

Article

# An Efficient Convolutional Blind Source Separation Algorithm for Speech Signals under Chaotic Masking

Shiyu Guo , Mengna Shi, Yanqi Zhou, Jiayin Yu and Erfu Wang \* 

Electrical Engineering College, Heilongjiang University, Harbin 150080, China; guoshiyubaby@163.com (S.G.); 2191310@s.hlju.edu.cn (M.S.); 2191294@s.hlju.edu.cn (Y.Z.); 2181234@s.hlju.edu.cn (J.Y.)

\* Correspondence: wangerfu@hlju.edu.cn; Tel.: +86-1383-6068-896

**Abstract:** As the main method of information transmission, it is particularly important to ensure the security of speech communication. Considering the more complex multipath channel transmission situation in the wireless communication of speech signals and separating or extracting the source signal from the convolutional signal are crucial steps in obtaining source information. In this paper, chaotic masking technology is used to guarantee the transmission safety of speech signals, and a fast fixed-point independent vector analysis algorithm is used to solve the problem of convolutional blind source separation. First, the chaotic masking is performed before the speech signal is sent, and the convolutional mixing process of multiple signals is simulated by impulse response filter. Then, the observed signal is transformed to the frequency domain by short-time Fourier transform, and instantaneous blind source separation is performed using a fast fixed-point independent vector analysis algorithm. The algorithm can preserve the high-order statistical correlation between frequencies to solve the permutation ambiguity problem in independent component analysis. Simulation experiments show that this algorithm can efficiently complete the blind extraction of convolutional signals, and the quality of recovered speech signals is better. It provides a solution for the secure transmission and effective separation of speech signals in multipath transmission channels.

**Keywords:** independent vector analysis; convolutional blind source separation; multipath transmission; chaotic masking



**Citation:** Guo, S.; Shi, M.; Zhou, Y.; Yu, J.; Wang, E. An Efficient Convolutional Blind Source Separation Algorithm for Speech Signals under Chaotic Masking. *Algorithms* **2021**, *14*, 165. <https://doi.org/10.3390/a14060165>

Academic Editor: Constantin Paleologu

Received: 6 April 2021  
Accepted: 24 May 2021  
Published: 26 May 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

With the rapid development of information technology, speech transmission technology brings more and more convenience to people's life. As the main form of human communication, speech information is widely used for its real-time and easy identification [1–3]. Whether it is indoor communication or outdoor transmission, the process of sending and receiving speech information will inevitably be affected by the effect of multipath transmission [4]. After the speech is sent out, it needs to reach the receiving end through multiple transmission paths, and the length and attenuation of each path are different, so the received signal presents the superposition of various speech signals affected by phase, delay and multipath [5]. In addition, an increasing number of fields and departments need safe and efficient speech transmission technology, and information security [6] is facing unprecedented development opportunities. Therefore, in the multipath channel, studying the safe transmission and effective extraction of speech information in indoor and outdoor environments is one of the scientific issues that many scholars are interested in.

In recent years, the chaotic phenomena of nonlinear systems have been studied more and more intensively [7–9]. Chaos is a nonperiodic bounded dynamic behavior caused by a deterministic nonlinear dynamic system. It has initial sensitivity, internal randomness and unpredictability [10], which make it play an important role in secure communication. Considering the secure transmission of speech information in wireless

transmission environment, the speech signal can be chaotic encrypted or masked at the originating end. For indoor environments, a classic problem of speech communication is the “cocktail party”. In a noisy indoor environment, the sound received by the information-collecting microphone is diverse. The sound received includes the speech, music and other sound sources of multiple people talking at the same time, and there is also the reflected sound generated by the reflection of these sounds by the walls and indoor objects. In fact, whether in indoor or outdoor environments, speech transmission under multipath channels can be abstracted as a convolutional mixed model of signals. At the same time, there is bound to be a lack of source information and channel information during the transmission process. Therefore, separating the desired speech signal from the unknown information is known as classic multichannel convolutional blind source separation (CBSS) [11–13].

There are two main types of methods to solve the CBSS problem in multipath channels: time domain and frequency domain. The idea behind time-domain approach [14,15] is the need to compute a separation filter in the time domain of equal or greater order than the impulse response filter. The method is often computationally complex, making the complexity of the algorithm high and convergence slow. The frequency-domain approach [16,17] is to transform the observed signal into the frequency domain using the short-time Fourier transform (STFT) [18]. The instantaneous BSS is performed separately for each frequency point in the frequency domain using the standard independent component analysis (ICA) method [19,20], which is the key to blind source separation in the frequency domain. Although this method avoids tedious convolution operations and reduces computational effort, the ICA algorithm causes amplitude and permutation ambiguity to be blurred at each frequency point, which reduces the accuracy of the separated signals. The amplitude ambiguity can be handled by normalizing the separation matrix, while the permutation ambiguity is relatively difficult to solve. For this reason, numerous scholars have proposed different methods to solve the permutation ambiguity problem in frequency-domain convolutional blind source separation (FD-CBSS) [13,21–23].

The independent vector analysis (IVA) [24] algorithm has been proposed as an extension of the ICA algorithm, which is mainly used to solve the permutation ambiguity problem in FD-CBSS. Considering the algorithm itself, the traditional independent univariate source prior is replaced by the source prior knowledge of multivariate correlations, using higher-order correlations across frequencies [25] to prioritize the associated multivariate super-Gaussian distribution over each vector source. This modeling imposes the independence of sources between vectors while preserving the correlation of sources within higher-order vectors, i.e., the structural correlation between the frequency components of each source. The dependent multi-source prior preserves the dependence between the different frequency units of each source and maximizes the independence between the frequency units of the different sources [26]. Thus, the IVA algorithm alleviates the permutation ambiguity problem in the learning process and does not require subsequent permutation processing. To improve the separation performance and faster convergence of this algorithm, the fast fixed-point IVA (Fast IVA) method [27] of the Newtonian learning algorithm is applied to better solve the case of CBSS. This has been studied in more depth by subsequent scholars. J. Hao proposed an online IVA algorithm for real-time audio separation and developed a two-channel hardware demonstration system [28]. M. Anderson applied the IVA model to solve the joint blind source separation problem for fMRI data and in the process derived gradient and Newton-based update rules [29]. Y. Liang used video information to provide an intelligent initialization for the optimization problem and proposed a fast video-based fixed-point analysis method that still has good separation performance in noisy environments [30].

To this end, this paper designed a speech information security transmission under chaotic masking and a FD-CBSS algorithm based on Fast IVA. The scheme is designed for the security of speech signals in wireless transmission environments. First, the speech signals to be transmitted were processed by chaotic masking and then sent to the channel for multipath transmission. Then, the encrypted convolutional speech signal in the time

domain was transformed into the frequency domain using the short-time Fourier transform (STFT), followed by blind signal extraction using the Fast IVA algorithm. The blind deconvolution process of the signal was completed when the separation matrix converged or when the number of iterations was reached. The algorithm is able to perform permutation simultaneously during the iterative update of the separation matrix and converges quickly.

This paper is organized as follows: in the second part, Chen chaotic systems for confidential transmission are firstly briefly introduced; and then the FD-CBSS under multipath transmission model is highlighted. In the third part, the theoretical basis of the Fast IVA algorithm is first introduced, and the specific implementation steps of the algorithm are listed. Furthermore, a safe and fast overall process of blind deconvolution based on Fast IVA algorithm is illustrated. Finally, the evaluation criteria used to judge the separation performance of the algorithm are presented. The fourth part is to simulate the algorithm experiment and analyze the algorithm performance. In the fifth part, the conclusions and future research prospects are given based on the analysis results.

## 2. Mathematical Modeling and Theoretical Foundations

This part briefly introduces the Chen chaotic system used for chaotic masking of speech signals in this paper and present the theoretical preparation for constructing secure transmission of speech signals. In addition, the convolutional model for multipath transmission of speech signals is constructed, and the BSS process under this model is illustrated to make the theoretical basis for the independent vector analysis algorithm.

