Data-Driven Optimization Strategy of Microphone Array Configurations in Vehicle Environments

Lehai Liu, Fengrong Bi, Jiewei Lin, Tongtong Qi, and Xin Li, *Member, IEEE*

***Abstract*— Microphone array (MA) speech enhancement is a crucial component of vehicle intelligence. However, the complex acoustic environments and the spatial constraints of array layouts present challenges for the design and implementation of MAs in intelligent vehicles. This study proposes a data-driven optimization strategy for constructing the optimal MA configuration in-vehicle environments. We first developed a novel in-vehicle noise model that considers azimuth and elevation angles by defining a search region for microphone elements in a plane. Subsequently, based on the in-vehicle noise model, we conducted sound field modeling to ensure the designed MA is compatible with the complex acoustic environments inside vehicles. Utilizing this sound field model, we formulated a specialized optimization algorithm to devise the optimal configuration of the MA. Finally, the designed array configuration was constructed using an MEMS MA acquisition system, and the array performance was evaluated in real driving environments. Compared to conventional MA configurations, comprehensive experiments indicate that the designed MA enhances performance by increasing the short-time objective intelligibility (STOI) scores by 13.9%, improving the output signal-to-noise ratio (SNR) levels by 53.3%, and ensuring robustness in complex in-vehicle acoustic environments.**

***Index Terms*— Array configuration design, in-vehicle acoustic environments, optimization strategy, planar microphone array (MA), speech enhancement.**

I. Introduction

As VEHICLE intelligence advances, speech interaction systems have become standard in modern cars, driving the evolution of vehicle intelligence. Speech interactive systems minimize driver distractions compared to manual controls, enhancing safety and driving enjoyment. However, environmental noise during driving compromises speech signal quality [1], impacting speech recognition accuracy and user experience. Advanced signal-processing techniques are necessary to capture and interpret speech commands correctly under driving conditions.

Advanced microphone arrays (MAs) and signal-processing algorithms are being integrated into modern vehicles to isolate the driver's speech from the surrounding noise. Beamforming, an essential MA signal-processing technique, enhances signal clarity and recognition by selectively boosting desired signals and reducing noise from other directions. This technique is widely applied in fields like hearing aids [2], speech recognition [3], and mobile communications [4]. Furthermore, compared to single-channel methods [5], [6], MA speech enhancement is superior in spatial filtering [7], [8], [9], making it the preferred choice for signal acquisition and speech enhancement in vehicular environments [1], [10], [11], [12], [13], [14]. Numerous studies, highlighted in references [15], [16], [17], [18], [19], [20], concentrate on optimizing the beamforming algorithm. However, we find that the geometry, position, and number of microphones significantly influence the system's performance. An optimal configuration can substantially improve system efficiency, as demonstrated in references [21], [22], [23]. Research on MA design has primarily focused on optimizing conventional arrays [24], [25], [26], [27], [28], such as linear and circular arrays. Nevertheless, irregular MAs outperform conventional arrays in broadband beamforming [29], [30]. On the other hand, vehicle interior space constraints limit the applicability of conventional configurations. In recent years, researchers have explored various innovative array configurations to achieve more effective signal acquisition, including triangular arrays [31], [32], arrays aligned with the direction of driving [33], and arrays designed along the contours of internal vehicle components, such as dashboards and speedometer panels [34]. However, the selection of these configurations is primarily based on the experience of researchers, and the suitability for the in-vehicle environment has yet to be validated. Therefore, investigating irregular array configurations in in-vehicle environments holds particular importance.

Recent research has made significant progress in optimizing irregular MAs in simulated environments [30], [35], [36]. However, simulations typically assume that environmental noise is uniform and controllable, while the acoustic environment in vehicles is much more complex. In previous research, MAs' design or optimization process in vehicular environments commonly relied on diffuse noise field models to simulate the acoustic environments inside the vehicle [1], [10], [11], [12], [13], [14]. However, the acoustic environments in the interior of a vehicle are notably more complex. As documented in the study [37], a recent investigation involving real noise data collection inside a vehicle under normal driving conditions revealed significant discrepancies between real in-vehicle noise fields and ideal diffuse fields. These discrepancies indicate that using diffuse field noise as a proxy for real in-vehicle noise is inaccurate. This could significantly reduce the performance of MAs

Received 6 September 2024; accepted 2 October 2024. Date of publication 25 October 2024; date of current version 12 November 2024. This work was supported by the National Key Research and Development Program of China under Grant 2021YFD2000303. The Associate Editor coordinating the review process was Dr. Rajan Sarkar. *(Corresponding author: Tongtong Qi.)*

Lehai Liu, Fengrong Bi, and Jiewei Lin are with the School of Mechanical Engineering, Tianjin University, Tianjin 300072, China.

Tongtong Qi is with the Institute of Internal Combustion Engines, Tianjin University, Tianjin 300072, China (e-mail: qitongtong0131@sina.com).

Xin Li is with the College of Computing and Data Science, Nanyang Technological University, Singapore 639798.

