k)$ where $s(k)$ and $n(k)$ are the clean signal and the additive noise in $x(k)$, respectively. Then denote a segment of noisy signal as $\{x(k), \text{for } 1 \le k \le L\}$ where $L = 1 + N_{ss}(K-1)$ is the total number of samples. Notation K is the number of samples used in GM(1,1) modeling and $N_{ss} = \lfloor L/(K-1) \rfloor$ is the number of subsets with one sample overlapped. The proposed grey filtering approach is given as follows.

Step 1. Divide $\{x(k), \text{for } 1 \le k \le L\}$ into N_{ss} subsets as $\{x_i(k), \text{for } 1 \le i \le N_{ss}\}$.

Step 2. For each subset *i*, find estimate of $x_i(k)$, $\hat{x}_i(k)$, by GM(1,1) model as stated in Subsection 2.1. Then consider $\hat{x}_i(k)$ as an estimate of $s_i(k)$, $\hat{s}_i(k)$. That is, $\hat{x}_i(k) = \hat{s}_i(k)$.

Step 3. Estimate noise component $n_i(k)$ as $\hat{n}_i(k) = a\,(x_i(k) - \hat{s}_i(k)) = a e_i(k)$ where $a > 0$ is a scaling parameter and $e_i(k)$ is the estimation error of GM(1,1) model for $x_i(k)$.

Step 4. Estimate mean m of additive noise $n(k)$ as

$$\hat{m} = \frac{1}{N_{ss}(K-1)} \sum_{i=1}^{N_{ss}} \sum_{k=2+(i-1)(K-1)}^{1+i(K-1)} \hat{n}_i(k) \tag{15}$$

Since $x_i(k)$ is of one sample overlapped, thus only $\hat{n}(1) = 0$ is excluded in (15).

Step 5. Estimate standard deviation s of $n(k)$ as

$$\hat{s} = \Big[\frac{1}{N_{ss}(K-1)} \sum_{i=1}^{N_{ss}} \sum_{k=2+(i-1)(K-1)}^{1+i(K-1)} (\hat{n}_i(k) - \hat{m})^2\Big]^{1/2} \tag{16}$$

3. Application of Grey Filtering to Speech Enhancement

In this section, the motivations for this paper are given in Subsection 3.1. Then the proposed spectral subtraction approach for speech enhancement, which is based on grey filtering, is described in Subsection 3.2.

3.1 Motivations

This paper is motivated by the following observations. First, as shown in [1] the estimation error of GM(1,1) model is zero or approximately zero when additive noise is absent and non-zero when additive noise is included. This is true both for constant and random exponential signal. This implies that the estimation error of GM(1,1) model can be related to additive noise. As demonstrated in [1], statistics of additive can be estimated accurately with an appropriate scaling factor a . Second, a clean speech signal generally consists of non-speech and speech portions. The non-speech portion can be considered as constant signal while the speech portion as random signal. Consequently, there is a hope that the estimation error of GM(1,1) model for clean speech is approximate to zero and non-zero when noisy speech is present. Moreover, there is no need to determine speech and non-speech portion as in [3] since both constant and random signals can be estimated appropriately by GM(1,1) model and the GM(1,1) modeling requires as few as four data samples. To demonstrate the idea just described, the clean speech b.wav (male speech of letter “b”) obtained from [9] is provided as an example which is shown in Figure 1(a). Since b.wav is within the range $(-1,1)$, it fails to meet the requirement of Condition (i) in Subsection 2.1. To make it satisfied, b.wav is level-shifted by 5 before it is put into GM(1,1) modeling. Condition (ii) is met in speech signal since adjacent samples does not change abruptly in general. The estimate of b.wav obtained from GM(1,1) model is given in Figure 1(b) where $K=4$. Obviously, the estimate of b.wav by GM(1,1) model retains b.wav appropriately as shown in Figure 1(b).

Note that the standard deviation s of additive noise $n(k)$ can be estimated accurately by grey filtering and that a spectral subtraction approach for speech enhancement depends heavily on the accuracy of the standard deviation of $n(k)$. An MSS [8] based on grey filtering is proposed in this paper. The proposed approach is described in the following subsection.