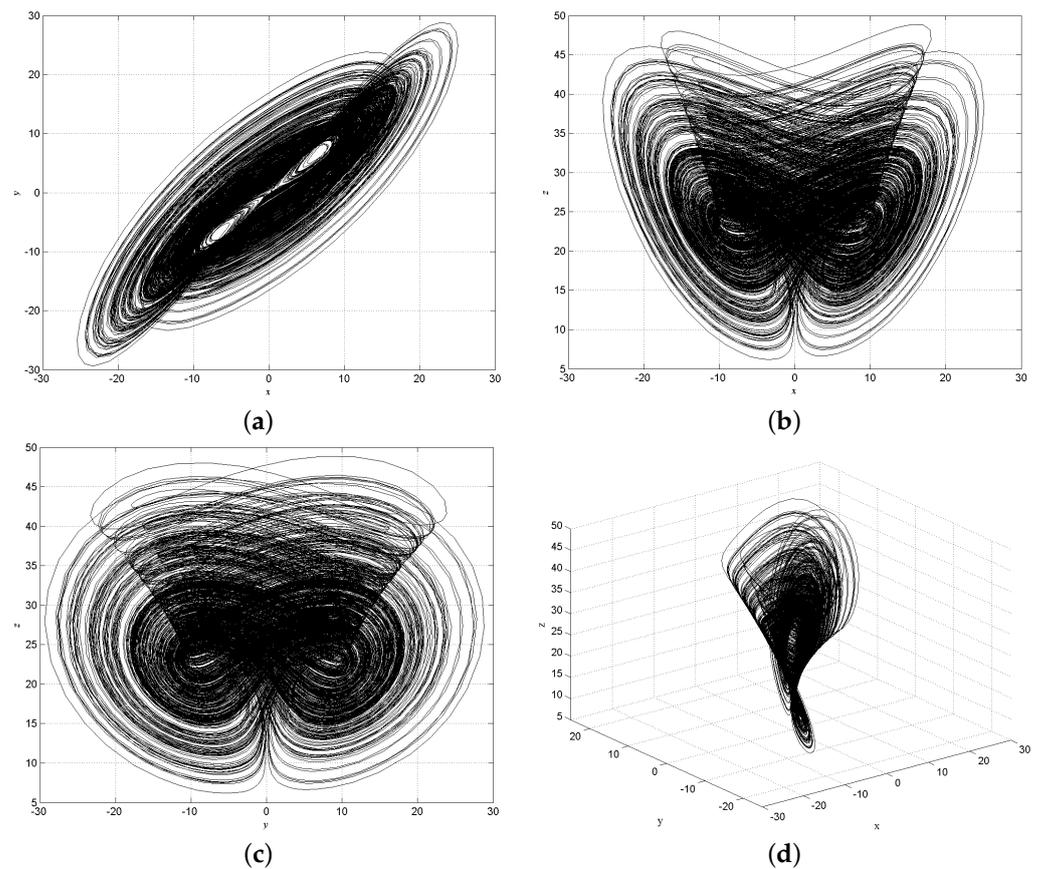
### 2.1. Chen Chaotic System

In the field of chaos research, the Chen system [31,32] is another type of classical chaos among continuous chaotic systems, which provides a theoretical basis for many researchers. Based on the the phase diagram of the system, it is similar to the Lorenz system, but its topological structure is more complex and has more research significance. The mathematical model of Chen's attractor is

$$\begin{cases} \frac{d^{\alpha} x_1}{dt^{\alpha}} = a(y - x) \\ \frac{d^{\alpha} x_2}{dt^{\alpha}} = (c - a)x - zx + cy \\ \frac{d^{\alpha} x_3}{dt^{\alpha}} = xy - bz \end{cases}, \quad (1)$$

where  $X = (x, y, z)^T \in R^3$  is the system state and  $a, b, c$  are the system variables and all are greater than 0. The Chen system enters a chaotic state when the parameters take values of  $a = 35, b = 3, c = 28$  and the time interval is at  $t_0 \in [0, 100]$ . Due to the initial value sensitivity of the chaotic system, the initial value  $x_0 = 1, y_0 = 1, z_0 = 1$  is selected here in order to observe the phase diagram and time sequence diagram of the Chen attractor at a certain state, as shown in Figure 1.

Chen chaotic systems have inherently better key space and system complexity. Chaotic masking refers to using the pseudo-randomness of chaotic tracks and broadband power spectrum characteristics as a modulating signal. At the transmitter side, the speech signals are hidden into the chaotic signal to form a seemingly noisy signals, and the useful information is completely masked by the chaotic signal to achieve confidential communication.



**Figure 1.** The three-dimensional phase diagram of Chen's chaotic motion state. (a) The x–y plane phase diagram; (b) The y–z plane phase diagram; (c) The x–z plane phase diagram; (d) The Spatial phase diagram of Chen chaos.

## 2.2. Frequency Domain Convolutional Blind Source Separation

The actual communication environment is more complex, and the speech information is reflected by obstacles to form multipath transmission, which propagate through multiple paths to reach the receiving end, thus obtaining the unknown observed signals. It should be noted that there are both direct and reflected waves such as obstacles in the observed signals. Due to the different multipath lengths, the observed signals at the receiver are the superposition of source signals affected by phase, time delay and multipath. Mathematical modeling of the above multipath transmission environment is performed. In the case where the source signal and channel are unknown, only a small amount of a priori knowledge of the observed signals is used to achieve an estimate of the observation matrix or source signals. Without considering the noise, the mathematical model is shown in Figure 2, and this model is referred to as convolutional blind source separation [12].

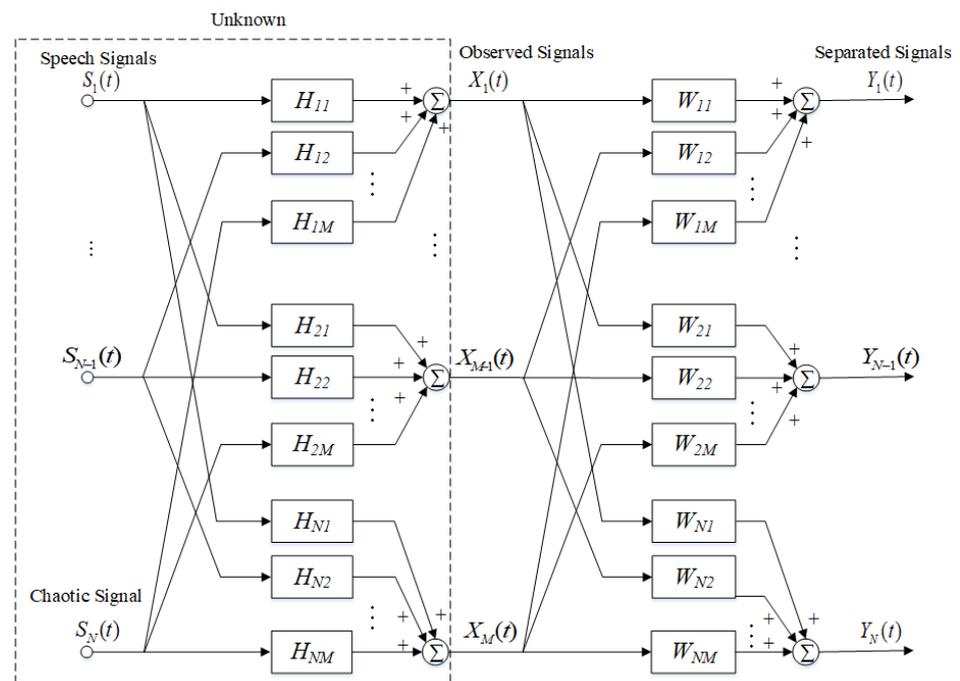


Figure 2. CBSS model for multiple source signals.

Suppose there are \$N\$ source signals  $s(t) = [s_1(t), s_2(t), \dots, s_N(t)]^T$  and \$M\$ observation signals  $x(t) = [x_1(t), x_2(t), \dots, x_M(t)]^T$  ( $N > M$ ) and all are discrete signals. The mathematical model for convolutional mixing of multiple speech signals is:

$$X_i(t) = \sum_{j=1}^M \sum_{p=0}^{P-1} h_{ij}(p) \cdot S_j(t - p) = h_{ij} * S_j(t) \quad i = 1, 2, \dots, M, \tag{2}$$

where  $h_{ij}(p)$  denotes the impulse response of the \$j\$th source signal to the \$i\$th microphone;  $S_j(t)$  denotes the \$j\$th source signal;  $X_i(t)$  denotes the \$i\$th observed signal;  $*$  denotes the convolution operation, and  $p$  denotes the order of the transfer function. That is, each observed signal is a convolutional mixture of the individual source signals and the corresponding impulse responses to them. The larger the filter order  $p$ , the more complex the received observation signal. In particular, when  $p = 1$ , the model degenerates into a linear instantaneous mixed model. To simplify, Equation (1) is expressed in vector form as:

$$\mathbf{X}(t) = \sum_{p=0}^{P-1} \mathbf{H}(p)\mathbf{S}(t - p). \tag{3}$$

where  $\mathbf{H}(p)$  is the vector representation of the transfer function. In practice, the transfer function  $h_{ij}(p)$  can be approximated by a finite impulse response filter.

