The Improvement and Implementation of Speech Enhancement Based on Mel frequency Wiener Filtering

Fan Binwen[1, a], Wang Yongjun,[2, b]

[1]Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen 518000, China;

[2]Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen 518000, China;

[a]645067060@qq.com, [b]491614034@qq.com

Keywords: Speech enhancement, Wiener filtering, Short time amplitude spectrum, AGC.

Abstract. The main purpose of speech enhancement is to eliminate the noise in noisy speech signal and extract pure speech signal, which has important significance to improve the performance of digital hearing aid. This paper mainly studies the speech enhancement technology in digital hearing aids. Using the improved second order Mel twisted Wiener filtering algorithm, introduced short-time amplitude spectrum of dynamic decision voice activity detection (VAD) algorithm, which solves the part deviation of estimation stationary noise. At the same time, add a priori SNR gain factor based on decomposition of pure noise frame to increase the degree of inhibition, and contain the speech frame is reduced the extent of the suppression. Update the SNR prediction and low SNR ratio and Wiener filtering gain coefficient, automatic gain control (AGC) effect is obvious. The experimental results show that the output SNR of the processed signal is obviously improved, and the speech intelligibility is good and the quality is high. Significantly improve the recognition ability of weak signal.

Introduction

In the speech enhancement algorithm, the method of adding the Mel frequency domain processing method, which can make the speech more in line with the human ear hearing. The basic theory of Wiener filtering is on the assumption that the input current filter for useful signal and noise, both of which are generalized stationary process, and knowing their second order statistics, Wiener according to the minimum mean square error criterion obtained the parameters of the linear filter, the obtained parameters are designed to filter known as Wiener filters. Wiener filtering is applied to the need to separate the signal from the noise is the whole signal (waveform), rather than its several components. Input hypothesis of the Wiener filter for random signal containing noise, the difference between the desired output and actual output error for the square error, error is small, the effect of filtering noise better. In order to minimize the mean square error, the key is to find the impulse response. If it is able to satisfy the Wiener Hof equation, the filtering effect can be achieved.

Speech enhancement based on Wiener filtering, although the quality of the processed sound be significantly improved, but it retains too much background noise, so that the output SNR decreased. The key is to estimate the accuracy of the noise spectrum. So in this paper, this improved the second order Mel twisted Wiener filtering algorithm and the design basis is two stages signal processing, each stage were Mel warped Wiener filtering and processing, output of the first stage as the input signal of the second stage, and after join based on a priori SNR gain factor, for each frame signal of coefficient of adjustment. A fast noise processing method is designed to improve the accuracy of the algorithm. And the robustness of the algorithm is also improved.

Mel frequency Wiener filtering basic model

In the application of speech enhancement, the input signal y(n) is a band of noise signal, which can be expressed as:

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

The x(n) is the pure speech signal, n(n) is the noise signal, The purpose of Wiener filter is through filtering the input signal, produce the estimation of pure speech signal x(n). We introduce the derivation process of frequency-domain Wiener filter.

In speech recognition and speaker recognition, the common speech feature is based on Mel frequency cepstral coefficients (Mel frequency cepstrum coefficient, MFCC). Due to the MFCC parameters is the human ear auditory perceptual characteristics and speech generation mechanism combination, so the current most speech recognition systems widely used this feature. Here we introduce the derivation process of the Wiener filter in the frequency domain. :

Due to the characteristics of human ears, ear perception for these pure tone frequency are nonlinear, we usually use the Mel scale to show, the relationship between the scale and the actual linear frequency can be approximated as:

$$Mel(f) = 2595 \times \log_{10}(1 + f/700)$$

The calculation formula of triangular window function for the Mel filter:

$$W(k) = \begin{cases} 0 & k < f(m-1) \\ \dfrac{k - f(m-1)}{f(m) - f(m-1)} & f(m-1) \le k \le f(m) \\ \dfrac{f(m+1) - k}{f(m+1) - f(m)} & f(m) \le k \le f(m+1) \\ 0 & k > f(m+1) \end{cases}$$

Among them, m is the Mel frequency, k is digital linear frequency.

