Research on Endpoint Detection Algorithm in High Altitude Explosion Point Location Technology

Yafei Wang

School of Communication Engineering
Chengdu University of Information Technology
Chengdu, China

Wen Bin

School of Communication Engineering
Chengdu University of Information Technology
Chengdu, China

Abstract: Aiming at the problem of inadequate positioning accuracy of sound endpoints by the dual-element single-threshold endpoint detection algorithm and the single-element dual-threshold endpoint detection algorithm in the process of locating the sound source of weather modification bombs at high altitudes during artificial weather modification, a multi-element dual threshold endpoint detection algorithm. First, according to the characteristics of high-altitude explosion of weather modification bombs and ground reception, the sound signal is filtered and denoised, divided into frames, and windowed. Then, the time-domain feature short-term energy, short-term zero-crossing rate and frequency domain feature short-term information entropy of each frame of the sound signal are calculated, and double thresholds are set for detection. In this way, the start and end points of the explosion sound in the collected sound signal are found, and the data is imported into the positioning algorithm for processing, and then the explosion point of the weather modification bombs in the high air is located. The test results show that the method can accurately distinguish the end point of effective explosion sound, and has practical application value for the location of the sound source of the high-altitude explosion point of the weather modification bombs.

Keywords: Endpoint detection; Multi-element double threshold; Short-term stability; Weather Modification; Sound source localization

1. INTRODUCTION

Artificial weather influencing is to artificially influence the physical process of the local atmosphere through anti-aircraft artillery operations under appropriate conditions to achieve the purpose of increasing rain and snow, preventing hail, eliminating rain, eliminating fog, and preventing frost^[1]. The use of sound source localization technology can effectively measure the high-altitude explosion location information of human shadow bombs, and at the same time effectively evaluate the effect of human shadow operations. When the sound source positioning system is used to locate the explosion point of the shadow bomb, only a small part of the large amount of sound data collected by the system belongs to the sound information of the shadow bomb explosion. A large amount of other redundant information makes the positioning system unable to transmit or process the data in real time. Reduce the performance of the system^[2].

Voice endpoint detection algorithm is a common method of voice signal preprocessing [3]. Literature [4] proposed an advanced quantum-based endpoint detection algorithm to process audio data in a quantum computer, but it is temporarily not suitable for embedded audio processing. Literature [5-7] starts with time domain features, and literature [8] starts with frequency domain features. They study detection algorithms based on short-term energy, detection algorithms based on short-term zero-crossing rate, and short-term energy and short-term energy. Zero-crossing rate detection algorithm and detection algorithm based on spectral entropy. Literature [9-11] combined frequency domain features and time domain features, and studied detection algorithms based on short-term energy and information entropy and detection algorithms based on Mel frequency cepstrum coefficients. Literature [12] uses wavelet transform convolution method to detect. These algorithms have achieved certain results in voice endpoint detection.

Aiming at practical application requirements, this paper proposes a multi-element double-threshold voice endpoint detection algorithm based on the time-domain and frequency-domain features of high-altitude explosion sounds. First, filter a segment of speech signal, divide it into frames, and preprocess it with windows, and then calculate the short-term average energy, short-term average zero-crossing rate and short-term information entropy of each frame of speech. Each element is set with dual thresholds. The start point and end point of the collected explosive voice information are determined. The multi-element dual-threshold voice endpoint detection algorithm is convenient for dynamic real-time processing or transmission of effective voice data, and it can also improve the timeliness of the positioning algorithm [13].

2. POSITIONING ALGORITHM

Aiming at the problem of estimating the position of the high-altitude bombing point of human shadow bombs in order to adjust the launch angle and improve the operation effect, a time-delay-based sound source positioning system with low computational complexity and strong real-time performance is designed. The system uses a four-element cross pickup array to collect sound signals in real time. And processing.

2.1 Time Difference Of Arrival

The sound source localization algorithm based on time delay calculates the time delay between the sound source reaching each sensor, and then uses the minimum distance mean square error criterion to solve the azimuth angle, pitch angle and distance of the sound source^[14].

Figure 1. Four-element cross localization model diagram

The construction of the four-element cross sound source localization model is shown in Figure 1. The four high-precision capacitive pickups are located at M1(d,0,0), M2(0,-d,0), M3(-d,0,0), M4(0,d,0). The origin coordinate is O(0,0,0), and the sound source is located at S(x,y,z). The projection of point S on the XY plane is S'. The azimuth angle of the sound source point S is φ, and the pitch angle is θ. The azimuth angle is the angle between the positive semi-axis of the X axis and the counterclockwise, and the pitch angle is the angle between the positive semi-axis of the Z axis.

