Improved Expectation-Maximization Algorithm for Unknown Reverberant Audio-Source Separation

Shaher Slehat [1,*]

[1] College of Artificial Intelligence, Balqa Applied University, Jordan
[1] Shaher5481561@yahoo.com*
* corresponding author

(Received: November 10, 2021; Revised: December 18, 2021; Accepted: January 2, 2022; Available online: January 25, 2022)

Abstract

The problem of undecided Separating reverberant audio sources is crucial for speech and audio processing. Numerous separation strategies have been developed to solve this problem; however, all of them estimate model parameters in the time–frequency domain, resulting in permutation ambiguity and poor separation performance. Additionally, one of the main challenges with existing expectation–maximization (EM) strategies is the time needed for each iterative step to update the model parameters. In this article, we offer an enhanced EM approach that combines nonnegative matrix factorization (NMF) with time differences of arrival (TDOA) estimations while eliminating time expenditure to the EM algorithm's starting values being appropriately selected. The suggested approach avoids permutation ambiguity by using the NMF source model, and acoustic localization is accomplished by converting the TDOA. Following that, model parameters are changed to improve separation outcomes. Finally, Wiener filters are used to separate the source signals. The experimental findings indicate that the suggested algorithm outperforms current blind separation approaches in terms of source separation.

Keywords: TDOA; Expectation-Maximization; Audio-Source Separation; Data Mining

1. Introduction

Source blind separation (BSS) aims to separate the original source signal from the recorded mix [1]. For acoustic signals in natural environments, for example to solve cocktail party problems [2], the mixing process is generally considered to be convolutive. A more difficult situation arises when the number of source signals exceeds the number of microphones; however, the situation is not highlighted in this case, and the mixing condition is repeatedly applied to the signals. In this context, the undefined convolutive BSS has been identified as a serious issue in speech and audio processing that requires more investigation.

ICA is a method that is the conventional strategy for handling BSS difficulties in the stated or prescribed situation, which is when the number of source signals is less than or equal to the number of microphones. ICA is also known as independent component analysis (ICA) in certain circles. ICA is based on the assumption of source component independence. Many ICA-based algorithms have been applied in various fields, such as biomedical, audio, mechanical engineering [3]. In the underdetermined case, sparse component analysis (SCA) is a feasible method based on the assumption that only one source is active in each time-frequency slot, and has been an effective method in blind deconvolution [4-5].

In addition, BSS problems are usually solved using a time-frequency domain approach [1,6,7], in which the observed signal in the time domain is converted into the frequency domain using a short time Fourier transform (STFT). This leads to permutation ambiguity and a source that is suboptimal splitting efficiency. To circumvent this permutation issue, multiple approaches based on non-negative matrix factorization (NMF) have been proposed for BSS [8], where the NMF model is used to simulate the resource spectrum density matrix This is because the NMF source model requires frequency band merging, which resolves the problem of permutation ambiguity. Additionally, an expectation-maximization (EM) technique is used to update and estimate the model parameters, and a Wiener filter is used to separate the source signals from the model parameters. EM, on the other hand, is a very sensitive technology to beginning values, resulting in suboptimal source separation performance. A combined TDOA and NMF estimation EM algorithm was developed. To begin, we used generalized cross-correlation with the phase transformation approach (GCC-PHAT) to estimate the source TDOA [3]. Following that, the mixing filter may be calculated using the TDOA approach. Additionally, the IS-NMF technique was used to generate the starting value for the NMF parameter [9]. We employ the fundamental truth parameter in the simulation, as it has been used in numerous comparable research' experiments [10]. The paper's primary contribution is to merge the TDOA estimating technique with the EM algorithm, eliminating the ambiguity of source permutations while addressing the EM algorithm's starting value issue. Additionally, one of the significant drawbacks of the present EM technique is the time required to update the model parameters at each iterative step. As a result, our suggested technique attempts to integrate the NMF and TDOA predictions in order to minimize time consumption associated with the EM algorithm's starting values selection. Finally, we compare our proposed algorithm's separation quality to that of advanced approaches.Meanwhile, we consider the defined cases and compare with the blind separation algorithm [9,10,16]. In the underdetermined case, we compare with the algorithm [11,12], and the source separation performance is improved according to the simulation results.

