Stochastic allocation strategy for irregular arrays based on geometric feature control

Jingjing Yu\textsuperscript{1}, Qi Xi\textsuperscript{1}, Runlei Li\textsuperscript{1}, Hui Tian\textsuperscript{2} and Yaxi Xie\textsuperscript{1}

Abstract
Irregularities in microphone distribution enrich the diversity of spatial differences to decorrelate interferences from the beamforming target. However, the large degrees of freedom of irregular placements make it difficult to analyse and optimize array performance. This article proposes fast and feasible optimal irregular array design methods with improved beamforming performance for human speech. Important geometric features are extracted to be used as the input vector of the neural network structure to determine the optimal irregular arrangements of sensors. In addition, a hyperbola design method is proposed to directly cluster microphones in the hyperbola areas to produce rich differential distance entropies and yield significant signal-to-noise ratio improvements. These methods can be easily applied to guide non-computer-aided optimal irregular array designs for human speech in acoustic scenes such as immersive cocktail party environments.

Keywords
Irregular array, beamforming, geometric feature, sensor distribution, cluster design

Date received: 12 January 2020; accepted: 2 April 2020
Handling Editor: Xi (Vincent) Wang

Introduction
Microphone array processing uses time, spatial and spectral diversity to capture target acoustic signals and suppress interference and noise. Although regular arrays with uniformly spaced elements have been well studied, their performance are typically limited by the problems derive from symmetrical array arrangement, such as spatial aliasing and inconsistent performance over signal spectral.\textsuperscript{1–3} It has been demonstrated that irregular arrays have the potential to outperform regular ones, especially for speech signals in immersive environments.\textsuperscript{4,5} However, the large degrees of freedom of irregular microphone placements make it difficult to analyse and optimize array performance. Although optimization approaches have been proposed for irregular arrays with minimized gain pattern residues from desired pattern shape, it is not clear what are the crucial geometric features to determine the beamforming performance of totally randomly distributed array, such as array aperture and element spacing for regular arrays.\textsuperscript{6,7} Irregular array synthesis algorithms for antenna design have been proposed in the literature.\textsuperscript{8–10,11–13} But they are not easily feasible for broadband human speech signals with limited knowledge of possible source locations, dynamic acoustic scenes and unknown sound propagation models. In the applications with moving

\textsuperscript{1}School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing, China
\textsuperscript{2}School of Information and Communication Technology, Griffith University, Gold Coast, QLD, Australia

Corresponding author:
Jingjing Yu, School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing 100044, China.
Email: jjyu@bjtu.edu.cn

Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (https://creativecommons.org/licenses/by/4.0/) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).
sources and various background noise, such as high-speed train and crowded public scene with irregular space, direct and rapid array design method with stochastic arrangement of sensors based on the prior knowledge of acoustic environment remains lacking.

Therefore, this article proposes fast and feasible optimal irregular array design methods with enhanced signal-to-noise ratio (SNR) performance for human speech applications in immersive environment. Important distribution features, which are demonstrated theoretically and experimentally to show great impacts on performance metrics, are extracted to be applied as the input of a pre-trained neural network (NN) structure to predict array performance without the use of time-consuming simulations. Optimal arrays with high probabilities of superior beamforming performance can be directly picked based on prior knowledge of acoustic scenes. Another cluster design method is also developed based on hyperbola theory. Optimal arrays can be directly generated for specified acoustic scenes by clustering microphones in the hyperbola areas to produce rich differential distance entropies and provide superior noise suppression capability, even without optimization.

**Problem formulation**

Assuming a three-dimensional (3D) space for the field of view (FOV), \( u(t; r_s) \) stands for the source signal transmitted from position vector \( r_s \). The received signal of the \( p \)th microphone can be given as

\[
v_p(t; r_s, r_p) = u(t; r_s) * h(t; r_s, r_p)
\]

(1)

\[
h(t; r_s, r_p) = \sum_{n=0}^{\infty} a_{qm}(t - \tau_{qm})
\]

(2)

where \( h(\cdot) \) is the impulse response of the propagation path from \( r_s \) to \( r_p \), \( a_{qm}(t) \) is the response of \( n \)th path propagation model and \( \tau_{qm} \) is the corresponding time delay.

