Computer Life (CPL) ISSN: 1819-4818
DELIVERING QUALITY SCIENCE TO THE WORLD

Research on Low SNR Speech Endpoint Detection Algorithm

Tingyu Yang [a], Hongbo Wang [b], Chenghua Fu [c]

School of Automation& Information Engineering, Sichuan University of Science& Engineering,

Zigong 643000, China

[a]1274720592@qq.com,[b]1920266248@qq.com,[c]1303676825@qq.com

---

Abstract: Under low SNR conditions, speech detection performance drops dramatically so that it is impossible to distinguish between noise and speech caused by false detection. In this paper, a new speech endpoint detection algorithm combining improved Wiener filtering with multi-window spectrum estimation and improved MFCC cepstrum distance is proposed. Firstly, the wavelet threshold multi-window power spectrum estimation method is used to reduce the variance of the speech power spectrum and the noise power spectrum, and the a priori SNR is smoothed. The Wiener filtering method is used to filter and denoise the noise. Finally, the two algorithms are combined to enhance the speech and noise discrimination improving the voice signal to noise ratio. Then the traditional short-term MFCC cepstrum distance threshold is improved to adapt to different signal-to-noise ratios. Finally, the endpoint detection is combined with the MFCC cepstral distance method. Through MATLAB simulation, the article improves the detection accuracy under low SNR conditions compared with other endpoint detection algorithms, and has strong stability and practicability.

Keywords: Endpoint detection, Multi-window spectrum estimation Wiener filtering, MFCC cepstrum distance, Threshold threshold.

---

1. INTRODUCTION

Speech endpoint detection(Video Activity Detection,VAD)is a technique for dividing a start and end points of a speech signal to separate speech segments from noise segments. As part of the speech recognition front-end signal processing technology, accurate endpoint detection can reduce subsequent noise processing, correctly extract speech signal features, and improve speech feature recognition accuracy. Endpoint detection in noisy environments is more complex than endpoint detection in quiet environments. Detection of speech endpoints in low SNR environments often leads to missed detection, thus attracting the attention of researchers. The classical speech endpoint detection algorithm is divided into time domain parameters and frequency domain parameters [1], in which time domain parameters have short-term average energy, short-term average zero-crossing rate [2], etc. Frequency domain parameters have correlation method, frequency band variance method, spectral distance method and spectral entropy method [3-4]. The time domain parameter method

algorithm is relatively simple and rapid, but the detection result is poor under low SNR conditions, and the frequency domain parameter method has higher precision, but the computational complexity and computation amount are relatively strong. The traditional speech endpoint detection algorithm has better detection results under higher SNR conditions, but the quality of endpoint detection is significantly lower after reducing the signal-to-noise ratio, and some sounds cannot be detected. Even at a signal-to-noise ratio of 10 dB, the endpoint detection results are not satisfactory. Therefore, an algorithm that combines the noise reduction of the speech signal with the classical speech endpoint detection algorithm has emerged. As in the literature [5], an improved information entropy algorithm is proposed to improve the speech noise immunity. In [6], the spectral subtraction method and the short-time zero entropy ratio method are used to denoise the noisy speech by power spectrum subtraction, and then the endpoint detection is performed The literature [7] proposes a speech enhancement algorithm based on the deep confidence network to combine with the traditional endpoint detection algorithm. All of the above studies have proved that the detection accuracy is improved to some extent by performing noise reduction processing on the noisy speech and then performing endpoint detection.

Wiener filtering [8] is an optimal estimator for stationary processes based on minimum mean square error criterion. The mean square error between the output of this filter and the desired output is minimal, so it is an optimal filtering system. It can be used to extract signals that are contaminated by stationary noise. In this paper, the Wiener filtering method is improved. The improved Wiener filtering method using multi-window spectrum estimation [9] improves the speech signal-to-noise ratio after noisy speech noise reduction, and then uses an improved MFCC cepstrum distance for noise-reduced speech. The detection algorithm performs endpoint detection on the speech. The simulation results by MATLAB show that the improved method is more robust and accurate than the traditional method in the low SNR environment.

