Sensor placement optimization of blind source separation by using hybrid heuristic approaches in a wireless acoustic sensor network

Qingzheng Wang[1], Siow Yong Low[2], Zhibao Li[3], Ka-fai Cedric Yiu [1,*]

Abstract

Blind source separation (BSS) method separates the desired signals from a mixed observed signal by making full use of spatial information. Spatial information refers the fact that the sources originate from different location in space and thus provides the diversity for the separation. Similarly, this diversity can be further enhanced by optimizing the sensor placement as opposed to a fixed location. This paper aims to fill this research gap by proposing a sensor placement optimization strategy to further improve the performance of BSS. As the problem is non-convex in nature, a new hybrid descent optimization method is proposed by embedding a gradient-based method into the genetic algorithm. The proposed method benefits from the robustness of the genetic algorithm and the fast convergence speed of the gradient-based method. Results show that the optimized sensor placement greatly improves the separation performance of the BSS system across the different reverberation times.

Keywords: Blind source separation; Sensor array network; Hybrid descent algorithm; Genetic algorithm.

1. Introduction

The cocktail party effect which was coined by Cherry in the early fifties illustrates humans' ability to focus on a specific talker in a multi-talker situations [1]. Part of the explanation for this focussing capability lies in the spatial sampling performed by the two human ears. This spatial diversity makes use of the fact that the origins of the desired and interfering signals originate from different locations in space. With the development of communication equipments, speech

---

[*]Corresponding author
   Email address: cedric.yiu@polyu.edu.hk (Ka-fai Cedric Yiu)
   [1]Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, PR China
   [2]University of Southampton Malaysia (UoSM), Iskandar Puteri, Johor, Malaysia
   [3]School of Mathematics and Statistics, Central South University, Changsha, Hunan, PR China

separation or extraction may be realized electronically by sampling in space. A classical and typical application scenario is when speech signals are recorded by pre-mounted microphones in the wireless acoustic sensor network (WASN). Two common popular methods to perform speech extraction are blind source separation (BSS) and beamforming [2, 3, 4, 5]. The latter is very sensitive to array model mismatch [6, 7, 3] and the former requires less assumptions of the acoustic scene or system.

BSS involves extracting and recovering the underlying source signals from multivariate statistical data [8]. It is widely used in the wireless digital communication, image processing and recognition as well as geological spatial information processing. Blind implies that the source localization information is not known. Generally, BSS manifests in two general cases. The first case is the underdetermined situation where the number of sensors is less than that of sources. Another is the determined or overdetermined situation where the number of sensors is greater than or equal to the number of sources. In the first case, nonnegative matrix factorization has received much attention [9]. For the second case, independent component analysis (ICA) is the method most commonly applied but requires use of higher-order statistics [10]. As opposed to ICA, the second-order statistics (SOS) method is generally preferred to separate nonstationary signals due to its computational simplicity [11, 4, 12]. For WASN settings, sensors can be placed in any suitable position without being restricted by the wired connections and large number of sensor nodes. This largely puts WASN in the overdetermined case. However, in any reverberant environment, the situation quickly reverts to the underdetermined case.

With the assumption that source signals are statistically independent, ICA methods separate source signals by expressing a set of random variables as a linear combination of statistically independent variables [13]. ICA methods have been widely applied in many field and a variety of the ICA methods have been proposed. A neural network based ICA algorithms was presented in [14]. Hyvarinen and Oja proposed an ICA algorithm based on the negentropy maximization of random variables method [15]. Masnadi-Shirazi and Rao proposed a state-based ICA method aimed to the non-stationarity of the signal [16]. With the contribution of these pioneering authors, the theoretical framework and algorithms for ICA methods have matured. However, there are still some unsolved problems of ICA method due to its restricted assumptions or lack of information. For instance, ICA algorithms perform poorly when the high-order statistics for original signals are dependent or the stationarity condition is violated.

A SOS method has been proposed to separate nonstationary signals by joint decorrelation [17, 4]. Compared to the higher-order method (ICA), this method makes no assumption about the cumulative densities of signals and puts itself to more robust second-order statistics. Owing to second-order processing only, this method is computationally efficient. A fast convergent BSS algorithm based on the SOS method was presented in [18]. An alternative beamspace SOS method was introduced by [19] where a priori spatial information was embedded in the formulation as a preprocessor to help improve the separation performance. The performance of SOS method has been investigated in the case of reverberant

acoustic environment [20].

Araki et al. established the equivalence between BSS and a set of adaptive beamformers [5]. This means that BSS is no other than a set of spatial filters, albeit different in its formulation compared to a spatial filter. As such, the performance of BSS is highly dependent on the configuration of the sensor arrays similar to a beamformer. If so, the research question here is to ascertain if there is an optimum sensor configuration for BSS. This is akin to viewing sensor placement as a preprocessor for BSS to improve its separation capability. However, the literature thus far have been limiting as the focus on BSS has been for a fixed set of linear and non-linear array configurations [19, 21]. The sensor placement investigation is particular relevant with the rise of wireless acoustic sensor networks (WASN). In such a WASN setting, the configuration of multiple cooperative devices may be exploited further to provide a better set of sensors for the task at hand.

This paper aims to fill the research gap by studying the effect of sensors placement on the performance of BSS. Such study will enable the incorporation of optimized placement for improved spatial information to enable the best source separation performance. As opposed to focusing on a specific configuration, a hybrid descent method is proposed to optimize the performance of BSS for the best sensor placement on the allowable placement space as the number of sensors and the reverberation time (RT) change. The heuristic algorithm has a good performance on solving the discrete optimization problem and provides a good methodology to more away from local optima, but it becomes very slow when approaching to stationary points. The hybrid descent method combining the heuristic algorithm and gradient descent method was proposed to efficiently solve the continuous optimization problem [25]. In this paper, we develop a suitable hybrid descent method to tackle the design problem.

The results show that there is largely a demarcation point where the sensors placement will have a certain configuration before and after a certain RT. The results also reveal that the sensors tend to cluster on the desired source when the RT is less than 200ms and spread out otherwise. Importantly, the results consistently show a marked increase in the separation performance of an optimized placement compared to a fixed linear placement. The converged spatial locations serve as a useful guide for increasing the performance of source separation.

The rest of this paper is organized as follows. The problem formulation is given in Section 2. The ICA-based approach is introduced in Section 3. The proposed hybrid descent optimization method is introduced in Section 4. The simulation study is demonstrated in Section 5. Conclusion and potential future extensions are presented in Section 6.

2. Notation and Problem formulation

The important symbols and scientific terms in this paper are listed on Table 1. We consider an enclosed room with acoustic reverberation. The locations of N speech sources are denoted by γ_n where $n = 0, \ldots, N - 1$. The location

3

variables of M-elements acoustic sensor array are defined as δ_m, where $m = 1, \ldots, M$. Given the room dimension, sound speed, locations of sources and sensors, the time domain room impulse responses (RIR) $\mathbf{h}(\delta_m, \gamma_n)$ from the n-th source to the m-th sensor can be generated by the image method [2]. Let $\mathbf{s}_n$ denote the signal at the source γ_n. At time t, the received signal from the n-th source to the m-th sensor is given as

$$s_{m,n}(t) = \sum_{\tau=0}^{L-1} h_{m,n}(\tau) s_n(t - \tau) = \mathbf{h}(\delta_m, \gamma_n) * \mathbf{s}_n \tag{1}$$

where $h_{m,n}(\tau)$ is the $(\tau + 1)$-th element in vector $\mathbf{h}(\delta_m, \gamma_n)$; L is the length of the RIR vector; $\mathbf{s}_n$ denotes the vector $[s_n(t), s_n(t-1), \ldots, s_n(t-L+1)]$; and $*$ is the convolution operator. The observed signal at the m-th sensor is

$$x_m(t) = \sum_{n=0}^{N-1} \mathbf{s}_{m,n} = \sum_{n=0}^{N-1} \mathbf{h}(\delta_m, \gamma_n) * \mathbf{s}_n. \tag{2}$$

