

Blind source separation of indoor mobile voice sources

Chunli Wang*, Quanyu Wang, Yuping Cao

College of Electronics and Information Engineering, Lanzhou Jiaotong University, Lanzhou
730070, China

Email: wcl@mail.lzjtu.cn

ABSTRACT

The mobile voice sources move freely in the indoor range of several to tens of square meters at a speed of lower than 10m/s. The reflection signal and the subsequent original sound signal are superposed in each space to generate reverberation and cause serious interference to the original sound signal. Through the comparison of three classical algorithms of blind source separation, the online algorithm can constantly update the separation system in real time according to the different positions of voice source signals, but it has no advantages in such performances as operation speed and convergence speed. The batch algorithm is fast but delayed, while the blind source separation algorithm based on independent component analysis of the frequency domain has less computation and fast convergence. The improved separation matrix algorithm is used to verify the effectiveness of the algorithm.

Keywords: Mobile Voice Sources, Reverberation, Blind Source Separation, Natural Gradient, Independent Component Analysis.

1. INTRODUCTION

If the indoor voice source signals continuously moves, the reflection signal generated in the early stage and the subsequent original sound signal are superposed in each space to generate reverberation and cause interference to the subsequent original sound signals. The reverberation is a form of convolution noise as well as a kind of echo, which is mainly caused by the reflection of objects such as indoor walls, ground and ceiling. Since the absorption attenuation of these hard reflecting surfaces is smaller, the reflected signal is similar in waveform and amplitude to the original sound signal but with different phases [1].

2. REVERBERATION

The existing blind source separation algorithm is not very effective for indoor mobile voice source separation, especially when the indoor impact response is greater, the separation effect is even more limited. The concept of reverberation time [2] was proposed by acoustician W. C. Sbani in 1900: after the speaker stops speaking, the residual acoustic energy is reflected repeatedly through hard reflecting surfaces and attenuated by absorption of other objects, and the acoustic energy density is reduced to one millionth of the original value, the time for which is shown in Equation (1).

$$T_R \approx \frac{KV}{A} = \frac{KV}{\alpha S} \quad (1)$$

Where, T_R is the reverberation time which is independent of the indoor measurement position [3], K is a constant, which is related to the indoor air temperature, V is the indoor volume, α is the sound absorption coefficient of the indoor related surface material, and S is the sum of the indoor surface area. The longer the reverberation time is, the greater the impact response is. Thus, the reverberation determines the length of the selected FIR filter.

In the blind source separation algorithm of frequency domain, the window length T of the short-time Fourier windowing is much longer than the length of the impact response, $P: T \gg P$ [4-5]. The frequency domain BSS is selected to study the relationship between them, and the learning rule is shown in Equation (2).

$$W_{i+1} = W_i + \eta \left\{ \text{diag} \left(\langle \varphi(Y) Y^H \rangle \right) - \langle \varphi(Y) Y^H \rangle \right\} W_i \quad (2)$$

Where, η is learning step, $\langle \cdot \rangle$ is mean operator and $\varphi(\cdot)$ is the nonlinear function in the frequency domain, as shown in Equation (3).

$$\varphi(Y) = \frac{1}{1 + \exp(-Y^{(R)})} + j \frac{1}{1 + \exp(-Y^{(I)})} \quad (3)$$

where, $Y^{(R)}$ and $Y^{(I)}$ are the real part and the imaginary part of the output signal in frequency domain respectively.

Under different reverberation time, the algorithm separation performance corresponding to different window lengths can be evaluated by noise reduction rate (NRR). Figure 1 is a graph comparing the separation performance at different reverberation time, with the definition of NRR shown in Equation (4).

$$\begin{aligned} NRR_i &= SNR_{O_i} - SNR_{I_i} \\ SNR_{O_i} &= 10 \log \frac{\sum_{\omega} |A_{i_i}(\omega) S_i(\omega)|^2}{\sum_{\omega} |A_{i_j}(\omega) S_j(\omega)|^2} \\ SNR_{I_i} &= 10 \log \frac{\sum_{\omega} |H_{i_i}(\omega) S_i(\omega)|^2}{\sum_{\omega} |H_{i_j}(\omega) S_j(\omega)|^2} \end{aligned} \quad (4)$$

where, SNR_{O_i} and SNR_{I_i} are the output and input signal to noise ratio respectively. For two source signals, one is used as a target signal and the other is used as a noise signal, that's, $i \neq j$. The greater the NRR value is, the better the separation effect and the better the performance of the algorithm are [6].

