International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. #, No.# (Mon-20..)

A Novel Blind Audio Source Separation Utilizing Adaptive Swarm Intelligence and Combined Negentropy-Cross Correlation Optimization

Pushpalatha G[1], B.Sivakumar[2]

[1] *Research Scholar, Department of Telecommunication Engineering, Dr. Ambedkar Institute of Technology, Bengaluru, Karnataka, India*

[2] *Professor, Department of Telecommunication Engineering, Dr. Ambedkar Institute of Technology, Bengaluru, Karnataka, India*

E-mail address: gpushpalatha2k17@gmail.com, sivabs2000@email.com

Received ## Mon. 20##, Revised ## Mon. 20##, Accepted ## Mon. 20##, Published ## Mon. 20##

Abstract: This paper presents a novel computational framework for blind audio source separation (BASS) that enhances existing Independent Component Analysis (ICA) with an adaptive swarm intelligence algorithm (ASIA). The proposed ASIA methodology addresses the challenges of optimal parameter determination in stochastic optimization process of swarm intelligence approach for an estimation of the precise unmixing matrix. In order to ensure the separated signals are as independent as possible in BASS task, a complex and non-convex optimization problem is formulated where the unmixing matrix is customized to minimize mutual information and maximize the non-Gaussianity of the signals. To solve our optimization problem the study introduces a weighted combination of negentropy and cross-correlation in the fitness function of the proposed ASIA. This unique approach of proposed framework ensures maximum statistical independence of the separated signals from the unknown mixed signals. Overall analysis of experimental outcome demonstrate that the proposed framework exhibits superior blind separation of mixed audio signals, showcasing enhanced computational efficiency and de-mixing accuracy compared to conventional baseline approaches. This paper has presented unique approach to blind audio source separation in over-determined scenario that combines adaptive PSO with ICA. The main goal of the proposed approach was to find an optimal de-mixing matrix that could efficiently separate mixed signals. The presented approach incorporates an adaptive inertia weight and velocity clamping mechanism into the traditional PSO, which effectively addresses the challenges associated with parameter determination in stochastic optimization techniques

Keywords: Audio Signal; Mixed Signal; Blind Source Separation, ICA; Swarm optimization

1. Introduction

Blind Audio Source Separation (BASS) is a powerful technique that is used to extract individual audio sources from a mixture of sounds. It is widely used in many auditory signal processing applications such as speech enhancement, music processing and Bioacoustics. However, separating audio signals from a mixture without knowing the source signals or mixing process poses significant challenges. This means that BASS algorithms must be able to handle the inherent complexity and non-stationarity of audio signals, each with its unique temporal signatures and frequent reverberations. Additionally, the blind nature of the separation task makes it inherently prone to inaccuracies, requiring careful design of BASS algorithms. Independent Component Analysis (ICA) has been a traditional method for Blind Audio Source Separation (BASS) due to its effectiveness in separating statistically independent sources. However, ICA's performance is often dependent on the optimal selection

E-mail:author's email

of various parameters and the underlying assumption that the sources are non-Gaussian and mutually independent. These limitations can hinder separation accuracy and computational efficiency. ICA-based source separation methods often get trapped in local optima and are not very robust against non-linear mixtures. Additionally, existing ICA-based methods often face slow convergence issues and lack precision in estimating the mixing matrix, affecting the accuracy of source separation [18].
In this paper, the research study proposes a novel computational framework that significantly enhances the conventional ICA methodology for BASS by integrating an adaptive swarm intelligence algorithm (ASIA). The proposed ASIA methodology is precisely designed to address the common pitfalls associated with optimal parameter determination in standard particle swarm optimization (PSO) algorithm and as well as to ensure reliable unmixing matrix estimation using classical ICA. The proposed ASIA employs an adaptive inertia weight and a velocity clamping parameter to fine-tune the optimization process with higher precision. To further strengthen the performance of the BASS task in the proposed framework, we introduce a weighted combination of negentropy and cross-correlation in the fitness function of the ASIA. This unique combination serves as a key-enabler in devising an effective objective function for solving our optimization problem of maximizing the statistical independence of the separated signals. The proposed objective function is derived in such a way that strategically operates the unmixing matrix to minimize mutual information while concurrently maximizing the non-Gaussianity of the signals. The incorporation of the cross-correlation is ensuring minimal similarity between the separated signals, thereby significantly enhancing the robustness and efficacy of the proposed BASS framework using ASIA driven ICA. The rest of the manuscript is organized as follows: Section II briefly review some related work in BASS context through signal processing and metaheuristic algorithms. In Section III, the proposed ASIA scheme is discussed in details. Then the obtained results are presented in Section IV and finally Section V concludes the paper with core findings and future research direction.

2. RELATED WORK

Numerous research studies have been done in the literature to address BASS problem. The existing studies includes methods such as time-frequency-masking [7], computational auditory scene analysis (CASA) [9], beamforming [8], independent component analysis (ICA) [10], and principal component analysis (PCA) [11]. Each of these methods have their own limitations and may not perform optimally in all situations. In addition, many new techniques have been introduced in audio BSS research, including non-negative matrix factorization (NMF) [12], sparse component analysis [13], dictionary learning [14], and the application of neural networks [15]. However, these methods are very sensitive to noise, an unavoidable aspect in many practical applications. Among the numerous existing methods, ICA has been widely recognized for its effectiveness in solving BASS problems. ICA aims to represent a set of mixed signals as a linear combination of statistically independent components [16, 17]. However, this assumption may not always hold true, particularly in complex audio environments. Kitamura [19] introduces a computational scheme integrating Independent Vector Analysis (IVA) and single-channel NMF to separate the mixed auditory signal into discrete components in the context of determined BSS problem. However, this approach may fail to converge to an optimal solution when introduced to complex speaker mixing problems. Leplat et al. [20], introduced an approach that combines NMF with β-divergences to measure the discrepancy between the mixed signal and its reconstruction. To encourage a compact representation in the dictionary matrix, a penalty term is employed, promoting basis vectors with reduced volume.

