

## Article

# Exploiting Time–Frequency Sparsity for Dual-Sensor Blind Source Separation

Jiajia Chen, Haijian Zhang \* and Siyu Sun

School of Electronic Information, Wuhan University, Wuhan 430072, China; chenjia@whu.edu.cn

\* Correspondence: haijian.zhang@whu.edu.cn

**Abstract:** This paper explores the important role of blind source separation (BSS) techniques in separating  $M$  mixtures including  $N$  sources using a dual-sensor array, i.e.,  $M = 2$ , and proposes an efficient two-stage underdetermined BSS (UBSS) algorithm to estimate the mixing matrix and achieve source recovery by exploiting time–frequency (TF) sparsity. First, we design a mixing matrix estimation method by precisely identifying high clustering property single-source TF points (HCP-SSPs) with a spatial vector dictionary based on the principle of matching pursuit (MP). Second, the problem of source recovery in the TF domain is reformulated as an equivalent sparse recovery model with a relaxed sparse condition, i.e., enabling the number of active sources at each auto-source TF point (ASP) to be larger than  $M$ . This sparse recovery model relies on the sparsity of an ASP matrix formed by stacking a set of predefined spatial TF vectors; current sparse recovery tools could be utilized to reconstruct  $N > 2$  sources. Experimental results are provided to demonstrate the effectiveness of the proposed UBSS algorithm with an easily configured two-sensor array.

**Keywords:** underdetermined blind source separation; dual-sensor; mixing matrix estimation; source number estimation; time–frequency sparsity



**Citation:** Chen, J.; Zhang, H.; Sun, S. Exploiting Time–Frequency Sparsity for Dual-Sensor Blind Source Separation. *Electronics* **2024**, *13*, 1227. <https://doi.org/10.3390/electronics13071227>

Academic Editors: Daniele Riboni, Fangqing Wen, Yangyang Dong and Hua Chen

Received: 2 February 2024

Revised: 19 March 2024

Accepted: 25 March 2024

Published: 26 March 2024



**Copyright:** © 2024 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

### 1.1. Background

With the rapid growth of wireless systems and their applications, frequency spectrum has become increasingly congested, which might bring unwanted co-channel interference (intentional or unintentional) to expected signals [1]. In order to further deal with the signal of interest given the received mixtures, blind source separation (BSS) techniques have exhibited potential ability to extract informative signals and suppress undesirable signals with the aim of improving spectrum efficiency. Stemming from the cocktail party problem, BSS attempts to reconstruct the original signals from observed mixtures without the prior information of mixing weights and original sources, and is widely applied in wireless communication [2,3], speech processing [4,5], image processing [6], biomedicine [7,8], and more.

Typical techniques, which have achieved considerable success in addressing the BSS problem, include but are not limited to independent component analysis (ICA) and its variations [9,10], sparse component analysis (SCA) [7,11,12], sparse bounded component analysis (SBICA) [13], and non-negative matrix factorization (NMF) [14,15]. Recently, there has been an increasing interest in deep learning-based data-driven approaches [16,17]. Among these techniques, sparsity-based methods [18–22] have been extensively utilized due to their versatility in various situations for both (over)determined and underdetermined mixtures. The success of sparsity-based methods depends on the fact that sources in different fields are either sparse or sparsely representable in a certain domain, e.g., the time–frequency (TF) domain [19,23] or wavelet domain [24].

Taking into account the implementation cost and space arrangement, BSS with a limited number of sensors [25] should be more attractive and appropriate for practical

scenarios, e.g., speech enhancement systems [26] and multiple-input–multiple-output radar or communication systems [27], where low-cost and small-space configuration are highlighted. Despite the extensive prior works and applications laying a foundation for BSS, dual-sensor BSS has not been effectively addressed in the literature. Therefore, our focus is devoted to the underdetermined BSS (UBSS) problem with a two-sensor array, aiming to solve the mixing matrix and source estimation problems by exploiting the TF sparsity of sources.

### 1.2. Related Work

Traditionally, the UBSS model faces two challenging issues: mixing matrix estimation (MME) and source recovery. In the literature, some BSS strategies determine both the mixing matrix and sources simultaneously or circumvent the MME problem entirely [28,29]. As opposed to these methods, we emphasize a two-step UBSS strategy [30], i.e., first estimating the mixing matrix and then recovering the sources.

For the first issue, various MME methods [18,21,27,31–38] consider the TF points where only single source occurs or possesses dominant energy. These single-source TF points (SSPs) provide a directional clustering property, which can be associated with each column of the mixing matrix. Hence, clustering the mixture TF vectors at correct SSPs significantly improves the estimation accuracy. Reju et al. [31] proposed a simple method for detecting SSPs by comparing the absolute directions of the real and imaginary parts of the mixture TF coefficient vectors. Estimation methods [32,33] have been reported for discovering SSPs by investigating the difference between the real and imaginary parts of mixture TF vectors at SSPs and multi-source TF points (MSPs). Recently, Zhen et al. [34] exploited the sparse coding technique to find SSPs with some one-dimensional (1D) subspaces from all the TF vectors of observed mixtures. However, the above methods are specifically designed for real-valued mixing matrix and are not suitable for complex-valued cases. In [18], those TF points having sufficient energy were regarded as SSPs, and the complex-valued mixing matrix was estimated by directly clustering them; however, the estimation performance becomes degraded when sources are highly overlapped in the TF domain. A noise-robust estimation method was developed in [21] by decomposing the signal energy of each TF point into two parts and selecting those TF points where only one source possesses dominant energy. In [35], agglomerative hierarchical clustering was applied for automatic clustering of SSPs. Guo et al. [36] proposed a complex-valued mixing matrix estimation method based on L-shaped arrays and uniform circular arrays, i.e., the arrangement of arrays needs to be fixed.

The second issue is to separate those sources with fewer sensors. A number of early works exploiting TF sparsity assumed the sources to be TF-disjoint, i.e., there most one active source exists at any auto-source TF point (ASP). In [25,39], the authors proposed the degenerate unmixing estimation technique (DUET) for separating more sources than mixtures by assuming that the sources are  $W$ -disjoint orthogonal in the TF domain. The TF ratio of mixtures (TIFROM) method was proposed in [40] to relax the restrictive  $W$ -disjoint assumption; however, this approach still requires adjacent TF regions where only one source occurs. Aissa-El-Bey et al. [18] applied the subspace projection technique to signal synthesis to separate TF-nondisjoint sources, with the number of sources existing at every ASP assumed to be less than the number of sensors. This sparse assumption was relaxed in [41] where the number of active sources at an ASP does not exceed that of sensors. A UBSS algorithm with free active sources was reported [20] based on the Wigner–Ville distribution and Khatri–Rao product, which further relaxes the sparse constraint of sources in the TF domain (see Table 1).

On the other hand, the source separation problem can be addressed based on sparse recovery theory. Bofill et al. [42] proposed a staged UBSS approach performed in the frequency domain (with much higher sparsity than the time domain), where the mixing matrix was first estimated by a potential function-based method, then the sources were inferred by minimal  $l_1$ -norm decomposition obtained by solving low-dimensional linear

programming problems. Subsequently, more and more attention has been paid to this paradigm for sparse signal reconstruction. Li et al. [43] analyzed the two-stage cluster-then- $l_1$ -optimization approach in [42] and its recoverability for UBSS. Later [24], they further extended and discussed the applications of the two-stage sparse representation approach. Recently, Xie et al. [7] proposed an improved  $l_1$ -norm minimization algorithm to estimate the source signals and employed preconditioned conjugate gradient technology to accelerate the convergence rate such that the computational load is reduced. With the development of compressed sensing (CS) [44], CS reconstruction methods, e.g., orthogonal matching pursuit (OMP) [45,46], have been applied to handle the source signal recovery problem. In order to reduce the sparsity limitation of signals, Liu et al. [47] proposed a source recovery method based on submatrix transformation and multi-source point compensation for the dual-channel underdetermined TF overlapped signals.

**Table 1.** Summary of existing two-step BSS algorithms.

Reference	Mixing Matrix Estimation	Source Recovery
[18]	$M \geq 2$ (Real/Complex)	$N_{\text{tf}} < M$
[21]	$M > 2$ (Real/Complex)	$N_{\text{tf}} < M$
[41]	$M \geq 2$ (Real/Complex)	$N_{\text{tf}} \leq M$
[20]	$M \geq 2$ (Real)	$N \leq 2M - 1$
[34]	$M \geq 2$ (Real)	$N_{\text{tf}} < M$
Proposed	$M \geq 2$ (Real/Complex)	$\sum N_{\text{tf}} < MN_{\text{asp}}$

$M$  is the number of sensors,  $N$  is the number of sources,  $N_{\text{asp}}$  is the number of ASPs, and  $N_{\text{tf}}$  denotes the number of sources at each ASP in the TF domain.

### 1.3. Motivation and Contribution

Although much effort has been made to solve the UBSS problem, potential remains for enhanced performance in the development of a UBSS model with a dual-sensor array, which is in accordance with many practical situations requiring low computational cost. Generally, the aforementioned works might have the following limitations. (i) Existing BSS algorithms exploiting TF sparsity are summarized in Table 1, where the sparsity conditions for both mixing matrix estimation and source recovery are specified. Note that an obstacle to the application of BSS is the strict constraint imposed on the sparse conditions, e.g., the MME method in [21] has the requirement that the number of sensors must be more than two. (ii) To a large extent, the separation performance relies heavily on the distinction between the number of sources and the number of sensors, i.e., when the number of sources is augmented with a fixed sensor array, the performance suffers significant degradation. (iii) Obtaining accurate sparse solutions while avoiding trivial solutions under a suitable sparsity constraint remains challenging for the source recovery problem [7]. The above difficulties are real challenges faced by researchers, and highlight the need for improved UBSS performance using low-dimensional arrays.

In this paper, we take advantage of a two-sensor array and propose a novel UBSS algorithm by exploiting auto-source TF sparsity (ASTFS-UBSS). The two-step paradigm is adopted to separate multiple sources by first estimating the mixing matrix and then separating the sources. In the first step, we design an MME method consisting of SSP detection and subsequent clustering. Direct clustering is carried out to generate a spatial vector dictionary, which is later employed to select SSPs with a high clustering property. This selection is implemented by assigning a clustering property threshold according to the principle of matching pursuit (MP) [48]. With the selected TF points, we then cluster them through a postprocessing operation to obtain the estimated mixing matrix. Next, a sparse recovery problem is formulated by reshaping all the mixture TF vectors at ASPs into a 1D vector based on the fact that only a limited number of sources are active at ASPs. As such, we formulate a dictionary matrix in which the TF domain sources can be sparsely represented, then exploit this sparsity to identify the expected sources by  $l_1$ -norm minimization. The formulated sparse representation permits current sparse recovery algorithms to acquire the reconstructed sources.