2. WIENER FILTERING NOISE REDUCTION

2.1 Wiener filtering method and its improvement

The Wiener filter is a linear filter that has a wide range of applications, regardless of whether the speech signal is continuous or not in a stationary random process. Assume that the input noisy speech signal is $y(n) = s(n) + d(n)$,Where s(n) is a pure speech signal; d(n) is a noisy signal. The signal we collected is only with noise y(n). The essence of Wiener filtering is to design a digital filter h(n). When the input signal y(n) is input to the filter, the output signal will be

$$\overset{\Lambda}{S}(n) = y(n) * h(n) = \sum_{m=-\infty}^{+\infty} y(n-m)h(m) \tag{1}$$

$\overset{\Lambda}{s}(n)$ can use the minimum mean square error criterion to get the minimum mean square error of $s(n)$ and $\overset{\Lambda}{s}(n)$ ($\varepsilon = E\{[s(n) - \overset{\Lambda}{s}(n)]^2\}$).

The orthogonality theorem is used, and the following conditions are required:

$$E[\{s(n) - \overset{\Lambda}{s}(n)\} \bullet y(n-m)] = 0 \tag{2}$$

The Wiener filter estimator $H(k)$ can be derived by substituting the formula 1 into the equation 2 and performing the Fourier transform:

$$H(k)=\frac{p_{sy}(k)}{p_y(k)} \tag{3}$$

In the above formula, $p_y(k)$ is the power spectral density of $y(n)$; $p_{sy}(k)$ is the mutual power spectral density of s(n) and y(n). Since the speech signal s(n) is not correlated with the noise signal d(n), Rsd(m)=0, and then derived

$$p_{sy}(k)=p_s(k) \tag{4}$$

$$p_y(k)=p_s(k)+p_d(k) \tag{5}$$

Further simplify the formula 3 to

$$H(k)=\frac{p_s(k)}{P_s(k)+p_d(k)} \tag{6}$$

Then use H(k) to calculate the estimate of the speech spectrum of $\overset{\Delta}{s(n)}$ in the frequency domain.

$$S(k)=\hat{H}(k)\bullet Y(k) \tag{7}$$

In the above formula, Y(k) is the spectral value of the noisy speech at the corresponding frequency point. Because in real life, speech is a short-term stationary signal and the speech power spectrum cannot be obtained, and thus

$$H(k)=\frac{E[|S(k)|^2]}{E[|S(k)|^2]+\lambda_d(k)} \tag{8}$$

The above equation 8 is divided by λd(k) and simultaneously obtained by λd(k).

$$H(k)=\frac{\xi(k)}{1+\xi(k)} \tag{9}$$

$$H(k)=1-\frac{1}{\gamma(k)} \tag{10}$$

Among them, the definition $\xi(k)=\frac{E[|S(k)|^2]}{\lambda_d(k)}$ is the a priori signal-to-noise ratio; $\gamma(k)=\frac{|Y(k)|^2}{\lambda_d(k)}$ is the a posteriori signal-to-noise ratio.

Improve the above formula and add adjustable parameters. α, β, make the transfer function H(k) controllable, and the above formula is changed to

$$H(k)=(\frac{\xi(k)}{\alpha+\xi(k)})^{\beta} \tag{11}$$

$$H(k)=(1-\frac{\alpha}{\gamma(k)})^{\beta} \tag{12}$$

This paper takes α=2.7, β=0.7.

Further introducing a smoothing parameter a, deriving

$$\begin{aligned}\xi_i(k)&=a\xi_i(k)+(1-a)\xi_i(k)\\&=a\xi_i(k)+(1-a)(\gamma_i(k)-1)\\&\approx a\xi_{i-1}(k)+(1-a)(\gamma_i(k)-1)\end{aligned} \tag{13}$$

In the above formula, the subscript i represents the i-th frame. From the above equation, we can see that the a priori SNR of the i-th frame and the a posteriori signal-to-noise ratio of the i-th frame can be obtained. By verifying the signal-to-noise ratio, we can get the transfer function of the Wiener filter by repeatedly deriving:

$$H_i(k) = \frac{\overset{\wedge}{\xi_i}(k)}{\overset{\wedge}{\xi_i}(k) + 1} \tag{14}$$

