PAN
ARCHIVES OF ACOUSTICS
Vol. **45**, No. 3, pp. 453–458 (2020)
DOI: 10.24425/aoa.2020.134061

Research Paper

The Combination of Spectrum Subtraction and Cross-power Spectrum Phase Method for Time Delay Estimation

Feng BIN*, Xu LEI

School of Opto-electronical Engineering
Xi'an Technological University
Xi'an 710021, China
*Corresponding Author e-mail: fengbin98@126.com

(*received February 2, 2020; accepted May 9, 2020*)

In order to solve the problem of large error of delay estimation in low SNR environment, a new delay estimation method based on cross power spectral frequency domain weighting and spectrum subtraction is proposed. Through theoretical analysis and MATLAB simulation, among the four common weighting functions, it is proved that the cross-power spectral phase weighting method has a good sharpening effect on the peak value of the cross-correlation function, and it is verified that the improved spectral subtraction method generally has a good noise reduction effect under different SNR environments. Finally, the joint simulation results of the whole algorithm show that the combination of spectrum subtraction and cross-power spectrum phase method can effectively sharpen the peak value of cross-correlation function and improve the accuracy of time delay estimation in the low SNR environment. The results of this paper can provide useful help for sound source localization in complex environments.

Keywords: sound source localization; time delay estimation; generalized cross correlation; phase of cross power spectrum; spectral subtraction.

1. Introduction

Sound source location technology is to use the sound sensors to obtain sound signals, and use digital signal processing technology to analyze and process them, so as to determine and track the location of sound source, which is widely used in video phone, video conference, monitoring system, intelligent robot, voice recognition and other fields (Wei, 2018). In addition, acoustic source location technology also plays an important role in seismic research, underwater target location, nondestructive testing of pressure vessels, measurement of cabin drop point in aerospace field, and measurement of impact point in military test (Danicki, 2005; Kai *et al.*, 2015).

There are three types of acoustic array source location algorithms:

1) a controlled beam-forming method based on maximum output power;
2) a sound source localization method based on high resolution spectral estimation;
3) localization method based on time difference of arrival (TDOA) (Zhang *et al.*, 2013).

Among the three kinds of positioning methods, the positioning technology based on time delay estimation is the most widely used in the industrial and military fields because of its simple principle and small calculation amount (Wang *et al.*, 2010). This method can be divided into two steps. The first step is time delay estimation, which uses the geometric relations of different array elements to solve the time delay of sound source arriving at different array elements. The second step is sound source location, which uses geometric algorithm or search algorithm combined with the time delay obtained in the previous step to locate the final sound source location (Cheng, Yang, 2015; Duan, Doa, 2014). The existing acoustic array positioning technology mainly adopts the cross-correlation method for time delay estimation. This method has certain anti-noise and reverberation ability under ideal conditions, but its performance declines significantly under real conditions such as low signal-noise ratio and strong reverberation (Wang, Wu, 2014; Dong, 2016; Yang *et al.*, 2018). Aiming at this problem, the traditional cross-correlation method was studied and improved, and spectrum subtraction and cross-power spectrum

phase (CSP) were introduced to achieve higher accuracy of sound source location.

2. Theoretical analysis of time delay estimation

Basic cross-correlation analysis is a method to compare the similarity of two signals in the time domain, and estimate the time delay by searching the maximum peak of the cross-correlation curve of two sound signals in the time domain.

Let S be the sound source, M_1 and M_2 are the two sound sensors in the array, and $x_1(t)$ and $x_2(t)$ are the signals picked up by M_1 and M_2, which can be expressed as:

$$x_1(t) = s(t-\tau_1) + n_1(t), \tag{1}$$

$$x_2(t) = s(t-\tau_2) + n_2(t). \tag{2}$$

In the above formulas, the sound source signal is represented by $s(t)$, $n_1(t)$ and $n_2(t)$ are used to represent the Gaussian white noise of the two sensors, τ_1 and τ_2 represent the time when the sound source reaches the sensor M_1 and M_2, so the time delay for the sound wave to reach the two sensors is $\tau_{12} = \tau_1 - \tau_2$. Suppose that $s(t)$ is independent of $n_1(t)$ and $n_2(t)$, then the cross-correlation function of $x_1(t)$ and $x_2(t)$ is:

