doi:10.3772/j.issn.1006-6748.2015.02.005

Research of multi-channel impulse noise detection algorithms based on template①

Zhao Wei (赵　薇)②, Xie Qianli, Wu Yin, Zhang Qin
(Faculty of Science and Technology, Communication University of China, Beijing 100024, P. R. China)

Abstract

Attenuating the undesired audio noise generated by impulse noise, such as shot and scream of brakes, is specially useful for real-time audio recording of TV or broadcasting live report. On the basis of impulse noise detection algorithms based on template, this paper improves the method of establishing the template by using multiple microphones to pick up noise corrupted signals and impulse noises in the environment. The universal of thresholds is found and a detection algorithm with slope as the characteristic is proposed by comparing a variety of feature extraction algorithms. The proposed algorithm gets a significant improvement in testing speed and accuracy, which means it is suitable for real-time processing of audio signals.

Key words: impulse noise detection, template, multi-channel, universal of thresholds, slop algorithm (SA)

0 Introduction

Impulse noises are short discontinuous pulses or noise peaks with large amplitude. Typical acoustic impulse noises include sounds of machine gun firing, of typing on the keyboard, of popping popcorn and so on[1]. Impulse noise is one of the most ruinous noises in audio signal[2].

Impulse noise detection locates pulse positions based on waveforms. It is one of the most important preparations of impulse noise cancellation. Many experts and scholars have made a lot of studies on impulse noise detection algorithms. The traditional threshold detection algorithms, for example, calculate the difference between the window element and noise signal when a sliding window glides past the noise corrupted signal[3,4]. If the difference value is larger than the threshold, then the position of the sliding window is the position of the impulse noise. The detection algorithm based on auto regressive (AR)[5] has a bad detection performance when the noise pulse and the glottis pulse happen at the same time. But the detection algorithm based on spare auto regressive (SAR)[6] has a good detection performance and interpolation result. The result of the bidirectional detection algorithm based on AR/SAR[7] is even better. The detection algorithm based on wavelet transform[8] is to apply the wavelet transform to the noise corrupted signal to get the wavelet coefficients as a feature. If the wavelet coefficients are larger than the threshold, then the detected point is the position of the impulse noise. The method based on the impulse noise template is widely used in recent years because of its simple algorithm and principle. Extracting the feature of impulse noise template and processing it in time domain, one can to find the start position of the impulse noise[9]. The impulse noises of the noise corrupted signal may be regarded as the waveform of the same kind of noise panning and zooming, due to the pulse waveform is similar to those produced by the same circumstances and only the amplitude is different. Therefore, for impulse noise generated in a particular environment, the pulse position can be detected by matching the template which can be got by extracting its waveform at the time of its generating, namely the template of impulse noise.

Feature extraction of audio signal is the most important step to the method based on impulse noise template. Features are extracted by transforming the signal in time domain and frequency domain and getting the related feature parameter. There are a variety of features extraction methods[10] including the Zero-crossing rate, Spectral Centroid, Spectral Roll-off, Spectral Flux, etc. Ref. [11] is a combination of the above characteristics, which improves the accuracy of retrieval. In Ref. [11], the authors introduce the pretreat-

① Supported by the Research on Multi-channel Audio Noise Reduction Algorithm (No. 3132014XNG1430).
② To whom correspondence should be addressed. E-mail: zhao_wei@cuc.edu.cn
Received on Jan. 27, 2014

ment and code the features to save them in the database. Knowing the audio coding which is the same as the coding of template, they calculate the correlation between the corresponding audio feature and the template feature. This method achieves good results.

Besides, extracting noise template is also one of the hot research spots. Kao et al constructed the noise template by picking up impulse noise for several times, and averaged the impulse noises[12]. On the basis of the wind noise template in audio library, Kuroiwa et al combined the characteristics of the wind noise and focused on the subtle differences in different periods to reconstruct the wind model[13]. This method is more effective than traditional methods.

In this work, several microphones are used to record the noise corrupted signal and the impulse noise in the environment, and improve the construction method of the impulse noise template on the basis of the impulse noise detection algorithm based on template. Several methods of feature extraction are comparied to obtain the universality of the threshold, and an impulse noise detection method based on slope feature is proposed. This algorithm has got significant improvement in detection speed and detection accuracy. Improving the accuracy and effectiveness of impulse noise detection algorithm helps to improve the efficiency of impulse noise cancellation. That means the proposed algorithm obtains a high practical value for real-time audio recording of television and radio live report.

