Received: September 14, 2022; Accepted: November 29, 2022

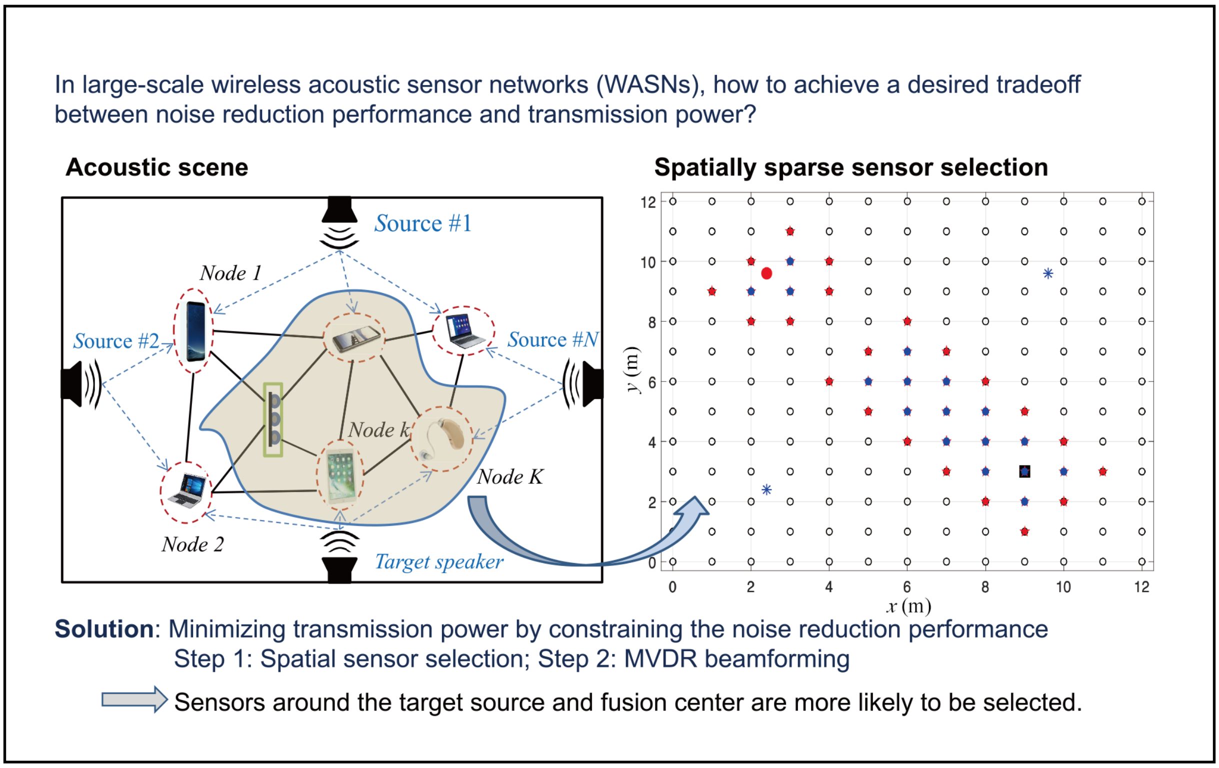
Low-complexity energy-aware sensor selection for noise reduction in distributed microphone networks

Jie Zhang ✉, Lu-Zhen Xu, and Li-Rong Dai

National Engineering Research Center for Speech and Language Information Processing, University of Science and Technology of China, Hefei 230027, China

✉Correspondence: Jie Zhang, E-mail: jzhang6@ustc.edu.cn

Graphical abstract



In large-scale wireless acoustic sensor networks (WASNs), how to achieve a desired tradeoff between noise reduction performance and transmission power?

Solution: Minimizing transmission power by constraining the noise reduction performance
Step 1: Spatial sensor selection; Step 2: MVDR beamforming

Sensors around the target source and fusion center are more likely to be selected.

Spatial sparse sensor selection for MVDR beamforming.

Public summary

- Sensor selection is an effective tool to optimize the geometry of microphone networks and reduce the transmission cost, where many sensors contributes marginally to the task performance at hand.
- Based on the existing semi-definite programming utility-based methods, in this work we propose three energy-efficient utilities (i.e., weighted utility, gradient and weight input SNR), based on which three corresponding low-complexity sensor selection approaches are proposed.
- Results show that sensors around sources and the fusion center are more informative in the sense of performance and the proposed narrowband methods converge more faster.

Citation: Zhang J, Xu L Z, Dai L R. Low-complexity energy-aware sensor selection for noise reduction in distributed microphone networks. *JUSTC*, **2023**, 53(4): 0402. DOI: 10.52396/JUSTC-2022-0121

JUSTC
Article
http://justc.ustc.edu.cn
Received: September 14, 2022; Accepted: November 29, 2022

Low-complexity energy-aware sensor selection for noise reduction in distributed microphone networks

Jie Zhang ✉, Lu-Zhen Xu, and Li-Rong Dai

National Engineering Research Center for Speech and Language Information Processing, University of Science and Technology of China, Hefei 230027, China

✉Correspondence: Jie Zhang, E-mail: jzhang6@ustc.edu.cn

Cite This: *JUSTC*, **2023**, 53(4): 0402 (10pp)   Read Online

Abstract: Noise reduction (NR) is a necessary front-end in many audio applications for improving signal quality. It was shown that sparsity-promoting sensor selection potentially makes a trade-off between energy consumption and NR performance, which is rather important for large-scale wireless acoustic sensor networks (WASNs), where many sensors contribute negligibly to NR but energy consumption affects the lifetime of WASNs. This paper presents a sensor selection approach for beamforming-based NR by minimizing the total energy consumption and constraining the output noise variance. Motivated by the optimal semi-definite programming (SDP) solution and the utility-based method, we propose three low-complexity selection metrics: weighted utility, gradient, and weighted input signal-to-noise ratio (SNR). It is shown that the proposed weighted utility and gradient-based methods are near-optimal in performance but much faster than the SDP-based method, and the weighted SNR method has the lowest time complexity with a tiny performance sacrifice. Numerical results using a simulated WASN validate the superiority of the proposed approaches over conventional methods.

Keywords: sensor selection; forward/backward algorithms; gradient; utility; MVDR beamformer; speech enhancement; distributed microphone network

CLC number: TN912.35   **Document code:** A

1 Introduction

Currently, with advanced microelectronics, wireless devices are commonly used in daily life. The organization of devices forms a distributed microphone network, or more generally called a wireless acoustic sensor network (WASN), which enables information exchange between neighboring nodes[1–3]. The devices can thus be manipulated remotely via wireless signal communication. From the perspective of system control, speech is one of the most natural signals for human-machine interaction[4], since most devices have microphone(s) equipped. However, in practice, environmental noise inevitably degrades the quality of microphone recordings, which heavily affects the speech interaction performance. For this, speech enhancement or noise reduction (NR) becomes necessary[5].

Speech enhancement has a broad range of applications, e.g., teleconferencing systems[6], hands-free telephony[7], speech recognition[8], human-robot interaction[9], and hearing aids (HAs)[10]. Conventional array-based speech processing systems usually physically link a fusion center (FC) to multiple microphones, implying that such wired array systems cannot be rearranged flexibly (e.g., the addition or removal of microphone nodes). As the location of microphone arrays is fixed, low-quality audio measurements will be recorded in case the target speaker is distant. The utilization of WASNs can potentially resolve these limitations. For example, due to the random placement of wireless devices, some nodes may be even around the target speaker, which can then provide high-quality audio data, resulting in a benefit for the NR. Although the HAs only allow for a small-sized microphone array, if the surrounding wireless devices can transmit the recorded data to the HAs, more data are then available, leading to a performance improvement[10]. Moreover, microphone nodes can be configured more flexibly, e.g., as centralized or distributed networks[3, 5, 11, 12].

For the multi-microphone NR problem, increasing the number of microphones leads to better performance[13] but also more data transmissions and a higher computational complexity. It was shown in Ref. [14] that multi-microphone recordings are highly redundant, which enables the exclusion of noninformative sensors (e.g., sensors that are far away from the target speaker) with an ignorable impact on the NR performance. On the other hand, the NR algorithms in the context of the WASN have to take the transmission power into account, which plays an important role in the network lifetime because portable devices usually only have a limited amount of battery resources. As the total energy consumption is determined by the number of microphones for NR, to improve the network efficiency, it would be promising to search an informative microphone subset from a large-scale candidate set, which can result in an expected NR performance. It was shown in Ref. [15] that sensor selection can facilitate a significant reduction in the cardinality of informative sensors with a negligible performance loss. In theory, it can be formu-

DOI: 10.52396/JUSTC-2022-0121

lated by minimizing the signal estimation error and constraining the number of selected subsets, or vice versa. Essentially, this is an NP-hard problem that can be solved by convex optimization with semi-definite relaxation[16] or greedy approaches (e.g., sub-modularity-based heuristics[17]). Sensor selection has been applied to, e.g., field estimation[18], source localization[16], target tracking[19], and speech enhancement[20–23], with a saving of data transmission.

