IEEE Access
Multidisciplinary ⁝ Rapid Review ⁝ Open Access Journal

Received November 14, 2019, accepted December 6, 2019, date of publication December 13, 2019, date of current version December 23, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2958955

Blind DOA Estimation in a Reverberant Environment Based on Hybrid Initialized Multichannel Deep 2-D Convolutional NMF With Feedback Mechanism

QIANG FU[1], BO JING[1], AND PENGJU HE[2,3]

[1]College of Aeronautics Engineering, Air Force Engineering University, Xi'an 710038, China
[2]Research and Development Institute of Northwestern Polytechnical University in Shenzhen, Shenzhen 518057, China
[3]School of Automation, Northwestern Polytechnical University, Xi'an 710072, China

Corresponding author: Qiang Fu (fuqiang931@126.com)

This work was supported in part by the project of China's Ministry of Industry and Information Technology under Grant MJ-2017-F-05, and in part by the Shenzhen's science and technology project under Grant JCYJ20170306154611415.

ABSTRACT The accuracy performance of traditional direction of arrival (DOA) estimation algorithms is seriously affected by the reverberation. Considering the advantage of the sparse characteristic of speech signal in time-frequency (T-F) domain, this paper presents a new blind DOA estimation method based on integrated deep learning and convolutional non-negative matrix factorization (NMF). Firstly, mathematic models of microphone array and room impulse response are built. In addition, we extracted blindly initialization parameters of 2-D convolutional NMF using k-means clustering algorithm and singular value decomposition algorithm, which can be used to accurately estimate the main components of desired sound source in the reverberation environment of multi-path propagation. Moreover, the feedback mechanism is introduced into deep 2-D convolutional NMF and correlation coefficient between the signal decomposed by NMF and the signal to be decomposed is used to select the best separated signal for DOA estimation, which make the separation algorithm simpler and more efficient. Finally, test of orthogonality of projected subspaces (TOPS) algorithm is used to validate the DOA estimation capability of this algorithm. Compared with the unprocessed reverberation speech, the estimation error is reduced, which shows that the proposed algorithm can effectively improve the estimation accuracy of DOA estimation when the received signals are in a reverberant environment.

INDEX TERMS Blind DOA estimation, speech dereverberation, array signal processing, convolutional non-negative matrix factorization (CNMF), depth extraction, feedback mechanism.

I. INTRODUCTION

Direction of arrival (DOA) is a fundamental theoretical problem in array signal processing (ASP), and is used widely in sonar, radar, mobile communication and electronic countermeasure [1]. Reverberation is a common phenomenon in daily life, and can be found in many indoor spaces such as airports, stations, classrooms, offices, etc. The existence of reverberation will seriously affect the accuracy of DOA [1]. Therefore, it is very important to improve DOA estimation accuracy in reverberation environment.

The commonly used methods of extracting DOA from measurement data can be divided into conventional beamforming and subspace-based techniques. The weight of conventional beamforming method is fixed, so it can not adapt to different noise and various interference. Compared the performance with other methods, this method has the best real-time performance. The subspace-based algorithm is represented by MUSIC and ESPRIT, Schmidt developed the multiple signal classification (MUSIC) based on spatial spectral estimation in 1979, and Roy and Kailath developed the estimation of signal parameter via rotational invariance techniques (ESPRIT) in 1986. Eigen-Value Decomposition (EVD) and traversing search within the whole space

The associate editor coordinating the review of this manuscript and approving it for publication was Mohammad Zia Ur Rahman.

spectrum in the traditional MUSIC algorithm is a time-consuming process. The disadvantage of the space spectrum estimation algorithm is that has a large computational cost. Reduce computation complexity and improve the efficiency is very important. It is easy to calculate without spectral peak search using ESPRIT, but the estimation accuracy is not guaranteed when the two-dimensional parameters are the same. MUSIC algorithm is more accurate, more stable and higher angular resolution than ESPRIT algorithm.

However, these methods mentioned above are based on the premise that the measured signal is a pure signal, and can not apply to signals in strong reverberation environment directly. In nature, almost many signals are in the presence of background noise and reverberant conditions. In practice, the problem of background noise has been well solved, but room reverberation still affects the performance of the algorithm to a large extent. However, the existing multi-source location algorithm has achieved good results in the environment of medium and low reverberation and high signal-to-noise ratio, but it is still unable to achieve accurate location in the environment of strong reverberation at present, which makes it difficult for a large number of algorithms to be applied in practice. When the room is quiet, the noise is small and the objective room reverberation becomes the main factor affecting the accuracy of DOA estimation. So far, many kinds of classical dereverberation methods have been developed, such as beamforming method, complex cepstrum filtering method, minimum phase decomposition method and wiener filtering method. These methods have great advantages when room reverberation is not serious, and inverse filtering method is effective only in a very short reverberation time of 200-400ms. However, since the impact response of the general room is non-minimum phase and has an unstable inverse, and the phase winding problem of reconstruction, the practical application of the classical method is very limited.

Nonnegative Matrix Factorization (NMF) is an effective tool for decomposing mixed audio signals in time-frequency domain. In recent years, NMF has also been tried to solve the problem of speech dereverberation. NMF has gained a certain effect in speech separation, which shows strong regression capabilities, and has been used to address the speech dereverberation issue [2], [3]. In [4], a single acoustic vector sensor is used to achieve spatial filtering for DOA estimation in multisource reverberant environments. But due to there are only three sensor elements have directional response in an AVS, the maximum number of sources is restricted to two. Research shows that the short-time Fourier transform coefficients of pure speech have certain sparseness. The short-term coefficient sparseness of the reverberant speech signal collected by the microphone is smaller than that of pure speech, so an output signal similar to pure speech can be generated by improving the sparseness of the output signal [5]. NMF is non-negative, which means the elements in the decomposition matrix are all non-negative, while NMF can make the result sparse. According to this characteristic, NMF can be applied to dereverberation. Non-negative tensor factorization (NTF) is applied to binaural sound source localization, which uses the sparse representation of multichannel audio signals in time, frequency, and space [6], but it can only realize the localization of less than two sound sources. Blind single-channel speech dereverberation method based on N-CTF model and NMF is presents to enhance the quality of speech signals that have been recorded in an enclosed space [7]. The algorithm requires a lot of prior knowledge of the spectrum characteristics of sound source and environmental noise, which is difficult to obtain in the actual environment. In addition, according to the number of microphones used, the dereverberation methods are mainly divided into single-channel speech dereverberation methods and multi-channel speech dereverberation methods. When using a single microphone to achieve speech reverberation, the position of the sound source and the microphone should be relatively fixed, and the distance between the sound source and the microphone must be very close. Because multi-channel algorithm can utilize spatial diversity, it is possible to gain more gain than single-channel algorithm. In multi-channel NMF method [8], different constraints affect the performance of dereverberation to some extent. How to add appropriate constraints is the problem of reverberation removal method based on NMF. To sum up, the problem lies in lacking efficient methods to deal with the constraints and NMF algorithm is sensitive to initial value.