Digital Object Identifier 10.1109/TIM.2024.3485461

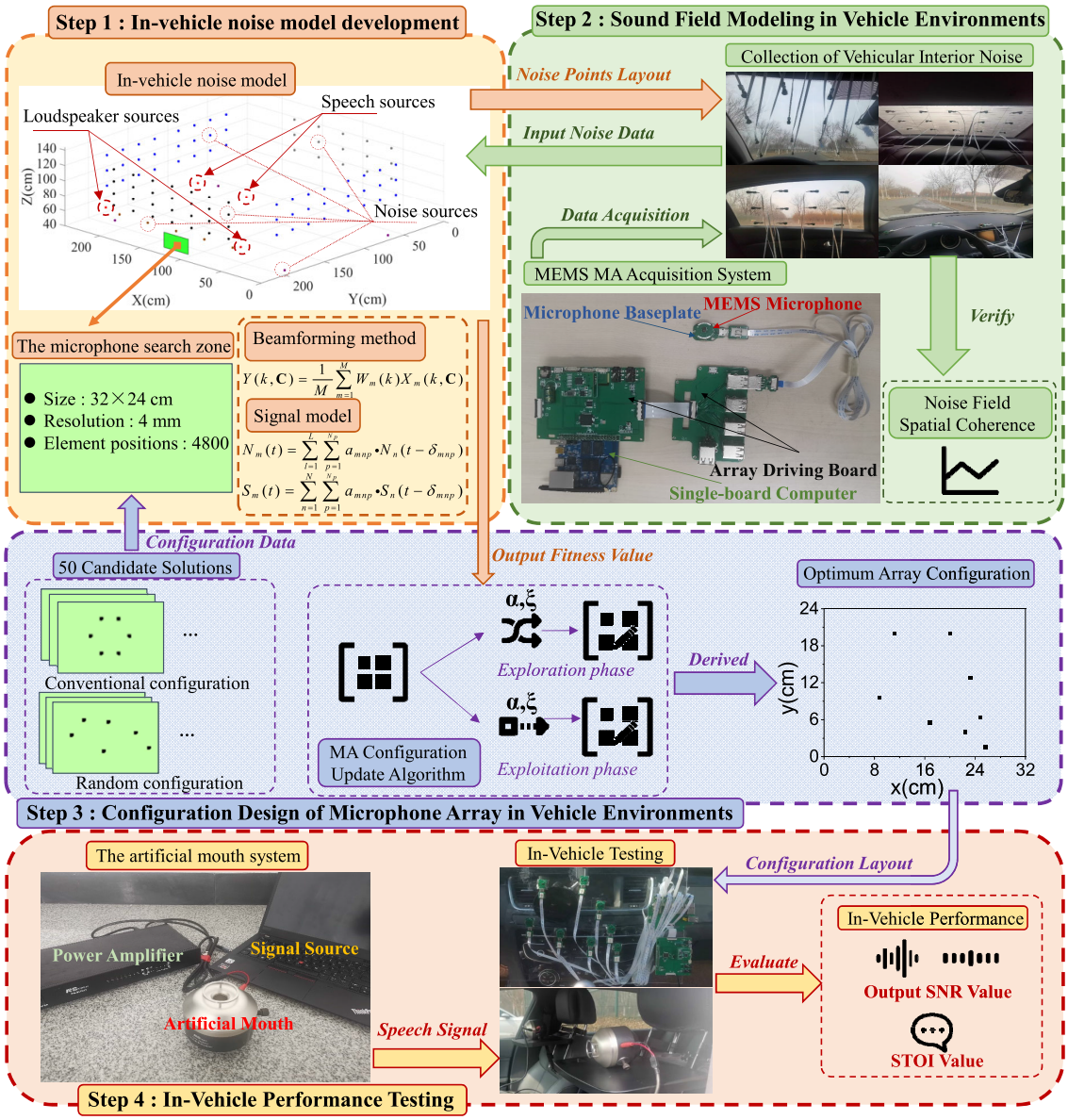


Fig. 1. Optimization strategy for MAs in the vehicle environment.

designed or optimized based on diffuse noise fields when deployed in real vehicular environments. Thus, it is essential to research the modeling of real sound fields in vehicular environments.

In this study, we propose a data-driven optimization strategy for MAs to build the optimal MA configuration in-vehicle environments. The proposed optimization strategy for the MAs incorporates actual signal collection experiments conducted in vehicles rather than relying solely on mathematical or computational models. Initially, we developed a novel in-vehicle noise model based on sound field modeling, capable of precisely replicating the interior noise of a vehicle under usual driving conditions. In addition to vehicle interior noise, this study also accounts for interference from loudspeakers and the co-driver. Subsequently, utilizing this model, we propose an MA configuration update algorithm to design the optimal MA configuration in a vehicle environment. Finally, the designed array configuration was constructed using an MEMS MA acquisition system, and the array performance was evaluated in real driving environments. The flowchart for the optimization strategy for MAs in-vehicle environments is shown in Fig. 1.

The contribution of this article can be primarily summarized into the following three aspects:

1) Collected real vehicle noise signals to model the sound field in in-vehicle environments accurately.
2) A data-driven optimization strategy for MAs in-vehicle environments was developed, focusing on performance in real-vehicle conditions.
3) Designed and constructed arbitrary MA configurations suitable for in-vehicle environments, considering spatial limitations.

The rest of this article is structured as follows. Section II introduces the signal model and the in-vehicle noise model employed in this study. Section III introduces in-vehicle sound field modeling. Section IV introduces the design of MA configuration in vehicular environments. Section V presents a comparative analysis of the simulation and

experimental results for the designed MA against conventional arrays. The article concludes with a summary in Section VI.

II. In-Vehicle Noise Model

In this section, we first introduce the signal model and beamforming algorithm, which are the foundations for the in-vehicle noise model. Subsequently, we will provide a detailed description of the in-vehicle noise model.

A. Signal Model and Beamforming Method

Assume that $\mathbf{C} = [\mathbf{c}_l, \ldots, \mathbf{c}_m, \ldots, \mathbf{c}_M]$ as an MA composed of M elements, $\mathbf{c}_m = (x_m, y_m, z_m)$ represents the mth microphone concerning the origin of coordinates.

1) Echoic Mixing Model: The criterion for classifying a sound source as near-field is shown as follows:

$$|r| < 2L^2 \setminus \lambda. \tag{1}$$

In the given context, $|r|$ represents the distance between the sound source and the MA, L denotes the effective length of the array, and λ signifies the wavelength. Given expression (I) and considering the dimensions of the vehicle interior space, the speech source is classified as a near-field source in this study.

Consider a scenario involving a set of M microphones designed to receive speech signals from sources $N_n(t)$, $n \in \{1, \ldots, L\}$, to generate M distinct mixtures $S_m(t)$, $m \in \{1, \ldots, M\}$ and $N_m(t), m \in \{1, \ldots, M\}$ at discrete time points represented by t.

Generally, the expression for the additive mixing model is described by

$$S_m(t) = \sum_{n=1}^{N} S_n(t) * h_{mn}(t), \quad m = 1, \ldots, M. \tag{2}$$

In this equation, the symbol $*$ represents linear convolution. However, owing to reflections in the vehicle, multiple delayed and attenuated signals originating from the same source signal are captured by the microphone. Consequently, an echoic mixing model was posited by

$$S_m(t) = \sum_{n=1}^{N} \sum_{p=1}^{N_P} a_{mnp}(t) \cdot S_n\big(t - \delta_{mnp}\big), \quad m = 1, \ldots, M \tag{3}$$

$$N_m(t) = \sum_{l=1}^{L} \sum_{p=1}^{N_P} a_{mnp}(t) \cdot N_n\big(t - \delta_{mnp}\big), \quad m = 1, \ldots, M. \tag{4}$$

In the given context, the symbol $\cdot$ denotes element-wise multiplication, N_P presents the number of distinct paths that signals traverse from the sources to the microphones, and a_{mnp} and δ_{mnp} denote the attenuation and delays introduced in the Pth path, respectively.

The signal that the mth element of the array $\mathbf{C}$ received can be described by

$$X_m(t) = S_m(t) + N_m(t). \tag{5}$$

2) Wideband Beamforming: The frequency spectrum of speech signals predominantly lies in the 300–3400-Hz range, indicating that the signal is wideband. The steering vectors of a wideband signal exhibit correlation with frequency; therefore, a frequency-domain model is employed in wideband signal processing. This article utilizes the subband minimum variance distortionless response (MVDR) beamforming method [38] to enhance the input speech signals.

The directivity pattern or frequency response of an array $\mathbf{C}$ composed of M elements is given by

$$D(k, \mathbf{s}, \mathbf{C}) = \sum_{m=1}^{M} W_m(k) A_m(k, \mathbf{s}, \mathbf{C}). \tag{6}$$

The vector $\mathbf{s}$ represents the position of the source relative to the coordinate origin, $\mathbf{s} = (r, \theta, \varphi)$, where r denotes the distance between the source and the coordinate origin, θ is the elevation angle, and φ is the azimuth angle. The symbol k signifies frequency. The complex weights applied to the mth element of the array are represented by $W_m(k)$, $A_m(k, \mathbf{s}, \mathbf{C})$ denotes the frequency response of the mth element of the array about the described source

$$A_m(k, \mathbf{s}, \mathbf{C}) = \frac{r}{d(\mathbf{s}, \mathbf{C})} \exp\{-j\beta[d(\mathbf{s}, \mathbf{C}) - r]\} \tag{7}$$

where $\beta = 2\pi k \setminus c$ and c denotes the sound speed. $d(\mathbf{s}, \mathbf{C})$ is the distance between the source and the mth element of the array, according to

$$d(\mathbf{s}, \mathbf{C}) = \big[(r\cos\varphi\sin\theta - x_m)^2 + (r\cos\varphi\cos\theta - y_m)^2 + (r\sin\varphi - z_m)^2\big]^{\frac{1}{2}} \tag{8}$$

where $\mathbf{w}_k = [W_1(k), \ldots, W_M(k)]^T$ is the array weight vector, $(\cdot)^H$ denotes the Hermitian transpose, and $\mathbf{a}_{k\mathbf{sC}}$ is the steering vector that contains the M microphone responses of the array $\mathbf{C}$, $\mathbf{a}_{k\mathbf{sC}} = [A_1(k, \mathbf{s}, \mathbf{C}), \ldots, A_M(k, \mathbf{s}, \mathbf{C})]^T$.