1 Multi-channel Audio

Attenuating the undesired audio impulse noise, such as shot and scream of brakes, is particularly useful for real-time audio recording of TV or broadcasting live report. In this method, an all directional microphone is used to record the noise corrupted signal, meanwhile a cardioid microphone or a microphone array is used to record the impulse noise in the environment, as shown in Fig. 1.

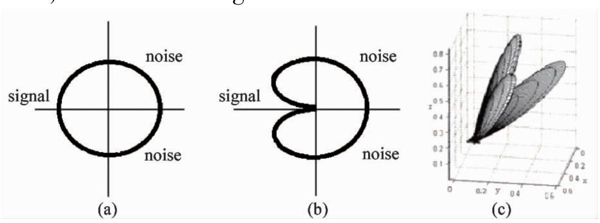

Fig. 1 Multi-channel microphone directional pattern

Fig. 1(a) is a directional pattern of an onidirectional microphone which records the noise corrupted signal. Fig. 1(b) is a directional pattern of a cardioid microphone. Fig. 1(c) is a directional pattern of a 4 × 4 microphone array using strong directional microphone to record impulse noise $y(n)$, which generally contains 20 impulse noises and makes full use of the directional characteristic of the microphone to avoid the disclosure of noise free signal while picking up noises.

2 Algorithms

2.1 Traditional algorithms

The traditional algorithms of impulse noise detection based on template are done based on the characteristic of the matching filter, when Extract the features from the template and the noise corrupted signal respectively (such as amplitude[14], frame energy[14], and the energy histogram[11,15,16], etc.), and process the feature values (correlation or similarity calculation, size comparison, etc.) to obtain a threshold value. Finally, they use the obtained threshold to detect the impulse noise in the noise corrupted signal.

2.1.1 Traditional template creation

Ref. [12] of various kinds of algorithms based on template shows all about algorithms without the extraction process of the template. The basic principle of the template creating is using strong directional microphone to pick up an audio signal $n(n)$ containing several impulse noises, and summing and averaging the intercepted single impulse noises from $n(n)$ as the noise template.

2.1.2 Traditional feature extraction algorithms

(1) Amplitude correlation coefficient algorithm (ACCA)

Waveforms of impulse noise template and the noise corrupted signals are compared. If the correlation coefficient is large enough, it can be considered that this part of signal is corrupted by impulse noise. If the noise free signal is $s(m)$ and impulse noise is $y(m)$, then the noise corrupted signal $x(m)$ is

$$x(m) = s(m) + y(m) \tag{1}$$

Assume that the length of the impulse template $\bar{y}(m)$ is L. Set the window length equal to the length L, and process the noise corrupted signal by window interception. Start from the first sampling point of the noise corrupted signal $x(m)$, intercept a window $x_w(m)$ with length L. Calculate the correlation coefficient cor_{ap} between the window $x_w(m)$ and the template $\bar{y}(m)$ as

$$\mathrm{cor}_{ap} = \frac{\sum_{m=1}^{L} x(m)\bar{y}(m)}{\left[\sum_{m=1}^{L} x_w^{\,2}(m) \sum_{m=1}^{L} \bar{y}(m)\right]^{\frac{1}{2}}} \tag{2}$$

After calculating the first correlation coefficient, one should glide the window past the noise corrupted signal point by point till the end, and calculate the correlation by each point. Then all the correlation coefficients could be got between the amplitude of the template and the amplitude of the noise corrupted signal.

(2) Energy correlation coefficient algorithm (ECCA)

Apart from the amplitude, another feature in time domain which can be considered to characterize the signal is the energy. Due to the impulse noise occuring with great amplitude, the energy of the pulse has great mutation.

Calculate the signal energy in the unit of frame. Set the frame length L_f. Frame the impulse noise template $\bar{y}(m)$ with the length L_f. The framing result is $\{f_1, f_2, \cdots, f_i, \cdots, f_u\}$. The number of frames is u.