Therefore, in view of the problems in the above research, this paper proposes a method based on deep 2D convolutional NMF to suppress reverberation in narrow enclosed space and achieve accurate DOA estimation. The paper is organized as follows. Wideband array and signal mixing model are briefly introduced in section 2. From the perspectives of dereverberation method, dereverberation degree and DOA estimation method, sections 3, 4 and 5 respectively propose 2-D convolutional NMF algorithm, depth extraction method and TOPS algorithm. Section 6 summarizes the overall algorithm steps and draws the corresponding algorithm flow chart. Simulation experiments are done to verify the correctness of the proposed algorithm in section 7. Conclusions and future works are discussed in Section 8.

II. PROBLEM DESCRIPTION

Compared with the traditional array model, besides noise, we also need to consider the influence of reverberation. Therefore, the microphone array needs to receive both voice signals and spatial interference sources (including noise and reverberation), which increases the complexity of the model.

The empirical condition for judging the far field is that the radial distance from the sound source to the array is greater than $2L^2/\lambda$, where L is the aperture of the array, λ is the wavelength [9]. Wideband signals simultaneously impinging on the far field of a uniform linear array (ULA), in which the microphones are uniformly spaced and the element spacing is d, Suppose the space has K far-field wideband sources, and signals inject on the array antenna in different directions.

For indoor applications, there are two kinds of models, one is the ideal model considering only environmental noise, the other is the reverberation model. Ideal uniform linear array can be described as Fig.1.

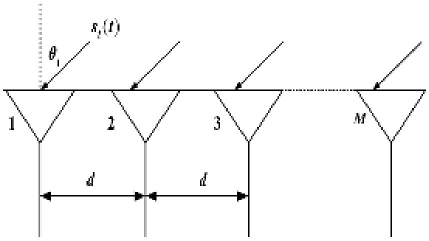


FIGURE 1. Ideal uniform linear array schematic.

The mathematical model of the output signal of ULA can be written as

$$X = AS + N \tag{1}$$

where, S is the sound source; A stands for the mixing matrix; X represents the observation matrix formed by the observation signals containing noise; N denotes the additive noise vector.

Consider a ULA of isotropic sensors equispaced by d, θ is the DOA of a wideband far field signal. Employing the first element of the ULA as the phase reference, the steering vector can be given by

$$a(\theta) = \left[1, \upsilon, \cdots, \upsilon^{M-1}\right]^T, \quad \upsilon = e^{-j2\pi d/\lambda \sin\theta} \tag{2}$$

where λ denotes the carrier wavelength of the signal. The superscript $(\cdot)^T$ stands for the matrix/vector transpose.

Because of the reflection of the room wall and other reasons, the speech signal propagates in the room through multiple paths, resulting in the phenomenon of amplitude attenuation and sound quality deterioration of the received signal, known as reverberation. These reverberation components are mixed with the signals transmitted over the direct path, which affects the quality of the collected speech signals and reduces the sound source location performance of the signal processing system. Reverberation uniform linear array can be described as Fig.2.

In the actual model, the signal received by the microphone includes not only the direct signal of the sound source and the environmental noise, but also includes signals that are repeatedly reflected between walls and other objects in the room before reaching the microphone. The mathematical model can be written as

$$x[n] = u[n] + \sum_{k=0}^{\infty} \rho_k s[n - kn_k] \tag{3}$$

where, s[n] is the sound source; $\mathrm{n_k}$ is the delay unit value of the signal after the k-th reflection; s[n-kn$_\mathrm{k}$] is the signal after the k-th reflection of the original signal, ρ_k is the reflection coefficient of the k-th reflection; u[n] is the noise signal in the current environment x[n] is the sum of the signal directly arriving at the sensor and the signal arriving at the sensor through all other paths.

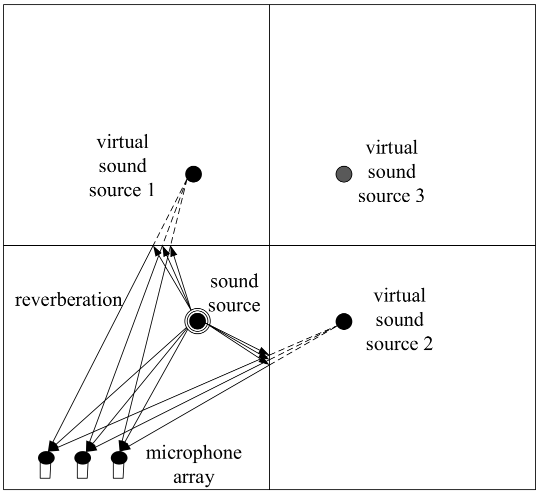


FIGURE 2. Actual uniform linear array schematic.

When the room is quiet, the noise is small and the room reverberation becomes the main factor affecting the estimation accuracy. Ignoring the influence of the noise signal u[n], the reverberation model represented by formula (3) can be rewritten into convolution form according to the convolution property of unit impulse function(n)

$$x[n] = s[n] * \sum_{k=0}^{\infty} \rho_k \delta[n - kn_k] = s[n] * h[n] \tag{4}$$

where, $*$ represents convolution in time-domain; h[n] is the impulse response of the room; h[n] equals the sum of a series of continuous impulse functions; δ[n] represents unit impulse function. The reverberation signal x[n] is equal to the convolution of the room impulse response h[n] and the pure sound source signal s[n].

$$h[n] = \sum_{k=0}^{\infty} \rho_k \delta[n - kn_k] \tag{5}$$

There are many methods to calculate room impulse response [10]. IMAGE method [11] is a typical one, and is used in many scenarios, such as detection of snoring in the human head. In this paper, an indoor reverberation model based on IMAGE method is used to calculate room impulse response. The impulse response function of room reverberation is generated by IMAGE model. The room is 6 meters long, 5 meters wide and 3.5 meters high. Take the bottom left corner of the top view as the origin, the sound source position is (5,5,1), the position of three sensors is (1,1,1),(2,1,1) and (3,1,1), and the sound velocity is 345m/s. Assuming all walls have the same reflection coefficient, and the reverberation time can be changed by adjusting the reflection coefficient of the room wall. The speech signal is convolved with the room impulse response to produce reverberation speech.

Reverberation time T60 is a parameter used to describe the attenuation rate of indoor sound. The channel response of different RIRS are shown in Fig.3. The RIRS labeled as channel 1, channel 2 and channel 3, respectively.

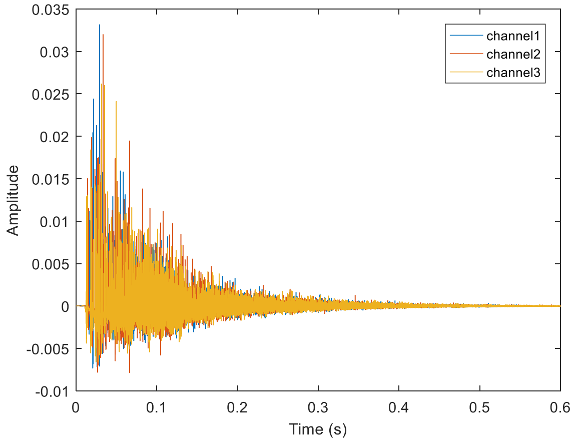


FIGURE 3. Room impulse response.