$$R_{x_1x_2}(\tau) = E[x_1(t)x_2(t-\tau)]. \tag{3}$$

$R_{x_1x_2}(\tau)$ represents the cross-correlation function of signal $x_1(t)$ and $x_2(t)$. By taking Eqs (1) and (2) into expansion, we can get:

$$\begin{aligned} R_{x_1x_2}(\tau) &= E[s(t-\tau_1)s(t-\tau_1-\tau)] \\ &\quad + E[s(t-\tau)n_2(t-\tau)] \\ &\quad + E[s(t-\tau_2-\tau)n_1(t)] \\ &\quad + E[n_1(t)n_2(t-\tau)]. \end{aligned} \tag{4}$$

Since $s(t)$, $n_1(t)$, and $n_2(t)$ are not related to each other, formula (4) can be simplified to

$$\begin{aligned} R_{x_1x_2}(\tau) &= E[s(t-\tau_1)s(t-\tau_1-\tau)] \\ &= R_{ss}(\tau-(\tau_1-\tau_2)). \end{aligned} \tag{5}$$

where R_{ss} is the autocorrelation function of the sound source signal $s(t)$, according to the property of self-correlation function, $R_{x_1x_2}(\tau)$ is at its maximum when $\tau-(\tau_1-\tau_2)=0$. In this case, the corresponding value of τ is the delay value between the two sound sensors. The estimated time difference between the sound waves reaching the sensors M_1 and M_2 can be expressed as:

$$\widehat{\tau} = \arg\max_{\tau} R_{x_1x_2}(\tau). \tag{6}$$

In the formula (6), $\tau \in [-\tau_{\max}, \tau_{\max}]$ is the maximum of possible delay values, $\arg\max_{tau} R_{x_1x_2}(\tau)$ is the value of τ when $R_{x_1x_2}(\tau)$ is maximized.

In practical application, the signals received by the sensor are all noisy, and there may be no obvious peak after cross-correlation estimation, which makes the delay estimation error larger.

The generalized cross-correlation method is based on the basic cross-correlation theory. When calculating the cross power spectrum between two signals, the signal and noise can be whitened by giving a weighted function in the frequency domain. At the same time, it is helpful to increase the frequency proportion of the signal-to-noise ratio, so as to suppress the noise power. Finally, the inverse Fourier transform is applied to the time domain to obtain the cross-correlation function between the two signals, which can make the peak value of the cross-correlation function sharper and improve the accuracy of the delay estimation. The steps of the generalized cross-correlation time-delay estimation method are shown in Fig. 1.

The Fourier transform of formula (5) is performed to obtain the cross power spectrum of signals $x_1(t)$ and $x_2(t)$:

$$G_{x_1x_2}(\omega) = G_{ss}(\omega)e^{-j\omega\tau_{12}}. \tag{7}$$

Formula (7) is weighted, and then inverse Fourier transform is carried out to the time domain, as shown in formula (8):

$$R^g_{x_1x_2}(\tau) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} G_{x_1x_2}(\omega)\psi_{12}(\omega)e^{j\omega\tau}\,\mathrm{d}\omega. \tag{8}$$

In formula (8), $G_{x_1x_2}(\omega)$ is the cross-power spectrum of signals $x_1(t)$ and $x_2(t)$, and $\psi_{12}(\omega)$ is the generalized cross-correlation weighting function. In practical applications, different weighting functions are selected for different types of noise and reverberation, so as to make the peak value of $R^g_{x_1x_2}(\tau)$ more obvious. But in this process, it is not stable because of strong noise and finite window, so the choice of weighted fun-

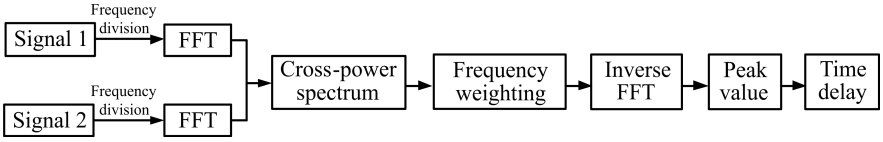

Fig. 1. Generalized cross correlation step diagram.

ction $\psi_{12}(\omega)$ becomes a difficulty. Table 1 shows three common weighting functions of SCOT (Smooth Coherent Transformation), Roth Processor and CSP (Cross-power Spectrum Phase).