We can further derive the spectrum estimation of the speech signal of the ith frame:

$$\overset{\wedge}{S_i}(k) = H_i(k) Y_i(k) \tag{15}$$

2.2 Multi-window spectrum estimation improved Wiener filtering algorithm

Multitaper Spectrum estimation [10] was proposed by Thomson in 1982. The estimation algorithm uses several orthogonal data windows for the same data sequence to calculate the corresponding direct spectra, and then uses these direct spectra to obtain the average. The value obtains a smaller estimated variance, and the process yields a more accurate spectral estimate [11].

The definition of multi-window spectrum is as follows:

$$S^{mt}(w) = \frac{1}{L} \sum_{k=0}^{L-1} S_k^{mt}(w) \tag{16}$$

Where L is the number of data windows; S_k^{mt} represents the spectrum of the Kth data window, which can be expressed by:

$$S_k^{mt}(w) = \left| \sum_{n=0}^{N-1} a_k(n) x(n) e^{-jnw} \right|^2 \tag{17}$$

Where x(n) represents the data sequence; N is set to the length of the sequence; ak(n) represents the Kth data window, which conforms to the orthogonal relationship between several data windows:

$$\begin{cases} \sum a_k(n) a_j(n) = 0 & k \neq j \\ \sum a_k(n) a_j(n) = 1 & k = j \end{cases} \tag{18}$$

Among them, we define the data window as a set of mutually orthogonal discrete ellipsoidal sequences (DPSS).

In the improved Wiener filtering algorithm for multi-window spectrum estimation, the multi-window spectrum power estimation function pmtm in the toolbox of MATLAB is used to calculate the power spectral density estimation value and the Wiener filtering algorithm to realize the speech denoising operation. The specific steps are as follows:

Step 1 sets the noisy speech to x(n), and performs windowing and framing to obtain a xi(m) sequence, in which two adjacent frames are superimposed.

Step 2 Perform Fourier transform on xi(m), find the amplitude spectrum |Xi(k)| and the phase spectrum θi(k), and smooth the adjacent frames to obtain the average amplitude spectrum :

$$|\overline{X_i}(k)| = \frac{1}{2M+1} \sum_{j=-M}^{M} |X_{i+j}(k)| \tag{19}$$

Take M frames from the left and right of the origin to obtain 2M+1 frames and then find the average value. In practical applications, M=1 is usually used to obtain the average value of the three frames.

Step 3 Multi-window spectrum power spectral density P(k,i) is obtained by multi-window spectrum estimation for xi(m) sequence

$$P(k,i) = PMTM[x_i(m)] \tag{20}$$

In the above formula, k represents a spectral line, i represents a frame, and PMTM represents a function.

Step 4 For the P(k,i) smoothing process obtained by Step3, calculate the smooth power spectral density Py(k,i):

$$P_y(k,i) = \frac{1}{2M+1}\sum_{j=-M}^{M} p(k,i+j) \tag{21}$$

Step 5 Calculate the average power spectrum of the speech-free segment (NIS), ie, the noise segment, by the speech signal:

$$P_n(k) = \frac{1}{NIS}\sum_{i=1}^{NIS} P_y(k,i) \tag{22}$$

For the above value, it is the noise power spectrum λd(k).

Step 6 The a posteriori signal-to-noise ratio $\gamma_i(k)$ and the a priori signal-to-noise ratio $\xi_i(k)$ are calculated according to the above equation, and the Wiener filter transfer function Hi(k) is calculated.