The phasing of the reverberation speech signal generated by the sound source will deviate from the source signal after it is reflected by various dielectric surfaces in the room and attenuated in the air. In IMAGE model, the sound source is idealized as a point, and reverberation signal is the convolution of pure speech signal and RIR [12]. Multichannel microphone systems are sensitive to interchannel phase, which has a great impact on the subsequent DOA estimation. So, the multi-channel reverberation signal collected by the array needs some reverberation suppression

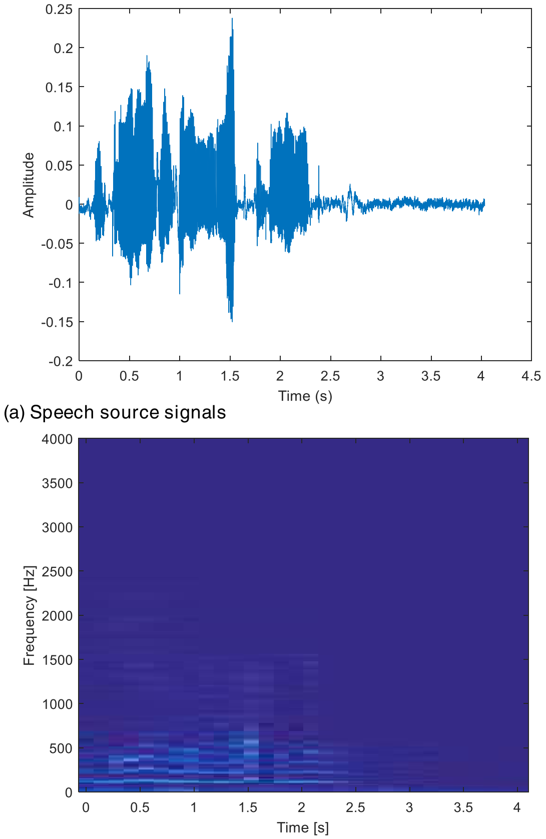
III. 2-D CONVOLUTIONAL NMF ALGORITHM

The source signal is a single source speech signal. The following experiments deal with a single source formed of a real speech sound sampled at 16000 Hz. Fig. 4 shows the waveform and spectrogram of pure source signals.

The speech signal waveform affected by reverberation are shown in Fig.5.

As can be seen from Fig.4 and Fig.5, excessive reverberation will seriously affect the clarity of the original signal.

Because there are a large number of local extremums, NMF needs to be carefully initialized to produce meaningful results. Through using k-means clustering method, clustering center is calculated as the initial value of the coefficient matrix H, avoiding traditional decomposition result's non-unity problem. Considering the number of base matrix W of 2-D convolutional NMF algorithm is more than one-dimensional nonnegative matrix decomposition, using singular value decomposition and principal component analysis method to iterative initial W matrix, avoiding the initialization error from a single algorithm. Because of its high sensitivity to initialization, initialization of NMF can provide significantly better results than previous NMF algorithms that directly responded to random parameters. In addition to the extended base matrix W, the coefficient matrix H is extended,



(a) Speech source signals

(b) Spectrogram

FIGURE 4. Clean source signals.

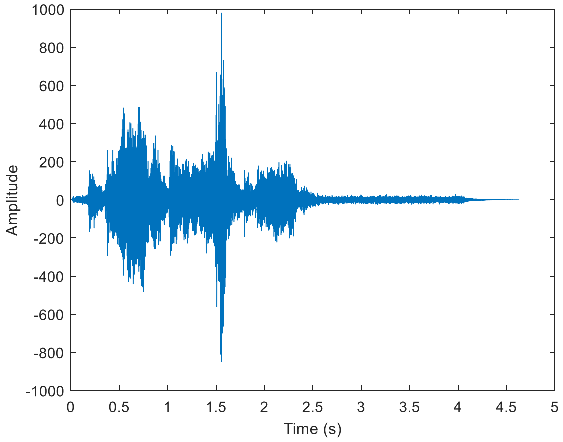
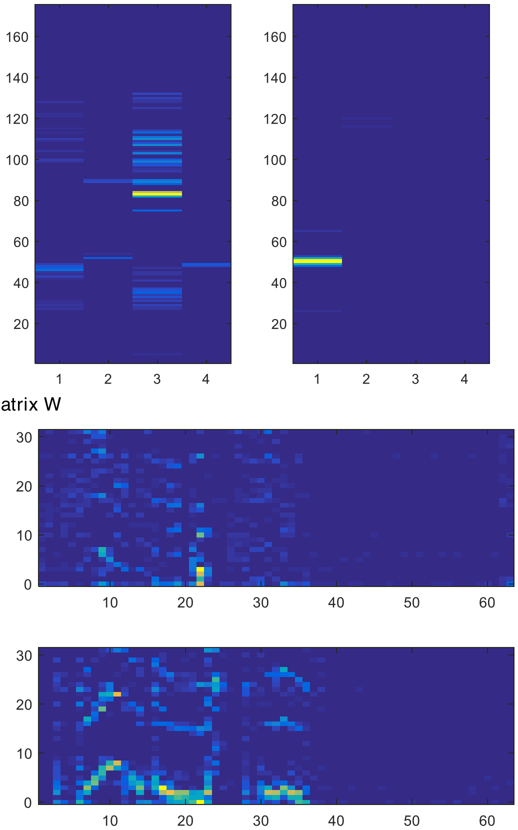


FIGURE 5. Reverberation speech of channel 1.

and the 2-D convolutional NMF algorithm formula is as follows.

$$V \approx \mathbf{\Lambda} = \sum_{\tau} \sum_{\phi} \overset{\phi\downarrow}{W_{\tau}} \cdot \overset{\tau\rightarrow}{H_{\phi}} \tag{6}$$

where, $\overset{\tau\rightarrow}{(\cdot)}$ is a matrix that moves τ points to the right by column, and the left empty column is filled with 0, $\overset{\phi\downarrow}{(\cdot)}$ is a matrix that moves ϕ points to the bottom by row, and the upper empty row is filled with 0.



(a) Matrix W

(b) Matrix H

FIGURE 6. The process of 2-D convolutional NMF.

TABLE 1. Correlation coefficient of channel 1.

Number of decomposition	Separated signal 1	Separated signal 2
1	-0.1033	**-0.3998**
2	**-0.4307**	-0.2318
3	**-0.5620**	-0.1172
4	**-0.7313**	-0.3688
5	**-0.8130**	-0.4290
6	**-0.8637**	-0.5765
7	**-0.8552**	-0.6340

TABLE 2. Correlation coefficient of channel 2.

Number of decomposition	Separated signal 1	Separated signal 2
1	0.1696	**-0.2519**
2	**-0.3469**	-0.1497
3	**-0.5072**	-0.0609
4	**-0.6291**	-0.1066
5	**-0.6463**	-0.3993
6	**-0.6796**	-0.1851
7	**-0.7036**	-0.4431
8	**-0.7574**	-0.4706
9	**-0.7644**	-0.5441

TABLE 3. Correlation coefficient of channel 3.

Number of decomposition	Separated signal 1	Separated signal 2
1	-0.0171	**0.1319**
2	**0.1349**	0.0991
3	0.1196	**0.3494**
4	0.0702	**0.5901**
5	0.1862	**0.6383**
6	0.1902	**0.6541**
7	0.2173	**0.6631**

In [13], initial values for the basis and the activations were obtained by performing NMF on the spectrogram of reverberated speech. Taking the first channel as an example, the matrix W and matrix H in the decomposition and the separation result are as Fig.6.