Mogami et al. [21] addresses blind multichannel mixed audio separation. The proposed approach combines ICA with deep learning to estimate the unknown mixing matrix and update the time-frequency structures of each source. However, the reliance on pre-trained learning models limits flexibility and generalizability of this approach. Moreover, this work lacks a thorough discussion of the computational complexity subjected to the proposed scheme. Eldin and Youssif [22] presented a hybrid scheme that combines hidden Markov model (HMM) and CASA to solve cochannel speech BSS. The HMM is applied as a preprocessing method to improve pitch tracking, pitch enrichment, and pitch grouping. Subsequently, CASA is utilized for speech separation. However, HMM is sensitive to initial conditions and assumes stationary statistical properties of the input signals. Therefore, it may lead to slow convergence issues. To improve convergence rate and obtained sub-optimal solution, Khalfa et al. [23] suggested a PSO with high-level exploration mechanism that incorporates additional operators namely crossover an application of genetic algorithms (GA) and a bee colony optimization (BCO) method, to update particle velocity and position.

The approach demonstrates robustness in BSS based on experimental results. However, the utilization of these additional operators, significantly increases algorithm complexity and increase algorithm response time. Salman et al. [24] suggested lightweight PSO driven ICA with the objective of optimizing the mutual information function for speech source separation. The PSO method implemented in this work is Quantum PSO which is very sensitive to initial parameters setup such as quantum gates and quantum rotations, which itself is time consuming empirical analysis. The work carried out by Zi and Lv [28] have studied performance of the several swarm-based optimization scheme to solve BASS problem. There are also many research works carried out in similar direction by applying metaheuristic such as Xia et al. [25], used butterfly optimization algorithm, Abbas, and Salman [26], adopted an approach of Quantum PSO, while Lee and yang [27] used gravitational PSO. Despite many significant research efforts in the field of audio BSS, there remains a significant gap between the theoretical capabilities of the above discussed methods and their practical performance in real-world. This is because of following reasons

- The ill-posed nature of the BASS: In the BASS task there are infinitely many possible solutions to the problem, given a mixed signal. This means that it is difficult to find a unique solution that is also accurate.
- The complexity of audio signals: Audio signals are complex and non-stationary, which makes it difficult to design algorithms that can effectively separate them.
- The presence of noise: In real-world applications, the mixed signal is often corrupted by noise and interference. This makes it even more difficult to separate the source signals.

The existing statistical techniques and metaheuristic optimization methods are versatile and can be applied to a wide range of problems. However, when used to solve the BASS problem, their effectiveness often diminished due to the need for significant modifications to handle the complexities of audio signals. A critical aspect of this challenge is accurately estimating the true mixing matrix, which is of significant practical importance in BASS. To address these limitations, we propose a comprehensive and flexible system and methodology explicitly tailored to address the complexities of BSS in audio environments. Our approach is meticulously designed to provide a precise, customized solution for BSS, with a particular focus on addressing the intricacies presented by complex audio environments.

3. Methodology

In this section, the study first details on the theoretical basis for the proposed ASIA algorithm and then discusses the implementation procedures to address BASS problem effectively.

A. *Mathematical Model*

Consider n source signals represented by $S(t) \in [s_1, s_2, s_3, \cdots, s_n]$ where s_i refers to the i^{th}source signal. If each signal in $S(t)$ is statistically independent of each other, then the mixing model is described as follows:

$$X(t) = A \times S(t) + N(t) \tag{1}$$

Where, $X(t) \in [x_1(t), x_2(t), x_3(t), \cdots, x_n(t)]^T$is a vector of observed mixed signals at time t, T is transpose operator, A is mixing-matrix of size $n \times n$, and $N(t)$ is a vector representing the noise at time t, such that: $N(t) \in [n_1(t), n_2(t), n_3(t), \cdots, n_n(t)]^T$. Thus, from Equ (1) it is clear that the observed mixed signal $X(t)$ is a linear combination of the signal $S(t)$ and noise $N(t)$. Therefore, the prime aim in the BASS task is to separate signal $X(t)$ and reconstruct or recover the $S(t)$, without knowing the actual mixing matrix A. The unmixing model in BASS task can be described as follows:

$$Y(t) = W \times X(t) \tag{2}$$

Where, $Y(t) \in [y_1(t), y_2(t), y_3(t), \cdots, y_n(t)]^T$ is an output vector consisting of recovered and separated signals $y_i(t)$ which should be similar to the original source signal $S(t)$, W is the mixing matrix, which is an approximated form of actual mixing matrix A. However, in the real-world, estimating W precisely is quite challenging task as A is unknown.

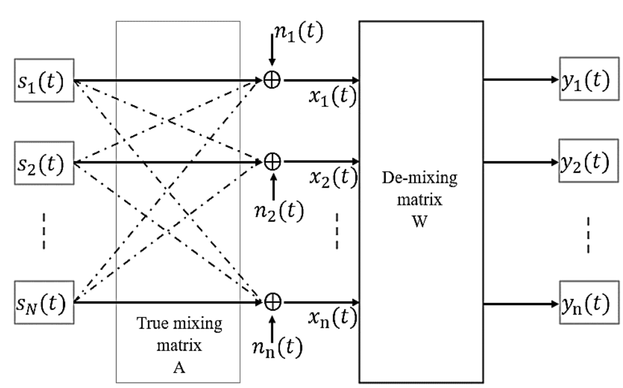

Figure 1. Schematic illustration of the BASS process

Figure 1 provides a visual representation of the above-described BASS model. It clearly depicts how multiple source signals, $s_i(t)$ are mixed with noise $n_i(t)$ through the true mixing matrix A to produce the observed mixed signals $x_i(t)$. The unmixing or de-mixing matrix W aims to recover the original source signal as $y_i(t)$ from the mixed signals.

B. ICA-based BASS Solution

ICA is one of the popular statistical methods for separating mixed signals into their individual components. It assumes that the mixed signals are statistically independent, and are non-Gaussian, meaning that they have no correlations or shared statistical properties. ICA algorithm employs two common statistical measures, namely negentropy (NE) and mutual information (MI), for the quantification of non-Gaussianity and statistical independence in the separated signal $Y(t)$. The non-Gaussianity is a vital attribute of ICA for the isolation of mixed signals into distinct components, while MI assesses the degree of mutual dependence between the separated signals. The process involves maximizing NE and minimizing MI to enhance the non-Gaussian characteristics and statistical dependence of the signal, thereby facilitating the separation procedure. The procedure of ICA is detailed in Algorithm-1.