Step 7 Calculating the output amplitude spectrum The $\hat{S}_i(k)$ phase spectrum θi(k) is subjected to IDFT conversion to replace the frequency domain signal into a time domain signal. Finally, the noise-reduced speech signal $\hat{S}_i(m)$ is obtained:

$$\hat{S}_i(m) = IDFT\{\hat{S}_i(k)\exp[j\theta_i(k)]\} \tag{23}$$

3. IMPROVED MFCC CEPSTRUM DISTANCE ENDPOINT DETECTION ALGORITHM

3.1 Improved MFCC cepstrum distance calculation

The MFCC is based on the auditory model and is obtained by the Delta frequency filtering process and the DCT cepstrum change. The Mel frequency can simulate the human ear hearing frequency, which has the following relationship with the actual sound frequency f:

$$Mel(f) = 1125\log(1+f/700) \tag{24}$$

The steps to improve the MFCC cepstrum distance are as follows:

Step 1 The above-mentioned noise-reduced speech $\hat{S}_i(m)$ is windowed and framed, and then subjected to fast Fourier transform to convert the time domain signal into the frequency domain signal X(i,k), and is calculated as a non-speech frame (about 15 frames) 25 ms before the speech. Average noise energy:

$$D(k) = \frac{1}{15}\sum_{i=1}^{15} |X(i,k)|^2 \tag{25}$$

Where $E(i,k) = |X(i,k)|^2$ represents the spectral line energy per frame; i, k represents the number of frames and spectral line values, respectively.

Step 2 Reduce the noise by subtracting the noise energy value from the spectral line energy value of each frame to obtain a relatively pure spectral line energy $\hat{E}(i,k)$:

$$\hat{E}(i,k) = \begin{cases} E(i,k) - a\times D(k), E(i,k) \geq a\times D(k) \\ b\times D(k), E(i,k) < a\times D(k) \end{cases} \tag{26}$$

In the above formula, a and b are constants, a is an over-subtraction factor, and b is a gain compensation factor. This paper takes a=4, b=0.001.

Step 3 Through a set of Mel-scale triangular filters, the center frequency of each filter is between f(m), m = 1, 2, 3, ..., M. M is the number of filters in this group, also called the order, usually between 24 and 40. Since the human ear has different degrees of perception of low frequency and high frequency, when setting the triangular filter bank, the selection is from dense to sparse, and the distance between each f(m) decreases as the value of m decreases, with m Increase and become larger. The transfer function of each filter is as follows:

$$H_m(k)\begin{cases} 0, & k < f(m-1) \\ \dfrac{2(k-f(m-1))}{(f(m+1)-f(m-1))(f(m)-f(m-1))}, & f(m-1) \le k \le f(m) \\ \dfrac{2(f(m+1)-k)}{(f(m+1)-f(m-1))(f(m+1)-f(m))}, & f(m) \le k \le f(m+1) \\ 0 & k \ge f(m+1) \end{cases}$$

f(m) in the above formula is determined by the following formula：

$$f(m)=(\frac{N}{f_s})F_{mel}^{-1}(F_{mel}(f_l)+\frac{F_{mel}(f_h)-F_{mel}(f_l)}{M+1}) \tag{27}$$

Where fl and fh are the lowest and highest frequencies of the Mel filter, N is the DFT length, fs is the sampling frequency, and F_{mel}^{-1} is the inverse of F_{mel} .

Step 4 The energy spectrum passes through the triangular filter bank to obtain the logarithmic energy.

$$E(m)=\ln(\sum_{k=0}^{N-1}\Lambda E(i,k)H_m(k)),0 \le m \le M \tag{28}$$

The obtained logarithmic energy is subjected to discrete cosine transform to obtain an L-order MFCC characteristic parameter.

$$C_l(n,j)=\sum_{m=0}^{M}E(m)\cos(\frac{\pi l(m-0.5)}{M}),0 \le l \le L \tag{29}$$

Step 5 The preamble speech-free frame NIS is used to obtain the MFCC cepstral feature parameter mean Cl(j) as the cepstrum coefficient of the noise signal, and then the improved MFCC cepstrum distance value is calculated:

$$d_{mfcc}(i)=\sqrt{\sum_{n=1}^{p}(C_l(i,j)-C_l(j))^2} \tag{30}$$

3.2 Threshold Threshold Estimation and Endpoint Detection Algorithm

The threshold threshold of the endpoint detection of the traditional MFCC cepstrum distance [12] is a fixed coefficient product relationship, but this fixed threshold threshold is difficult to adapt to different signal-to-noise ratios. Therefore, a new dynamic threshold is proposed in this paper. The calculation steps are as follows:

Step 1 Calculate the new log line energy according to [13]:

$$LE(i)=\lg(E(i,k)+a)-\lg a \tag{31}$$

In the above formula, a is a constant, which is 1 in this paper.