Array gain, the improvement in the signal-to-noise ratio (SNR) between a reference sensor and the array output, is expressed as $G = G_d \setminus G_n$, where G_d represents the gain toward the desired signal and G_n corresponds to the average gain toward all noise sources, contingent upon the characteristics of the noise field. Assuming the target source is positioned at a specific location, and there exists a finite number of noise sources, each with equivalent power, interfering with the target source, the array gain can be quantified according to

$$G(k, \mathbf{s}_0, \mathbf{C}) = \frac{\big|D(k, \mathbf{s}_0, \mathbf{C})^2\big|}{\frac{1}{N}\sum_{n=1}^{N=1}\big|D(k, \mathbf{s}_n, \mathbf{C})^2\big|} = \frac{\big|\mathbf{w}_k^H \cdot \mathbf{a}_{k\mathbf{s}_0\mathbf{C}}\big|^2}{\mathbf{w}_k^H \cdot \mathbf{H}_\mathbf{c} \cdot \mathbf{w}_k} \tag{9}$$

where $\mathbf{s}_n$ represents the position of the nth noise source and $\mathbf{H}_\mathbf{c}$ is the noise cross-spectral matrix.

The fundamental principle of the MVDR approach is the optimization of array gain, according to

$$\min_a{}^b \big\{\mathbf{w}_k^H \cdot \mathbf{H}_\mathbf{c} \cdot \mathbf{w}_k\big\} \quad \text{s.t. } \mathbf{w}_k^H \cdot \mathbf{a}_{k\mathbf{s}_0\mathbf{C}} = 1. \tag{10}$$

The optimization problem is solved using Lagrange multipliers, resulting in

$$\mathbf{w}_k = \frac{\mathbf{H}_\mathbf{c}^{-1}\mathbf{a}_{k\mathbf{s}_0\mathbf{C}}}{\mathbf{a}_{k\mathbf{s}_0\mathbf{C}}^H \mathbf{H}_\mathbf{c}^{-1}\mathbf{a}_{k\mathbf{s}_0\mathbf{C}}}. \tag{11}$$

Finally, the array output $Y(k, \mathbf{C})$ is expressed as the combination of the weighted input channels, according to

$$Y(k, \mathbf{C}) = \frac{1}{M} \sum_{m=1}^{M} W_m(k) X_m(k, \mathbf{C}) \tag{12}$$

where $X_m(k, \mathbf{C})$ is the discrete Fourier transform (DFT) of the input signal received by the mth microphone of the array $\mathbf{C}$.

B. Model Development

After determining the necessary signal model and beamforming technique, we will develop the in-vehicle noise model, which is fundamental to the MA optimization strategy for vehicles outlined in this study.

In the present study, we employ a room impulse response generator (RIRG) to simulate the microphone echo responses in the vehicular environment. We have utilized the simplified image method proposed by Allen and Berkley [39] to compute the acoustic impulse response between any two points in the vehicle's interior. The reflection coefficient in the simulation environment was established at -1, signifying that the sound signal would experience infinite reflections and attenuation in the simulated space.

The in-vehicle noise model, presented in Step 1 of Fig. 1, primarily consists of the vehicle interior space, microphone search area, noise points, sound sources of the loudspeaker, and the sound sources of the driver and co-driver. Drawing upon the actual internal spatial dimensions in the test vehicle, the proposed in-vehicle noise model simplifies the vehicle's internal space into a cuboid with dimensions of 2300 mm in length, 2100 mm in width, and 1200 mm in height.

We assume that the internal car noise is generated by a finite number of points located predominantly on the surface areas of the vehicles, such as tires, windows, and engines. Thus, noise collection points are strategically positioned in these regions. The locational relationship between the noise points and the inner space of the in-vehicle noise model is shown in Fig. 2. In this depiction, black spheres represent noise points on the front windshield, gray for the rear windshield, blue for side windows, violet for near the wheels, and brown for the engine section. The distribution of noise points is as follows: 28 on the front windshield, 15 on the rear, 12 on each side window, 1 at each wheel area, and 5 at the engine, totaling 100 points.

Moreover, the in-vehicle noise model also accounts for disturbances from loudspeakers and the co-driver (passenger in the front seat). The source direction and energy of the co-driver speech, being proximate to that of the driver, constitute a significant disturbance that should be considered. Fig. 2 also marks the locations of both the target and interfering sound source points. The heights for the driver and passenger speech sources are calibrated for an individual of 1.75 m. The speech source is derived from the TIMIT database [40]. The green rectangle indicates the microphone search zone on the central control panel, measuring 32 × 24 cm with a grid resolution of 4 mm, offering 4800 potential microphone positions. We have precalculated the impulse responses between each sound source and all potential microphone positions to avoid repetitive calculations.

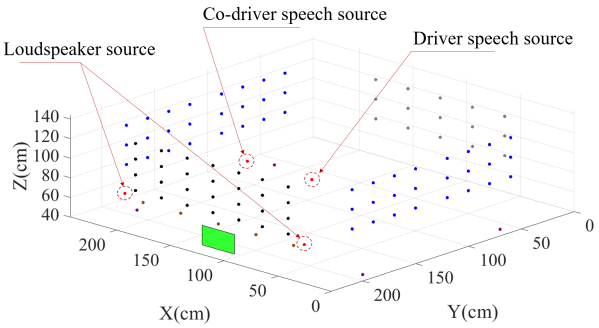


Fig. 2. Arrangement of noise points in the in-vehicle noise model.

III. In-Vehicle Sound Field Modeling

This section is dedicated to developing a precise model of the acoustic field inside a vehicle. Utilizing an MEMS MA acquisition system, noise data from within the vehicle are gathered at specific measurement points predetermined by the in-vehicle noise model under actual driving conditions. The noise data collected are then fed as input signals into corresponding noise points in the in-vehicle noise model to simulate the interior sound field.

A. Noise Data Collection Experiment

MEMS technology, known for its compactness, energy efficiency, affordability, and integration ease, has been widely adopted. Its MAs offer flexibility and configurability, fitting well in confined spaces such as car interiors. The collection of interior vehicle noise was performed using an MEMS MA acquisition system. The MEMS MA acquisition system employs an MEMS microphone, microphone baseplate, driving board, and an open-source single-board computer, as depicted in Step 2 of Fig. 1. The single-board computer connects to the driving board via a 26-pin interface, while the two driving boards are linked through a 28-pin flexible printed circuit (FPC). The MEMS microphone is connected to the driving board via a 5-pin FPC. This system supports the simultaneous acquisition of up to 16 channels. During operation, signals from the MEMS microphone capture are transmitted to the single-board computer through the driving board. The single-board computer is then connected to another computer via an Ethernet cable for system control and data storage. The MEMS microphone used in this study is a digital model with dimensions of 5 × 4 × 0.98 mm, a frequency response range of 20 Hz–20 kHz, and a sensitivity of -38 ±3 dB. The acquisition of audio signals is facilitated through an audio interface card, which boasts a sampling rate of 48 kHz and a resolution of 24 bits.