Transform the observed signals in the time domain to the time-frequency domain through STFT:

$$\mathbf{X}^{(k)}[n] = \sum_{k=0}^{K-1} X(nJ + t)win(t)e^{-j\omega_k t} \quad k = 1, 2, \dots, K, \tag{4}$$

where  $\omega_k = 2\pi(k - 1)/K$   $K$  is the number of frequency points and  $win(t)$  is the window function—the Hamming window is usually selected because it has better low-pass characteristics so that it can better approximate the short-term speech frequency spectrum. Furthermore,  $x_i^{(k)}[n]$  denotes the sample of the \$n\$th random variable of  $x_i^{(k)}$ . It should be noted that, unlike real-time  $t$ ,  $n$  represents a random variable in the discrete time domain, where each value of  $n$  corresponds to the STFT of each frame. For convenience, the time

symbols  $t$  and  $n$  are mostly omitted. When the length of the window function  $win(t)$  is much larger than the order  $p$  of the impulse response filter, the convolutional model can be approximated as an instantaneous model for each frequency band [28,33]:

$$\mathbf{X}^{(k)} = \mathbf{H}^{(k)} \mathbf{S}^{(k)}, \quad (5)$$

where  $\mathbf{X}^{(k)} = [x_1^{(k)}, x_2^{(k)}, \dots, x_M^{(k)}]^T$  and  $\mathbf{S}^{(k)} = [s_1^{(k)}, s_2^{(k)}, \dots, s_N^{(k)}]^T$  are the observation matrix and the source signal matrix in the frequency domain, respectively. Furthermore,  $x_i$  represents the  $i$ th observation vector composed of  $K$  frequency points  $[x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(K)}]^T$ , and  $x^{(k)}$  denotes the observation vector composed of  $M$  observation signals at the  $k$ th frequency point  $[x_1^{(k)}, x_2^{(k)}, \dots, x_M^{(k)}]^T$ . Furthermore,  $H^{(k)} \equiv \{h_{ij}^{(k)}\}$ , this implies that  $h_{ij}^{(k)} = [h_{ij}^{(1)}, h_{ij}^{(2)}, \dots, h_{ij}^{(K)}]$  is denoted as the element of the  $i$ th row and  $j$ th column of the  $k$ th mixing matrix.

Note that we use lowercase bold letters for vector variables and uppercase letters for matrix variables, and each value of the superscript  $k$  is a frequency point and  $K$  is the number of frequency points.

If the separation filter matrix exists, i.e., the inverse or pseudo-inverse matrix of the mixing matrix at each frequency point exists, the separated source signal is:

$$\mathbf{Y}^{(k)} = \mathbf{W}^{(k)} \mathbf{X}^{(k)}, \quad (6)$$

where  $\mathbf{W}^{(k)}$  is the separation matrix in the frequency domain. The purpose of BSS is to find the corresponding separation matrix  $\mathbf{W}^{(k)}$  from the observed signal  $\mathbf{X}^{(k)}$  to achieve the extraction of the source signal. The frequency-domain separated signal is transformed to the time-domain by ISTFT, which is the estimation of the source signal.

### 3. An Efficient Speech CBSS Algorithm in Multipath Channels

In the traditional frequency-domain convolutional blind source separation (FD-CBSS) algorithm, there will inevitably be problems of permutation ambiguity. There are usually two methods to solve such problems: one is a geometric method based on the direction of arrival (DOA) that has strict requirements on the spatial position of the sensor, and another method is based on the correlation of adjacent frequency bands with mutual parameters. These two methods have their own advantages and disadvantages in solving the ambiguity problem of separating sub-signal permutation. However, both methods require an additional permutation step after the CBSS algorithm, which inevitably increases the time complexity and computational complexity of the FD-CBSS process.

For this reason, this paper considers the existing CBSS itself and proposes the Fast IVA algorithm using Newtonian learning method to solve the multipath speech CBSS problem. The algorithm is essentially an extension of the one-dimensional random variables in the ICA algorithm to multi-dimensional random variables. It looks at the sampled data of the same source signal at different frequency points as a vector and performs CBSS of the signals in order, thus solving the permutation ambiguity case without designing additional permutation algorithms. The innovation of this paper is that the optimization algorithm process introduces Taylor series polynomials to quickly approximate the contrast function of Fast IVA algorithm. This step efficiently accomplished the CBSS without the need to design additional sequencing algorithms.

#### 3.1. Fast IVA Algorithm

The Fast IVA algorithm uses the negative entropy criterion to measure the strength of the non-Gaussian and minimizes the contrast function over each frequency band to complete the estimation of the source signals. The advantage of this algorithm is that it performs CBSS in order at the ordered frequency points, thus solving the permutation ambiguity problem. On this basis, this paper uses Taylor series in the algorithm optimiza-

tion process to approximate the contrast function of Fast IVA and then performs fast and ordered CBSS of speech signals at each frequency point.

Here, the Kullback–Leibler divergence [34] between the two functions is defined as a measure of high-order dependence, which is the cost function of multivariate random variables. The relative entropy [24] is expressed as follows:

$$C = KL\left(p(s_1, s_2, \dots, s_N) \parallel \prod_i q(s_i)\right) = const. + \sum_k \log|\det(H^{(k)})| - \sum_i E_{s_i}(\log q(s_i))' \tag{7}$$

where  $p(s_1, s_2, \dots, s_N)$  represents the exact joint probability density function of a single source vector;  $\prod_i q(s_i)$  represents the product of approximate marginal probability distribution functions;  $k$  is the number of frequency points;  $\det(\cdot)$  represents the matrix determinant operator;  $E(\cdot)$  denotes the expectation, and  $KL(\cdot)$  denotes the Kullback–Leibler scatter calculation.

For simplicity, suppose that  $x_i$  is zero-mean and the separated signal  $y_i$  is pre-whitened so that the rows of the mixing matrix  $H^{(k)}$  or the separation matrix  $W^{(k)}$  are orthogonal in each dimension, i.e.,  $A^{(k)\top} A^{(k)} = I$  or  $W^{(k)} W^{(k)\top} = I$ . Therefore, Equation (7) can be simplified as:

$$C = const. + \sum_i E_{s_i}(\log q(s_i)). \tag{8}$$

In this paper, the fast fixed-points method is chosen to optimize the contrast function. Compared with the natural gradient method, the Fast IVA algorithm avoids the selection of learning rate and has the advantage of fast convergence. During the optimization of the algorithm, we introduce a Taylor series approximation to the contrast function of Lagrangian residue type in the notation of the complex variables in order to be able to quickly obtain the optimal approximate solution. Therefore, the contrast function of the Fast IVA algorithm with Lagrange multipliers  $\beta$  is obtained as follows:

$$C = - \sum_i E_{s_i}(\log q(s_i)) - \sum_k \beta \left( W^{(k)\top} W^{(k)} - I \right). \tag{9}$$

Since the Hessian matrix of the contrast function [35] is a diagonal matrix under the whiteness constraint, the following simple learning rule can be obtained by applying Newton’s method:

$$w_i^{(k)} \leftarrow w_i^{(k)} - \frac{E\left[\varphi^{(k)}\left(y_i^{(1)}, y_i^{(2)}, \dots, y_i^{(K)}\right)x^{(k)}\right] + \beta w_i^{(k)}}{E\left[\varphi^{(k)'}\left(y_i^{(1)}, y_i^{(2)}, \dots, y_i^{(K)}\right)\right] + \beta}. \tag{10}$$

It can be known that the equilibrium point of Equation (10) above is the local minima of the contrast function, since the demixing matrix is no longer updated. When the equation holds, the equilibrium point is found. The Lagrange multiplier is eliminated to obtain the following fixed point iteration algorithm:

$$\begin{cases} w_i^{(k)} \leftarrow E\left[\varphi^{(k)'}\left(y_i^{(1)}, y_i^{(2)}, \dots, y_i^{(K)}\right)\right] w_i^{(k)} - E\left[\varphi^{(k)}\left(y_i^{(1)}, y_i^{(2)}, \dots, y_i^{(K)}\right)x^{(k)}\right] \\ \varphi^{(k)'}\left(y_i^{(1)}, y_i^{(2)}, \dots, y_i^{(K)}\right) = -\frac{\partial\varphi^{(k)}\left(y_i^{(1)}, y_i^{(2)}, \dots, y_i^{(K)}\right)}{\partial y_i^{(k)}} \end{cases}. \tag{11}$$

In addition to normalization, the rows of demixing matrix  $w_i^{(k)}$  need to be decorrelated. The symmetric decorrelation is calculated as:

$$w_i^{(k)} \leftarrow \left( w_i^{(k)} \left( w_i^{(k)} \right)^H \right)^{-1/2} w_i^{(k)}. \tag{12}$$

In the algorithm, in order to avoid local optimal solutions, we use the unit matrix as the initial demixing matrix in all frequency points. It can be seen from the derivation process of the algorithm that demixing matrix has no orthogonality constraints and does not produce separation error accumulation, which ingeniously solves the permutation ambiguity in the frequency-domain algorithm. We introduce a Taylor series to optimize the contrast function in the optimization algorithm, which makes it possible to obtain the optimal solution quickly in the iterative update of the separation matrix, simplify the iterative update process and reduce the computational complexity of the algorithm. The specific implementation steps of the core algorithm in this paper are as follows in Algorithm 1.

