



# Research on Speech Endpoint Detection Algorithm with Low SNR

Jian Wei, Xiang'e Sun

College of Electronics and Information Engineering, Yangtze University, Jingzhou, China

Email: 1553766394@qq.com

**How to cite this paper:** Wei, J. and Sun, X.E. (2017) Research on Speech Endpoint Detection Algorithm with Low SNR. *Open Access Library Journal*, 4: e3487. <https://doi.org/10.4236/oalib.1103487>

**Received:** March 2, 2017

**Accepted:** March 27, 2017

**Published:** March 30, 2017

Copyright © 2017 by authors and Open Access Library Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

The detection of speech endpoint is an important application for speech signal processing. Although there are variance methods, the endpoint can't be detected accurately in low SNR (Signal to Noise Ratio). The paper points out an endpoint detection algorithm combining two methods together: the one is improved spectral subtraction based on multitaper spectral estimation, and the other is BARK subband variance in frequency domain. Firstly, the noisy speech signal is processed through the improved spectral subtraction based on multitaper spectral estimation. It can achieve the purpose of noise reduction through this step. Then the noisy speech signal is detected using the method of BARK subband variance in frequency domain. Compared with the common endpoint detection algorithm, it is concluded that endpoint detection accuracy by new method can be improved in low SNR.

## Subject Areas

Multimedia/Signal Processing

## Keywords

Endpoint Detection, Multitaper Spectral Estimation, Improved Spectral Subtraction, BARK Subband Variance

## 1. Introduction

Speech is the most natural form of human-human communications and is related to human physiological capability. And it is the most important, the most effective and most convenient form of information exchange. Speech signal processing is a comprehensive subject and a popular research field, which involves a wide range of content [1]. Speech endpoint detection which can identify the starting point and endpoint of speech signal accurately also plays an important role in speech signal processing. So, it is widely used in speech coding,

speech recognition, speech enhancement and echo cancellation [2] [3]. Sun *et al.* presented an improved double threshold method to detect speech endpoint. The experiment results show that the accuracy of endpoint detection can be improved greatly. When the SNR is very low (about 5 dB), the research needs to be continued [4]. Wang *et al.* proposed a new speech endpoint detection algorithm based on combination spectrum variance with spectral subtraction. It has better detection capability and good adaptability and anti-noise ability in the case of low SNR [5]. Haigh and Mason proposed the method voice activity detection using cepstral features [6]. In Ref. [7], the authors presented a method combined Logarithmic Energy with Spectral Entropy to detect the endpoint. EN De *et al.* proposed a new measure of time series complexity, fuzzy entropy, and applied it to the characterization of speech. Then dual threshold method is used to detect endpoint [8]. Zheng *et al.* make a comparative study for detection method including the double threshold method, frequency variance method and information entropy. The experiment shows that three methods can't meet the requirements in the case of low SNR [9].

Though there are a lot of methods to detect speech signal endpoint, it is a challenging research for low SNR condition. With the help of the bark wavelet transforming knowledge [10] and spectral subtraction based on multitaper spectral estimation [11], a speech signal endpoint detection algorithm is proposed in this paper. The algorithm combines the improved spectral subtraction based on multitaper spectral estimation with BARK subband variance in frequency domain. The improved spectral subtraction based on multitaper spectral estimation can effectively suppress noise. The application of BARK wavelet transforming can make us know that it is also a method of endpoint detection. The BARK subband theory is combined with the frequency variance method and it is an extension of the frequency variance. With the help of BARK subband variance in frequency domain, the endpoint can be better detected.

The simulation experiment and result analysis are carried out by using MATLAB software. The results show that this method can improve the accuracy for the situation of low SNR.

## 2. The Improved Spectral Subtraction Based on Multitaper Spectral Estimation

### 2.1. Multitaper Spectral Estimation Introduction

Multitaper spectral estimation was proposed by Thomson in 1982. While the traditional periodogram method only uses one data window for the same data sequence, the multitaper spectral proposed by Thomson uses multiple orthogonal data windows to obtain the direct spectrum for the same data sequence. Then the spectrum estimation could be got by seeking average of the direct spectrum before. Therefore, it is possible to obtain smaller estimation variance [12] [13].