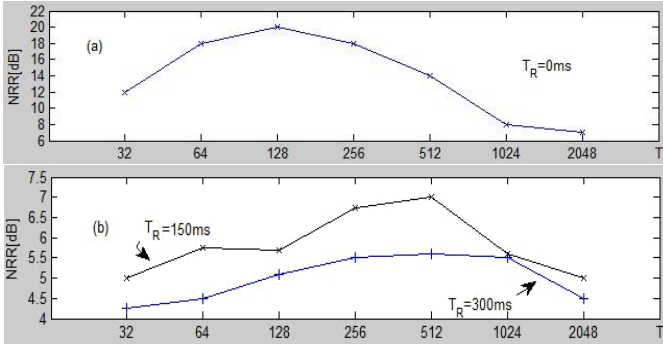


Figure 1. NRR value of different window lengths under different reverberation time

As shown in Figure 1, without reverberation, when $T=128$, NRR can reach the maximum value of 20.8 dB, and the reverberation time are 150 ms and 300 ms; when $T=512$, NRR can reach the maximum values of 7.1 dB and 5.8 dB; meanwhile, the separation effect of short window length ($T=128, 512$) is better than that of the long window length since when the reverberation time is longer, the original signals on each frequency point are no longer independent of each other and do not conform to the independence assumption in the blind source separation algorithm, and thus the above-mentioned limitation condition of the frequency domain BSS is meaningless.

3 BLIND SOURCE SEPARATION ALGORITHM FOR MOBILE SPEECH SIGNALS BASED ON FREQUENCY DOMAIN ICA

3.1 Frequency domain ICA

For mobile signals, its hybrid system is time-varying. In this paper, the frequency domain ICA method based on mutual information minimization is used to enhance mobile speech.

Set N source signals: $s(t)=[s_1(t), s_2(t), \dots, s_N(t)]^T$, get the observed signals through M microphones: $x(t)=[x_1(t), x_2(t), \dots, x_M(t)]^T$, separation signal: $y(t)=[y_1(t), y_2(t), \dots, y_N(t)]^T$, BSS can be represented by Equations (5) and (6).

$$x_j(t) = \sum_{k=1}^N (h_{jk} * s_k)(t) \quad (5)$$

$$y_i(t) = \sum_{j=1}^M (w_{ij} * x_j)(t) \quad (6)$$

where, h_{jk} is the frequency response of the i source signal to j microphone, w_{ij} is separation filter and $*$ represents the operation of convolution. The short-time Fourier transform is performed on Equation (5), and the model is approximated to Equation (7).

$$x(f, n) = H(f) s(f, n) \quad (7)$$

where, f is frequency, n is frame number, $x(f, n)=[x_1(f, n), x_2(f, n), \dots, x_M(f, n)]^T$ is the Fourier transform of the n frame at the f frequency of the received signal, and $H(f)$ is the mixing matrix at the f frequency.

The separation process of each frequency domain is as shown in Equation (8).

$$y(f, n) = W(f) x(f, n) \quad (8)$$

where, $y(f, n)=[y_1(f, n), y_2(f, n), \dots, y_N(f, n)]^T$ is the estimated signal at the f frequency of the n frame, and $W(f)$ is separation matrix. By adopting the idea of mutual information minimization, the optimal separation matrix can be obtained by the natural gradient iterative method as shown in Equation (9).

$$W_{i+1}(f) = W_i(f) + \mu [I - \langle \varphi(y(f, n)) y^H(f, n) \rangle] W_i(f) \quad (9)$$

where, i is number of iterations, μ is step parameter, $\langle \cdot \rangle$ is mean operator, I is unit matrix, and $\varphi(\cdot)$ is nonlinear function, with the definitions shown in Equation (10).

$$\begin{aligned} \varphi(y(f, n)) &= [\varphi(y_1(f, n)), \varphi(y_2(f, n)), \dots, \varphi(y_N(f, n))] \\ \varphi(y_1(f, n)) &= \tanh(\text{Re}(y_1(f, n))) + i \cdot \tanh(\text{Im}(y_1(f, n))) \end{aligned} \quad (10)$$

where, $\text{Re}(\cdot)$ represents the real part and $\text{Im}(\cdot)$ represents the imaginary part.

3.2 On-line algorithm

On-line algorithm, also called adaptive algorithm, can update the separation system according to the different positions of source signals, that's, real-time processing of mobile speech with adaptive algorithm [7]. It is an algorithm for processing single observation sample. Its advantage is that the separation matrix can be adjusted and updated according

to the time-varying characteristic of hybrid system. With good real-time processing performance, it's generally applicable to the field with relatively high requirements in the real-time processing performance.