The contribution of the proposed ASTFS-UBSS algorithm is twofold. First, the proposed high clustering property SSP-based MME (HCP-SSP-MME) method shows the potential of both real-valued and complex-valued mixing matrices. Second, according to the particular properties and structure distribution of different sources in the TF domain, e.g., regular harmonic pattern of speech sources, exploiting TF sparsity from the perspective of ASPs enables us to separate multiple sources using only  $M = 2$  sensors with the sparsity constraint  $\sum N_{\text{tf}} < MN_{\text{asp}}$  in Table 1. Theoretical and experimental analyses demonstrate the effectiveness and feasibility of the ASTFS-UBSS algorithm with a simply configured dual-sensor array.

The remainder of the paper is organized as follows: we elaborate the proposed ASTFS-UBSS algorithm in Section 2, where the HCP-SSP-MME method is described based on MP and a sparse recovery model is established by exploiting the sparsity of source TF vectors at the ASPs; in Section 3, the proposed ASTFS-UBSS is evaluated in terms of mixing matrix estimation and source separation; finally, Section 4 concludes the paper.

## 2. Proposed ASTFS-UBSS Algorithm

This paper considers the UBSS problem including  $M$  instantaneous mixtures and  $N$  sources, which is modeled as

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t), \quad (1)$$

where  $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_n, \dots, \mathbf{a}_N]$  denotes the mixing matrix and  $\mathbf{a}_n = [a_{n1}, \dots, a_{nm}, \dots, a_{nM}]^T$  is the steering vector of the  $n$ -th source. Here,  $\mathbf{x}(t) = [x_1(t), \dots, x_m(t), \dots, x_M(t)]^T$  are observation mixtures, where  $T$  denotes the transpose operator, while  $\mathbf{s}(t) = [s_1(t), \dots, s_n(t), \dots, s_N(t)]^T$  are original sources and  $\mathbf{n}(t)$  is the additive white noise vector. It should be highlighted that a dual-sensor array, i.e.,  $M = 2$ , is advocated in our model ( $M > 2$  mixtures are also considered for simulations). The objective is to estimate  $\mathbf{s}(t)$  by learning  $\mathbf{A}$  using the received  $\mathbf{x}(t)$ .

The above UBSS model is generally built without any prior information about the mixing process and the sources, which leads to infinite solutions of (1) even though the mixing matrix  $\mathbf{A}$  is available. Thanks to the sparse property of signal TF distributions, we exploit the sparsity of ASP vectors to address the dual-sensor UBSS problem. Because the short-time Fourier transform (STFT) is easy to implement and does not involve cross-terms in the TF domain, we transform the time domain mixtures in (1) into the STFT domain, which provides

$$\mathbf{S}_x(t, f) = \mathbf{A}\mathbf{S}_s(t, f) + \mathbf{S}_n(t, f), \quad (2)$$

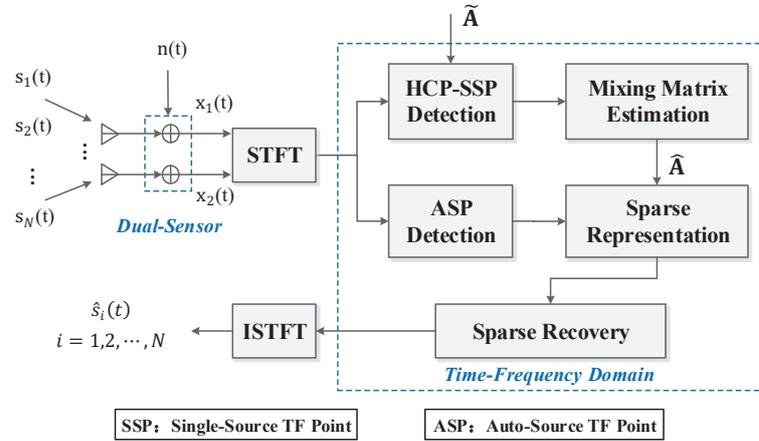
where  $\mathbf{S}$  denotes the STFT operator and  $\mathbf{S}_x(t, f) \in \mathbb{C}^{M \times 1}$ ,  $\mathbf{S}_s(t, f) \in \mathbb{C}^{N \times 1}$ , and  $\mathbf{S}_n(t, f) \in \mathbb{C}^{M \times 1}$  are the mixture TF vector, source TF vector, and noise TF vector, respectively.

The block diagram of the proposed dual-sensor ASTFS-UBSS algorithm is illustrated in Figure 1. In the MME stage, a spatial vector dictionary  $\tilde{\mathbf{A}}$  is generated by directly clustering the mixture TF vectors at some detected ASPs with strong energy. This initial  $\tilde{\mathbf{A}}$  is utilized to identify a set of HCP-SSPs with the principle of MP, then these HCP-SSPs are clustered to obtain an accurate estimation of the mixing matrix  $\hat{\mathbf{A}}$ . In the source recovery stage, a sparse recovery model is established by exploiting the sparsity of auto-source TF vectors. Detailed descriptions of the two-step ASTFS-UBSS algorithm are presented in the following subsections.

### 2.1. HCP-SSP-Based Mixing Matrix Estimation (MME)

Auto-source TF points (ASPs), including SSPs and MSPs, represent the TF points where at least one source exists [18,21]. The principal challenge in the design of an SSP-based MME method is for the accurate detection of some SSPs at which the mixture TF vectors embody a superior clustering property with respect to each steering vector, i.e., each column vector of the mixing matrix  $\mathbf{A}$ . Mixture TF vectors in (2) for ideal SSPs would provide helpful directional information about the steering vectors of each source, whereas this clustering property tends to be distorted for MSPs. In addition, the superiority of

the clustering property on SSPs will be weakened due to interference and strong noise, resulting in a low clustering property. Therefore, in order to accurately estimate  $\mathbf{A}$ , the high clustering property SSPs (HCP-SSPs) need to be preserved while eliminating the MSPs and low clustering property SSPs (LCP-SSPs).



**Figure 1.** Block diagram of the proposed ASTFS-UBSS algorithm ( $\tilde{\mathbf{A}}$  and  $\hat{\mathbf{A}}$  are the initial estimation and final estimation of mixing matrix  $\mathbf{A}$ , respectively).

In this subsection, we elaborate a simple yet efficient MME method by accurately locating a group of HCP-SSPs (hereinafter termed as HCP-SSP-MME), which follows the two assumptions below.

**Assumption 1.** For each source, there are some TF points where only this source exists or where its energy is dominant.

**Assumption 2.** Any  $M$  column vectors of the mixing matrix  $\mathbf{A}^{M \times N}$  are linearly independent. All the column vectors of  $\mathbf{A}$  have a unit norm, i.e.,  $\|\mathbf{a}_n\| = 1, n = 1, \dots, N$ .

Note that **Assumption 1**, which extends the definition of SSPs by considering noise influence and threshold operations, can be easily satisfied by a wide range of practical scenarios, as it allows the corresponding SSPs belonging to each source to be arbitrarily distributed in the TF domain rather than requiring a connected SSP region for every source. **Assumption 2** guarantees that all sources can be successfully recovered, and is a common assumption in existing BSS algorithms [11,24,31,34,49]. Furthermore, it avoids indeterminacies due to scaling and permutation [18].

### 2.1.1. Detection of ASPs with Strong Energy

To increase noise robustness while reducing computational complexity, we prefer to deal with those ASPs having significant enough energy for MME. Specifically, a set of strong-energy TF points (SEPs) in the STFT domain can be selected using the following criterion:

$$\text{If } \frac{\|\mathbf{S}_{\mathbf{x}}(t, f)\|}{\max_{\tau, v} \{\|\mathbf{S}_{\mathbf{x}}(\tau, v)\|\}} > T_0, \text{ then } (t, f) \in \Omega_{\text{sep}} \quad (3)$$

where  $\|\cdot\|$  denotes the Euclidean norm and  $T_0 \in (0, 1)$  is an empirical threshold value for selecting SEPs. All the ASPs meeting the criterion in (3) are included in the set  $\Omega_{\text{sep}}$ . The above process further enhances the signal sparsity and suppresses the influence of noise.

Because the SEPs have negligible noise energy, the signal TF model in (2) for an SEP  $(t, f)$  is changed into

$$\mathbf{S}_x(t, f) \approx \mathbf{A}\mathbf{S}_s(t, f) = \sum_{n=1}^N S_s^{(n)}(t, f)\mathbf{a}_n, \tag{4}$$

where  $S_s^{(n)}(t, f)$  denotes the STFT value of the  $n$ -th source and can be regarded as the coefficient of the basis  $\mathbf{a}_n$ . Thus, the term  $\sum_{n=1}^N S_s^{(n)}(t, f)\mathbf{a}_n$  represents a linear expansion of  $\mathbf{S}_x(t, f)$ . Assuming a desirable HCP-SSP on which only the  $n$ -th source is active, Equation (4) degrades into

$$\mathbf{S}_x(t, f) \approx S_s^{(n)}(t, f)\mathbf{a}_n, \quad n \in \{1, \dots, N\}, \tag{5}$$

which means that the steering vector  $\mathbf{a}_n$  is co-linear with the observed mixture TF vector  $\mathbf{S}_x(t, f)$  at the TF point  $(t, f)$ . Clearly, the vector  $\mathbf{S}_x(t, f)$  at an HCP-SSP possesses favourable directional information about  $\mathbf{a}_n$ ; thus, we can estimate  $\mathbf{a}_n$  by averaging the mixture TF vectors of the  $n$ -th source-related HCP-SSPs:

$$\hat{\mathbf{a}}_n = \frac{1}{\#\Omega_{\text{hcp}}^{(n)}} \sum_{(t,f) \in \Omega_{\text{hcp}}^{(n)}} \frac{\mathbf{S}_x(t, f)}{S_{x_{\text{ref}}}(t, f)}, \quad n = 1, \dots, N \tag{6}$$

where  $S_{x_{\text{ref}}}(t, f)$  represents the TF value of the mixture reference channel,  $\#\Omega_{\text{hcp}}^{(n)}$  denotes the total number of HCP-SSPs included in the set  $\Omega_{\text{hcp}}^{(n)}$ ,  $n = 1, \dots, N$ , and  $\cup \Omega_{\text{hcp}}^{(n)} = \Omega_{\text{hcp}}$ ; in addition, note that  $\Omega_{\text{hcp}} \subset \Omega_{\text{sep}}$ . The aim of the proposed MME lies in determining the set  $\Omega_{\text{hcp}}$ , i.e., locating a set of HCP-SSPs from the detected SEPs in (3), where the MSPs and LCP-SSPs are regarded as outliers to be eliminated.