---

**Algorithm 1:** CBSS based on Fast IVA algorithm

---

- Step 1: Transform the time-domain convolutional mixed signal by STFT into a complex-valued signal at each frequency point in the complex frequency domain;  
 Step 2: Centralize and pre-process the mixed signal at each frequency point;  
 Step 3: Initialize the demixing matrix  $w_0^{(k)}$  for each frequency point;  
 Step 4: Estimate the separated signal  $y_i^{(k)}$  according to Equation (6);  
 Step 5: Calculate and optimize the contrast function and nonlinear function based on the estimated source signal;  
 Step 6: Update the demixing matrix  $w_i^{(k)}$  according to Equation (11);  
 Step 7: Decorrelation of the demixing matrix according to Equation (12);  
 Step 8: Normalize the separation matrix to resolve the amplitude uncertainty of the separated sub-signals;  
 Step 9: Determine whether the separation matrix converges, and if it converges, execute step 11;  
 Step 10: If the maximum number of iterations is reached, output the final separated signal at each frequency point, otherwise return to step 4;  
 Step 11: Restore the separated signal of each frequency point into a time-domain separated signal by ISTFT; that is, extract the source speech signal.
- 

### 3.2. BSS of Speech Signals with Chaotic Masking in Multipath Channels Based on Fast IVA Algorithm

Considering the information security transmission of speech signals under multipath channels in the wireless transmission environment, this paper processes the confidentiality of speech signal before sending. Due to the characteristics of initial sensitivity and internal randomness of the chaotic system itself, the chaotic sequence of its output is unpredictable, which provides a new design idea for secure communication. In this paper, the message signal and the chaotic signal are superimposed on each other, and the useful signal is covered by the pseudo-random and noise-like characteristics of the chaotic system to realize the confidential transmission of the message signal. Since the speech signal is a small signal, the magnitude of its energy and amplitude is far from that of a chaotic signal. Even though they are both multi-frequency signals, the broadband power spectrum characteristics of the chaotic signal still ensure a good masking effect on the speech signal.

In this paper, chaotic signals are introduced into the source signal for chaotic masking to achieve the secure transmission of speech signals. In addition, this paper also proposes using a Fast IVA algorithm to solve the CBSS in the case of multipath speech secure transmission. Taking three signals (two speeches and one chaotic signal) as an example, we simulate the multipath transmission situation of speech signals and the whole process of blind observation signal separation to extract the source speech signal in the actual environment, as shown in Figure 3.

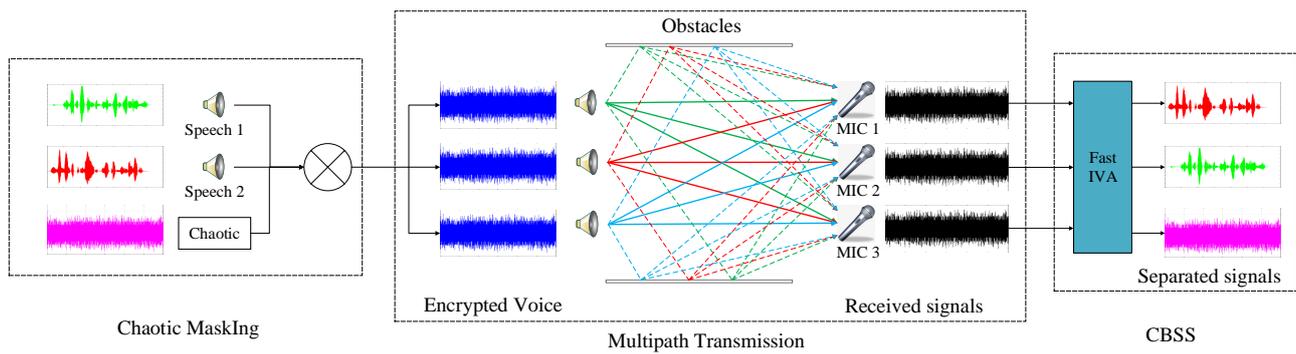


Figure 3. CBSS model for secure transmission of multiple speech signals.

From the above model, it can be seen that this paper introduces chaotic masking technology to hide the multiplexed speech signals into the random signals generated by the Chen chaotic dynamical system to ensure the transmission security of speech information. Secondly, the masked speech signals are transmitted via multipath channels to reach the receiving end. The multipath effect made the observed signals inevitably affected by phase and time delay and became a superposition of multiple source signals. The Fast IVA algorithm is applied to solve the FD-CBSS of the unknown observed signals, which efficiently achieved blind separation or extraction of speech signals without an additional sorting process. This greatly reduced the computational complexity of the algorithm and improved the efficiency of the FD-CBSS.

### 3.3. Evaluation Criteria

During the simulation experiment of CBSS, the signal estimation still differs somewhat from the source signal even in the absence of noise. In this paper, evaluation criteria such as signal distortion ratio (SDR) [36,37], signal interference ratio (SIR) [36,37] and correlation coefficient [23] were used to quantitatively analyze the separation performance of the algorithm.

#### 3.3.1. SDR and SIR

The estimated signal can be represented by four parts, namely, the real source part  $s_{ture}(t)$ , the filtering distortion part  $e_{filt}(t)$ , the interference part of other sources  $e_{interf}(t)$  and the false value  $e_{artif}(t)$ , where  $s_{target}(t) = s_{ture}(t) + e_{filt}(t)$  indicates the part of estimated signal that belongs to the source signal (indicating that the sensor acquires information about the target source containing transmission effects);  $e_{interf}(t)$  indicates that the estimated signal does not satisfy the source signal and belongs to the mixed signal, which is the residual after the separation of other sources;  $e_{artif}(t)$  indicates the external noise generated by the algorithm. The specific definition equations for SDR and SIR are

$$SDR = 10\log_{10} \frac{\|s_{target}(t)\|^2}{\|e_{filt}(t) + e_{interf}(t) + e_{artif}(t)\|^2}, \tag{13}$$

$$SIR = 10\log_{10} \frac{\|s_{target}(t)\|^2}{\|e_{interf}(t)\|^2}. \tag{14}$$

The SDR reveals the ratio of the true source to the other components. The larger the indicator value, the less the separated signal is affected by distortion, interference and artifacts. Furthermore, the larger the SIR value, the less the components of the separated signal are separated from other sources, and the better the performance of the separated signal.

### 3.3.2. Correlation Coefficient

The correlation coefficient is usually the degree of similarity between the separated signal and the source signal if  $s(t)$  and  $y(t)$  are used to represent the source signal and the estimated signal, respectively. The mathematical expression is as follows:

$$\rho(y_i, s_i) = \frac{\left| \sum_t y_i(t) s_i(t) \right|}{\sqrt{\sum_t y_i^2(t) \sum_t s_i^2(t)}}. \quad (15)$$

The similarity coefficient between the source signal and the estimated signal is between 0 and 1. When two signals are perfectly correlated, the similarity between the source and estimated signals is high. The closer the similarity coefficient is to 1, the better the separation performance of the algorithm. Conversely, the smaller the similarity coefficient, the worse the separation performance of the algorithm.

## 4. Simulation Experiment and Result Analysis

One male and one female speech signal from the TIMIT database randomly selected as the speech message to be sent, and the signal length is 3 s. The x-component of Chen chaotic system is chosen as the input chaotic signal, and the signal length of the intercepted x-component is the same as the length of speech signals, and the sampling frequency is  $f_s = 16$  KHz. First, the speech signals are hidden into the chaotic signal to ensure the security of information transmission. Next, the impulse response filter is used to simulate the multipath effect generated by the speech transmission.