IV. DEPTH EXTRACTION METHOD

The deep learning model is complex and non-convex. Combining deep learning with non-negative matrix factorization (NMF), many algorithms are proposed, such as multi-layer NMF, sparse multi-layer NMF, deep NMF, deep Semi-NMF, etc. Scholars find better deep matrix decomposition results by adjusting each W matrix of each decomposition in the model. Based on these models and the idea of feedback mechanism, we study a deep 2-D convolutional nonnegative matrix decomposition model. Feedback separation [14] means that in the process of separation of mixed signals, the signal with the best separation effect is fed back to the input of mixed signals, and the signal is subtracted from the input to form a new mixed signal. In this paper, the feedback mechanism is introduced into the dereverberation NMF to form a new depth extraction algorithm.

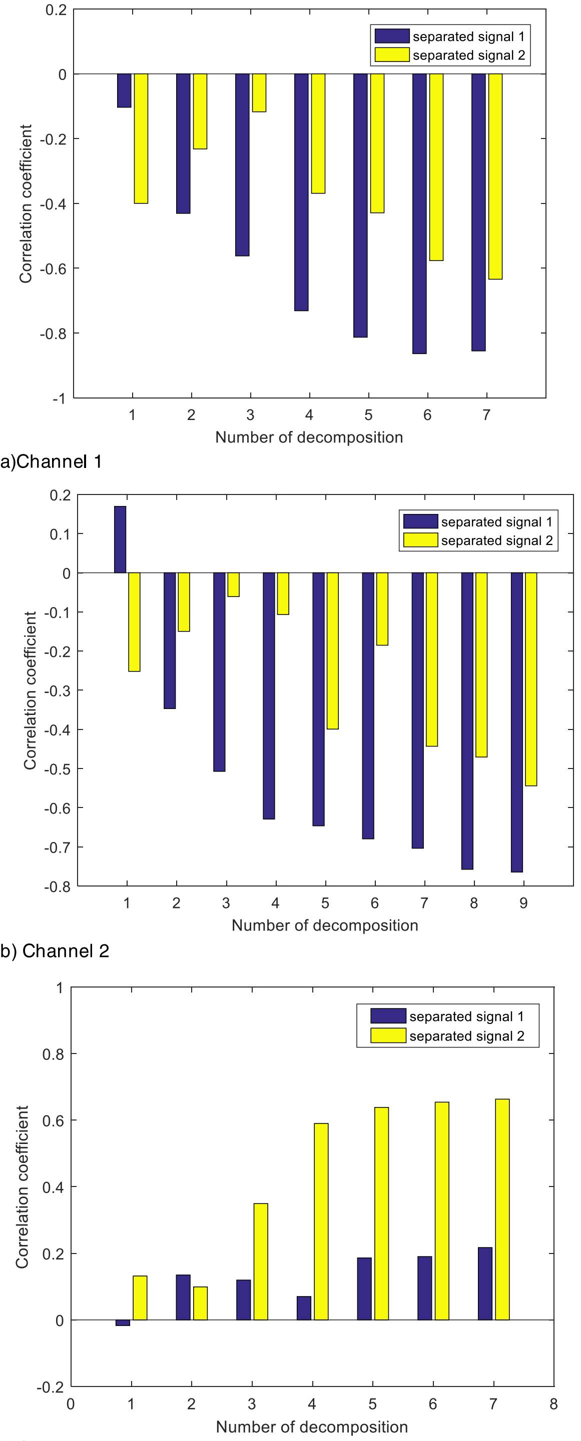
The correlation coefficient between the separated source signal and reverberation signal is used to measure the purity of the separated source signal, and the signal with larger correlation coefficient in separated signals is selected for further decomposition. The Formula for calculating the correlation coefficient of each separated signal and the signal to be separated is as follows.

$$\rho_{XY_i} = \frac{Cov(X, Y_i)}{\sqrt{D(X)}\sqrt{D(Y_i)}} \tag{7}$$

where, X is received data of the array in reverberation environment; Y_i is the i-th separated signal; the $D(\cdot)$ function stands for the variance.

The signal decomposition will stop until the difference between the correlation coefficient of the separated signal at this step and that of the separated signal at the previous step is less than 0.01, and the purest separated signal is extracted for the next DOA estimation. The correlation coefficients of different channels under the number of iterations required by the depth extraction algorithm proposed in this paper are as follows.

As can be seen from Tab. 1-3 and Fig. 7, NMF, as a new blind source separation algorithm, is not monotonic in terms of the number of decomposition due to the inherent sequence uncertainty of blind source separation. However, when choosing the next decomposition object, the method in this paper is to choose a decomposition signal with a large correlation coefficient to continue the decomposition. Therefore, on the whole, the correlation coefficient is monotonic in terms of the number of decompositions.



(a)Channel 1

(b) Channel 2

(c) Channel 3

FIGURE 7. Correlation coefficient.

The correlation coefficient between the signal decomposed by NMF and the signal to be decomposed is used to select the separated signal for DOA estimation. It can also be seen from H matrix in Fig.6(b) that compared with the other separated signals, the selected signal is cleaner in the absence of voice signal, and has obvious sparsity in the presence of voice signal, which further verifies the reliability of the algorithm in this paper. After deep non-negative matrix decomposition, the higher-level features of the signal are extracted layer by layer. It is worth pointing out that the stronger the reverberation, the more layers need to be decomposed.

A depth extraction method based on correlation coefficient is proposed for reverberation signal, and the next step is to estimate DOA using the extracted signal.

V. TOPS ALGORITHM

Speech signal is a non-stationary process, which can not be directly analyzed and processed by digital signal processing technology for stationary signals. Although speech signal has time-varying characteristics, its characteristics remain basically unchanged in a short time range. Therefore, it can be regarded as a quasi-steady-state process, namely speech signal has short-term stability [15]. All speech signal analysis and processing methods divide speech signal into segments to analyze its characteristic parameters, each of which is called a frame. TOPS algorithm uses multiple frequency components of broadband signals to process data at one time, which can avoid the instability of independent subband estimation performance in non-correlation methods [16]. TOPS algorithm also does not need the focus angle required by relevant estimation methods, which can avoid the DOA estimation errors introduced by the default focus angle errors in the focusing process of coherent signal subspace method (CSSM). To solve the problem of low performance in low signal-to-noise ratio of TOPS algorithm, many improved methods are proposed. However, like most existing DOA estimation algorithm, the TOP algorithm does not explicitly consider reverberation.

TOPS algorithm is an effective method for wideband incoherent sources, which takes advantage of all frequency points in the bandwidth range and does not need angle coarse estimation for focusing. The noise subspace is projected onto the signal subspace of the reference frequency, and a new matrix $D(\theta)$ is constructed by these projections in TOPS algorithm. The diagonal element of the transformation matrix is

$$\Phi(\omega_k, \theta) = \exp(-j\omega_k \tau) \tag{8}$$

$$U = \mathrm{diag}(\Phi(\Delta w_k, \theta))F_0 \tag{9}$$

where, $\Delta w_k = w_k - w_0$, F_0 is the reference signal subspace obtained at the reference frequency, θ is the possible azimuth.

$$D(\theta) = [\, U_1'^H W_1 \; | U_2'^H W_2 \; | \cdots \; | U_K'^H W_K \,] \tag{10}$$

When the new matrix $D(\theta)$ is rank deficient, the angle of wideband signal can be estimated by one dimensional angle traversal search.