Calculate the energy of the signal in each frame. Characterize the impulse noise template by energy. The energy is $E_{yf} = \{E_{yf_1}, E_{yf_2}, \cdots, E_{yf_i}, \cdots, E_{yf_u}\}$, where the frame energy of the i frame E_{yf_i} can be expressed as

$$E_{yf_i} = \sum_{a=(i-1)*L_f+1}^{i*L_f} f_i^2(a) \tag{3}$$

Similar to ACCA, the noise corrupted signal is procened by window interception first. The window length is still equal to the template length L. Then frame the noise corrupted signal in the window with the same frame length L_f, and calculate the frame energy of the noise corrupted signal in the window. The energy is $E_{xf} = \{E_{xf_1}, E_{xf_2}, \cdots, E_{xf_i}, \cdots, E_{xf_u}\}$. Calculate the correlation coefficient cor_{en} between the frame energy E_{yf} of the template and the frame energy E_{xf} of the impulse noise signal.

$$\mathrm{cor}_{en} = \frac{\sum_{i=1}^{L_f} E_{yf}(i) E_{xf}(i)}{\left[\sum_{i=1}^{L_f} E_{yf}^2(i) \sum_{i=1}^{L_f} E_{xf}^{\ 2}(i)\right]^{\frac{1}{2}}} \tag{4}$$

Glide the window past the noise corrupted signal point by point till the end, and calculate the correlation by each point. Then we can get all the correlation coefficients between the energy of the template and the energy of the noise corrupted signal.

(3) Energy histogram similarity algorithm (ESCA)

Energy histogram is distribution of the energy. After the impulse noise template is framed, the frame energy and the energy histogram are calculated. While dividing the frame energy into several energy levels, the energy histogram is the statistics on the appearance frequency (the number of frame) of each energy level.

According to ECCA, the frame energy of the template and the noise corrupted signal are calculated, respectively as E_{yf} and E_{xf}. When the energy is divided into B levels, namely the columns number of the histogram is B, the histogram h_y of the template energy and the histogram h_x of the noise corrupted signal energy are obtained which can be expressed as

$$\begin{aligned} h_y &= \{h_{y1}, h_{y2}, \cdots, h_{yi}, \cdots, h_{yB}\} \\ h_x &= \{h_{x1}, h_{x2}, \cdots, h_{xj}, \cdots, h_{xB}\} \end{aligned} \tag{5}$$

where h_{yi} represents the appearance frequency (the number of frame) of the corresponding energy of the i energy column in the energy histogram of the template, and h_{xi} represents the appearance frequency (the number of frame) of the corresponding energy of the i energy column in the energy histogram of the noise corrupted signal window.

the similarity between the two corresponding histograms are calculated as[14,15]

$$S(h_{yi}, h_{xi}) = \sum_{i=1}^{B} \min(h_{yi}, h_{xi}) \tag{6}$$

Similarity S is defined as the summation of the minimum appearance frequency (the number of frame) of the energy of the corresponding energy column in two corresponding histogram.

2.1.3 Traditional impulse noise detection

Due to the corrupted parts of the noise corrupted signal getting great characteristic value, the characteristic value and the threshold value of the characteristic are compared. Set T (experience value) to be the threshold value of the characteristic value. When the characteristic value is larger than T, the corresponding part of the impulse noise in window is considered to be impulse noise. Otherwise, it is not considered to be impulse noise. Threshold formula is shown as

$$\begin{cases} cor > T & \text{impulse} \\ \text{otherwise} & \text{not impulse} \end{cases} \tag{7}$$

2.2 Improved algorithms

The features in traditional algorithms above are not related to the amplitude of the noise free signal (m), so we obtain different thresholds for detection according to the different noise free signals in actual detecting situation. Accuracy of the threshold is directly related to the detection effect. Once the noise free signal transforms, we must modify the corresponding threshold value. In the improved detection algorithm, we use the slope as the feature, and find out the universality of the threshold, thus improving the efficiency of the detection.

2.2.1 Improved template creation

Although the traditional algorithm can quickly get the length, shape and amplitude of the template, it will eliminate the maximum pulse information and other details because of the average. In addition, the traditional interception process is done manually, increasing the complexity of the work and working time. In this paper, we improve the method of establishing the impulse template.