3.2 MSS based on grey filtering

Assume that the additive signal model is appropriate for the noisy speech and that the noisy speech signal is stored in the wave file format whose range is within $(-1,1)$. The diagram block for the proposed magnitude spectral subtraction (MSS) approach based on grey filtering is depicted in Figure 6. Given a noisy speech signal $x_o(k)=s_o(k)+n_o(k)$, the implementation steps are described in the following where additive noise $n_o(k)$ is assumed known and the length of $x_o(k)$ is assumed as a multiple of *L*:

Step 1. Shift up the level of $x_o(k)$ by an appropriate constant *C*, $x_o(k)\leftarrow x_o(k)+C$, such that Condition (i) in Subsection 2.1 is met.

Step 2. Divide $x_o(k)$ into *M* segments of length *L* and each segment is denoted as $x(k)$. Then Steps 3 to 9 are performed for each speech segment $x(k)$.

Step 3. Obtain $X(f)=\mathrm{FFT}_L\{x(k)\}=S(f)+N(f)$ where $\mathrm{FFT}_L\{\cdot\}$ denotes as *L*-point fast Fourier transform (FFT).

Step 4. Estimate additive noise $n(k)$ as $\hat{n}(k)$ by the grey filtering approach as in Subsection 2.2.

Step 5. Perform L-point FFT on $\hat{n}(k)$ to find the magnitude of $\hat{N}(f)$, $\left|\hat{N}(f)\right|$, where $\hat{n}(1)=\hat{n}(2)$ is used.

Step 6. Estimate the standard deviation of $\left|\hat{N}(f)\right|$, $s_{\left|\hat{N}(f)\right|}$.

Step 7. Perform MSS [7] as

$$\left|\hat{S}(f)\right| = \begin{cases} D = \left|X(f)\right| - b\,s_{\left|\hat{N}(f)\right|}, & \text{if } D > 0 \\ 0, & \text{else} \end{cases} \tag{17}$$

where $\left|\hat{S}(f)\right|$ is an estimate of $\left|S(f)\right|$ and b is a scaling factor.

Step 8. Find estimate of $s(k)$ as $\hat{s}(k) = \mathrm{IFFT}_L\{\left|\hat{S}(f)\right| e^{j\angle X(f)}\}$ where $\mathrm{IFFT}_L\{\cdot\}$ is the inverse of L-point FFT.

Step 9. Shift down the level of $\hat{s}(k)$ by the constant C.

Step 10. Concatenate M segments $\hat{s}(k)$ to find estimate of $s_o(k)$, $\hat{s}_o(k)$.

Step 11. Obtain residual noise $n_r(k) = \hat{s}_o(k) - s_o(k)$.

Step 12. Calculate input, output, and improvement signal to noise ratios, SNR_{in}, SNR_{out}, and SNR_{imp}, as follows:

$$SNR_{in} = 10\log\frac{\sum_{k=1}^{M\times L} s_o^2(k)}{\sum_{k=1}^{M\times L} n_o^2(k)} \tag{18}$$

$$SNR_{out} = 10\log\frac{\sum_{k=1}^{M\times L} \hat{s}_o^2(k)}{\sum_{k=1}^{M\times L} n_r^2(k)} \tag{19}$$

$$SNR_{imp} = SNR_{out} - SNR_{in} \tag{20}$$

The block diagram for the proposed MSS approach described above is depicted in Figure 2.

4. Simulation Results

Using MATLAB, the proposed MSS approach depicted in Figure 2 is programmed. The speech file f0101s.wav in [9], a female speech counting from one to ten in English, is used in the simulation. For more details, one may consult in the Appendix 4 of [9]. The sampling rate for f0101s.wav is 10 KHz and the length of samples is 98,000. In the simulation, the speech file f0101s.wav is level-shifted by 5, i.e., $C = 5$ and the segment length is set to 1,000. That is, $L = 1{,}000$ and therefore the number of segments $M = 98{,}000/1{,}000 = 98$. The number of samples used in GM(1,1) modeling is 4, i.e., $K = 4$. And the scaling factor $b = 5$ in (17) is used. In the additive signal model $x_o(k) = s_o(k) + n_o(k)$, the file f0101s.wav shown in Figure 3 is considered as clean speech $s_o(k)$ and the additive noise $n_o(k)$ is artificially generated which is uniform and distributed within the range $g(-0.5, 0.5)$ where g is a scaling factor. The SNR_{in}, SNR_{out}, and SNR_{imp}, for several values of g are given in Table 1. The noisy speech $x_o(k)$ and the enhanced speech $\hat{s}_o(k)$ for $g = 0.4$ are depicted in Figures 4(a) and 4(b), respectively. Simulation results indicate that noisy speeches have been significant improved in terms of SNR, varied from 6.13 dB up to 13.20 dB. Consequently, the objective of speech enhancement is achieved by the proposed MSS approach.