To consider the impact of microphone positions, delay-and-sum beamforming algorithm is applied with inverse distance weighting in this article. The expected optimal geometries should statistically enhance the performance of array, regardless of the beamforming types. The power gain leaked between beamforming focal point \( r_i \) and spatial points in FOV can be expressed in frequency as

\[
S(r_i, r_s) = \int \sum_{p=1}^{P} \sum_{q=1}^{Q} B_q B_p^* \hat{V}_p^*(\omega; r_s, r_p) \hat{V}_q(\omega; r_s, r_q) \exp(j\omega(\tau_{ip} - \tau_{iq}))d\omega
\]

where \( r_s \) is the possible sound source location in FOV, \( \hat{V}_p \) is the Fourier transform of received signal of \( p \)th microphone \( (p = 1, 2, ..., P) \), \( B_{ip} = 1/d_{ip} \) and \( \tau_{ip} = d_{ip}/c \), where \( d_{ip} \) is the propagation distance from beamforming target at \( r_i \) to microphone position \( r_p \) and \( c \) is the real-time speed of sound measured in application scenes.

When searching for the optimal geometry features, the coefficients of delay-and-sum beamformer can be considered as the function of microphone coordinates. The distribution of differential path from all pairwise microphones to the potential source positions is the important statistical factor to determine the array beamforming ability for noise suppression.\(^7,14,15\) By applying the expected operations in equation (3), assuming the attenuation factors are uncorrelated with pairwise distance differences of microphones and considering only direct path propagation,\(^14-16\) the output power of beamformer for sound sources at and away from focal points can be expressed as

\[
E(S(r_i, r_s)) = P^2 \int \langle |\hat{U}(\omega; r_i)|^2 \rangle E(\hat{B}_{iq} B_{ip} A_{iq} A_{ip}^*)
\]

\[
E\left(\exp\left(j2\pi\left(\frac{d_{iq} - d_{ip}}{\lambda} + \frac{d_{ip} - d_{iq}}{\lambda}\right)\right)\right) d\omega
\]

(4)

where the angular brackets represent the average power of source signals. As seen in equation (4), for the target source located at \( r_i = r_s \), the complex exponential terms become 1, and the target signal is enhanced by the coherent addition of differential path of pairwise microphones, regardless of array geometry. And for the interfering sources, weaker average output power of beamformer is generated, due to incoherent phases of the exponential terms.

As shown in equation (4), the key point for noise source suppression is to increase the incoherent level of transmission phases, which are related to the differential-path distance (DPD) distribution of overall pairwise microphones to the interfering sources and focal point. With fixed signal spectral and possible source distribution, limited range of DPD levels results in stronger partial coherence for multichannel signals received from interfering sources and might degrade the SNR performance of beamformer. Therefore, when searching for the optimal array geometry, instead of identifying exact positions of each microphone, the diversity and spread of DPDs are important for achieving incoherence to suppress noise signals. DPD distribution with wide range and rich diversity (such as uniform distribution) with the phase terms spreading from \(-\pi\) to \(\pi\) can results in a near zero power gain for non-target source positions.

As shown in Table 1, combining with the typical array parameters of aperture and centroid, statistics based on DPD distributions can be considered as important geometric features to characterize similar
arrays and predict the beamforming performance of arrays without any Monte Carlo experiments. Table 1 also lists results from multi-way analysis of variance (ANOVA) to further demonstrate the strong correlation between geometric features and key performance matrices of array, such as mainlobe width (MLW) and mainlobe-to-peak-sidelobe ratio (MPSR). The proposed geometry features \{L, a, s, J\} show highly significant impacts \(p < 0.01\) on the performance matrices of beampattern. They can explain over 80% of the array performance variance (as shown in \(R^2\)) when beamforming for human speech signals.

In the next section, proposed geometric features are applied as the input vector for array optimization algorithms (e.g. a NN structure) to rapidly predict array SNR performance. Considering mutable acoustic applications, such as high-speed train and crowded public scenes, feasible cluster design method for stochastic arrays is proposed to directly generate optimal microphone clusters with good values of proposed features and to guide fast non-computer-aided optimal array design.

### Optimal geometries for stochastic arrays

#### NN method

Because the relationship between irregular array features and beamforming performance is complex and nonlinear, a deep NN, which is good at nondeterministic mapping, is applied in this section. Geometry features extracted from the acoustic scene along with microphone number are applied as the first layer of a NN structure to rapidly predict the array beamforming performance for human speech signals.