In practical applications, orthogonal projection matrix $P(\omega_k, \theta)$ is often used to correct the test matrix $D(\theta)$.

$$P(\omega_k, \theta) = I - (a^H(\omega_k, \theta)a(\omega_k, \theta)^{-1}a(\omega_k, \theta)a^H(\omega_k, \theta)) \tag{11}$$

$$U'(\omega_k, \theta) = P(\omega_k, \theta)U(\omega_k, \theta) \tag{12}$$

where, I is the identity matrix, U' is the projection matrix.

Square TOPS method improves the resolution of DOA estimation by modifying matrix $D(\theta)$. The matrix is modified to

$$D'(\theta) = [U_1'^H W_1 W_1'^H U_1 | U_2'^H W_2 W_2'^H U_2 | \cdots | \\ U_K'^H W_K W_K'^H U_K] \quad (13)$$

Finally, DOA estimation is obtained by spectral peak search.

$$\hat{\theta} = \arg\max_{\theta} \frac{1}{\sigma_{\min}(\theta)} \quad (14)$$

where, $\sigma_{\min}(\theta)$ is a minimum singular value of the matrix $D'(\theta)$.

From the calculation process of the TOPS algorithm, it can be seen that generating new and suitable array signals is very important for the accuracy of the algorithm. Therefore, the idea of this paper is to decompose the reverberation signal into a non-negative matrix and regenerate a new array signal as the input of the TOPS algorithm.

VI. TECHNIQUE PROCESS OF PROPOSED ALGORITHM

The steps of the algorithm for blind DOA estimation in a strong reverberation environment can be given by follows.

1) The received data X of the array in reverberation environment is obtained.

2) Use k-means clustering method to calculate the initial value of the coefficient matrix H, and use singular value decomposition and principal component analysis method to iterative initial matrix W.

3) The 2-D convolutional NMF method was used to separate multi-channel array signal V.

4) Calculate the correlation coefficients between the separated signal and reverberation signal. If it is less than 0.01, then the obtained W and H are the final matrix. The dereverberation step in the algorithm is finished and turn to step 8, otherwise turn to the next step.

5) By calculating the correlation value between each separated signal and the signal to be separated, the separated signal corresponding to the maximum value is selected, and also means that the signal which retains the most important information of the original signal is separated.

6) Reconstruct the best separation signal as the observation signal matrix V′.

7) According to the new matrix V′., W and H are recalculated again by 2-D convolutional NMF method, and return to step2.

8) Generate new array signals as the input of TOPS algorithm and estimate DOA.

Taking the array of N microphone as an example, the detailed flow chart of the algorithm presented in this paper is shown in Fig.8.

VII. SIMULATION ANALYSES AND DISCUSSIONS

In this paper, all simulations are conducted in ULA, the analyses include estimation error analysis and robustness analysis.

TABLE 4. Presents the global parameters used in simulations.

Number of array elements	3
Distance between elements	0.5λ
DOA optimization resolution	0.1deg
frequency ranges	[45,300] Hz
sampling frequency	16000 Hz
snapshots	20

In the case of channel 1, the signal produced by each iteration decomposition is as follows.

It can be seen from the Fig.15 that the phenomenon of amplitude attenuation and time trailing after each syllable is well solved from the time domain, and the energy is more concentrated and the resonances are less overlap with each other from the frequency domain.

The speech performance index can be divided into objective index and subjective perception index, which mainly includes SAR (Sources to Artifacts Ratio), SDR (Source to Distortion Ratio), SIR (Source to Interferences Ratio), PESQ (Perceptual Evaluation of Speech Quality) and STOI (Short-time Objective Intelligibility). SAR, SDR and SIR mainly evaluate the separation results of multiple speakers, PESQ mainly evaluates the auditory perception characteristics of speech, and STOI mainly evaluates the intelligibility of separated speech under noise interference. Because this paper focuses on the speech enhancement effect in the case of reverberation, STOI and PESQ are used to evaluate the separated speech.

PESQ [17] is ITU-T standards for intrusive speech quality measurement, and STOI [18] metric is an intrusive speech intelligibility metric based on the correlation of normalized filterbank envelopes in short-time frames of speech.

$$PESQ = 4.5 - 0.1D_{ind} - 0.0309A_{ind} \quad (15)$$

where, D_{ind} is mean disturbance value; A_{ind} is Symmetric interference.

STOI range is [0,1], PESQ range is [−0.5,4.5]. The higher the value, the higher the intelligibility and quality of the speech. By calculation, the STOI of unprocessed speech received by the three channel sensors is 0.1728, 0.1866 and 0.1892 respectively. PESQ is 2.571,2.479 and 2.591, respectively. After pretreated, the STOI of signal is 0.1988, 0.216 and 0.2205 respectively, and PESQ is 2.811,2.705 and 2.833, respectively.

In the presence of reverberation, improvements are up to 0.25 on PESQ and 0.3 on STOI by using the method proposed in this paper. The improvement of STOI value indicates that the improvement effect of voice quality is obvious. The increase of PESQ indicates a certain improvement in auditory perception.

In this paper, the signal frequency is concentrated in the low frequency, and signal spectrum after reverberation suppression is concentrated in 45 to 300Hz, so the signal frequency range is selected as 45 to 300Hz. The basic parameters in the experiment are shown in Table 4.