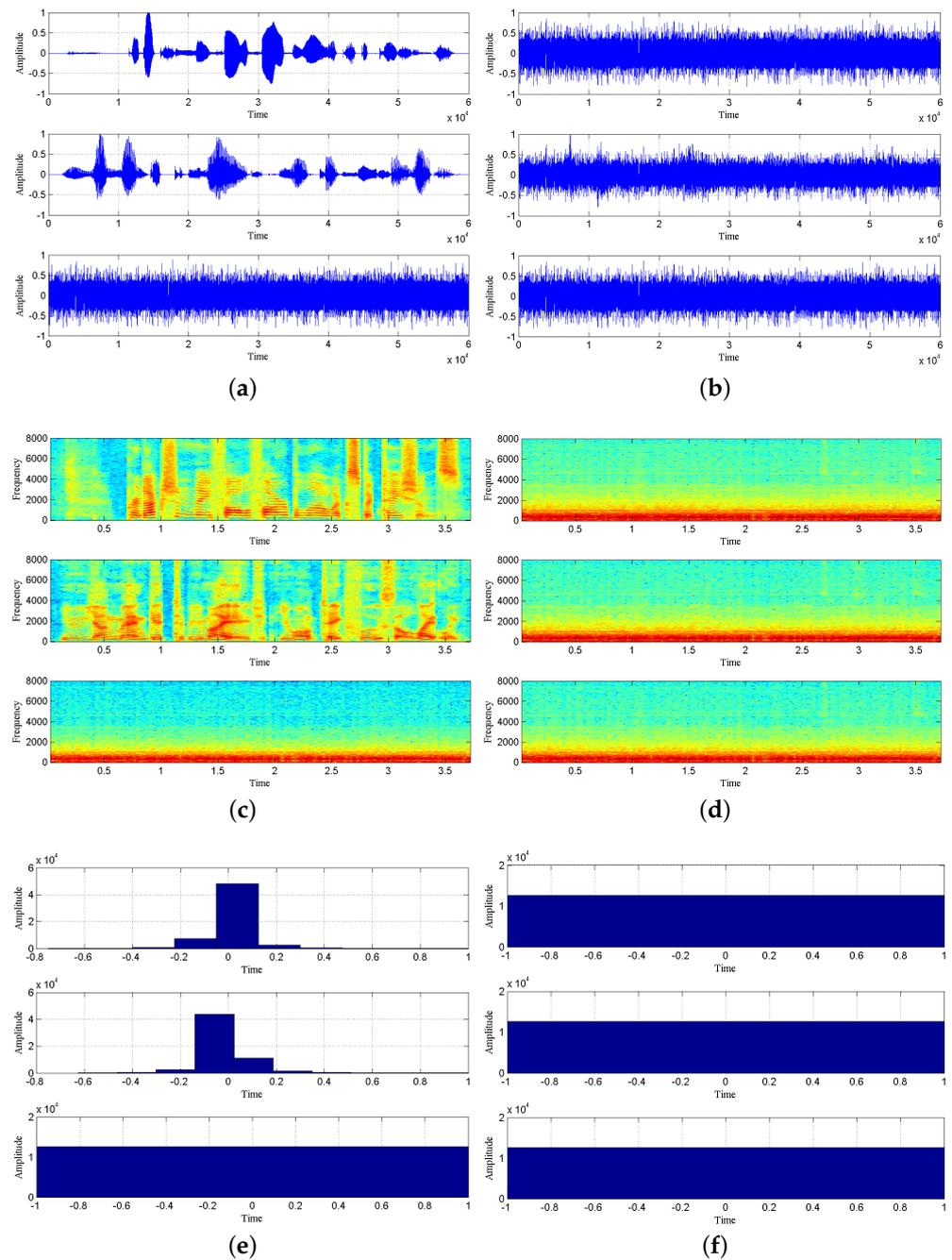
It is important to note here that the filter length directly affects the degree of convolutional mixing of the source signal. The longer the filter, the longer the process of mixing response of the speech signal in the channel, and consequently the more complex the received observation signal and the more difficult the separation or extraction of the source speech signal. In order to simulate the actual transmission environment and for the signals safety, the filter length should be increased. Furthermore, taking into account the difficulty of the separation process, the filter length should be reduced to ensure that the Fast IVA algorithm can obtain better separation results. Considering the above, the filter length set in the experiment of this paper is  $p = 20$ .

Then, the Fast IVA algorithm is used to separate and extract the speech signals from the observed signal, and the number of algorithm iterations is set to 1000 times. To ensure that the signal has good spectral characteristics in the frequency domain, the Hamming window function is chosen for the STFT transform, as in Section 2.2. Furthermore, the appropriate number of frequency points is  $K = 512$ , and the data frame is  $H = 192$ .

Based on the above data and parameter selection, the simulation experiment is carried out and the selection of frequency points is considered in the experiment. The following is a quantitative analysis of the security of the speech signal, the separation results and separation performance of the Fast IVA algorithm, the evaluation of the speech quality, and the complexity of the algorithm from multiple perspectives.

### 4.1. Security Analysis

Chen chaotic system has good obfuscation and diffusion properties, and it is better able to resist statistical attacks. Figure 4 shows the waveforms, spectrograms and histograms of the source and observation signals obtained by simulation, and the security of transmission system can be seen intuitively.

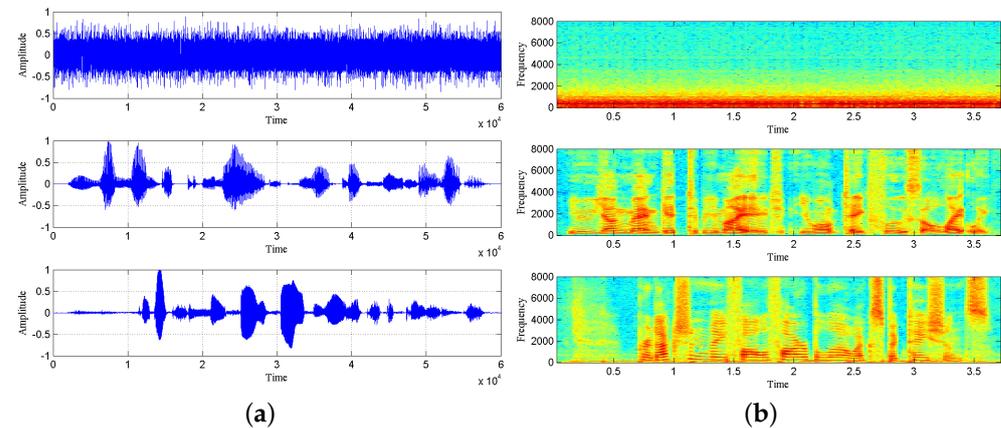


**Figure 4.** The graphical information of the source and observed signals. (a) Time-domain waveform of the source signals; (b) Time-domain waveform of the observed signal. (c) Spectrogram of the source signal; (d) Spectrogram of the observed signal; (e) Histogram of the source signal; (f) Histogram of the observed signal.

The image information of the source signal and the observed signal after chaotic masking is shown in Figure 4. The time-domain waveform, spectrograms and histogram of the observed signal are given in Figure 4b–f, respectively, to compare the graphical information of the observed signal with that of the source signal. It is obvious that the source speech information is no longer found in the observed signal, which is enough to show that the chaotic signal plays a good masking effect on the speech signal and achieves the confidential transmission of speech information.

#### 4.2. Correlation Coefficient Analysis

The selection of the source signal, experimental conditions and parameter settings are the same as described previously. Then, the blind observation signal after chaotic masking are processed at the receiver to achieve the extraction of source signal. In this paper, the Fast IVA algorithm is proposed to use it to perform FD-CBSS for the blind observation signal, and the waveforms of the separation result are shown in Figure 5.



**Figure 5.** The separated signal graphic information is obtained by using the Fast IVA algorithm. (a) Time-domain waveform of separated signals; (b) Spectrogram of separated signals.

In Figure 5, it can be seen that the waveform shapes of the separated signals obtained using the Fast IVA algorithm are almost identical when compared with the source signal graphical information, both in time domain waveform and spectrogram. Therefore, from a subjective point of view, it can be considered that the algorithm realizes the blind extraction of the source speech signal. However, it is not enough to rely on visual judgment only. Here we use the correlation coefficient as the objective evaluation criterion of the separated signal, and the data in Table 1 are the average of 20 experiments.