The algorithm-1 takes mixed audio signals $X(t)$ and number of iterations T as its input and takes mixed audio signals $Y(t)$. It begins with preprocessing operations namely data centering and whitening. Data centering eliminates first-order statistics from $X(t)$ and whitening ensures the data has unit variance. Next, an unmixing matrix $\mathcal{W}$ of size $N \times N$ is initialized randomly due to the blind nature of the problem. The algorithm then estimates the separated signals $Y_E(t)$, computes NE i.e., $\mathcal{J}(y_i)$ for each signal in $Y_E(t)$, where, $H(y_i^G)$ is the entropy of a Gaussian random variable with the same covariance as the separated signal y_i and $H(y_i)$ is the entropy of the separated signal y_i. It then updates the values in $\mathcal{W}$, and normalizes the columns of $\mathcal{W}$. The algorithm checks for convergence by computing the MI between the separated signals and checking if the maximum number of iterations has been reached. The quality of the estimated separated signals $Y(t)$ is ensured by checking the convergence criteria, which include computing MI between separated signals and checking if the maximum number of iterations and convergence criteria have been met.

Algorithm-1 Estimation of separated signals using ICA

Input: $X(t)$ (mixed audio signals), T (number of iteration)
Output: $Y(t)$ (separated audio signals)
Start

1.Preprocessing
Center and whiten the mixed signals $X(t)$
2. Initialize random unmixing matrix: $\mathcal{W}$ of size $N \times N$
3. For each ICA iteration i = 1:T, do
4. Estimate the separated signals: $Y_E(t) = \mathcal{W} \times X(t)$
5. Compute NE ($\mathcal{J}$) for each separated signal to measure
non-Gaussianity: $\mathcal{J}(y_i) = H(y_i^G) - H(y_i)$
6. Update the unmixing matrix: $\mathcal{W} = \arg max_{\mathcal{W}} \sum_i \mathcal{J}(y_i)$
7. Normalize the column of $\mathcal{W}$
8.Check convergence

9. COMPUTE THE MI BETWEEN THE SEPARATED SIGNALS
$MI(Y_E(t)) = \sum_i H(y_i) - H(Y_E(t))$
10. CHECK: THE MAXIMUM NUMBER OF ITERATIONS IS REACHED AND CONVERGENCE CRITERIA ARE MET

If yes
go to step 11
Otherwise,
go back to step 4.

11. **RETURN THE SEPARATED SIGNALS:** $Y(t)= Y_E(t)$
End

C. Challenges with ICA in BASS

The prime objective of ICA is to find the unmixing that minimizes mutual information (MI), thereby maximizing independence and enhancing the non-Gaussianity of the separated signals. This objective can be mathematically represented as an optimization problem:

Optimization problem 1: Maximize the sum of NE ($\mathcal{J}$) for the *i*-th separated signal, such that:

$$\max_{W} \sum_{i=1}^{N} \mathcal{J}(y_i(t)) \tag{3}$$

Where, $\mathcal{J}(y_i(\mathrm{t})) = \mathrm{H}(y_i{}^G(\mathrm{t})) - \mathrm{H}(y_i(t))$ and $\mathrm{H}(y_i{}^G(\mathrm{t}))$ is the entropy of a Gaussian random variable with the same covariance as $y_i(t)$ and $\mathrm{H}(y_i(t))$, is the entropy of $y_i(t)$.

Optimization problem 2*:* Minimize the MI between the separated signals, such that:

$$\min_{W} MI(Y(t)) \tag{4}$$

Where, $MI(Y(t) = \sum_{i=1}^{N} \mathrm{H}(y_i(t)) - H(Y(t))$ is the mutual information between the separated signals which measures their statistical dependence. The solution to this optimization problem will give the optimal unmixing matrix that separates the mixed signals into their independent components. However, achieving solution to these optimization problems to find an optimal unmixing matrix is not without its challenges, especially due to the following reasons.

- *Blind nature of solution*: ICA is a powerful technique for BASS, but it is often difficult to find an optimal unmixing matrix. This is because the problem is blind (lack of access to ground truth source signals). The accuracy of the unmixing matrix significantly affects the quality of the recovered signals.
- *Sensitivity to Initialization*: As shown in the Algorithm-1, the unmixing matrix is initialized randomly, which can affect the solution that the algorithm converges to, in complex search space with multiple local optima.
- *Getting Stuck in Local Optima*: The search space for ICA is complex with multiple local optima. If the algorithm converges to a local optimum, the separated signals may still exhibit mixtures of the original sources, compromising the independence of the separated signals.
- *Parameter Selection*: The performance of ICA can be sensitive to the choice of algorithm parameters such as the number of components and learning rate.
- *Assumption of Non-Gaussianity*: ICA relies on the non-Gaussian nature of the source signals. If the signals are Gaussian or close to Gaussian, the separation might be inaccurate

D. Need for Optimization in ICA

The challenges faced in ICA, primarily the problem of local optima, highlight the need for robust optimization techniques. Optimization is vital in ICA for BASS as it helps navigate the complex landscape to find the global optimum to estimate optimal unmixing matrix, that yields independent components that align with the true underlying sources. Considering Equ (3) and (4), the optimization problem can be re-formulated as follows:

$$\begin{cases} \underset{W}{\text{maximize}} \sum_{i=1}^{N} \mathcal{J}(y_i(t)) \\ \text{subjected to: } MI(Y(t)) \leq \in \end{cases} \qquad (5)$$

Where, $\in$ is a small threshold value. The problem arises when this optimization landscape is highly non-convex with many local optima.

In an ideal situation, the optimization landscape is subjected to a single global maximum that corresponds to the true underlying sources. However, in practice, the optimization landscape or solution space have multiple local maxima and minima, making it difficult for the optimization algorithm to find the global maximum. Therefore, the problem of ICA getting stuck in local optima can be understood in terms of the solution space that ICA is trying to navigate. Mathematically, a local optimum refers to a point W_{local} such that there exists $\in > 0$ where:

$$O(W_{local}) \geq O(W), \forall W \in \mathcal{B}(W_{local}) \qquad (6)$$

But there exists some global optimum W_{global} such that:

$$O(W_{global}) > O(W_{local}) \qquad (7)$$

Where, $O = \sum_{i=1}^{N} \mathcal{J}(y_i)$, $\mathcal{B}(W_{local})$ is the ball of radius $\in$ centered at W_{local}. This means that W_{local} is a local maximum within a small neighborhood, but not necessarily the global maximum. When the algorithm gets stuck at W_{local}, it fails to find the true optimal solution W_{global} (optimal unmixing matrix) that maximizes the non-Gaussianity of the separated signals. This means that the separated signals $Y = W_{local} X$ may not represent the true underlying sources, thus affecting the performance of ICA in blind source separation.