Step 2 The log line energy value obtained by Step1 is multiplied by the improved MFCC cepstrum distance value and smoothed to obtain a new parameter value LD(i).

$$LD(i) = smooth(LE(i) \times d_{mfcc}(i)) \tag{32}$$

Step 3 Calculate the average value MLD of LD(i), and then find the initial threshold thresholds T1, T2, which are calculated as follows:

$$MLD = \sum_{i=1}^{NIS} LD(i) \tag{33}$$

$$\begin{cases} T_1 = a \times MLD + p\delta \\ T_2 = b \times MLD + p\delta \end{cases} \tag{34}$$

Where δ represents the standard deviation and can be obtained by:

$\delta^2 = \frac{1}{NIS}\sum_{i=1}^{NIS}(LD(i) - \frac{MLD}{NIS})^2$, In the formula, NIS takes 15 and corresponds to 25 ms; a, b is the upper and lower limit coefficients which can be measured experimentally, and p takes a value of 3. In order to make the measurement results more accurate, the values of T1 and T2 are updated by the following formula.

$$\begin{cases} T_1 = T_1 \times \partial + LD(i)(1-\partial) \\ T_2 = T_2 \times \partial + LD(i)(1-\partial) \end{cases} \tag{35}$$

among them, $\partial = 0.95$.

Step 4 The speech endpoint is detected using a single parameter double threshold threshold [14], when LD(i)>T2 is determined as speech, then when LD(i)>T1 speech frame starts, until LD(i)<T1 ends the speech frame, thus The speech segment and the noise segment are detected.

4. SIMULATION EXPERIMENTS AND ANALYSIS

In this paper, the algorithm is verified by MATLAB simulation software. The speech material used in the experiment is a pure speech "blue sky, white clouds, turquoise sea" recorded by Cool Edit Pro under quiet laboratory environment. The pure speech is added with a signal-to-noise ratio of -10dB, -5dB, 0dB, 5dB of noise, and the noise is selected from white noise and Babble noise in the NOISEX-92 standard noise library. Experiments are carried out by adding white noise speech and babble noise speech without noise reduction. Figure 1 is an unreduced noise white noise endpoint detection map, where the red dotted line represents the detected speech start point and the green solid line represents the detected speech endpoint.

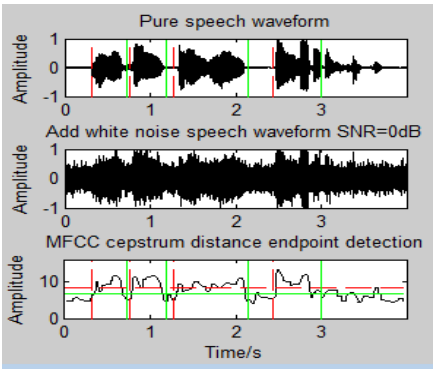

Fig.1. SNR=0dB unnoise white noise endpoint detection

It can be seen from Fig. 1 that when the noise signal-to-noise ratio is 0 db and the noise reduction processing is not performed on the noisy speech, the endpoint detection of the short-term MFCC cepstral distance is directly detected, and the words “sea” cannot be detected. The multi-window spectrum improved Wiener filtering algorithm is used to denoise the noisy speech, and the classical

short-time MFCC cepstrum distance method and the improved algorithm endpoint detection method are used to compare the noisy speech. The results are shown in Figures 2-4.