In the experiment, a commonly used family sedan was selected as the test vehicle to ensure the relevance and applicability of the findings to a wide range of consumer vehicles. MEMS microphones were employed to collect acoustic noise data in the vehicle. The precise placement of the microphones was informed by the noise point configuration presented in the in-vehicle noise model illustrated in Fig. 2. By this diagram, an MEMS microphone was installed at each predetermined noise point in the vehicle to ensure the accurate acquisition of sound information from the specified locations. Due to space constraints, only a selection of images depicting the

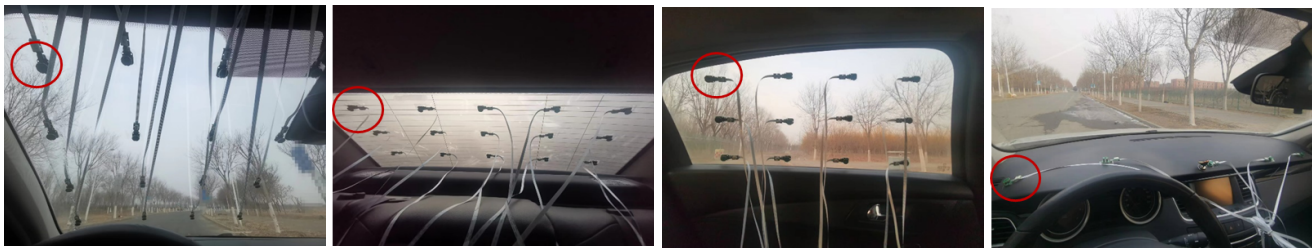


Fig. 3. Arrangement of noise points in the vehicle. (a) Front windshield, (b) rear windshield, (c) side window, and (d) engine area.

arrangement of MEMS microphones in the vehicle measurement areas is presented in this article, as shown in Fig. 3. The red circles in the figure indicate the MEMS microphones in the measurement area. The signal collection experiment is conducted on a closed road test track, ensuring the vehicle remains operational during data acquisition while maintaining a 40 km/h speed.

B. Verification of Noise Field Consistency

By sequentially inputting the collected noise data into the in-vehicle noise model according to the arrangement of noise measurement points, we can accomplish the modeling of the vehicle's internal noise field. A measure to characterize the noise environment is the complex coherence of a noise field, represented by two signals x_i and x_j, at discrete time index t. In the frequency domain, this coherence is defined as

$$\Gamma_{ij}(k) = \frac{\phi_{x_i x_j}(k)}{\sqrt{\phi_{x_i x_i}(k)\phi_{x_j x_j}(k)}}. \tag{13}$$

Herein, k denotes the frequency and $\Gamma_{ij}(k)$ represents the coherence function between x_i and x_j at the frequency k. The term $\phi_{x_i x_j}(k)$ refers to the cross-power spectral density, indicating the mutual variations of x_i and x_j at frequency k, $\phi_{x_i x_i}(k)$ and $\phi_{x_j x_j}(k)$, on the other hand, denote the auto-power spectral densities of x_i and x_j, respectively.

To validate the accuracy of the noise field generated by the in-vehicle noise model, a uniform linear MA composed of eight microphones with a spacing of 35 mm was employed to capture the interior noise of a car traveling at a constant speed of 40 km/h. Fig. 4 illustrates the average spatial coherence of the noise collected inside the vehicle, in conjunction with the spatial coherence results of the noise field produced by the model we proposed, evaluated at a microphone spacing of 35 mm. To compare the performance of the proposed model, Fig. 4 also displays the spatial coherence results of an ideal diffuse noise field and the noise field generated by the model developed by Ayllón et al. [37], each evaluated with an identical microphone spacing of 35 mm.

The result reveals that, for frequencies below 1000 Hz, the average spatial coherence of the noise measured in the vehicle diminishes progressively; however, for frequencies exceeding 1000 Hz, the spatial coherence of the interior noise begins to ascend. The ideal diffuse noise field exhibits a high correlation with the actual vehicular noise field at lower frequencies, yet its correlation becomes significantly weaker for frequencies above 1000 Hz. The noise field generated by David's model, although accounting for the correlation across

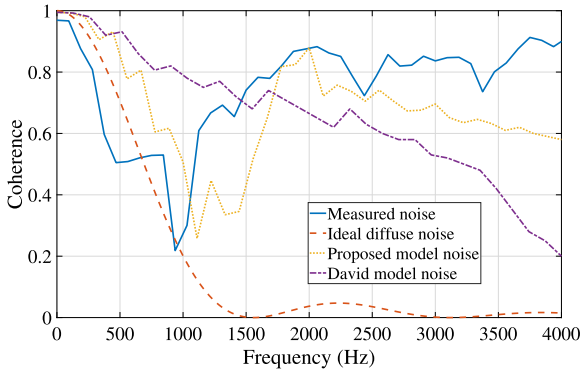


Fig. 4. Spatial coherence results.

frequencies below and above 1000 Hz, still demonstrates considerable deviations on the whole and fails to represent accurately the actual noise environment in the vehicle. In contrast, the noise field generated by the model introduced in this article highly correlates with the actual vehicular noise field in the low-frequency range. It maintains a commendable correlation for frequencies above 1000 Hz. Consequently, the noise field produced by the model proposed in this study more closely approximates the actual noise environment in the vehicle compared to other methods.

IV. In-Vehicle Configuration Design of MA

We have designed MA configurations for in-vehicle environments to leverage the sound field modeling-based in-vehicle noise model. This study has 4800 installation positions for microphones, yielding many possible array configurations. Given the vast array of possibilities, we have implemented heuristic optimization algorithms for an efficient solution. To address the design challenges of MAs in vehicular environments, we have developed an MA configuration update algorithm based on an innovative optimization algorithm proposed by Mohammed and Rashid [41].

A. Initialization Strategy

Initializing solutions is vital for the performance of heuristic optimizations. Hence, we developed a novel initialization strategy that uses conventional MA setups like uniform rectangular array (URA), uniform circular array (UCA), and uniform L-shaped array (ULsA), with added random variations to build and evaluate a set of initial candidate MA solutions for the in-vehicle acoustic field environment. The original population combines solutions based on prior knowledge and random solutions that may offer new potential solutions, ensuring the

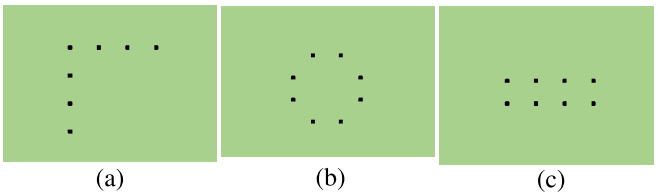


Fig. 5. Configuration of conventional MAs. (a) ULsA with seven microphones, (b) UCA with eight microphones, and (c) URA with eight microphones.

diversity and quality of the solutions. Each MA configuration, varying from four to eight elements, is a possible solution. We have set up a population of 50 candidate solutions to balance search performance and convergence speed. Due to space limitations, this article only displays a selected conventional setup in the measurement area, as shown in Fig. 5. Based on the research by Grenier [42], we standardized the spacing between elements in the array at 35 mm to accommodate the frequency range of speech signals.

Each candidate in the population was assessed to ensure compliance with three criteria: 1) MA elements must be situated on predefined grid coordinates; 2) the minimum interspacing between microphones must be at least 35 mm; and 3) candidate solution positions must fall in the designated search area; if a microphone's position lies outside this area, it is repositioned to the nearest boundary of the search zone.

The choice of an appropriate fitness function is crucial in our algorithm application. We have selected short-time objective intelligibility (STOI) [43], primarily used to evaluate the quality of speech signals in noisy environments. The STOI metric is quantified on a scale from 0 to 1, with higher values indicating better speech intelligibility, making it ideal for assessing the quality of speech signals captured by the in-vehicle MA.

The initialization phase process is illustrated in Fig. 6. First, the number of microphones M is determined, and then an initial set of 50 candidate solutions, including conventional and randomly generated arrays, each containing M microphones, is automatically generated. Each candidate is represented as a 2-D array containing the 2-D positional data of the MA, with a data volume of $2M$. All candidate solutions undergo a dimensionality reduction to reduce the computational load, transforming them into 1-D arrays of length $2M$. Then, using the in-vehicle noise model, the STOI for each MA group is calculated, evaluating the performance of each candidate solution. The optimal microphone configuration $\mathbf{C}_{\text{opt}}$ is determined by comparing fitness values, and then the algorithm proceeds to the optimization stage.