E. PSO as an Optimization Technique

PSO is a meta-heuristic optimization algorithm based on the intelligence of swarms that have ability to explore the search space effectively and avoid getting stuck in local optima. PSO achieves this by maintaining a population of swarm of particles $\{\mathcal{P}_i\}_{i=1}^{N}$, traverse the search space with each particle's position representing a candidate solution. Each particle moves through the search space based on its own experience and the experience of neighboring particles, allowing for a balance between exploration and exploitation. The movement of each particle is guided by its personal best-known position, $pbest_i$, and the global best-known position $gbest$ among all particles in the swarm.

PSO utilizes a fitness function that evaluate how good a solution is. Here, the objective function captures the essence of the optimization problems in Equ (3) and (4) as follows:

$$f(W) \leftarrow \max_{W} = \left(\sum_{i=1}^{N} \mathcal{J}\big(y_i(\mathrm{t})\big) - \lambda MI(Y(t))\right) \qquad (8)$$

Where, λ is a weight parameter that balances the two objectives of maximizing non-Gaussianity $\mathcal{J}$ and minimizing MI. The fitness function $f(W)$ then becomes evaluating the quality of each candidate solution, i.e., the value of W that maximizes this objective function. Then the optimal unmixing matrix, W_{global}, corresponds to the global best position in the swarm, such that:

$$W_{global} = arg \max_{W} f(W)$$

(9)

The iterative process of PSO facilitates a balance between exploration and exploitation in the search space, converging towards W_{global} by leveraging the collective intelligence of the swarm. The rationale behind considering PSO algorithm for optimization is that it is simple to implement and often converges to the optimum solution faster than other optimization algorithms.

However, PSO is not without its own shortcomings, with a significant drawback being the risk of particles overshooting the global optimum. In addition to addressing these challenges, it is also crucial to consider the limitations that arise from relying exclusively on NE and MI as objective functions in the optimization process. While NE and MI are vital for ensuring that the separated signals are non-Gaussian and independent, respectively, they do not necessarily guarantee that these signals are accurate representations of the original sources. This discrepancy can lead to a false positive scenario, where the algorithm might incorrectly suggest successful source separation, despite the separated signals lacking meaningful correlation with the original signals. Therefore, the proposed algorithm ASIA considers incorporating mechanism of adaptive inertia weight and velocity clamping in the PSO algorithm.

F. Proposed ASIA Algorithm For BASS

To mitigate the challenges associated with ICA and PSO in BASS, this research study proposes ASIA that integrates adaptive swarm intelligence and combined negentropy cross-correlation (CC) optimization approach. Incorporating CC helps to validate that the separated signals are true representations of the original sources. If the cross-correlation between separated signals and original mixed signals is low, it indicates that the separated signals accurately represent the original sources without any mixing. The computation of CC to ensures better separation quality is given as follows:

$$C(Y,X) = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}|corr(y_i(t), x_j(t)) \tag{10}$$

Where, $corr(y_i(t), x_j(t)$ is the CC between i-th component of separated signal $Y(t)$ and j-th component of original mixed signals $X(t)$, and N denotes the number of components in the signals.

1) Proposed Fitness Function: The fitness function of proposed ASIA algorithm is a weighted combination of (NE), (MI), and CC to find the optimal unmixing matrix W that maximizes the objective function, thereby ensuring that the separated signals are as independent, non-Gaussian, and accurate representations of the original sources as possible. Therefore, using Equ (10), the fitness function is updated as follows:

$$f(W) \rightarrow \max_{W} = (\sum_{i=1}^{N}\mathcal{J}(y_i) - \lambda_1 MI(Y) - \lambda_2 C(Y,X)) \tag{11}$$

Where, λ_1 and λ_2 are the regularization parameters that control the trade-off between maximizing the non-Gaussianity, minimizing the MI, and minimizing the CC.

2) Adaption to PSO: The performance of PSO depends on the number of particles, and the number of iterations, with the potential for variations in particle velocities to result in overshooting the global optimum and lacks a convergence guarantee to W_{global}. The proposed ASIA addresses this by incorporating adaptive inertia weight and a velocity clamping parameters to fine-tune the optimization process of swarm intelligence with higher precision, thereby ensuring a balanced exploration-exploitation trade-off. The adaptive inertia weight w, is dynamically adjusted during the optimization process to balance global and local search abilities, calculated as follows:

$$w(t) = w_{max} - \left(\frac{(w_{max} - w_{min})}{iter_{max}}\right) \times iter_{current} \tag{12}$$

Where, w_{max} and w_{min} are the maximum and minimum bounds for the inertia weight, respectively. The velocity clamping is used to restrict the particle's velocity within a predefined range to prevent overshooting. The updated velocity is calculated as follows:

$$v_i' = \min\left(\max\left(v_i - v_{clamp}\right), v_{clamp}\right) \tag{13}$$

Where, v_i is the original velocity and v_i' is the clamped velocity of particle i and v_{clamp} is a velocity clamping parameter.

Algorithm-2 Estimation of separated signals using ASIA

Input: Mixed audio signals (*X*), Maximum iterations ($iter_{max}$), Number of particles (*N*)
Output: Separated audio signal (*Y*)
Start

1. **Initialize Parameters:**
   w_{max} (maximum inertia weight), w_{min} (minimum inertia weight), λ_1 and λ_2 (regularization parameters), v_{clamp} (velocity clamping parameter), c_1 and c_2 (learning rates), W (random unmixing matrix)

2. **Initialize Swarm:**

Initialize a swarm of N particles $\{\mathcal{P}_i\}_{i=1}^{N}$ with random positions and velocities.