After the calculation of Wiener filter, Wiener filter coefficients can be changed Mel domain:

$$H_{Mel}(i) = \frac{1}{\sum_{k=0}^{N} W_i(k)} \sum_{k=0}^{N} W_i(k) \times H(k)$$

The central frequency of the Mel filter is defined as：

$$f(m) = \frac{N}{Fs} B^{-1}(B(f_l) + m\frac{B(f_h) - B(f_l)}{M+1})$$

$$B^{-1}(b) = 700(e^{\frac{b}{1125}} - 1)$$

f_k and f_l are the highest frequency and lowest frequency of the filter group, F_s is the sampling frequency. M is the number of filter banks, N for the FFT transform of the points.

The output of the logarithmic energy for each filter group:

$$S(m) = \ln(\sum_{k=0}^{N-1} | X_a(k) |^2 H_m(k)) \quad ,0 \le m < M$$

The cosine transform MFCC coefficient:

$$C(n) = \sum_{m=0}^{M-1} S(m) \cos(\pi n(m + 0.5)/M) \quad , \quad 0 \le n < M$$

Improvement of Mel's distorted Wiener filtering algorithm

After the Wiener filtering algorithm, the estimation of the speech signal is in line with the characteristics of the human ear. Wiener filtering is sensitive to the speech parameters, but the estimation of the parameters is less accurate. In this paper, The Wiener filtering modified based on using the prior SNR gain factor, and design a fast and effective noise estimation method, the algorithm to ensure voice quality at the same time, to further improve the output signal to noise ratio (SNR). In order to improve the robustness of the algorithm, this paper adopts two steps, as shown in Figure 1. Figure 1 is a Wiener filter denoising module structure, there are two structures are basically the same module cascaded, the first stage of the Wiener filter is used to whiten the non white noise, and the second stage is to remove all traces of white noise.

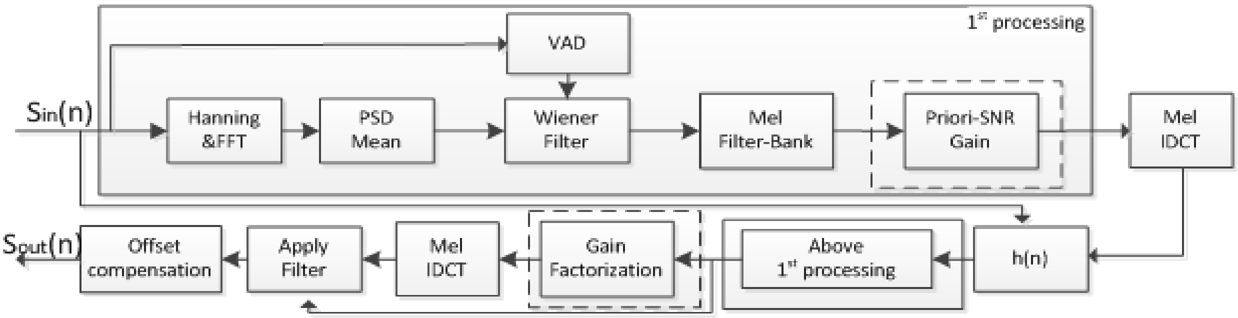

Fig. 1 Schematic diagram of improved algorithm

By the basic theory of Wiener filtering known, Wiener filtering method is looking for a linear filter, recover the target signal from the additive noise interference in the sequence. Reasonable and efficient estimate the noise spectrum is the key to the Wiener filtering algorithm and filtering of the key lies in the calculation the SNR, and the letter of the SNR depending on the effect of VAD. So we need to find out the non speech frames and the noise spectrum estimation. For a short time stable speech signal, VAD can be considered as the background noise from the signal.

Dynamic decision based on short-time amplitude spectrum VAD. The main purpose of the speech activity detection is to distinguish the speech signal from a sound area (pure speech segment or a noisy speech segment) and a quiet area (pure noise or no audio segment).

By the basic knowledge of Wiener filtering, we can know that the logarithmic energy of the last 80 sampling points is:

$$frameEn = 0.5 + \frac{16}{\ln 2} \times \ln\left(\frac{64 + \sum_{i=0}^{M-1} S_{in}(n)^2}{64}\right)$$

For the detection of VAD need to define the position and energy of the segmentation to find the voice frame energy, each frame of the average energy estimates of the signal refreshed according to different conditions of dynamic. Because the energy spectrum parameters is calculate after second operation ,more sensitive of high level part, whereas the amplitude spectrum compared with other feature parameters extraction and much simpler, and the estimated noise effect is good. The algorithm is as follows:

(1)The overlapping speech signals with noise y framing after add window after FFT transform, calculate each frame of speech spectrum:

$$E_m = \sum_{i=0}^{L-1} \left|Y_m(i)\right|$$

In the formula, L is the frame length; m is the frame number; i is the each frame of the voice;