Establish a system of equations, and then according to the minimum distance mean square error criterion, we can get:

$$\text{distance: } SO \approx \frac{c \cdot (\tau_{12}^2 + \tau_{14}^2 - \tau_{13}^2)}{2 \cdot (\tau_{13} - \tau_{12} - \tau_{14})} \quad (1)$$

$$\text{Azimuth: } \varphi \approx \arctan\left(\frac{\tau_{14} - \tau_{12}}{\tau_{13}}\right) \quad (2)$$

$$\text{Pitch angle: } \theta \approx \arcsin\left(\frac{c \cdot \sqrt{\tau_{13}^2 + (\tau_{14} - \tau_{12})^2}}{2 \cdot d}\right) \quad (3)$$

Among them, $\tau_{12}, \tau_{13}, \tau_{14}$ are the time delays of the sound source reaching sensor 1 and reaching sensors 2, 3, 4, c is the speed of sound, and d is the distance between the sensors.

2.2 Time Delay Estimation Method

When the sound source is located $r > 2 \cdot d^2 / \lambda$, the sound source is considered to be in the far field position, and the sound source is a plane wave when it reaches the sensor, which is the maximum wavelength of the sound source. Ignoring the amplitude difference between the received signals of each array element, it is approximately considered that there is a simple time delay relationship between the received signals. Suppose any two sensors and the received signals are respectively

$$x_i(m) = s(m - \tau_i) + n_i(m) \quad (4)$$

$$x_j(m) = s(m - \tau_j) + n_j(m) \quad (5)$$

They are N-point discrete signals sampled by ADC. Among them s(m) are the sound source signal, $n_i(m)$ and $n_j(m)$ white noise, τ_i and τ_j the time that the sound propagates to the sensor M_i and M_j , the time delay between the sensor M_i and M_j is $\tau_{ij} = \tau_i - \tau_j$. The cross-correlation function of the received signal can be expressed as:

$$R_{ij}(\tau) = E[x_i(m) \cdot x_j(m + \tau)] \quad (6)$$

Because s(m), $n_i(m)$, $n_j(m)$ are not related to each other, we can get equations (4) and (5) into equation (6) and simplify

$$R_{ij}(\tau) = E[s(m - \tau_i) \cdot s(m - \tau_j - \tau)] = R_{ss}(\tau - \tau_{ij}) \quad (7)$$

It can be known from the nature of the autocorrelation function that at that time $\tau = \tau_{ij}$, the cross-correlation function has the largest peak. However, in the actual environment, the cross-correlation function will produce many peaks due to non-Gaussian white noise. When the reflected wave is too strong, the cross-correlation peak of the direct signal will be overwhelmed, resulting in inaccurate experimental estimation. In order to suppress the influence of noise, the generalized cross-correlation method can be used to weight the cross-correlation function in the frequency domain to make the peak of the cross-correlation function more obvious^[15].

3. SIGNAL PREPROCESSING

When the sound signal is used for positioning, the final sample value matrix of the sound signal imported into the positioning algorithm should be the sound fragment of the human shadow bomb explosion in the air. The proportion of the effective sound segment in this sound has a great influence on the positioning accuracy. Therefore, it is necessary to distinguish between a section of sound containing noise (unused sound section) and explosive sound (effective sound section), and extract the effective sound section containing only explosive sound or as little noise sound section as possible. The preprocessing of sound signals mainly includes filtering, framing, windowing.^[16]

3.1 Filter

The signal contains various noises that will cause interference with short-term energy, short-term zero-crossing rate, and short-term information entropy. Therefore, it is necessary to use a filter to filter out some noise. According to the sound attenuation, the sound of the explosion can be known from the sound attenuation after being spread over a long distance in the air. The energy of the explosion wave reaching the acquisition system is mainly concentrated below 1000 Hz. Therefore, the filter is a Butterworth bandpass filter with a passband frequency of 50Hz~1000Hz. As shown in the band-pass filter diagram in Figure 2, the upper half is the original signal collected by the first sensor, and the lower half is the original signal after the band-pass filter. The filtering effect is ideal.