Table 1. Three common weighting functions.

Algorithm	Weighted function
1. SCOT	$\psi_{12}(\omega) = \frac{1}{\sqrt{G_{x_1x_1}(\omega)G_{x_2x_2}(\omega)}}$
2. Roth	$\psi_{12}(\omega) = \frac{1}{G_{x_1x_2}(\omega)}$
3. CSP	$\psi_{12}(\omega) = \frac{1}{\lvert G_{x_1x_2}(\omega)\rvert}$

3. Experimental simulation and verification

To verify the actual performance of the algorithm, a voice is collected by a recorder with a sampling frequency of 8 kHz to simulate the source signal $x_1(t)$ received by the sensor M_1. Signal $x_2(t)$ is obtained by delaying $x_1(t)$ by D sampling cycles to simulate the acoustic signal received by sensor M2. The random white Gaussian noise is added to the signal $x_1(t)$ and $x_2(t)$ respectively. We assume that the noise is independent of the sound source and of the noise each other, so Hamming window is adopted for frame division, in which frame length is 1024 and frame shift is 512. The performances of the four weighted algorithms are compared under different SNR. The simulation results are expressed by the cross-correlation function, and the precision of the estimated delay can be shown by the peak sharpness of the cross-correlation function. The actual delay is 800 sampling cycles. Figure 2 shows the performance comparison of the four weighted methods at a SNR of 10 dB.

As can be seen from Fig. 2, under the condition of little noise interference, the time delay can be estimated more accurately by these four methods. Fig. 2c and Fig. 2d have obvious sharpening peak effect, and the sharpening effect of Fig. 2b and Fig. 2d is better. In particular, the CSP weighting method eliminates the amplitude information of the signal and only retains the phase characteristics of the signal by normalizing the cross-power spectrum of the signal, which has a good suppression effect on noise and reverberation. However, in practical applications, the signals received by the sensor are often interfered by large noises. In order to compare the performance of the four methods in a high-noise environment, the SNR is adjusted to 0 dB and then the time delay estimation is carried out. The simulation results are shown in Fig. 3.

As shown in Fig. 3, with the reduction of SNR of 0 dB, the performance of the four methods decreased to varying degrees, making it difficult to accurately estimate the time delay value. The performance degradation of methods Fig. 3a and Fig. 3c is relatively obvious, and the peak value of basic cross-correlation function is almost annihilated in the interference. The main