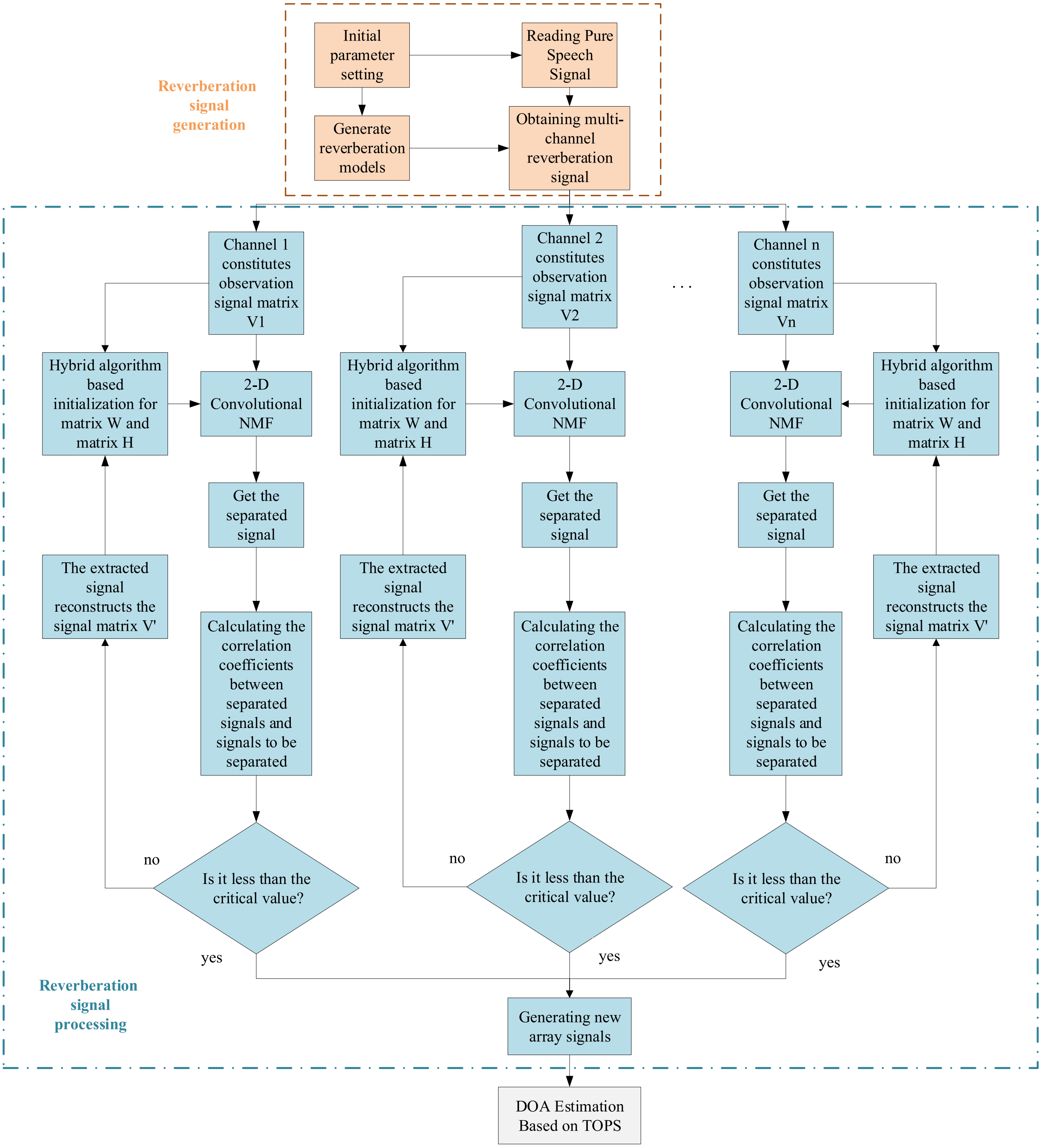
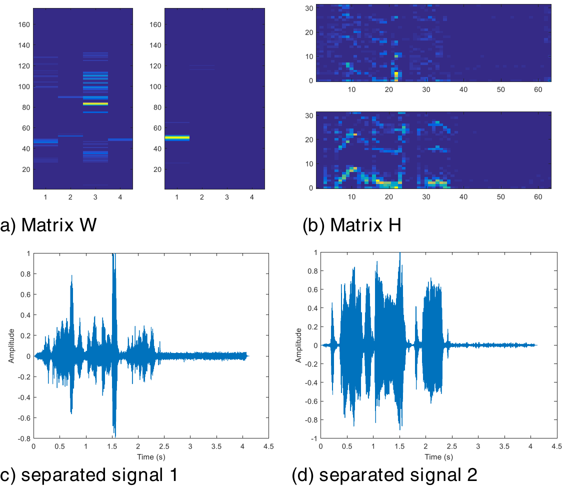

Reverberation signal generation
Initial parameter setting
Reading Pure Speech Signal
Generate reverberation models
Obtaining multi-channel reverberation signal
Channel 1 constitutes observation signal matrix V1
Channel 2 constitutes observation signal matrix V2
Channel n constitutes observation signal matrix Vn
. . .
Hybrid algorithm based initialization for matrix W and matrix H
2-D Convolutional NMF
Get the separated signal
The extracted signal reconstructs the signal matrix V'
Calculating the correlation coefficients between separated signals and signals to be separated
Is it less than the critical value?
no
yes
Reverberation signal processing
Generating new array signals
DOA Estimation Based on TOPS

FIGURE 8. Flowchart of algorithm for blind DOA estimation in reverberant environment.

In order to better verify the effectiveness of the proposed algorithm in reverberation environment, comparisons are made between the improved algorithm based on multichannel deep 2-D convolutional NMF and the TOPS algorithm. The influence of noise is not considered in this section, and the reverberation time is selected as 60ms in the experiment.

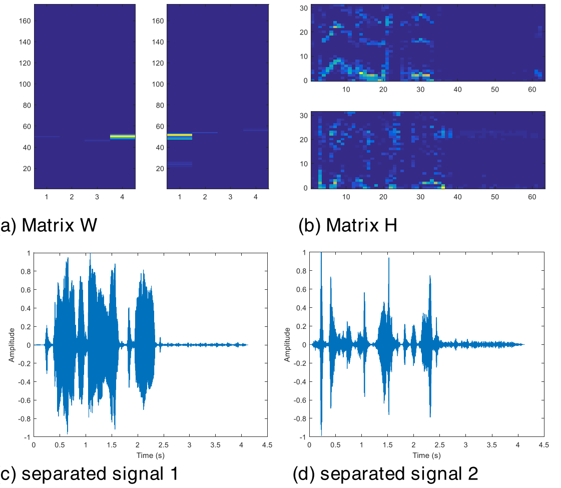
It can be seen from Fig. 16 that the reverberation environment causes the peak of the beam to be expanded, which makes it difficult to determine its maximum value and enlarges the positioning error. Worse still, the false peak amplitude is higher than the real signal amplitude, so it is impossible to estimate the direction of arrival effectively.



(a) Matrix W (b) Matrix H

(c) separated signal 1 (d) separated signal 2

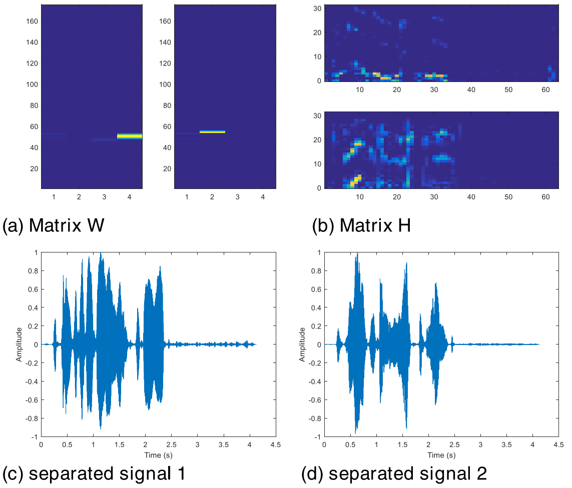
FIGURE 9. **The signal generated by the first decomposition.**



(a) Matrix W (b) Matrix H

(c) separated signal 1 (d) separated signal 2

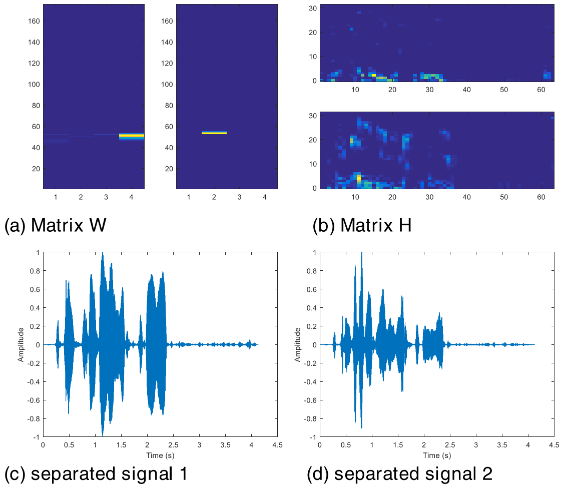
FIGURE 10. **The signal generated by the second decomposition.**