Figure 2. Bandpass filter diagram

3.2 Framing and Windowing

Fourier transform is to study the relationship between the entire time domain and the frequency domain. When processing sound signals, it is impossible to measure and calculate infinitely long signals. Therefore, the sound signal is divided into frames, and the long sound signal in a period of time is divided into multiple shorter frames, and then the short-term energy, short-term zero-crossing rate and short-term information entropy can be calculated for each frame. . After an infinite signal is truncated, its spectrum is distorted, which is also called spectral energy leakage. In order to reduce spectrum leakage, different interception functions can be used to intercept the signal. The interception function is called a window function, and the sound signal is windowed. More stable [17]. The sound signal is a random signal. According to the recommendation in [18], the Hanning window is selected to process the sound signal, the frame length is set to 1024 sampling points, and the frame shift is 512 sampling points.

4. ENDPOINT DETECTION

4.1 Short-term Energy

The difference between the sound segment and the noise segment is particularly obvious in terms of energy. The energy of the sound segment is obviously higher than the energy of the noise segment. At the same time, the energy of the sound segment is the sum of the noise energy superimposed on the sound energy [18]. As shown in Figure 3, when the signal-to-noise ratio is very high, only the short-term average energy can effectively distinguish the sound segment from the background noise segment. Within the error allowed by the positioning algorithm.

The short-term energy is defined as:

$$E_n = \sum_{m=-\infty}^{\infty} [x(m) \cdot w(n-m)]^2 \quad (8)$$

It can also use the short-term average amplitude to replace the short-term energy. The short-term average amplitude is defined as:

$$M_n = \sum_{m=-\infty}^{\infty} |x(m)| \cdot w(n-m) \quad (9)$$

Among them, $w(n)$ is the window function.

Figure 3. The effect of noise on short-term energy

4.2 Short-term Zero crossing Rate

The number of times a frame of sound crosses the zero point, that is, the horizontal axis, is defined as a short-term zero crossing. For continuous sound signals, zero crossing means that the time-domain waveform crosses the time axis. For discrete signals, if the sign of adjacent sampling points changes and the product of two adjacent sampling points is less than zero, it means zero crossing. The zero-crossing rate is the number of sign changes.

The short-term average zero-crossing rate is defined as:

$$Z_n = \frac{1}{2} \sum_{m=-\infty}^{\infty} |\text{sgn}[x(n)] - \text{sgn}[x(m-1)]| \cdot w(n-m) \quad (10)$$

Among them, $\text{sgn}(n)$ is the symbolic function.

From the definition of the zero-crossing rate, it can be seen that the zero-crossing rate is easily interfered by low-frequency signals, and part of the noise is concentrated in the low-frequency part. Therefore, before calculating the short-term zero-crossing rate, a high-pass filter can be used to filter out the low-frequency noise, and then the zero-crossing rate can be calculated. You can also perform median filtering on the data, then modify the traditional zero point and set the zero point threshold T , so that zero crossing means crossing the positive and negative threshold T [19]. It can effectively reduce the false zero-crossing rate caused by noise and improve the efficiency of endpoint detection.

This article chooses the second method, and through research, it is found that if the DC component is not removed during the filtering process, or a small DC component is added after removing the DC component of the original signal, its effect is the same as setting the zero threshold. . As shown in Figure 4, the DC component is very small, and not using the average filter is equivalent to setting the zero threshold T . However, as the signal-to-noise ratio decreases, the false zero-crossing rate caused by noise will make it impossible to distinguish the start and end points of the sound.

Figure 4. The effect of noise on the short-term zero-crossing rate

4.3 Information Entropy

For a discrete source, when it is composed of several random events, the uncertainty of the occurrence of a random event is described by the probability of its occurrence. The smaller the probability of an event, the smaller the probability and the greater the amount of information contained. On the contrary, the greater the likelihood of an event, the greater the probability and the smaller the amount of information

contained. Therefore, the short-term average amount of information provided by the source is expressed as:

$$H(X) = -\sum_{i=1}^n p_i \log_2 p_i \quad (11)$$

Since the amplitude of the sound signal has a large dynamic range relative to the amplitude of the background noise, it can be considered that the random event of the sound signal in the range $(-M, M)$ is large, that is, the information entropy is large, and the invalid sound segment only contains the amplitude of the background noise Small, the distribution is relatively concentrated, so the information entropy is small. As shown in Figure 5, as the signal-to-noise ratio decreases, the short-term information entropy fluctuates to a certain extent, but the start and end points of the sound can still be distinguished. According to the above principles, after constructing the entropy function, the information entropy of each frame of signal can be calculated.