The observed mixtures from the WASN is denoted by

$$\mathbf{x}(t) = \mathbf{H} \odot \mathbf{s} \tag{3}$$

where $\odot$ is the element-wise convolution operator, $\mathbf{x}(t) = [x_1(t), x_2(t), \ldots, x_M(t)]^{\mathrm{T}}$ and $\mathbf{s}$ denotes the vector $[\mathbf{s}_0, \mathbf{s}_1 \ldots, \mathbf{s}_{N-1}]^{T}$. The unmixing matrix $\mathbf{H}$ consists of RIR vectors,

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}(\delta_1, \gamma_0)^{\mathrm{T}} & \cdots & \mathbf{h}(\delta_M, \gamma_{N-1})^{\mathrm{T}} \\ \vdots & \ddots & \vdots \\ \mathbf{h}(\delta_M, \gamma_0)^{\mathrm{T}} & \cdots & \mathbf{h}(\delta_M, \gamma_{N-1})^{\mathrm{T}} \end{bmatrix}.$$

Since the convolution operation in the time domain becomes multiplication in the frequency domain [22], the problem can be elegantly transformed to the simple instantaneous case if the Fourier transform (STFT) is applied on the observed signals. Rewriting Equation (3) in the frequency domain gives

$$\mathbf{X}(f, k) = \mathbf{H}(f)\mathbf{S}(f, k) \tag{4}$$

where $\mathbf{X}(f, k)$ and $\mathbf{S}(f, k)$ are the f-th subband transformations of $\mathbf{x}(t)$ and $\mathbf{s}$ respectively. $\mathbf{H}(f)$ is a matrix containing the elements of the mixing matrix $\mathbf{H}$ at the f-th subband and k is the frequency sampling index.

The objective here is to unmix the mixtures so as to recover the original signal $\mathbf{S}(f, k)$ from $\mathbf{X}(f, k)$. Assuming that the f-th subband of the mixing matrix is invertible, then the unmixing process is

$$\mathbf{Y}(f, k) = \mathbf{W}(f)\mathbf{X}(f, k) \tag{5}$$

where $\mathbf{W}(f)$ is the unmixing matrix in the f-th subband and $\mathbf{Y}(f, k)$ is the estimated separated signals in the frequency domain. Substituting Equation (4) into Equation (5), we derive

$$\mathbf{Y}(f, k) = \mathbf{W}(f)\mathbf{H}(f)\mathbf{S}(f, k). \tag{6}$$

Equation (6) shows that the separated signals $\mathbf{Y}(f, k)$ not only depends on the unmixing matrix $\mathbf{W}(f)$ in the separation system but also relates to the mixing matrix $\mathbf{H}(f)$ in the mixing matrix.

Notation	Definition
$s.(\cdot)$	source signals in time domain
$x.(\cdot)$	received signals in time domain
$\gamma.$	locations of speech sources
$\delta.$	locations of acoustic sensors
$\mathbf{h}(\cdot, \cdot)$	room impulse responses (RIR) in time domain
$\mathbf{H}$	mixing matrix in time domain
$\mathbf{H}(\cdot)$	mixing matrix in frequency domain
$\mathbf{W}(\cdot)$	unmixing matrix in frequency domain
$\mathbf{X}(\cdot, \cdot)$	observed signals in frequency domain
$\mathbf{Y}(\cdot, \cdot)$	separated signals in frequency domain
$\mathbf{R}_X(\cdot, \cdot)$	covariance matrix of the observed signals in frequency domain
$\mathbf{R}_Y(\cdot, \cdot)$	covariance matrix of the separated signals in frequency domain
$\rho(\cdot, \cdot)$	correlation function
$\boldsymbol{\delta}$	the vector representing location variables for all the sensors
$\mathbf{L}_{\boldsymbol{\delta}}$	lower bound of location variables
$\mathbf{U}_{\boldsymbol{\delta}}$	upper bound of location variables
$\hat{P}.(\cdot)$	spectral power estimate
$\mathrm{SUPP}_{\mathrm{I}}$	suppression measure
PESQ	perceptual evaluation of speech quality
T_{60}	reverberation time (RT)
T_{algo}	iteration index in Algorithm 1
J	iteration index in Algorithm 2

Table 1: The list of symbols and scientific terms

3. Second-order separation method

In order to derive the optimal unmixing matrix $\mathbf{W}(f)$, the BSS with second-order statistics method is adopted [19, 4]. The SOS method exploits the non-stationarity of the source signals to provide more information to separate the sources [4, 5]. To be specific, the covariance matrix of the non-stationary source signals, at different time intervals, are assumed to be linearly independent. As such, the additional diversity can be used to perform the separation process. In this case, the covariance matrix $\mathbf{R}_X(f, k)$ of the observed mixtures can be estimated on N successive intervals as

$$\mathbf{R}_X(f, n) = \frac{1}{P} \sum_{k=0}^{P-1} \mathbf{X}(f, nP + i)\mathbf{X}(f, nP + i)^H \tag{7}$$

where $n = 0, \ldots, N-1$; P is the length of the interval for estimating the covariance matrix; and $(\cdot)^H$ is the Hermitian transpose operator. By the projection

5

of the unmixing matrix $\mathbf{W}(f)$, the covariance matrix $\mathbf{R}_Y(f,k)$ of the separated output is calculated by

$$
\begin{aligned}
\mathbf{R}_Y(f,n) &= \frac{1}{P}\sum_{k=0}^{P-1}\mathbf{Y}(f,nP+i)\mathbf{Y}(f,nP+i)^H \\
&= \frac{1}{P}\sum_{k=0}^{P-1}\mathbf{W}(f)\mathbf{X}(f,nP+i)\mathbf{X}(f,nP+i)^H\mathbf{W}(f)^H \\
&= \mathbf{W}(f)\mathbf{R}_X(f,n)\mathbf{W}(f)^H.
\end{aligned}
\tag{8}
$$

If the source signals are successfully separated, the estimated source signal will be statistically independent which results in the covariance matrix $\mathbf{R}_Y(f,k)$ to be a diagonal matrix. Then, the BSS problem can be posed as a least square optimization problem which is formulated as follows [4],

$$
\mathbf{W}_{\mathsf{opt}}(f) = \arg\min_{\mathbf{W}(f)}\sum_{n=1}^{N}\|\mathbf{E}(f,n)\|_{\mathrm{F}}^2
\tag{9}
$$

where

$$
\begin{aligned}
\mathbf{E}(f,n) &= \mathbf{R}_Y(f,n) - \mathbf{\Lambda}(f,n) \\
&= \mathbf{W}(f)\mathbf{R}_X(f,n)\mathbf{W}(f)^H - \mathbf{\Lambda}(f,n)
\end{aligned}
$$

and $\|\cdot\|_{\mathrm{F}}^2$ is the squared Frobenius norm operator. The diagonal matrix $\mathbf{\Lambda}(f,n)$ represents the power of each source. If the field is homogeneous, all sources have the same power. Then, $\mathbf{\Lambda}(f,n) = S(f,n)\mathbf{I}$ where $S(f,n)$ is the power constant of the source and $\mathbf{I}$ is the identity matrix. The homogeneous assumption can be achieved by pre-whitening operation in [23].