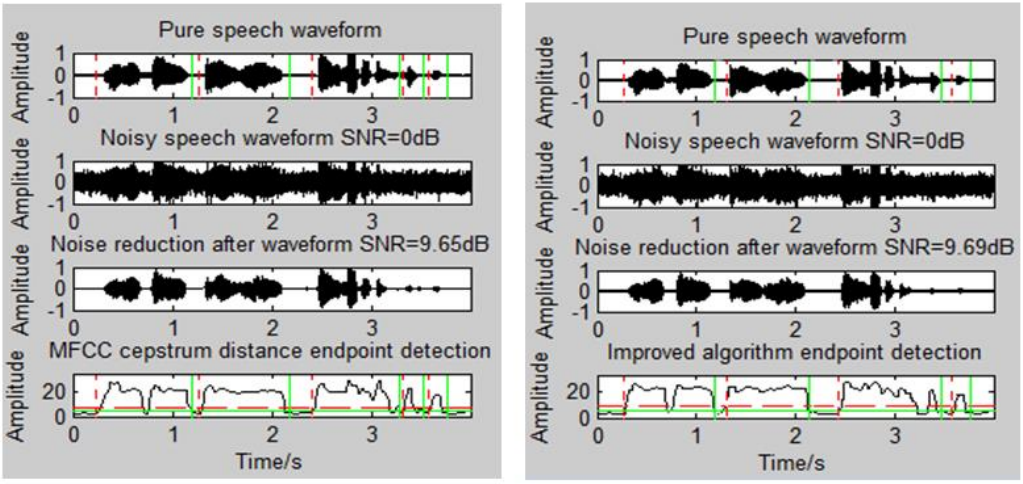

Fig.2. SNR=0dB endpoint detection in white noise environment

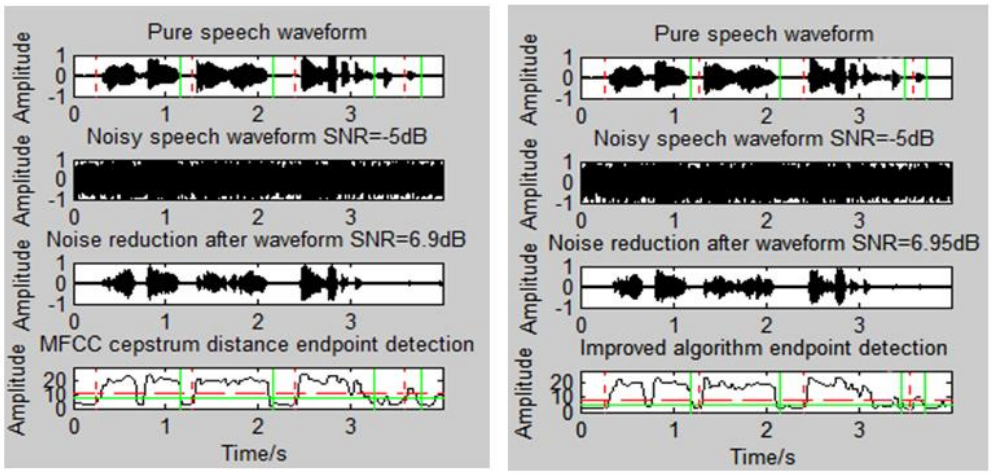

Fig.3. SNR=-5dB endpoint detection in babble noise environment

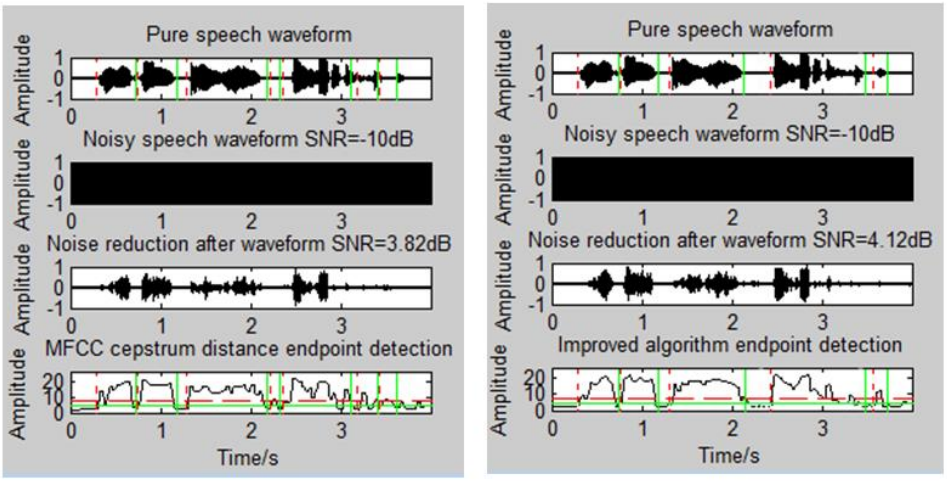

Fig.4. SNR=-10dB endpoint detection in white noise environment

It can be seen from Fig. 2-4 and Table 1 that the short-time MFCC cepstral distance method (A) without noise reduction processing and the short-time MFCC under the premise of noise reduction by multi-window spectrum estimation improved Wiener filtering method The cepstrum distance method

(B) and the endpoint detection method (C) of the improved algorithm are compared. It is found that the improved algorithm has the highest accuracy. When the noise signal is at 0dB, the A method begins to detect falsely; the B method and the C method can both. The starting point of the speech is detected, but the accuracy of the former method is gradually lowered as the signal-to-noise ratio is lowered, and the accuracy of the C method is still higher. When the signal is in the white noise environment with signal-to-noise ratio of -10dB, the B method is not detected and the C method can detect it accurately. This proves the practicability of the improved algorithm.

Table 1 Comparison of speech endpoint detection accuracy

Endpoint detection method	White noise				Babble noise			
	5dB	0dB	-5dB	-10dB	5dB	0dB	-5dB	-10dB
Method A	91.8	87.6	71.7	Invalid	85.2	76.9	61.7	Invalid
Method B	93.5	91.2	83.4	75.3	89.7	84.9	79.7	Invalid
Method C	98.4	95.1	86.5	81.6	94.9	90.6	85.2	76.7

The A method, which is a classic short-time MFCC cepstrum distance endpoint detection method, has weak anti-noise performance because it uses the average of the MFCC cepstral coefficients as the background noise MFCC cepstrum coefficient by using the leading speechless frame as the background noise frame. The estimated value, if the noise fluctuates strongly, its estimated value will deviate from the ideal value so that the endpoint cannot be detected correctly. The improved algorithm firstly uses the improved Wiener filtering algorithm of multi-window spectrum estimation to denoise the noisy speech, thus overcoming the weak anti-noise