B. Optimization Strategy

The optimization process is divided into exploration and exploitation to achieve optimal search performance and avoid convergence to suboptimal solutions. Each phase has a 50 percent probability, balancing discovering new solutions and refining known ones. Fig. 7 illustrates the tailored process of the exploration and exploitation phases.

During the exploration phase, the optimization target randomly navigates in the solution space to uncover potential superior solutions. At this stage, the array configuration $\mathbf{C}$ is adjusted based on the optimal fitness value and the optimal array configuration $\mathbf{C}_{\text{opt}}$ from the previous iteration. In the exploitation phase, the optimization target engages in directional movement in the solution space to pinpoint the optimal solution, signifying a shift to a more meticulous search. In this phase, the optimization target demonstrates two movement patterns: if p exceeds 0.18, the target traverses a larger distance in the solution space, leading to more significant modifications in the array configuration $\mathbf{C}$. Conversely, when p falls below 0.18, the target covers a shorter distance in the solution space, resulting in more subtle changes to the array configuration $\mathbf{C}$. In our experimental research, we observed that updating the position of array $\mathbf{C}$ with a constant step size may result in significant fluctuations or minor variations in the fitness value, depending on the configuration of the array. When the changes in fitness values are minimal, the algorithm may waste considerable time on arrays with poor fitness, thereby diminishing optimization efficiency. To address this issue, we introduced a novel optimization target update strategy in the optimization process to enhance the algorithm's search efficiency. We established two new parameters to control the magnitude of the optimization target update: the scaling factor α, which governs the magnitude of position updates for array $\mathbf{C}$, and the rate of change limit ξ, which determines whether the rate of change in fitness values meets the algorithm's requirements. By appropriately setting α and ξ, it is possible to prevent the algorithm from expending excessive time on arrays with poor fitness while ensuring that the update magnitude is sufficient to maintain search precision. The pseudocode for the proposed MA configuration update algorithm is presented in Algorithm 1.

Algorithm 1 MA Configuration Update Algorithm

Input: Scaling factor: α; Change rate limit: ξ
Output: Array configuration: $\mathbf{C}$

for $i \leftarrow 0$ **to** $size - 1$ **do**
  $(r, p) \leftarrow \text{random}(0, 1), \text{random}(0, 1)$;
  $\mathbf{C}[i] \leftarrow MACUA(\mathbf{C}, r, p, i, \alpha, \xi, \Theta)$;
end
Function *MACUA*($\mathbf{C}$, r, p, i, α, ξ, Θ)**:**
  $Dist \leftarrow Move\_Distance(r, p, \mathbf{C}[i], \Theta)$ $\alpha_{update} \leftarrow \alpha$, $\xi_{change} \leftarrow 0$;
  while $\xi_{change} < \xi$ **do**
    $\mathbf{C}[i] \leftarrow \Theta + Dist \times (1 + \alpha_{update})$;
    $fit \leftarrow \text{Vehicle\_model}(\mathbf{C}[i])$;
    $\xi_{change} \leftarrow |fit - Bestfit|/Bestfit$;
    $\alpha_{update} \leftarrow \alpha_{update} + \alpha$;
  end
  return $\mathbf{C}[i]$
end
Remarks: *Move_Distance*() is the displacement update method used in Mohammed's research [41], *Vehicle_model*() is the performance evaluation method for the MA proposed in this study, and Θ represents the optimal array configuration from the last iteration.

After the exploitation or exploration phase, the algorithm generates a new set of array configurations $\mathbf{C}$. Subsequently, this new set of candidate solutions is dimensionally escalated into a 2-D array encapsulating microphone configuration information. It autonomously adjusts its values according to

TABLE I
STOI AND OUTPUT SNR OBTAINED BY THE DESIGNED ARRAYS, ULA, AND UCA

M	ULA				UCA				Designed array			
	STOI		SNR		STOI		SNR		STOI		SNR	
	A	B	A	B	A	B	A	B	A	B	A	B
4	0.536	0.542	-7.35	-7.28	0.611	0.619	-6.12	-6.31	0.695	0.685	-2.88	-3.28
5	0.548	0.551	-6.87	-6.69	0.628	0.622	-5.33	-5.22	0.703	0.693	-2.37	-2.83
6	0.561	0.564	-6.22	-6.37	0.638	0.633	-4.86	-4.87	0.718	0.705	-2.29	-2.57
7	0.574	0.571	-5.84	-6.12	0.642	0.644	-4.28	-4.36	0.733	0.718	-2.14	-2.53
8	0.583	0.582	-5.35	-5.43	0.654	0.655	-4.09	-4.35	0.745	0.728	-1.91	-2.21

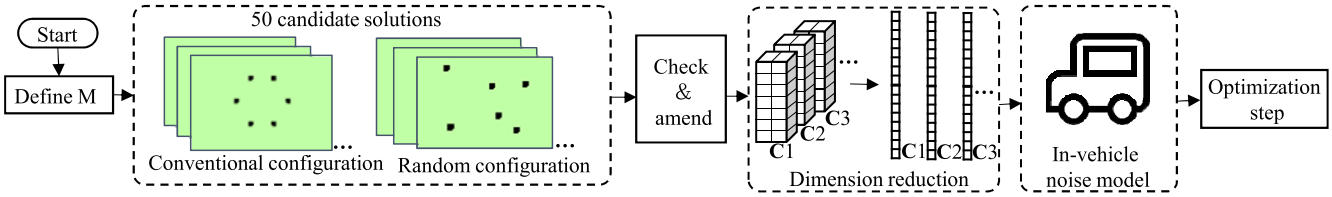


Fig. 6. Initialization process.

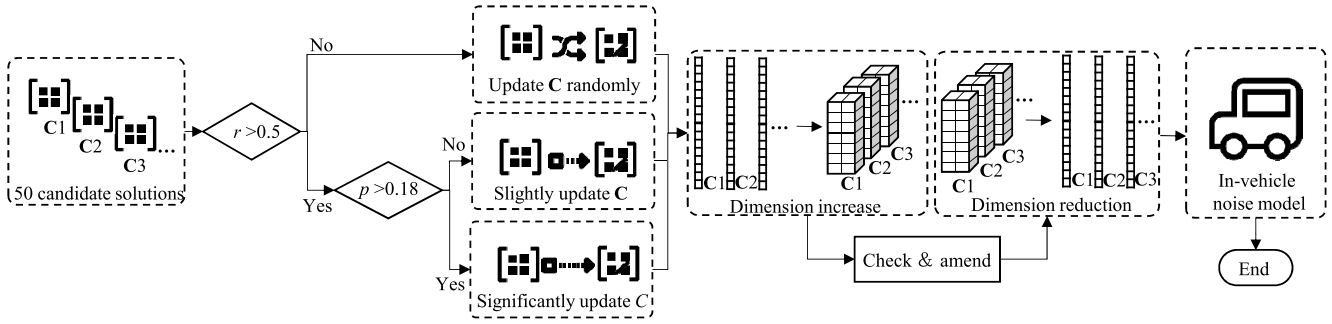


Fig. 7. Optimization process.

predefined constraints. Then, the candidate solutions undergo dimensionality reduction, and their array configuration information is sequentially fed to the in-vehicle noise model to calculate the fitness value, aiming to find the optimal fitness value and array configuration $\mathbf{C}_{\text{opt}}$ for this iteration. In this study, the number of iterations is set at 100. Finally, using optimization, the optimal placement of the M components in the MA $\mathbf{C}$ in a vehicular environment can be determined. The subsequent section will conduct a comparative performance analysis between the designed array and conventional MAs.