As shown in Figure 1, microphone positions and prior knowledge about the acoustic scene are considered as the inputs, including probability density functions of possible target and noise source locations, related to the usual moving tracks and speaking manners of sources’ behaviour. If no prior knowledge is available, uniform distribution is the default setting to evenly consider all the spatial points in FOV as the possible source location. The objective function is expressed as

---

**Table 1.** Key geometry features related to array beamforming performance.

| Features            | Definition                                                                 | Correlation with performance (ANOVA)               |
|---------------------|---------------------------------------------------------------------------|---------------------------------------------------|
| Array centroid offset \{L\} | The distance from array centroid to the focal point of beamformer         | MLW: \(p < 0.0001\)  \(MLW - R^2 = 0.997\)  
                        | MPSR: \(p < 0.001\)  \(MPSR - R^2 = 0.817\) |
| Array dispersion \{a\}   | The standard deviation of microphone coordinates about array centroid      | MLW: \(p < 0.0001\)  
                        | MPSR: \(p < 0.0366\) |
| DPD statistics \{\sigma, J\} | The standard deviation and normalized Shannon entropy of DPDs\(^{17}\) | MLW: \(p < 0.0031\)  
                        | MPSR: \(p < 0.0001\) |

ANOVA: analysis of variance; MPSR: mainlobe-to-peak-sidelobe ratio; MLW: mainlobe width; DPD: differential-path distance.
where $G$ represents the microphone distribution with specified geometric features, $\Phi(G, r_i, r_s)$ represents the relations between geometric features and key performance matrices in specified scene and $p(r_i)$ and $p(r_s|r_i)$ represent the probability density functions of desired target and interfering source locations. The criterion searching for optimal array geometry can be given by

$$G_{opt} = \arg\min_{G \in \text{mic space}} \langle F(G) \rangle$$

The first layer of the optimization structure extracts five geometric features from the input vector, which are \{L, a, \sigma, J\} and microphone number. Two pre-trained sub-NNs are applied, one to serve as an array-noise-suppression metric and the other as a metric of spatial resolution. Each subnetwork is a two-layer feed-forward network, trained with Bayesian regularization based on a data set of 3D array gain patterns collected using Monte Carlo experiments with human speech signals. Thirty-five neurons are applied in each hidden layer of the subnetwork. For both the training and testing data sets, the regression $R$ values reach 91%-96%, representing successful mappings from the selected array features to the key beamforming performance metrics. The outputs of each subnetwork are combined under probabilistic rules and constraints are derived from the acoustic scene. The experimental results of this optimization scheme are presented in the later section.

### Hyperbola cluster design

It has been demonstrated that high entropy and wide spread of DPD distribution derived from array geometric statistics and acoustic scene can increase the incoherence of noise components in received multi-channel signals and further improve beamforming SNR. However, because DPD statistics do not have intuitive simple geometric interpretations that can be used to guide allocation of microphone distribution directly for mutual application environment, a cluster design method based on hyperbola area is proposed in this section for non-computer-aided optimal array design.

By defining the hyperbola areas based on knowledge of acoustic scene, the hyperbola cluster (HC) method can be used to directly generate an optimal array with good values of geometry features and further guide non-computer-aided optimal microphone placements. As mentioned in equation (4), with pairwise microphones $\{p, q\}$ and two spatial positions $\{r_i, r_s\}$ in FOV, the DPD term can be rewritten as

$$\Delta_{pq}(r_i, r_s) = (d_{pq} - d_{ip}) - (d_{pq} - d_{iq})$$

where different value of $(d_{pq} - d_{iq})$ can be marked by hyperbola curve with two focuses at $\{r_i, r_s\}$.

As shown in Figure 2, hyperbola curve is explained as the locus of points with a constant absolute value of differential distances to two focuses. With given two spatial positions $\{r_i, r_s\}$ for sound sources, microphones located on one hyperbola curve show the same value of $(d_{pq} - d_{iq})$. When microphones move towards the outside of hyperbola pair (in the grey area), the absolute value of $(d_{pq} - d_{iq})$ increases. Therefore, in

![Figure 2. Optimal array geometries. The blue circles represent distributed microphones. The red crosses represent the possible noise source locations. The red triangle represents desired target as focal point of beamformer. (a) Computer-aided GA array and (b) HC array.](image-url)
order to generate a large spread of DPDs in equation (7), microphone clusters should be distributed in both hyperbola areas. In addition, there is no need to place microphones in the middle area of \( r_1, r_2 \), because the small values of DPD distribution can be generated by the nearby microphones in the same hyperbola area. Therefore, with prior knowledge of possible source distribution in acoustic scene, a large spread and rich entropy of DPDs for each target and noise source pair can be generated by simply placing small microphone clusters over the hyperbola areas, which would provide superior noise suppression ability of beamformer.