2. Literature Review

Let's use $sI\ (f\ ,\ n), i\ =\ 1,..., I\ dan\ x(f\ ,\ n)\ =\ [X_1(f\ ,\ n),\ X_2(f\ ,\ n)]^{\ T}$ to express the STFT of the ith source and its mixing, , where I signifies the number of source signals and f = 1,..., F denotes the frequency of the source signals.and time frame bin indexes, respectively. As a result, the model of source mixing may be represented as [3].

$$x(f\ ,\ n)\ =\ \sum_{I=1}^{I} d\,(f\ ,\tau_i\ \ )S_i\ \ \ (f\ ,\ n)\ +\ b(f\ ,\ n) \qquad (1)$$

Where

$$d(f\ ,\tau_i\ \ )\ =\ [1,\ e^{-2\,j\pi\, f\ \tau i}\ ]^{\ T} \qquad (2)$$

is a vector for mixing, and $b(f\ ,\ n)$ encapsulates additive noise based on Gaussian independent assumptions, stable and spatially uncorrelated noise for simplicity, such that $b(f\ ,\ n)\ {\sim}N_c(0,\ \sigma^2)$. The purpose of this research is to reconstruct the source signal using the observed signal's time-frequency domain. $x(f\ ,\ n)\ =\ [x_1(f\ ,\ n),\ x_2(f\ ,\ n)]^{\ T}$ without knowledge of the mixing vector in advance $d(f\ ,\ \tau_i)$. In the last step of source recovery, the obtained source signal, $S_I\ (f\ ,\ n), i\ =\ 1,..., I,$ is required to be converted into areversing STFT operation in the time domain.

In order to get started, let's refer to S as the audio source, and M1 and M2 as the two microphones that were utilized throughout the recording process. owing to the distance between the two microphones and the source, the sound waves produced by the S source arrive to the two microphones with a delay (TDOA). If we take the far-field assumption, the delay is proportional to the angle of incidence (considering sound waves as "plane waves")

$$\alpha\ =\ arccos\left(\frac{c\tau}{d}\right) \qquad (3)$$

where c denotes the sound speed and d denotes the microphone separation (Fig. 1). Depending on the number of microphones and the angle of arrival, the TDOA is calculated as follows.

$$\tau i \ = \ \frac{dn}{c} \, cos(\alpha si) \tag{4}$$

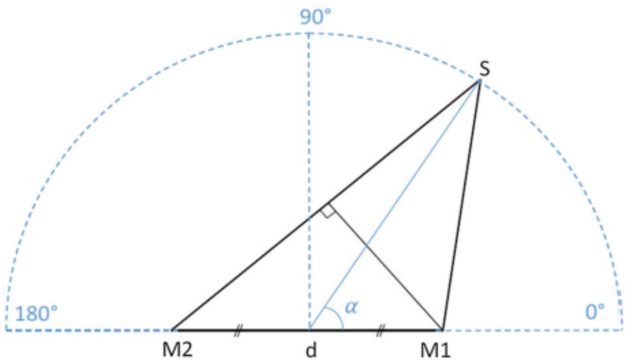

Figure. 1. Arrival angle of a sound source as determined by TDOA

As a result, the mixing vector may be expressed as follows:

$$d(f\, ,\, \tau_1) \ = \ [1,\ e^{-2\, j\pi\, f\, \tau 1}]^{\ T}$$

$$d(f\, ,\ \tau_2) \ = \ [1,\ e^{-2\, j\pi\, f\, \tau 2}\ ]^{\ T} \tag{5}$$

$$d(f\, ,\ \tau_3) \ = \ [1,\ e^{-2\, j\pi\, f\, \tau 3}\ ]^{\ T}$$

As a result, the mixing matrix may be calculated as follows:

$$d(f\, ,\tau) \ = \ [d(f\, ,\ \tau_1),\ .\ .\ .\ ,\ d(f\, ,\ \tau_1)] \tag{6}$$

Moreover, We begin by assuming that the underlying model may be stated as

$$|s_i|^{\ 2} \approx W_i H_i \tag{7}$$

where $s_I \in R^{F \times N}$ represents the STFT matrix of the i-th source, $W_i \in R^{F \times N}$ is the fundamental dictionary matrix, $H_I \in R^{K \times N}$ is the matrix of activation. Additionally, every source confirms this.

$$s_i(f\, ,\ n) \sim N_C\left(0,\ \sum_{K\, =\, 1}^{K} w_i(f\, ,\ k)hi(k,\ n)\right) \tag{8}$$

The elements of W_i and H_i are represented by the variables W_i (for fibonacci) and H_i (for kernel). As a result, we make the assumption that the components are independent throughout the frequency bins of interest (f and n). As a result, the maximum likelihood estimates wi(f, k) and hi(k, n) of Si are obtained by reducing the maximum likelihood estimates.