(a) Matrix W (b) Matrix H

(c) separated signal 1 (d) separated signal 2

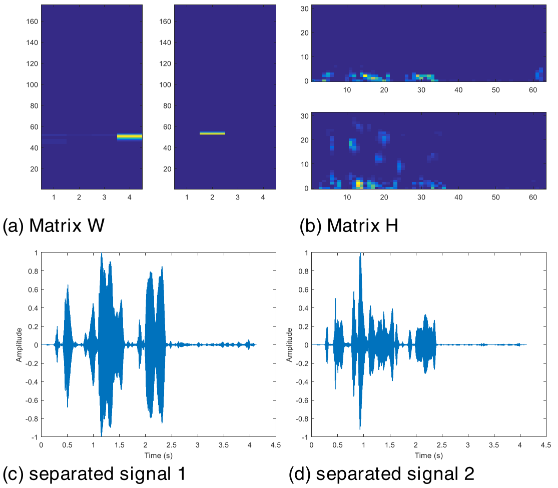
FIGURE 11. **The signal generated by the third decomposition.**



(a) Matrix W (b) Matrix H

(c) separated signal 1 (d) separated signal 2

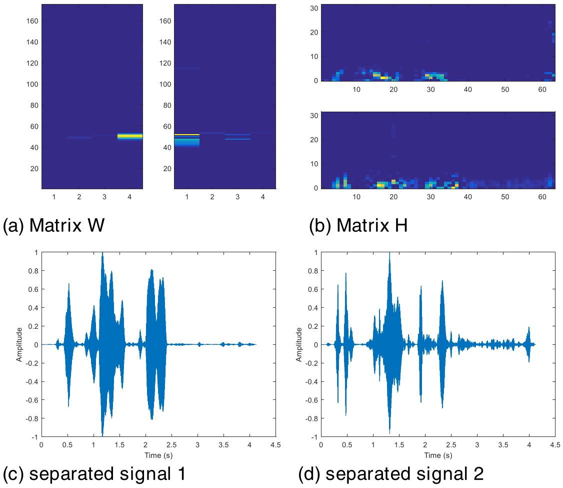
FIGURE 12. **The signal generated by the fourth decomposition.**



(a) Matrix W (b) Matrix H

(c) separated signal 1 (d) separated signal 2

FIGURE 13. **The signal generated by the fifth decomposition.**

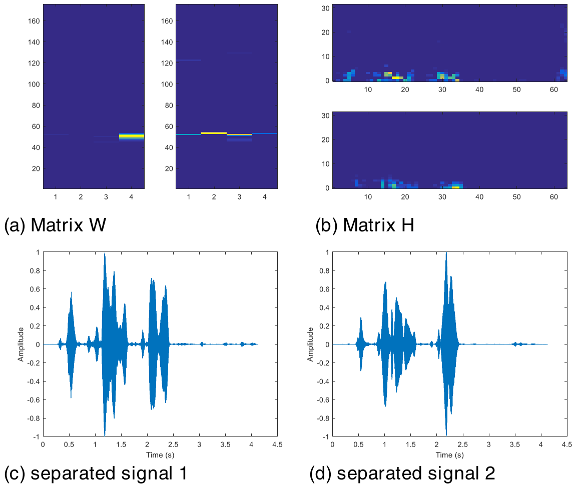


(a) Matrix W (b) Matrix H

(c) separated signal 1 (d) separated signal 2

FIGURE 14. **The signal generated by the sixth decomposition.**

In the presence of reverberation, the spatial maximum generated by the wrong location makes the spatial peak value of the real sound source position not obvious, which makes the search of the global peak error and leads to the failure of DOA estimation. The real peak has a great decline, which is not suitable for DOA estimation. Therefore, reverberation suppression should be carried out before DOA estimation.



(a) Matrix W (b) Matrix H

(c) separated signal 1 (d) separated signal 2

FIGURE 15. The signal generated by the seventh decomposition.

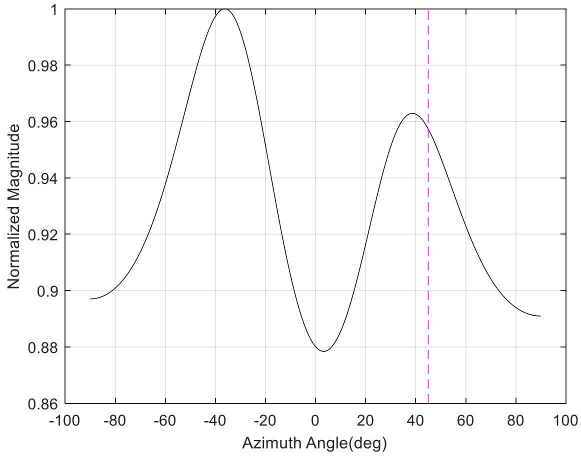


FIGURE 16. Azimuth estimation of source signals with NMF.

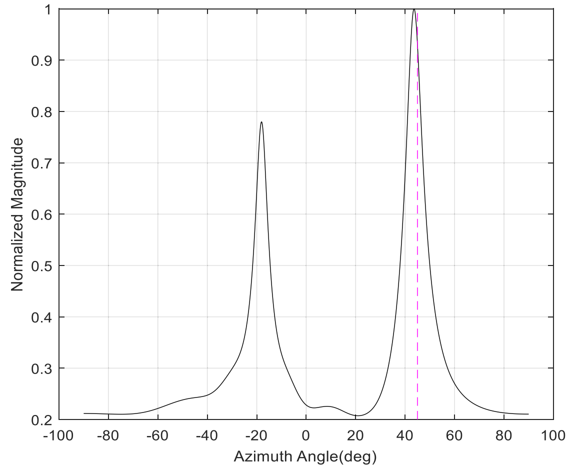


FIGURE 17. Azimuth estimation of source signal.

It can be seen from the Fig.17 that this method weakens the interference peak obviously and highlights the real peak value in the strong reverberation environment. Compared to Fig.16, not only the peak value is more prominent, but also the accuracy is obviously improved.

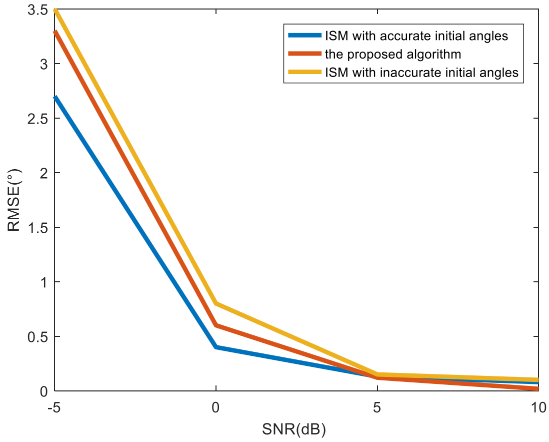


FIGURE 18. DOA estimation performance comparison versus SNR.

Wax [19] proposed the Incoherent Signal-Subspace Method (ISM) to solve the DOA estimation of broadband incoherent signal sources. To validate the validity of the algorithm above, the mean-square error of the proposed algorithm is compared with that of the ISM under different noise conditions.

The root mean square error (RMSE) under different conditions is used to evaluate the performance of the algorithm. RMSE of azimuth angle and elevation angle is

$$\theta_{RMSE} = \sqrt{\frac{1}{TK}\sum_{j=1}^{T}\sum_{i=1}^{K}\left(\hat{\theta}_{ij}-\theta_i\right)^2} \qquad (16)$$

where $\hat{\theta}$ is the estimated DOA of signals, K is the number of signal sources, T is the number Monte Carlo simulations.