To improve the energy efficiency of WASNs, we consider sensor selection for the minimum variance distortionless response (MVDR) beamformer-based NR in this work. The initial problem is built upon optimizing the total energy consumption in terms of the selection status of sensors subject to a constraint on the NR performance in output noise variance. In general, we can design model-driven and data-driven strategies to solve this problem. Given the noise covariance matrix of the complete network, the proposed sensor selection problem was converted into a semi-definite program (SDP) using convex optimization techniques (optSdpRemoval)[23]. To avoid the dependence on the statistics of the whole network, a greedy data-driven optSdpAddition method was further proposed in Ref. [23], which gradually increases the candidate set of sensors and executes a smaller-scaled SDP problem at each iteration. However, optSdpAddition is rather computationally complicated and has to execute an SDP problem at each iteration. The impact of the acoustic transfer function (ATF) mismatch on the MVDR sensor selection problem and the extension to the linearly-constrained minimum variance (LCMV) beamformer were presented in Ref. [24]. In Ref. [25], a frequency-invariant sensor selection-based MVDR method was proposed, which can avoid switching the selection status across frequencies but requires a much higher-complexity solver. In Ref. [26], by defining the contribution of each sensor to NR as utility, a backward model-based utilityRemoval method was proposed, which excludes the sensor with the smallest utility from the current candidate set in each iteration. Given an initial selected subset, a forward data-driven utilityAddition can be designed by adding the sensor with the largest utility to the selected subset[26]. Although utilityAddition is generally faster than optSdpAddition, it does not take the transmission energy into account, as a sensor might have a large utility but requires a large transmission energy to the FC. Furthermore, due to the dependence on the noise covariance of the complete network, model-driven methods are impractical due to the unknown number of available sensors in practice. We therefore consider low-complexity sensor selection approaches and take power consumption into account in this work.

The contribution of this paper is threefold. (i) We define the contribution of a sensor to NR over the corresponding transmission power as the weighted utility, based on which we propose a model-driven weightedUtilityRemoval method and a data-driven weightedUtilityAddition approach similar to utility-based methods. (ii) Based on the original optimization problem, we calculate the gradient of the objective function with respect to the selection variables and use the gradient to measure the contribution of sensors to the optimality, as the gradient measures the changing trend of a function along a certain direction. Applying the gradient, we design gradient-based sensor selection algorithms, namely, gradRemoval and gradAddition. The time complexity of the weighted utility and gradient-based approaches is square in terms of the number of candidate sensors, which is lower than that of optSdpAddition. (iii) Since it was shown in Ref. [23] that sensors with a high signal-to-noise ratio (SNR) are useful for target estimation and sensors with a low transmission cost are helpful for saving energy, we thus define the input SNR over the sensor-FC transmission power as the weighted SNR. Using the weighted SNR as the utility, we can design weightedSnrRemoval and weightedSnrAddition sensor selection algorithms, which has a linear time complexity since no covariance matrix needs to be updated. Numerical simulations using a large-scale microphone network validate the proposed methods. The sensors proximal to the target signal and the FC are more likely to be included. The proposed weighted utility and gradient-based approaches also choose some sensors next to the coherent interfering sources, as they might be useful for cancelling noise sources even with a very low SNR, which, however, cannot be observed from the weighted SNR-based methods.

The rest of this paper is structured as follows. Section 2 introduces fundamental knowledge on the signal model and MVDR beamforming. Section 3 presents the sensor selection model, problem formulation, and related works. In Section 4, we present the proposed weightedUtility, gradient, weightedSnr-based methods. Section 5 presents the numerical results using a simulated WASN. Finally, Section 6 concludes this work.

2 Signal model and MVDR beamformer

2.1 Signal model

In this work, we consider a distributed microphone network consisting of M sensor nodes in the acoustic scene for the estimation of a single target source, where each node is equipped with a single microphone and the FC is one of the nodes without loss of generality (w.o.l.g.). In the short-time Fourier transform (STFT) domain, the k-th microphone signal is given by

$$Y_k(\omega,l) = X_k(\omega,l) + N_k(\omega,l),\ k = 1,\cdots,M, \tag{1}$$

where $X_k(\omega,l)$ denotes the speech component at microphone k, $N_k(\omega,l)$ includes all noise components (e.g., competing speakers, late reverberations, sensor self-noise), and l and ω are the frame and angular frequency indices, respectively.

Let $a_k(\omega)$ represent the ATF relating the source position to the k-th microphone. The signal component then equals

$$X_k(\omega,l) = a_k(\omega)S(\omega,l), \tag{2}$$

where $S(\omega,l)$ denotes the target signal at the source position. Furthermore, assuming that the FC is the first node w.o.l.g. and taking it as the reference position, we can define the relative ATF (RTF) as

$$h_k(\omega) = a_k(\omega)/a_1(\omega), \tag{3}$$

such that $X_k(\omega,l) = h_k(\omega)X_1(\omega,l)$, where $X_1(\omega,l) = a_1(\omega)S(\omega,l)$. Note that the assignment of reference microphone might affect the multi-microphone NR performance[27], which is beyond the scope of this paper. The utilization of RTF does not affect the NR performance, which can be estimated using the methods in Ref. [28].

For notational brevity, the frame and frequency indices will be omitted in the sequel. Let $\boldsymbol{y} = [Y_1, Y_2, \ldots, Y_M]^{\mathrm{T}}$ with $(\cdot)^{\mathrm{T}}$ denoting matrix/vector transpose store the STFT coefficients for each time-frequency bin. Vectors $\boldsymbol{n}$, $\boldsymbol{a}$, $\boldsymbol{h}$ and $\boldsymbol{x} = \boldsymbol{h}X_1 = \boldsymbol{a}S$ are defined similarly to, respectively stack the noise components, ATF, RTF, and signal components, such that we can rewrite the signal model in (1) in a vector form as

$$\boldsymbol{y} = \boldsymbol{h}X_1 + \boldsymbol{n}. \tag{4}$$

2.2 MVDR beamformer

Typically, the MVDR beamformer is optimized by minimizing the output noise power (or variance) under a linear constraint associated with the RTF (or ATF) of the target source, which can be mathematically given by

$$\boldsymbol{w}_{\mathrm{MVDR}} = \arg\min_{\boldsymbol{w}} \boldsymbol{w}^{\mathrm{H}}\boldsymbol{\varPhi}_{\boldsymbol{nn}}\boldsymbol{w} \qquad \text{s.t.} \quad \boldsymbol{w}^{\mathrm{H}}\boldsymbol{h} = 1, \tag{5}$$

where $\boldsymbol{\varPhi}_{\boldsymbol{nn}} = \mathbb{E}\{\boldsymbol{n}\boldsymbol{n}^{\mathrm{H}}\}$ being the noise covariance matrix with $\mathbb{E}[\cdot]$ the mathematical expectation. We define $\boldsymbol{\varPhi}_{\boldsymbol{xx}} = \mathbb{E}\{\boldsymbol{x}\boldsymbol{x}^{\mathrm{H}}\} = \sigma_S^2\boldsymbol{a}\boldsymbol{a}^{\mathrm{H}} = \sigma_{X_1}^2\boldsymbol{h}\boldsymbol{h}^{\mathrm{H}}$ as the signal covariance matrix, where σ_S^2 and $\sigma_{X_1}^2$ denote the power spectral densities (PSDs) of the target speech source at the source and reference positions, respectively, and $(\cdot)^{\mathrm{H}}$ is the conjugate transpose. With the linear constraint in (5), it holds that $\boldsymbol{w}_{\mathrm{MVDR}}^{\mathrm{H}}\boldsymbol{\varPhi}_{\boldsymbol{xx}}\boldsymbol{w}_{\mathrm{MVDR}} = \sigma_{X_1}^2$, implying that the power of the desired signal component at the reference position is preserved, and the noise power can be reduced after MVDR beamforming. It can be easily verified that the MVDR beamformer is given by [29, 30]

$$\boldsymbol{w} = (\boldsymbol{h}^{\mathrm{H}}\boldsymbol{\varPhi}_{\boldsymbol{nn}}^{-1}\boldsymbol{h})^{-1}\boldsymbol{\varPhi}_{\boldsymbol{nn}}^{-1}\boldsymbol{h}. \tag{6}$$

Using the filter $\boldsymbol{w} = [w_1, w_2, \ldots, w_M]^{\mathrm{T}}$, the estimated target signal component can be obtained through beamforming as

$$\hat{X}_1 = \boldsymbol{w}^{\mathrm{H}}\boldsymbol{y}. \tag{7}$$

After filtering, the residual noise power can be computed as

$$\boldsymbol{w}^{\mathrm{H}}\boldsymbol{\varPhi}_{\boldsymbol{nn}}\boldsymbol{w} = (\boldsymbol{h}^{\mathrm{H}}\boldsymbol{\varPhi}_{\boldsymbol{nn}}^{-1}\boldsymbol{h})^{-1}, \tag{8}$$

and the output SNR (oSNR) is given by

$$\text{oSNR} = \frac{\boldsymbol{w}^{\mathrm{H}}\boldsymbol{\varPhi}_{\boldsymbol{xx}}\boldsymbol{w}}{\boldsymbol{w}^{\mathrm{H}}\boldsymbol{\varPhi}_{\boldsymbol{nn}}\boldsymbol{w}} = \sigma_{X_1}^2\boldsymbol{h}^{\mathrm{H}}\boldsymbol{\varPhi}_{\boldsymbol{nn}}^{-1}\boldsymbol{h}. \tag{9}$$

Note that in practice the noise covariance matrix in the MVDR design needs to be replaced by its estimate $\hat{\boldsymbol{\varPhi}}_{\boldsymbol{nn}}$. If the covariance matrices are perfectly estimated, the objective function in (5) is equivalent to minimizing the output signal power, which is called the minimum power distortionless response (MPDR) beamformer[31].