### 2.1.2. Estimation of Initial Dictionary $\tilde{\mathbf{A}}$

To identify a set of HCP-SSPs which satisfy the condition in (5), an initial estimation of the steering vectors  $\{\mathbf{a}_n\}_{n=1, \dots, N}$  is required. To address this problem, we generate a spatial vector dictionary  $\tilde{\mathbf{A}}$  by clustering the mixture TF vectors of all the SEPs defined in (3) using  $k$ -means clustering. As a result, each mixture TF vector of  $\tilde{\mathbf{A}}$  can be estimated the same way as in (6), i.e.,

$$\tilde{\mathbf{a}}_n = \frac{1}{\#\Omega_{\text{sep}}^{(n)}} \sum_{(t,f) \in \Omega_{\text{sep}}^{(n)}} \frac{\mathbf{S}_x(t, f)}{S_{x_{\text{ref}}}(t, f)}, \quad n = 1, \dots, N_0, \tag{7}$$

where  $\#\Omega_{\text{sep}}^{(n)}$  denotes the number of SEPs in the  $n$ -th cluster set  $\Omega_{\text{sep}}^{(n)}$ ,  $n = 1, \dots, N_0$ , and  $\cup \Omega_{\text{sep}}^{(n)} = \Omega_{\text{sep}}$ . Herein,  $N_0$  is assigned a larger value than  $N$  to ensure that the steering vectors of all sources are included in  $\tilde{\mathbf{A}}$ .

### 2.1.3. Identification of HCP-SSPs

Next, we attempt to identify a group of HCP-SSPs from the SEPs in (3) while removing the outliers, i.e., MSPs and LCP-SSPs. In this paper, we assume that there are at most two sources contributing dominant energy at each MSP.

Based on the decompositions in (4) and (5), the observed mixture TF vector  $\mathbf{S}_x(t, f)$  can be considered as a linear expansion of several bases (columns) belonging to  $\tilde{\mathbf{A}}$ . In this sense, we resort to the attractive property of MP [48], which defines a linear expansion of  $\mathbf{S}_x(t, f)$  to best match its inner structure by successive approximation of  $\mathbf{S}_x(t, f)$  with orthogonal projection on the bases in  $\tilde{\mathbf{A}}$ . According to this property of MP, the mixture TF vector of each SEP in the set  $\Omega_{\text{sep}}$  can be decomposed as

$$\mathbf{S}_x(t, f) = \left\langle \mathbf{S}_x(t, f), \tilde{\mathbf{a}}_{n_1} \right\rangle \tilde{\mathbf{a}}_{n_1} + \mathbf{r}^{(1)}(t, f), \tag{8}$$

where  $\tilde{\mathbf{a}}_{n_1}$  is the  $n_1$ -th column in  $\tilde{\mathbf{A}}$  and  $\langle \cdot, \cdot \rangle$  is the inner product operator. The term  $\mathbf{r}^{(1)}(t, f)$  denotes the residual vector after approximating  $\mathbf{S}_x(t, f)$  in the direction of  $\tilde{\mathbf{a}}_{n_1}$ .

Assuming an SSP with the  $\alpha$ -th source having the dominant energy, according to (8) the steering vector  $\mathbf{a}_\alpha$  of this SSP can be estimated by minimizing the norm of the residual vector  $\|\mathbf{r}^{(1)}(t, f)\|$ , expressed as follows:

$$\hat{\mathbf{a}}_\alpha = \arg \min_{\tilde{\mathbf{a}}_{n_1}} \left\{ \left\| \mathbf{S}_x(t, f) - \langle \mathbf{S}_x(t, f), \tilde{\mathbf{a}}_{n_1} \rangle \tilde{\mathbf{a}}_{n_1} \right\| \right\}, \tag{9}$$

where  $n_1 = 1, \dots, N_0$ .

This becomes more complicated for an MSP, as  $\mathbf{S}_x(t, f)$  in (8) involves multiple bases. Fortunately, MP is an iterative process that continuously subdecomposes the residue  $\mathbf{r}^{(1)}(t, f)$  by projecting it on another base in  $\tilde{\mathbf{A}}$  in the same way as in (8) and (9). Defining  $\mathbf{S}_x(t, f) = \mathbf{r}^{(0)}(t, f)$ , the residual vector  $\mathbf{r}^{(i)}(t, f)$  in the  $i$ -th level of decomposition is expressed as

$$\mathbf{r}^{(i)}(t, f) = \langle \mathbf{r}^{(i)}(t, f), \tilde{\mathbf{a}}_{n_{i+1}} \rangle \tilde{\mathbf{a}}_{n_{i+1}} + \mathbf{r}^{(i+1)}(t, f), \tag{10}$$

where  $\tilde{\mathbf{a}}_{n_{i+1}}$  is the  $n_{i+1}$ -th column in  $\tilde{\mathbf{A}}$ . Thus, the expression in (8) for a  $D$ -level decomposition is changed into

$$\mathbf{S}_x(t, f) = \sum_{i=0}^{D-1} \langle \mathbf{r}^{(i)}(t, f), \tilde{\mathbf{a}}_{n_{i+1}} \rangle \tilde{\mathbf{a}}_{n_{i+1}} + \mathbf{r}^{(D)}(t, f), \tag{11}$$

where  $\{\tilde{\mathbf{a}}_{n_{i+1}}\}_{i=0, \dots, D-1}$  are  $D$  bases from  $\tilde{\mathbf{A}}$ , which can be estimated in a similar way as in (9). This implies that for each SEP in  $\Omega_{\text{sep}}$  we can determine  $D$  sources which contribute significant energy to  $\mathbf{S}_x(t, f)$  of this SEP.

It has been demonstrated that setting  $D = 2$  in (11) would be capable of distinguishing between MSPs, HCP-SSPs, and LCP-SSPs; in this way, Equation (11) can be rewritten as

$$\mathbf{S}_x(t, f) = \langle \mathbf{S}_x(t, f), \tilde{\mathbf{a}}_{n_1} \rangle \tilde{\mathbf{a}}_{n_1} + \langle \mathbf{r}^{(1)}(t, f), \tilde{\mathbf{a}}_{n_2} \rangle \tilde{\mathbf{a}}_{n_2} + \mathbf{r}^{(2)}(t, f), \tag{12}$$

where  $\tilde{\mathbf{a}}_{n_1}$  and  $\tilde{\mathbf{a}}_{n_2}$  are two bases from  $\tilde{\mathbf{A}}$  which correspond to the steering vectors of two contributing sources.

Because the resulting approximation of MP after any finite number of iterations might be suboptimal, as an alternative we apply orthogonal matching pursuit (OMP) [45,50,51] to compute the residue  $\mathbf{r}^{(1)}(t, f)$  in (12):

$$\mathbf{r}^{(1)}(t, f) = \mathbf{S}_x(t, f) - \tilde{\mathbf{a}}_{n_1} \tilde{\mathbf{a}}_{n_1}^\dagger \mathbf{S}_x(t, f) \tag{13}$$

where  $\dagger$  denotes the Moore–Penrose pseudo-inversion operator. Then, Equation (12) is further expressed as

$$\mathbf{S}_x(t, f) = \mathcal{C}_1 \tilde{\mathbf{a}}_{n_1} + \mathcal{C}_2 \tilde{\mathbf{a}}_{n_2} + \mathbf{r}^{(2)}(t, f), \tag{14}$$

and

$$\begin{cases} \mathcal{C}_1 = \langle \mathbf{S}_x(t, f), \tilde{\mathbf{a}}_{n_1} \rangle \\ \mathcal{C}_2 = \langle \mathbf{S}_x(t, f) - \tilde{\mathbf{a}}_{n_1} \tilde{\mathbf{a}}_{n_1}^\dagger \mathbf{S}_x(t, f), \tilde{\mathbf{a}}_{n_2} \rangle, \end{cases} \tag{15}$$

where  $\tilde{\mathbf{a}}_{n_1}$  is the first base in  $\tilde{\mathbf{A}}$  determined by (9) and the second base  $\tilde{\mathbf{a}}_{n_2}$  can be obtained via

$$\tilde{\mathbf{a}}_{n_2} = \arg \min_{\tilde{\mathbf{a}}_{n_2}} \left\{ \left\| \mathbf{S}_x(t, f) - \mathcal{C}_1 \tilde{\mathbf{a}}_{n_1} - \mathcal{C}_2 \tilde{\mathbf{a}}_{n_2} \right\| \right\}, \tag{16}$$

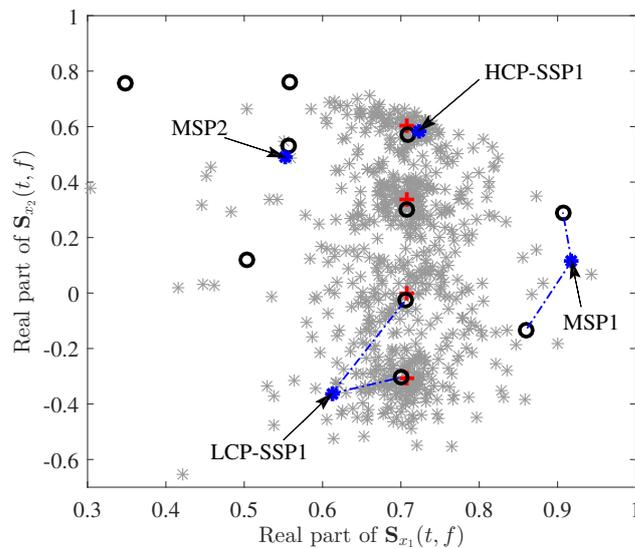
where  $n_1, n_2 \in \{1, 2, \dots, N_0\}$  and  $\mathcal{C}_1 \geq \mathcal{C}_2$ .