With the implementation of the gradient descent method, the gradient update scheme with respect to the optimization problem in Equation (9) is denoted by

$$
\mathbf{W}^{(m+1)}(f) = \mathbf{W}^{(m)}(f) - 2\mu\sum_{n=0}^{N-1}\left[\mathbf{W}^{(m)}(f)\mathbf{R}_X(f,n)\mathbf{W}^{(m)^{\mathrm{H}}}(f) - \mathbf{\Lambda}(f,n)\right]\mathbf{W}^{(m)}(f)\mathbf{R}_X(f,n)
\tag{10}
$$

where μ is the step-size and m is the iteration index. We set The convergence analysis of this method has been well studied in [19].

4. Sensor placement optimization method

4.1. Proposed optimization strategy

The hybrid descent method is proposed to solve the long standing sensor placement problem for a blind signal separation system. The result of the investigation shows that there is an optimized placement for the sensors for BSS to
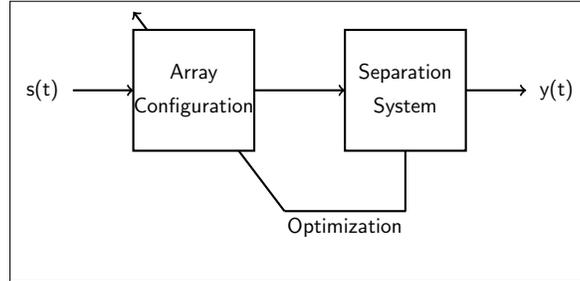
Figure 1: The proposed sensor placement optimization process where the separation system is optimized through the optimal spatial placement.

best separate the target signal. As the nature of the problem is continuous and highly non-convex, the hybrid descent method was proposed to prevent local optima. To the best of the authors' knowledge, this paper is the first to apply the hybrid descent method to solve the sensor placement optimization for the BSS system.

As mentioned in Section II, the output of the convolutive BSS model depends on both the mixing system and the separation system. The mixing system highly depends on array configurations. Given a fixed array configuration, the separation system is optimized using the SOS method. The aim of this paper is to propose an optimization technique to optimize sensor placement so as to improve the separation performance. By doing so, the full diversity of the spatial information can be fully utilized. To do so, we introduce a hybrid descent method to optimize the mixing system via the change of the sensor locations. As we can see from Figure 1, the performance of separation system is used to optimize the array configuration. The optimized array in return can improve the separation performance. The propagation of speech sound in reverberant environment is not instantaneous but rather a convolutive mixtures as a result of different room impulse responses (RIRs) between each source speech and sensor. Given the information of the speaker locations, sensor locations and other coefficients of acoustic scene, RIRs related to a small room can be modelled by using the image method [24]. By the adoption of RIRs, the mixing process of acoustic signals can be represented in Equation (2). In reality, however, we can only receive the mixed signals rather than derive the explicit form of RIRs due to the lack of speaker locations or coefficients of acoustic scene.

4.2. Optimization problem formulation

In order to optimize the mixing system without the knowledge of RIRs, we formulate the optimization problem as

$$\boldsymbol{\delta}_{\text{opt}} = \underset{\boldsymbol{\delta} \in \boldsymbol{\Delta}}{\arg\max} \ g(\boldsymbol{\delta}) \tag{11}$$

$$\text{s.t.} \quad \mathbf{L}_{\boldsymbol{\delta}} \leq \boldsymbol{\delta} \leq \mathbf{U}_{\boldsymbol{\delta}} \tag{12}$$

7

where $g(\delta) = \rho(\mathbf{y}_m, \mathbf{s}_0)$, $\boldsymbol{\delta} = [\delta_1, \delta_2, \ldots, \delta_M]^{\mathrm{T}}$ is the vector representing the location variables for all the sensors, $\mathbf{L}_{\boldsymbol{\delta}}$ and $\mathbf{U}_{\boldsymbol{\delta}}$ are respectively the lower bound and upper bound of location variables which depend on the room dimension, $\rho(\mathbf{y}_m, \mathbf{s}_0)$ is the correlation function between $\mathbf{y}_m$ and $\mathbf{s}_0$, $\mathbf{y}_m$ is m-th output of the convolutive BSS model in which $m \in [1, 2, \ldots, M]$, and $\mathbf{s}_0$ is the desired speech. Given a candidate solution $\tilde{\boldsymbol{\delta}}$, convolutive BSS method generates the optimal separated signals using the corresponding mixtures $\mathbf{x}_{\tilde{\boldsymbol{\delta}}}(t)$. Since the mixing matrix is unknown in the optimization process, the gradient related to the problem in Equation (11) cannot be expressed in an explicit form. Therefore, its global maxima cannot be directly located using a gradient-based approach and a better global optimization technique should be developed [25].

There are two major points to note when finding the global optima. The solution should avoid local optima and the optimization technique has the speed of convergence to approach stationary points. The heuristic method could find the global optima due to its capability of avoiding the local optima [26]. In reality, it is used to find the global optima without pre-defined precision due to its slow convergence rate. One of the reasons for the slow convergence is because the heuristic method becomes slow when it tries to approach or descent to stationary points. On the other hand, a gradient-based numerical method is much more efficient in finding a stationary point such that it can be used to speed up the local search. Since the location variables impact the objective value via the change of the RIR vector, it is difficult to write an explicit form of Equation (11) with respect to the location variables. The optimization problem is highly nonlinear and is essentially nonconvex with respect to the location variables. To solve the aforementioned issues, this paper proposes an efficient hybrid descent algorithm embedding a gradient-based numerical optimization algorithm into the heuristic method. The heuristic method is used to locate a descent point from a previous converged local solution and the gradient-based numerical algorithm is used to find better local optima. By doing so, we can retain the robustness of the heuristic method and the convergence speed of gradient-based numerical method.

4.3. Proposed hybrid descent algorithm

There is a variety of heuristic methods such as the genetic algorithm [27], Tabu search algorithm [28], simulated annealing algorithm [29] and so forth. The main difference between these methods is the way to traverse the whole parametric space to reach a global peak in the case of unevenly distributed, nonuniform, multiple-peak space [30]. In this paper, we propose the use of genetic algorithm. The genetic algorithm was developed in [31] to solve the sensor placement problem in the beamformer configuration design. Comparing to others, the genetic algorithm has a nice parallel computation structure in which the candidate solution is generated by the random perturbation of a population rather than by moving from one point to the next. Apart from that, a large number of independent individuals in one generation enable the genetic algorithm to traverse the whole parametric space efficiently. The proposed sensor

placement optimization method is stated in Algorithm 1. The details of genetic algorithm are illustrated in Algorithm 2. The hyperparameters used in Algorithm 1 and 2 are listed on Table 2.

Hyperparameter	2 sensors	3 sensors	4 sensors
Population size	100	120	200
StallGenerations (J_{stall})	10	10	10
MaxGenerations ($J_{\max}$)	40	20	20
Elite rate		0.05	
Crossover rate		0.8	
Mutation rate		0.2	
Function Tolerance (ζ)		1×10^{-6}	
Maximum iteration in Algorithm 1 ($T_{\max}$)		20	

Table 2: Hyperparameters in Algorithm 1 and 2

Algorithm 1 Hybrid descent method for the BSS model

1: Set the hyperparameters for the genetic algorithm and $T_{\mathrm{algo}} = 0$. Setup an initial solution $\boldsymbol{\delta}^{\mathrm{Initial}}$.
2: **repeat**
3:     Execute the genetic algorithm (Algorithm 2) to get the improved solution $\tilde{\boldsymbol{\delta}}$.
4:     Starting from $\tilde{\boldsymbol{\delta}}$, execute the sequential quadratic programming (SQP) method to derive the local optima $\boldsymbol{\delta}_{\mathrm{opt}}$.
5:     **if** $g(\boldsymbol{\delta}^{\mathrm{Initial}}) < g(\boldsymbol{\delta}_{\mathrm{opt}})$ **then**
6:         $\boldsymbol{\delta}^{\mathrm{Initial}} = \boldsymbol{\delta}_{\mathrm{opt}}$;
7:         $g(\boldsymbol{\delta}^{\mathrm{Initial}}) = g(\boldsymbol{\delta}_{\mathrm{opt}})$;
8:         $T_{\mathrm{algo}} = T_{\mathrm{algo}} + 1$;
9:     **else**
10:         break;
11:     **end if**
12: **until** $T_{\mathrm{algo}} > T_{\max}$

In Step 17 of Algorithm 2, there are two stopping criteria stated below.