The definition of multitaper spectral is shown as follows:

$$S^{mt}(\omega) = \frac{1}{L} \sum_{k=0}^{L-1} S_k^{mt}(\omega) \quad (1)$$

where  $L$  expresses the number of data window.  $S^{mt}$  means the spectrum of the  $k^{\text{th}}$  data window:

$$S_k^{mt}(\omega) = \left| \sum_{n=0}^{N-1} a_k(n) x(n) e^{-jn\omega} \right|^2 \quad (2)$$

where,  $x(n)$  expresses data series.  $N$  expresses the length of series.  $a_k(n)$  means the  $k^{\text{th}}$  data window which satisfies the mutually orthogonal. Its formula is expressed as follows:

$$\begin{cases} \sum a_k(n) a_j(n) = 0 & k \neq j \\ \sum a_k(n) a_j(n) = 1 & k = j \end{cases} \quad (3)$$

The data window is a set of discrete ellipsoidal sequences with mutually orthogonal and also known as Slepian windows.

## 2.2. The Improved Spectral Subtraction Based on Multitaper Spectral Estimation

Both the amplitude spectrum and phase spectrum should be calculated by FFT when the noisy speech had been processed to frames by Slepian windows firstly. Then the average amplitude spectrum is calculated by smoothing based on the adjacent frames. At the same time, the power spectral density of multitaper spectral is estimated for data frame. The estimated values can also be disposed by smoothing based on the adjacent frames and the smoothed power spectrum density can be calculated. With the condition that the number of preamble frames without words segment (noise) is known, the average power spectral density of the noise could be gotten. Then the gain factor of spectral subtraction can be obtained by taking advantage of spectral subtraction relationship. If the condition of amplitude spectral and phase spectral is known beforehand, the speech signal could be restored to time domain by IFFT. Furthermore, the enhancement of spectral subtraction speech is realized [14] [15] [16].

## 3. The BARK Subband Variance in Frequency Domains

The function of basilar membrane of human ear is similar to that of the frequency analyzer based on the frequency group content of the auditory masking effect. The frequency between 20 Hz and 22,050 Hz is divided into 25 frequency group, while the basilar membrane of human ear is divided into a lot of parts by our brain. Each part of basilar membrane is corresponded to a frequency group which is also known as the unequal bandwidth (BARK) subband [17] [18] [19].

The principle of endpoint detection algorithm based on BARK subband variance is described briefly. First of all, the speech signal is added with window to be frames. Secondly, it is processed with FFT. The total number of  $(N/2 + 1)$  of positive frequency spectral lines is obtained sequentially. The spectral line will be extended by interpolating. The average amplitude of the BARK subband in the BARK is calculated by Equation (4).

$$E_i(j) = \frac{1}{f_{j,h} - f_{j,l} + 1} \sum_{f_{j,l} \leq f_k \leq f_{j,h}} |X_i(k)| \quad j = 1, 2, \dots, q \quad (4)$$

where the  $f_{j,l}$  and  $f_{j,h}$  are the  $j^{\text{th}}$  critical frequencies of BARK subband at low frequency and high frequency respectively..

The mean and variance of the BARK subband can be obtained as followed. Using preamble without words segment, the average value of noise is obtained. The speech signal endpoint will be detected by using the double threshold method with the single parameter after the threshold is set.