Figure .5 The effect of noise on short-term information entropy

4.4 Multi-element and Double threshold Detection

The multi-element double-threshold method is used to locate the sound source of the explosion point of the shadow bomb to find out the part of the speech that contains the explosion sound. The multi-element double-threshold method is realized by using the three-level judgment of the short-term energy of the sound, the short-term zero-crossing rate and the short-term information entropy.

The first level of judgment, as shown in Figure 6, a higher threshold TH1 is selected on the short-term energy envelope of the sound, and the one higher than the TH1 gate, that is, between the AB segments must be speech, and the start and end points of the speech should be located at this threshold. Outside the time point corresponding to the point of intersection with the short-term energy envelope. Determine a lower threshold TH2 on the average energy, and search from point A to the left and from point B to the right to find the two points C and D where the short-term energy envelope intersects the threshold TH2, so the CD segment is used. The double-threshold method determines the start and end positions of the sound segment based on the short-term energy.

The second level of judgment is based on short-term information entropy. Two thresholds TH3 and TH4 are set on the short-term information entropy envelope. Using the same judgment method as the short-term energy, the double-threshold method of short-term information entropy can be obtained. The starting and ending point positions of the

obtained voice. The start and end points obtained by the short-term energy and short-term information entropy are weighted and averaged to obtain the dual-element dual-threshold speech start and end points.

The third level of judgment is based on the short-term average zero-crossing rate. From the weighted average starting and ending points, search from point C to the left and from point D to the right, and find two short-term average zero-crossing rates lower than the threshold TH5. Points M and N are the starting and ending points of the sound segment determined by the multi-element double-threshold algorithm.

Figure .6 Voice endpoint detection diagram

5. EXPERIMENT ANALYSIS

5.1 Experimental Environment

The experimental environment was chosen in an empty playground to reduce the reflection and reverberation of sound. The sound source chooses cowhide drums to replace the explosion sound, the sound frequency is 200Hz~300Hz, and the sound source is 10m away from the center of the pickup array. Start at 0° and collect twice every 45°. In the first collection, hit the cowhide drum twice, and in the second collection, hit the cowhide drum once.

5.2 Data Collection and Analysis

The sampling rate of the data acquisition system is set to 100kHz, and the four channels are collected simultaneously. A total of sixteen sets of data are obtained, and each set of data has four channels.

Using a multi-element double-threshold sound endpoint detection algorithm in MATLAB, each set of data can be used to obtain the sound starting point and ending point of the signals received by the four sensors. The starting point is used as the starting point for the sensor array to receive the sound signal, and the largest ending point is used as the ending point of the sound. The starting point and ending point of the sound signal received by the sensor array can be obtained.

Table 1 compares the results of single-sound four-channel sound endpoint detection and manual marking of the start and end points of the sound, and the error is within one frame. When locating the sound source, you only need to know the starting point of the sound, and then take 5-10 frames of data from the starting point backwards, and then you can use generalized cross-correlation to get the time delay between each sensor for positioning. Moreover, the sound has a certain length of tail, and it is difficult to manually mark the end point, so the end point of the sound is not manually marked.

Table 1 Starting and ending point test results table

	Detection algorithm		Manual marking	
	Starting point	End point	Starting point	Error
Ch.1	123905	135680	123960	-55
Ch.2	123905	134656	123990	-90
Ch.3	123905	133632	124290	-385
Ch.4	123905	134656	124000	-95
Ch_All	123905	135680	123960	-55

Figure 7. Voice endpoint detection diagram

Figure 8. Sixteen data sets of sound starting point errors

Figure 7 Voice endpoint detection The picture shows the voice endpoint detection including two voices. It can be seen that the algorithm accurately marked the start and end points of the two speech signals. Figure 8 shows the error between the 16 sets of data voice endpoints detecting the start and end points of the marked voice and the start point of the manual marking. It can be seen that the multi-element double-threshold algorithm effectively marks the starting point of the speech, and the detection error meets the needs of the positioning algorithm.

