Semi-Coprime Microphone Arrays for Estimating Direction of Arrival of Speech Sources

Jiahong Zhao and Christian Ritz
School of Electrical, Computer and Telecommunication Engineering, University of Wollongong, NSW, Australia
E-mail: jz262@uowmail.edu.au and critz@uow.edu.au

***Abstract*— This paper evaluates the performance of semi-coprime microphone arrays (SCPMAs) for speech source direction of arrival (DOA) estimation based on the steered response power – phase transform (SRP-PHAT) algorithm. The SCPMA is an extension of the coprime microphone array (CPMA), which combines the outputs of three sub-arrays to reduce the impact of spatial aliasing and achieves performance comparable to that obtained from arrays using much larger numbers of microphones. The proposed approach considers two different processors to calculate the outputs from the sub-arrays and adapts the SRP-PHAT approach to these arrays. Simulations are conducted under anechoic and reverberant scenarios in a noisy room. Beam pattern and array gain results indicate that the SCPMA works better than the conventional CPMA at reducing the peak side lobe (PSL) level and total side lobe area while increasing the capability of amplifying the desired target signal and restraining noise from all other directions for typical frequencies of speeches. DOA Estimation results also show that the SCPMA achieves accurate DOA estimates in anechoic and low reverberant conditions, which is comparable to the equivalent full ULA, while the large side lobes in the beam pattern of the SCPMA lead to less accurate results in the highly reverberant environment.**

I. INTRODUCTION

Microphone array geometries and related signal processing algorithms have been investigated for a few decades to solve problems in the acoustics context, including estimating the number of sound sources, direction of arrival (DOA) estimation and enhancing the intelligibility of humans' speeches. Compared with methods applied to conventional antenna arrays, microphone arrays require broadband processing, and the performance across a wide band of frequencies need to be evaluated [1], [2].

At present, the most commonly used geometries of microphone arrays are the uniform linear array (ULA) and uniform circular array (UCA), while more and more sophisticated microphone arrangements are used to better analyze the acoustic scenes or capture the sound field such as ad hoc microphone arrays, B-format microphones and spherical microphone arrays [3], [4], [5]. For broadband recording, one of the most crucial issues is spatial aliasing, which occurs when the inter-element spacing of a uniform microphone array is larger than half of the wavelength of recorded signal according to the spatial Nyquist theorem. The spatial aliasing results in grating lobes in the beam pattern and causes ambiguity in differentiating the desired source with sources propagating from the directions of the grating lobes. Coprime arrays (CPA) are proved to present capabilities to cancel the grating lobe, while having narrower main lobe and smaller side lobe than the beam pattern of a ULA with same number of elements [6]. The coprime microphone array (CPMA) uses this sensor structure and interleaves two uniform linear sub-arrays, which has been applied to speech sources to estimate the DOA [7], [8]. The accuracy is improved, but there are also larger side lobes in the coprime beam pattern, which is a potential problem of the CPMA. Additionally, these previous work do not consider the effects of noise to DOA estimation, which is a normal issue in the real world. To decrease the peak side lobe (PSL) level and side lobe area in the beam pattern, a new type of sensor array is proposed in [9], which is the semi-coprime array (SCA). The SCA introduces an extra sub-array compared with the CPA, resulting in the interleaving of three sub-arrays. This array geometry has been shown to possess strength in having smaller side lobes in the beam pattern than that of a CPA with same number of elements, leading to better DOA estimation results. However, the application of SCA in acoustics and the DOA estimation of broadband signals are not discussed. Besides, the proposed structure of SCA uses a min processor, proposed in [10], to merge the sub-array outputs, while other processors can also be applied.

This paper investigates the semi-coprime microphone array (SCPMA) and its application in DOA estimation of speech sources. Two most popular processors are applied to the SCPMA and CPMA, which are the product processor and min processor, to explore probabilities in further canceling the side lobe in beam pattern while increasing the array gain. In addition, adaption of the conventional SRP-PHAT algorithm used for estimating the DOA is proposed to match the characteristics of the processors. The performance is evaluated through simulations under different levels of noise and reverberation.

The following chapters are organized as follows. Section II briefly reviews the CPMA and then generalizes the array geometry of SCPMA, followed by formulating the signal model of recording. Common workflow of the processors used for sub-array processing and performance measures of microphone arrays are also introduced. In Section III, equations for adapting SRP-PHAT to the processors are derived, which are then applied to a histogram-based DOA estimation method. Simulation results for different testing scenarios are shown and discussed in Section IV with conclusions provided in Section V.

