The application of Dual microphone array and Voice Activity Detection Method Based on Blind Source Separation

Mei Wang[a, *], Jiaxing Ren[b] and Xin Liu[c]

College of Information Science and Engineering, Guilin University of Technology, Guilin 541006, China

[a]glwm@qq.com, [b]5178573@qq.com, [c]liuxin2017125@glut.edu.cn

*Corresponding author

Keywords: Noise, Dual microphone array, Spectral Subtractive, Blind Source Separation, Reduce Noise

Abstract: In real life, there are more and more outdoor positioning fields, which are higher in the quality of multi-source collection, separation of multiple sound sources, positioning of multiple sound sources, accuracy of positioning, volume and power consumption of equipment. Claim. In order to obtain a high-quality sound source in a noisy background environment, this paper proposes a dual-microphone noise reduction system to cope with indoor and outdoor complex noise environments, combining time domain and frequency domain processing, and spectral noise reduction and blind source separation noise reduction. Two methods are used to eliminate noise to improve the accuracy and quality of the sound source.

1. The characteristics of noise

With the continuous development of sound source signal processing technology, sound source identification, positioning and the like are widely used in various electronic devices. At present, many researchers at home and abroad have developed some products commonly used in conferences, noise detection, smart home, artificial intelligence, etc. in the field of sound source localization technology. The sound source system has been applied to various intelligent electronic products. In the process of sound source identification, it is inevitable to be affected by other environmental sounds, such as daily ambient sounds, animal sounds, noise, information sounds, etc., and finally receive noise containing interference, not a pure sound source signal. In order to solve the effects of various ambient sounds, it is usually solved by enhancing and increasing the sound source signal, such as increasing the volume or shortening the distance. Therefore, in order to improve the stability and reliability of the whole system, the paper proposes to improve the quality of the sound source after multiple times of noise reduction on the collected sound.

The characteristic analysis of the acoustic signal can be performed from both the time domain and the frequency domain. From the time domain, it can be divided into transient and steady-state sounds, but most of the noise is transient, because the transient sound duration is very short, and the time is strong and weak. The steady-state sound lasts for a long time, and the signal characteristics remain basically the same type of sound [1]. From the perspective of probability and statistics, the acoustic signal can be divided into two categories: pure tone and polyphony. Pure tone is a deterministic signal, which is a line spectrum sound, and a polyphonic sound is a random signal. For noise, the actual ambient noise can be divided into additive and non-additive [2], and noise is usually used to solve and analyze the problem.

2. Dual microphone noise reduction scheme

2.1 Dual Microphone Array

At present, most of the electronic products mobile phones adopt a dual microphone scheme, which can eliminate noise from the two characteristics of directional beamforming and non-stationary noise cancellation [3], and realize clear calls. That is, the two microphones are on the front or back of the mobile phone, and the distance between the two microphones is about 5-10cm. The above technology is sensitive to the distance between the microphones, and the actual system design and installation requirements are high, and the use is not flexible. The effect is not ideal, the price is more expensive, only suitable for a small range of near sound field. For outdoor or large-scale environmental sound recognition, the use of mobile phone dual microphone technology is not ideal, and the purpose of real noise reduction is not achieved. Therefore, on this basis, the two microphones used for sound collection are omnidirectional microphones for monitoring and collecting sounds. The noise-reduction microphones can be flexibly placed and placed as far as possible from the main microphone, so that the difference between the two microphone pickup signals is as large as possible.

2.2 Hardware Composition

The hardware composition is shown in Figure 1. A chip with two or more microphone input signal channels is used. The signals received by the two microphones are subjected to a dual microphone array noise reduction algorithm by a digital signal processor to output a noise-reduced sound source signal.

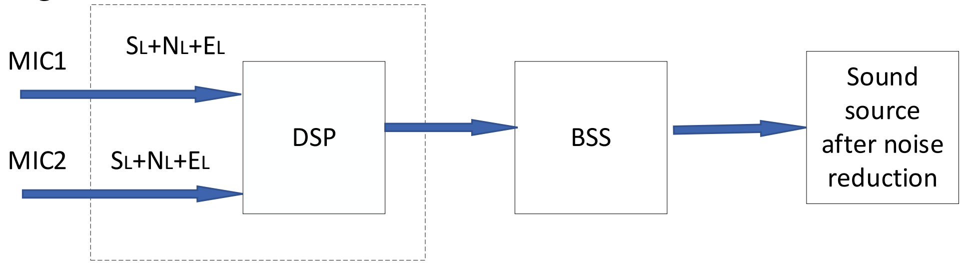


Figure 1. Dual microphone noise reduction hardware.