#### 4. The Block Diagram of Algorithm Principle

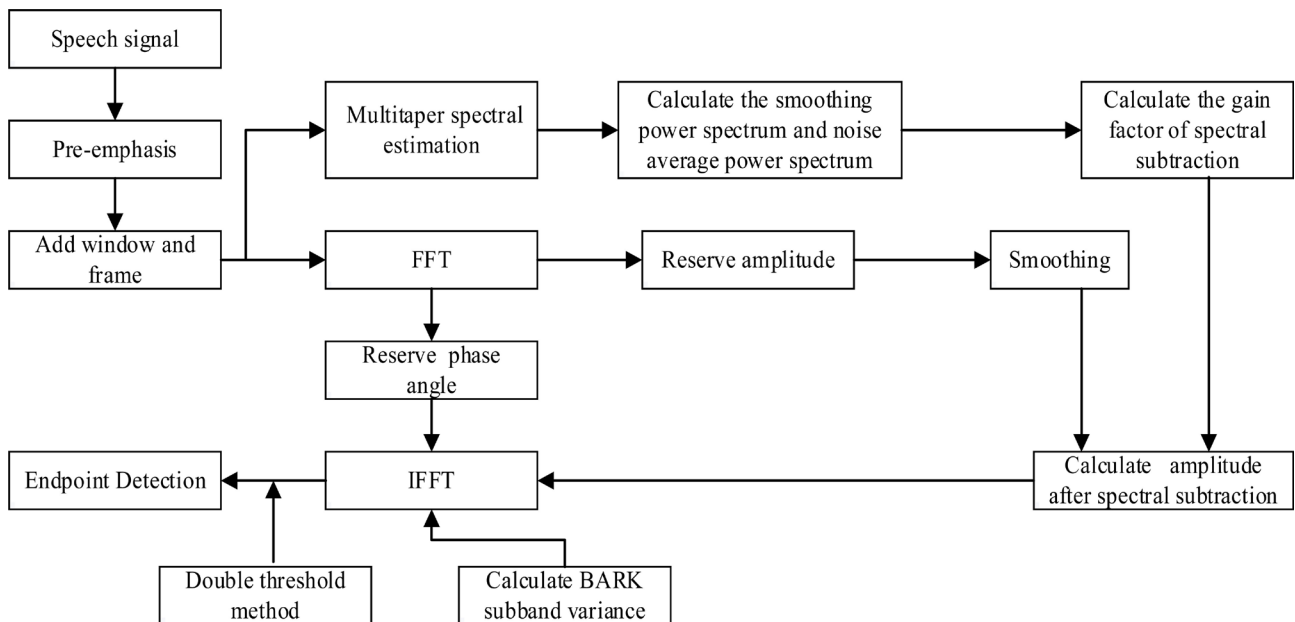
There are three key steps for the proposed algorithm in this paper. The first step is doing spectrum analysis by FFT to obtain the characteristic of speech signal. Consequently, the signal to noise of speech signal can be improved by using the method of improved spectral subtraction based on multitaper spectral estimation. Finally, the endpoint can be detected though calculating the BARK subband variance in frequency domain. The flow diagram is indicated in **Figure 1**.

#### 5. Analyzing the Results of Experiment in Low SNR

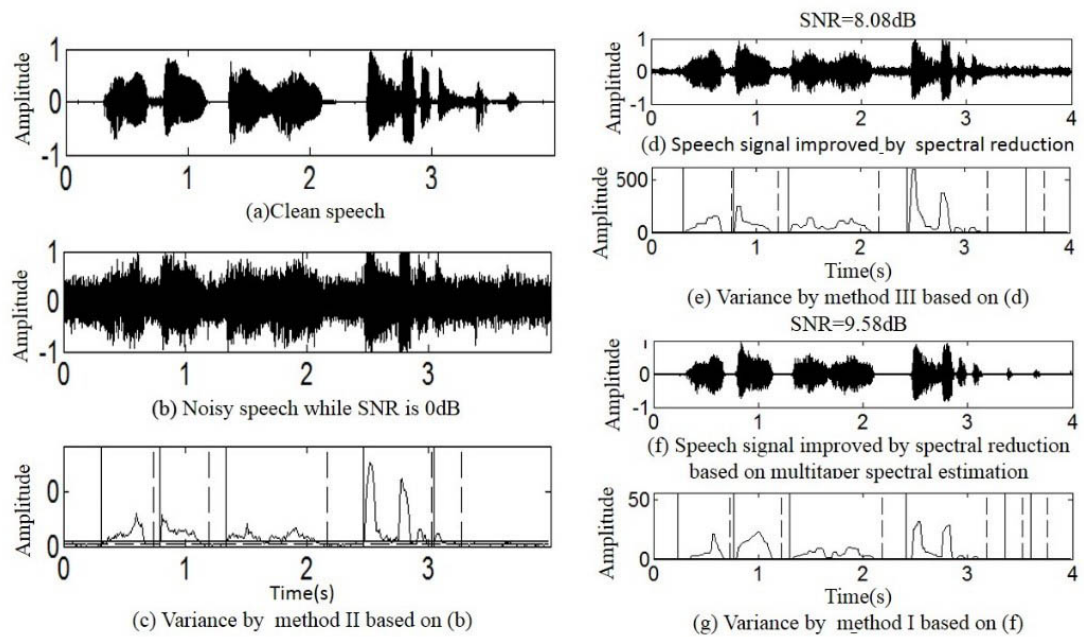
##### 5.1. Anti-Noise Capability Analysis

The proposed algorithm is simulated with the simulation software MATLAB. There are 4 parameters to be defined for the experiment. The sampling frequency of clean speech is 8kHz and Hamming window is selected. Meanwhile, the length of preamble without words segment is 0.25 seconds and frame shift is 80 sample points. The clean speech is a clean Chinese phrases that “lan tian, bai yun, bi lv de da hai” indicated in **Figure 2(a)**. Gauss noise is added to it.