V. RESULTS AND ANALYSIS

In this section, we compare the speech enhancement performance of the designed MA to that of the conventional MA using STOI and the output SNR. For the SNR, the signal refers to the speech signal emitted by the target speech source (such as the driver or co-driver), while the noise pertains to the background noise inside the vehicle. The analysis is divided into two segments: the analysis of the simulation results and the examination of the experimental results.

A. Simulation Results and Analysis

In this section, based on the in-vehicle noise model, we conducted a comparative analysis of the speech enhancement performance of designed MAs, ULA, and UCA in a vehicular environment. The study encompassed configurations with four, five, six, seven, and eight microphones. The speech signals from the driver and the co-driver were emitted with equal acoustic power levels, while the power level of the loudspeaker source was set at 3 dB lower. The input SNR was maintained at -15 dB and the sampling frequency was fixed at 8000 Hz.

While the co-driver speech was initially classified as an interfering noise source during optimization, it concurrently serves as an essential target sound source in the actual driving conditions. Consequently, this section also analyzed the MA's performance, enhancing the co-driver's speech. The performance outcomes of the designed arrays, ULA, and UCA with varying quantities of microphones are detailed in Table I. Here, A represents the speech of the driver and B represents the speech of the co-driver.

As Table I shows, the designed arrays, ULA, and UCA with varying microphone quantities have different performance outcomes. STOI and output SNR positively correlate with the number of microphones M. Compared to ULA and UCA, the designed arrays yield the best STOI and output SNR, indicating superior performance. Taking the case with eight microphones as an example, compared to ULA and UCA, the designed MA increased the STOI scores by 27.9% and 13.9%, respectively, and also improved the output SNR levels by 64.3% and 53.3%, respectively. The designed array and UCA configurations outperform ULA due to their planar structure, allowing for using both azimuth and elevation angles in steering vector estimation. In contrast, the ULA can only employ

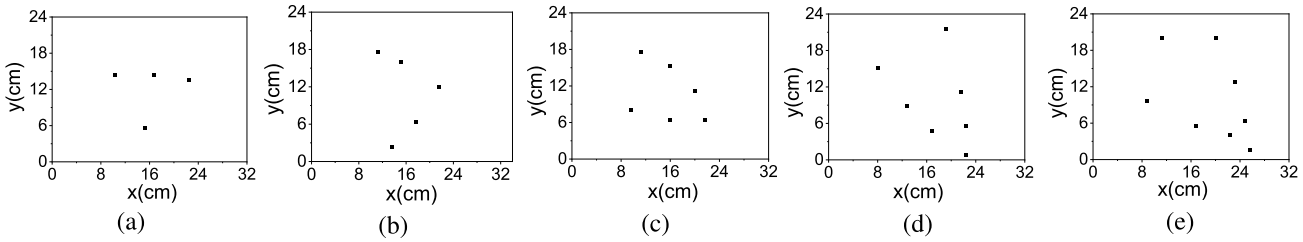


Fig. 8. Designed MAs. (a) Four-element, (b) five-element, (c) six-element, (d) seven-element, and (e) eight-element.

the azimuth angle for this purpose, resulting in less accurate steering vector estimations and reduced speech enhancement quality. Furthermore, the designed arrays are the product of iterative refinements tailored to vehicular environments, which accounts for their exemplary speech enhancement attributes. Notably, the speech enhancement performance for the driver is marginally superior to that of the co-driver when utilizing the designed arrays. This discrepancy arises because optimizing the MAs prioritizes the speech enhancement of the driver's speech.

The designed MA configuration derived from this research is depicted in Fig. 8. It can be discerned that the designed array arrangement adheres to specific transformation rules related to the number of microphones, with each increment in the number of microphones resulting in an enhancement of the array performance. Starting with a fundamental configuration of four microphones, we have already observed the initial formation of configuration characteristics, demonstrating that even the simplest of arrays can be improved through our algorithm. When the number of microphones was increased to five, we designed a structure similar to a ULsA, which underscored the pivotal role of the placement orientation of the MA in augmenting performance. Further addition of microphones to six revealed an acute-angled ULsA configuration, which bolstered the capability of directional capture of sound waves. The configuration with seven microphones was further optimized by merging two Minimum Redundancy Linear Arrays [44], which maximized the array's pick-up range and significantly improved work efficiency in multisource environments. Ultimately, the configuration with eight microphones, achieved by adding an extra microphone between the endpoints of the ULsA, formed a closed geometric structure. This configuration further optimized the uniformity of the sound field coverage. It validated the performance improvement brought about by the increase in microphones, maintaining commendable performance even in complex acoustic environments.

B. In-Vehicle Performance Testing

To validate the simulation results and assess the performance of the MA in an actual vehicle environment, we conducted signal acquisition experiments under actual driving conditions. We evaluated the impact of changes in the input SNR and source position on the MA performance to verify its applicability under complex acoustic conditions of driving environments. Notably, under the same operating conditions, the error between the actual vehicle test results and the simulation results was within 5%, which fully validates the accuracy of the model. Subsequent sections will delve into these essential points in detail.

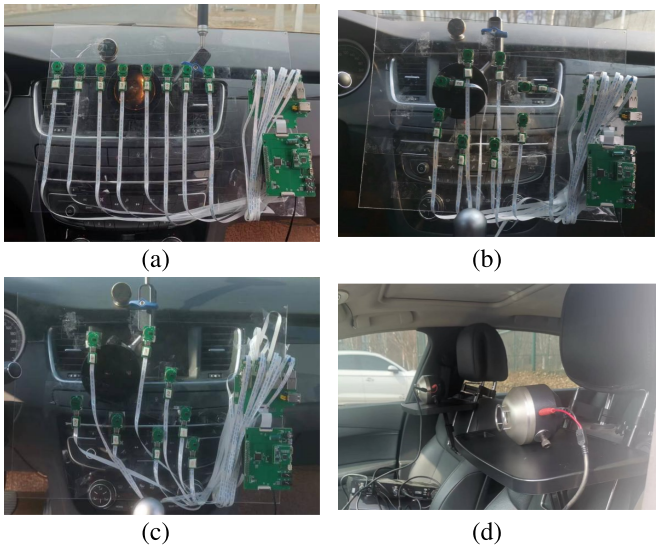


Fig. 9. Arrangement of the experiment equipment in a vehicle. (a) Eight-element uniform linear array, (b) eight-element uniform circular array, (c) eight-element designed MA, and (d) artificial mouths.

1) In-Vehicle Data Collection Experiments: The vehicle data collection experiment comprises two main components: interior noise collection and speech acquisition. For noise collection, an 8-element designed array, an 8-element ULA, and an 8-element UCA are utilized to capture interior vehicle noise at a speed of 40 km/h, as illustrated in Fig. 9(a)–(c), respectively. In terms of speech acquisition, this study proposes a novel method employing an artificial mouth to mimic natural speech, avoiding the signal instability and safety risks inherent in conventional in-vehicle speech acquisition techniques that capture human voices while driving. The artificial mouth, designed to emulate human vocal production, offers a secure, manageable, and efficient way to produce steady, reproducible signals crucial for audio device testing. The setup comprises the artificial mouth, a power amplifier, and a signal source, detailed in Step 4 of Fig. 1. The artificial mouths are strategically positioned and secured at the locations corresponding to the driver's and co-driver's mouths using custom fixtures to emulate their speaking activities, as shown in Fig. 9(d). The TIMIT corpus supplies the target and interference sounds. The experiment is performed with the engine turned off. The power of the signals from the driver and co-driver is equal, while the power of the loudspeaker is 3 dB lower relative to the driver and co-driver signals. After collection, the clean speech is synthesized with the noise to simulate speech in a vehicular environment.