Due to the short impulse duration, it can be considered that the greatest slope is at the most volatile place. Therefore, we process the impulse noise $y(n)$ recorded by the strong directional microphone to find the location of the maximum slope in each impulse, namely obtain the rough position of the impulse noise in this audio signal to be referred to as a reference point. It is considered that the duration of a continuously rising or declining impulse is the time of the slope calculation. Since the impulse fluctuations have a great change in very short time, the continuous period time of rising or falling must be small enough. The change of the amplitude will be calculated according to the amplitudes of the start point and the end point of the corresponding period of time. For the impulse signal $y(n)$, the rising or falling time is

$$\Delta n = n_2 - n_1 \tag{8}$$

where n_2 is the end position of the duration of a continuously rising or declining impulse n_1 is the start position of the duration of a continuously rising or declining impulse. When n_2 corresponds to amplitude $y(n_2)$, and n_1 corresponds to amplitude $y(n_1)$, the amplitude change is

$$\Delta y = y(n_2) - y(n_1) \tag{9}$$

Slope d_n of the continuously duration can be expressed as

$$k_n = \frac{\Delta y}{\Delta n} = \frac{y(n_2) - y(n_1)}{n_2 - n_1} \tag{10}$$

After calculating slope d_n, it selects several maxima from them, namely get several reference points of the impulse.

According to the principle of the correlation, the same kinds of impulses have large correlation. With the declining of the amplitude of the impulse, the correlations gradually decrease when the amplitude fluctuation range becomes small in the end part of the impulse. Frame each impulse, and calculate the correlation. Set p reference points $\{n_1, n_2, \cdots, n_i, \cdots, n_p\}$ Frame the impulse from the reference point. Glide L sample points from the reference point to be the first frame which length is L, and set it to be $\{f_{11}, f_{21}, \cdots, f_{i1}, \cdots f_{p1}\}$, where f represents frame, the first subscript number represents the reference point, and the second subscript number represents the number of frames on the right of the reference point. For example, f_{i1} represents the first frame on the right of the i reference point.

Calculating the correlation of the first frame, the correlation coefficients cor_{f_1} between the i reference point and the first frame on the right of the j reference point can be expressed as

$$\mathrm{cor}_{f_1} = \frac{\sum_{a=1}^{L} f_{i1}(a) f_{j1}(a)}{\left[\sum_{a=1}^{L} f_{i1}^{2}(a) \sum_{a=1}^{L} f_{j1}^{2}(a)\right]^{\frac{1}{2}}} \tag{11}$$

where $f_{i1}(a)$ and $f_{j1}(a)$ are the i and j reference points of the first frame respectively. If the correlation cor_f is larger than a certain value, then continue to compare with the next frame. Otherwise, the frame is the end frame, namely it is the end position of the impulse. The start position is obtained in the same way. Knowing the reference point, the start position, and the end position of the impulse, the impulse fragment can be intercepted automatically.

For extracting the template from the impulse noise more accurately, the impulse position should be modified. The largest amplitude in the impulse is positioned to ensure that the impulse peak is not damaged. By comparing the fragment of the impulse, we obtain the start and the end positions which have the shortest distance from the peak of the impulse. If p impulses are found, the peak position is $\{q_1, q_2, \cdots, q_i, \cdots, q_p\}$, the corresponding start position of each impulse is $\{st_1, st_2, \cdots, st_i, \cdots, st_p\}$, and the corresponding end position is $\{ed_1, ed_2, \cdots, ed_i, \cdots, ed_p\}$. Then we set

$$\begin{aligned} st &= \min\{st_1, st_2, \cdots, st_i, \cdots, st_p\} \\ ed &= \min\{ed_1, ed_2, \cdots, ed_i, \cdots, ed_p\} \quad i \in [1, p] \end{aligned} \tag{12}$$

The start position of each impulse is $q_i - st$, and the end position of each impulse is $q_i + ed$, where $i \in [1, p]$, which makes sure that each impulse gets the shortest distance from the peak to the start and end point.

Then a lot of new impulse fragments are created. In the end, average these impulse fragments to get the waveform. The waveform is considered to be the template of the impulse noise.

2.2.2 Slope algorithm (SA)

Impulse fluctuation changes a greatly in short time. Slope is the ratio of the amplitude difference to the duration difference. According to Eq. (10), the slope is not related to the noise free signal $s(m)$, but only related to the amplitude difference and the duration of the impulse noise. Therefore, SA is applicable on a wider scale. And the feature extraction of SA is more accurate.