6. CONCLUSIONS

This paper uses the short-term stability characteristics of the sound signal and proposes a multi-element double-threshold algorithm using short-term energy, short-term zero-crossing rate and short-term information entropy to determine the start and end points of the sound. Experimental results show that the algorithm can effectively mark the start and end points of the sound, and can provide effective speech fragments for the sound source localization algorithm. In this paper, the algorithm processing of adaptive threshold is not implemented. The threshold is only derived from a large amount of experimental data, and when the signal-to-noise ratio drops below 5dB, a large number of misjudgments will occur in the short-term zero-crossing rate and lead to errors in the results. In the next step, the adaptive threshold sound endpoint detection will be studied to achieve more accurate sound endpoint detection.

7. REFERENCES

- [1] Huanbin Xu, Jinfang Yin. Key issues in developing numerical models for artificial weather modification[J]. Journal of Meteorological Research, 2017, 31(6).
- [2] Nan Jiang, Ting Liu, Thomas Hanne. An Improved Speech Segmentation and Clustering Algorithm Based on SOM and K-Means[J]. Mathematical Problems in Engineering, 2020, 2020.
- [3] Changqin Quan, Bin Zhang, Xiao Sun, Fuji Ren. A combined cepstral distance method for emotional speech recognition[J]. International Journal of Advanced Robotic Systems, 2017, 14(4).
- [4] Jian Wang, Han Wang, Yan Song. Quantum Endpoint Detection Based on QRDA[J]. International Journal of Theoretical Physics, 2017, 56(10).
- [5] Wang Huasong, Zhao Xin. An accurate time delay extraction method based on endpoint detection and its positioning accuracy analysis [J/OL]. Acoustic technology: 1-6 [2021-05-18].
- [6] Li Xuezheng. Design of whistle vehicle identification and positioning system for car whistle monitoring [D]. Hebei University,
- [7] Yao Lina, Lu Xutao, Sun Yunqiang, Chai Jinqiang. Research on Endpoint Detection Algorithms in the Positioning Technology of Rain Enhancement Bomb and Blast[J]. Foreign Electronic Measurement Technology, 2020, 39(01): 6-9.
- [8] Ling Yu. Design and implementation of sound source positioning system [D]. Hebei University of Science and Technology, 2015.
- [9] Xuan Zhangjian, Cai Xiaoxia, Chu Dingli. Research on a voice endpoint detection method based on improved information entropy[J]. Communication Technology, 2018, 51(06): 1302-1306.
- [10] Danyang Cao, Xue Gao, Lei Gao. An Improved Endpoint Detection Algorithm Based on MFCC Cosine Value[J]. Wireless Personal Communications, 2017, 95(3).
- [11] Yongwon Hwang, Mun-Ho Jeong, Sang-Rok Oh, Il-Hwan Kim. Applying the Bi-level HMM for Robust Voice-activity Detection[J]. Journal of Electrical Engineering & Technology, 2017, 12(1).
- [12] Tanmoy Roy, Tshilidzi Marwala, Snehashish Chakraverty. Precise Detection of Speech Endpoints Dynamically: A Wavelet Convolution based approach[J]. Communications in Nonlinear Science and Numerical Simulation, 2018, 67.
- [13] Hall J I, Lozano M, Estradapetrocelli L, et al. The present and future of cough counting tools[J]. Journal of Thoracic Disease, 2020, 12(9): 5207-5223.

[14] Zhang Xiaoguang, Lu Haifeng, Lu Chuanmao, Wang Puhao, Han Dong, Chen Longhu. Research on the sound source localization algorithm of four-element cross microphone array [J]. China Test, 2020, 46(02): 96-102.

[15] Zhou Hefeng, Tian Zhangfu, Zeng Yang, Gong Changchao, Zeng Xinwu, Jiang Weikang. Improved time delay estimation based on generalized cross-correlation algorithm [A]. Editorial Department of Acoustic Technology, 2020: 3.

[16]Zeng Jianfei. Research on Speech Endpoint Detection Algorithm under Low SNR[D]. South China University of Technology, 2019.

[17] Xing Lizhao. Research on Speech Signal Endpoint Detection Algorithm [D]. Zhengzhou University, 2016. Piñal Moctezuma Fernando, Delgado Prieto Miguel, Romeral Martínez Luis. An acoustic emission activity detection method based on short-term waveform features: Application to metallic components under uniaxial tensile test[J]. Mechanical Systems and Signal Processing, 2019, 142.

[18]X Zhang, Cheng D, Jia P, et al. An Efficient Android-Based Multimodal Biometric Authentication System with Face and Voice[J]. IEEE Access, 2020, PP(99):1-1.