2) Impact of the Input SNR: Vehicular environments exhibit variability in background noise and speech volume due to speed, road conditions, and acoustics, leading to an uncertain

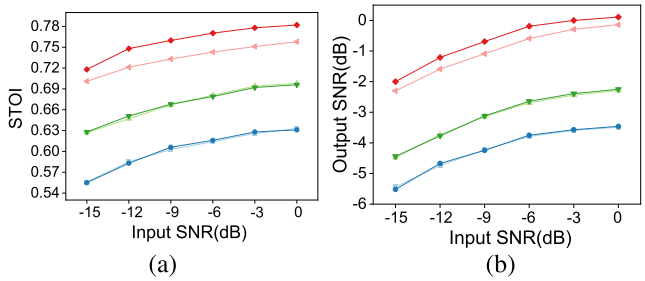


Fig. 10. Results of the MAs under different input SNRs for $M = 8$. (a) Output results of STOI. (b) Output results of the output SNR. (Blue: ULA driver speech. Light blue: ULA co-driver speech. Green: UCA driver speech. Light green: UCA co-driver speech. Red: Designed array driver speech. Light red: Designed array co-driver speech.)

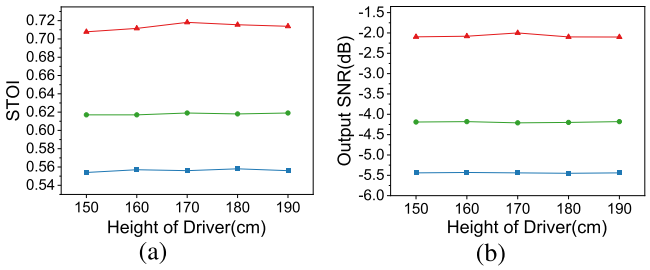


Fig. 11. Results of the MAs under different sound source positions for $M = 8$. (a) Output results of STOI. (b) Output results of the output SNR. (Blue: ULA driver speech. Green: UCA driver speech. Red: Designed array driver speech.)

input SNR for MAs. Assessing the effect of input SNR variations on array performance is crucial. Speech signals were combined with noise at various SNRs (-15 to 0 dB) to create noisy signals. The performance outcomes of the designed array, ULA, and UCA at different input SNRs are depicted in Fig. 10. The data reveal that the designed array outperforms ULA and UCA in STOI and output SNR for both driver and co-driver sources. On the other hand, While STOI and output SNR generally rise with the input SNR, the increase slows and nears a limit. Analysis of the enhanced signal processed by the MAs reveals that arrays reduce uncorrelated noise but slightly mitigate interfering speech. Although background noise decreases, the output still contains some interference, suggesting arrays partially counter interference. Nonetheless, the designed array delivers substantial speech enhancement, with STOI over 0.7 and output SNR above -2.3 dB even at low input SNRs.

3) Impact of Source Position: The diversity of drivers further compounds the complex acoustic environment in vehicles. Driver height variations lead to vehicle source position changes, affecting MA performance. Therefore, it is crucial to investigate source position variability's impact on MAs' efficacy in vehicular environments. To study this, we simulated source positions for drivers of different heights by adjusting seat-to-wheel distances and artificial mouth heights, collecting speech signals for drivers ranging from 150 to 190 cm. We compared the performance of an 8-element designed array, an 8-element ULA, and an 8-element UCA at an input SNR of -15 dB, as shown in Fig. 11. Results show that the designed array consistently outperforms others in STOI and output SNR across various source positions, with stable performance indicating strong adaptability to source position changes, making it suitable for complex vehicular environments.

VI. Conclusion and Outlook

The article presents a data-driven optimization approach for MA design to address the challenges in vehicle interiors. We developed an in-vehicle noise model to accurately simulate the driving sound field, providing a foundation for designing array configurations. Subsequently, we combined sound field reconstruction with a tailored optimization algorithm to achieve the optimal array configuration for vehicular environments. Finally, we implemented and tested the designed design using an MEMS system and an artificial mouth system, demonstrating that the configuration meets spatial and acoustic demands while maintaining high performance in complex vehicle acoustics.

Future research should address the "cocktail party problem," a significant challenge in vehicular environments where multiple sound sources interfere with speech intelligibility. Our forthcoming work will leverage the in-vehicle noise model and experimental validation methods presented in this study to tackle the challenges of speech separation in these contexts. In addition, the sound field modeling method employed in this study applies to various enclosed acoustic environments, such as train carriages, aircraft cabins, and specialized vehicle interiors, thereby contributing to research across multiple fields.

References

[1] W. Li, K. Takeda, and F. Itakura, "Adaptive log-spectral regression for in-car speech recognition using multiple distributed microphones," *IEEE Signal Process. Lett.*, vol. 12, no. 4, pp. 340–343, Apr. 2005.

[2] Y.-H. Lai and W.-Z. Zheng, "Multi-objective learning based speech enhancement method to increase speech quality and intelligibility for hearing aid device users," *Biomed. Signal Process. Control*, vol. 48, pp. 35–45, Feb. 2019.

[3] D.-H. Yang and J.-H. Chang, "Attention-based latent features for jointly trained end-to-end automatic speech recognition with modified speech enhancement," *J. King Saud Univ. Comput. Inf. Sci.*, vol. 35, no. 3, pp. 202–210, Mar. 2023.

[4] P. Počta and S. Isabelle, "Quality aspects of music used as a background noise in speech communication over mobile network," *Appl. Acoust.*, vol. 134, pp. 125–130, May 2018.

[5] J. Fan, J. Yang, X. Zhang, and Y. Yao, "Real-time single-channel speech enhancement based on causal attention mechanism," *Appl. Acoust.*, vol. 201, Dec. 2022, Art. no. 109084.

[6] H. Ping and W. Yafeng, "Single-channel speech enhancement using improved progressive deep neural network and masking-based harmonic regeneration," *Speech Commun.*, vol. 145, pp. 36–46, Nov. 2022.

[7] J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing*, vol. 1. Berlin, Germany: Springer, 2008.

[8] M. Arrays, "Signal processing techniques and applications," in *Microphone Arrays: Signal Processing Techniques and Applications*. Berlin, Germany: Springer-Verlag, 2001.

[9] S. Chakrabarty and E. A. P. Habets, "On the numerical instability of an LCMV beamformer for a uniform linear array," *IEEE Signal Process. Lett.*, vol. 23, no. 2, pp. 272–276, Feb. 2016.

[10] X. Chen, J. Benesty, G. Huang, and J. Chen, "On the robustness of the superdirective beamformer," *IEEE/ACM Trans. Audio, Speech, Lang., Process.*, vol. 29, pp. 838–849, 2021.

[11] W. Jin, J. Wei, and X. Zhong, "Multi-channel speech enhancement in driving environment," in *Proc. ICSPCS*, 2017, pp. 1–5.

[12] S. M. Kim and H. K. Kim, "Probabilistic spectral gain modification applied to beamformer-based noise reduction in a car environment," *IEEE Trans. Consum. Electron.*, vol. 57, no. 2, pp. 866–872, May 2011.

[13] J. Cho and A. Krishnamurthy, "Speech enhancement using microphone array in moving vehicle environment," in *Proc. Intell. Vehicles Symp.*, 2003, pp. 366–371.

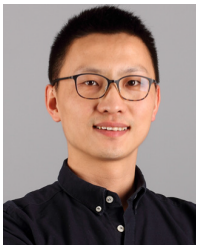
[14] J. H. L. Hansen and X. Zhang, "Analysis of CFA-BF: Novel combined fixed/adaptive beamforming for robust speech recognition in real car environments," *Speech Commun.*, vol. 52, no. 2, pp. 134–149, Feb. 2010.