In this section, $K = 1$, $T = 200$.

So, in this section, signals with different signal to noise ratio (SNR) are tested with the ISM in this paper. The selected noise is additive complex white Gaussian noise with SNR of -5dB, 0dB, 5dB and 10dB, respectively. The results are shown below in Fig. 18.

Comparing with traditional ISM, the method in this paper realizes more accurate DOA estimation with high SNR, and the ISM realizes more accurate DOA estimation with poor SNR. But the initial angles of ISM must be accurate, the RMSE of DOA estimation from this method can be smaller than that of ISM when the initial angles are inaccurate. The proposed algorithm has more stability.

Through analysis, we can see that there are serious errors in sound source location in strong reverberation environment and the direction-finding accuracy after reverberation suppression is obviously improved. The improved TOPS algorithm can estimate the angle of the signal accurately and the spectral peak is sharp. It can be obtained more accurate DOA estimation by using the algorithm presented in this paper.

VIII. CONCLUSION

In order to solve the problem of the DOA estimation in a reverberation environment, and considering the room impulse response is usually unknown and varies with sound source movement or room state in practical application scenarios, this paper studies a new blind DOA estimation method. In this paper, a direct method of using deep 2D convolutional non-negative matrix factorization to remove reverberation is proposed for the first time. The availability is verified by simulations and tests. The novelty of the proposed method is as follows.

(1) A new multichannel NMF model is designed. 2-D convolutional NMF is a single channel algorithm, this paper successfully used it into the multi-channel. Experimental results show that the algorithm has high reliability in speech recognition, and its accuracy is improved compared with the untreated reverberation speech recognition.

(2) Reinitialization technique, which is used in the iterative process of algorithm, reduces the degree of speech distortion while improving the effect of dereverberation. It also can realize self-adaptive adjustment according to actual sample characteristics.

(3) The number of decomposition cycles is determined by calculating the correlation between the original signal to be decomposed and the decomposed signal. Thus, the algorithm does not need to set the decomposition times in advance and the decomposition adaptability is improved. The algorithm only chooses signals with high correlation, and can also solve the problem of sequence uncertainty of separated signals.

The algorithm and calculation process presented in this paper have certain universality, which could also be extended to multi-source location system. This paper contributes a new technique for blind DOA estimation in a reverberant environment, that is, the azimuth of the reverberation signal can be estimated by using the mixed initialization-2D convolution NMF method to generate new array signal and substituting it into the commonly used DOA estimation method.

REFERENCES

[1] A. Jukic, N. Mohammadiha, and T. V. Waterschoot, "Multi-channel linear prediction-based speech dereverberation with low-rank power spectrogram approximation," Tech. Rep., 2015.

[2] H. Kameoka, T. Nakatani, and T. Yoshioka, "Robust speech dereverberation based on non-negativity and sparse nature of speech spectrograms," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Apr. 2009, pp. 45–48.

[3] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, and N. L. Dahlgren, *TIMIT Acoustic-Phonetic Continuous Speech Corpus*. Philadelphia, PA, USA: Linguistic Data Consortium, 1993.

[4] W. Kai, R. V. Gopalan, and A. W. H. Khong, "Multisource DOA estimation in a reverberant environment using a single acoustic vector sensor," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 26, no. 10, pp. 1848–1859, Oct. 2018.

[5] S. Arberet, A. Ozerov, N. Q. K. Duong, E. Vincent, R. Gribonval, F. Bimbot, and P. Vandergheynst, "Nonnegative matrix factorization and spatial covariance model for under-determined reverberant audio source separation," in *Proc. 10th Int. Conf. Inf. Sci., Signal Process. Appl.*, 2011, pp. 1–4.

[6] E. L. Benaroya, N. Obin, M. Liuni, A. Roebel, W. Raumel, and S. Argentieri, "Binaural localization of multiple sound sources by non-negative tensor factorization," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 26, no. 6, pp. 1072–1082, Jun. 2018.

[7] N. Mohammadiha and S. Doclo, "Speech dereverberation using non-negative convolutive transfer function and spectro-temporal modeling," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 24, no. 2, pp. 276–289, Feb. 2016.

[8] H. Sawada, H. Kameoka, S. Araki, and N. Ueda, "New formulations and efficient algorithms for multichannel NMF," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust.*, Oct. 2011, pp. 153–156.

[9] Q. Huang and T. Song, "DOA estimation of mixed near-field and far-field sources using spherical array," in *Proc. IEEE 11th Int. Conf. Signal Process.*, Oct. 2013, pp. 382–385.

[10] N. Mohanan, R. Velmurugan, and P. Rao, "Speech dereverberation using NMF with regularized room impulse response," Tech. Rep., 2017.

[11] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Amer.*, vol. 65, no. 4, pp. 943–950, 1979.

[12] P. M. Peterson, "Simulating the response of multiple microphones to a single acoustic source in a reverberant room," *J. Acoust. Soc. Amer.*, vol. 80, no. 5, pp. 1527–1529, 1986.

[13] M. N. Schmidt, *Nonnegative Matrix Factor 2-D Deconvolution for Blind Single Channel Source Separation*, 2006.

[14] H. Wenwen, "Research on blind signal separation based on nonnegative matrix factorization," M.S. thesis, Hangzhou Dianzi Univ., Hangzhou, China, 2013, pp. 46–53.

[15] S. Yoo, J. R. Boston, and J. D. Durrant, "Relative energy and intelligibility of transient speech components," in *Proc. Eur. Signal Process. Conf.*, 2015.

[16] Y.-S. Yoon, L. M. Kaplan, and J. H. McClellan, "TOPS: New DOA estimator for wideband signals," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 1977–1989, Jun. 2006.

[17] *Perceptual Evaluation of Speech Quality: An Objective Method for End-to-End Speech Quality Assessment of Narrow-Band Telephone Network and Speech Coders*, document ITU-T P.862, ITU Telecommunication Standardization Sector (ITU-T), 2001.

[18] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "An algorithm for intelligibility prediction of time–frequency weighted noisy speech," *IEEE Trans. Audio Speech Language Process.*, vol. 19, no. 7, pp. 2125–2136, Sep. 2011.

[19] M. Wax, T.-J. Shan, and T. Kailath, "Spatio-temporal spectral analysis by eigenstructure methods," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-32, no. 4, pp. 817–827, Aug. 1984.

QIANG FU received the B.Sc. degree in electrical engineering and automation and the M.Sc. degree in control science and technology from Air Force Engineering University, Xi'an, China, in 2013 and 2016, respectively, where he is currently pursuing the Ph.D. degree with the College of Aeronautics Engineering. His research interests include signal processing and fault diagnosis.

BO JING received the M.Sc. degree from Air Force Engineering University, in 1996, and the Ph.D. degree from Northwestern Polytechnical University, in 2002. She is currently a Professor with Air Force Engineering University. Her main research interests include prognostics and health management, design for testability, sensor networks, and information fusion.

PENGJU HE received the Ph.D. degree from Northwestern Polytechnical University, in 2004. He is currently an Associate Professor with Northwestern Polytechnical University. His main research interests include sensor networks and signal processing.