$$-\ log\ \ P(S_i|W_i\, ,H_i)$$

$$= \sum_{n=1}^{N} \sum_{f=1}^{F} d_{IS}(|s_i(f\, ,\ n)\ |^{\ 2}|\sum_{K=1}^{K} w_i(f\, ,\ k)hi(k,\ n) + cst \tag{9}$$

where P represents the probability density function, cst represents the constant term, and dIS represents the Itakura Saito divergence [11]

$$d_{is}(a|b \;=\; \frac{a}{b} - log\,\frac{a}{b} - 1 \qquad (10)$$

Therefore, Widan ML estimation represents the NMF decomposition from its power spectrogram. To get better source separation results, model parameters need to be updated using the EM algorithm [13]. To keep things simple, we've omitted (f, n) from each item. The following table summarizes all revised formulations:

$$d_{ji} \;=\; \frac{\frac{1}{N}\sum_{n} x_j S_{ji}}{\frac{1}{N}\sum_{n} S_{ji} S_{ji}^{H} + R_{Sji} - R_{Sji} d_{ji}^{H} R_{xj}^{-1} d_{ji} \;\; R_{Sji}} \qquad (11)$$

$$w_i \;=\; \frac{1}{N}\sum_{n} \frac{{}_{n}S_{ji} S_{ji}^{H} + R_{Sji} - R_{Sji} d_{ji}^{H} R_{xj}^{-1} d_{ji} \;\; R_{Sji}}{h_i} \qquad (12)$$

$$h_i = \frac{1}{N}\sum_{n} \frac{S_{ji} S_{ji}^{H} + R_{Sji} - R_{Sji} d_{ji}^{H} R_{xj}^{-1} d_{ji} \;\; R_{Sji}}{w_i} \qquad (13)$$

Where $R_{sji} \;=\; E[\,S_{ji} S_{ji}^{H}]$ and $R_{xj} \;=\; E[\,x_j x_j^{H}]$. The source signal is then reconstructed using Wiener filtering.

$$S_{ji}(f,n) \;=\; \frac{d_{ji}(f,f)(\sum_{K=1}^{K} w_i(f,k) h_i(k,n)) x_j(f,n)}{v_{j(f,n)}} \qquad (14)$$

is the $S_{ji}(f,n)$ component of the j mixture's reconstructed I source. $v_j(\,f\,,\,n)$ is the (f,n) element of $v_j = \sum_{i}^{1} = 1 d_{ji} W_i H_i$. A more detailed explanation is given in [14]. Finally, the source signal is converted to a time domain signal. obtained by using the inverse Fourier transform.

3. Research Model

The separation performance is mainly discussed using the signal-to-distortion ratio (SDR) and signal-to-interference ratio (SIR) [15], which are two effective evaluation measures for comparing source separation performance. We need information.the actual source of the image $S_{ij}^{img}(t)$ for theI and microphone are sources with I = 1,... j = 1,..., J, followed by the approximate $S_{ij}^{img}(t)$ can be expressed as:

$$S_{ij}^{img}(t) = S_{ij}^{img}(t) + \; y_{ij}^{spat}(t) + y_{ij}^{interf}(t) + y_{ij}^{artif}(t) \qquad (15)$$

where $y_{ij}^{spat}(t)$, $y_{ij}^{interf}(t)$ and $y_{ij}^{interf}(t)$ are error components that, respectively, indicate spatial distortion, interference, and artifacts. As a result, the SDR is represented as follows:

$$SDR_i \;=\; 10log \; 10\frac{\sum_{J=1}^{I}\sum_{t} S_{ij}^{img}(t) \;\;^{2}}{\sum_{J=1}^{I} y_{ij}^{spat}(t) + y_{ij}^{interf}(t) + y_{ij}^{artif}(t) \;\;^{2}} \qquad (16)$$

and SIR is expressed as:

$$SIR_i = 10log_{10} \frac{\sum_{J=1}^{I} \sum_t (S_{ij}^{img}(t) + y_{ij}^{spat}(t))^2}{\sum_{J=1}^{I} \sum_t y_{ij}^{interf}(t))^2} \tag{17}$$

3.2. Conditions of Experimentation and Separation Outcomes

3.2.1. Conditions of Experiment

We conducted numerical tests to assess the suggested algorithm's performance. The data were taken from the UCI Machine Learning Repository's "Underdetermined speech and mixed music" data set [16]. Table 1 summarizes the most frequently used experimental parameters.