3. **Evaluate Fitness:**
   For each particle, evaluate its fitness using
   $$f(W) = \left(\sum_{i=1}^{N} \mathcal{J}(y_i) - \lambda_1 MI(Y) - \lambda_2 C(Y, X\right)$$
   Where $Y = W \times X$.
4. **Initialize Best Positions:**
   Initialize personal best positions $pbest_i$ for each particle
   Initialize global best positions $gbest$
5. **Optimization loop**
   For $iter$ = 1 to $iter_{max}$
   For each particle i= 1 to N:
   Update the inertia weight $w(t)$ using Equ (12)
   Update particle velocity:
   $v_i(t+1) = w(t) \times v_i(t) + c_1 \times \alpha + c_2 \times \beta$
   where, $\alpha = rand() \times (pbest_i - pos_i)$
   $\beta = rand() \times (gbest - pos_i)$
   Clamp velocity:
   $v_i(t+1) = \min\big(\max\big(v_i(t+1) - v_{clamp},, v_{clamp}\big)\big)$
   Update particle position
   $pos_i(t+1) = pos_i(t) + v_i(t+1)$
   Evaluate new fitness $f(W)$ updated position
   Update $pbest_i$ if new fitness is better.
   Update $gbest$ if new fitness is better than current
   If stopping criteria are met, break the loop
   End of For
6. **Optimal Unmixing Matrix:**
   The optimal unmixing matrix W_{global} is the position corresponding to $gbest$
7. **Separate Audio signals using** $Y = W_{global} \times X$
8. **Return Output:** separated audio signals Y

End

The above-mentioned algorithm-2 begins by setting up the necessary parameters and the fitness of each particle is assessed by calculating the weighted sum of negentropy, mutual information, and cross-correlation between the separated and mixed signals. The algorithm utilizes a loop structure where, for a specified number of iterations, the inertia weight, particle velocity, and position are updated in a way that optimally balances exploration and exploitation of the search space. Once the optimization loop is completed, the global best position is used to extract the optimal unmixing matrix. This matrix is then used to separate the mixed audio signals into individual audio signals.

IV. Result And Discussion

The design and development of the proposed ASIA model is done using python programming language executed on Anaconda distribution installed on windows 10 machine. This presents the outcomes for the experimental analysis carried with different test cases of mixed auditory signals. The study considers male and female voice signal from SiSEC-08, dev2 dataset. The performance assessment is conducted in terms of both visual analysis and numerical outcome analysis. For numerical analysis, the study considers three statistical parameters namely SIR (Signal-to-Interference Ratio), SAR (Signal-to-Artifacts Ratio) and SDR (Signal-to-Distortion Ratio).

SIR: This metric quantifies the level of the desired signal in relation to the interference caused by other signals, computed as follows:

$$SIR = 10 \times log_{10}\left(\frac{P_{desired}}{P_I}\right) \tag{14}$$

Where, $P_{desired}$ is the power of desired signal, and P_I is the power of the interference from other signals

SAR: This metric assesses the quality of the separated signal by measuring the ratio of the desired signal to the artifacts introduced during the separation process.

$$SAR = 10 \times log_{10}\left(\frac{P_{desired}}{P_A}\right) \tag{15}$$

Where, P_A is the power of the artifacts introduced during the separation process

SDR: This metric provides a comprehensive evaluation by measuring the ratio of the desired signal to the distorted signal post-separation.

$$SDR = 10 \times log_{10}\left(\frac{P_{desired}}{P_{distortion}}\right) \tag{16}$$

Where, P_D is the power of the distorted signal post-separation.

A. *Test Case 1: Three male speaker*

In this test case, a mixed signal was created using three male speech signals. The proposed ASIA algorithm was then applied to separate the mixed signal, and its performance was compared with the Fast-ICA algorithm, ICA-PSO and the NMF method. Fast-ICA is a computational method used to separate a multivariate signal into additive, independent non-Gaussian signals.

NMF is a group of algorithms in multivariate analysis and linear algebra, where a matrix *V* is factorized into two matrices *W* and *H,* with the property that all three matrices have no negative elements. Table I illustrates the comparative analysis of the SDR for three male audio mixed signals.

TABLE I. ANALYSIS OF SDR FOR 3 MALE AUDIO MIXED SIGNAL

Methods/Signals	S1	S2	S3
Fast-ICA	34.75	36.16	30.98
NMF	-19.8	-25.18	-3.55
ICA-PSO	40.5	37.2	35.5
ASIA (PROPOSED)	47.86	38.19	43.02

The results demonstrate that the proposed ASIA algorithm outperformed both Fast-ICA and NMF across all the three signals. The SDR scores of the proposed method for S1, S2, and S3 were 47.86, 38.19, and 43.02, respectively. These scores are notably higher compared to Fast-ICA and significantly surpass the negative scores achieved by NMF. The lower scores of NMF indicate a considerable amount of distortion in the separated signals. This substantial difference in performance is attributed to the augmented capabilities of the ASIA algorithm. By integrating adaptive swarm intelligence and combined negentropy cross-correlation (CC) optimization approach, ASIA ensures not only the non-Gaussianity and independence of the separated signals but also verifies that the separated signals are true representations of the original sources.

TABLE II. ANALYSIS OF SIR FOR 3 MALE MIXED SPEECH SIGNALS

Methods/Signals	S1	S2	S3
Fast-ICA	34.78	36.16	30.99
NMF	9.28	3.4	-12.46
ICA-PSO	45	40	39
ASIA (PROPOSED)	49	39	45

The above-mentioned Table II illustrates the comparative analysis of the SIR for three male audio mixed signals. SIR scores demonstrate the effectiveness of ASIA in minimizing interference, outperforming Fast-ICA and significantly outpacing NMF.

TABLE III. ANALYSIS OF SAR FOR 3 MALE MIXED SPEECH SIGNALS

Methods/Signals	S1	S2	S3
Fast-ICA	46.56	69.56	65.36
NMF	-19.31	-23.54	-21
ICA-PSO	70	75	72
ASIA (PROPOSED)	66.77	94.09	83.27

The above-mentioned Table III illustrates the comparative analysis of the SAR for three male audio mixed signals. The SAR scores confirm that ASIA minimizes artifacts in the separated signals, with the highest scores across all signals. In contrast, NMF yielded substantially lower scores, implying significant artifacts in the separated signals.