Under the noiseless situation and assuming known  $\mathbf{A}$ , an SSP with the  $\alpha$ -th active source can be easily identified according to the value of  $\mathcal{C}_2$  and (5), derived as follows:

$$\begin{aligned}
 \mathcal{C}_2 &= \left\langle \mathbf{S}_x(t, f) - \tilde{\mathbf{a}}_{n_1} \tilde{\mathbf{a}}_{n_1}^\dagger \mathbf{S}_x(t, f), \tilde{\mathbf{a}}_{n_2} \right\rangle \\
 &= \left\langle \mathbf{S}_x(t, f) - \mathbf{a}_\alpha \mathbf{a}_\alpha^\dagger \mathbf{S}_x(t, f), \tilde{\mathbf{a}}_{n_2} \right\rangle \\
 &= \left\langle \mathbf{S}_x(t, f) - S_s^{(\alpha)}(t, f) \mathbf{a}_\alpha \mathbf{a}_\alpha^\dagger, \tilde{\mathbf{a}}_{n_2} \right\rangle \\
 &= \left\langle \mathbf{S}_x(t, f) - S_s^{(\alpha)}(t, f) \mathbf{a}_\alpha, \tilde{\mathbf{a}}_{n_2} \right\rangle \\
 &= \left\langle \mathbf{0}, \tilde{\mathbf{a}}_{n_2} \right\rangle = 0.
 \end{aligned}
 \tag{17}$$

It is clear that  $\mathcal{C}_2 = 0$  for SSPs, while  $\mathcal{C}_2 > 0$  for MSPs in ideal conditions.

However, practically noisy environments with estimated  $\tilde{\mathbf{A}}$  show great disparity from the above theoretical analysis. Figure 2 presents noisy mixture TF vectors  $\mathbf{S}_x(t, f)$  of SEPs (gray asterisks) detected by (3); the red plus signs and black circles are related to the real parts of columns in  $\mathbf{A}$  and HC-based  $\tilde{\mathbf{A}}$ , respectively. We select four samples of SEPs, including one HCP-SSP, one LCP-SSP, and two MSPs, denoted by the blue asterisks in Figure 2. First, note that HCP-SSP1, which approaches one of steering vectors in  $\mathbf{A}$ , can be accurately identified, as one of columns in  $\tilde{\mathbf{A}}$ , i.e.,  $\tilde{\mathbf{a}}_{n_1}$  in (14), is close to the ideal steering vector. Unlike HCP-SSP1, LCP-SSP1 is located far from the ideal steering vectors, either due to strong noise or weak source energy, although it falls within the range of SSPs. For this LCP-SSP1,  $\tilde{\mathbf{a}}_{n_1}$  and  $\tilde{\mathbf{a}}_{n_2}$  in (14) are the two most relevant columns (the closest black circles) in  $\tilde{\mathbf{A}}$ , as indicated by the dotted lines in Figure 2. In this case, the values of  $\mathcal{C}_1$  and  $\mathcal{C}_2$  are both greater than zero. In addition, MSP1 is a typical multi-source TF point; similar to LCP-SSP1, its two bases in (14) correspond to the two black circles closest to it. Nevertheless, MSP2 is a special MSP located very close to one of the black circles, resulting from the randomness of noise and cluster distribution. According to the previous analysis, the  $\mathcal{C}_2$  of MSP2 is near zero, and MSP2 has a high probability of being erroneously identified as an SSP. Thus, based on the above discussion, we can make the following remarks.



**Figure 2.** 2D plot depicting real parts of the observed mixture TF vectors of the SEPs (denoted by gray asterisks) in  $\Omega_{\text{sep}}$  with  $M = 2, N = 4, N_0 = 10, T_0 = 0.2$ , and  $\text{SNR} = 0$  dB. The blue asterisks indicate selected samples of MSP, HCP-SSP, and LCP-SSP, the red plus signs correspond to ideal steering vectors of  $\mathbf{A}$ , and the black circles are related to columns of  $\tilde{\mathbf{A}}$  obtained using the  $k$ -means method).

**Remark 1** (HCP-SSP). The base  $\tilde{\mathbf{a}}_{n_1}$  in (14) comes from the closest column in  $\tilde{\mathbf{A}}$ , and  $C_1 > 0$ . The base  $\tilde{\mathbf{a}}_{n_2}$  in (14) corresponds to the second closest black circle to HCP-SSP; as such, it can be another column of  $\tilde{\mathbf{A}}$ , and  $C_2 \approx 0$ .

**Remark 2** (LCP-SSP). The two bases  $\tilde{\mathbf{a}}_{n_1}$  and  $\tilde{\mathbf{a}}_{n_2}$  in (14) correspond to the two black circles closest to LCP-SSP; as such, they can be any two columns of  $\tilde{\mathbf{A}}$ , and  $C_1 \geq C_2 > 0$ .

**Remark 3** (MSP). The two bases  $\tilde{\mathbf{a}}_{n_1}$  and  $\tilde{\mathbf{a}}_{n_2}$  in (14) correspond to the two black circles closest to MSP; as such, they can be any two columns of  $\tilde{\mathbf{A}}$ , and  $C_1 \geq C_2 > 0$ . Occasionally,  $C_2 \approx 0$ .

In view of the above remarks, we propose using the ratio of  $C_1$  to  $C_2$  as the detection criterion for HCP-SSPs:

$$T_{\text{hcp}} = \frac{C_1}{C_2} = \frac{\langle \mathbf{S}_x(t, f), \tilde{\mathbf{a}}_{n_1} \rangle}{\langle \mathbf{S}_x(t, f) - \tilde{\mathbf{a}}_{n_1} \tilde{\mathbf{a}}_{n_1}^+ \mathbf{S}_x(t, f), \tilde{\mathbf{a}}_{n_2} \rangle} \quad (18)$$

and identifying a set of HCP-SSPs by assigning a proper threshold  $\beta_{\text{hcp}}$ , i.e.,

$$\text{If } T_{\text{hcp}}(\tilde{\mathbf{a}}_{n_1}, \tilde{\mathbf{a}}_{n_2}) > \beta_{\text{hcp}}, \text{ then } (t, f) \in \Omega_{\text{hcp}}. \quad (19)$$

All the SEPs in  $\Omega_{\text{sep}}$  that satisfy (19) would then be identified as HCP-SSPs and included in the  $\Omega_{\text{hcp}}$  mentioned in (6).

Finally, we apply the  $k$ -means clustering method to classify the HCP-SSPs in  $\Omega_{\text{hcp}}$  into  $N$  clusters; each steering vector of  $\mathbf{A}$  is estimated by averaging all of the mixture TF vectors belonging to the corresponding cluster, as explained earlier in (6). The proposed HCP-SSP-MME method is summarized in **Algorithm 1**.

---

#### Algorithm 1 Proposed HCP-SSP-MME Method

---

**Input:**  $T_0$ , and  $\beta_{\text{hcp}}$ .

- 1: Detect a group of SEPs by (3), and obtain the set  $\Omega_{\text{sep}}$ ;
- 2: Use the  $k$ -means clustering method to classify all the SEPs in  $\Omega_{\text{sep}}$ , and estimate an initial dictionary  $\tilde{\mathbf{A}}$  by (7);
- 3: For each SEP in  $\Omega_{\text{sep}}$ , compute  $T_{\text{hcp}}$  by (18), where  $\tilde{\mathbf{a}}_{n_1}$  and  $\tilde{\mathbf{a}}_{n_2}$  are determined by (9) and (16); Identify a group of HCP-SSPs by (19), and obtain the set  $\Omega_{\text{hcp}}$ ;
- 4: Classify the HCP-SSPs in  $\Omega_{\text{hcp}}$  via  $k$ -means clustering; Each column of estimated  $\hat{\mathbf{A}}$  is computed by (6), and the number of retained clusters gives an estimated  $\hat{N}$ .

**Output:** Estimated  $\hat{\mathbf{A}}$ .

---

#### 2.2. Source Separation with Sparse Recovery Model

Based on the estimated  $\hat{\mathbf{A}}$  and  $\hat{N}$ , we next delineate the process of source separation, i.e., recovering  $\mathbf{s}(t)$  in (1). Existing UBSS algorithms have demonstrated that investigating the sparsity in different ways plays a significant role in separating observed mixtures, which motivates us to exploit the sparsity of ASPs. As STFT is used to transform time domain mixtures into the TF domain, the objective of the proposed ASTFS-UBSS algorithm in the stage of source separation is to estimate TF vectors of each source, i.e.,  $\mathbf{S}_s(t, f)$  in (2), based on which the time domain sources  $\mathbf{s}(t)$  can be recovered by inverse STFT (ISTFT) [52]. Note that there is no need to estimate  $\mathbf{S}_s(t, f)$  in the entire TF domain; we only deal with ASPs where at least one active source exists, thereby relieving the computational burden.

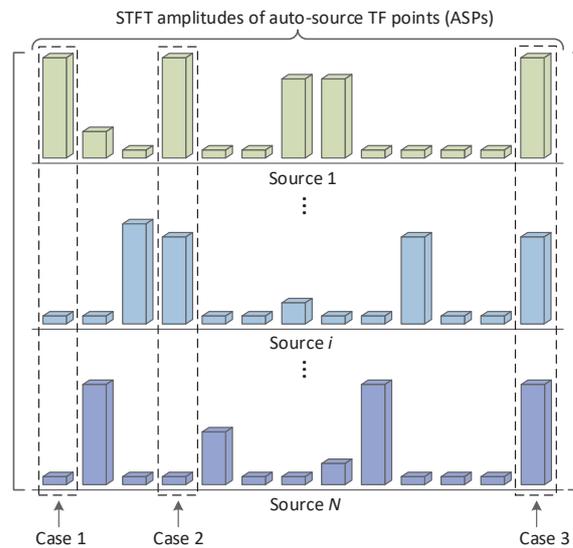
Because STFT is free of cross-terms, ideal ASPs are essentially the TF points with nonzero STFT values in the noiseless case. In actual situation, ASPs are detected by

assigning an energy threshold in the STFT domain. Specifically, we can select a set of ASPs by applying a noise thresholding procedure in the same way as (3):

$$\text{if } \frac{\|\mathbf{S}_x(t, f)\|}{\max_{\tau, v}\{\|\mathbf{S}_x(\tau, v)\|\}} > T_1, \text{ then } (t, f) \in \Omega_{\text{asp}} \quad (20)$$

where  $T_1$  is a small threshold value (empirically  $T_1 = 0.03$ ) used to guarantee accurate selection of ASPs. The difference from  $T_0$  in (3) should be noted; it is relatively larger than  $T_1$ , as the estimate of  $\mathbf{A}$  prefers ASPs with strong energy. All the TF points which satisfy the above criterion are included in the set  $\Omega_{\text{asp}}$ . It can be concluded that the different sets obtained thus far satisfy the relationship:  $\Omega_{\text{asp}} \supset \Omega_{\text{sep}} \supset \Omega_{\text{hcp}}$ .