3.2.2. Separation as a Result of a Specific Case

To begin, we examine the specified situation, that is,I equals 2, J equals 2. The microphone is located 5 cm from the source, the reverberation time is between 50 and 750 milliseconds, and the source is located 50 centimeters from the microphone. Figures 2 and 3 illustrate the simulation findings.

Table. 1. Typical experimental parameterization

Sources	*I* = 2, 3 or 4
Channel count	*J*= 2
Rate of sampling	16 kHz
Spacing between microphones	5 cm or 1 m
Types of sources	Speech and music
Time of reverberation	T60 = 50 ms ~ 750 ms
Function of the window	Hanning window
STFT frame dimensions	2048 sample (128 ms)
The rate of propagation	343 m/s

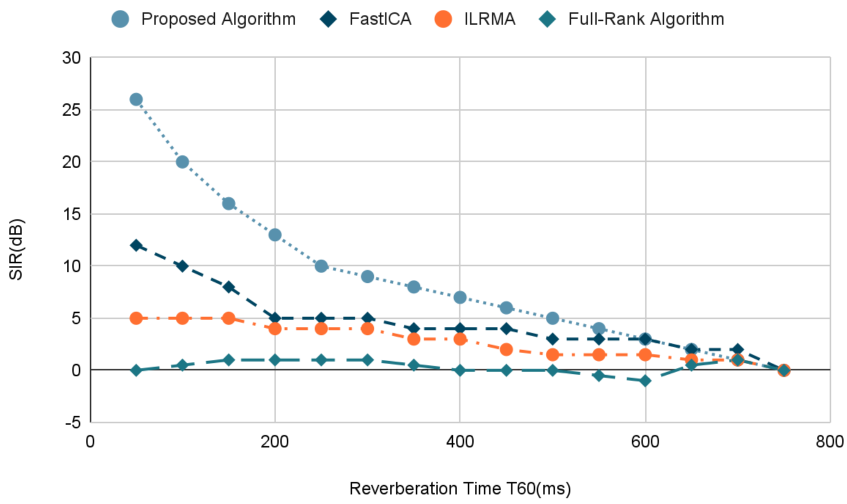

Figure. 2. The average SDR of two source mixes is 5 cm, while the standard deviation is 5 cm. as is the microphone spacing.

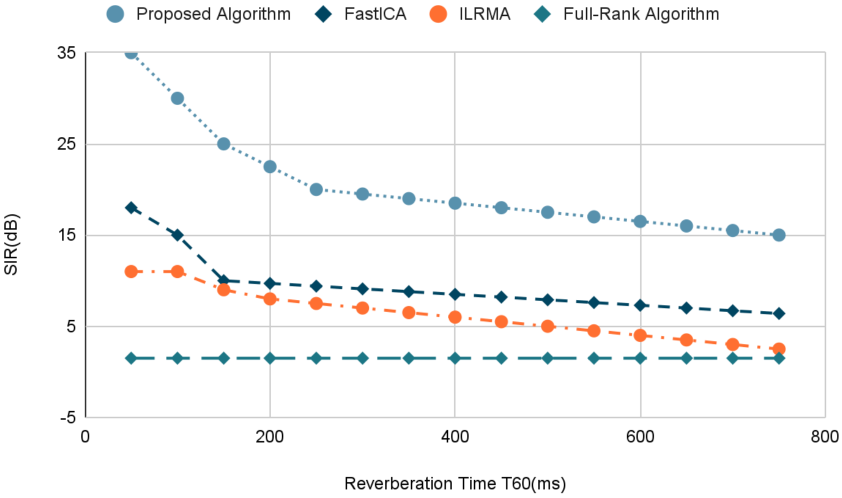

Figure. 3. Two source combinations averaged at five centimeters SIR.

So, in this study we assume the distance between the source and the microphone is 1 m. The results of the simulation are shown in Figures 4 and 5.