One extracts the slope characteristics from the noise corrupted signal $x(m)$, uses Eq. (1) to find the slope k_i (i is the number of rising or declining) of the continuously rising or declining waveform. Assuming that the maximum slope of the impulse noise template is k_0, and T is the threshold of the impulse noise, the corresponding detection position is the reference point of the impulse noise if the ratio of k_i to k_0 is larger than T. Otherwise, it is noise free signal.

$$\begin{cases} k_i/k_0 \geqslant T & \text{reference point} \\ \text{otherwise} & \text{noise free signal} \end{cases} \quad (13)$$

2.2.3 Universality of threshold

Due to the slope is not related to the noise free signal $s(m)$, one can only process the impulse noise $y(n)$ to obtain the threshold. In this work, $y(n)$ and Eqs(11,12) are used to obtain several impulse fragments, and Eq. (10) is used to calculate the maximum slope k_{mi}. All the slope maximums are compared to and find out the minimum k_m of them. The maximum slope k_0 of the impulse noise template is extracted. Then T is the ratio of k_m to k_0.

$$T = k_m/k_0 \quad (14)$$

According to Eq. (13), T is only related to the impulse with the slowest change in the impulse noise when the template is the same (same kind of impulse noise). As is known to all, the same kind of impulse noise has the similar waveform. Once $y(n)$ contains enough impulse noises, the threshold can be obtained.

Set $y(n)$ respectively contains 5, 8, 10, 15, 20, 25, and 40 impulse noises. The threshold result is shown in Table 1.

Table 1 Threshold the slope algorithm (SA)

Impulse number in $y(n)$	T
5	1
8	0.997
10	0.981
15	0.974
20	0.962
25	0.960
40	0.962

Analysis of Table 1 shows the threshold value is decreased. And it exists a certain universal value. When $y(n)$ contains more than 20 pulses, the threshold value equals to the universal value. Hence we set $y(n)$ contains 20 impulse noises in simulation, because Eq. (13) has a universal value in this case.

The threshold universality of the three traditional algorithms are deduced in the same way. The result is as shown in Table 2. However, the threshold of EHSA can't be well convergent.

Table 2 Threshold of the traditional algorithms

Impulse Number	T		
	ACCA	ECCA	EHSA
5	0.902	0.856	0.761
10	0.710	0.764	0.522
15	0.610	0.474	0.458
20	0.598	0.434	0.428
25	0.598	0.434	0.438

3 Simulation

The simulation verifies the performance of the improved detection algorithm by comparing the traditional detection algorithms and the proposed algorithm. In this paper, studies are carried out in two ways. First, all of the algorithms are used to simulate the detection for the same noise corrupted signal $x(m)$ with different SNR, and then all of the algorithms are used to simulate the detection for the same noise corrupted signal $x(m)$ with different number of impulse noises.

3.1 Material

The simulation environment is shown as Fig. 2. Speaker 1 plays the prelude music $s1(m)$ of the song "Liberty". Speaker 2 plays a period of voice signal $s2(m)$. Speaker 3 plays the additive white Gaussian noise $s3(m)$. Speaker 4 plays the impulse noise $y(m)$. Noise source is a Machine gun noise from the Noisex92 which is a standard noise library. The noise is generated by a. 50 caliber gun fired repeatedly. In the center of the pattern, the omnidirectional microphone is to record the mixed sound (noise corrupted signal) $x(m)$ and the microphone array is to record the impulse noise $y(n)$.

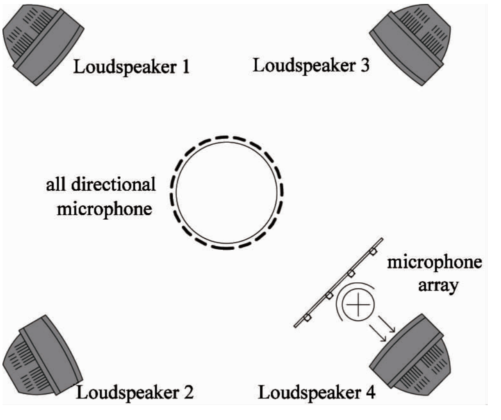

Fig. 2 Simulation environment

3.2 Evaluation

We define the Right Detection Percentage (RDP) and the Efficient Detection Percentage (EDP)[2] are defined as the main evaluation indexes for the impulse noise detection processing.