ability of the MFCC cepstrum distance in the noise environment with SNR of 0~10dB. Compared with the B method, the improved C method has strong anti-noise ability, and can detect the endpoint correctly under a certain signal-to-noise ratio. However, when the signal-to-noise ratio is lower than -5dB, the speech endpoint detection will be disordered. Furthermore, in the case where the signal-to-noise ratio is low, the threshold thresholds T1 and T2 of the endpoint detection are appropriately adjusted to satisfy the endpoint detection of the adaptive signal-to-noise ratio. Therefore, the improved C method further improves the anti-noise ability of the A method. When the pure speech is at a lower signal-to-noise ratio, the threshold thresholds T1 and T2 will be adaptively adjusted to make the endpoint detection result more accurate.

5. CONCLUSION

When endpoint detection is performed on noisy speech, when the speech signal-to-noise ratio is low, the detection performance of the A method is drastically reduced or even the speech cannot be detected. In real life, speech is often in a noisy environment, so there is a certain value for endpoint detection of low SNR speech. In this paper, the noise reduction processing of low SNR speech is performed by LMS adaptive filtering [15], spectral subtraction noise reduction [16], Wiener filtering method noise reduction [17-18] and the improvement of multi-window spectrum in this paper. Wiener filtering method for noise reduction, found that the improved Wiener filtering method with multi-window spectrum has the best effect on low SNR speech processing. The noise reduction of the noisy speech will be caused by the noise reduction after the noise reduction. The improved algorithm will reduce the false detection caused by the speech distortion to some extent. The simulation results

of MATLAB show that compared with the classical short-time MFCC cepstrum distance and the MFCC cepstrum distance method after improving the Wiener filtering denoising algorithm through multi-window spectrum, the improved algorithm has stronger anti-noise ability and more stability. High, the accuracy has also reached a certain level of improvement, so it has a strong practicality.