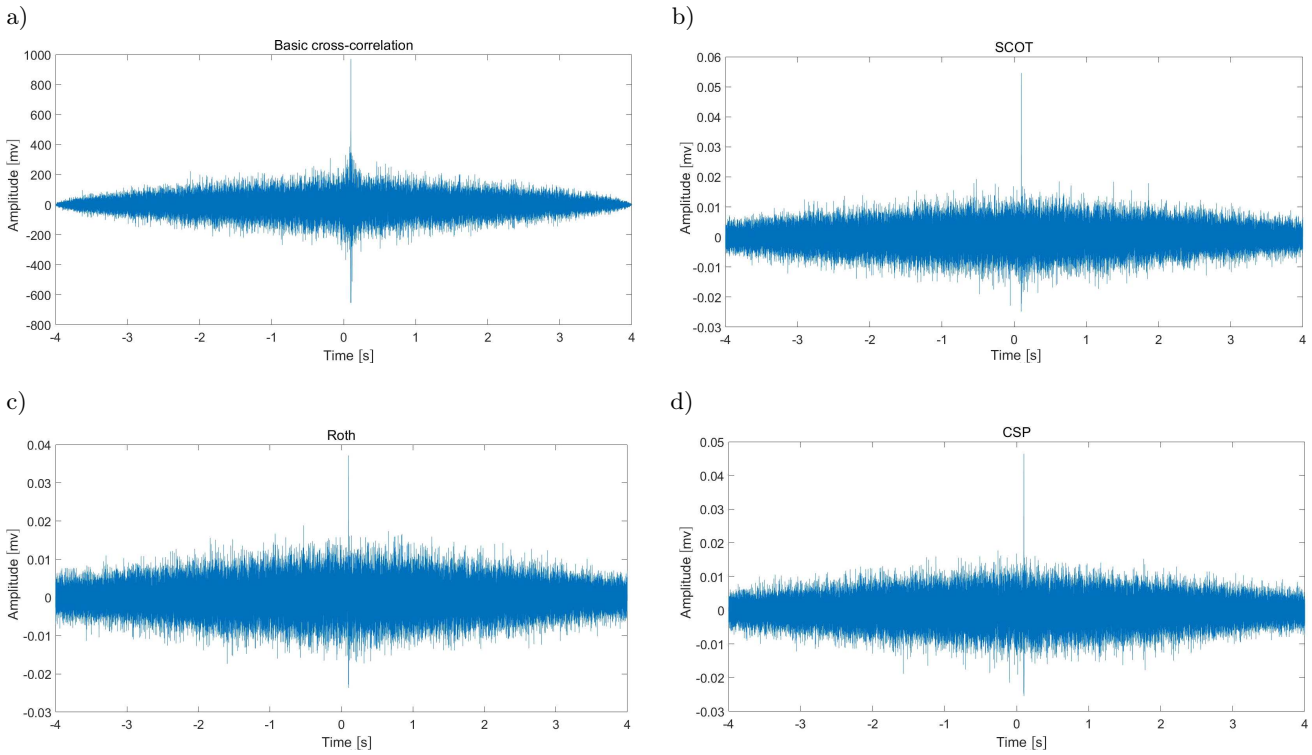

Fig. 2. SNR = 10 dB cross-correlation results.

