

A Modified Method for Acoustic Pulse Extraction under Noisy Environments

Menghua Cao^{1,a}, Zhiqiang Sun^{*2,b}, Erling Gong^{3,c}, Hongwei Xie^{4,d}

College of Mechatronic Engineering and Automation, National University of Defense Technology

Changsha, 410073, China

^aemail: upcnuadt, ^bemail: sunzq39, ^cemail: gel314, ^demail: xhwei65}@163.com

Keywords: Acoustic Pulse; Spectral Subtraction; Frequency Band Variance; Cross-correlation

Abstract. This paper addresses a modified method of extracting acoustic pulse based on the multi-band spectral subtraction and the frequency band variance. In this method, the acoustic pulse is detected, the noise spectrum is estimated and subtracted from the acoustic pulse spectrum. By abandoning the late reverberation and decayed part of acoustic pulse, effective part of signal is obtained. Using this method, signal-to-noise ratio (SNR) is considerably improved and cross-correlation between different channels can be well optimized.

Introduction

Acoustic pulse is a very common class of nature signal, and is characterized by high frequency, short duration and great impulse. Detecting and extracting the effective part of acoustic pulse accurately can help improve the precision of source position and voiceprint recognition. However, because of the existence of noise in the testing field, sometimes it's difficult to detect the signals in real-time exactly. So distinguishing the acoustic pulse from the background noise is an urgent need.

In this paper, a modified method based on frequency band variance[1] and multi-band spectral subtraction[2] is proposed to get the effective acoustic pulse signals for further study. Signal detection is the first step. Frequency band variance of signals collected from microphones in real-time should be calculated in every frame. In practice, there are two types of signal features as the indicators of occurrence of acoustic pulse which are time domain and frequency domain features respectively. Short-time energy and short-time average zero-crossing rate are the most commonly used time domain features which have the characteristics of less calculation and bad anti-interference performance. Frequency band variance, as a kind of frequency domain feature, has less computational complexity and a certain degree of anti-noise performance. These characteristics make it available to satisfy the need of real-time detection. And the modified frequency band variance proposed in this paper can further enhance the anti-noise performance through the zoom spectrum analysis.

The second step is to confirm the endpoints for extracting the effective component of signals by reducing the influence of noise and reverberation. Spectral subtraction algorithm[3,4,5] is one of the first algorithms proposed for noise reduction, which is characterized by simplicity. And the spectral subtraction algorithm is based on two major components: the estimation of noise spectrum and the estimation of acoustic pulse spectrum. In this paper, multi-band spectral subtraction is chosen as the performance for reducing the white noise simply and easily. And the multi-band spectral subtraction algorithm is found that it can perform better than other subtractive-type algorithms in real environment with highly random noise[2].

This paper is organized as follows. A modification of the frequency band variance is detailed in Section II. The conception of multi-band spectral subtraction is presented in Section III. Experimental results are presented and discussed in Section IV. Finally, the work is concluded in Section V.

Modification of Frequency Band Variance

The acoustic pulse signals behave differently from the noise in frequency domain. In general, the energy of acoustics pulse varies greatly with the change of frequency band, whereas the energy of noise varies steady relatively. And we found that the frequency of acoustic pulse is mainly distributed in the interval over than 2 kHz in the experiment. We can use this characteristic to tell the difference between the acoustic pulse and the noise. So, a modified method based on frequency band variance[1] is proposed in this section.

Frequency Band Variance

Let $x(n)$ denotes an acoustic pulse signal with background noise, and $x_i(m)$ is given as

$$x_i(m) = w(m) * x(iT + m) \quad 1 \leq m \leq N \quad (1)$$

where $w(m)$ is the window function; $i=0,1,2,\dots$; N is the frame length; T is the frame shift length.

Using Discrete-time Fourier Transform (DTFT), spectrum can be got as

$$X_i(k) = \sum_{m=0}^{N-1} x_i(m) \exp\left[-j\frac{2\pi km}{N}\right], \quad 0 \leq k \leq N-1 \quad (2)$$

And the mean amplitude of $X_i(k)$ is

$$E_i = \frac{1}{N} \sum_{k=0}^{N-1} |X_i(k)| \quad (3)$$

The variance is

$$D_i = \frac{1}{N-1} \sum_{k=0}^{N-1} \left[|X_i(k)| - E_i \right]^2 \quad (4)$$

where E_i and D_i respectively present the mean amplitude and variance of i^{th} frame.