- $J > J_{\max}$ where $J_{\max}$ is the hyperparameter denoting the maximum generation number in the genetic algorithm.

- $J_{\mathrm{stall}} < J \leq J_{\max}$ and $\frac{1}{J - J_{\mathrm{stall}}} \sum_{j=J-J_{\mathrm{stall}}}^{J} [g(\boldsymbol{\delta}^{(j)}) - g(\boldsymbol{\delta}^{(j-1)})] < \zeta$ where ζ is the hyperparameter denoting the function tolerance of the average change of the objective function value in the last J_{stall} generations.

It is noted that, in Step 2 of Algorithm 2, the genetic algorithm is kicked off from different initial points. Compared to the simulated annealing method, the genetic algorithm has a stronger ability to traverse the whole parametric

---
Algorithm 2 Genetic algorithm

1: $g(\tilde{\boldsymbol{\delta}}) = g(\boldsymbol{\delta}^{\text{Initial}})$; $\tilde{\boldsymbol{\delta}} = \boldsymbol{\delta}^{\text{Initial}}$; $J = 1$.
2: Generate the initial population with the constraint in Inequality 12 via the random perturbations of $\tilde{\boldsymbol{\delta}}$.
3: **repeat**
4:      Optimize the unmixing matrix $\mathbf{W}(f)$ for each individuals in the population using Equation 10.
5:      Calculate the maxima of $\rho(\mathbf{y}_m, \mathbf{s}_0)$ for each individuals where $m \in [1, 2, \ldots, M]$.
6:      Scale the values of objective function for the J-th generation.
7:      Store the optimize solution of the J-th generation $\delta^{(J)}$.
8:      Store the elite population $\mathcal{E}$ consisting of the individuals with the objective function value great than or equal to the 5-th percentile.
9:      Recombine pairs of individuals whose objective function values are between the 5-th and 81-th percentiles and store this sub-population $\mathcal{C}$.
10:      Randomly permutate the individuals whose objective function values are less than 81-th percentile and store this sub-population $\mathcal{M}$. The permutated individuals should satisfy the constraint in Inequality 12.
11:      Construct the population of the next generation as $\{\mathcal{E}; \mathcal{C}; \mathcal{M}\}$.
12:      **if** $g(\tilde{\boldsymbol{\delta}}) < g(\boldsymbol{\delta}^{(J)})$ **then**
13:         $\tilde{\boldsymbol{\delta}} = \boldsymbol{\delta}^{(J)}$;
14:         $g(\tilde{\boldsymbol{\delta}}) = g(\boldsymbol{\delta}^{(J)})$;
15:      **end if**
16:      $J = J + 1$.
17: **until** Stopping criteria have been satisfied.
---

space because the genetic algorithm can explore the search space in many directions simultaneously and it can be parallelized [32]. Three key parts of the genetic algorithm includes the evaluation of the objective function (Step 4 and 5), crossover operation (Step 9) and mutation operation (Step 10). The three parts above can be parallel computed by assigning a fraction of the population to different threads [33]. The genetic algorithm can speed up by using more threads.

5. Experimental Results and Discussions

5.1. Experimental settings

In this section, we illustrate the performance of our convolutive BSS method in a simulated room. We consider a 6m × 6m × 3m square office room with the RT (T_{60}) chosen as 0ms, 100ms, 200ms, 300ms, 400ms and 500ms. The size of our simulated room is a common specification for a conference room or a tiny lecture theater. This mimics the use of a separation system in those settings. Note that the proposed algorithm is not limited by any given dimensions of a room. The absorption coefficients for all room boundaries are set to be the same.
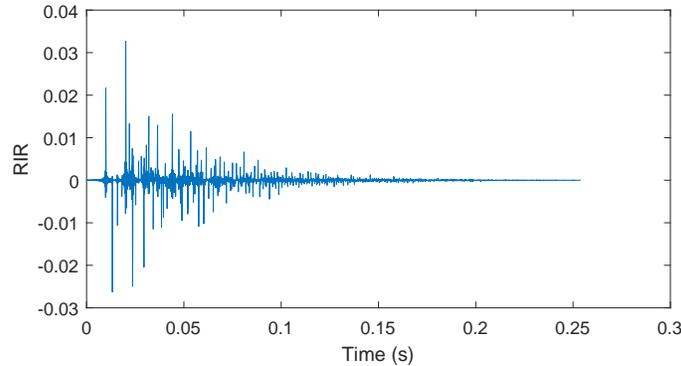
Figure 2: (Example) One RIR vector when $T_{60} = 300ms$

The sound velocity is equal to 343 m/s. The sensors and speakers are modelled by the image-source method [24]. The acoustic transfer between a source and a receiver can be characterized by the RIR vector which is shown in Figure 2. Figure 3a displays our acousitc scene. The heights of sensors and sources are 1.5 meters. The horizontal coordinates of signal of interest (SOI) and interference signal (INT) are $(1.5, 4.5)$ and $(1.5, 1.5)$ respectively. The number of sensors in the BSS model is chosen as 2, 3 and 4. The horizontal coordinates of sensors are varied in the area encircled by the black dashed line. Furthermore, both source speech and interference speech contain 6.25s voice signals sampled at 8kHZ. The objective function of optimizing convolutive BSS model is the interference suppression. Let $\hat{P}_Y(\omega)$ be the spectral power estimate of the source signal. Let $\hat{P}_{\tilde{Y}}(\omega)$ be the spectral power estimate of the output of BSS. Let $\hat{P}_{Y_I}(\omega)$ be the spectral power estimate of the interference speech. Let $\hat{P}_{\tilde{Y}_I}(\omega)$ be the spectral power estimate of the output of BSS when the interference speech is active alone. The normalized interference suppression is defined by

$$\mathrm{SUPP_I} = 10 \times \log_{10} \frac{\int_{-\pi}^{\pi} \hat{P}_{\tilde{Y}_I}(\omega)d\omega}{C_d \int_{-\pi}^{\pi} \hat{P}_{Y_I}(\omega)d\omega}$$

where C_d is defined by

$$C_d = \frac{\int_{-\pi}^{\pi} \hat{P}_Y(\omega)d\omega}{\int_{-\pi}^{\pi} \hat{P}_{\tilde{Y}}(\omega)d\omega}.$$

5.2. Preliminary Results

To investigate the behaviour of BSS in reverberant environments, consider a convolutive BSS model containing two acoustic sensors. One sensor is fixed at centre of the acoustic scene where the horizontal coordinate is $(3, 3)$ and the other sensor is varied among the the area encircled by the black dashed line in Figure 3a. Figures 3b, 3c and 3d show the suppression capability ($\mathrm{SUPP_I}$) as

11

a function of sensor location variables when T_{60} is chosen as $0ms$, $100ms$ and $200ms$ respectively. The mesh plots provide a clear relationship between the position of the sensor and the suppression as RT varies.