There are many on-line algorithms. In this paper, the natural gradient descent method is adopted to minimize the objective function.

$$\min. J(y(f, n), W(f)) = KL(p_Y(y(f, n))) \parallel (q_Y(y(f, n))) \\ = -H(y(f, n)) - \sum_{i=1}^n E \left\{ \log(q_Y(y_i(f, n))) \right\} \quad (11)$$

Where, $q_Y(y(f, n))$ is marginal probability of $y_i(f, n)$, and $p_Y(y(f, n))$ is joint probability. From $y(f, n) = W(f)x(f, n)$, Equation (12) is obtained by the definition of entropy.

$$H(y(f, n)) = H(x(f, n)) + \log|\det W(f)| \quad (12)$$

Equation (13) can be obtained by substituting Equation (12) into Equation (11) and simultaneously calculating the gradients $W(f)$ of both sides of the equation.

$$\frac{\partial J(y(f, n), W(f))}{\partial W(f)} \\ = \frac{\partial}{\partial W(f)} \left(-H(x(f, n)) - \log|\det(W(f))| - \sum_{i=1}^n E \left\{ \log(q_Y(y_i(f, n))) \right\} \right) \\ = -\frac{\partial \log|\det(W(f))|}{\partial W(f)} - \frac{\partial \sum_{i=1}^n E \left\{ \log(q_Y(y_i(f, n))) \right\}}{\partial W(f)} \\ = -W(f)^{-T} + E[\phi(y(f, n))x(f, n)^T] \quad (13)$$

Where, $W(f)^{-T} = (W(f, n)^{-1})^T$ and $\phi(y(f, n))$ are activation functions, and

$$\phi(y(f, n)) = \left[-\frac{q'_Y(y_1(f, n))}{q_Y(y_1(f, n))}, \dots, -\frac{q'_Y(y_n(f, n))}{q_Y(y_n(f, n))} \right]^T \quad (14)$$

The online adjustment formula thus obtained is shown in Equation (5.15).

$$\Delta W(f, n) = -\eta \frac{\partial J(y(f, n), W(f))}{\partial W} - \eta (W(f)^{-T}) - \\ E[\phi(y(f, n))x(f, n)^T] \quad (15)$$

where, η is step parameter.

This algorithm has such disadvantages as slow operation speed and poor performance, and is not suitable for the separation system in which both the source signals and the interfering signals are moving, so the batch algorithm is introduced to improve its performance.

3.3 Batch algorithm

The batch algorithm corresponds to the on-line algorithm, and mainly processes with some acquired data and hypotheses, instead of by the recursive method.

To track time-varying hybrid systems, the separation matrix should be updated for each block of $B_m = \{t: (m-1)T_b \leq t < mT_b\}$, where, T_b represents the size of block and m represents block exponent with $m \geq 1$.

Koutras et al. have proposed similar methods in the time domain [8]. When T_b is equal to the frame length of the short-time Fourier transform, the method can be considered as an on-line algorithm in the frequency domain. The last block separation matrix is used as the initial iteration value of a new block, i.e., $W_{m+1}^{(0)}(f) = W_m^{(N_I)}(f)$, where N_I is the number of iterations of Equation (15).

Batch algorithm has inherent delay, because the calculation of W needs to wait for the arrival of next data block, and the required calculation time is longer, but if the corresponding calculation is completed within T_b , the separation matrix can be calculated with W_{m-2} to avoid waiting and calculation delay, thus it is suitable for delay insensitive situation.

4. OPTIMIZATION OF ICA BLIND SOURCE SEPARATION ALGORITHM

4.1 Uncertainty of arrangement and amplitude

In the classical ICA blind source separation algorithm, there are generally two kinds of uncertainties: arrangement uncertainty and amplitude uncertainty [9]. Assuming that there are two source signals and two receivers, the separation matrix at f of each frequency is obtained using Equation (8) as shown in Equation (16).

$$\begin{bmatrix} y_1(f) \\ y_2(f) \end{bmatrix} = \begin{bmatrix} \lambda_1 s_1(f) \\ \lambda_2 s_2(f) \end{bmatrix} \text{ or } \begin{bmatrix} y_1(f) \\ y_2(f) \end{bmatrix} = \begin{bmatrix} \lambda_2 s_2(f) \\ \lambda_1 s_1(f) \end{bmatrix} \quad (16)$$

where, λ_1 and λ_2 are arbitrary plurals. It can be seen that the arrangement order and amplitude scale are uncertain. If the estimated signal is not dearranged and descaled, the signal after the short-time Fourier transform is still a mixed signal, and thus the blind source separation algorithm cannot be applied to it.