5. Conclusive Remarks

In this paper, the grey filtering approach proposed in [1] is applied to speech enhancement whose noise removal technique is MSS. This paper is motivated by the following observations. For constant signal and random signal, GM(1,1) model has zero or approximately zero estimation error when additive noise is absent and non-zero when additive noise is present. Therefore, in [1] the estimation error of GM(1,1) model is related to additive noise. Next, note that the speech signal generally consists of non-speech and speech portions. The non-speech portion can be considered as constant signal while speech portion as random signal. Thus, an MSS-based speech enhancement approach based on grey filtering is proposed. The proposed MSS approach is justified by simulation where uniform noise is considered. This simulation results indicate that the proposed MSS approach works well for the case.

Acknowledgement

This work was supported by the National Science Council of the Republic of China under grant NSC 90-2213-E-270-006.

References

[1] Cheng-Hsiung Hsieh, "Grey filtering," *Journal of Grey System*, Vol. 13, No. 4, pp. 339-346, 2001.

[2] S. Haykin, *Adaptive Filtering Theory*, 3rd Edition, Prentice Hall, 1996.

[3] J.S. Lim and A.V. Oppenheim, "Enhancement and Bandwidth Compression of Noisy Speech," *Proc. of the IEEE*, Vol. 67, No. 12, pp. 1689-1697, Nov. 1962.

[4] W. Kim and H. Ko, "Noise Variance Estimation for Kalman Filtering of Noisy Speech," *IEICE Trans. Inf. & Syst.*, Vol. E84-D, No. 1, Jan. 2001.

[5] J. Yamauchi and T. Shimamura, "Noise Estimation Using High Frequency Regions for Spectral Subtraction," *IEICE Trans. Fundamentals*, Vol. E85-A, No. 3, Mar. 2002.

[6] L. Singh and S. Sridharan, "Speech Enhancement Using Pre-processing," *Proc. of IEEE*

TENCON, pp. 755-758, 1997.

[7] J. Deng, "Introduction to Grey System Theory," *Journal of Grey System*, Vol. 1, pp. 1-24, 1989.

[8] S.V. Vaseghi, *Advanced Digital Signal Processing and Noise Reduction*, 2nd Edition, John Wiley and Sons Ltd., 2000.

[9] D.G. Childers, *Speech Processing and Synthesis Toolboxes*, John Wiley & Sons, Inc., 1999.