Modified Sub-band Frequency Variance

In this paper, in order to eliminate the disturbance of some additive noise with high energy and low frequency, we just analyze a certain frequency band where the acoustic pulse frequency exists. And Chirp Z transform can solve the problem successfully. This kind of transform samples the spectrum using the helix instead of a unit circle[6]. So the spectrum distribution of signal is focused on the certain frequency band with higher resolution. Let

$$z_k = AW^{-k} \quad (5)$$

Then there is

$$\begin{cases} A = A_0 e^{j\theta_0} \\ W = W_0 e^{-j\varphi_0} \end{cases} \quad (6)$$

where A_0 denotes the radius length of initial sampling point, and usually is less than 1; W_0 denotes the extensive rate of the helix, when $W_0=1$, the helix equals to a portion of a circle which radius is A_0 ; θ_0 denotes the phase angle of initial sampling point z_0 , and φ_0 denotes the angle between adjoining points.

And the formula of Z transform[7] is

$$X(z_k) = \sum_{n=0}^{N-1} x(n) z_k^{-n} \quad (7)$$

So we can get that

$$\begin{aligned} X(z_k) &= \sum_{n=0}^{N-1} x(n) A^{-n} W^{nk} \\ &= W^{\frac{k^2}{2}} \sum_{n=0}^{N-1} x(n) A^{-n} W^{\frac{n^2}{2}} W^{-\frac{(k-n)^2}{2}} \end{aligned} \quad (8)$$

Define that

$$\begin{cases} g(n) = x(n) A^{-n} W^{\frac{n^2}{2}} \\ h(n) = W^{\frac{n^2}{2}} \end{cases} \quad (9)$$

Thus

$$\begin{aligned} X(z_k) &= W^{\frac{k^2}{2}} \sum_{n=0}^{N-1} g(n) h(k-n) \\ &= W^{\frac{k^2}{2}} g(k) * h(k) \quad k = 0, 1, \dots, M-1 \end{aligned} \quad (10)$$

This operation can be got by FFT more effectively. And then we can calculate the variance of Chirp Z transform by (3) and (4). After getting the modified variance, we can obtain the initial points of the signals using the double threshold method[8]. As for the end points, there are the problems of attenuation and reverberation involved. The next section will give the interpretation.

Multi-band Spectral Subtraction Algorithm

When an acoustic pulse happened, it will decay rapidly and be confused by the noise and the reverberation from the room environment. In fact, it is impractical to separate the effective signals from the reverberation and the noise accurately. But, we can reduce the influence of reverberation and the noise by analyzing the components of the signals obtained[9,10,11].

Model of the Signals

The signal received by a microphone can be expressed as

$$x(n) = \hat{x}(n) * h(n) \quad (11)$$

and

$$h(n) = w(n) e^{-\tau n} u(n) \quad (12)$$

where $\hat{x}(n)$ represents the real acoustic pulse, $h(n)$ is the room impulse response, $w(n)$ represents a white zero-mean Gaussian noise, $u(n)$ is the unit step function, and “*” denotes the linear convolution operation. And τ is a constant related to the reverberation time T_{60} , as

$$\tau = 3 \ln(10) / T_{60} \quad (13)$$

This model can be further decomposed into a direct component $h_d(n)$ with background noise and a few early reflection, and a component $h_r(n)$ with late reverberation, so that

$$h(n) = \begin{cases} 0 & n < 0 \\ h_d(n) & 0 \leq n \leq T_d \\ h_r(n) & n \geq T_d \end{cases} \quad (14)$$

Target of this work is to remove the noise and late reverberation and get the direct component $h_d(n)$. Spectral subtraction is taken into count to de-noising and de-reverberation.

Multi-band Spectral Subtraction Algorithm

Multi-band spectral subtraction algorithm is an improvement of spectral subtraction. Simplicity of calculation is the main advantage fitting for real-time processing. Here, spectral over-subtraction is chosen because of the ability of eliminating the musical noise. Considering a signal with noise

$$x(n) = \hat{x}(n) + d(n) \quad (15)$$

Providing that acoustic pulse and noise are uncorrelated, so that the de-noising signal can be obtained by[5]

$$|\hat{X}_i(\omega)|^2 = \begin{cases} |X_i(\omega)|^2 - \alpha_i \delta_i |\hat{D}(\omega)|^2 & k_i < \omega < k_{i+1} \\ \beta |\hat{D}(\omega)|^2 & \text{otherwise} \end{cases} \quad (16)$$

where δ is the over-subtraction factor of i^{th} band; β_i is the noise spectral floor; δ_i is an additional band; $\hat{D}(\omega)$ is the estimation of noise spectrum. k_i and k_{i+1} are the start and end frequency.