The signal collected by the main microphone MIC1 is $S_L + N_L + E_L$, where S_L collects a valid signal component for the main microphone, N_L is the noise signal component, and *EM* is the echo component. The signal collected by the noise reduction microphone MIC2 is $S_R + N_R + E_R$, where S_R is the effective signal component collected by the noise reduction microphone, N_R is the noise signal component, and E_R is the echo component. The S_L and S_R have large differences in signal amplitude due to the difference in position from the target sound source. N_L and N_R are essentially the same for far-field noise signals.

2.3 Spectral Subtraction

Because the microphone has a limited range of pickup, when the sound source leaves the microphone array at a certain distance, the signal collected by the microphone is weak, and the acquired signal needs to be amplified. When the signal is amplified, the background noise is also amplified, so the background noise is removed by pre-processing. With noise source signal $x[m]$:

$$\mathrm{x}[m] = a[m] + b[m] \tag{1}$$

In the above formula, $a[m]$ is a pure source signal, $b[m]$ is a noise signal independent of $b[m]$, and the short-term spectral amplitude $|X(\omega)|$ of the mixed observed signal and the estimated value of the noise spectrum. The difference between $B\hat{}(\omega)|$ is the estimate of the short-term spectral amplitude $|A(\omega)|$ of the enhanced source signal, i.e:

$$|\mathrm{A}(\omega)|^2 = \begin{cases} |\mathrm{X}(\omega)| - |\hat{B}(\omega)|^2, \text{if } |\mathrm{X}(\omega)|^2 > |\hat{B}(\omega)|^2 \\ 0, \end{cases} \tag{2}$$

Among them, $|B\hat{}(\omega)|$ can be estimated without a segment, and the correction is repeated during the call. Since the human ear is insensitive to the phase of the sound source signal, the phase of the noise source is directly multiplied by $|B\ \widehat{}(\omega)|$ to recover the noise-removed sound source signal, resulting in:

$$\hat{A}\ [\mathrm{m}] = \mathrm{IFFT}\{|\mathrm{A}(\omega)| \times \exp[\mathrm{j}\arg(\mathrm{X}(\omega))]\} \tag{3}$$

The pre-processing of the spectral cancellation method cannot completely filter out the overlapping noise and signal noise collected by the microphone array, but it can also play a certain role to provide assistance for signal reception. The spectral cancellation method is relatively complex in algorithm, which is not conducive to implementation, and will result in longer sound processing time.

2.4 Blind Source Separation Noise Reduction

In the noise reduction algorithm, the digital signal processor needs to use the blind source separation technology based on the ICA algorithm in addition to the steady-state noise filtering algorithm for the single microphone. Blind Source Separation (BSS), also known as blind source separation, refers to the process of separating the source signals from the mixed signal in the theoretical model where the signal cannot determine the source signal. The basic model of blind source separation is shown in Figure 2.

Blind source signal separation is a powerful signal processing method that is the best estimate of the source signal [4]. In the case where the direction of the sound source signal is known, beamforming can significantly improve the quality of the sound source and improve the signal to noise ratio. If the mixed sound source is separated without knowing the direction of the sound source signal, and the sound source signal of interest is extracted to achieve the purpose of sound source enhancement, this is Independent Component Analysis (ICA). It is a commonly used method for blind source signal separation [5], and is also a commonly used tool for array signal processing. In the last century, scholars have proposed that ICA independent component analysis is a new way to solve problems. The ICA blind source separation is shown in Figure 3.

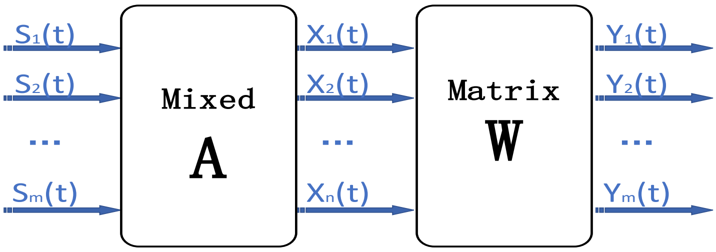


Figure 2. Basic model of blind source separation

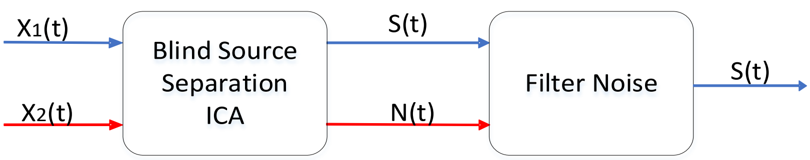


Figure 3. ICA blind source separation diagram

The blind signal separation technique does not require the number of source signals to be the same as the number of observed signals. In Figure 1, it is assumed that the signal sources $s_1(t), s_2(t), \cdots, s_m(t)$ are independent of each other, and $s(t) = [s_1(t), s_2(t), \cdots, s_m(t)]^T$, the mean of $s_1(t), s_2(t), \cdots, s_m(t)$ is 0, after channel A to get n mixed signal independent the source signal $x(t) = [x_1(t), x_2(t), \cdots, x_n(t)]^T$, where $x_1(t), x_2(t), \cdots, x_n(t)$ is *an* observation signals, and $n \geq m$. The above process can be expressed in mathematical terms:

$$\begin{cases} x_1(t) = a_{11}s_1(t) + \cdots + a_{1n}s_n(t) \\ \cdots \\ x_m(t) = a_{m1}s_1(t) + \cdots + a_{mn}s_n(t) \end{cases} \tag{4}$$

Where $A = (a_{\mathrm{ij}})_{m\times n}$ is a mixed matrix, let A be a mixed matrix of $m\times n$, and an instantaneous mixed signal model [6], which:

$$x(t) = A_s(t) \tag{5}$$

Where A is a hybrid matrix. Considering the interference x (t) =A_s (t) +n (t), n (t) in the equation is a noise vector, which is mixed with other signals under the action of the mixing matrix, and n can be eliminated in the expression. Blind signal processing is to solve the source signal s (t) estimated from the observed signal x (t), and find a separation matrix W in the separation, so that

$$\mathrm{y(t)} = w_x(t) = WA_{S(t)} \approx \mathrm{s}(t) \tag{6}$$

In the formula, y (t) represents the separated signal vector (estimated source signal), and the principle of source signal separation is the known observation signal x (t), and the original signal s (t) is recovered by the separation matrix W. The algorithm first performs signal preprocessing and independent component extraction. Preprocessing refers to de-averaging and whitening the signal x (t) in order to simplify the correlation between data processing and the removal of x (t). Go to mean:

$$\widetilde{x_i}(t) = \widetilde{x_i}(t) - \frac{1}{m}\sum_{i=1}^{m} x_i(t), i = 1,2,\cdots,m \tag{7}$$

Whitening process:

$$\mathrm{z}(t) = \mathrm{V}\tilde{x}(t) \tag{8}$$

Z (t) is the whitened signal and V is the whitening matrix.

One of the separated signals is the target sound source, and the other is the superimposed noise that may be multiple far-field noise sources. The dual microphone array is a signal that cannot separate more than two sources, but due to the noise of the far field, the waveform formed on the two microphones is a superposition of multiple noise sources, and their signals on the two microphones are basically the same. It can be considered that these noise sources are already a single source noise after superposition, so a noise signal is separated. For non-stationary noise signal characteristics, like the sound source, the two signals separated from each other should be observed in combination with the characteristics of the dual microphone array. For far-field signal noise, the energy in the two microphone signals is substantially equal, the attenuation is relatively fixed, and the sound source signal has a large energy difference, thus, the sound source signal conforming to this feature is retained.

Since the signal energy P is inversely proportional to the square of the distance, it is obtained:

$$\mathrm{P} = \propto \frac{1}{d^2} \tag{9}$$

The energy attenuation in the two microphones is:

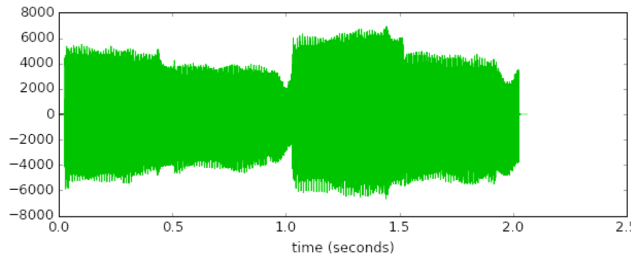
$$\mathrm{A} = 10\mathrm{lg}\frac{P_1}{P_2} = 10\mathrm{lg}\frac{d_2^2}{d_1^2} \tag{10}$$

2.5 Simulation experiment

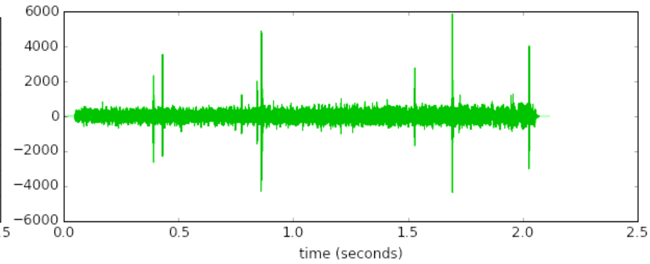
The experiment uses a signal with a completely similar noise and sound source signal. Consider using a dual microphone array to filter the unsteady noise. The experiment uses two sound signals to make an ambulance alarm sound signal as the sound source s_1 (t), another under The sound of rain as noise s_2 (t), after mixing, becomes a mixed sound signal, and then separates the signal y_1 (t), y_2 (t) from the mixed sound by the blind source separation algorithm, and separates the sound source signal from the original signal s_1 (t) Relatively close, noise, s_2 (t) is quite different from the separated noise. See Figure 4 below. After separation, it is **necessary** to determine which signal is noise, which signal is the sound source, and the noise is removed by noise reduction to obtain an effective high quality sound source.