For a lower RT ($T_{60} = 0$ or $100ms$), the 2-elements acoustic sensor array achieves a good suppression capability when the unfixed sensor is closer to the interference source. This is consistent with the fact that BSS behaves as a nulling beamformer where the sensor is closest to the interference to perform nulling. For the case of $T_{60} = 200ms$, we observe a different fluctuation pattern on SUPP$_\text{I}$. The suppression capability changes rapidly with the change of the sensor locations. With the increment of T_{60} from 0 to $200ms$, more and more local minima or maxima can be observed. It is evident that SUPP$_\text{I}$ has a spurious behaviour in the presence of acoustic reverberation. This is because as RT increases, the number of reflected signals increases and the problem becomes highly underdetermined. In this case, it is difficult to find the global optima because the sensor placement optimization can get trapped in a state of local optimum [34]. Therefore, we use the proposed hybrid descent method in Algorithm 1 to address the problem. The following subsection presents the complete experimental analysis of the proposed system.

5.3. Experimental Results

5.3.1. Setup

We consider the convolutive BSS model containing 2, 3 or 4 sensors. The initial horizontal coordinates of the acoustic sensor array are shown in Table 3. In the optimization process, the initial horizontal coordinates can be randomly chosen because the genetic algorithm can simultaneously generate many initial guesses for the horizontal coordinates. For the comparison of the separated speech by the BSS system, the initial positions of the acoustic sensor array are deliberately placed close the original sources. All the sensors can be varied among the area encircled by the black dashed line in Figure 3a.

2 sensors	3 sensors	4 sensors
(3,1.5)	(3,1.5)	(3,1.5)
(3,4.5)	(3,3)	(3,2.17)
-	(3,4.5)	(3,3.83)
-	-	(3,4.5)

Table 3: Initial horizontal coordinates of the acoustic sensor array.

5.3.2. Computational Time

A comparison is made between the computational time for both the heuristic algorithm (GA) and the gradient-based method (SQP). This comparison was made for the case of 2 sensors and with a reverberation time $T_{60} = 0$ and the experiment was repeated twenty times. From Table 4, we can see that the heuristic method requires longer average time to converge compared to the gradient-based
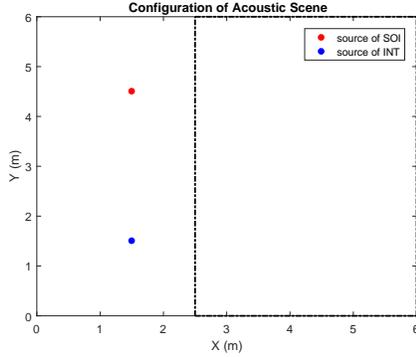
(a) Configuration of acoustic scene
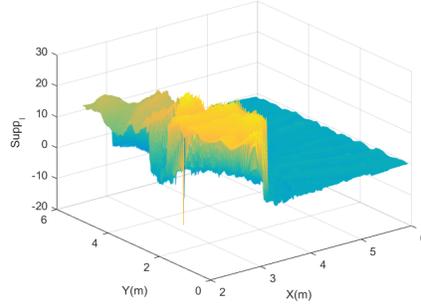
(b) Mesh plot with $T_{60} = 0ms$
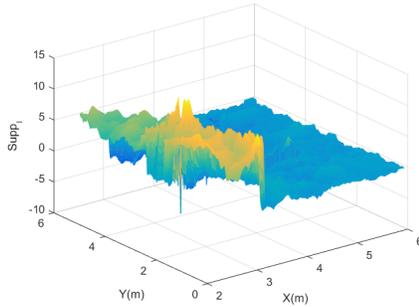
(c) Mesh plot with $T_{60} = 100ms$
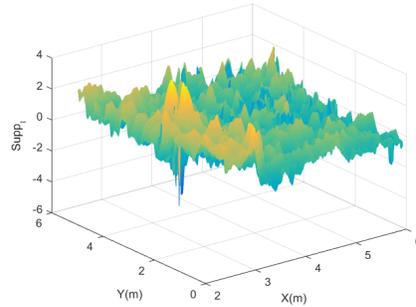
(d) Mesh plot with $T_{60} = 200ms$

Figure 3: Configuration of acoustic scene and mesh plots for SUPP$_I$ value when T_{60} is chosen as $0ms$, $100ms$ and $200ms$.

method. However, for non-convex optimization, the faster convergence time of the gradient-based method comes at the expense of local optima.

	GA	SQP
Mean	3113.64 (s)	739.67 (s)
Std.	731.18	433.68

Table 4: Statistics of computational time

5.3.3. Suppression Performance

Figures 4a, 5a and 6a plot the suppression measure of the interference signals when the convolutive BSS model is applied with both the optimized position and the initialized position of the acoustic sensor array. Evidently, the proposed optimized BSS system outperforms the unoptimized case by at least 10dB across the RTs. It is interesting to note that the performance improvement is fairly consistent even up the case of RT=500ms. To further compare the quality

13

improvement of the enhanced speech, Figures 4b, 5b and 6b plot the perceptual evaluation of speech quality (PESQ) [35] measure for the received and separated speech (enhanced speech). Again, the results demonstrate that the PESQ score of the proposed system is greatly improved especially for lower values of RTs.
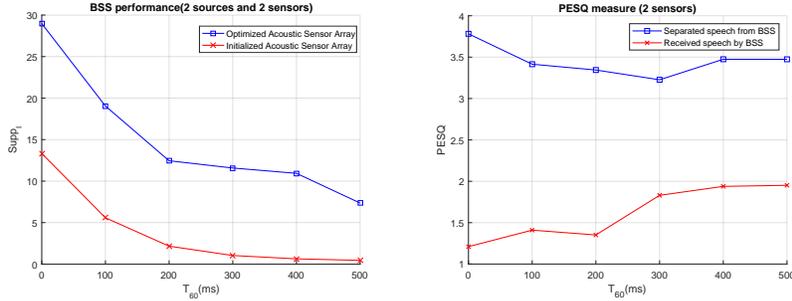
(a) Suppression measure of interference signal.

(b) PESQ measure of the received and separated signal.

Figure 4: Suppression measure and PESQ measure of the WASN with 2 sensors.
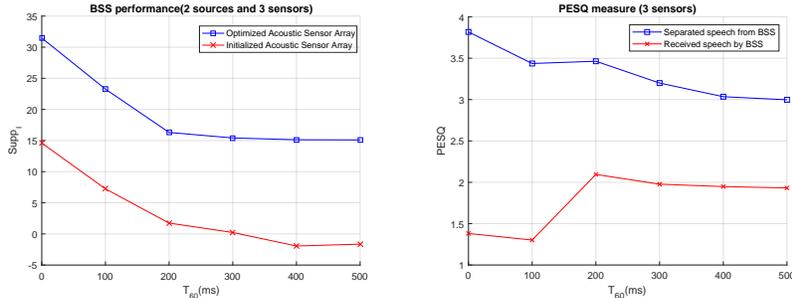
(a) Suppression measure of interference signal.

(b) PESQ measure of the received and separated signal.

Figure 5: Suppression measure and PESQ measure of the WASN with 3 sensors.