3 Problem description

In this section, the general problem description for the sensor selection MVDR beamforming based NR issue within a large-scale distributed microphone network will be presented.

3.1 Definitions

To reflect the microphone selection status, let $\boldsymbol{p} = [p_1, p_2, \cdots, p_M]^{\mathrm{T}} \in \{0,1\}^M$,where $p_k = 1, \forall k$ represents the selected status and $p_k = 0$ the unselected status of microphone k, respectively. Based on $\boldsymbol{p}$, the selected subset is given by $\mathcal{S}_+ = \{k|p_k = 1\}$, and the unselected subset is denoted by $\mathcal{S}_- = \{k|p_k = 0\}$. Furthermore, $K = |\mathcal{S}_+|$ and $M - K = |\mathcal{S}_-|$ represent the numbers of the *selected* and *unselected* nodes, respectively. Letting $\mathrm{diag}(\boldsymbol{p})$ denote the diagonal matrix whose diagonal entries are given by $\boldsymbol{p}$, the selection matrix $\boldsymbol{\varSigma}_{\boldsymbol{p}} \in \{0,1\}^{K\times M}$ can thus be constructed by removing the all-zero rows of $\mathrm{diag}(\boldsymbol{p})$. Clearly, two properties related to the selection matrix hold:

$$\boldsymbol{\varSigma}_{\boldsymbol{p}}\boldsymbol{\varSigma}_{\boldsymbol{p}}^{\mathrm{T}} = \boldsymbol{I}_K, \quad \boldsymbol{\varSigma}_{\boldsymbol{p}}^{\mathrm{T}}\boldsymbol{\varSigma}_{\boldsymbol{p}} = \mathrm{diag}(\boldsymbol{p}), \tag{10}$$

where $\boldsymbol{I}_K$ denotes the K-dimensional identity matrix. Applying $\boldsymbol{\varSigma}_{\boldsymbol{p}}$, the microphone measurements of the selected sensors are then given by

$$\boldsymbol{y}_{\boldsymbol{p}} = \boldsymbol{\varSigma}_{\boldsymbol{p}}\boldsymbol{y} = \boldsymbol{\varSigma}_{\boldsymbol{p}}\boldsymbol{x} + \boldsymbol{\varSigma}_{\boldsymbol{p}}\boldsymbol{n} \in \mathbb{C}^K. \tag{11}$$

The RTF and noise covariance matrix related to the selected sensors can be similarly written as

$$\boldsymbol{h}_{\boldsymbol{p}} = \boldsymbol{\varSigma}_{\boldsymbol{p}}\boldsymbol{h} \in \mathbb{C}^K, \ \boldsymbol{\varPhi}_{\boldsymbol{nn},\boldsymbol{p}} = \boldsymbol{\varSigma}_{\boldsymbol{p}}\boldsymbol{\varPhi}_{\boldsymbol{nn}}\boldsymbol{\varSigma}_{\boldsymbol{p}}^{\mathrm{T}} \in \mathbb{C}^{K\times K}. \tag{12}$$

In line with (6), the resulting selected subset of sensors dependent MVDR beamformer is then given by

$$\boldsymbol{w}_{\boldsymbol{p}} = (\boldsymbol{h}_{\boldsymbol{p}}^{\mathrm{H}}\boldsymbol{\varPhi}_{\boldsymbol{nn},\boldsymbol{p}}^{-1}\boldsymbol{h}_{\boldsymbol{p}})^{-1}\boldsymbol{\varPhi}_{\boldsymbol{nn},\boldsymbol{p}}^{-1}\boldsymbol{h}_{\boldsymbol{p}}. \tag{13}$$

Notably, in general $\boldsymbol{\varPhi}_{\boldsymbol{nn},\boldsymbol{p}}^{-1} \neq \boldsymbol{\varSigma}_{\boldsymbol{p}}\boldsymbol{\varPhi}_{\boldsymbol{nn}}^{-1}\boldsymbol{\varSigma}_{\boldsymbol{p}}^{\mathrm{T}}$, unless $\boldsymbol{\varPhi}_{\boldsymbol{nn}}$ is a diagonal matrix with only uncorrelated noise present.

3.2 Energy-aware sensor selection based NR formulation

In large-scale WASNs, the energy consumption is an important performance indicator to measure the network efficiency of data processing. To improve this, it is thus expected to optimize the total energy consumption over a WASN by constraining the NR performance. To do this, we use $c_k, k \in \mathcal{M} = \{1, 2, \cdots, M\}$ to denote the transmission power per sample from microphone k to the FC. By this definition, it should be noted that the power for keep sensors activated is neglected, which exists in practice. Given the RTF, the proposed energy-aware MVDR sensor selection problem is formulated as

$$\begin{aligned} \min_{\boldsymbol{p},\boldsymbol{w}_{\boldsymbol{p}}} \ & \sum_{k=1}^{M} p_k c_k \\ \text{s.t.} \ & \boldsymbol{w}_{\boldsymbol{p}}^{\mathrm{H}}\boldsymbol{\varPhi}_{\boldsymbol{nn},\boldsymbol{p}}\boldsymbol{w}_{\boldsymbol{p}} \leqslant \frac{\beta}{\alpha}, \\ & \boldsymbol{w}_{\boldsymbol{p}}^{\mathrm{H}}\boldsymbol{h}_{\boldsymbol{p}} = 1, \ \boldsymbol{p} \in \{0,1\}^M, \end{aligned} \tag{14}$$

where β denotes the minimum noise power, and $0 < \alpha \leqslant 1$ controls the expected NR performance[1]. To focus on the sensor selection problem, in this work, we consider an ideal transmission scheme where the transmission rate is constant for every sensor and delays in the WASN are ignored, such

[1] In theory, the minimum output noise variance can only by calculated if all sensors are involved. However, this might be impossible because it is reasonable that in large-scale WASNs, the total number of microphones might be even unknown. In this case, we can set a specific value for β/α for the proposed model to indicate the expected NR performance, e.g., 40 dB.

that the transmission power from sensor k to the FC can be written as[32]

$$c_k = \kappa d_k^{\gamma}, \tag{15}$$

where κ is a constant ($\kappa \approx 10^{-10}$ J/(m$^{-\gamma}\cdot$bit)), γ is an attenuation factor (typically between 2 and 6), and d_k is the Euclidian transmission distance. It is clear that (14) is a nonconvex optimization problem due to the nonlinear selection operation and the Boolean variables $\boldsymbol{p}$. For simplicity, we consider the Lagrangian function of (14) as

$$\mathcal{L}(\boldsymbol{p}, \boldsymbol{w}_p, \lambda, \mu) = \sum_{k=1}^{M} p_k c_k + \lambda\left(\boldsymbol{w}_p^{\mathrm{H}} \boldsymbol{\Phi}_{nn,p} \boldsymbol{w}_p - \frac{\beta}{\alpha}\right) + \mu\left(\boldsymbol{w}_p^{\mathrm{H}} \boldsymbol{h}_p - 1\right),$$

where λ and μ are the non-negative Lagrange multipliers. The gradient with respect to $\boldsymbol{w}_p^*$ with $(\cdot)^*$ being the conjugate of complex numbers can be computed as

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}_p^*} = \lambda \boldsymbol{\Phi}_{nn,p} \boldsymbol{w}_p + \mu \boldsymbol{h}_p.$$

Setting $\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}_p^*}$ to zero, we obtain

$$\boldsymbol{w}_p = -\mu \boldsymbol{\Phi}_{nn,p} \boldsymbol{h}_p / \lambda. \tag{16}$$

Substituting $\boldsymbol{w}_p$ into the constraint $\boldsymbol{w}_p^{\mathrm{H}} \boldsymbol{h}_p = 1$ leads to

$$\frac{\mu}{\lambda} = -\frac{1}{\boldsymbol{h}_p^{\mathrm{H}} \boldsymbol{\Phi}_{nn,p} \boldsymbol{h}_p}. \tag{17}$$

It is clear that plugging (17) into (16) results in the sensor selection-dependent MVDR beamformer in (13). As a result, we can substitute the MVDR beamformer from (13) into (14) to avoid simultaneous optimization over the filter variable $\boldsymbol{w}_p$ and the selection unknown $\boldsymbol{p}$, resulting in a pure sensor selection problem:

$$\min_{\boldsymbol{p} \in \{0,1\}^M} \sum_{k=1}^{M} p_k c_k \quad \text{s.t.} \quad \boldsymbol{h}_p^{\mathrm{H}} \boldsymbol{\Phi}_{nn,p}^{-1} \boldsymbol{h}_p \geqslant \frac{\alpha}{\beta}, \tag{18}$$

which is still a nonconvex optimization problem but is constrained by SNR because the left-hand side of the constraint is the output SNR of the MVDR beamformer in (13).