3. Conclusion

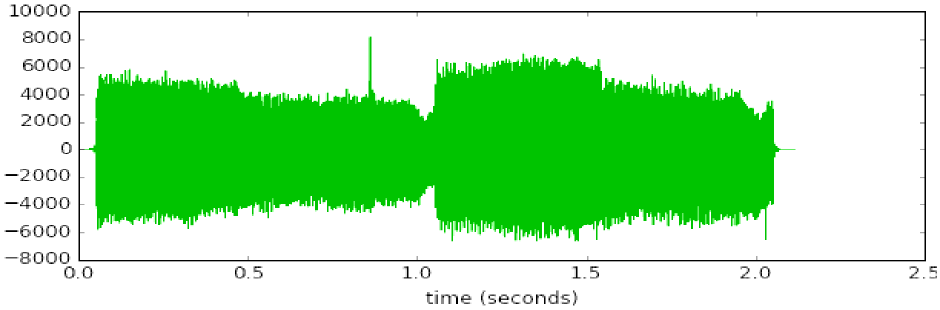
Combining the characteristics of noise reduction and noise characteristics of mobile phones, the noise reduction method of two microphones for sound collection outside the room is proposed. The dual-microphone noise reduction scheme uses an ordinary independent monitoring pickup. Since only one microphone channel is added, the algorithm is simple and effective, but when a plurality of signals of the same loudness and the same tone are mixed at the same time, the effect of using this separation noise is improved. Poor, further research is needed later.



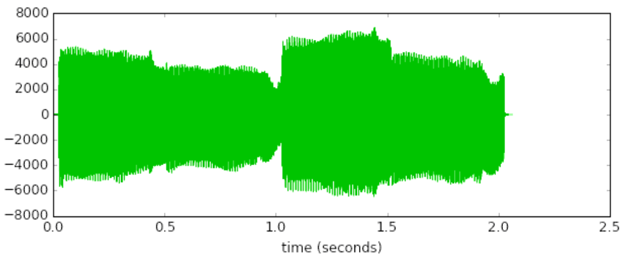
Ambulance alarm signal



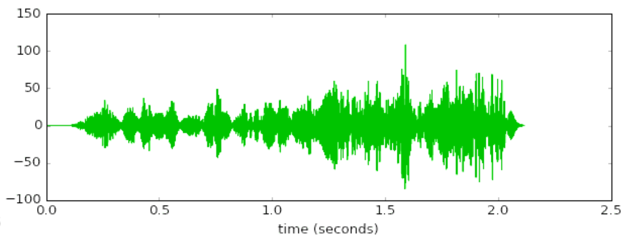
Raining sound



Mixed acoustic signal



Isolated ambulance siren signal



Separated rain

Figure 4. Original Sound waveform.

Acknowledgments

This research work was supported by the National Natural Science Foundation of China (No.61961010), the Guangxi Key Research and Development (R&D) Program (Guike AB17292058), the Guangxi key Laboratory Fund of Embedded Technology and Intelligent System under Grant No.

2018B-1, the Key Laboratory Found of Cognitive Radio and Information Processing, Ministry of Education (Guilin University of Electronic Technology) under Grant No. CRKL180203.

References

[1] Chen Kean. Auditory perception and automatic recognition of environmental sound. [D]. Science Press, 2014.

[2] Dong Hu. Research on endpoint detection method of sound source signal under strong noise environment [Master Thesis]. Hunan: Hunan Normal University.

[3] He Zhijian, Zheng Huming. Microphone array principle and application, electroacoustic technology November 2009 issue: 20 ~ 30.

[4] Ou Xudong. Research on Convolutional Blind Source Separation Algorithm for Speech Signals. [D]. Chongqing: Chongqing University of Posts and Telecommunications, 2016.

[5] Yu Xianchuan, Hu Dan. Theory and application of blind source separation [M]. Beijing: Science Press, 2011.

[6] Wang Kang. Research on blind source separation method. [D]. Huainan: Anhui University of Science and Technology, 2018.