[15] M. Souden, J. Benesty, and S. Affes, "A study of the LCMV and MVDR noise reduction filters," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4925–4935, Sep. 2010.
[16] M. Karimi and L. Maxit, "Acoustic source localisation using vibroacoustic beamforming," *Mech. Syst. Signal Process.*, vol. 199, Sep. 2023, Art. no. 110454.
[17] J. Zhang, G. Squicciarini, D. J. Thompson, W. Sun, and X. Zhang, "A hybrid time and frequency domain beamforming method for application to source localisation on high-speed trains," *Mech. Syst. Signal Process.*, vol. 200, Oct. 2023, Art. no. 110494.
[18] C.-C. Shen, Y.-A. Chen, and H.-Y. Ku, "Improved source localization in passive acoustic mapping using delay-multiply-and-sum beamforming with virtually augmented aperture," *Ultrasonics*, vol. 135, Dec. 2023, Art. no. 107125.
[19] C. W. Lee and W. Ma, "Simulation investigation of spatial interpolations in virtual rotating array beamforming with different array configurations for rotating sound source localization," *J. Sound Vibrat.*, vol. 560, Sep. 2023, Art. no. 117784.
[20] C. Zhang, R. Wang, L. Yu, and Y. Xiao, "Order domain beamforming for the acoustic localization of rotating machinery under variable speed working conditions," *Appl. Acoust.*, vol. 205, Mar. 2023, Art. no. 109290.
[21] M. S. Brandstein, J. E. Adcock, and H. F. Silverman, "Microphone-array localization error estimation with application to sensor placement," *J. Acoust. Soc. Amer.*, vol. 99, no. 6, pp. 3807–3816, Jun. 1996.
[22] B. Yang, "Different sensor placement strategies for TDOA based localization," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2007, pp. 1093–1096.
[23] Z. G. Feng, K. F. C. Yiu, and S. E. Nordholm, "Placement design of microphone arrays in near-field broadband beamformers," *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1195–1204, Mar. 2012.
[24] A. Trucco and V. Murino, "Stochastic optimization of linear sparse arrays," *IEEE J. Ocean. Eng.*, vol. 24, no. 3, pp. 291–299, Jul. 1999.
[25] A. Trucco, "Weighting and thinning wide-band arrays by simulated annealing," *Ultrasonics*, vol. 40, nos. 1–8, pp. 485–489, May 2002.
[26] M. F. Delgado, J. A. R. Gonzalez, R. Iglesias, S. Barro, and F. A. Pena, "Fast array thinning using global optimization methods," in *Proc. Eur. Conf. Antennas Propag.*, 2010, pp. 1–3.
[27] K. Chen, X. Yun, Z. He, and C. Han, "Synthesis of sparse planar arrays using modified real genetic algorithm," *IEEE Trans. Antennas Propag.*, vol. 55, no. 4, pp. 1067–1073, Apr. 2007.
[28] A. Razavi and K. Forooraghi, "Thinned arrays using pattern search algorithms," *Prog. Electromagn. Res.*, vol. 78, pp. 61–71, 2008.
[29] P. L. Son, "Irregular microphone array design for broadband beamforming," *Signal Process.*, vol. 193, Apr. 2022, Art. no. 108431.
[30] X. Chen, C. Pan, J. Chen, and J. Benesty, "Planar array geometry optimization for region sound acquisition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, 2021, pp. 11–5.
[31] K. Goto, L. Li, R. Takahashi, S. Makino, and T. Yamada, "Study on geometrically constrained IVA with auxiliary function approach and VCD for in-car communication," in *Proc. Asia–Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA ASC)*, Dec. 2020, pp. 858–862.
[32] H. Segawa, R. Takahashi, R. Jinzai, S. Makino, and T. Yamada, "Applying virtual microphones to triangular microphone array in in-car communication," in *Proc. Asia–Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA ASC)*, Dec. 2020, pp. 421–425.
[33] M. Tsujikawa, A. Sugiyama, K. Hanazawa, and Y. Kajikawa, "Linear microphone array parallel to the driving direction for in-car speech enhancement," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2023, pp. 1–5.
[34] C. T. Ishi, A. Utsumi, and I. Nagasawa, "Analysis of sound activities and voice activity detection using in-car microphone arrays," in *Proc. IEEE/SICE Int. Symp. Syst. Integr. (SII)*, Jan. 2020, pp. 640–645.
[35] Y. Konforti, I. Cohen, and B. Berdugo, "Array geometry optimization for region-of-interest broadband beamforming," in *Proc. Int. Workshop Acoustic Signal Enhancement (IWAENC)*, Sep. 2022, pp. 1–5.
[36] R. Moisseev, G. Itzhak, and I. Cohen, "Array geometry optimization for region-of-interest near-field beamforming," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, vol. 17, Apr. 2024, pp. 576–580.
[37] D. Ayllón, R. Gil-Pita, M. Utrilla-Manso, and M. Rosa-Zurera, "An evolutionary algorithm to optimize the microphone array configuration for speech acquisition in vehicles," *Eng. Appl. Artif. Intell.*, vol. 34, pp. 37–44, Sep. 2014.
[38] B.-C. Kim and I.-T. Lu, "High resolution broadband beamforming based on the MVDR method," in *Proc. OCEANS MTS/IEEE Conf. Exhibition. Conf.*, vol. 2, Jun. 2000, pp. 1025–1028.
[39] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Amer.*, vol. 65, no. 4, pp. 943–950, Apr. 1979.
[40] List of Contributors, *Recent Research Towards Advanced Man-machine Interface Through Spoken Language*. Amsterdam, The Netherlands: Elsevier, 1996, pp. 7–12, doi: 10.1016/B978-044481607-8/50046-3. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780444816078500463
[41] H. Mohammed and T. Rashid, "FOX: A FOX-inspired optimization algorithm," *Appl. Intell.*, vol. 53, no. 1, pp. 1030–1050, Jan. 2023.
[42] Y. Grenier, "A microphone array for car environments," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Jun. 1992, pp. 305–308.
[43] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "An algorithm for intelligibility prediction of time-frequency weighted noisy speech," *IEEE Trans. Audio. Speech Lang., Process.*, vol. 19, no. 7, pp. 2125–2136, Dec. 2011.
[44] M. Ebrahimi, M. Karimi, and M. Modarres-Hashemi, "Optimal sparse linear array design with reduced mutual coupling effect," *AEU Int. J. Electron. Commun.*, vol. 170, Oct. 2023, Art. no. 154781.

Lehai Liu is currently pursuing the Ph.D. degree with Tianjin University, Tianjin, China, focusing on microphone array processing and speech enhancement.

Fengrong Bi received the M.S. degree in construction machinery from the Hebei University of Technology, Tianjin, China, in 1994, and the Ph.D. degree in power machinery and engineering from Tianjin University, Tianjin, in 2003.

He is currently a Professor with Tianjin University. His research interests mainly focus on the control of vibration and noise and the fault diagnosis of vehicle and power machinery.



Jiewei Lin received the B.S. degree in energy and power engineering, the M.S. degree in power machinery and engineering, and the Ph.D. degree from Tianjin University, Tianjin, China, in 2007, 2009, and 2013, respectively.

He is a Professor at the State Key Laboratory of Engines, Tianjin University. His research interests mainly focus on the mechanical system vibration and fatigue.

Tongtong Qi is with the Institute of Internal Combustion Engines, Tianjin University, Tianjin, China.

Xin Li (Member, IEEE) is a Research Fellow at the College of Computing and Data Science, Nanyang Technological University, Singapore. His research interests include machine learning, signal processing, and wireless sensing.