5.3.4. Effect of the Number of the Sensors

Figure 7 shows the separation performance between the proposed systems and the fixed arrays as the number of sensors vary. The dash lines denote the SUPP_I value with respect to the initial acoustic sensor arrays in Table 3. The solid lines denote the SUPP_I value with respect to the optimized acoustic sensor arrays derived from Algorithm 1. For both unoptimized and optimized cases, the performance increases gradually as the number of sensors grows. Consistent with the previous results, the suppression capability reduces as RT rises. However, it is clear that for all cases considered, the optimized acoustic sensor array outperform the unoptimized array.
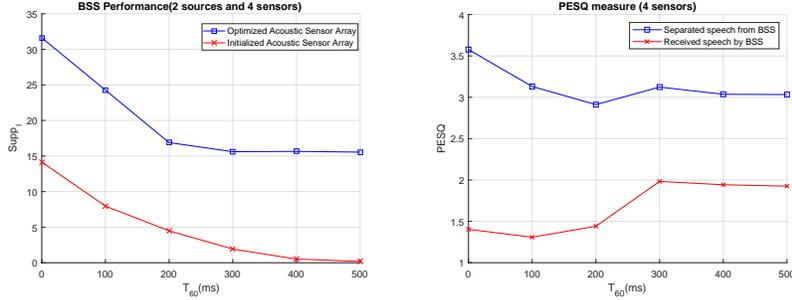
(a) Suppression measure of interference signal.

(b) PESQ measure of the received and separated signal.

Figure 6: Suppression measure and PESQ measure of the WASN with 4 sensors.

5.3.5. Optimized Sensor Positions

Figure 8-10 show the optimized locations of sensors when the number of sensors is chosen as 2, 3 and 4. It is noted that the optimized acoustic sensor array tends to cluster around the interference source when T_{60} is less than or equal to 200ms. Taking advantage of close placement, the BSS system can achieve a significant suppression for the interference signal. When T_{60} is greater than 200ms, more and more acoustic signals are reflected by the walls and the optimized acoustic sensor array spreads out. This is because more and more reverberant signals need to be suppressed. This is consistent to that reported by [5] where BSS tries to suppress the interference signal by forming a null on it. Moreover, It also coincides with the different pattern of SUPP_I fluctuation in Figures 3b, 3c and 3d. As the RT increases, then the array element spreads itself. Interestingly, the optimized geometry is consistent with the suppression performance for the case of T_{60} equaling 400ms and 500ms, where very little improvement was observed. This can be understood from the optimized location for 2, 3 and 4 sensors for both RTs as they are very similar with the converged geometry.

5.3.6. Convergence Analysis

Figure 11 shows the evaluation of the objective function as a function of iterations for the case of 4 microphones with $T_{60} = 0ms$ or $100ms$. Interestingly, Figure 11a reveals that there is a plateauing effect at around iteration $4-6$, suggesting the solution is trapped in a local optima. However, a huge jump is registered at the next iteration. This clearly illustrates how the proposed hybrid method untraps itself from a local optima in search of a global optima, hence increased performance. From both the figures, we can observe the significant improvement of the objective function value when the iteration is less than 10. The objective function then starts to plateau and converge. When the iteration is greater than 15, the change of objective function value is only incremental, which suggests that the hybrid heuristic method is converged.
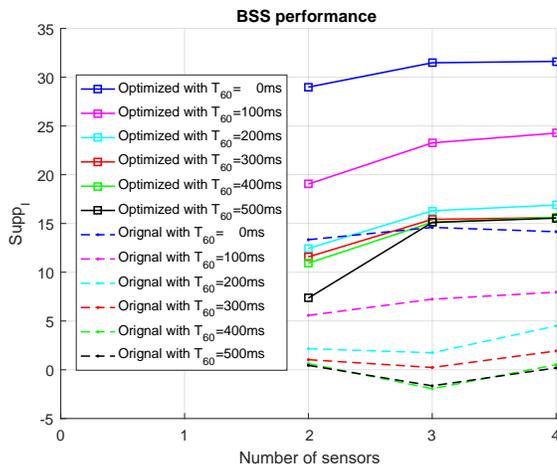
Figure 7: The suppression measure of the interference signal between the proposed method and the fixed array for different reverberation time.

5.4. General Discussions and Analysis

From the experimental results, we can observe that a higher RT leads to a worse separation performance.Figure 3b, 3c and 3d show the mesh plots as a function of suppression capability versus the sensor position for different reverberation times. The plots clearly show as the RT increases, there exists spurious number of optimal points. This is because as the reverberation time increases, the number of reflected signal increases. These reflected signals act as additional sources, which invariably cause the problem to be underdetermined as there are now many more source compared to the number of sensors. The spurious points are a result of BSS extracting those reflected signals as opposed to the direct path signals. It can be seen that when the RT is lower, then the set of location corresponding to the highest suppression is more clustered. On the other words, the separation problem becomes more difficult as many more sources need to be separated in the case of a higher RT.

Given a fixed T_{60}, it can be observed that the performance of optimized BSS system becomes better with the increment of number of sensors. However, without the optimization of sensor locations, system performance cannot be simply improved by increasing the number of sensors. Moreover, an inappropriate placement of sensors can undermine the BSS system. For example, when T_{60} is chosen as $200ms$, $300ms$, $400ms$ and $500ms$, the initial acoustic sensor array with 3 sensors has a poorer performance compared with 2 sensors. On the other hand, the optimized acoustic sensor array consisting of 2 sensors can easily outperform the initialized acoustic sensor array containing 4 sensors when the acoustic reverberation exists. Regardless, all suppression performance improves as the spatial information is optimized.

Interestingly, the optimized acoustic sensor arrays also exhibit some similar

(a) $T_{60} = 0ms,\ 100ms$

(b) $T_{60} = 200ms,\ 300ms$

(c) $T_{60} = 400ms,\ 500ms$
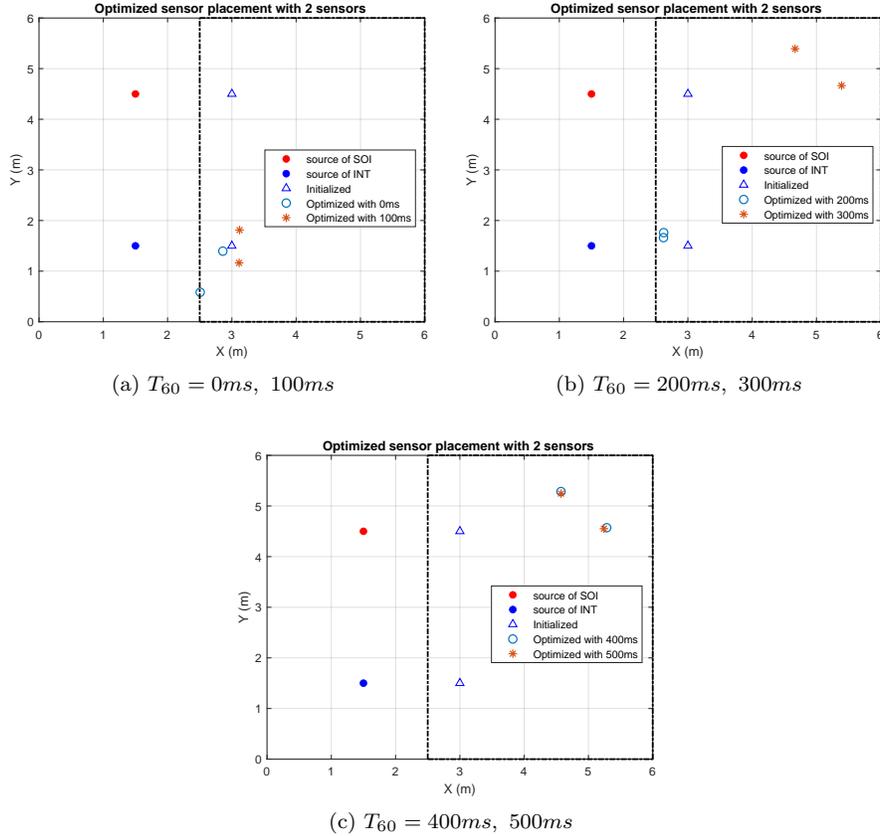
Figure 8: Optimized locations of 2 sensors with T_{60} chosen as $0ms$, 100ms, 200ms, 300ms, 400ms and 500ms.

placement patterns. For a lower RT ($T_{60} = 0$ or $100ms$), the sensors tend to cluster closer to the interference source. This is because the BSS model behaves like a null beamformer. As RT increases, the sensors are more scattered as more and more reverberant signals are deemed to be the interference sources. The results also show that, up to a certain RT ($T_{60} = 400ms$ or $500ms$), the change in the placement of the optimized acoustic sensor arrays is minimal. This is primarily attributed to the overdetermined situation as RT increases, which completely saturates the separation performance akin to a plateauing effect. This also corroborates for the cases with less number of sensors as the placement movement saturates even quicker since there is now even less degrees of freedom.