Firstly, the signal to noise ratio is set to 0 dB to verify the anti-noise capabili-



**Figure 1.** Algorithm flow diagram.



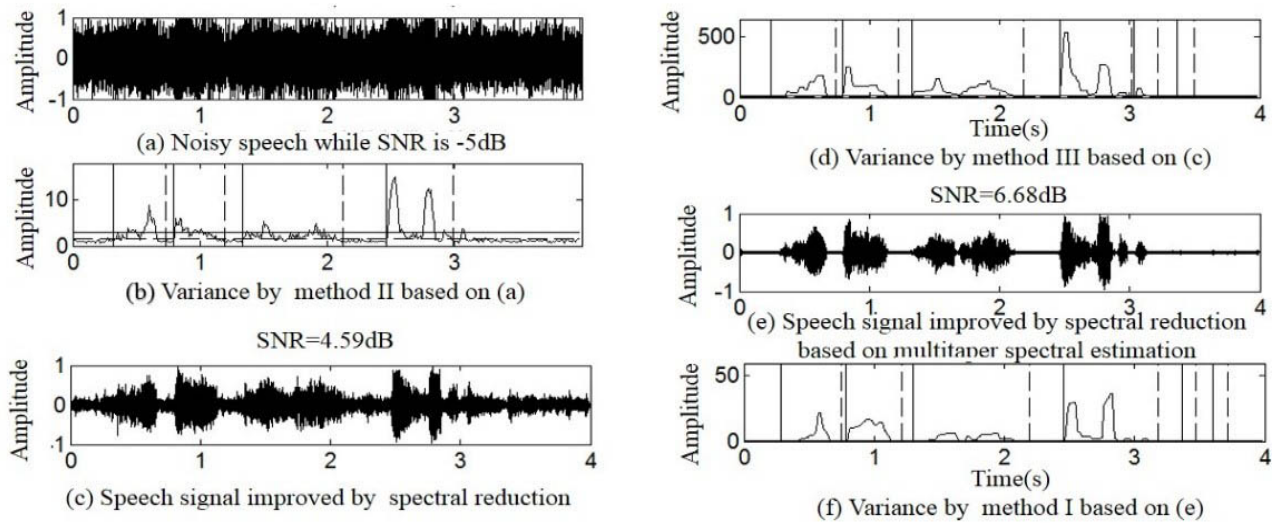
**Figure 2.** The detection results of three algorithms (SNR = 0 dB). (a) Clean speech. (b) Noisy speech while SNR is 0 dB. (c) Variance of BARK subband in frequency domain (method II) of noisy speech indicated in (b). (d) Speech signal improved by the method of spectral reduction. (e) Variance of short-time uniform subband (method III) of speech improved in fig. (d). (f) Speech signal improved by the method of improved spectral subtraction based on multitaper spectral estimation. (g) Variance of BARK subband in frequency domain of speech improved (method I) in (f).

ty. The result of the paper algorithm (method I) is compared with the result of the method of BARK subband variance in frequency domain (method II) and that of spectral subtraction short-time uniform subband variance [20] (method III) respectively. All the results are shown in **Figure 2**.

Analyzing the result of BARK subband variance in frequency domain, as indicated in **Figure 2(c)**, the Chinese word “hai” can’t be detected by it. In contrast, the word segment can be detected correctly by the other two methods. Because the SNR is equal to 8.08 dB in **Figure 2(d)** and it is 9.58 dB in **Figure 2(f)** respectively, the accuracy of detection is better than that of the methods spectral subtraction in **Figure 2(c)**. In the case of detecting the word segment with the two algorithms correctly, the SNR of this paper algorithm improves obviously because its SNR is larger than that of the spectral reduction about 1.5 dB. This result shows that the anti-noise capability effect of paper algorithm is better.