Figure 2 gives the optimal arrays resulted from computer-aided heuristic searching\(^ {19} \) and hyperbola cluster design method. The hyperbola areas are marked by dashed lines with different colours. In Figure 2(a), it can be seen that the optimal geometries resulted from genetic algorithm (GA) searching,\(^ {5,19,20} \) most of the microphones are actually clustered in the hyperbola grey areas, which demonstrates the effectiveness of hyperbola analysis. Figure 2(b) provides a corresponding HC array directly generated by the HC method. Simulations and real-case experiments with human speech signals have demonstrated that the designed HC arrays show comparable or even better beamforming SNR, when compared with computer-aided optimized GA arrays.

### Experimental results

Experiments in three acoustic scenes with different potential source distributions/spaces were performed to evaluate SNR performance for human speech signals. Audio cage with the size of \( 10 \times 10 \times 2 \text{ m}^3 \) was applied to simulate the indoor immersive environment for multi-source audio surveillance application cases. Three types of optimized arrays were employed: optimized arrays obtained by 100 GA iterations, arrays directly generated by HC and arrays selected from random distributions by a NN structure. In addition, the SNRs of a relevant random array set and regular array with the same centroid and dispersion are also provided for comparisons.

Table 2 compares the SNR results of the random array set and regular arrays in cocktail party experiments. Sound sources transmitting broadband human speech signals are distributed in the audio cage and are randomly selected as the target and noise sources. For specified geometry sets with similar aperture, average and top SNRs over 100 arrays are computed to demonstrate the effectiveness of proposed array geometry optimization method. All three types of optimal irregular geometries revealed enhanced beamforming performance, which demonstrates the feasibility of the array design methods proposed in this article and the effectiveness of proposed geometric features. Due to the statistical parameters and probabilistic rule applied in the optimization, as the acoustic scene becomes more complicated (more potential speakers and more microphones in an overlapping noise/target space), an even stronger SNR improvement can be observed.

Table 2. SNR (dB) comparisons for cocktail party experiments.

| Acoustic scenes          | Sixty-four microphones with continuous noise space | Sixty-four microphones with discrete noise sources | Nine microphones with discrete noise sources |
|-------------------------|---------------------------------------------------|---------------------------------------------------|---------------------------------------------|
|                         | Top three SNR                                      | Top three SNR                                     | Top three SNR                                |
| GA-optimized array set  | 9.03, 8.99, 8.98                                   | 26.33, 25.78, 25.63                               | 18.24, 17.71, 17.65                          |
|                         | Average SNR                                        | Average SNR                                       | Average SNR                                  |
|                         | 8.45                                               | 23.83                                             | 16.56                                        |
| HC array set            | Top three SNR                                      | Top three SNR                                     | Top three SNR                                |
|                         | 8.92, 8.87, 8.71                                   | 28.04, 27.38, 27.03                               | 21.83, 21.07, 20.69                          |
|                         | Average SNR                                        | Average SNR                                       | Average SNR                                  |
|                         | 6.81                                               | 24.76                                             | 17.93                                        |
| NN array set            | Top three SNR                                      | Top three SNR                                     | Top three SNR                                |
|                         | 10.73, 9.43, 8.74                                 | 24.42, 24.10, 23.13                               | 19.14, 16.67, 16.65                          |
|                         | Average SNR                                        | Average SNR                                       | Average SNR                                  |
|                         | 6.96                                               | 20.05                                             | 10.00                                        |
| Random array set        | Top three SNR                                      | Top three SNR                                     | Top three SNR                                |
|                         | 6.47, 5.67, 5.49                                   | 22.61, 22.18, 22.09                               | 17.78, 17.62, 17.20                          |
|                         | Average SNR                                        | Average SNR                                       | Average SNR                                  |
|                         | 3.83                                               | 17.96                                             | 9.01                                         |
| Regular array           | SNR                                                |                                                   | 8.89                                         |

HC: hyperbola cluster; NN: neural network; SNR: signal-to-noise ratio.
without time-consuming optimization or heuristic searching by GA, as the direct design method, HC and NN directly generate optimal geometries with higher probability to show good beamforming performance. These direct-designed optimal arrays show comparable or even better SNR results than computer-aided GA arrays. And meanwhile, large SNR improvements are observed compared with corresponding regular arrays. In Figure 3, the top-view gain patterns for real-case cocktail party experiments when targeting the top source are given. It can be seen that our optimal arrays showed superior noise suppression abilities in this scene, whereas the spatial resolutions are also improved in comparison with the regular ones.