The calculation formulas are as follows:

$$\text{RDP} = (N_R/N_N) \times 100\% \quad (15)$$

$$\text{EDP} = (N_R/N_D) \times 100\% \quad (16)$$

where N_R is the number of the correctly detected impulse noise. N_N is the total number of the impulse noises in the noise resorce. N_D is the number of all the detected impulse signals. The RDP value indicates the accuracy of the detection while the EDP value indicates the effectiveness of the detection. We intend to have the maximum RDP, and at the same time to have the EDP as large as possible. Ideally, in the context of all the impulse noises are detected and no noise free signal is detected as the impulse noise signal. That is $N_R = N_N = N_D$, and RDP = EDP = 100%.

For a more comprehensive evaluation of the algorithm performance, the simulation processing duration of the computer (CPU Time) is used to be the reference evaluation of the real-time impulse noise detection algorithm.

3.3 Create impulse noise template

The microphone array records a period of impulse noise signal $y(n)$ which contains 20 Machine gun pulses. The impulse noise template is created by processing the signal $y(n)$ as shown in Fig. 3.

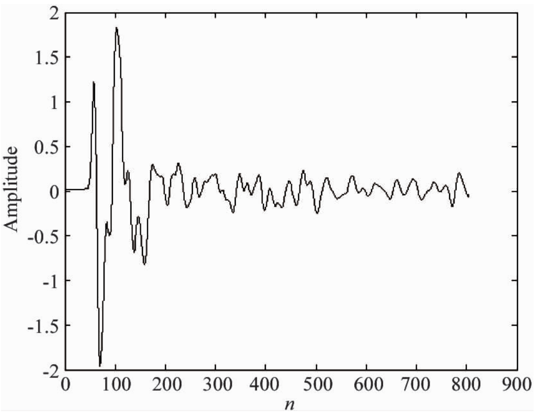

Fig. 3 Impulse noise template

3.4 Simulation Result

3.4.1 Different SNR

For the same noise corrupted signal which contains 20 pulses, the ratio of the signal to noise SNR is changed by changing the magnification of the impulse noise signal. The simulation results are shown in Table 3. Since $y(n)$ contains 20 pulses, the threshold of ACCA is 0.598, the threshold of ECCA is 0.434, the threshold of EHSA is 0.438, and the threshold of the SA is 0.962.

ACCA can accurately detect all the impulse noises by 100% when SNR is less than 0.685. But there are some error detections when the SNR is less than −8.858.

ECCA can also detect all the impulse noises when the SNR is less than 0.685. But EDP values are quite low.

Table 3 Simulation results with different SNR

Algorithms		ACCA	ECCA	EHSA	SA
x	SNR	RDP			
3	0.685	100%	100%	85.0%	100%
4	−1.814	100%	100%	90.0%	100%
5	−3.752	100%	100%	90.0%	100%
6	−5.336	100%	100%	90.0%	100%
7	−6.675	100%	100%	100%	100%
8	−7.835	100%	100%	100%	100%
9	−8.858	100%	100%	100%	100%
10	−9.773	100%	100%	95.0%	100%
Algorithms		ACCA	ECCA	EHSA	SA
x	SNR	EDP			
3	0.685	100%	69.5%	81.0%	100%
4	−1.814	100%	67.7%	79.3%	100%
5	−3.752	100%	66.8%	78.3%	100%
6	−5.336	100%	66.7%	78.3%	100%
7	−6.675	100%	62.5%	80. %	100%
8	−7.835	100%	60.6%	80. %	100%
9	−8.858	90.9%	60.6%	76.9%	100%
10	−9.773	88.3%	60.6%	76.0%	100%

EHSA cannot detect all the impulse noises and EDP values are also low.

SA can accurately detect all the impulse noise by 100% when the SNR is less than 0.685 and all EDP can reach 100%.

3.4.2 Different number of impulse noise

For the same noise corrupted signal , simulate the three traditional algorithms and the proposed slope algorithm when the impulse noise signal contains different number of impulse noises. The detection performance of the algorithms is compared. The result is shown in Table 4:

Table 4 Simulation results with different number of impulse noises

Impulse Number	ACCA	ECCA	EHSA	SA
	RDP			
5	60%	80%	40.0%	100%
10	100%	80%	80.0%	100%
15	100%	100%	93.3%	100%
20	100%	100%	95.0%	100%
25	100%	100%	96.0%	100%
Impulse Number	ACCA	ECCA	EHSA	SA
	EDP			
5	100%	100%	33.3%	100%
10	100%	88.9%	57.1%	100%
15	93.7%	46.9%	66.7%	100%
20	95.2%	51.3%	73.1%	100%
25	92.6%	59.5%	77.4%	100%

In ACCA, it can accurately detect all the impulse noises by 100%. But the detection result emergences a few wrongly detected impulses. The EDP is above 90%.