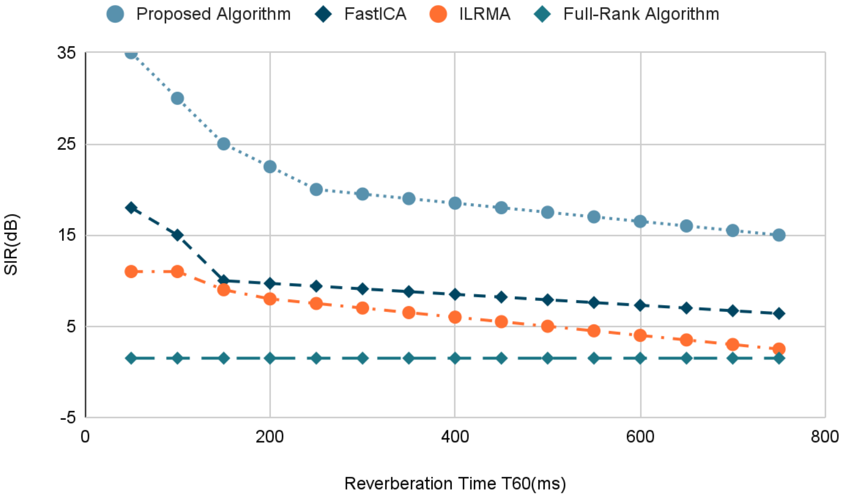

Figure. 4. The average SDR of two source mixes is 1 m, as is the microphone spacing

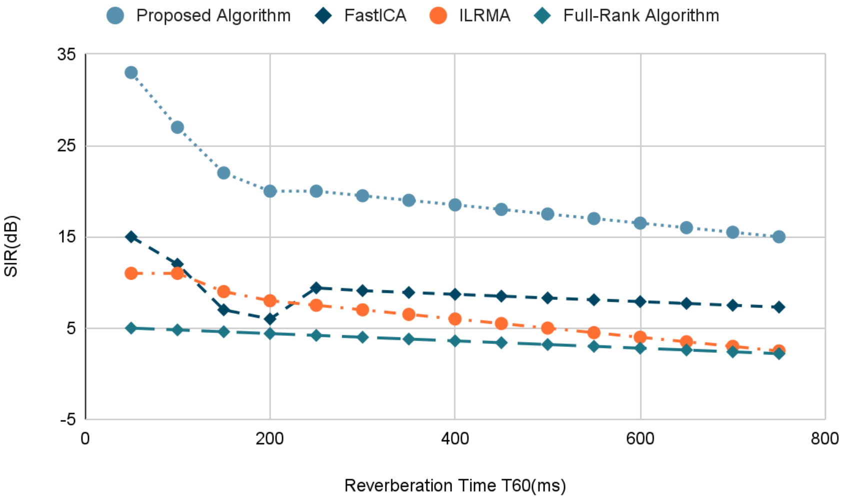

Figure. 5. The average SIR of two source mixes is one meter, and the microphone spacing is one meter as well

3.2.3. Separation Occurs in Uncertain Cases

Second, we investigate the undecided scenario, with RT60 = Reverberation time is either 130 ms or 250 ms. Tables 2, 3, and 4 illustrate the outcomes of the simulations, respectively.

Table. 2. SDR assessment of the SiSEC 2017 dataset on an average basis for the situation of three speech source mixes

Data	Range	Casing	RT60	Speech type					Proposed
Dat1	8 cm	I, J = (3, 2)	130 ms	Woman	1.17	7.27	6.50	7.10	9.81
				Man	3.33	6.42	5.83	7.11	8.20
			250 ms	Woman	3.19	5.80	5.10	5.63	9.13
				Man	3.50	4.82	4.23	5.41	7.82

Table. 3. An average-based SDR evaluation of the SiSEC 2017 dataset was performed for the condition of four different speech source combinations

Data	Range	Casing	RT60	Speech type					Proposed
Dat1	8 cm	I, J = (4, 2)	130 ms	Woman	2.52	4.31	3.1	4.50	9.45
				Man	2.32	3.76	3.37	4.16	7.52
			250 ms	Woman	1.48	3.56	3.43	3.55	8.16
				Man	1.12	3.13	2.68	3.59	6.70

Table. 4. Evaluation of the SiSEC 2017 dataset on an average basis for three different audio source combinations using SDR techniques

Data	Range	Casing	RT60	Speech type					Proposed
Dat2	3 cm	I, J = (3, 2)	Sync	Woman	-2.75	1.31	2.67	4.10	8.12
				Man	0.36	1.16	2.60	4.50	9.37
			Live	Woman	1.18	6.21	7.22	8.29	8.60
				Man	0.48	2.31	3.70	5.71	9.50