(1) Arrangement uncertainty optimization

In the two uncertainties, the arrangement sequence uncertainty has greater effect, and determines the success or failure of the blind deconvolution algorithm, because it may cause the signals separated at adjacent frequencies to be misplaced when they are connected, and may connect two different signals together, resulting in erroneous separation results.

Using the method of smoothing the separation matrix in the frequency domain, the length of the filter is restricted in the time domain, or the average separation matrix is used by the adjacent frequency. The arrangement matrix $P(f)$ can be obtained, so the separation matrix can be modified with Equation (17).

$$W(f) \leftarrow P(f)W(f) \quad (17)$$

(2) Amplitude uncertainty optimization

The split spectrum method is used to divide each array signal into two different performance indicators [10].

Matsuoka proposes a minimum distortion principle [11] and the separation matrix is modified by Equation (18).

$$W(f) \leftarrow \text{diag}[W(f)^{-1}]W(f) \quad (18)$$

After solving the amplitude problem by this method, the separated signal is obtained as shown in Equation (19).

$$y_n(f, t) = W_n(f) S_n(f, t), n = 1, \dots, N \quad (19)$$

Equation (19) is a part of the separated signals obtained by the split spectrum method, which selects the main path signal with a relatively good performance, and also reduces the calculation amount.

5. SIMULATION EXPERIMENT AND ANALYSIS OF RESULTS

In order to verify the effectiveness of the improved algorithm as mentioned above, Matlab software is used for functional simulation. A general laboratory with 7 computers working and no professional sound insulation equipment is selected as the indoor environment for inputting speech. Two speakers who walk around at will are selected, and two receivers respectively receive speech signals. According to the measurement and calculation, the reverberation time is 150 ms, the sampling frequency of data is 8000Hz, the sampling time is 10s, and the number of samples is 10,000. The waveforms of the acquired original and aliased speech signals are shown in Figures 2 (a) and 2 (b).

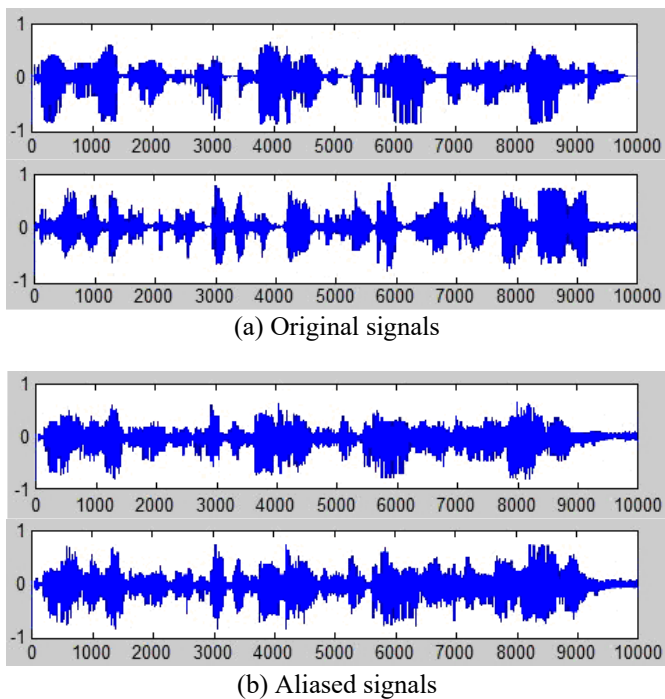


Figure 2. Waveforms of the original and aliased speech signals

After the short-time Fourier transform of the aliasing signals in Figure 2, the blind source separation is performed on each frequency point, and the modified separation matrix can be quickly calculated by using the improved frequency domain ICA through the simulation software as follows:

$$W = \begin{bmatrix} 8.2134 & -2.5978 \\ 3.4552 & 5.7001 \end{bmatrix}$$

The waveform of the blind source separation signal realized by using the separation matrix is shown in Figure 3.

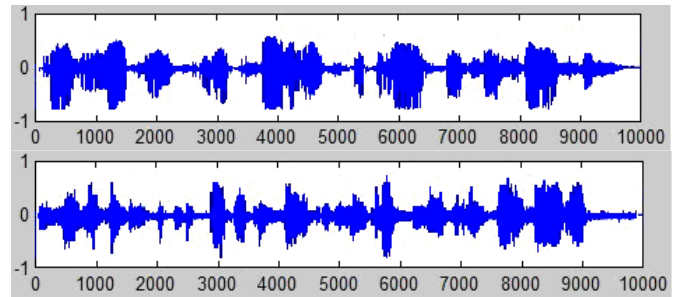


Figure 3. Waveform of separated signals

Comparing with Figures 2 (a), 2 (b) and 3, the optimized blind source separation algorithm is used to suppress the reverberation of the mobile voice source, and the waveform of the speech signal separated from the aliased signal is close to that of the original speech signal, which verifies that the algorithm can achieve higher separation degree through less computation and achieves the effect of speech enhancement.