In ECCA, it can detect almost all the impulse noises. But the detection result emergences a lot of wrongly detected impulses.

In EHSA, it cannot detect few impulse noises. The more number of impulses is, the higher RDP and EDP the algorithm has. But the detection result emergences a lot of wrongly detected impulses either.

SA shows the best detection performance. It can accurately detect all the impulse noises by 100%.

3.4.3 Detection duration

For different SNR and different number of impulse noises, the detection durations of the same algorithm are similar. Average CPU time of each algorithm is used to describe the detection duration, which is shown as Table 5.

Table 5 Average CPU time of each algorithm

Algorithm	CPU Time (s)
ACCA	10.731
ECCA	2.006
EHSA	2.366
SA	2.75

Averaging the simulation result of the same kind of algorithm one can get the average detection duration, which is shown as Table 4.

3.4.4 Analysis of simulation result

In ACCA, the algorithm uses the correlation coefficient of amplitude, which is deeply affected by SNR. When SNR changes, in another word, the noise free signal or the amplitude of impulse noise changes, there will have an impact on the detection results because of the threshold, which is obtained by the comparison of the template and impulse noise signal. If impulse noises are obvious, EDP and RDP can both reach 100%. When the impulse noise amplitude is much smaller than the amplitude of the noise free signal (SNR = −8.858), EDP decreases quickly. In addition, the detection result is also affected by impulse template. The accuracy of detection algorithm will be reduced when the template extraction is inaccurate or impulse noise changes a lot. Clapping sounds, for example, may have a subtle changes varying from different clapping locations of hands, which make the result change. ACCA uses the longest time because of using point detection.

Both ECCA and EHSA use frames as units in comparison, which greatly accelerate the speed of operation. As these two algorithms are based on the energy and both of them ignore the details of the impulse signal, some parts of the noise free signal would be misjudged when they have the similar energy distribution with the impulse noise template. EHSA can only present the energy distribution without the time information, so the RDP cannot reach 100% but only to find a higher value of RDP and EDP relatively. When the impulse noise is less in the noise corrupted signal, the error of detection may cause EDP and RDP lower.

Based on the slope, which is the variation in unit time signal magnitude, SA has the least relationship with SNR, especially the amplitude of the noise free signal does not change dramatically. The applicability of this algorithm is wider than the former three, and the threshold is obtained more easily. The simulation results show that all EDP can arrive at 100%. That is to say, all the impulse noises are detected, and there is no error detection. SA aims at a period when signal rises or falls, so the operation speed is also very quick.

4 Conclusions

Conclusions are drawn below by analysis and research of the simulation.

4.1 Algorithms comparison

The comparison result of the algorithms is as follows.

In ACCA, it can accurately detect all the impulse noises by 100% when SNR is less than 0.685. But the detection result emergences a few wrongly detected impulses when the SNR is less than −8.858. The EDP is above 90%. However, the CPU time is the longest, about 10 seconds. ECCA can detect almost all the impulse noises when SNR is less than 0.685. But the detection result emergences a lot of wrongly detected impulses, namely it has low EDP. However, the CPU time is short, about 2 seconds. In EHSA, one can detect almost all the impulse noises when SNR is less than −1.814. The more number of impulses there is, the higher the RDP and the EDP the algorithm has. But the detection result emergences a lot of wrongly detected impulses either, namely it has the lowest EDP. ACCA has the best performance in among these three traditional algorithms. However the CPU time is the longest.

The proposed SA shows the best detection performance. It gets the better detection result as well as the better stability than ACCA. It can 100% accurately detect all the impulse noise when the SNR is less than 0.685. Meanwhile, the CPU time of it is about 3 seconds, 3 times less than the CPU time of ACCA. The proposed SA can help to improve the processing efficiency of impulse noise detection, so it has a high practical value of the real-time audio recording of television and radio.