3.3 An overview of existing approaches

To guide the reader, in this section, we will briefly review two related sensor selection approaches for (18).

optSdp-based method: In Ref. [23], a convex relaxation based approach was proposed for solving (18), where the noise covariance matrix $\boldsymbol{\Phi}_{nn}$ is first decomposed as

$$\boldsymbol{\Phi}_{nn} = \eta \boldsymbol{I} + \boldsymbol{G}, \tag{19}$$

where η is positive and $\boldsymbol{G}$ is positive definite. This decomposition can be obtained if η is smaller than the minimum eigenvalue of $\boldsymbol{\Phi}_{nn}$, because $\boldsymbol{\Phi}_{nn}$ is always positive definite due to the existence of correlated and uncorrelated noises. With this decomposition, it holds that

$$\boldsymbol{\Phi}_{nn,p} = \boldsymbol{\Sigma}_p \boldsymbol{\Phi}_{nn} \boldsymbol{\Sigma}_p^{\mathrm{T}} = \eta \boldsymbol{I}_K + \boldsymbol{\Sigma}_p \boldsymbol{G} \boldsymbol{\Sigma}_p^{\mathrm{T}}. \tag{20}$$

Based on (19), we can reformulate $\boldsymbol{h}_p^{\mathrm{H}} \boldsymbol{\Phi}_{nn,p}^{-1} \boldsymbol{h}_p$ as

$$\boldsymbol{h}_p^{\mathrm{H}} \boldsymbol{\Phi}_{nn,p}^{-1} \boldsymbol{h}_p = \boldsymbol{h}^{\mathrm{H}} \underbrace{\boldsymbol{\Sigma}_p^{\mathrm{T}} \left(\eta \boldsymbol{I}_K + \boldsymbol{\Sigma}_p \boldsymbol{G} \boldsymbol{\Sigma}_p^{\mathrm{T}}\right)^{-1} \boldsymbol{\Sigma}_p}_{\boldsymbol{Q}} \boldsymbol{h}, \tag{21}$$

where $\boldsymbol{Q}$ can be further reformulated as

$$\boldsymbol{Q} = \boldsymbol{G}^{-1} - \boldsymbol{G}^{-1} (\boldsymbol{G}^{-1} + \eta^{-1} \mathrm{diag}(\boldsymbol{p}))^{-1} \boldsymbol{G}^{-1}, \tag{22}$$

by applying the matrix inversion lemma[33]

$$\boldsymbol{C}(\boldsymbol{B}^{-1} + \boldsymbol{C}^{\mathrm{T}} \boldsymbol{A}^{-1} \boldsymbol{C})^{-1} \boldsymbol{C}^{\mathrm{T}} = \boldsymbol{A} - \boldsymbol{A}(\boldsymbol{A} + \boldsymbol{C}\boldsymbol{B}\boldsymbol{C}^{\mathrm{T}})^{-1} \boldsymbol{A}.$$

Substituting $\boldsymbol{Q}$ from (22) into $\boldsymbol{h}^{\mathrm{H}} \boldsymbol{Q} \boldsymbol{h} \geqslant \frac{\alpha}{\beta}$ results in

$$\boldsymbol{h}^{\mathrm{H}} \boldsymbol{G}^{-1} \boldsymbol{h} - \frac{\alpha}{\beta} \geqslant \boldsymbol{h}^{\mathrm{H}} \boldsymbol{G}^{-1} (\boldsymbol{G}^{-1} + \eta^{-1} \mathrm{diag}(\boldsymbol{p}))^{-1} \boldsymbol{G}^{-1} \boldsymbol{h},$$

which is equivalent to the linear matrix inequality (LMI) on the basis of the Schur complement[34]:

$$\begin{bmatrix} \boldsymbol{G}^{-1} + \eta^{-1} \mathrm{diag}(\boldsymbol{p}) & \boldsymbol{G}^{-1} \boldsymbol{h} \\ \boldsymbol{h}^{\mathrm{H}} \boldsymbol{G}^{-1} & \boldsymbol{h}^{\mathrm{H}} \boldsymbol{G}^{-1} \boldsymbol{h} - \frac{\alpha}{\beta} \end{bmatrix} \geq \boldsymbol{0}_{M+1}, \tag{23}$$

since the matrix $\boldsymbol{G}^{-1} + \eta^{-1} \mathrm{diag}(\boldsymbol{p})$ is always positive definite.

Finally, the Boolean variables are relaxed using continuous surrogates, i.e., $0 \leqslant p_k \leqslant 1, \forall k$. The final MVDR sensor selection problem is given by

$$\begin{aligned} &\min_{\boldsymbol{p} \in [0,1]^M} \sum_{k=1}^{M} p_k c_k \\ &\text{s.t.} \begin{bmatrix} \boldsymbol{G}^{-1} + \eta^{-1} \mathrm{diag}(\boldsymbol{p}) & \boldsymbol{G}^{-1} \boldsymbol{h} \\ \boldsymbol{h}^{\mathrm{H}} \boldsymbol{G}^{-1} & \boldsymbol{h}^{\mathrm{H}} \boldsymbol{G}^{-1} \boldsymbol{h} - \frac{\alpha}{\beta} \end{bmatrix} \geq \boldsymbol{0}_{M+1}, \end{aligned} \tag{24}$$

which is a standard SDP problem and can be efficiently solved using off-the-shelf solvers, e.g., CVX[34]. Note that the Boolean selection variable has to be recovered by rounding techniques. The computational complexity of (24), e.g., using interior-point method, is of the order of $O(M^3)$. Because the fact that this method requires the noise covariance matrix of the complete network, it is called the model-based optSdpRemoval method.

To alleviate the dependence on the complete noise covariance matrix, in Ref. [23], a data-based optSDP method was further proposed. In detail, given an initial point in the WASN (e.g., the FC) and a transmission range R_0, the sensors that are within the transmission range are considered as the candidate set $\mathcal{S}_R$. For the candidate set, the data-based method executes the SDP problem in (24) to find the best subset, where note that β needs to be replaced by the minimum noise power using all sensors in $\mathcal{S}_R$ (i.e., local constraint). Based on the selected subset $\mathcal{S}_+$ and the transmission range, the candidate set $\mathcal{S}_R$ is increased by including all R_0-closest sensors with respect to $\mathcal{S}_+$, and (24) is executed again. This procedure will be terminated when the local constraint converges. As the local constraint is somehow worse than the global constraint β depending on the whole network, the algorithm will switch to examine the global constraint, which requires several extra iterations. Since the sensors are gradually added to the selected subset $\mathcal{S}_+$, the data-based method is called optSdpAddition. In case J_1^+ iterations are required by optSdpAddition for convergence, the time complexity becomes $O(J_1^+ |\mathcal{S}_R|^3)$, where $|\mathcal{S}_R| < M$.