6. CONCLUSIONS

For the indoor mobile speech signal enhancement, the relationship between reverberation and separation filter in practical environment is analyzed, which lays a theoretical foundation for realizing blind source separation algorithm, and then three classical blind source separation algorithms are compared and analyzed. The on-line algorithm can update the separation system in real time according to the positions of the source signals. However, it doesn't have advantage in performances such as operation speed and convergence speed. Batch algorithm has the inherent delay, and longer computation time, so it is suitable for the situation where the delay is insensitive. The ICA algorithm has a small amount of computation and is suitable for processing convolution mixed speech signals similar to the actual noise situation, but it has uncertainties of inherent arrangement and amplitude. By modifying the separation matrix transfer function, the software simulation results show that the effectiveness of ICA algorithm is improved to a certain extent.

ACKNOWLEDGMENTS

Supported by the Youth Foundation of Lanzhou Jiaotong University Project NO.2014003; Supported by the National Nature Foundation Project of China NO.61461024; Supported by the Graduate educational reform project of Lanzhou Jiaotong University (Training and practice in the ability of innovation and experiment based on the course of the modern electronic technology for graduate students).

REFERENCES.

- [1] Liao T. (2012). Researches of the blind source separation algorithm based on an independent component analysis, *Master dissertation of Hunan*

- Normal University*, pp. 15-21, DOI: [10.7666/d.y2149207](https://doi.org/10.7666/d.y2149207)
- [2] Zhou G.X., Yang Z.Y., Xie S.L. (2011). Mixing matrix estimation from sparse mixtures with unknown number of sources, *IEEE Transactions on Neural Networks*, Vol. 22, No. 2, pp. 211-221. DOI: [10.1109/TNN.2010.2091427](https://doi.org/10.1109/TNN.2010.2091427)
- [3] Cao Y.P. (2012). Speech enhancement algorithm based on the signal subspace, *Electronic Test*, Vol. 6, pp. 54-57.
- [4] Zhou J. (2016). Research of underdetermined source estimation and blind extraction method for mechanical fault signals, *Doctoral dissertation of Kunming University*, pp. 38-45.
- [5] Yang P. (2008). Blind source separation study on the aero engine aliasing vibration signal, *Master dissertation of Nanjing University of Aeronautics and Astronautics*, pp. 13-15. DOI: [10.7666/d.d053213](https://doi.org/10.7666/d.d053213)
- [6] Guo W., Yu F.Q. (2015). Improved speech music signal separation based on negative entropy maximization, *Computer Engineering and Application*, Vol. 51, No. 4, pp. 209-212,
- [7] Lu J.T., Cheng W. (2015). Adaptive blind source separation algorithm in a changed or equivariant step, *Journal of Xi'an Jiaotong University*, Vol. 49, No. 12, pp. 83-89. DOI: [10.7652/xjtub201512014](https://doi.org/10.7652/xjtub201512014)
- [8] Zhang Y.Y., Xin J.H., Liu G.B. (2016). Applications of combined with cumulant slice joint diagonalization of blind source separation, *Journal of Huazhong University of Science and Technology (Natural Science)*, Vol. 44, No. 7, pp. 86-90. DOI: [10.13245/j.hust.160717](https://doi.org/10.13245/j.hust.160717)
- [9] Dermoune A., Wei T. (2013). Fast ICA algorithm: five criteria for the optimal choice of the nonlinearity function, *IEEE Transactions on Signal Processing*, Vol. 61, No. 8, pp. 2078-2087. DOI: [10.1109/TSP.2013.2243440](https://doi.org/10.1109/TSP.2013.2243440)
- [10] Chen G.Q. (2015). Blind source separation algorithm based on kurtosis cumulant proportional differential control vector, *Journal of Electronics*, Vol. 12, No. 5, pp. 929-934. DOI: [10.3969/j.issn.0372-2112.2015.05.015](https://doi.org/10.3969/j.issn.0372-2112.2015.05.015)
- [11] Yang J.M., Qi H.Y. (2015). Improved nonlinear blind source separation algorithm based on the minimization of mutual information, *Electric Measurement and Instrument*, Vol. 52, No. 9, pp. 66-69. DOI: [10.3969/j.issn.1001-1390.2015.09.013](https://doi.org/10.3969/j.issn.1001-1390.2015.09.013)