**Conclusion**

This article has proposed feasible irregular array design methods with improved beamforming performance for cocktail party applications. Important geometric features have been proposed for use as NN structure inputs to predict array performance and directly pick optimal irregular geometries with good beamforming performance. In addition, in order to generate rich DPD entropy to better suppress noise signals, HC arrays derived from hyperbola areas can be directly generated based on prior knowledge of acoustic scene, providing improved SNR performance comparable with other complex optimization methods. Proposed method can be easily applied to guide non-computer-aided optimal irregular array design in dynamic multi-source acoustic applications such as mobile platforms with changing acoustic scenes and high-speed trains/aircraft with irregular spaces.

**Declaration of conflicting interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Funding**

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article.
This work was supported by Fundamental Research Funds for Central Universities (Grant No. 2018JBM008), National Natural Science Foundation of China (Grant No. 61501025) and Beijing Natural Science Foundation (Grant No. 4172045).

ORCID iD
Jingjing Yu https://orcid.org/0000-0002-1491-0377

References
1. Benesty J, Chen J and Huang Y. Microphone array signal processing. Berlin: Springer, 2008.
2. Rabinkin D, Renomeron R and French J. Optimum sensor placement for array sound capture. Proc SPIE 1997; 3162: 227–239.
3. Christensen J and Hald J. Technical review beamforming. Naerum: Bruel & Kjaer Sound & Vibration Measurement A/S, 2004, pp.18–54.
4. Yoon H and Song T-K. Sparse rectangular and spiral array designs for 3D medical ultrasound imaging. Sensors 2020; 20(1): E173.
5. Yu J. Microphone array optimization in immersive environments. PhD thesis, University of Kentucky, Lexington, KY, 2013.
6. Rabinkin D. Optimum sensor placement for microphone arrays. PhD thesis, Rutgers, The State University of New Jersey, New Brunswick, NJ, 1998.
7. Cheng D. Optimization techniques for antenna arrays. Proc IEEE 1971; 59(12): 1664–1674.
8. Bucci OM, D’Urso M, Isernia T, et al. Deterministic synthesis of uniform amplitude sparse arrays via new density taper techniques. IEEE Trans Antennas Propag 2010; 58(6): 1949–1958.
9. Aslan Y, Puskely J, Roederer A, et al. Multiple beam synthesis of passively cooled 5G planar arrays using convex optimization. IEEE Trans Antennas Propag. Epub ahead of print 9 December 2019. DOI: 10.1109/TAP.2019.2955885.
10. Oliveri G, Gottardi G and Massa A. A new metaparadigm for the synthesis of antenna arrays for future wireless communications. IEEE Trans Antennas Propag 2019; 67(6): 3774–3788.
11. Pal P and Vaidyanathan PP. Nested arrays: a novel approach to array processing with enhanced degrees of freedom. IEEE Trans Signal Process 2010; 58(8): 4167–4181.
12. Chi Y. Exploitation of geometry in signal processing and sensing. PhD thesis, Princeton University, Princeton, NJ, 2012.
13. Haykin S. Adaptive filter theory. 5th ed. Essex: Pearson Education, 2017, pp. 20–31.
14. Yu J and Donohue K. Geometry descriptors of irregular microphone arrays related to beamforming performance. EURASIP J Adv Signal Process Epub ahead of print 27 November 2012. DOI: 10.1186/1687-6180-2012-249.
15. Yu J and Donohue K. Performance for randomly described arrays. In: Proceedings of the 2011 IEEE workshop on applications of signal processing to audio and acoustics, New Paltz, NY, 16–19 October 2011. Piscataway, NJ: IEEE.
16. Donohue K, Saghaian Nejad Esfahani S and Yu J. Constant false alarm rate sound source detection with distributed microphones. EURASIP J Adv Signal Process. Epub ahead of print 15 February 2011. DOI: 10.1155/2011/656494.
17. Pielou EC. The measurement of diversity in different types of biological collections. J Theor Biol 1966; 13: 131–144.
18. Townsend P. Enhancements to the generalized sidelobe canceler for audio beamforming in an immersive environment. MS thesis, University of Kentucky, Lexington, KY, 2009, pp.5–10.
19. Yu J and Donohue K. Optimal irregular microphone distributions with enhanced beamforming performance in immersive environments. J Acoust Soc Am 2013; 134: 2066–2077.
20. Akbari R and Ziarati K. A multilevel evolutionary algorithm for optimizing numerical functions. Int J Ind Eng Comp 2010; 2: 419–430.