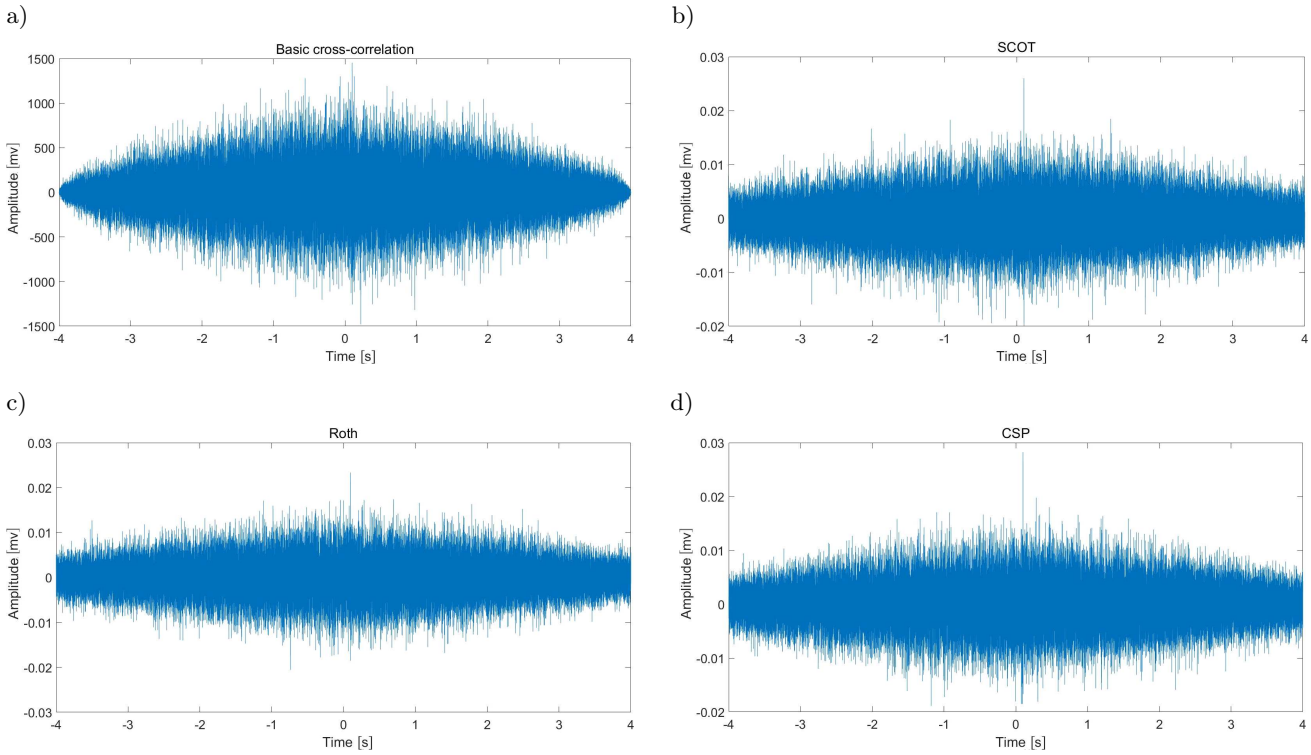

Fig. 3. SNR = 0 dB cross-correlation results.

peak of the Roth algorithm also becomes less obvious as shown in Fig. 3c. Methods SCOT and CSP have relatively little performance degradation as shown in Fig. 3b and Fig. 3d.

The principle of spectrum subtraction is to take the short-time Fourier transform of the initial audio signal and save the amplitude and phase angle obtained. By using the condition that there is only noise in the front part of the signal, the average energy of the noise section can be calculated. By subtracting the average noise energy from the original signal energy, we can get the amplitude after noise reduction by taking the square root of the result, and then we can get the signal after noise reduction by adding the initial phase angle. Its schematic diagram is shown in Fig. 4.

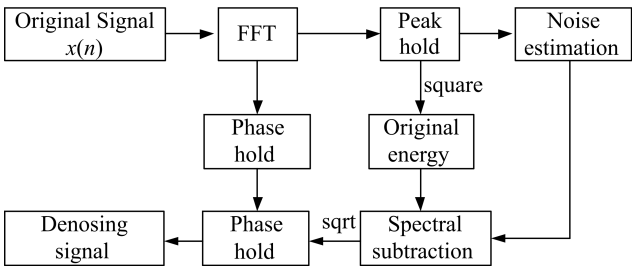

Fig. 4. Spectral subtraction schematic.

There will be obvious "music noise" after denoising with basic spectrum subtraction. An appropriate increase in the over-subtractive factor α can sometimes reduce the "music noise", but too much will distort the waveform. In order to reduce the noise and ensure the waveform is not distorted, this paper mainly uses the following improved spectral subtraction.

$$\left|\widehat{X}_i(k)\right|^{\gamma} = \begin{cases} |X_i(k)|^{\gamma} - \alpha \times D(k) \\ \qquad \text{for} \quad |X_i(k)|^{\gamma} \geq \alpha \times D(k), \\ \beta \times D(k) \\ \qquad \text{for} \quad |X_i(k)|^{\gamma} < \alpha \times D(k). \end{cases} \tag{9}$$

Compared with the traditional spectral subtraction, Eq. (9) in addition to the use of original signal spectrum amplitude $|X_i(k)|^{\lambda}$ minus the ambient noise power spectrum $D(k)$ pure purpose is obtained through the signal, but also avoids the individual point of signal spectrum amplitude is smaller than the noise power spectrum in the $|X_i(k)|^{\gamma} < \alpha \times D(k)$, caused by the residual noise is the "music noise", an improved spectral subtraction is proposed in keeping maximum noise, thereby reducing noise as much as possible. In Eq. (9), α is the over-subtractive factor, and β is the gain compensation factor. $|X_i(k)|^{\lambda}$ is the amplitude of a spectral line; $D(k)$ is the mean value of a certain noise spectrum line:

$$D(k) = \frac{1}{NIS} \sum_{i=1}^{NIS} |X_i(k)|^{\gamma}. \tag{10}$$

When $\gamma = 1$, it is equivalent to using spectral amplitude for spectral subtraction; when $\gamma = 2$, it is equivalent to using power spectrum for spectral subtraction.