In this paper, different with the usually method using the noise before the acoustic pulse coming, $\hat{D}(\omega)$ is computed by the component after endpoint detection using the frequency band variance. By setting the proper threshold, the end point can be confirmed approximately, which is similar to the T_d in Section II. But this is an interval, not a certain point.

Smoothing processing can reduce the error of spectral estimation, thus

$$P_i(k) = \frac{1}{2M+1} \sum_{j=-M}^M X_{i+j}(\omega) \quad (17)$$

when $M=1$, 3 frames of signal will be averaged. And the over-subtraction factor α_i is computed as

$$\alpha_i = \begin{cases} 5 & SNR_i < -5 \\ 4 - \frac{3}{20} SNR_i & -5 \leq SNR_i \leq 20 \\ 1 & SNR_i > 20 \end{cases} \quad (18)$$

where SNR_i is the signal-to-noise ratio of i^{th} frame. And values of δ_i are set to

$$\delta_i = \begin{cases} 1 & f_i < 1 \text{ KHz} \\ 2.5 & 1 \text{ KHz} \leq f_i \leq (F_s/2) - 2 \text{ KHz} \\ 1.5 & f_i > (F_s/2) - 2 \text{ KHz} \end{cases} \quad (19)$$

where f_i is the upper bound frequency of i^{th} band and the f_s is the sampling frequency.

To summarize, block diagram of this method is showed below

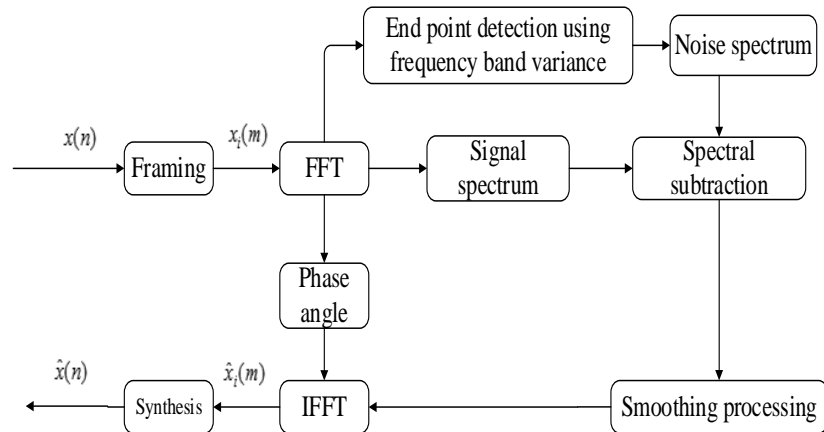


Fig.1. Block diagram of spectral subtraction

Test results

Acoustic Pulse Detection

The signals are sampled at the rate of 50 kHz in a room which size is $4 \times 5 \times 3(m^3)$. When the signals are detected in real-time, we set length of each frame is $N=512$, and there is no frame shift. The original dual-channel signals and the spectrogram are showed in Fig.2. We can see that most of the frequency components of acoustic pulse signals are over than 2 kHz.

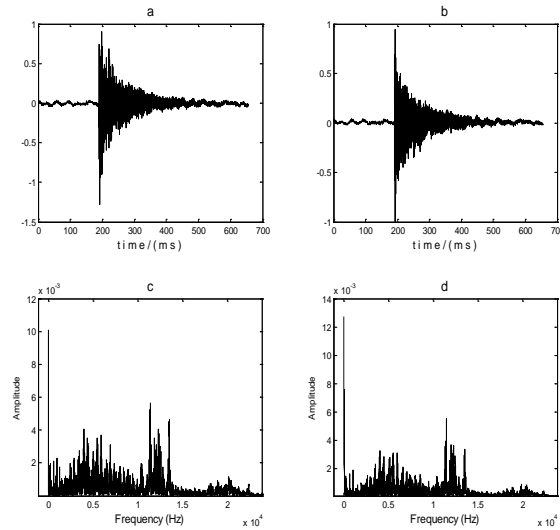


Fig.2. (a) Original signal of channel 1. (b) Original signal of channel 2. (c) Spectrogram of signal 1. (d) Spectrogram of signal 2.