3.2.4. A Robust Investigation of the Movement of the Light Source

We next tested the dependability of our suggested approach for recording source movement bits by recording the three sources s1, s2, and s3 in the simulation space shown in Figure 6 and comparing them to the results. The space measures 6,046 m4,562 m2,535 m and has a reverberation time that has been defined. 250 milliseconds and a microphone distance of 5 cm were used in this experiment. Positions of microphones are as follows: [3,000 meters 2,000 meters 1,500 meters], [3,050 meters 2,000 meters 1,500 meters], starting positions of sources are as follows: [3.015 meters 3.226 meters 1.193 meters], [2.415 meters 3.226 meters 2.183 meters], and [2.405 meters 2.226 meters 1.183 meters], and final positions of sources are as follows: [4.015 meters 3.226 meters 1.193 meters], [3.415 meters 3.226 meters 2.183 meters], and Additionally, we studied the separation performance by measuring SDR and SIR in a short 200 ms segment and comparing the results to the baseline. The scores from each segment are averaged over all segments, and the resulting average is used to assess overall performance. The metrics are referred to as segmental SDR (SSDR) and segmental SIR (segmental SIR) (SSIR). Short-term objective clarity (STOI) measures [17] and frequency-weighted segmental radio signal-to noise ratios [18] are two more metrics that are not segmentation-dependent but may be used in conjunction with each other. The results of the experiment are shown in Figure 7.

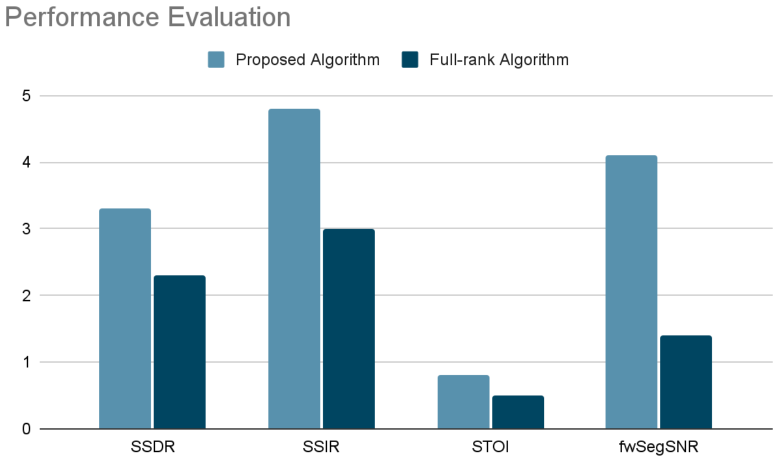

Figure. 6. SSDR, SSIR, STOI, and fwSegSNR were used to quantify separation performance.

4. Result and Discussion

To begin, we examine the described scenarios and compare the proposed approach to the full-rank algorithm [19], the FastICA algorithm [20], and the independent low-level matrix analysis (ILRMA) algorithm [21]. According to Figs. 2, 3, 4, and 5, the The mean SDR and SIR drop with increased reverberation the mean SDR and SIR decrease with increasing reverberation the reverberation the reverberation duration. Other methods' splitting performance is noticeably worse when the reverberation period exceeds 250 ms. Nonetheless, our technique is successful, indicating that our method achieves a higher level of separation performance.

Second, we investigate the undecided scenario and compare the proposed approach to the current algorithm [22-24]. The tables 2, 3, and 4 summarize the mean SDR values derived using the various approaches. It is observed that our suggested technique improves the SDR by an average of 1.77 dB, 1.08 dB, 3.18 dB, and 2.41 dB for the three sound source mixes, respectively, when compared to the existing best results. The SDR rose by an average of 4.91 dB, 3.39 dB, 4.63 dB, and 3.11 dB for the four mixed sound sources. They increased the SDR by an average of 4.92 dB, 4.86 dB, 0.31 dB, and 3.81 dB for the three music source mixes, respectively.