**Table 1.** Correlation coefficient between each separated signal and the source signal.

Correlation Coefficient	s1	s2	Chaotic Signal
$y_1$	0.0018	0.0016	0.9971
$y_2$	0.0027	0.9632	0.0120
$y_3$	0.9278	0.0137	0.0163

In Table 1, s1 and s2 represent the source speech signals, while  $y_1$ ,  $y_2$  and  $y_3$  represent the separated signals obtained by the Fast IVA algorithm. As you can see, the correlation coefficient between the separated signal  $y_1$  and the chaotic signal is as high as 0.9991, which is close to 1. In addition, the correlation coefficients between the separated signals  $y_2$  and  $y_3$  and the corresponding source speech signals s2 and s1 are 0.9932 and 0.9278, respectively, which proves that the algorithm achieves the recovery of speech signals. The algorithm not only achieved the estimation of source speech signal, but also enabled a better recovery of the wide-spectrum chaotic signal. Thus, a high degree of reduction in the separated signal is illustrated from an objective point of view, and the blind extraction effect of the noise-like signal is guaranteed.

This paper proposes using the Fast IVA algorithm, which exploits both the statistical independence between multi-source signals and the internal dependence of each signal. The algorithm is not only fast and effective but also solves the permutation ambiguity problem in the FD-CBSS process and eliminates the permutation design in the standard ICA frequency-domain algorithm. To verify the superiority of this algorithm for convolutional blind signal separation and extraction, it is compared with two algorithms from the

literature [38,39]. To this end, we conducted a large number of repeated experiments to reduce the randomness and improve the reliability of the results. To assess the stability of these algorithms, this experiment controlled for the consistency of the input signals and parameters, and 20 Monte Carlo experiments were conducted with each separation algorithm as a single variable only. The three blind separation algorithms are compared in terms of separation accuracy (average correlation coefficient). Table 2 shows the average values of the above indicators obtained after 20 trials.

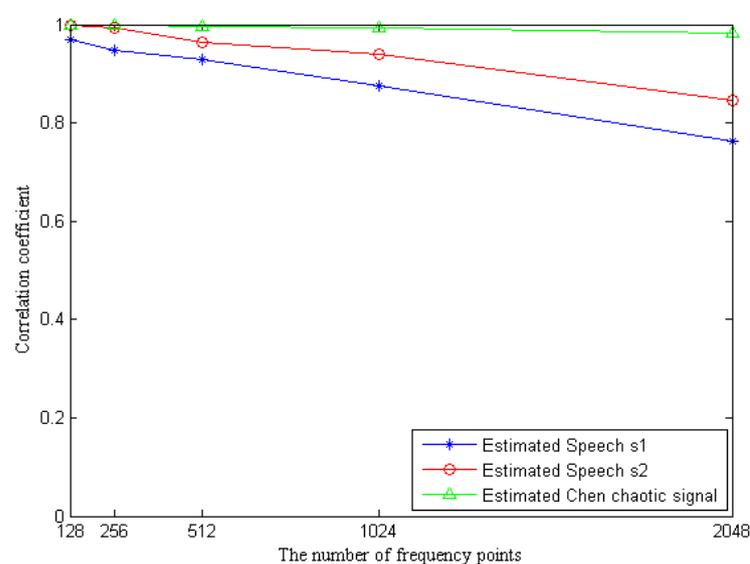
**Table 2.** Correlation coefficient of each separated signal with the source signal.

Algorithms	EFICA [38]	IF Algorithm [39]	Fast IVA
Correlation Coefficient	0.9532	0.9541	0.9732

From the comparative analysis of the experimental results, it can be seen that the Fast IVA algorithm has the highest correlation coefficient and has better separation accuracy. By comparing with EFICA algorithm and IF algorithm, it is proved that the algorithm has better convergence speed and separation performance. The Fast IVA algorithm used in this paper not only eliminates the tedious permutation process among the separated sub-signals and reduces the computational complexity of the algorithm but also has better separation performance.

#### 4.3. The Influence of Frequency Points in STFT

When setting the experimental parameters of the STFT, we mention one of the more important parameters, namely the number of frequency points. This is also the number of samples in the short-time window, and this parameter has an important impact on the STFT. The number of frequency points is too small, making the process of STFT too tedious and complicated, increasing the computational complexity of the algorithm. If the number of frequency points is too large, many features of the speech signal will be lost, which will affect the separation or extraction results of the blind extraction algorithm. Therefore, it is especially important to choose a suitable number of frequency points. We will discuss this parameter next, setting the number of frequency points to  $K = [256, 512, 1024, 2048]$ . The other experimental conditions are guaranteed to be constant, and the simulation experiments were performed sequentially using the Fast IVA algorithm, with the correlation coefficient as the evaluation criterion. The experimental results are the average data of 20 trials, as shown in Figure 6.

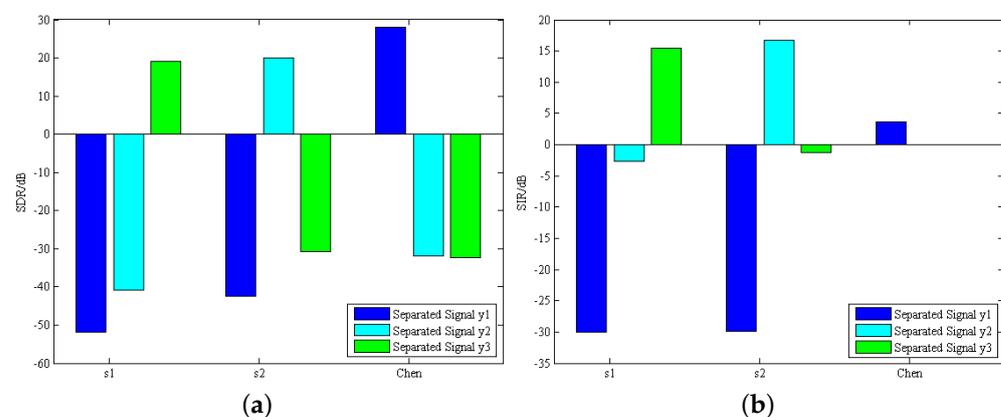


**Figure 6.** The relationship between frequency points and correlation coefficient.

The effect of the parameter selection of frequency points on the separation result is intuitively reflected in Figure 6. The larger the number of frequency points, the larger the number of samples in the short time window. For a speech signal with a certain length, the fewer frames are obtained by adding windows, which makes more speech features lost and results in a poorer recovered speech signal. With a smaller the number of frequency points, the opposite result is obtained. The experimental results in Figure 6 also confirm this relationship. When the number of frequency points  $K = 128$ , the average correlation coefficients of the speech estimation signals extracted by this algorithm are all above 0.96, and the separation effect is very good. When the number of frequency points  $K = 512$ , the correlation coefficients of the estimated signals of speech s1 and s2 can reach about 0.92 and 0.96, respectively. Although the correlation coefficients have decreased, the separation results are still better. When the number of frequency points  $K = 1024$  or even larger, the correlation coefficient of the speech estimation signal decreases more obviously, which makes the separation result worse. To ensure that the separation results obtained by this algorithm are better and to avoid too much computation in the STFT process, we compromise by choosing a frequency point  $K = 512$ . In addition, it can be seen from Figure 6 that the change of frequency points has little effect on the separation results of Chen's chaotic signal. Therefore, the superiority of the algorithm used in this paper is that it can guarantee the quality of the separated speech signal of good quality while still extracting the chaotic signal of noise class well.

#### 4.4. SDR/SIR Analysis

This section analyzes the degree of distortion and interference of the estimated speech signal. Here, the quantitative evaluation is based on SDR and SIR, which verifies the goodness of the separation results of the Fast IVA algorithm. The performance index is shown in Figure 7, where the SDR and SIR data values are the average values obtained from 10 tests.



**Figure 7.** The evaluation of algorithm separation performance with different metrics. (a) SDR value; (b) SIR value.

Figure 7 represents the performance evaluation of the separation results by different metrics; that is, the SDR and SIR values between each separated signal and the source signal, where s1 and s2 represent the two source speech signals and Chen is the chaotic signal. As shown in Figure 7a, the dark blue bars indicate the SDR values obtained after comparing the separated signal  $y_1$  with s1, s2 and s3. The SDR values with s1 and s2 are negative, and the SDR value with Chen is 27.9537 dB, which determines that the separated signal  $y_1$  is chaotic. Furthermore, the SDR values of the estimated signals of source speech s1 and s2 are also both close to 20 dB, which shows that the separated signals are less affected by distortion and interference. Similarly, in Figure 6b, the SIR values with s1 and s2 are negative due to the higher energy and larger bandwidth of the chaotic signal, and the SIR with s3 is calculated to be 3.6834 dB. The SIR values of both s1 and s2 estimated

signals are calculated to be around 15 dB, while the SIR value of the chaotic signal is almost 0, indicating that the estimated signal has few components separated from other sources, which in turn proves the superior performance of the separated signal.