REFERENCES

[1] Liu Huan, Wang Jun, Lin Qiguang, Wang Shitong. A new method for speech endpoint detection based on the fusion of time domain and frequency domain features [J].Journal of Jiangsu University of Science and Technology (Natural Science Edition), 2017, 31 (01): 73-78.

[2] Song Zhiyong. MATLAB Speech Signal Analysis and Synthesis (Second Edition) [M]. Beijing: Beijing University of Aeronautics and Astronautics Press, 2017.

[3] Zhang Chao. Research on speech endpoint detection method [D]. Dalian University of Technology, 2016.

[4] WANG Wei, HU Guiming, YANG Li, HUANG Dongfang, ZHOU Yang.Endpoint Detection Based on Spectral Subtraction and Uniform Subband Band Variance[J]. Electroacoustic Technology, 2016, 40(05): 40-43+66.

[5] Xuan Zhangjian, Cai Xiaoxia, Zhai Dingli. Research on a speech endpoint detection method based on improved information entropy[J].Communication Technology, 2018, 51(06): 1302-1306.

[6] Wu Peng,Zhang Xiaobing,Ding Wu.Analysis of speech enhancement based on spectral subtraction and Wiener filtering[J].Computer Applications and Software,2017,34(03):67-70+118.

[7] Chen Yingying,Bi Chunyan,Long Jianzhong.Study on speech endpoint detection algorithm under low SNR[J].Tv Engineering,2018,42(06):9-12+27.

[8] Cai Ping.A Speech Enhancement Algorithm Based on Wiener Filter[J].Guangdong Communication Technology,2016,36(06):63-66.

[9] ZHANG Qing,WU Jin.A Speech Endpoint Detection Algorithm Based on Improved Entropy Ratio Method for Multi-Window Spectral Estimation[J].Journal of Chaohu University,2016,18(06):80-85.

[10] Thomson D J.Spectrum estimation and harmonic analysis[J].IEEE, 1982, 70 (9) :1055-1096.

[11] ZHAO Fa.A Speech Endpoint Detection Algorithm Based on Multi-Window Spectral Estimation Spectral Subtraction and Entropy Ratio Method[J].Journal of Chaohu University,2016,18(06):80-85.

[12] Zhang Tao,Zhang Xiaobing,Zhu Mingxing.Improved cepstrum distance speech endpoint detection algorithm in low SNR environment[J].Electronic Acoustic Technology,2017,41(Z2):108-112+125.

[13] Zeng Shuhua,Lü Jingxiang,Nie Xiaowu.A Speech Endpoint Detection Method Based on MFCC Cepstral Distance and Logarithm[J].Electronic Acoustic Technology,2016,40(09):51-55.

[14] HAN Fang, ZHAI Zong-xin. Research on Endpoint Detection Algorithm Based on Low Signal-to-Noise Ratio[J]. Journal of Northwest Normal University(Natural Science), 2016, 52(05): 55-59.

[15] Sun Jing, Tao Zhi, Gu Jihua, Zhao Heming. Research on Ear Speech Enhancement Based on LMS Adaptive Filtering[J].Communications Technology,2007(12):394-396.

[16] Jin Xuedong,Li Dongxin.Improved algorithm for speech signal denoising based on spectral subtraction[J].Foreign Electronic Measurement Technology,2018,37(05):63-67.

[17] Li Zhanming,Shang Feng.A Wiener Filtering Speech Enhancement Algorithm Based on Speech Endpoint Detection[J].Electronic Design Engineering,2016,24(02):42-44.

[18] Bao Wujie,Huang Hao.Voice Endpoint Detection Based on Speech Enhancement Method[J].Modern Electronic Technique,2017,40(22):1-4+9.