Consequently, the SNR is reduced to  $-5$  dB indicated in **Figure 3(a)**, and the anti-noise capability is verified again. All the results of three methods are shown in **Figure 3**.

It is obviously that BARK subband variance in frequency domain is failed to meet the detection requirements as shown in **Figure 3(b)**. On the contrary, the other two methods can detect the word segment correctly in **Figure 3(d)** and **Figure 3(f)**. Comparing the value of SNR in **Figure 3(c)** and **Figure 3(e)**, the SNR of this paper algorithm is larger than that of the spectral reduction about 2.09 dB. It means that the paper algorithm can remove a lot of noise, and its



**Figure 3.** The detection results of three algorithms (SNR = -5 dB). (a) Noisy speech while SNR is -5 dB. (b) Variance of BARK subband in frequency domain (method II) of noisy speech indicated in (a). (c) Speech signal improved by the method of spectral subtraction. (d) Variance of short-time uniform subband (method III) of speech improved in (c). (e) Speech signal improved by the method of improved spectral subtraction based on multitaper spectral estimation. (f) Variance of BARK subband in frequency domain of speech improved (method I) in (e).

ability of anti-noise is much better under the condition of low SNR.

## 5.2. Analysis Accuracy of Detection

Another speech “ci hen mian mian wu jue qi” in Chinese poetry was applied to this detection experiment. The speech signal is added with Gauss noise and white, volvo of NOISE-92 Library respectively. This paper algorithm is compared with the method of BARK subband variance in frequency domain and spectral subtraction short-time uniform subband variance [20] in different SNR (-10 dB, 0 dB, 5 dB) to analyze accuracy of detection.

Accuracy is defined as follow:

$$\text{Accuracy} = \frac{\text{Total Frames} - \text{Error Frames}}{\text{Total Frames}} \quad (5)$$

$$\text{Error Frames} = \text{Speech misjudges noise frames} + \text{noise misjudges speech frames} \quad (6)$$

The accuracy of three methods with different noise and different SNR value is shown in **Table 1**.

As shown in **Table 1**, when the SNR is 5 dB, this paper algorithm does not reflect the detection accuracy for the same speech signal because the signal is much stronger than noise. Considering Gauss noise with SNR is -10 dB and 0 dB, the accuracy rate of this paper algorithm improves obviously. The improvement is also evident when white noise and volvo noise with -10 dB are added to the signal, the improvement of accuracy rate increases obviously with Volvo noise.

All of the above, it is concluded that this paper algorithm can show better anti-noise performance and higher accuracy in low SNR. Though the result detection is affected by several factors, such as the speed, environmental noise, and

**Table 1.** Endpoint detection accuracy.

Endpoint detection algorithm	Gauss noise white noise Volvo noise								
	-10 dB	0 dB	5 dB	-10 dB	0 dB	5 dB	-10 dB	0 dB	5 dB
BARK subband variance in Frequency domain	Null	42.9	42.9	Null	42.9	42.9	14.3	42.9	42.9
Spectral subtraction short-time uniform subband variance [20]	57.1	42.9	42.9	28.6	42.9	42.9	14.3	28.6	42.9
This paper algorithm	71.4	57.1	42.9	57.1	42.9	42.9	57.1	42.9	42.9

the performance of the algorithm and so on, it needs to be improved in the future.

## 6. Conclusion

Considering that the detection of endpoint is one of the most important aspects of speech signal processing, a speech endpoint detection algorithm with low SNR condition is proposed in this paper. Firstly, the noisy speech is processed with the method of improved spectral subtraction based on multitaper spectral estimation in order to improve the signal to noise ratio. Then the method of BARK subband variance in frequency domain is applied to detect the speech endpoint. According to the results of simulation, it is distinct that the algorithm mentioned in the paper can detect speech endpoint correctly in low SNR condition, and it has a good anti-noise performance. The method should make a part in application of speech endpoint detection because of its high efficiency in the condition of low SNR. However, there are some elements, for instance, the type of noise, which have some influences with the algorithm. It is necessary to make some deep research to improve for accuracy of detection. This will be the focus of the work in the future.

## Acknowledgements

This work was supported in part by the National Science Foundation of China under Grants 51504039.