B. Analysis of Test Case 2: Three female speaker

In test case 2, we evaluated the performance of our proposed ASIA algorithm on a mixed signal composed of three female speech signals. We analyzed the results using the SDR, SIR, and SAR metrics, with the outcomes presented in the following tables. The results in Table IV demonstrate that the proposed method significantly outperformed both Fast-ICA and NMF in terms of SDR scores, indicating its superior ability to reduce signal distortion.

TABLE IV. ANALYSIS OF SDR FOR 3 FEMALE AUDIO MIXED SIGNAL

Methods/Signals	S1	S2	S3
Fast-ICA	30.15	35.23	26.46
NMF	4.49	6.35	8.48
ICA-PSO	38	42	31
ASIA (PROPOSED)	45.25	55.32	34.62

The results in Table IV demonstrate that the proposed method significantly outperformed both Fast-ICA and NMF in terms of SDR scores, indicating its superior ability to reduce signal distortion.

TABLE V. ANALYSIS OF SIR FOR 3 FEMALE MIXED SPEECH SIGNALS

Methods/Signals	S1	S2	S3
Fast-ICA	31.61	35.95	26.1
NMF	5.21	7.39	10.16
ICA-PSO	40	45	32
ASIA (PROPOSED)	42.26	55.5	34.63

As shown in Table V, the proposed method exhibited superior interference minimization capabilities, achieving higher SIR scores compared to Fast-ICA and significantly outperforming NMF.

TABLE VI. ANALYSIS OF SAR FOR 3 MALE MIXED SPEECH SIGNALS

Methods/Signals	S1	S2	S3
Fast-ICA	53.54	59.71	55.53
NMF	13.78	13.79	13.8
ICA-PSO	63	68	60
ASIA (PROPOSED)	68.87	69.23	65.42

Table VI further validates the efficacy of our method, with the proposed ASIA algorithm achieving the highest

SAR scores, indicating minimal artifacts in the separated signals. The combined results from Tables IV, V, and VI confirm the consistent superior performance of the proposed ASIA method in extracting and preserving the quality of female speech signals, effectively

C. Analysis of Test Case 3 : Single female and male speaker

This section presents a detailed analysis of test case 3, where the mixed signal was composed of a single female and a single male speech signal. The SDR, SIR, and SAR metrics were employed to gauge the performance of the different algorithms. The outcome analysis shown in Table VII, the proposed algorithm distinctly outshines both Fast-ICA and NMF with regards to the SDR metric, thereby affirming its superiority in separating speech signals with minimal distortion.

TABLE VII. ANALYSIS OF SDR FOR SINGLE MALE AND FEMALE SPEAKER

Methods/Signals	S1	S2
Fast-ICA	20.15	25.23
NMF	6.86	10.34
ICA-PSO	30	35
ASIA (PROPOSED)	33.54	42.32

As illustrated in Table VIII, the proposed algorithm demonstrates impressive results in terms of SIR, with scores of 33.95 and 47.74 for S1 and S2, respectively. This underscores the algorithm's capability to effectively minimize interference in mixed-gender speech signals.

TABLE VIII. ANALYSIS OF SIR FOR SINGLE MALE AND FEMALE SPEAKER

Methods/Signals	S1	S2
Fast-ICA	21.01	27.55
NMF	8.11	12.36
ICA-PSO	32	40
ASIA (PROPOSED)	33.95	47.74

As evident in Table IX, the proposed method surpasses in the SAR metric as well, achieving scores of 43.98 and 49.79 for S1 and S2, respectively. This indicates a lower presence of artifacts in the separated signals obtained through the proposed algorithm.

TABLE IX. ANALYSIS OF SAR FOR SINGLE MALE AND FEMALE SPEAKER

Methods/Signals	S1	S2
Fast-ICA	29.56	23.49
NMF	18.25	19.06
ICA-PSO	42	47
ASIA (PROPOSED)	43.98	49.79

The extensive evaluation across three different test cases demonstrates the robustness and exceptional performance of the proposed ASIA for BASS task.

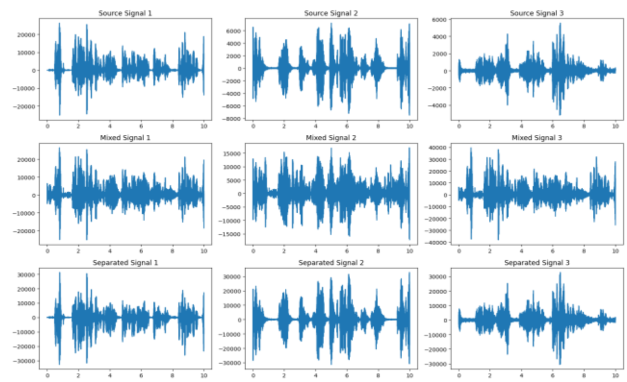

Figure 2. Visual Analysis for 3 male voice

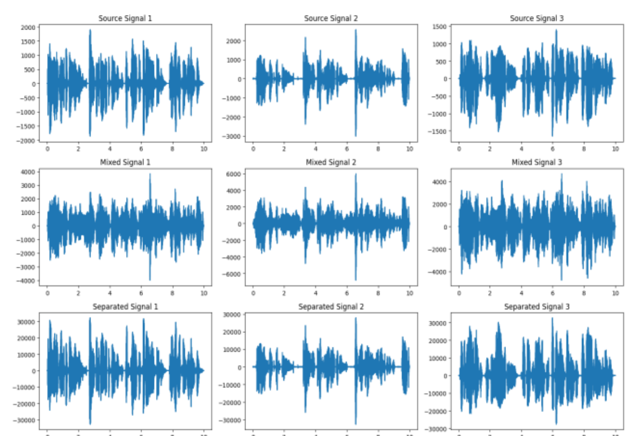

Figure 3. Qualitative analysis of test case-2: 3 female voice signals

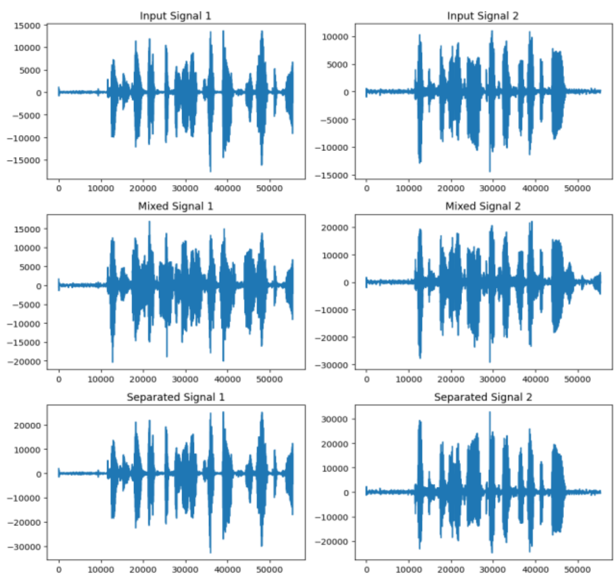

Figure 4. Qualitative analysis of test case-3:

The above-mentioned graphical representations in Figure 2, 3 and 4 depicts the progression of distinct voice signals: from their original source states, through a mixed phase, and ultimately to their separated states after being processed through an audio separation algorithm. the visual analysis of these waveforms provides valuable insights into the efficiency and precision of the audio separation algorithm. While the mixed signals manifest the intertwined complexities of the source signals, the separated signals, post-processing, bear a striking resemblance to their original counterparts. This underscores the algorithm's capability to effectively segregate individual male voice signals from a convoluted acoustic mixture, preserving the inherent characteristics of each voice with high fidelity.

Analysis of the processing time

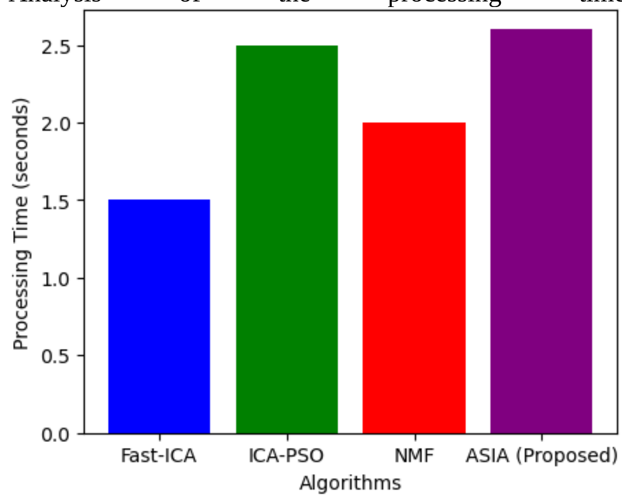

Figure 5. Analysis of the processing time

The above figure 5 illustrates the comparative processing times of four different algorithms used for BASS. The algorithms compared are Fast-ICA, ICA-PSO, NMF, and the proposed ASIA method. Based from the graph trend Fast-ICA has the shortest processing time, taking just 1.5 seconds. This is expected as Fast-ICA is known for its computational efficiency due to its simpler, less iterative approach to Independent Component Analysis. ICA-PSO shows a longer processing time of 2.5 seconds. The increased time can be attributed to the hybrid nature of this algorithm, which combines the ICA method with Particle Swarm Optimization, adding to the computational load, while NMF is represented with a processing time of 2.0 seconds, which is quicker than ICA-PSO but slower than Fast-ICA. NMF's time reflects its own iterative process of factorizing matrices, which, while complex, appears to be less so than the hybrid ICA-PSO approach in this instance. On the other hand, the proposed ASIA is the most time-consuming, with a processing time of 2.8 seconds. This suggests that the advanced integrations and optimizations within ASIA, aimed at enhancing the quality of audio separation, come with a trade-off in terms of computational time with slight variation.

V. CONCLUSION

This research offered a novel approach to blind audio source separation in an overdetermined scenario that combines adaptive PSO and ICA. The primary purpose of the suggested method was to identify an effective de-mixing matrix capable of efficiently separating mixed signals. The proposed method adds an adaptive inertia weight and velocity clamping mechanism into the standard PSO, successfully addressing the issues associated with parameter determination in stochastic optimization techniques. The ICA method is used to maximize the statistical independence of separated signals. The optimization problem is formulated as finding an unmixing matrix that minimizes mutual information or maximizes non-Gaussianity of separated signals. The enhanced PSO is then used to globally minimize this objective function. Experimental results showed that the proposed enhanced PSO-ICA approach exhibited superior performance in separating mixed audio signals compared to conventional methods. This makes it an effective and efficient solution for audio BSS problems. Future work will focus on the enhancing proposed algorithm to solve underdetermined BSS problem using more optimized approach. The results clearly highlight the enhanced computational efficiency and de-mixing accuracy of our methodology in comparison to conventional baseline approaches, thereby affirming its potential as a groundbreaking solution in the realm of blind audio source separation.

References

[1] Ortiz-Echeverri, C. J., Daniela, P. K., Galindo-Burgos, B. G., & Rodríguez-Reséndiz, J. (2018, May). Blind source separation problem algorithms for audio and biomedical signals. In 2018 XIV International Engineering Congress (CONIIN) (pp. 1-7). IEEE.

[2] Deville, Y., & Duarte, L. T. (2015, August). An overview of blind source separation methods for linear-quadratic and post-nonlinear mixtures. In International Conference on Latent Variable Analysis and Signal Separation (pp. 155-167). Cham: Springer International Publishing.

[3] Naik, G. R., & Wang, W. (2014). Blind source separation. Berlin: Springer, 10, 978-3.

[4] Luo, Z., Li, C., & Zhu, L. (2018). A comprehensive survey on blind source separation for wireless adaptive processing: Principles, perspectives, challenges and new research directions. IEEE Access, 6, 66685-66708.

[5] Bronkhorst, A. W. (2015). The cocktail-party problem revisited: early processing and selection of multi-talker speech. Attention, Perception, & Psychophysics, 77(5), 1465-1487.

[6] Qian, Y. M., Weng, C., Chang, X. K., Wang, S., & Yu, D. (2018). Past review, current progress, and challenges ahead on the cocktail party problem. Frontiers of Information Technology & Electronic Engineering, 19, 40-63.

[7] Unnikrishnan, H., Donohue, K. D., & Hannemann, J. (2014, March). Time-frequency masking for speaker of interest extraction in an immersive environment. In IEEE SOUTHEASTCON 2014 (pp. 1-8). IEEE.