(2)The average amplitude spectrum of the first 40 frames of the noisy speech is E_{mean};

(3) find the first 40 frames values of minimum and maximum of amplitude spectrum $E_{\max}$, $E_{\min}$;

(4) according to the type to determine the threshold

$T = \min\left[0.03\left(E_{\max} - E_{\min}\right) + E_{mean}, 4E_{mean}\right]$

to each frame of the amplitude spectrum compared with the threshold, if is greater than the threshold value judgment for speech frame, else for noise, conversely method is as follows:

$$\hat{N}(k,m) = \begin{cases} Y(k,m) & Y(k,m) < T \\ T & otherwise \end{cases}$$

$\hat{N}(k,m)$noise estimation is obtained by VAD decision frame; K for the frequency.

Gain factoring based on the prior SNR. AGC provide voice level adjustment ability, it has weakened and strengthened the signal is mapped to a user to define the ideal level. In this paper, AGC is a two-way street, and is not only a fixed level of gain or decay. AGC is to distinguish the strengthened the ability of the signals and the soft, avoid this problem, at the same time maintaining coherent and let voice easy to hear the tone of the dialogues. Requirement to distinguish from the voice signal amplification, and only amplified voice without any low level background noise amplification or echo.

After Mel filter coefficient is obtained, this paper adopts a method based on prior SNR gain adjustment coefficient, after Mel filter coefficient is obtained, according to the different speech frame of SNR, places a adjustment coefficient for all frequencies, there is obvious improvement in denoising effect. Gain coefficient of the solving process is shown in figure2.

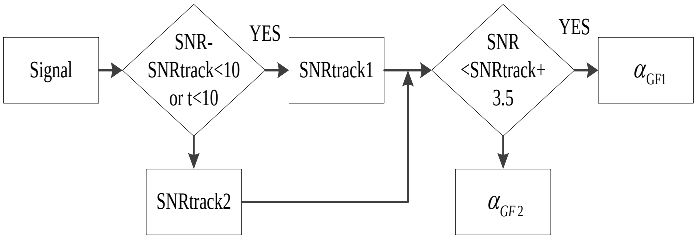

Fig. 2 the process of gain coefficient based on prior SNR

Among them, SNR is the priori SNR, t is the number of frame, SNRtrack is the signal to noise ratio monitoring value, the calculating method is the following formula:

$$SNR_{track1}(t)=\begin{cases}(1-\frac{1}{t})\times SNR_{track}(t-1)+\frac{1}{t}\times SNR(t) & t<10\\ 0.95\times SNR_{track}(t-1)+0.05\times SNR(t) & else\end{cases}$$

$$SNR_{track2}(t)=SNR_{track}(t-1)$$

α_{GF} is the gain coefficient, according to the different judgment results, the method is as follows:

$$\alpha_{GF1}=\alpha_{GF}(t-1)+0.15 \qquad \alpha_{GF2}=\alpha_{GF}(t-1)-0.3$$

α is in the range from 0.1 to 0.8. The following filter coefficients are calculated and finally we can get: $H_{Mel\_GF}(k,t)=(1-\alpha_{GF}(t))+\alpha_{GF}(t)\times H_{Mel}(k,t)$

The gain coefficient is updated by comparing the signal to noise ratio monitoring value and each frame signals of prior SNR, the algorithm is simple. For prior high SNR of speech signal, the value will decrease continuously, the original signal attenuation is small; for the prior low SNR of speech signal, the value will continue to increase. The original signal attenuation is large, so that to achieve the purpose of the dynamic adjustment. Coefficient of GF (t) values of 0.1 to 0.8, this means that in the second stage wiener filtering noise attenuation for voice and frame was 10%, the noise signal frame was 80.

Gain in the second stage is the main function of Factorization of pure noise frame increasing degree of noise suppression, to include voice frame, reduce noise grafting degree. At the same time, the current frame SNR prediction and low SNR, and wiener filtering gain coefficient is updated.

Experiments and results analysis

During the experiment, we adopt different background noises, and recording the pure speech file, add signal to noise in the different background noise, to generate a noisy speech files, all of the audio file sampling rate are 16KHz to speech signal. The first one is the speech signal processing results of different stages of the design algorithm, including AGC processing, and the other is not containing AGC.