In the literature [18,41], the number of active sources  $N_{\text{tf}}$  at each ASP is generally required to be less than or equal to the number of sensors, i.e.,  $N_{\text{tf}} \leq M$  (summarized in Table 1). Clearly, this assumption significantly limits the application of UBSS, especially when  $M = 2$ . Figure 3 illustrates the source TF amplitudes of a set of ASPs, which is a standard sparse matrix with the rows corresponding to STFT amplitudes of  $N$  sources and columns representing different ASPs in the STFT domain. Based on the ASPs from Figure 3, we can make the following remarks.



**Figure 3.** TF sparsity illustration of  $|\mathbf{S}_s(t, f)|$  in (2) for different ASPs. *Case 1:* SSP with a dominant source. *Case 2:* MSP with a limited number of sources. *Case 3:* MSP with  $N$  sources.

**Remark 4 (Case 1).** SSPs with dominant energy contributed by one single source. These SSPs are randomly distributed in the STFT domain at either the TF-point level or the TF-zone level [53].

**Remark 5 (Case 2).** MSPs with energy mainly contributed by a limited number of sources, i.e.,  $1 < N_{\text{tf}} < N$ . This kind of MSPs frequently emerges in a variety of applications (e.g., speech signals) due to the randomness of the TF superposition.

**Remark 6 (Case 3).** MSPs with energy contributed by almost all  $N$  sources, i.e.,  $N_{\text{tf}} \approx N$ . These MSPs rarely occur; however, they may be significant in certain applications, e.g., image and multipath signals.

Based on the above cases, it is reasonable to assume that the energy at most of the ASPs mainly comes from a limited number of sources, while allowing a handful of ASPs where almost  $N$  sources are active. Therefore, we define the following assumption for the proposed ASTFS-UBSS algorithm.

**Assumption 3.** *The majority of the ASPs detected in (20) belong to Case 1 and Case 2, i.e., only a limited number of sources contribute their energy at these ASPs. Meanwhile, we allow the existence of a minority of ASPs satisfying Case 3. Under these circumstances, the requirement on the number of active sources  $N_{tf}$  relative to the number of sensors  $M$  at the ASPs can be greatly relaxed.*

With **Assumption 3**, the 2D TF structure in Figure 3 would be a typical sparse matrix. It is worth stressing that the constraints required by [18,41] in Table 1 are strictly imposed on every ASP in order to guarantee the sparsity; however, in general there are always some ASPs that do not meet these constraints, as analyzed in Figure 3. On the contrary, the conditions in **Assumption 3** are imposed on all the ASPs from an overall perspective; that is, for a small number of ASPs where  $N_{tf} > M$  becomes feasible, as long as most of ASP vectors are sparse we are still able to guarantee the sparse solvability of the 2D matrix.

In order to successfully recover the sources, these 2D sparse matrix data are treated by vectorizing them to fit the compressed sensing model, i.e., a 1D sparse vector is formed by connecting all the rows of the 2D matrix. The 1D vector generated in this way can be recovered by implementing current sparse recovery tools. Therefore, the essence of the proposed ASTFS-UBSS algorithm lies in its reformulation of the UBSS problem as a sparse signal recovery problem involving estimating the STFT value matrix  $\mathbf{S}_s(t, f)$  in (2).

For convenience of derivation, the noise term is ignored for the ASPs in  $\Omega_{asp}$ ; thus, the expression in (2) degrades into

$$\mathbf{S}_x(t, f) = \widehat{\mathbf{A}}\mathbf{S}_s(t, f) = \widehat{\mathbf{A}} \begin{bmatrix} S_s^{(1)}(t, f) \\ \vdots \\ S_s^{(N)}(t, f) \end{bmatrix}, \quad (t, f) \in \Omega_{asp}. \quad (21)$$

Letting  $\mathbf{Z}$  and  $\mathbf{Y}$  represent the combinations of  $\mathbf{S}_x(t, f)$  and  $\mathbf{S}_s(t, f)$  at all the ASPs in  $\Omega_{asp}$ , respectively, we then obtain the following simplified expression:

$$\mathbf{Z} = \widehat{\mathbf{A}}\mathbf{Y} \quad (22)$$

and

$$\begin{cases} \mathbf{Z}^{M \times N_{asp}} = \begin{bmatrix} S_x^{(1)}(t_1, f_1) & \cdots & S_x^{(1)}(t_{N_{asp}}, f_{N_{asp}}) \\ S_x^{(2)}(t_1, f_1) & \cdots & S_x^{(2)}(t_{N_{asp}}, f_{N_{asp}}) \\ \vdots & \ddots & \vdots \\ S_x^{(M)}(t_1, f_1) & \cdots & S_x^{(M)}(t_{N_{asp}}, f_{N_{asp}}) \end{bmatrix} \\ \mathbf{Y}^{N \times N_{asp}} = \begin{bmatrix} S_s^{(1)}(t_1, f_1) & \cdots & S_s^{(1)}(t_{N_{asp}}, f_{N_{asp}}) \\ S_s^{(2)}(t_1, f_1) & \cdots & S_s^{(2)}(t_{N_{asp}}, f_{N_{asp}}) \\ \vdots & \ddots & \vdots \\ S_s^{(N)}(t_1, f_1) & \cdots & S_s^{(N)}(t_{N_{asp}}, f_{N_{asp}}) \end{bmatrix}, \end{cases}$$

where  $N_{asp}$  is the number of ASPs in the set  $\Omega_{asp}$ . The  $M \times N_{asp}$  matrix  $\mathbf{Z}$  is composed of observed mixtures in the STFT domain. The  $N \times N_{asp}$  matrix  $\mathbf{Y}$  is actually the sparse structure illustrated in Figure 3, where *Case 1* and *Case 2* occur at most of the ASPs and *Case 3* is relatively scarce, which guarantees the sparsity of  $\mathbf{Y}$ .

As explained before, we can treat  $\mathbf{Z}$  and  $\mathbf{Y}$  by vectorizing them into two 1D vectors, which is expressed as follows:

$$\begin{aligned} \mathbf{z} &= \text{vec}\{\mathbf{Z}\} = \text{vec}\{\widehat{\mathbf{A}}\mathbf{Y}\} \\ &= \mathbf{\Psi}\text{vec}\{\mathbf{Y}\} \\ &= \mathbf{\Psi}\mathbf{y} \end{aligned} \quad (23)$$

and

$$\left\{ \begin{aligned} \mathbf{\Psi} &= \widehat{\mathbf{A}} \otimes \mathbf{I}^{N_{\text{asp}} \times N_{\text{asp}}} = \begin{bmatrix} \Delta_{11} & \Delta_{12} & \cdots & \Delta_{1N} \\ \Delta_{21} & \Delta_{22} & \cdots & \Delta_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \Delta_{M1} & \Delta_{M2} & \cdots & \Delta_{MN} \end{bmatrix} \\ \mathbf{z} &= \begin{bmatrix} S_x^{(1)}(t_1, f_1) \cdots S_x^{(1)}(t_{N_{\text{asp}}}, f_{N_{\text{asp}}}) S_x^{(2)}(t_1, f_1) \\ \cdots S_x^{(M)}(t_1, f_1) \cdots S_x^{(M)}(t_{N_{\text{asp}}}, f_{N_{\text{asp}}}) \end{bmatrix}^T \\ \mathbf{y} &= \begin{bmatrix} S_s^{(1)}(t_1, f_1) \cdots S_s^{(1)}(t_{N_{\text{asp}}}, f_{N_{\text{asp}}}) S_s^{(2)}(t_1, f_1) \\ \cdots S_s^{(N)}(t_1, f_1) \cdots S_s^{(N)}(t_{N_{\text{asp}}}, f_{N_{\text{asp}}}) \end{bmatrix}^T, \end{aligned} \right.$$

where  $\text{vec}\{\cdot\}$  means the vectorization operation that stacks the rows of the corresponding matrix to form a long vector, ' $\otimes$ ' denotes the tensor (Kronecker) product,  $\mathbf{I}^{N_{\text{asp}} \times N_{\text{asp}}}$  denotes an  $N_{\text{asp}} \times N_{\text{asp}}$  identity matrix,  $\Delta_{mn}$ ,  $m = 1, \dots, M$  and  $n = 1, \dots, N$ , and denotes an  $N_{\text{asp}} \times N_{\text{asp}}$  diagonal matrix with its diagonal elements equal to  $\widehat{a}_{mn}$  in  $\widehat{\mathbf{A}}$ .

The UBSS model in (23) is equivalent to a sparse signal recovery model, where  $\mathbf{\Psi} \in \mathbb{R}^{MN_{\text{asp}} \times NN_{\text{asp}}}$  ( $M < N$ ) can be regarded as a basis or dictionary matrix and  $\mathbf{z} \in \mathbb{R}^{MN_{\text{asp}} \times 1}$  is the available measurement vector. The task of estimating STFT values can be accomplished by recovering the underlying sparse vector  $\mathbf{y} \in \mathbb{R}^{NN_{\text{asp}} \times 1}$ . The indeterminacy of the solution in (23) is eliminated using the sparse prior information of  $\mathbf{y}$ . Unlike the commonly used approach in which the dictionary matrix is randomly generated, the dictionary  $\mathbf{\Psi}$  in our work is a very sparse matrix formulated from the elements of the estimated  $\widehat{\mathbf{A}}$ .