Utility-based method: In Refs. [20–22], backward and forward utility-based sensor selection approaches were proposed for linear minimum mean-square-error (LMMSE) estimator based signal enhancement. For the backward method, the most informative microphone subset is formed by gradually removing the sensor that has the smallest contribution to the NR performance, which is thus called utilityRemoval. In detail, for initialization, the selected subset $\mathcal{S}_+$ is set to be $\mathcal{M}$. Given the noise covariance matrix $\boldsymbol{\Phi}_{nn}$ of the complete network, at the first iteration, the utility is defined by

$$U_k = \boldsymbol{w}_{-k}^H \boldsymbol{\Phi}_{nn,-k} \boldsymbol{w}_{-k} - \boldsymbol{w}^H \boldsymbol{\Phi}_{nn} \boldsymbol{w} = \left(\boldsymbol{h}_{-k}^H \boldsymbol{\Phi}_{nn,-k}^{-1} \boldsymbol{h}_{-k}\right)^{-1} - (\boldsymbol{h}^H \boldsymbol{\Phi}_{nn}^{-1} \boldsymbol{h})^{-1}, \tag{25}$$

where $\boldsymbol{w}_{-k}$, $\boldsymbol{\Phi}_{nn,-k}$, and $\boldsymbol{h}_{-k}$ denote the MVDR filter, the noise covariance matrix and the RTF of the sensors contained in $\mathcal{S}_+$ excluding node k. The calculation of $\boldsymbol{\Phi}_{nn,-k}^{-1} \in \mathbb{C}^{(M-1)\times(M-1)}$ on the basis of $\boldsymbol{\Phi}_{nn}^{-1}$ has a time complexity of $\mathcal{O}(M^2)$. The sensor m that is to be removed can be obtained by searching the minimum U_k from $\mathcal{S}_+$, which has a complexity of $\mathcal{O}(M \log M)$. The selected subset is updated by excluding node m from $\mathcal{S}_+$. Then, the utility of the sensors contained in $\mathcal{S}_+$ has to be calculated again, and the second sensor has to be excluded similarly. This iterative procedure is stopped when the performance constraint is unsatisfied. Suppose that J_2^- iterations are required by utilityRemoval for convergence, the total time complexity is thus of the order of $\mathcal{O}(J_2^-(M^2 + M \log M))$.

To avoid dependence on complete noise statistics, a forward data-driven sensor selection approach was further proposed in Ref. [20], which is called utilityAddition. This is an opposite procedure as compared to utilityRemoval. In detail, given an initial point (e.g., FC) and a transmission range R_0, the selected subset $\mathcal{S}_+$ and the candidate set of sensors $\mathcal{S}_R$ that includes all R_0-closest sensors with respect to $\mathcal{S}_+$ can be initialized. The utility of the sensors contained in $\mathcal{S}_R$ can be calculated as

$$U_k = (\boldsymbol{h}_{\mathcal{S}_+}^{\mathrm{H}} \boldsymbol{\Phi}_{nn,\mathcal{S}_+}^{-1} \boldsymbol{h}_{\mathcal{S}_+})^{-1} - (\boldsymbol{h}_{\mathcal{S}_+,+k}^{\mathrm{H}} \boldsymbol{\Phi}_{nn,\mathcal{S}_+,+k}^{-1} \boldsymbol{h}_{\mathcal{S}_+,+k})^{-1}, k \in \mathcal{S}_R. \tag{26}$$

Searching the maximum value in $\boldsymbol{u} = [U_1, U_2, \ldots, U_{|\mathcal{S}_R|}]$ can then reveal the sensor in $\mathcal{S}_R$ that has the largest contribution to NR with respect to $\mathcal{S}_+$. The selected subset $\mathcal{S}_+$ is then updated by adding this sensor. Subsequently, the selected subset $\mathcal{S}_+$ and the candidate set $\mathcal{S}_R$ are updated similarly. This procedure will be similarly terminated until the residual noise power is smaller than or equal to the predefined threshold. In case J_2^+ iterations are executed, the time complexity of utilityAddition is of the order of $\mathcal{O}(J_2^+(|\mathcal{S}_R|^2 + |\mathcal{S}_R| \log |\mathcal{S}_R|))$, because the time complexities of calculating $\boldsymbol{\Phi}_{nn,\mathcal{S}_+,+k}^{-1}$ based on $\boldsymbol{\Phi}_{nn,\mathcal{S}_+}^{-1}$ and searching the maximum value are $\mathcal{O}(|\mathcal{S}_R|^2)$ and $\mathcal{O}(|\mathcal{S}_R| \log |\mathcal{S}_R|)$, respectively. Roughly, the time complexity of both utilityAddition and utilityRemoval are cubic in terms of M, as only one sensor is removed or added at each iteration (i.e., J_2^+ and J_2^- are linear in M).

4 Proposed low-complexity approaches

As the time complexity of optSdp is cubic and the utility-based approach does not consider energy usage, we propose three low-complexity energy-aware methods in this section.

4.1 Proposed weighted utility-based approach

From the perspective of optimization, (18) can be equivalently reformulated as

$$\max_{\boldsymbol{p} \in \{0,1\}^M} g(\boldsymbol{p}) = \frac{\boldsymbol{h}_p^{\mathrm{H}} \boldsymbol{\Phi}_{nn,p}^{-1} \boldsymbol{h}_p}{\sum_{k=1}^{M} p_k c_k} \quad \text{s.t.} \quad \boldsymbol{h}_p^{\mathrm{H}} \boldsymbol{\Phi}_{nn,p}^{-1} \boldsymbol{h}_p \geqslant \frac{\alpha}{\beta}, \tag{27}$$

where the numerator can be decomposed as

$$\boldsymbol{h}_p^{\mathrm{H}} \boldsymbol{\Phi}_{nn,p}^{-1} \boldsymbol{h}_p = \boldsymbol{h}_{-k}^{\mathrm{H}} \boldsymbol{\Phi}_{nn,p,-k}^{-1} \boldsymbol{h}_{-k} + \boldsymbol{h}_k^{\mathrm{H}} \boldsymbol{\Phi}_{nn,k}^{-1} \boldsymbol{h}_k, \tag{28}$$

in here the two terms on the right-hand side represent the output SNR using a sensor subset $\mathcal{S}_+$ and the SNR gain by adding sensor k to $\mathcal{S}_+$. Therefore, we can define the ratio between the SNR gain and the individual energy consumption as the contribution to the NR problem, which is called weighted utility in this work. Similar to the optSdp and utility-based methods, we can also design model- and data-driven sensor selection approaches using the weighted utility metric.

For backward selection, in the first iteration, the selected subset $\mathcal{S}_+$ is initialized by $\mathcal{M}$. Given the noise covariance matrix $\boldsymbol{\Phi}_{nn}$, the weighted utility is thus defined by

$$U_k = \frac{\boldsymbol{h}^H \boldsymbol{\Phi}_{nn}^{-1} \boldsymbol{h} - \boldsymbol{h}_{-k}^{\mathrm{H}} \boldsymbol{\Phi}_{nn,-k}^{-1} \boldsymbol{h}_{-k}}{c_k}, \tag{29}$$

where the numerator denotes the SNR loss by removing node k. Then, the m-th sensor that has the minimum weighted utility can be removed from $\mathcal{S}_+$. Clearly, the sensor that has the minimum SNR loss and maximum energy consumption will be excluded, i.e., $\mathcal{S}_+ \leftarrow \mathcal{S}_+ \backslash m$. Then, the weighted utility of the sensors contained in $\mathcal{S}_+$ needs to be calculated again. This procedure will be repeated until the constraint in (27) is unsatisfied any more, which is thus called weightedUtilityRemoval. By inspection, the time complexity of the model-driven weightedUtilityRemoval method is the same as that of the utility-based counterpart.

In contrast, given an initial point (e.g., FC) and a transmission range R_0, it is straightforward to design a forward data-driven weighted utility-based sensor selection approach, which is called weightedUtilityAddition in this work. The candidate set of sensors $\mathcal{S}_R$ can be initialized with R_0-closest sensors with respect to the initial point, and the m-th sensor to be added to $\mathcal{S}_+$ can be searched using

$$m = \arg\max U_k, \quad k \in \mathcal{S}_R, \tag{30}$$

where U_k is given by

$$U_k = \frac{\boldsymbol{h}_{\mathcal{S}_+ +k}^{\mathrm{H}} \boldsymbol{\Phi}_{nn,\mathcal{S}_+,+k}^{-1} \boldsymbol{h}_{\mathcal{S}_+,+k} - \boldsymbol{h}_{\mathcal{S}_+}^{\mathrm{H}} \boldsymbol{\Phi}_{nn,\mathcal{S}_+}^{-1} \boldsymbol{h}_{\mathcal{S}_+}}{c_k}. \tag{31}$$

The candidate set $\mathcal{S}_R$ with respect to $\mathcal{S}_+$ and the corresponding utility must be updated accordingly. Clearly, at each iteration one sensor will be added to $\mathcal{S}_+$. Similar to the utilityAddition method, the stopping criterion of the proposed iterative process can be checked if the performance bound is satisfied as

$$\boldsymbol{h}_{\mathcal{S}_+}^{\mathrm{H}} \boldsymbol{\Phi}_{nn,\mathcal{S}_+}^{-1} \boldsymbol{h}_{\mathcal{S}_+} \geqslant \frac{\alpha}{\beta}. \tag{32}$$

The time complexity of weightedUtilityAddition is similar to that of utilityAddition; however, weightedUtilityAddition takes energy consumption into account.