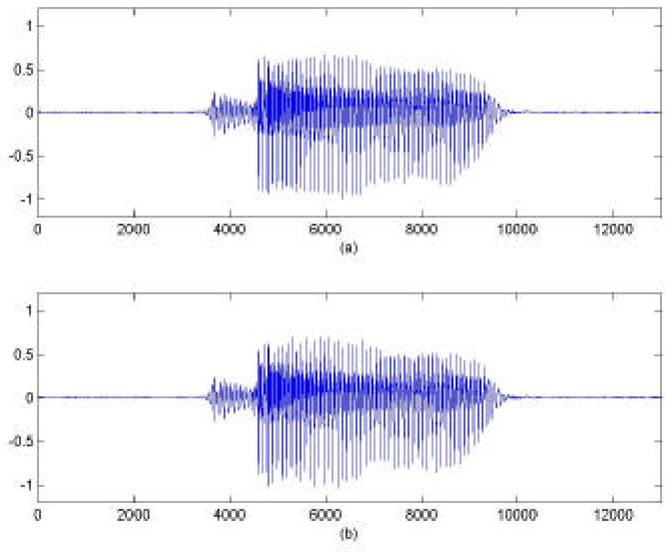

Figure 1. (a) Clean speech b.wav (b) Estimate of b.wav by GM(1,1) model

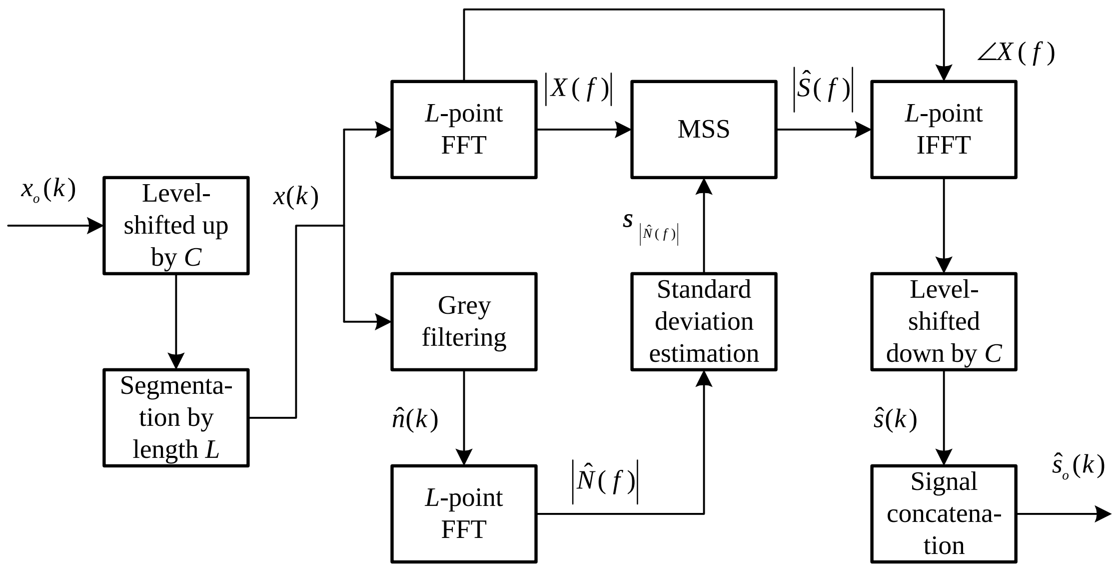

Figure 2. The block diagram for the proposed MSS approach

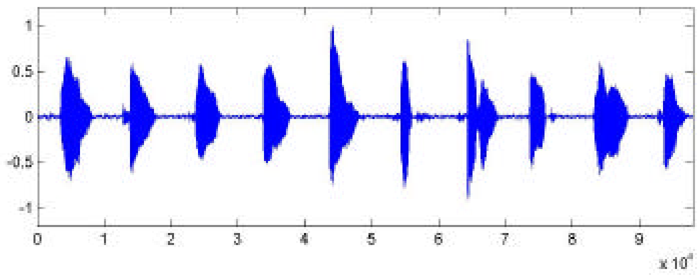

Figure 3. Clean speech of f0101s.wav

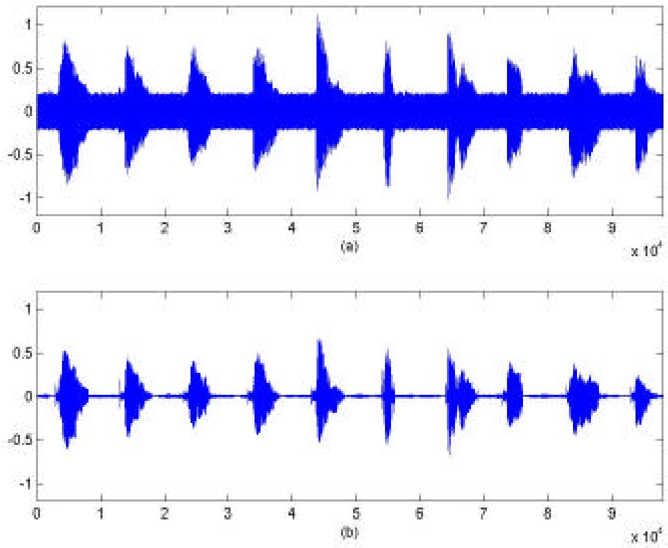

Figure 4. (a) Noisy Speech ($g = 0.4$)   (b) Enhanced speech

Table 1. SNR_{in}, SNR_{out}, and SNR_{imp}, for uniform noise with different g

Value of g	SNR_{in}	SNR_{out}	SNR_{imp}
0.1	5.47	11.87	6.13
0.2	-0.26	9.22	9.49
0.3	-3.79	6.98	10.77
0.4	-6.31	5.60	11.91
0.5	-8.25	4.16	12.41
0.6	-9.83	2.89	12.71
0.7	-11.15	1.76	12.91
0.8	-12.29	0.91	13.20