[8] Zeremdini, J., Ben Messaoud, M. A., & Bouzid, A. (2015). A comparison of several computational auditory scene analysis (CASA) techniques for monaural speech segregation. Brain informatics, 2, 155-166.

[9] Wang, L., Ding, H., & Yin, F. (2014). Speech separation and extraction by combining superdirective beamforming and blind source separation. In Blind Source Separation: Advances in Theory, Algorithms and Applications (pp. 323-348). Berlin, Heidelberg: Springer Berlin Heidelberg.

[10] Kulchandani, J., & Dangarwala, K. J. (2014). Blind source separation via independent component analysis: algorithms and applications. International Journal of Computer Science and Information Technologies, 5(5), 6739-674.

[11] Islam, R., & Tarique, M. (2020). Blind source separation of fetal ECG using fast independent component analysis and principle component analysis. International Journal of Scientific and Technology Research, 9(11), 80-95.

[12] Nag, N. C., & Shah, M. S. (2021). Learning Optimum Number of Bases for Indian Languages in Non-negative Matrix Factorization based Multilingual Speech Separation. International Journal of Advanced Computer Science and Applications, 12(8).

[13] Wang, G., Wang, Y., Min, Y., & Lei, W. (2022). Blind Source Separation of Transformer Acoustic Signal Based on Sparse Component Analysis. Energies, 15(16), 6017.

[14] Zhao, X., Zhou, G., Dai, W., & Wang, W. (2014). Blind source separation based on dictionary learning: a singularity-aware approach. Blind Source Separation: Advances in Theory, Algorithms and Applications, 39-59.

[15] Drude, L., & Haeb-Umbach, R. (2019). Integration of neural networks and probabilistic spatial models for acoustic blind source separation. IEEE Journal of Selected Topics in Signal Processing, 13(4), 815-826.

[16] Duarte, L. T., & Jutten, C. (2014). Design of smart ion-selective electrode arrays based on source separation through nonlinear independent component analysis. Oil & Gas Science and Technology–Revue d'IFP Energies nouvelles, 69(2), 293-306.

[17] Gultepe, E., & Makrehchi, M. (2018). Improving clustering performance using independent component analysis and unsupervised feature learning. Human-centric Computing and Information Sciences, 8, 1-19.

[18] Du, K. L., Swamy, M. N. S., Du, K. L., & Swamy, M. N. S. (2019). Independent component analysis. Neural networks and statistical learning, 447-482.

[19] Kitamura, D., Ono, N., Sawada, H., Kameoka, H., & Saruwatari, H. (2016). Determined blind source separation unifying independent vector analysis and nonnegative matrix factorization. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 24(9), 1626-1641.

[20] Leplat, V., Gillis, N., & Ang, A. M. (2020). Blind audio source separation with minimum-volume beta-divergence NMF. IEEE Transactions on Signal Processing, 68, 3400-3410.

[21] Mogami, S., Sumino, H., Kitamura, D., Takamune, N., Takamichi, S., Saruwatari, H., & Ono, N. (2018, September). Independent deeply learned matrix analysis for multichannel audio source separation. In 2018 26th European Signal Processing Conference (EUSIPCO) (pp. 1557-1561). IEEE.

[22] Eldin, A. M. M., & Youssif, A. A. (2013). A Hybrid Approach for Co-Channel Speech Segregation based on CASA, HMM Multipitch Tracking, and Medium Frame Harmonic Model. International Journal of Advanced Computer Science and Applications, 4(7).

[23] Khalfa, A., Amardjia, N., Kenane, E., Chikouche, D., & Attia, A. (2019). Blind audio source separation based on high exploration particle swarm optimization.

[24] Li, M., Chang, Z., Zhang, L., Xu, H., Luo, Z., & Guo, R. (2022). Blind separation for wireless communication convolutive mixtures based on denoising iva. IEEE Access, 10, 113756-113766.

[25] Xia, Q., Ding, Y., Zhang, R., Liu, M., Zhang, H., & Dong, X. (2022). Blind Source Separation Based on Double-Mutant Butterfly Optimization Algorithm. Sensors, 22(11), 3979.

[26] Abbas, N. A., & Salman, H. M. (2018). Independent component analysis based on quantum particle swarm optimization. Egyptian Informatics Journal, 19(2), 101-105.

[27] Lee, S. H., & Yang, C. S. (2018). GPSO-ICA: Independent component analysis based on gravitational particle swarm optimization for blind source separation. Journal of Intelligent & Fuzzy Systems, 35(2), 1943-1957.

[28] Zi, J., Lv, D., Liu, J., Huang, X., Yao, W., Gao, M., ... & Zhang, Y. (2021). Improved Swarm Intelligent Blind Source Separation Based on Signal Cross-Correlation. Sensors, 22(1), 118.

[29] E. Vincent, S. Araki, P. Bofill, The 2008 signal separation evaluation campaign: a community-based approach to large-scale evaluation, in: Independent Component Analysis and Signal Separation, Springer, Paraty, Brazil, 2009, pp. 734–741.

Pushpalatha G currently working as Assistant Professor in the department ECE, SJB Institute of Technology, Bangalore, Karnataka. Presently pursuing doctoral program from Visveswaraiah University, Belagavi, Karnataka. She has got to her credit 2 International/National in peer-reviewed Journals and 5 Conference Papers. Also filed 1 patent. She has organized several academic programs.

Dr. B. Sivakumar has got the academic qualification as B.E., M.E., PGDBA.,Ph.D. Presently as Professor Dept. of ETE (Govt. Aided), Graduated from Madurai Kamaraj University in the field of Electronics and Communication. Also obtained his Masters degree from PSG College of Technology, Bharathiar University in the field of Applied Electronics. Has been awarded Doctoral Degree in the field of Information & Communication Engineering from Anna University, Chennai. Has got a rich teaching experience of 31 years. Has to his credit 44 International/National in peer-reviewed Journals and 120 Conference Papers. He has already guided 3 Ph.D Scholars and presently guiding 5. He has filed 2 International patent to his credit. He has completed 2 AICTE Research grant projects to the tune of 20 lakhs. Presently carrying out 3 AICTE- AQIS programs. He is an Editor /Reviewer for 4 International Journals.