4.2 Threshold universality

Four methods in this research can all find the optimal threshold. For each algorithm, the corresponding optimal threshold, as well as the SNR of the detection algorithm, is decreased with the growth in the number of impulse noise in a certain range. However, the same impulse noise has a specific universal threshold with the increase of the number of the impulse noise, namely the threshold is not related to SNR within a certain rang of the number of the impulse noises. Therefore, once the number of impulse noises is enough, we can get a universal threshold value based on the improved algorithm.

4.3 Future work

The four algorithms in this paper are researched by simulations. By improving the extraction of the threshold, we will record more impulse noise signals to create a library of impulse noise template and the corresponding threshold database. In further study, a subjective evaluation[17] is introduced for the impulse noise detection to research the detection performance of the proposed algorithms.

References

[1] Nongpiur R C. Impulse noise removal in speech using wavelets. In: International Conference on Acoustics, Speech and Signal Processing, Las Vegas, USA, 2008. 1593-1596

[2] Wang J Y, Zhang X, Zhu J M, et al. Impulse noise suppression method based on time frequency spectrum. *Journal of Vibration and Shock*, 2010, 29(2):149

[3] Rajagopalan R, Subramanian B. Removal of impulse noise from audio and speech signals. In: Proceedings of the International Symposium on Signal, Circuits and Systems, lasi, Romania, 2003. 161-163

[4] Vaseghi S V. Advanced Digital Signal Processing and Noise Reduction. USA: John Wiley & Sons Publication, 2008. 341-358

[5] Kauppinen I. Methods for detecting impulsive noise in speech and audio signals. In: Proceedings of the 14th International Conference on Digital Signal Processing, Santorini, Greece, 2002, 2: 967-970

[6] Niedzwiecki M, Ciotek M. Elimination of clicks from archive speech signals using sparse autoregressive modeling. In: Proceedings of the 20th European on Signal Processing Conference, Bucharest, Romania, 2012. 2615-2619

[7] Niedzwiecki M, Ciolek M. Elimination of impulsive disturbances from archive audio signals using bidirectional processing. *IEEE Transactions on Audio, Speech, and Language Processing*, 2013, 21(5): 1046-1059

[8] Nongpiur R C, Shpak D J. Impulse-noise suppression in speech using the stationary wavelet transforms. *The Journal of the Acoustical Society of America*, 2013, 133: 866

[9] Li M, Lv P, Han J, et al. Implementation of the regular audio retrieval. In: The National Network and Information Security Technology Conference, Beijing, China, 2004. 90-92

[10] Burred J J, Lerch A. Hierarchical automatic audio signal classification. *Proc of J Audio Eng Soc*, 2005, 52: 724-739

[11] Kim K M, Kim S Y, Jeon J K, et al. Quick audio retrieval using multiple feature vectors. *IEEE Transactions on Consumer Electronics*, 2006, 52(1):200-205

[12] KAO W C, Hung Y W, Hsu C C, et al. adaptive fixed Pattern audio noise cancellation for digital camera applications. In: Proceedings of the 10th International Symposium on Consumer Electronics, Dallas, USA, 2006. 1-5

[13] Kuroiwa S, Mori Y, Tsuge S, et al. Wind noise reduction method for speech recording using multiple noise templates and observed spectrum fine structure. In: International Conference on Communication Technology, Guilin, China, 2006. 1-5

[14] Zhen W, Zhang J D, Lu Y. Research on background noise suppression method of embedded digital recording. *Journal of Electronic Measurement and Instrument*, 2006, 20(5): 56-59

[15] Tan H X, Chen F, Li S M. Quick regular audio retrieval method based on template-subspace. *Computer Engineering*, 2012, 38(20): 260-263

[16] Yao J C, Wan W G, Yu X Q, et al. A quick specific audio retrieval algorithm based on general prediction. In: International Conference on Audio, Language and Image Processing, Shanghai, China, 2008. 1180-1184

[17] Zhao W, Wu Y, Jin C, et al. Comparative study of DMS and conventional stereo based on subjective evaluation experiment. In: International Conference on Intelligent Control and Information Processing, Beijing, China, 2013. 236-240

Zhao Wei, born in 1981. She is currently a Ph. D candidate working in the National Key Laboratory of Media Audio & Video funded by the Ministry of Education. She received her M. S. degrees in Communication University of China in 2007. She also received her B. S. degrees from Beijing Polytechnic University in 2003. Her research interests include network information technology, audio signal processing.