When attempting to find sparse solutions in (23), the original approach is to solve the following  $l_0$ -norm minimization problem (P<sub>0</sub>):

$$(P_0) : \min \|\mathbf{y}\|_0 \quad \text{s.t. } \mathbf{z} = \mathbf{\Psi}\mathbf{y} \tag{24}$$

where  $\|\cdot\|_0$  denotes the  $l_0$ -norm operator. It is revealed that if  $\mathbf{y}$  is sufficiently sparse, the solution in (24) is equal to the solution of the  $l_1$ -norm minimization problem (P<sub>1</sub>) [54]:

$$(P_1) : \min \|\mathbf{y}\|_1 \quad \text{s.t. } \mathbf{z} = \mathbf{\Psi}\mathbf{y} \tag{25}$$

where  $\|\cdot\|_1$  denotes the  $l_1$ -norm operator. Because the practical observation is always contaminated by noise, the traditional equality constraint in (25) is relaxed by introducing a noise-aware variant. Thus, our task of estimating the STFT values of the sources is realized by solving the following optimization problem (P<sub>2</sub>):

$$(P_2) : \min_{\mathbf{y}} \left\{ \frac{1}{2} \|\mathbf{z} - \mathbf{\Psi}\mathbf{y}\|_2^2 + \lambda \|\mathbf{y}\|_1 \right\} \tag{26}$$

where  $\lambda \geq 0$  is a regularization parameter and its value governs the sparsity of the solution.

For the sparsity-regularized problem in (26), its solutions have been widely available from the wealth of knowledge in the literature. It should be noted that although choosing ASPs for source recovery reduces computational burden to an extent, the model in (23) is still a large-scale underdetermined problem. Hence, we solve (26) by applying the spectral projected gradient for  $l_1$  minimization (SPGL1) [55] due to its efficiency for large-scale problem and its suitability for complex-valued domains. Finally, each time domain source is recovered using ISTFT based on the estimated STFT values in  $\mathbf{y}$ . The overall procedure of the proposed ASTFS-UBSS algorithm is presented in **Algorithm 2**.

---

**Algorithm 2** The Proposed ASTFS-UBSS Algorithm
 

---

**Input:**  $N_0$ ,  $T_0$ ,  $T_1$ , and  $\beta_{\text{hcp}}$ .

- 1: Estimate the mixing matrix  $\mathbf{A}$  by **Algorithm 1**;
- 2: Detect a set of ASPs by (20), and obtain the set  $\Omega_{\text{asp}}$ ;
- 3: Establish the sparse recovery model by (23);
- 4: Reconstruct  $\mathbf{S}_s(t, f)$  by (26) via SPGL1;
- 5: Recover time domain  $\mathbf{s}(t)$  via ISTFT.

**Output:** Recovered  $\hat{\mathbf{s}}(t)$ .
 

---

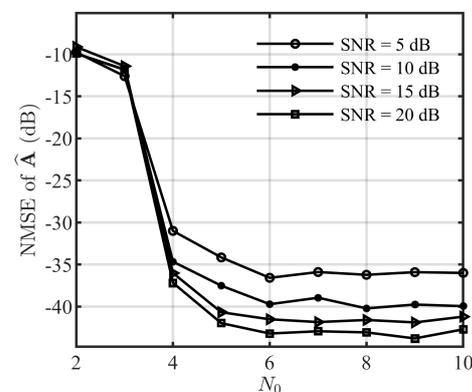
### 3. Experimental Results

In this section, we evaluate the proposed ASTFS-UBSS algorithm on synthetic data. We consider a uniform linear array (ULA) with its sensors separated by a half-wavelength spacing, and use  $M = 2$  sensors; the estimation accuracies of the mixing matrix and separated sources are both assessed via the normalized mean square error (NMSE). In addition, the quality of the separated sources is measured by the evaluation criteria reported in [56], termed the signal-to-distortion ratio (SDR). Time domain sources are simulated using speech signals randomly chosen from the TIMIT Corpus [57], and different sources are overlapped in the STFT domain. The duration of each source is 3 s, the sampling rate is 16 kHz, the window length of STFT is 1024 samples, and the overlapping size is 256 samples.

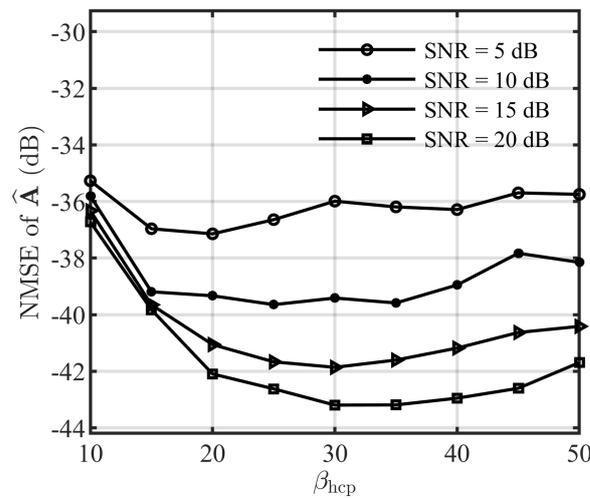
In the following, we first demonstrate the influence of  $N_0$  and  $\beta_{\text{hcp}}$  on the performance of the proposed HCP-SSP-MME, and thereby determine the optimal parameter setting in **Algorithm 1**. Second, various MME methods are compared and evaluated in terms of their NMSE performance. Lastly, performance comparisons with different evaluation metrics are presented to validate the effectiveness of our proposed ASTFS-UBSS in **Algorithm 2**.

#### 3.1. Optimization of Parameter Settings

We separately discuss the optimal selection of  $N_0$  and  $\beta_{\text{hcp}}$  for the mixing matrix  $\mathbf{A}$  generated in [18,21], where we assume that  $N = 4$  sources come from different directions of arrival (DOA).  $N_0$  is the first determined to achieve optimal estimation of  $\hat{\mathbf{A}}$ . Figure 4 presents the NMSE results of estimated  $\hat{\mathbf{A}}$  under different SNR situations. It can be seen that the proposed HCP-SSP-MSE achieves significant performance under both low and high SNRs. Figure 4 shows that the proposed HCP-SSP-MME performs better when  $N_0$  reaches 6, then improves extremely slightly with continuously increasing  $N_0$ . Thus, we set  $N_0 = 6$  for MME, as a larger value of  $N_0$  leads to high computational complexity. Similar experiments were carried out to determine the optimal threshold  $\beta_{\text{hcp}}$ , as exhibited in Figure 5. It can be noted that the proposed HCP-SSP-MME performs best when the value of  $\beta_{\text{hcp}}$  equals 30. It was further verified that the above settings are suitable for the cases involving different number of sources, e.g.,  $N = 3$  and  $N = 5$ .



**Figure 4.** NMSEs of estimated  $\hat{\mathbf{A}}$  with  $M = 2$ ,  $N = 4$ , and  $T_0 = 0.2$  under different SNR levels versus different values of  $N_0$ .



**Figure 5.** NMSEs of estimated  $\hat{\mathbf{A}}$  with  $M = 2$ ,  $N = 4$ , and  $T_0 = 0.2$  under different SNR levels versus different values of  $\beta_{\text{hcp}}$ .

### 3.2. Evaluation of Mixing Matrix Estimation (MME)

We evaluated the estimation of the mixing matrix  $\mathbf{A}$  generated using the approach from [18,21]. Each source was assigned a DOA randomly picked from the set  $\{15^\circ, 30^\circ, 45^\circ, 60^\circ, 75^\circ\}$  over every trial. Table 2 shows the performance comparison of estimated  $\hat{\mathbf{A}}$  in terms of NMSE using ET-MME [18], ED-MME [21], and the proposed HCP-SSP-MME under different noise scenarios. Clearly, the proposed HCP-SSP-MME method outperforms the others within the SNR range from 0 dB to 20 dB. Compared with ET-MME, the superiority of our method becomes more outstanding when  $\text{SNR} > 10$  dB. This performance gain is attributed to the detected HCP-SSP locations, which are more beneficial for clustering. Moreover, ED-MME performs poorly, and its NMSE results fluctuate without regularity when SNR changes. The reason for this lies in ED-MME requiring the number of sensors  $M > 2$  at each ASP in order to decompose the STFT energy into two parts. Hence, it is not suitable for two-sensor arrays.

**Table 2.** Comparison of mixing matrix estimated via different methods.

$N$	SNR (dB)	NMSEs of $\hat{\mathbf{A}}$ (dB)		
		ET-MME [18]	ED-MME [21]	HCP-SSP-MME
3	20	-34.6	-12.6	<b>-44.9</b>
	10	-34.2	-13.6	<b>-42.1</b>
	0	-32.5	-14.7	<b>-34.5</b>
4	20	-29.9	-10.3	<b>-41.5</b>
	10	-29.2	-10.9	<b>-38.9</b>
	0	-28.8	-11.0	<b>-31.5</b>
5	20	-21.2	-8.2	<b>-37.4</b>
	10	-20.9	-9.1	<b>-34.8</b>
	0	-20.6	-9.4	<b>-28.7</b>

The bold values indicate the best performance.

### 3.3. Evaluation of Source Separation

In this subsection, we evaluate the source recovery performance of the proposed ASTFS-UBSS in Algorithm 2 in comparison to the UBSS algorithm in [18]. Similarly, we separately analyze the separation of original sources utilizing the estimated mixing matrices. The remaining parameters are set the same as in previous experiments.

In addition to the gain obtained from the  $\hat{\mathbf{A}}$  estimated by **Algorithm 1**, it is necessary to validate the benefit from the proposed  $l_1$ -norm sparse representation (L1-SR) in (26). To achieve this, we only compare the second-stage source recovery performance of different algorithms provided by the known ideal mixing matrix. The proposed L1-SR is implemented in comparison with SP-SR [18] based on subspace projection. In Table 3, ‘SP-UBSS<sub>I</sub>’ and ‘ASTFS<sub>I</sub>’ mean that the SP-SR and L1-SR methods are respectively implemented to recover sources with the known actual mixing matrix  $\mathbf{A}$ . The results of ‘SP-UBSS<sub>I</sub>’ and ‘ASTFS<sub>I</sub>’ provide the performance comparison of the source recovery methods in terms of NMSE and SDR. It can be clearly seen from Table 3 that ‘ASTFS<sub>I</sub>’ outperforms ‘SP-UBSS<sub>I</sub>’ under different SNRs. Note that the performance of SP-SR is limited when  $N$  sources are highly overlapped in the TF domain, which is due to the constraint that the number of active sources at each ASP should be less than the number of sensors  $M$ . This sparse constraint is relaxed by the sparse recovery model in (26); therefore, L1-SR has an obvious advantage for separating sources when  $M = 2$ .