#### 4.5. Perceptual Evaluation of Speech Quality

Perceptual evaluation of speech quality (PESQ) [40] is an objective, full-reference speech quality evaluation method with the International Telecommunication Union labeling code ITU-T P.863. PESQ was specially developed for simulating subjective tests commonly used in telecommunications to evaluate human voice quality. Therefore, PESQ uses real speech samples as test signals, based on comparative measurements between the original reference signal and the extracted signal. PESQ was created to provide a subjective Mean Opinion Score (MOS) [41] predictive value for objective speech quality evaluation, and it can be mapped to a scaled range of MOS values. The PESQ score ranges from  $-0.5$  to  $4.5$ , with higher scores indicating better speech quality.

The experimental conditions are the same as described above, and here the quality of separated speech is evaluated. The PESQ values of each estimated speech  $y_1$  and  $y_2$  compared with the original reference speech  $s_1$  and  $s_2$  are given in Table 3. Among them, the PESQ value calculated between  $y_1$  and  $s_2$  is as high as 3.9232, while the PESQ value between  $y_2$  and  $s_1$  is also 3.9754. By mapping the PESQ values to MOS, the quality of separated speech can be called “good”. It is again verified that the Fast IVA algorithm has good separation performance for blind deconvolution of chaos-obscured speech signals and ensures the quality of separated speech signals.

**Table 3.** PESQ value between each separated speech and source speech signal.

PESQ	s1	s2
$y_1$	0.3681	3.9754
$y_2$	3.9232	0.3563

To illustrate the high perceptual quality of the speech signal extracted by this algorithm, the results are analyzed here in comparison with other algorithms. The same experimental conditions as set as those in the literature [42], and 20 repeated experiments are performed to reduce the randomness of the results and obtain more stable and accurate results. The calculation result in Table 4 is the average value of PESQ obtained after 20 tests.

**Table 4.** PESQ averages with different algorithms.

Algorithms	DTW [42]	Fast IVA
PESQ	3.17	4.04

In Table 4, the PESQ value obtained by DTW algorithm is 3.17, and the PESQ value obtained by Fast IVA algorithm in this paper is relatively high, which can reach 4.04. Compared with the PESQ value obtained by DTW algorithm, it is 0.87 higher, which is enough to show that the quality of the extracted speech signal is good. This shows that the algorithm has better separation results than other traditional blind separation algorithms. The reason is that the Fast IVA algorithm is a blind extraction process performed sequentially on each vector frequency point without the need for a subsequent permutation process, thus ensuring high-quality estimated speech.

#### 4.6. Complexity Analysis

In this section, the computational complexity of the algorithm is analyzed and the time complexity of the key operational steps is now calculated for the FD-CBSS algorithm based on Fast IVA. Let the number of frequency points of STFT be  $K$  and the total number of frames of data be  $B$ . Assume that the number of source signals  $N$  is the same as the number

of observed signals  $M$ . For convenience, only multiplication operations are considered when calculating the complexity, and the complex-valued multiplication is a four-fold relation of the real-valued multiplication [43]. The real-valued multiplication operations required for the main procedure are shown in Table 5, where  $N_p$  represents the number of iterations of the Fast IVA algorithm.

**Table 5.** Calculation of the time complexity required by the main program.

Step	Time Complexity
STFT	$T_1 = O(NBK)$
Whitening Process	$T_2 = O(N_pNBK)$
Iterative Process	$T_3 = O(NBK)$
Normalization Process	$T_4 = O(NBK)$
ISTFT	$T_5 = O(NBK)$

Combining the above analysis, the time complexity of the main steps of the Fast IVA algorithm is calculated. Therefore, the time complexity of the overall process of the algorithm is the sum of above processes, which is

$$T = O(N_pNBK). \quad (16)$$

In order to compare the computational complexity of various algorithms, the running time of the frequency domain permutation algorithm in the literature [43] is given. Choose the same equipment and operating environment, the number of speech signals is  $N = 4$ , the length of the speech signal is 10 s and the sampling frequency is  $f_s = 8000$  Hz. In Table 5, the running time required for the completion of each algorithm is compared to illustrate the time complexity of the Fast IVA algorithm.

As shown in Table 6, the running times of the algorithm designed in [43] and in this paper are 13.7 s and 7.9 s, respectively. The significantly smaller running times indicate the low complexity of the algorithms. The biggest advantage of this algorithm is that the signal permutation is done during the signal separation process, eliminating the need for subsignal permutation algorithm design. Based on the computational complexity analysis of the above algorithms and the comparative analysis of the experimental results, the CBSS algorithm designed in this paper cannot only achieve high quality extraction of speech signals, but it can also complete the separation of signals quickly with low time complexity.

**Table 6.** The running time required for the algorithm to complete.

Method	Reference [43]	Fast IVA
Running Time	13.7 s	7.9 s

## 5. Conclusions

This paper introduced chaotic masking technology to hide multiple speech signals into random signals generated by the Chen chaotic dynamic system, which provided a guarantee for the secure communication of speech signals in a wireless environment. In the process of multipath channel transmission, the observation signals at the receiving end are the convolutional mixed signals of the multipath source signals due to the influence of the multipath effect. To extract the source speech signal efficiently, the CBSS algorithm was explored in depth. The traditional FD-CBSS algorithm is accompanied by permutation and amplitude ambiguity. For this reason, an additional permutation algorithm for separating sub-signals is needed to improve the separation accuracy of the algorithm. This will undoubtedly increase the computational complexity of the algorithm. Therefore, this paper proposed using an efficient Fast IVA algorithm to achieve the separation or extraction of source speech signals. The algorithm relies on the order of correlation between the frequency points and completes the signal separation according to the order of each frequency

point, eliminating the need for additional sorting steps. The simulation results show that the algorithm is not only effective and fast for blind separation of convolutional mixed signals but also suitable for extraction of noise-like signals. In addition, the application of this algorithm can reduce the overhead and eliminate the permutation operation in the traditional frequency-domain algorithm and reduce the algorithm complexity. In the next work, the algorithm applicability will be extended to investigate the underdetermined model for solving the practical problem that the number of microphones is less than source signals in the process of speech sending and receiving. Therefore, the length of the filter directly affects the degree of convolutional mixing of the source signal.

**Author Contributions:** Conceptualization, E.W.; methodology, S.G. and M.S.; software, S.G.; validation, S.G., M.S. and J.Y.; literature and data curation, Y.Z. and M.S.; writing—original draft preparation, S.G.; writing—review and editing, S.G. and E.W.; project administration, E.W.; funding acquisition, E.W. and S.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Natural Science Foundation of China, grant number 61801173; the Natural Science Foundation of Heilongjiang Province, China, grant number No.LH2019F048 and The Outstanding Youth Fund Project of Heilongjiang University, grant number No.YJSCX2020-167HLJU.