4.2 Gradient-based approach

As the gradient measures the changing rate of a function along a certain direction, it can be used to define the contribution of an individual sensor to the energy-aware NR performance. From Section 3.3, we know that

$$\boldsymbol{h}_p^{\mathrm{H}}\boldsymbol{\Phi}_{nn,p}^{-1}\boldsymbol{h}_p=\boldsymbol{h}^{\mathrm{H}}\boldsymbol{G}^{-1}\boldsymbol{h}-\boldsymbol{h}^{H}\boldsymbol{G}^{-1}\left(\boldsymbol{G}^{-1}+\eta^{-1}\mathrm{diag}(\boldsymbol{p})\right)^{-1}\boldsymbol{G}^{-1}\boldsymbol{h}. \tag{33}$$

Considering the objective function of (27), the partial gradient with respect to p_k can be calculated as

$$\frac{\partial g(\boldsymbol{p})}{p_k}=\left(\sum_{k=1}^{M}p_kc_k\right)^{-2}\left(\frac{|v_k|^2}{\eta}\sum_{k=1}^{M}p_kc_k-c_k\boldsymbol{h}_p^{\mathrm{H}}\boldsymbol{\Phi}_{nn,p}^{-1}\boldsymbol{h}_p\right),$$

where v_k denotes the k-th element of vector $\boldsymbol{v}$, given by

$$\boldsymbol{v}=(\boldsymbol{G}^{-1}+\eta^{-1}\ \mathrm{diag}(\boldsymbol{p}))^{-1}\boldsymbol{G}^{-1}\boldsymbol{h}.$$

Considering the case that only sensor m in $\mathcal{M}=\{1,\cdots,M\}$ is unselected, i.e., $p_k=1,\forall k$ except $k=m$, $\boldsymbol{h}_p^{\mathrm{H}}\boldsymbol{\Phi}_{nn,p}^{-1}\boldsymbol{h}_p$ in the gradient can be interpreted as the output SNR using the remaining $M-1$ sensors. The complexity of the calculation of the gradient is of the order of $O(M^2)$ due to the computation of the inversion $\boldsymbol{\Phi}_{nn,p}^{-1}$ using $\boldsymbol{\Phi}_{nn}^{-1}$.

Based on the gradient, we can design a model-driven microphone subset selection method, termed by gradRemoval in this work. In detail, we initialize $\boldsymbol{p}$ using an all-ones vector, meaning that all sensors are selected at the beginning. At the first iteration the sensor with the smallest gradient has to be removed from the selected set $\mathcal{S}_+=\mathcal{M}$. By calculating the gradient, such a sensor can be determined by

$$m=\arg\min\frac{\partial g(\boldsymbol{p})}{p_k},\quad k\in\mathcal{S}_+, \tag{34}$$

because the vector $\boldsymbol{p}$ with $p_m=0$ and ones elsewhere gives the optimal direction for keeping the objective function of (27) unchanged. In other words, the m-th sensor contributes least to increasing $g(\boldsymbol{p})$. Next, the gradients for the remaining $M-1$ sensors in $\mathcal{S}_+$ need to be updated, and the sensor to be removed is searched similarly. During the first few iterations, many sensors are included in the selected subset, and the obtained performance will be greater than the threshold. Therefore, the proposed method can be stopped if the performance is smaller than the given threshold, i.e., $\boldsymbol{h}_{\mathcal{S}_+}^{\mathrm{H}}\boldsymbol{\Phi}_{nn,\mathcal{S}_+}^{-1}\boldsymbol{h}_{\mathcal{S}_+}\leqslant\dfrac{\alpha}{\beta}$.

The forward data-driven gradient-based microphone selection method, which is called gradAddition, can be designed similarly to utilityAddition and weightedUtilityAddition. Specifically, at each iteration given the selected subset $\mathcal{S}_+$ and the candidate set $\mathcal{S}_R$, the sensor in $\mathcal{S}_R$ that has the largest gradient is added to $\mathcal{S}_+$ until the constraint in (27) is satisfied. The time complexity of the proposed gradient-based methods can be analyzed in line with the utility-based or weighted utility-based counterparts.

4.3 Weighted SNR-based sensor selection approach

Intuitively, in practice, the sensors around the target speaker have a higher SNR, i.e., better speech quality, which is beneficial for the estimation of the speech source; the ones around the FC have a smaller transmission distance, i.e., transmission power, which is useful for reducing the network resource consumption. This observation can also be found in Ref. [23]. In the proposed weighted utility-based approach, the calculation of the utility is based on the output SNR, which requires a squared time complexity. Motivated by Ref. [23] and to reduce the complexity of utility-based methods, we can define a weighted input SNR to approximate (29), which is given by

$$U_k=\frac{\mathrm{inSNR}_k(\omega)}{c_k},\quad k\in\{1,\cdots,M\}, \tag{35}$$

where the narrowband input SNR is defined as

$$\mathrm{inSNR}_k(\omega)=\frac{\sum_l|X_k(\omega,l)|^2}{\sum_l|N_k(\omega,l)|^2}. \tag{36}$$

The sensor SNRs can be computed efficiently locally without data transmission using noise PSD estimators in Ref. [35]. Clearly, when applying the weighted SNR to sensor selection, the sensors that have a high SNR and/or a low transmission energy are more likely to be selected. The proposed weighted input SNR can then be used to design backward model-driven and forward data-driven sensor selection approaches similarly to the utility and weighted utility-based approaches, which are called weightedSnrRemoval and weightedSnrAddition. Note that for weightedSnrRemoval and weightedSnrAddition, at each iteration, we only need to search the minimum or maximum element from the weighted inSNR vector, which has a complexity of $O(K\log K)$. Suppose that J_3^- and J_3^+ iterations are required by weightedSnrRemoval and weightedSnrAddition, the time complexities are of the order of $O(J_3^-M\log M)$ and $O(J_3^+K\log K)$.

4.4 Summary and discussion

Based on the previous analysis, we summarize the considered model-driven and data-driven sensor selection approaches in Table 1. Note that in practice, the number of required iterations for the utility, weighted utility and gradient-based approaches might be different from each other. For notational conciseness, we merely use J_2^- and J_2^+ for the model-based and data-based approaches, respectively. In each iteration, the time complexity is dependent on the cardinality of the candidate set $\mathcal{S}_R$, which is linear in terms of the number of selected sensors K. Hence, for ease of comparison, we use M and K to measure the time complexities of the backward and forward methods, respectively.

It is clear that if the number of required sensors is much smaller than M, that is, the performance requirement is low, the data-based method is much more computationally efficient than the model-based counterpart. The proposed SNR-based method is computationally cheapest. For the model-based approaches, in case $J_2^-\ll M$, i.e., the performance re-

Table 1. The summary of the considered sensor selection approaches.

Type	Method	Time complexity
Model-based	optSdpRemoval[23]	$O(M^3)$
	utilityRemoval[21]	$O(J_2^-(M^2+M\log M))$
	weightedUtilityRemoval	$O(J_2^-(M^2+M\log M))$
	gradRemoval	$O(J_2^-(M^2+M\log M))$
	weightedSnrRemoval	$O(J_3^- M\log M)$
Data-based	optSdpAddition[23]	$O(J_1^+ K^3)$
	utilityAddition[21]	$O(J_2^+(K^2+K\log K))$
	weightedUtilityAddition	$O(J_2^+(K^2+K\log K))$
	gradAddition	$O(J_2^+(K^2+K\log K))$
	weightedSnrAddition	$O(J_3^+ K\log K)$

quirement is high, utilityRemoval, weightedUtilityRemoval, and gradRemoval might be computationally faster. Moreover, in general, the implementation of utilityAddition, weightedUtilityAddition, and gradAddition is more efficient than optSdpAddition, as both J_1^+ and J_2^+ are roughly linear in terms of K. Therefore, we can conclude that compared to the optSdp-based method, which is optimal in terms of performance, the proposed weighted utility and gradient-based approaches are more computationally efficient; compared to the utility-based method, the proposed methods are more energy efficient, as the optimization of transmission power is taken into account. It is worth noting that the forward data-based methods are more practical.

5 Numerical results

In this section, the proposed model- and data-driven sensor selection MVDR beamforming approaches will be validated via numerical simulations.

Fig. 1 depicts the employed microphone setup, where we place 169 microphones uniformly in a 2D room with dimensions of (12, 12) m. The target speaker and the FC are located at (3, 9) m and (9, 3) m. Two coherent noise sources are placed at (3, 3) m and (9, 9) m. The speech source is derived from the TIMIT database[36], and the noise signals are from the NoiseX-92 database[37], respectively. We utilize the image method[38] to generate the room impulse responses (RIRs) from sources to microphones, which are convolved with the source signals to produce the time-domain components in the signal model. The uncorrelated noise is assumed to be the sensor thermal noise and modeled as a white Gaussian random process. The microphone signal is generated by summing the signal component, the correlated noise component at a signal-to-interferer ratio (SIR) of 0 dB and the uncorrelated noise component at an SNR of 50 dB. The sampling frequency and the reverberation time are set to 16 kHz and 200 ms, respectively. The power for keeping sensors active is assumed to be the same for all sensors, and the transmission power from sensors $k, \forall k$ to the FC is initialized using the squared distance, i.e., $\gamma = 2$ in (15).