**Table 3.** Comparison of source separation via different methods.

N	SNR (dB)	NMSE/SDR of $\hat{\mathbf{s}}$ (dB)			
		SP-UBSS [18]	SP-UBSS <sub>I</sub>	ASTFS	ASTFS <sub>I</sub>
3	20	−5.8/4.2	−6.1/5.1	<b>−8.1/7.8</b>	−8.2/8.6
	15	−5.6/3.8	−5.9/4.8	<b>−7.9/7.6</b>	−7.9/7.6
	10	−5.6/4.2	−5.7/4.5	<b>−7.4/7.0</b>	−7.4/7.7
4	20	−4.5/1.9	−4.8/3.1	<b>−5.9/5.0</b>	6.0/5.0
	15	−4.3/2.3	−4.6/2.9	<b>−5.8/4.9</b>	−5.8/4.9
	10	−4.3/1.9	−4.5/2.6	<b>−5.5/4.3</b>	−5.6/4.4
5	20	−3.0/−1.9	−3.4/0.8	<b>−4.3/2.1</b>	−4.3/2.4
	15	−2.9/−2.2	−3.1/0.6	<b>−4.0/1.8</b>	−4.1/2.2
	10	−2.8/−2.2	−3.1/0.2	<b>−3.8/1.3</b>	−3.9/1.8

The bold values indicate the best performance.

We now evaluate the total separation performance of the proposed ASTFS-UBSS in **Algorithm 2** compared with SP-UBSS [18] in terms of NMSE and SDR. Table 3 shows the overall separation performance of the above algorithms with different numbers of sources under various levels of SNR by exploiting the mixing matrix. It can be seen intuitively that all of the algorithms show a declining trend with the increase in the number of sources and the decrease in SNR, as expected. The results show that the proposed ASTFS-UBSS achieves superior performance over the others across different SNRs. In addition, the performance of the proposed ASTFS-UBSS is close to that of the ideal ASTFS<sub>I</sub>, which verifies the accuracy of mixing matrix estimation utilizing the proposed HCP-SSP-MME method. As analyzed before, the overall performance improvement of our algorithm stems from the accurate estimation of  $\mathbf{A}$  as well as from the model proposed in (26), which can be seen from the comparison among ASTFS<sub>I</sub> and SP-UBSS<sub>I</sub>.

### 3.4. Experiments with Real-Valued Mixing Matrices

The experiments above were all based on complex-valued mixing matrices. To further demonstrate the effectiveness of the proposed algorithm, we conducted source recovery experiments using real-valued mixing matrices and compared them with baseline models. The real-valued mixing matrix is provided by [31]

$$\mathbf{A} = \begin{bmatrix} 0.2588 & 0.8192 & 0.9962 & 0.7071 & 0.9962 \\ -0.9659 & -0.5736 & 0.0872 & 0.7071 & 0.0872 \end{bmatrix}. \quad (27)$$

The experimental results are presented in Table 4. Comparing the results presented in Table 4 with those in Table 3 reveals a clear trend: the separation performance achieved with real-valued mixing matrices consistently surpasses that attained with complex-valued

mixing matrices across various scenarios involving different numbers of sources and different signal-to-noise ratios. Notably, our proposed algorithm consistently outperforms the baseline models under the settings involving real-valued mixing matrices. This observation is in line with the trends observed in the experiments conducted with complex-valued mixing matrices.

**Table 4.** Source separation with real-valued mixing matrices.

N	SNR (dB)	NMSE/SDR of $\hat{s}$ (dB)			
		SP-UBSS [18]	SP-UBSS <sub>I</sub>	ASTFS	ASTFS <sub>I</sub>
3	20	−5.8/4.2	−6.8/6.0	<b>−9.1/8.7</b>	−9.1/8.8
	15	−5.8/4.0	−6.7/5.9	<b>−9.0/8.7</b>	−9.0/8.7
	10	−5.6/3.7	−6.7/5.8	<b>−8.7/8.3</b>	−8.8/8.4
4	20	−5.5/4.1	−5.6/4.2	<b>−7.4/6.6</b>	−7.4/6.6
	15	−5.5/4.1	−5.6/4.2	<b>−7.4/6.5</b>	−7.4/6.6
	10	−5.4/4.0	−5.5/4.1	<b>−7.2/6.4</b>	−7.2/6.4
5	20	−4.1/2.0	−4.1/2.0	<b>−5.6/4.2</b>	−5.6/4.4
	15	−4.1/1.9	−4.1/2.0	<b>−5.6/4.2</b>	−5.6/4.4
	10	−4.0/1.6	−4.0/1.8	<b>−5.5/4.1</b>	−5.5/4.1

The bold values indicate the best performance.

#### 4. Conclusions

In this paper, the UBSS problem is addressed with a low-cost dual-sensor array under the framework of sparse recovery. The contribution of the proposed ASTFS-UBSS algorithm is twofold. First, an efficient MME method is designed to accurately locate HCP-SSPs, which have better clustering property, thereby significantly improving the mixing matrix estimation accuracy. From the view of practical applications, the proposed HCP-SSP-MME is verified to be effective for mixing matrices. Second, we build a novel sparse recovery model by exploiting the TF sparsity of ASPs. Currently available sparse recovery tools can be applied to obtain the sparse solution. More importantly, the sparse constraint is greatly relaxed for active sources at ASPs, allowing more robust separation of  $N > 2$  sources when  $M = 2$ . Comparative results with state-of-the-art algorithms validate the above contribution of the proposed ASTFS-UBSS algorithm and show its obvious advantage when the number of sources increases. Notably, our proposed ASTFS-UBSS remains based on the assumption of linear mixing of signals; however, various scenarios such as convolutive mixing, nonlinear mixing, and similar instances present heightened complexities in blind source separation tasks. Consequently, we will consider separation under more complex scenarios in future work to enhance the applicability of our algorithm.

**Author Contributions:** Conceptualization, H.Z. and S.S.; methodology, H.Z.; validation, J.C.; formal analysis, J.C. and S.S.; software, J.C. and S.S.; writing—original draft, J.C.; writing—review and editing, J.C. and H.Z.; visualization, J.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** Natural Science Foundation of Hubei Province (2022CFB084).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data is contained within the article.

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

## Abbreviations

The following abbreviations are used in this manuscript:

BSS	blind source separation
UBSS	underdetermined blind source separation
TF	time–frequency
ASP	auto-source time–frequency point
SSP	single-source time–frequency point
MSP	multi-source time–frequency point
HCP-SSP	high clustering property SSP
LCP-SSP	low clustering property SSP
SEP	strong energy TF point
MP	matching pursuit
OMP	orthogonal matching pursuit
MME	mixing matrix estimation
CS	compressed sensing
ICA	independent component analysis
SCA	sparse component analysis
SBCA	sparse bounded component analysis
NME	non-negative matrix estimation
ASTFS-UBSS	auto-source TF sparsity-based UBSS algorithm
HCP-SSP-MME	high clustering property SSP-based MME
STFT	short-time Fourier transform
ISTFT	inverse short-time Fourier transform
ULA	uniform linear array
NMSE	normalized mean square error
SDR	signal-to-distortion ratio
DOA	direction of arrival
L1-SR	$l_1$ -norm sparse representation

## References

1. Ansari, S.; Alatrany, A.S.; Alnajjar, K.A.; Khater, T.; Mahmoud, S.; Al-Jumeily, D.; Hussain, A.J. A survey of artificial intelligence approaches in blind source separation. *Neurocomputing* **2023**, *561*, 126895. [[CrossRef](#)]
2. Li, M.; Chang, Z.; Zhang, L.; Xu, H.; Luo, Z.; Guo, R. Blind separation for wireless communication convolutive mixtures based on denoising iva. *IEEE Access* **2022**, *10*, 113756–113766. [[CrossRef](#)]
3. Alaghbari, K.A.; Lim, H.S.; Jin, B.; Shen, Y. Source Separation in Joint Communication and Radar Systems Based on Unsupervised Variational Autoencoder. *IEEE Open J. Veh. Technol.* **2023**, *5*, 56–70. [[CrossRef](#)]
4. Xie, K.; Zhou, G.; Yang, J.; He, Z.; Xie, S. Eliminating the Permutation Ambiguity of Convolutional Blind Source Separation by Using Coupled Frequency Bins. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *31*, 589–599. [[CrossRef](#)] [[PubMed](#)]
5. Ikeshita, R.; Nakatani, T. Independent vector extraction for fast joint blind source separation and dereverberation. *IEEE Signal Process. Lett.* **2021**, *28*, 972–976. [[CrossRef](#)]
6. Du, B.; Wang, S.; Xu, C.; Wang, N.; Zhang, L.; Tao, D. Multi-Task Learning for Blind Source Separation. *IEEE Trans. Image Process.* **2018**, *27*, 4219–4231. [[CrossRef](#)] [[PubMed](#)]
7. Xie, Y.; Xie, K.; Xie, S. Underdetermined Blind Source Separation for Heart Sound Using Higher-Order Statistics and Sparse Representation. *IEEE Access* **2019**, *7*, 87606–87616. [[CrossRef](#)]
8. Song, R.; Zhang, S.; Cheng, J.; Li, C.; Chen, X. New insights on super-high resolution for video-based heart rate estimation with a semi-blind source separation method. *Comput. Biol. Med.* **2020**, *116*, 103535. [[CrossRef](#)] [[PubMed](#)]
9. Naik, G.R.; Kumar, D.K. An overview of independent component analysis and its applications. *Informatica* **2011**, *35*, 63–81.
10. Tharwat, A. Independent component analysis: An introduction. *Appl. Comput. Inform.* **2021**, *17*, 222–249. [[CrossRef](#)]
11. Georgiev, P.; Theis, F.; Cichocki, A. Sparse component analysis and blind source separation of underdetermined mixtures. *IEEE Trans. Neural Netw.* **2005**, *16*, 992–996. [[CrossRef](#)] [[PubMed](#)]
12. Erichson, N.B.; Zheng, P.; Manohar, K.; Brunton, S.L.; Kutz, J.N.; Aravkin, A.Y. Sparse principal component analysis via variable projection. *Siam J. Appl. Math.* **2020**, *80*, 977–1002. [[CrossRef](#)]
13. Babatas, E.; Erdogan, A.T. Time and frequency based sparse bounded component analysis algorithms for convolutional mixtures. *Signal Process.* **2020**, *173*, 107590. [[CrossRef](#)]
14. Gan, J.; Liu, T.; Li, L.; Zhang, J. Non-negative matrix factorization: a survey. *Comput. J.* **2021**, *64*, 1080–1092. [[CrossRef](#)]
15. Fathi Hafshejani, S.; Moabferd, Z. Initialization for non-negative matrix factorization: a comprehensive review. *Int. J. Data Sci. Anal.* **2023**, *16*, 119–134. [[CrossRef](#)]