The proposed approach has shown that the optimization of placement of the sensors greatly improves the separation performance for a blind source separation system. Assuming that the position of the source largely remains the

17

(a) $T_{60} = 0ms, \ 100ms$

(b) $T_{60} = 200ms, \ 300ms$

(c) $T_{60} = 400ms, \ 500ms$
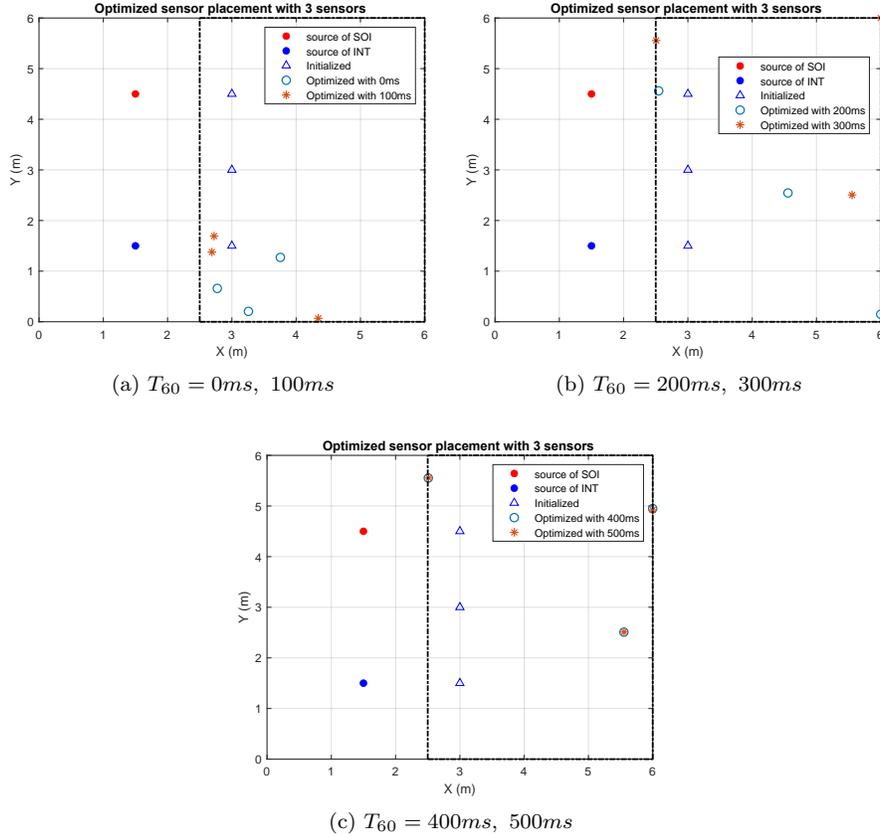
Figure 9: Optimized locations of 3 sensors with T_{60} chosen as $0ms$, 100ms, 200ms, 300ms, 400ms and 500ms.

same, e.g., source is sitting on a chair, then the sensor optimization will need to be performed initially during the calibration stage. Once the positions are identified, then any existing real-time BSS systems [36, 37, 38] can be readily deployed.

6. Conclusion

The paper presented a hybrid heuristic method to optimize the sensor placement for a BSS system. This is because the sensor optimization problem is highly non-linear and a conventional search algorithm will fail. The hybrid approach provides the rapid convergence of the gradient descent method coupled with local optima protection from the genetic algorithm. Clearly, the findings show that the sensors positions greatly impact the performance of a BSS system. The results indicate that the sensors tend to cluster on the desired source

(a) $T_{60} = 0ms, \ 100ms$

(b) $T_{60} = 200ms, \ 300ms$

(c) $T_{60} = 400ms, \ 500ms$
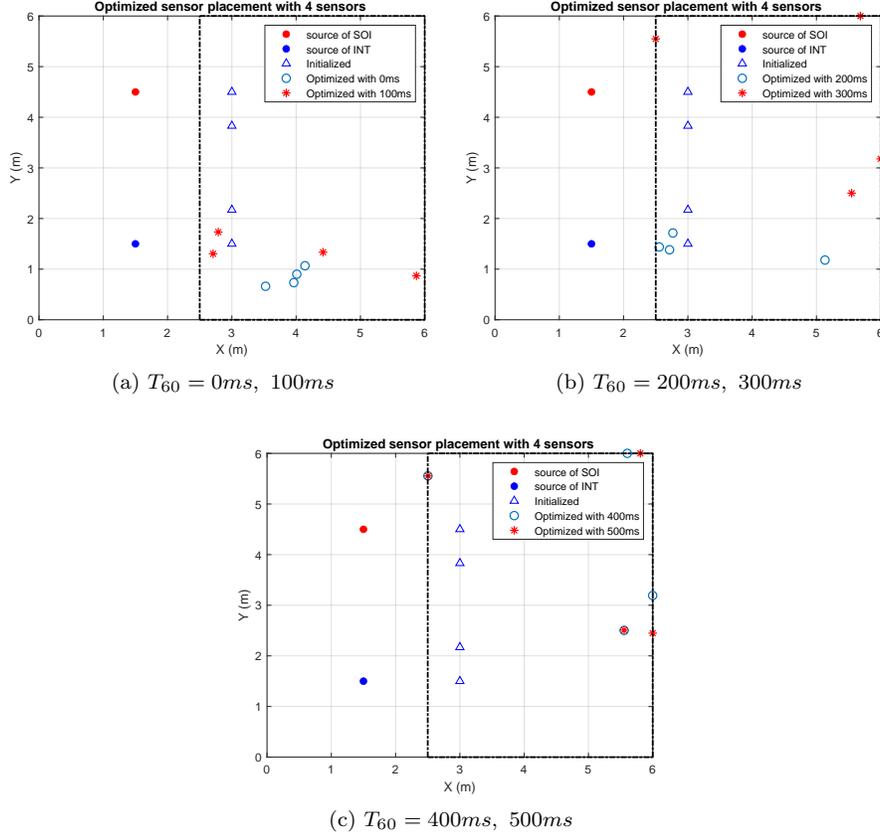
Figure 10: Optimized locations of 4 sensors with T_{60} chosen as $0ms$, 100ms, 200ms, 300ms, 400ms and 500ms.

when the RT is low. This is however less obvious when the RT increases from 200 ms onwards as the problem becomes underdetermined. All in all, the experiments show that there is a marked increase in the separation performance as the spatial information is optimized.