Fig. 1 shows some typical selection cases obtained by the model-based and data-based approaches for α = 0.6, where the number of iterations required for achieving the desired performance is also marked. For the model-driven methods, it is clear that optSdpRemoval, the proposed weightedSnrRemoval and gradRemoval methods obtain similar selection results, as some sensors around the target source, some around the FC and two sensors next to the noise sources are selected. The inclusion of these informative sensors is meaningful for source estimation, saving transmission energy and cancelling interferers. The utilityRemoval method does not choose sensors close to the FC, leading to more energy consumption compared to other methods. The proposed weightedSnrRemoval method cannot choose the microphones next to the interfering sources since these sensors do not have a high input SNR or a low transmission distance. Given the FC as the initial point for data-driven methods, the selected sensor subset of all methods increases from the FC to the target source. The data-driven methods fail to sparsely select two sensors next to the interfering sources, as they are never activated in the candidate set before the desired performance is achieved. We can conclude that in general, the selection result of data-based methods will not converge to that of the model-driven counterparts, as the selection is only constrained on the NR performance.

Fig. 2 shows the residual noise power in dB and the resource consumption of the forward data-driven methods in terms of α. The results for the backward model-based methods can be shown similarly. Obviously, all methods satisfy the performance requirement. The optSdpAddition method and weightedSnrAddition consume the lowest and highest transmission energy, respectively. The proposed weightedUtilityAddition and gradAddition methods achieve comparable performance, and the energy cost is lower than that of utilityAddition.

Finally, we compare the execution time in terms of α in Fig. 3. The simulations are conducted using a MacBook Pro with an Intel Core i5 processor. For the model-driven methods, the low-complexity weightedSnrRemoval method is fastest. The runtime of optSdpRemoval remains constant, as it executes the same SDP optimization problem for any α. The proposed weightedUtilityRemoval and gradRemoval methods are slower than optSdpRemoval for small α-values, since a smaller α means a lower performance bound and more sensors that have to be excluded, i.e., more iterations. However, when α is large, weightedUtilityRemoval and gradRemoval become faster. For the data-driven methods, the low-complexity weightedSnrAddition method is still the fastest. The optSdpAddition method becomes the slowest because it runs an SDP of cubic complexity at each iteration, even though it requires much fewer iterations than the proposed data-driven methods, as shown in Fig. 1. Clearly, the proposed methods are computationally much more efficient than optSdpAddition. Note that for small α-values, the data-driven method is more efficient than the model-driven counterpart, as much fewer sensors need to be added to the selected subset.

6 Conclusions

In this work, we presented several narrowband model-based and data-driven sensor selection methods for MVDR beamforming-based NR problems in large-scale distributed microphone networks. The energy-aware NR problem was built

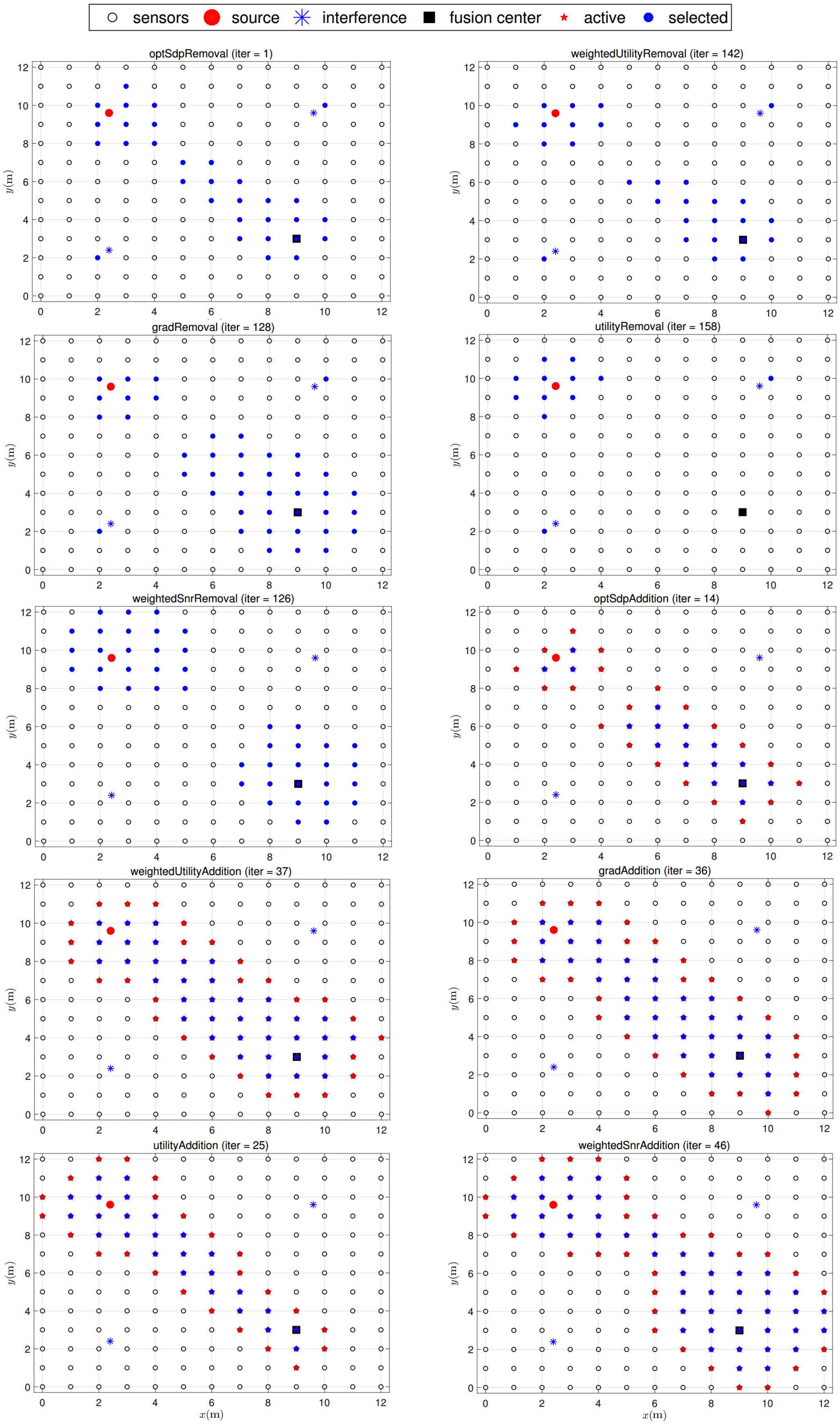


Fig. 1. Sensor selection examples of the model- and data-driven approaches for $\alpha = 0.6$. Note that active sensors are required by the data-driven methods, but are not required by the model-based counterparts.

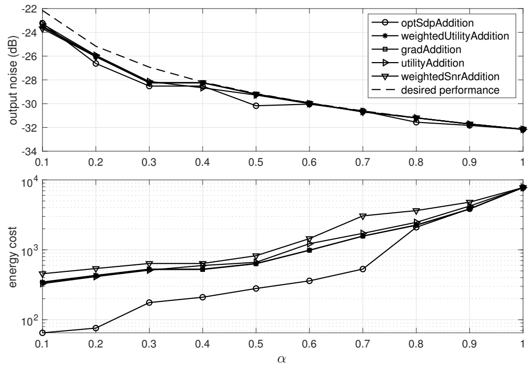


Fig. 2. The output noise and energy cost of data-driven approaches vs α.

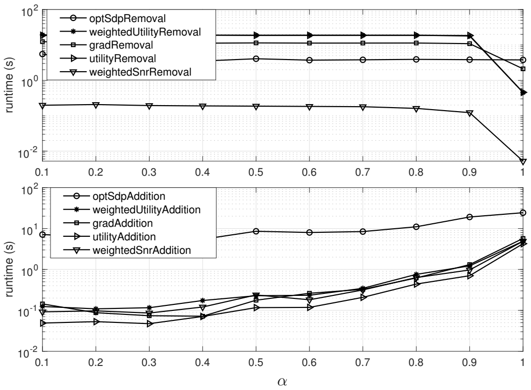


Fig. 3. The time consumption for performance requirement vs α.

upon minimizing the power consumption and constraining the desired NR performance such that the network efficiency can be optimized. Motivated by the SDP solver and utility-based approach, we proposed using the weighted utility, gradient and weighted input SNR to select an informative microphone subset. For each metric, we designed backward model-based and forward data-driven selection approaches, and the former is based on the noise statistics of the complete network. It was shown that the proposed methods can effectively choose the sensors around the target speaker and those around the FC for target enhancement and energy optimization. The proposed methods are computationally more efficient than the SDP-based method and more energy efficient than the utility-based method. We can conclude that in large-scale WASNs, many sensor measurements are redundant for NR and choosing a subset of sensors can be sufficient for performance requirements. Data-driven sensor selection methods are more efficient and practical for the design of energy-aware WASNs.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (62101523), Hefei Municipal Natural Science Foundation (2022012), Fundamental Research Funds for the Central Universities (WK2100000016), and USTC Research Funds of the Double First-Class Initiative (YD2100002008).