16. Wang, D.; Chen, J. Supervised Speech Separation Based on Deep Learning: An Overview. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2018**, *26*, 1702–1726. [[CrossRef](#)] [[PubMed](#)]
17. Agrawal, J.; Gupta, M.; Garg, H. A review on speech separation in cocktail party environment: challenges and approaches. *Multimed. Tools Appl.* **2023**, *82*, 1–33. [[CrossRef](#)]
18. Aissa-El-Bey, A.; Linh-Trung, N.; Abed-Meraim, K.; Belouchrani, A.; Grenier, Y. Underdetermined Blind Separation of Nondisjoint Sources in the Time-Frequency Domain. *IEEE Trans. Signal Process.* **2007**, *55*, 897–907. [[CrossRef](#)]
19. Belouchrani, A.; Amin, M.G.; Thirion-Moreau, N.; Zhang, Y.D. Source Separation and Localization Using Time-Frequency Distributions: An Overview. *IEEE Signal Process. Mag.* **2013**, *30*, 97–107. [[CrossRef](#)]
20. Xie, S.; Yang, L.; Yang, J.; Zhou, G.; Xiang, Y. Time-Frequency Approach to Underdetermined Blind Source Separation. *IEEE Trans. Neural Netw. Learn. Syst.* **2012**, *23*, 306–316.
21. Zhang, H.; Hua, G.; Yu, L.; Cai, Y.; Bi, G. Underdetermined blind separation of overlapped speech mixtures in time-frequency domain with estimated number of sources. *Speech Commun.* **2017**, *89*, 1–16. [[CrossRef](#)]
22. Wang, L.; Ohtsuki, T. Underdetermined Blind Source Separation With Multi-Subspace for Nonlinear Representation. *IEEE Access* **2019**, *7*, 84545–84557. [[CrossRef](#)]
23. Belouchrani, A.; Amin, M.G. Blind source separation based on time-frequency signal representations. *IEEE Trans. Signal Process.* **1998**, *46*, 2888–2897. [[CrossRef](#)] [[PubMed](#)]
24. Li, Y.; Amari, S.; Cichocki, A.; Ho, D.W.C.; Xie, S. Underdetermined blind source separation based on sparse representation. *IEEE Trans. Signal Process.* **2006**, *54*, 423–437.
25. Jourjine, A.; Rickard, S.; Yilmaz, O. Blind separation of disjoint orthogonal signals: Demixing N sources from 2 mixtures. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Istanbul, Turkey, 5–9 June 2000; pp. 2985–2988.
26. Li, K.; Sun, G.; Xiao, M. Dual channel delay speech denoising based on underdetermined BSS. In Proceedings of the International Conference on Mechatronics and Control (ICMC), Jinzhou, China, 3–5 July 2014; pp. 689–692.
27. Liu, Z.; Li, L.; Zheng, Z. Mixing matrix estimation method for dual-channel time-frequency overlapped signals based on interval probability. *Etri J.* **2019**, *41*, 658–669. [[CrossRef](#)]
28. Xu, T.; Wang, W.; Dai, W. Sparse coding with adaptive dictionary learning for underdetermined blind speech separation. *Speech Commun.* **2013**, *55*, 432–450. [[CrossRef](#)]
29. Weiss, A.; Yeredor, A. A Maximum Likelihood-Based Minimum Mean Square Error Separation and Estimation of Stationary Gaussian Sources From Noisy Mixtures. *IEEE Trans. Signal Process.* **2019**, *67*, 5032–5045. [[CrossRef](#)]
30. Zhu, Z.; Chen, X.; Lv, Z. Underdetermined Blind Source Separation Method Based on a Two-Stage Single-Source Point Screening. *Electronics* **2023**, *12*, 2185. [[CrossRef](#)]
31. Reju, V.; Koh, S.N.; Soon, I.Y. An algorithm for mixing matrix estimation in instantaneous blind source separation. *Signal Process.* **2009**, *89*, 1762–1773. [[CrossRef](#)]
32. Sun, J.; Li, Y.; Wen, J.; Yan, S. Novel mixing matrix estimation approach in underdetermined blind source separation. *Neurocomputing* **2016**, *173*, 623–632. [[CrossRef](#)]
33. Li, Y.; Nie, W.; Ye, F.; Lin, Y. A Mixing Matrix Estimation Algorithm for Underdetermined Blind Source Separation. *Circuits Syst. Signal Process.* **2016**, *35*, 3367–3379. [[CrossRef](#)]
34. Zhen, L.; Peng, D.; Yi, Z.; Xiang, Y.; Chen, P. Underdetermined Blind Source Separation Using Sparse Coding. *IEEE Trans. Neural Netw. Learn. Syst.* **2017**, *28*, 3102–3108. [[CrossRef](#)] [[PubMed](#)]
35. Zhang, L.; Yang, J.; Lu, K.; Zhang, Q. Modified subspace method based on convex model for underdetermined blind speech separation. *IEEE Trans. Consum. Electron.* **2014**, *60*, 225–232. [[CrossRef](#)]
36. Guo, Q.; Ruan, G.; Qi, L. A Complex-Valued Mixing Matrix Estimation Algorithm for Underdetermined Blind Source Separation. *Circuits Syst. Signal Process.* **2018**, *37*, 3206–3226. [[CrossRef](#)]
37. Li, Y.; Ramli, D.A. Research on Mixed Matrix Estimation Algorithm Based on Improved Sparse Representation Model in Underdetermined Blind Source Separation System. *Electronics* **2023**, *12*, 456. [[CrossRef](#)]
38. Luo, W.; Jin, H.; Li, X.; Li, H.; Liu, K.; Yang, R. A Novel Complex-Valued Blind Source Separation and Its Applications in Integrated Reception. *Electronics* **2023**, *12*, 3954. [[CrossRef](#)]
39. Scott, R. The DUET blind source separation algorithm. In *Blind Speech Separation*; Makino, S., Sawada, H., Lee, T.W., Eds.; Springer: Dordrecht, The Netherlands, 2007; pp. 217–241.
40. Abrard, F.; Deville, Y. A Time-Frequency Blind Signal Separation Method Applicable to Underdetermined Mixtures of Dependent Sources. *Signal Process.* **2005**, *85*, 1389–1403. [[CrossRef](#)]
41. Peng, D.; Xiang, Y. Underdetermined Blind Source Separation Based on Relaxed Sparsity Condition of Sources. *IEEE Trans. Signal Process.* **2009**, *57*, 809–814. [[CrossRef](#)]
42. Bofill, P.; Zibulevsky, M. Underdetermined blind source separation using sparse representations. *Signal Process.* **2001**, *81*, 2353–2362. [[CrossRef](#)]
43. Li, Y.; Cichocki, A.; Amari, S. Analysis of Sparse Representation and Blind Source Separation. *Neural Comput.* **2004**, *16*, 1193–1234. [[CrossRef](#)]
44. Donoho, D.L. Compressed sensing. *IEEE Trans. Inf. Theory* **2006**, *52*, 1289–1306. [[CrossRef](#)]

45. Tropp, J.A.; Gilbert, A.C. Signal Recovery From Random Measurements Via Orthogonal Matching Pursuit. *IEEE Trans. Inf. Theory* **2007**, *53*, 4655–4666. [[CrossRef](#)]
46. Sahoo, S.K.; Makur, A. Signal Recovery from Random Measurements via Extended Orthogonal Matching Pursuit. *IEEE Trans. Signal Process.* **2015**, *63*, 2572–2581. [[CrossRef](#)]
47. Liu, Z.; Li, L.; Lv, D.; Pan, N. Novel Source Recovery Method of Underdetermined Time-Frequency Overlapped Signals Based on Submatrix Transformation and Multi-Source Point Compensation. *IEEE Access* **2019**, *7*, 29610–29622. [[CrossRef](#)]
48. Mallat, S.G.; Zhifeng Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Trans. Signal Process.* **1993**, *41*, 3397–3415. [[CrossRef](#)]
49. Zhou, G.; Yang, Z.; Xie, S.; Yang, J. Mixing Matrix Estimation From Sparse Mixtures With Unknown Number of Sources. *IEEE Trans. Neural Netw.* **2011**, *22*, 211–221. [[CrossRef](#)] [[PubMed](#)]
50. Pati, Y.C.; Rezaifar, R.; Krishnaprasad, P.S. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In Proceedings of the 27th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, 1–3 November 1993; Volume 1, pp. 40–44.
51. Tropp, J.A.; Wright, S.J. Computational Methods for Sparse Solution of Linear Inverse Problems. *Proc. IEEE* **2010**, *98*, 948–958. [[CrossRef](#)]
52. Griffin, D.; Lim, J. Signal estimation from modified short-time Fourier transform. *IEEE Trans. Acoust. Speech Signal Process.* **1984**, *32*, 236–243. [[CrossRef](#)]
53. Wu, K.; Reju, V.G.; Khong, A.W. Multisource DOA estimation in a reverberant environment using a single acoustic vector sensor. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2018**, *26*, 1848–1859. [[CrossRef](#)]
54. Donoho, D.L. For most large underdetermined systems of linear equations the minimal  $\ell_1$ -norm solution is also the sparsest solution. *Commun. Pure Appl. Math.* **2006**, *59*, 907–934. [[CrossRef](#)]
55. van den Berg, E.; Friedlander, M.P. Probing the Pareto frontier for basis pursuit solutions. *SIAM J. Sci. Comput.* **2008**, *31*, 890–912. [[CrossRef](#)]
56. Vincent, E.; Gribonval, R.; Fevotte, C. Performance measurement in blind audio source separation. *IEEE Trans. Audio Speech Lang. Process.* **2006**, *14*, 1462–1469. [[CrossRef](#)]
57. Garofolo, J.S. *TIMIT Acoustic-Phonetic Continuous Speech Corpus*; Web Download; Linguistic Data Consortium: Philadelphia, PA, USA, 1993.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.