Thirdly, we analyze the suggested algorithm's resilience to minor Source motions are described in detail below. According to Fig. 6, the separation achieved by our recommended strategy is larger than that achieved by the full-rank algorithm. SSDR, SSIR, STOI, and fwSegSNR, for example, get a 0.80 dB boost, a 1.14 dB boost, a 0.11 dB boost, and a 2.27 dB boost, depending on the signal. These findings demonstrate that the suggested approach is resilient to source relocation of a minor magnitude. Finally, we evaluate the suggested algorithm's computational cost and some of its shortcomings. Multiplying matrices, for example, is one of the most time-consuming processes in many EM algorithms., $R_{xj} = d_{ji}.\ R_{sji}.d_{ji}$ in (11), whose complexity of computation is$O(J^3)$, and J denotes the microphone's serial number. We have lowered the computational cost of our approach correspondingly.O (I^2), where I cite the source. The following explains why. Calculating the inversion of the current EM algorithms R_{xj} the matrix in each time-frequency slot takes approximately $O(J^3)$. The explicit matrix inverse formula developed by Gamers was modified by reversing the order of the coefficients, in contrast to the current technique $R_{xj} \epsilon R^{2 \times 2}$ matrix per time-frequency slot and simply vectorizing the resulting matrix. The diagonal elements of the matrix are the only ones that need to be calculated in this manner $R_{sji} = E[S_{ji}S_{ji}^H] \epsilon R^{1 \times 1}$ on (11) and the computational complexity is related linearly to the number of sources I. Thus, the computational cost of $R_{xj} = d_{ji}.\ R_{sji}.d_{ji}$ can be reduced to $O(I^2)$. It is associated with the source code I. To show our computational advantage, we used synthetic tests to compare the running time of the proposed approach to that of the classic EM algorithm. For these situations, the suggested technique takes around 18.5 minutes and the classic EM algorithm takes approximately 23.5 minutes per 500 iterations, respectively. It is shown that our suggested approach outperforms the conventional EM algorithm in terms of time and cost.

5. Conclusion

We provide an improved EM technique for undefined convolutive BSS that incorporates TDOA and NMF estimate approaches in this article. The mixing filter was detected using the TDOA estimation approach, and permutation ambiguity was handled using the NMF source model. A series of experiments demonstrate that the upgraded EM algorithm outperforms the comparable separation method in terms of source separation performance. Additionally, we evaluate a variety of potential study avenues. To begin, we compare the proposed technique, as well as the resilience of the full-rank algorithm to even little source mobility. Despite the fact that the path from source to microphone has changed dramatically over time, distinguishing moving sources has remained a tough challenge.

Second, more investigation into the difficulties of monitoring a large number of speakers who are always moving in an interior setting is equally challenging. To summarize, this will be the subject of our future study.

References

[1] A. Tharwat, “Independent component analysis: An introduction,” Appl. Comput. Informatics, vol. 17, no. 2, pp. 222–249, Jan. 2021, doi: 10.1016/j.aci.2018.08.006.

[2] S. Piepenburg, “Disc‐Based Audio‐Video Technology,” Libr. Hi Tech News, vol. 23, no. 6, pp. 27–33, Jan. 2006, doi: 10.1108/07419050610692307.

[3] N. Liu, J. Li, Q. Liu, H. Su, and W. Wu, “Blind source separation using higher order statistics in kernel space,” COMPEL Int. J. Comput. Math. Electr. Electron. Eng., vol. 35, no. 1, pp. 289–304, Jan. 2016, doi: 10.1108/COMPEL-04-2015-0172.

[4] D. Tesendic and D. Boberic Krsticev, “Web service for connecting visually impaired people with libraries,” Aslib J. Inf. Manag., vol. 67, no. 2, pp. 230–243, Jan. 2015, doi: 10.1108/AJIM-11-2014-0149.

[5] A. Maity, P. Prakasam, and S. Bhargava, “Robust dual-tone multi-frequency tone detection using k-nearest neighbour classifier for a noisy environment,” Appl. Comput. Informatics, vol. ahead-of-print, no. ahead-of-print, Jan. 2021, doi: 10.1108/ACI-10-2020-0105.

[6] A. Zimmermann and A. Lorenz, “Creating audio‐augmented environments,” Int. J. Pervasive Comput. Commun., vol. 1, no. 1, pp. 31–42, Jan. 2005, doi: 10.1108/17427370580000111.

[7] C. Todd, S. Mallya, S. Majeed, J. Rojas, and K. Naylor, “Haptic-audio simulator for visually impaired indoor exploration,” J. Assist. Technol., vol. 9, no. 2, pp. 71–85, Jan. 2015, doi: 10.1108/JAT-06-2014-0016.

[8] D. Zhang, X. Song, X. Wang, K. Li, W. Li, and Z. Ma, “New agent-based proactive migration method and system for Big Data Environment (BDE),” Eng. Comput., vol. 32, no. 8, pp. 2443–2466, Jan. 2015, doi: 10.1108/EC-03-2015-0050.

[9] S. Gul, S. Bano, and T. Shah, “Exploring data mining: facets and emerging trends,” Digit. Libr. Perspect., vol. 37, no. 4, pp. 429–448, Jan. 2021, doi: 10.1108/DLP-08-2020-0078.