Because the noise is random, the spectral line amplitude may be greater than the average value of the noise at a certain time, and the residual noise is called "music noise". In the improved spectral subtraction method, the maximum of the retained noise is proposed to reduce the residual noise as much as possible. In the simulation experiment, the clean sound data with noise was read in, and the white noise of 0 dB was superimposed. The improved spectral subtraction method was used to reduce the noise of the sound signal with noise, and the waveform was obtained as shown in Fig. 5.

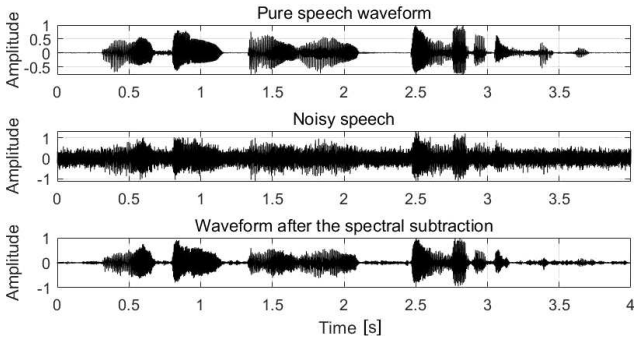

Fig. 5. The signal comparison processed by improved spectral subtraction.

In the Fig. 5, the initial SNR is 0 dB, and the SNR after spectrum subtraction is 8.489 dB, that is, the SNR increases by 8.489 dB. In order to prove the noise reduction ability of spectrum subtraction, simulation experiments were conducted on signals with different SNR, and the data obtained are shown in Table 2.

It can be seen from Table 2 that the improved spectral subtraction with different signal-to-noise ratios has good denoising effects, especially when the signal-to-noise ratio is relatively low. This is because the improved spectral subtraction method keeps the maximum noise in the process of noise reduction, so as to reduce the residual noise as much as possible. The second to fourth columns represent the original signal signal-to-noise ratio, the signal-to-noise ratio after spectrum subtraction and the signal-to-noise ratio improvement, respectively. The relation between them is:

$$\text{ascension SNR} = \text{processed SNR} - \text{original SNR}.$$

From ascension SNR changes it can be seen that the lower the signal-to-noise ratio, the better its noise reduction ability.

In order to further prove the effect of spectral subtraction in generalized cross-correlation, spectral subtraction is added to the previous generalized cross-correlation simulation. For the signal with SNR of 0 dB, spectrum subtraction was performed first, then generalized cross-correlation was performed, and CSP and Scot weighting functions were selected respectively. The results are shown in Fig. 6.

It can be seen from Fig. 6, the peak value of the generalized cross-correlation function weighted by CSP and SCOT is very prominent for the signal processed by spectral subtraction. Compared with the previous signal without noise reduction, the time delay performance is significantly improved. This indicates that the spectral subtraction method can reduce the noise without damaging the time delay information, and has a good effect on improving the peak sharpening of the cross-correlation function.

Table 2. Noise removal under different SNR.