Conflict of interest

The authors declare that they have no conflict of interest.

Biography

Jie Zhang received the B.S. degree, master's degree, and Ph.D. degree in Electrical Engineering from the Yunnan University, Peking University, and the Delft University of Technology in 2012, 2015, and 2020, respectively. He is currently an Associate Researcher in the National Engineering Research Center for Speech and Language Information Processing (NERC-SLIP), Faculty of Information Science and Technology, University of Science and Technology of China. He received the Best Student Paper Award for his publication at the 10th IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM). His team also won several awards in speech-related academic competitions (e.g., DiCOVA-ICASSP2022, NIST OpenASR2021, L3DAS23). His current research interests include multi-microphone speech processing, binaural auditory, speech recognition, and wireless (acoustic) sensor networks.

References

[1] Haller S, Karnouskos S, Schroth C. The Internet of things in an enterprise context. In: Domingue J, Fensel D, Traverso P, editors. Future Internet–FIS 2008. Berlin: Springer, **2008**.

[2] Adulyasas A, Sun Z, Wang N. Connected coverage optimization for sensor scheduling in wireless sensor networks. *IEEE Sensors Journal,* **2015**, *15* (17): 3877–3892.

[3] Turchet L, Fazekas G, Lagrange M, et al. The Internet of audio things: State of the art, vision, and challenges. *IEEE Internet of Things Journal,* **2020**, *7* (10): 10233–10249.

[4] Meng Y, Wang Z, Zhang W, et al. WiVo: Enhancing the security of voice control system via wireless signal in IoT environment. In: Proceedings of the Eighteenth ACM International Symposium on Mobile Ad Hoc Networking and Computing. New York: ACM, **2018**: 81–90.

[5] Wang Q, Guo S, Yiu K F C. Distributed acoustic beamforming with blockchain protection. *IEEE Transactions on Industrial Informatics,* **2020**, *16* (11): 7126–7135.

[6] Zou Q, Zou X, Zhang M, et al. A robust speech detection algorithm in a microphone array teleconferencing system. In: 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Salt Lake City, USA: IEEE, **2001**: 3025–3028.

[7] Gustafsson S, Martin R, Vary P. Combined acoustic echo control and noise reduction for hands-free telephony. *Signal Processing,* **1998**, *64* (1): 21–32.

[8] Moore D C, McCowan I A. Microphone array speech recognition: Experiments on overlapping speech in meetings. In: 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Hong Kong, China: IEEE, **2003**: V–497.

[9] Lee S C, Chen B W, Wang J F. Noisy environment-aware speech enhancement for speech recognition in human-robot interaction application. In: 2010 IEEE International Conference on Systems, Man and Cybernetics. Istanbul: IEEE, **2010**: 3938–3941.

[10] Amini J, Hendriks R C, Heusdens R, et al. Spatially correct rate-constrained noise reduction for binaural hearing aids in wireless acoustic sensor networks. *IEEE/ACM Transactions on Audio, Speech, and Language Processing,* **2020**, *28*: 2731–2742.

[11] Zeng Y, Hendriks R C. Distributed delay and sum beamformer for speech enhancement via randomized gossip. *IEEE/ACM Transactions on Audio, Speech, and Language Processing,* **2014**, *22* (1): 260–273.

[12] Guan Q, Ji F, Liu Y, et al. Distance-vector-based opportunistic routing for underwater acoustic sensor networks. *IEEE Internet of Things Journal,* **2019**, *6*: 3831–3839.

[13] Benesty J, Makino S, Chen J. Speech Enhancement. Berlin: Springer, **2005**.

[14] Zhang J, Heusdens R, Hendriks R C. Rate-distributed spatial filtering based noise reduction in wireless acoustic sensor networks. *IEEE/ACM Transactions on Audio, Speech, and Language*

Processing, **2018**, *26* (11): 2015–2026.

[15] Joshi S, Boyd S. Sensor selection via convex optimization. *IEEE Transactions on Signal Processing,* **2009**, *57* (2): 451–462.

[16] Chepuri S P, Leus G. Sparsity-promoting sensor selection for non-linear measurement models. *IEEE Transactions on Signal Processing,* **2015**, *63* (3): 684–698.

[17] Golovin D, Faulkner M, Krause A, Online distributed sensor selection. In: IPSN '10: Proceedings of the 9th ACM/IEEE International Conference on Information Processing in Sensor Networks. New York: ACM, **2010**: 220–231.

[18] Zhang H, Moura J M F, Krogh B. Dynamic field estimation using wireless sensor networks: Tradeoffs between estimation error and communication cost. *IEEE Transactions on Signal Processing,* **2009**, *57* (6): 2383–2395.

[19] Liu S, Chepuri S P, Fardad M, et al. Sensor selection for estimation with correlated measurement noise. *IEEE Transactions on Signal Processing,* **2016**, *64*: 3509–3522.

[20] Bertrand A, Moonen M. Efficient sensor subset selection and link failure response for linear MMSE signal estimation in wireless sensor networks. In: 2010 18th European Signal Processing Conference. Aalborg, Denmark : IEEE, **2010**: 1092–1096.

[21] Szurley J, Bertrand A, Moonen M, et al. Energy aware greedy subset selection for speech enhancement in wireless acoustic sensor networks. In: 2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO). Bucharest, Romania: IEEE, **2012**: 789–793.

[22] Bertrand A. Utility metrics for assessment and subset selection of input variables for linear estimation [tips & tricks]. *IEEE Signal Processing Magazine,* **2018**, *35* (6): 93–99.

[23] Zhang J, Chepuri S P, Hendriks R C, et al. Microphone subset selection for MVDR beamformer-based noise reduction. *IEEE/ACM Transactions on Audio, Speech, and Language Processing,* **2018**, *26* (3): 550–563.

[24] Zhang J, Du J, Dai L R. Sensor selection for relative acoustic transfer function steered linearly-constrained beamformers. *IEEE/ACM Transactions on Audio, Speech, and Language Processing,* **2021**, *29*: 1220–1232.

[25] Zhang J, Zhang G, Dai L. Frequency-invariant sensor selection for MVDR beamforming in wireless acoustic sensor networks. *IEEE Transactions on Wireless Communications,* **2022**, *21*: 10648–10661.

[26] Bertrand A, Szurley J, Ruckebusch P, et al. Efficient calculation of sensor utility and sensor removal in wireless sensor networks for adaptive signal estimation and beamforming. *IEEE Transactions on Signal Processing,* **2012**, *60* (11): 5857–5869.

[27] Zhang J, Chen H, Dai L R, et al. A study on reference microphone selection for multi-microphone speech enhancement. *IEEE/ACM Transactions on Audio, Speech, and Language Processing,* **2021**, *29*: 671–683.

[28] Zhang J, Heusdens R, Hendriks R C. Relative acoustic transfer function estimation in wireless acoustic sensor networks. *IEEE/ACM Transactions on Audio, Speech and Language Processing,* **2019**, *27* (10): 1507–1519.

[29] Frost O L. An algorithm for linearly constrained adaptive array processing. *Proceedings of the IEEE,* **1972**, *60* (8): 926–935.

[30] Van Veen B, Buckley K. Beamforming: A versatile approach to spatial filtering. *IEEE ASSP Magazine,* **1988**, *5* (2): 4–24.

[31] Capon J. High-resolution frequency-wavenumber spectrum analysis. *Proceedings of the IEEE,* **1969**, *57* (8): 1408–1418.

[32] Ciullo D, Celik G D, Modiano E. Minimizing transmission energy in sensor networks via trajectory control. In: 8th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks. Avignon, France: IEEE, **2010**: 132–141.

[33] Petersen K B, Pedersen M S. The Matrix Cookbook. Technical University of Denmark, **2008**: 15.

[34] Boyd S, Vandenberghe L. Convex optimization. Cambridge, UK: Cambridge University Press, **2004**.

[35] Hendriks R C, Heusdens R, Jensen J. MMSE based noise PSD tracking with low complexity. In: 2010 IEEE International Conference on Acoustics, Speech and Signal Processing. Dallas, USA: IEEE, **2010**: 4266–4269.

[36] Garofolo J, Lamel L, Fisher W, et al. DARPA TIMIT acoustic-phonetic speech database. National Institute of Standards and Technology (NIST), **1988**, 15: 29–50.

[37] Varga A, Steeneken H J M. Assessment for automatic speech recognition II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems. *Speech Communication,* **1993**, *12* (3): 247–251.

[38] Allen J B, Berkley D A. Image method for efficiently simulating small-room acoustics. *The Journal of the Acoustical Society of America,* **1979**, *65* (4): 943.

DOI: 10.52396/JUSTC-2022-0121