**Data Availability Statement:** Data sharing not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Shahadi, H.I. Covert communication model for speech signals based on an indirect and adaptive encryption technique. *Comput. Electr. Eng.* **2018**, *68*, 425–436. [[CrossRef](#)]
- Qi, D.; Nan, L.M.; Xu, J.F. A Speech Privacy Protection Method Based on Sound Masking and Speech Corpus. *Procedia Comput. Sci.* **2018**, *131*, 1269–1274. [[CrossRef](#)]
- Ntantogian, C.; Veroni, E.; Karopoulos, G.; Xenakis, C. A survey of voice and communication protection solutions against wiretapping. *Comput. Electr. Eng.* **2019**, *77*, 163–178. [[CrossRef](#)]
- Cao, X.L.; Jiang, W.H.; Tong, F. Time reversal MFSK acoustic communication in underwater channel with large multipath spread. *Ocean Eng.* **2018**, *152*, 203–209. [[CrossRef](#)]
- Leglaive, S.; Badeau, R.; Richard, G. Student's t Source and Mixing Models for Multichannel Audio Source Separation. *IEEE-ACM Trans. Audio Speech* **2018**, *26*, 1150–1164. [[CrossRef](#)]
- Abro, F.I.; Rauf, F.; Mobeen-ur-Rehman; Chowdhry, B.S.; Rajarajan, M. Towards Security of GSM Voice Communication. *Wirel. Pers. Commun.* **2019**, *108*, 1933–1955. [[CrossRef](#)]
- Dutta, M.; Roy, B.K. A new memductance-based fractional-order chaotic system and its fixed-time synchronization. *Chaos Solitons Fract.* **2021**, *145*, 110782. [[CrossRef](#)]
- Liao, X.; Zhou, G.; Yang, Q.; Fu, Y.; Chen, G. Constructive proof of Lagrange stability and sufficient—Necessary conditions of Lyapunov stability for Yang-Chen chaotic system. *Appl. Math. Comput.* **2017**, *309*, 205–221. [[CrossRef](#)]
- Musanna, F.; Kumar, S. Image encryption using quantum 3-D Baker map and generalized gray code coupled with fractional Chen's chaotic system. *Quantum Inf. Process.* **2020**, *19*, 1–31. [[CrossRef](#)]
- Chen, Y.J.; Chou, H.G.; Wang, W.J.; Tsai, S.H.; Tanaka, K.; Wang, H.O.; Wang, K.C. A polynomial-fuzzy-model-based synchronization methodology for the multi-scroll Chen chaotic secure communication system. *Eng. Appl. Artif. Intell.* **2020**, *87*, 103251. [[CrossRef](#)]
- Leung, C.T.; Siu, W.C. A general contrast function based blind source separation method for convolutively mixed independent sources. *Signal Process.* **2007**, *87*, 107–123. [[CrossRef](#)]
- Rahbar, K.; Reilly, J.P. A frequency domain method for blind source separation of convolutive audio mixtures. *IEEE Trans. Audio Speech* **2005**, *13*, 832–844. [[CrossRef](#)]
- Cheng, W.; Jia, Z.; Chen, X.; Gao, L. Convolutional blind source separation in frequency domain with kurtosis maximization by modified conjugate gradient. *Mech. Syst. Signal Process.* **2019**, *134*, 106331. [[CrossRef](#)]
- Zhang, K.; Chan, L.W. Convolutional blind source separation by efficient blind deconvolution and minimal filter distortion. *Neurocomputing* **2010**, *73*, 2580–2588. [[CrossRef](#)]
- Zhang, H.; Wang, G.; Cai, P.; Wu, Z.; Ding, S. A fast blind source separation algorithm based on the temporal structure of signals. *Neurocomputing* **2014**, *139*, 261–271. [[CrossRef](#)]
- Mei, T.; Mertins, A.; Yin, F.; Xi, J.; Chicharo, J.F. Blind source separation for convolutional mixtures based on the joint diagonalization of power spectral density matrices. *Signal Process.* **2008**, *88*, 1990–2007. [[CrossRef](#)]
- Pinchas, M. A New Efficient Expression for the Conditional Expectation of the Blind Adaptive Deconvolution Problem Valid for the Entire Range of Signal-to-Noise Ratio. *Entropy* **2019**, *21*, 72. [[CrossRef](#)]

18. Mateo, C.; Antonio Talavera, J. Short-Time Fourier Transform with the Window Size Fixed in the Frequency Domain (STFT-FD): Implementation. *Software* **2018**, *8*, 5–8. [[CrossRef](#)]
19. Comon, P. Independent component analysis, a new concept. *Signal Process.* **1994**, *36*, 287–314. [[CrossRef](#)]
20. Hyvarinen, A.; Karhunen, J.; Oja, E. *Independent Component Analysis*; John Wiley Sons: New York, NY, USA, 2001.
21. Xie, Y.; Xie, K.; Xie, S. Underdetermined convolutive blind separation of sources integrating tensor factorization and expectation maximization. *Digit. Signal Process.* **2019**, *87*, 145–154. [[CrossRef](#)]
22. Pedersen, M.S.; Larsen, J.; Kjems, U.; Parra, L.C. A survey of convolutive blind source separation methods. *Spring Handb. Speech Process. Commun.* **2007**, *8*, 1–34.
23. Murata, N.; Ikeda, S.; Ziehe, A. An approach to blind source separation based on temporal structure of speech signals. *Neurocomputing* **1998**, *41*, 1–24. [[CrossRef](#)]
24. Kim, T.; Lee, I.; Lee, T.W. Independent vector analysis: Definition and algorithms. In Proceedings of the 2006 Fortieth Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, 29 October–1 November 2006; Volume 6.
25. Kim, T.; Attias, H.T.; Lee, S.Y.; Lee, T.W. Blind source separation exploiting higher-order frequency dependencies. *IEEE Trans. Audio Speech* **2007**, *15*, 70–79. [[CrossRef](#)]
26. Waqas, R.; Syed, M.N.; Jonathon, A.C. Mixed Source Prior for the Fast Independent Vector Analysis Algorithm. In Proceedings of the 2016 9th IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM 2016), Rio de Janeiro, Brazil, 10–13 July 2016; pp. 1–5.
27. Lee, I.; Kim, T.; Lee, T.W. Fast fixed-point independent vector analysis algorithms for convolutive blind source separation. *Signal Process.* **2007**, *87*, 1859–1871. [[CrossRef](#)]
28. Kim, T. Real-Time Independent Vector Analysis for Convolutive Blind Source Separation. *IEEE Trans. Circuits Syst. I* **2010**, *57*, 1431–1438.
29. Anderson, M.; Adali, T.; Li, X. Joint Blind Source Separation With Multivariate Gaussian Model: Algorithms and Performance Analysis. *IEEE Trans. Signal Process.* **2012**, *60*, 1672–1683. [[CrossRef](#)]
30. Liang, Y.; Naqvi, S.M.; Chambers, J.A. Audio video based fast fixed-point independent vector analysis for multisource separation in a room environment. *EURASIP J. Adv. Signal Process.* **2012**, *183*. [[CrossRef](#)]
31. Chen, G.R.; Ueta, T. Yet Another Chaotic Attractor. *Int. J. Bifurc. Chaos* **1999**, *73*, 3. [[CrossRef](#)]
32. Lü, J.H.; Zhou T.S.; Chen G.R. The compound structure of Chen’s attractor. *Int. J. Bifurc. Chaos* **2002**, *12*, 855–858. [[CrossRef](#)]
33. Wang, P.; Li, J.; Zhang, H. Decoupled Independent Vector Analysis Algorithm for Convolutive Blind Source Separation without Orthogonality Constraint on the Demixing Matrices. *Math. Probl. Eng.* **2018**, *2018*. [[CrossRef](#)]
34. Hershey, J.R.; Olsen, P.A. Approximating the Kullback Leibler divergence between Gaussian mixture models. In Proceedings of the 2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP, Honolulu, HI, USA, 15–20 April 2007.
35. Chen, P.; Peng, Y.; Wang, S. The Hessian matrix of Lagrange function. *Linear Algebra Appl.* **2017**, *531*, 537–546. [[CrossRef](#)]
36. Vincent, E.; Gribonval, R.; Févotte, C. Performance measurement in blind audio source separation. *IEEE Trans. Audio Speech* **2006**, *14*, 1462–1469. [[CrossRef](#)]
37. Vincent, E.; Araki, S.; Theis, F.; Nolte, G.; Bofill, P.; Sawada, H.; Ozerov, A.; Gowreesunker, V.; Lutter, D.; Duong, N.Q.K. The signal separation evaluation campaign (2007–2010): Achievements and remaining challenges. *Signal Process.* **2012**, *92*, 1928–1936. [[CrossRef](#)]
38. Koldovsky, Z.; Tichavsky, P.; Oja, E. Efficient variant of algorithm FastICA for independent component analysis attaining the Cramer-Rao lower bound. *IEEE Trans. Neural Netw.* **2006**, *17*, 1265–1277. [[CrossRef](#)] [[PubMed](#)]
39. Bo, X.; He, Y.; Yin, B.; Fang, G.; Fan, X.; Li, Z. Algorithm to eliminate permutation of frequency domain blind source separation based on influence factor. *Acta Electron. Sin.* **2014**, *42*, 360–365.
40. Srinivasarao, V.; Ghanekar, U. Speech enhancement—An enhanced principal component analysis (EPCA) filter approach. *Comput. Electr. Eng.* **2020**, *85*, 106657. [[CrossRef](#)]
41. Mahesh, V.; Madhubalan, V. Measuring speech quality for text-to-speech systems: development and assessment of a modified mean opinion score (MOS) scale. *Comput. Speech Lang.* **2005**, *19*, 55–83.
42. Lv, Z.; Zhang, B.B.; Wu, X.P.; Zhang, C.; Zhou, B.Y. A permutation algorithm based on dynamic time warping in speech frequency-domain blind source separation. *Speech Commun.* **2017**, *92*, 132–141. [[CrossRef](#)]
43. Fang, K.; Feiran, Y.; Jun, Y. A low-complexity permutation alignment method for frequency-domain blind source separation. *Speech Commun.* **2019**, *115*, 88–94.