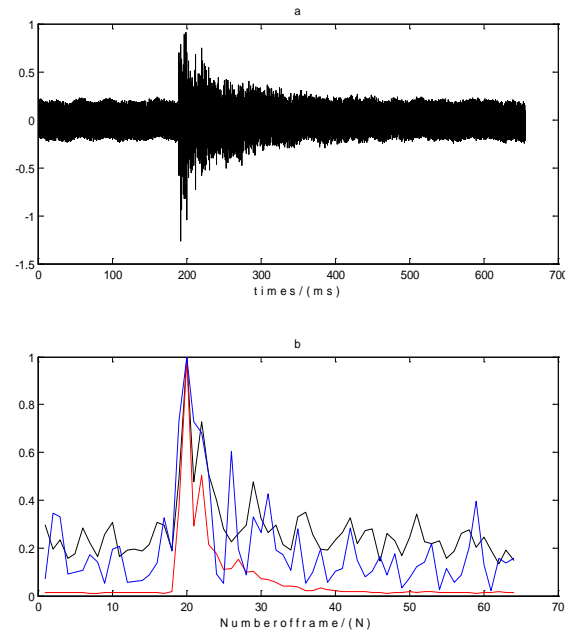


Fig.3.(a) Signal with noise which SNR is 2dB. (b) Comparison of different algorithms (Red line is the modified frequency band variance; Black line is the traditional frequency band variance; Blue line is the average amplitude of wavelet coefficients).

The signal of channel 1 showed in Fig.2 is added by the white noise at the SNR level of 2 dB. So, we can see that the acoustic pulse signal is partially masked by noise in Fig.3. And in this paper, we detect the acoustic pulse using 3 methods, they are traditional frequency band variance, modified frequency band variance and the average amplitude of wavelet coefficients[13]. The last method is widely used in many research, but its performance is not good in this work in the Fig.3. Similarly, using the traditional frequency band variance, we can't get the explicit sign that the acoustic pulse has occurred. The low SNR can severely impact the detection results. Whereas, modified frequency band variance has the sharp curve after the acoustic pulse appeared. This means that the modified frequency band variance proposed in this paper is workable.

De-noising and Signal Extraction

After detecting the acoustic pulse, we need extract the effective part of signal with less noise and reverberation by multi-band spectral subtraction and double threshold method[8,12].

In the test, we calculate the mean value of noise's frequency band variance in 10 frames. And the double threshold are respectively set as 2 and 3 times of the mean value. From Fig.4, we can get the accurate initial interval and the blurry end interval before de-noising and de-reverberation. Then we cut out 10 frames of the signal from the end interval as the noise for spectral subtraction. Here, we set the length of frame as 256, and frame shift length is 32. Selecting the less frame shift length can enhance the effect of this algorithm with the cost of increasing the computation, but this impact can be ignored.

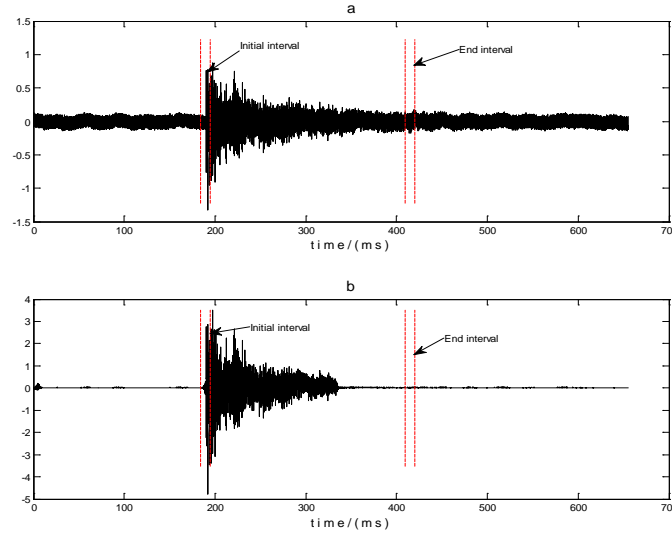


Fig.4. (a) Result of endpoint detection. (b) Result of de-noising and dereverberation.

After de-noising and de-reverberation using multi-band spectral subtraction, we can see that the signal with noise has been considerably improved in the Fig.4. The SNR has been also changed from 2 dB to 42 dB. Thus far, the effective acoustic pulse can be extracted easily.