After adding the automatic gain, the energy of the signal is changed greatly, and the contrast can be found, not only reduce the noise, but also the energy of the speech signal is amplified. We can get the first 500000 points of the signal in the financial and concrete numerical analysis can be obtained as shown below:

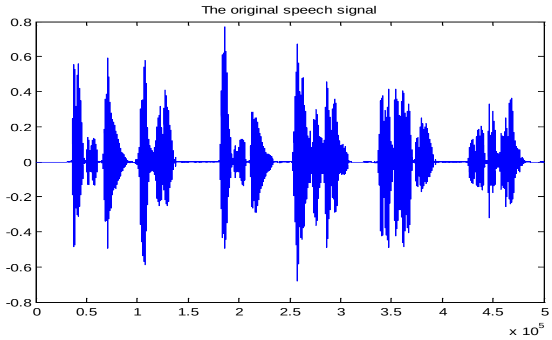

Fig. 3 the original signal

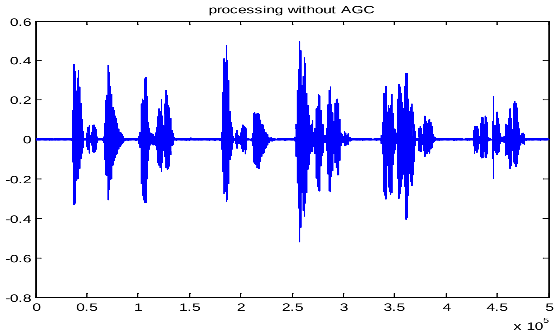

Fig. 5 processing without AGC

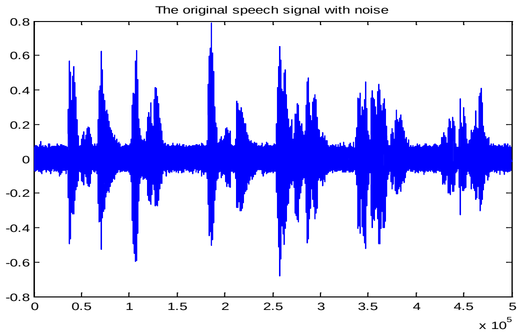

Fig. 4 the original signal with noise

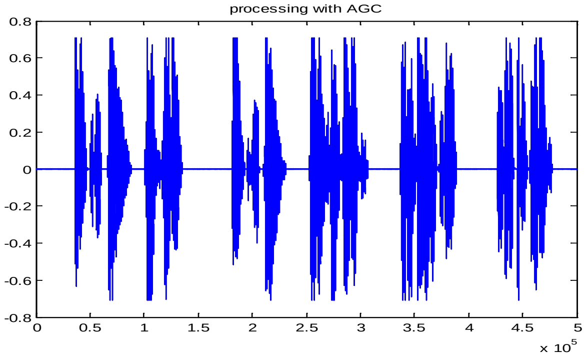

Fig. 6 processing with AGC

In the process of the experiment, we found that the overall quality of the two Mel Wiener filtering process is higher than that of the individual. By adding an improved a priori SNR gain factor, the noise is suppressed, the background noise is reduced, and the SNR is improved. Experiments show that the SNR is improved by 5dB. Therefore, short-time amplitude spectrum dynamic decision of VAD and a priori signal to noise ratio gain factor must be reasonable design combination to improve the quality of speech and dynamic gain not only did not amplify the noise signal, but enlarged powerful voice signal, which in digital hearing aids are very practical.

References

[1] Zhong L, Rafik G, Richard M, Noise estimation using speech/non-speech frame decision and subbed spectral tracking[J], Speech Communication, 2008, 49(7): 542-557.

[2] Zhi T, He M. Z, Noise reduction in whisper speech based on the auditory masking model[C], International Conference on Information, Network and Automation, 2010, 2: 272-277.

[3] Ch. V. Rama Rao, M. B. Rama Murthy, K. Srinivasa, Speech enhancement using a modified a priori SNR and adaptive spectral gain control[J], IEEE International Journal of Computer Applications, 2011, 12(12): 13-17.

[4] Hu Y. and Loizou P.(2006). Subjective comparison of speech enhancement algorithms. Proc.IEEE Int.Conf.Acoust., Speech, Signal Processing, I,153-156.

[5] Raúl Vicen-Bueno, Almudena Martínez-Leira. Modified LMS based feedback reduction subsystems in digital hearing aids based on WOLA filter bank[J]. IEEE Trans, 2009, 58(09): 3177-3189.