[10] E. Fersini and F. Sartori, “Semantic storyboard of judicial debates: a novel multimedia summarization environment,” Program, vol. 46, no. 2, pp. 199–219, Jan. 2012, doi: 10.1108/00330331211221846.

[11] M. D. Petković, Z. H. Perić, and A. Ž. Jovanović, “An iterative method for optimal resolution‐constrained polar quantizer design,” COMPEL - Int. J. Comput. Math. Electr. Electron. Eng., vol. 30, no. 2, pp. 574–589, Jan. 2011, doi: 10.1108/03321641111101087.

[12] S. Spagnol, M. Geronazzo, D. Rocchesso, and F. Avanzini, “Synthetic individual binaural audio delivery by pinna image processing,” Int. J. Pervasive Comput. Commun., vol. 10, no. 3, pp. 239–254, Jan. 2014, doi: 10.1108/IJPCC-06-2014-0035.

[13] R. R. A., S. Reddy, and V. K. V., “Multi-path selection based on fractional cuckoo search algorithm for QoS aware routing in MANET,” Sens. Rev., vol. 39, no. 2, pp. 218–232, Jan. 2019, doi: 10.1108/SR-08-2017-0170.

[14] H. B. Valiveti, A. K. B., L. C. Duggineni, S. Namburu, and S. Kuraparthi, “Soft computing based audio signal analysis for accident prediction,” Int. J. Pervasive Comput. Commun., vol. 17, no. 3, pp. 329–348, Jan. 2021, doi: 10.1108/IJPCC-08-2020-0120.

[15] F. J. Farsana, V. R. Devi, and K. Gopakumar, “An audio encryption scheme based on Fast Walsh Hadamard Transform and mixed chaotic keystreams,” Appl. Comput. Informatics, vol. ahead-of-print, no. ahead-of-print, Jan. 2020, doi: 10.1016/j.aci.2019.10.001.

[16] M. Yasin and P. Akhtar, “Design and performance analysis of live model of Bessel beamformer for adaptive array system,” COMPEL Int. J. Comput. Math. Electr. Electron. Eng., vol. 33, no. 4, pp. 1434–1447, Jan. 2014, doi: 10.1108/COMPEL-04-2013-0117.

[17] L. Xiao, H. Kim, and M. Ding, “An Introduction to Audio and Visual Research and Applications in Marketing,” in Review of Marketing Research, vol. 10, N. K. Malhotra, Ed. Emerald Group Publishing Limited, 2013, pp. 213–253.

[18] S. Ding, A. Cichocki, J. Huang, and D. Wei, “Blind source separation of acoustic signals in realistic environments based on ICA in the time‐frequency domain,” Int. J. Pervasive Comput. Commun., vol. 1, no. 2, pp. 89–100, Jan. 2005, doi: 10.1108/17427370580000115.

[19] G. Maguolo, M. Paci, L. Nanni, and L. Bonan, “Audiogmenter: a MATLAB toolbox for audio data augmentation,” Appl. Comput. Informatics, vol. ahead-of-print, no. ahead-of-print, Jan. 2021, doi: 10.1108/ACI-03-2021-0064.

[20] C. Grecos and Q. Wang, “Advances in video networking: standards and applications,” Int. J. Pervasive Comput. Commun., vol. 7, no. 1, pp. 22–43, Jan. 2011, doi: 10.1108/17427371111123676.

[21] D. N. Kanellopoulos, “Multimedia networking issues for digital video libraries,” Electron. Libr., vol. 32, no. 6, pp. 898–922, Jan. 2014, doi: 10.1108/EL-01-2013-0009.

[22] B. Kumaraswamy and P. P G, “Recognizing ragas of Carnatic genre using advanced intelligence: a classification system for Indian music,” Data Technol. Appl., vol. 54, no. 3, pp. 383–405, Jan. 2020, doi: 10.1108/DTA-04-2019-0055.

[23] \B. J. Jansen, M. Zhang, and A. Spink, “Patterns and transitions of query reformulation during web searching,” Int. J. Web Inf. Syst., vol. 3, no. 4, pp. 328–340, Jan. 2007, doi: 10.1108/17440080710848116.

[24] D. Kanellopoulos, “Semantic annotation and retrieval of documentary media objects,” Electron. Libr., vol. 30, no. 5, pp. 721–747, Jan. 2012, doi: 10.1108/02640471211275756.