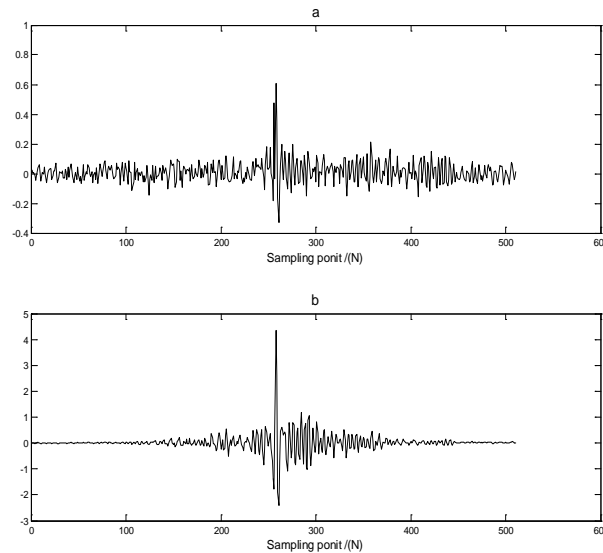


Fig.5. (a) Cross-correlation of dual-channel signals with noise which SNR is 2dB. (b) Cross-correlation of dual-channel signals after de-noising and dereverberation.

Finally, we calculate the cross-correlation of dual-channel signals using the first 256 points from the initial interval. We can see that the cross-correlation peak value of effective signals obtained is more prominent than noisy signals. And pseudo peaks has been removed effectively in Fig.5. This can be applied to the acoustic source localization based on time delay.

Conclusion

An improved method for acoustic pulse extraction is proposed in this paper. This method is based on the modified frequency band variance and multi-band spectral subtraction. Instead of estimating the spectrum of noise and reverberation by modeling, we find the end interval of signal and estimate the spectrum of the late noise and reverberation. Using this method, we can considerably reduce the impact of noise and reverberation and obtain the relatively pure signal with little distortion. This method can also meet the need of real-time because of its low computation.

References

- [1] BK. Reaves, "Speech detection device for the detection of speech end points based on variance of frequency band limited energy," *Scandinavian Journal of Gastroenterology Supplement*, 1997,102(6): 3252.
- [2] S. Kamath, P. Loizou, "A Multi-Band Spectral Subtraction Method for Enhancing Speech Corrupted by Colored Noise", *IEEE Int. Conf.on Acoustics, Speech, and Signal Processing*, Orlando, USA, vol. 4, pp. 4160–4164, May (2002).
- [3] N Upadhyay, A Karmakar, "Speech Enhancement using Spectral Subtraction-type Algorithms: A Comparison and Simulation Study," *Procedia Computer Science*, 2015,54:574-584.
- [4] Steven F. boll, "Suppression of Acoustic Noise in Speech Using Spectral Subtraction," *IEEE Transactions on Acoustics, Speech, and Signal processing*, vol. assp-27, pp.113-120, April 1979.
- [5] S.Ogata, T.Shimamura, "Spectral Subtraction method to enhance speech signal," *IEEE International Conference on Electrical and Electronic Technology*,2001, vol-I, pp.242-245.
- [6] L Rabiner, "The chirp z-transform algorithm-a lesson in serendipity, " *IEEE Signal Processing Magazine*,2004,21(2):118-119.
- [7] LR Rabiner, RW Schafer, CM Rader, "The Chirp z-Transform Algorithm and Its Application," *IEEE Transactions on Instrumentation and Measurement*, 1990, 39(2):318-323.
- [8] PAA Esquef, LWP Biscainho, PSR Diniz, and FP Freelanci, "A Double-Threshold-Based Approach to Impulsive Noise Detection in Audio Signals," *Pmc. EUSIPCO*, 2000, pp. 2041-2044.
- [9] E. A. P. Habets, "Single-channel speech dereverberation based on spectral subtraction," in *Proc. of the 15th Annual Workshop on Circuits, Systems and Signal Processing (ProRISC'04)*, Veldhoven, Netherland, Nov. 2004, pp. 250–254.
- [10] K. Lebart and J.M. Boucher, "A new method based on spectral subtraction for speech dereverberation," *Acta Acoustica*, vol. 87, pp. 359-366, 2001.
- [11] M. Berouti, R.Schwartz,and J. Makhoul, "Enhancement of Speech Corrupted by Acoustic Noise," *Proc ICASSP 1979*,pp.208-211.
- [12] Y Wang, L Zhang, "Effective Speech Endpoint Detection Algorithm For Voiceprint Recognition," *International Conference on Information Sciences*, 2015.
- [13] TR Downie, BW Silverman, "The discrete multiple wavelet transform and thresholding methods[J] ", *IEEE Trans on Signal Processing*.1998.46(9):2558-2561.