Noise type	Original SN [dB]	Processed SNR [dB]	Ascension SNR [dB]
White Gaussian noise	–5	6.058	11.058
	0	8.489	8.489
	5	11.495	6.495
	10	14.756	4.756

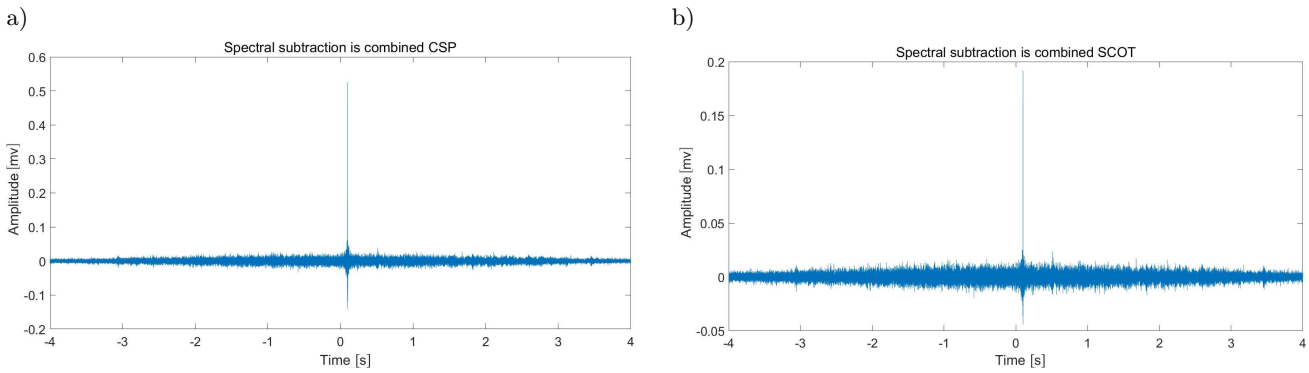

Fig. 6. Spectral subtraction is combined with generalized cross-correlation.

4. Conclusions

In this paper, the traditional time delay estimation technique for sound source location is studied and improved. We find that the combination of spectral subtraction and weighted generalized cross-correlation function can improve the accuracy of delay estimation, solve the problem of large delay estimation error under strong noise environment, and enhance the robustness and accuracy of sound source localization system. The results of computer simulation show that this method can make the peak value of the cross-correlation function of the signal more prominent in the environment of strong noise, which is very helpful to get more accurate time delay estimation, and has great reference value for practical application.

Acknowledgments

This work was supported by the Industrial Research Program, and funded by Shaanxi Science and Technology Department (Program No. 2020GY-158).

References

1. Cheng Y., Yang S.Y. (2015), Sound source localization algorithm and implementation based on arrival time difference, *Journal of Tianjin University of Technology*, **31**(2): 50–54.
2. Danicki E. (2005), Acoustic sniper localization, *Archives of Acoustics*, **30**(2): 233–245.
3. Dong H. (2016), *Study on speech recognition based on spectral subtraction in noise environment*, p. 15, Harbin Engineering University Press, China.
4. Duan L.P., Doa (2014), *Algorithm based on four-microphone array three-dimensional source localization research*, p. 21, Lanzhou Technology University Press, China.
5. Kai S., Xia C.Q., Zhang C.W. (2015), Passive acoustic localization of projectile landing point, Military Automation, *Military Automation*, **34**(6): 1–4.
6. Wang C.Y., Fan G.M., Meng J. (2010), A time delay estimation algorithm for acoustic array based on generalized cross correlation function, *Electroacoustic Technology*, **34**(8): 37–39.
7. Wang X., Wu Z. (2014), Sound source localization based on discrimination of cross-correlation functions, *Applied Acoustics*, **74**(1): 28–37, doi: 10.1016/j.apacoust.2012.06.006.
8. Wei L. (2018), Time delay estimation of sound source localization algorithm research, *Computer Knowledge and Technology*, **14**(7): 220–222.
9. Yan S.Y., Qu X.X., Lou J.Y. (2018), Spectral subtraction speech enhancement algorithm based on continuous noise spectrum estimation [in Chinese], *Communication Technology*, **51**(6): 1296–1301.
10. Zhang C.Y., Mi C.W., Yao P.Y. (2013), Research on sound source localization system based on time delay estimation, [in:] *Advanced Design and Manufacturing Technology III*, Series: *Applied Mechanics and Materials*, Vol. 397–400, pp. 2209–2214, Trans Tech Publications Ltd., doi: 10.4028/www.scientific.net/AMM.397-400.2209.