## References

- [1] Zhuo, G. and Bian-Ba, W.-D. (2015) A Study of Tibetan Speech Pitch Detection Algorithm Based on Matlab. *Modern Electronics Technique*, **10**, 20-22.
- [2] Zhao, L. (2016) *Speech Signal Processing*. 3rd Edition, China Machine Press, Beijing.
- [3] Han, L.H., Wang, B. and Duan, S.F. (2010) Development of Voice Activity Detection Technology. *Application Research of Computers*, **4**, 1220-1226.
- [4] Sun, Y.M., Wu, Y.Y. and Li, P. (2016) Research on Speech Endpoint Detection Based on the Improved Dual-Threshold. *Journal of Changchun University of Science and Technology*, **1**, 92-95.
- [5] Wang, L.L., Xia, X., Feng, L. and Liu, G.C. (2014) New Speech Endpoint Detection



- Algorithm Based on Spectrum Variance and Spectral Subtraction. *Computer Engineering and Applications*, **8**, 194-197.
- [6] Haigh, J.A. and Mason, J.S. (1993) Robust Voice Activity Detection Using Cepstral Features. *Proceedings of Computer, Communication, Control and Power Engineering*, Vol. 3, Beijing, 19-21 October 1993, 321-324.  
<https://doi.org/10.1109/tencon.1993.327987>
- [7] Zhao, H., Wang, G.J. and Zhao, L.X. (2010) A New Voice Activity Detection Using Logarithmic Energy Spectral Entropy. *Journal of Hunan University*, **7**, 72-77.
- [8] En, D., Zhang, F.L., Zhang, Z. and Hu, S.Q. (2016) Application of Fuzzy Entropy in Speech Endpoint Detection in Car Environments. *Computer Engineering and Applications*, **10**, 147-150.
- [9] Zheng, J.H., Huang, H.M., Zhong, M.H., Cao, N.W. and Chen, Y.L. (2007) Comparative Study of Several Speech Signal Endpoint Detection Methods. *Guanagxi Wuli*, **4**, 20-23.
- [10] Yin, X.C., Guo, Y., Zhang, B.F. and Liu, X. (2011) Voice Activity Detection Algorithm Based on Bark Wavelet. *Computer Engineering*, **12**, 276-278.
- [11] Wang, Y., Feng, Y., Ding, X.B. and Chen, D.Y. (2016) Endpoint Detection Algorithm for Noisy Speech Based on Time-Frequency Combination. *Journal of Natural Science of Heilongjiang University*, **3**, 410-415.
- [12] Wu, P.P., Zhao, G. and Zou, M. (2008) An Improved Spectral Subtraction Method Based on Multitaper Estimation. *Modern Electronics Technique*, **12**, 150-152.
- [13] Thomson, D.J. (1982) Spectrum Estimation and Harmonic Analysis. *Proceedings of the IEEE*, **70**, 1055-1096. <https://doi.org/10.1109/PROC.1982.12433>
- [14] Han, F. and Jin, Z.X. (2016) Study of Endpoint Detection Algorithm in Low SNR. *Journal of Northwest Normal University*, **5**, 55-59.
- [15] Hu, Y. and Loizou, P.C. (2004) Incorporating a Psycho Acoustical Model in Frequency Domain Speech Enhancement. *IEEE Signal Processing Letters*, **11**, 270-273.  
<https://doi.org/10.1109/LSP.2003.821714>
- [16] Dong, H. (2016) Improved Speech Endpoint Detection under Low SNR Environment. *Computer Technology and Development*, **3**, 71-74.
- [17] Zhang, C.L., Zeng, X.Y. and Wang, S.G. (2012) A Voice Activity Detection Algorithm Based on the Variance of Critical Band Power Spectrum. *Technical Acoustics*, **2**, 204-208.
- [18] Gao, M.M., Chang, T.H., Yang, G.T. and Li, M. (2009) Speech Feature Extraction Algorithm Based on Subband Dominant Frequency Information. *Computer Engineering*, **18**, 161-163.
- [19] Wang, X.H., Qu, L., Zhang, C. and Jian, X.W. (2016) Speech Feature Extraction Algorithm Based on the Bark Wavelet Packet Transform with Fisher. *Journal of Xi'an Polytechnic University*, **4**, 453-457.
- [20] Wang, W., Hu, G.M., Yang, L., Huang, D.F. and Zhou, Y. (2016) Research of Endpoint Detection Based on Spectral Subtraction and Uniform Subband Spectrum Variance. *Audio Engineering*, **5**, 40-43.