7. Acknowledgements
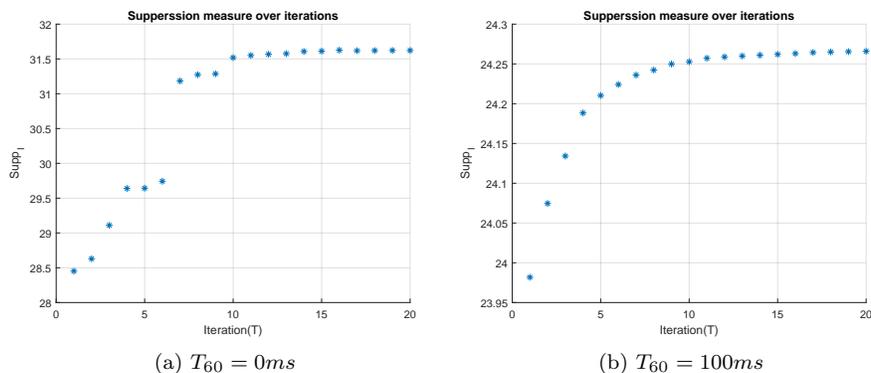
Figure 11: Objective function value over iterations in the case of 4 microphones

References

[1] E. C. Cherry, "Some experiments on the recognition of speech, with one and with two ears," *The Journal of the acoustical society of America*, vol. 25, no. 5, pp. 975–979, 1953.

[2] Z. Li, K. F. C. Yiu, and S. Nordholm, "On the indoor beamformer design with reverberation," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, vol. 22, no. 8, pp. 1225–1235, 2014.

[3] K. Y. Chan, K. F. C. Yiu, and S. Nordholm, "Microphone configuration for beamformer design using the taguchi method," *Measurement*, vol. 96, pp. 58–66, 2017.

[4] L. Parra and C. Spence, "Convolutive blind separation of non-stationary sources," *IEEE transactions on Speech and Audio Processing*, vol. 8, no. 3, pp. 320–327, 2000.

[5] S. Araki, S. Makino, Y. Hinamoto, R. Mukai, T. Nishikawa, and H. Saruwatari, "Equivalence between frequency-domain blind source separation and frequency-domain adaptive beamforming for convolutive mixtures," *EURASIP Journal on Applied Signal Processing*, vol. 2003, pp. 1157–1166, 2003.

[6] J. Benesty, S. Makino, and J. Chen, *Speech enhancement*. Springer Science & Business Media, 2005.

[7] S. Y. Low, S. Nordholm, and R. Togneri, "Convolutive blind signal separation with post-processing," *IEEE Transactions on Speech and Audio Processing*, vol. 12, no. 5, pp. 539–548, 2004.

[8] X. Yu, D. Hu, and J. Xu, *Blind source separation: theory and applications*. John Wiley & Sons, 2013.

[9] H. Sawada, H. Kameoka, S. Araki, and N. Ueda, "Multichannel extensions of non-negative matrix factorization with complex-valued data," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 5, pp. 971–982, 2013.

[10] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural networks*, vol. 13, no. 4-5, pp. 411–430, 2000.

[11] U. A. Lindgren and H. Broman, "Source separation using a criterion based on second-order statistics," *IEEE Transactions on Signal Processing*, vol. 46, no. 7, pp. 1837–1850, 1998.

[12] M. S. Pedersen, J. Larsen, U. Kjems, and L. C. Parra, "Convolutive blind source separation methods," in *Springer handbook of speech processing*. Springer, 2008, pp. 1065–1094.

[13] P. Comon, C. Jutten, and J. Herault, "Blind separation of sources, part ii: Problems statement," *Signal processing*, vol. 24, no. 1, pp. 11–20, 1991.

[14] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural computation*, vol. 7, no. 6, pp. 1129–1159, 1995.

[15] A. Hyvärinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural computation*, vol. 9, no. 7, pp. 1483–1492, 1997.

[16] A. Masnadi-Shirazi and B. Rao, "Independent vector analysis incorporating active and inactive states," in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2009, pp. 1837–1840.

[17] A. Souloumiac, "Blind source detection and separation using second order non-stationarity," in *1995 International Conference on Acoustics, Speech, and Signal Processing*, vol. 3. IEEE, 1995, pp. 1912–1915.

[18] H. H. Dam, S. Nordholm, S. Y. Low, and A. Cantoni, "Blind signal separation using steepest descent method," *IEEE Transactions on Signal Processing*, vol. 55, no. 8, pp. 4198–4207, 2007.

[19] S. Y. Low, K. F. C. Yiu, and S. Nordholm, "Beamspace blind signal separation for speech enhancement," *Optimization and engineering*, vol. 10, no. 2, pp. 313–330, 2009.

[20] S. Y. Low, K. F. C. Yiu, and S. Nordholm, "Second-order-based blind signal separation in reverberant environments," *International Journal of Electronics*, vol. 102, no. 9, pp. 1583–1593, 2015.

[21] M. Souden, S. Araki, K. Kinoshita, T. Nakatani, and H. Sawada, "A multichannel mmse-based framework for speech source separation and noise reduction," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 9, pp. 1913–1928, 2013.

[22] J. O. Smith, *Introduction to digital filters: with audio applications.* Julius Smith, 2007, vol. 2.

[23] A. Cichocki and S. i. Amari, *Adaptive blind signal and image processing: learning algorithms and applications.* John Wiley & Sons, 2002.

[24] E. A. Lehmann and A. M. Johansson, "Diffuse reverberation model for efficient image-source simulation of room impulse responses," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 6, pp. 1429–1439, 2009.

[25] K. F. C. Yiu, Y. Liu, and K. L. Teo, "A hybrid descent method for global optimization," *Journal of Global Optimization*, vol. 28, no. 2, pp. 229–238, 2004.

[26] R. Storn and K. Price, "Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces," *Journal of global optimization*, vol. 11, no. 4, pp. 341–359, 1997.

[27] K. Chen, X. Yun, Z. He, and C. Han, "Synthesis of sparse planar arrays using modified real genetic algorithm," *IEEE Transactions on Antennas and Propagation*, vol. 55, no. 4, pp. 1067–1073, 2007.

[28] K. Akdagli, "Null steering of linear antenna arrays using a modified tabu search algorithm," *Progress In Electromagnetics Research*, vol. 33, pp. 167–182, 2001.

[29] G. Doblinger, "Optimized design of interpolated array and sparse array wideband beamformers," in *2008 16th European Signal Processing Conference.* IEEE, 2008, pp. 1–5.

[30] E. A. Silver, "An overview of heuristic solution methods," *Journal of the operational research society*, vol. 55, no. 9, pp. 936–956, 2004.

[31] Z. Li and K. F. C. Yiu, "Beamformer configuration design in reverberant environments," *Engineering Applications of Artificial Intelligence*, vol. 47, pp. 81–87, 2016.

[32] X. S. Yang, *Nature-inspired optimization algorithms.* Academic Press, 2020.

[33] M. Nowostawski and R. Poli, "Parallel genetic algorithm taxonomy," in *1999 Third International Conference on Knowledge-Based Intelligent Information Engineering Systems. Proceedings (Cat. No. 99TH8410).* Ieee, 1999, pp. 88–92.

[34] N. V. Banichuk, *Introduction to optimization of structures.* Springer Science & Business Media, 2013.

[35] Z. Li, K. F. C. Yiu, Y. H. Dai, and S. Nordholm, "Distributed LCMV beamformer design by randomly permuted admm," *Digital Signal Processing*, vol. 106, p. 102820, 2020.

[36] R. Mukai, H. Sawada, S. Araki, and S. Makino, "Robust real-time blind source separation for moving speakers in a room," in *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03).*, vol. 5.   IEEE, 2003, pp. V–469.

[37] L. Parra and C. Spence, "On-line convolutive blind source separation of non-stationary signals," *Journal of VLSI signal processing systems for signal, image and video technology*, vol. 26, no. 1, pp. 39–46, 2000.

[38] K. F. C. Yiu and S. Y. Low, "On a real-time blind signal separation noise reduction system," *International Journal of Reconfigurable Computing*, vol. 2018, 2018.