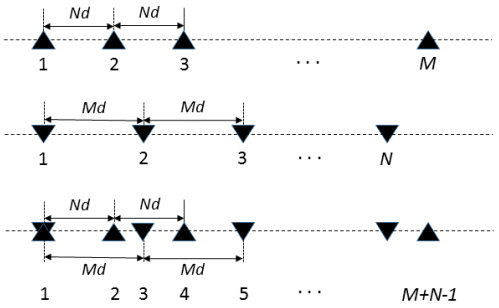

Fig. 1 Generalized geometry of coprime microphone arrays

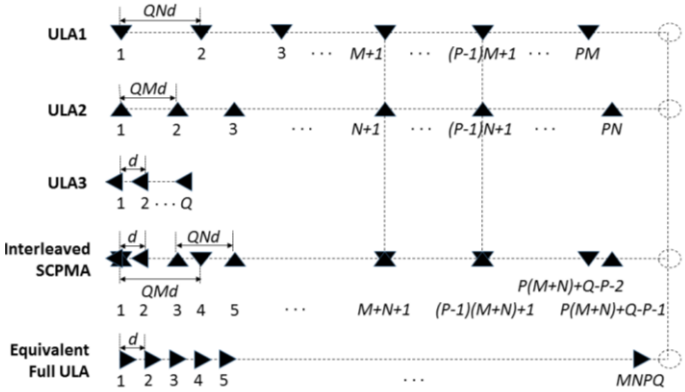

Fig. 2 Generalized structure of semi-coprime microphone arrays

II. Coprime and Semi-Coprime Microphone Arrays and Their Processors

A. Mathematical Model

A conventional CPMA is formed by interleaving two uniform linear sub-arrays, with the number of microphones of both being coprime numbers, and the geometry is illustrated in Fig. 1 [8]. Two coefficients are involved in the CPMA, which are the numbers of microphones of the two sub-arrays, *M* and *N*. The inter-microphone spacing of the two sub-arrays is also coprime, with the *M*-element sub-array being *Nd* and the other being *Md*. The two sub-arrays share the first microphone so that the aperture of both is the same, leading to an overall array with $M + N - 1$ microphones. To express the microphone numbers of the CPMA differently from that of the SCPMA, the following sections denote the *M* and *N* as M_c and N_c, respectively.

A SCPMA is a sparse array that interleaves three uniform linear sub-arrays, and the array structure is determined by setting four parameters, *M*, *N*, *P* and *Q*, as shown in Fig. 2. The *M* and *N* are a pair of coprime numbers, where the only positive common divisor is one. The number of microphones of the first two interleaved ULAs are *PM* and *PN*, with the inter-element spacing being *QNd* and *QMd*, where *d* is the spacing of the equivalent full ULA having equal resolution with the SCPMA. A third sub-array is then interleaved, which has *Q* microphones and the inter-microphone distance is *d*. This arrangement leads to microphone overlaps in the first two sub-arrays, while the three uniform sub-arrays also share the first microphone. As a result, the overall microphone number of the SCPMA can be obtained as $U = P(M + N) + Q - P - 1$.

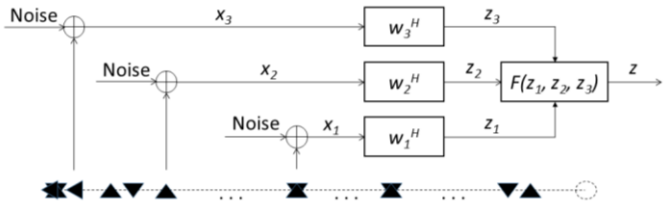

Fig. 3 Common workflow of sub-array processors

In addition, the apertures of ULA1, ULA2, SCPMA and the equivalent full ULA in Fig. 2 are identical, with a virtual microphone located at the rightmost of each array.

Compared with a conventional CPMA, the SCPMA involves coefficients *P* and *Q*, which change the microphone numbers and spacing, and an extra short sub-array is considered [9].

Assuming *K* uncorrelated acoustic sources (narrowband or wideband) are propagating as plain waves at the speed of sound (c = 343 *m/s*) and impinging on a microphone array from different DOAs θ_i (i = 1, 2, … , *K*). The signal model of recording can be expressed as

$$y_u(t) = \sum_{i=1}^{K} h_{u,i}(t) * s_i(t) + v_u(t), \quad (1)$$

where u = 1, 2, …, *U*, $y_u(t)$ is the output in time domain of each individual microphone, and $h_{u,i}(t)$ is the impulse response of source i received by microphone u. $s_i(t)$ and $v_u(t)$ are the ith source signal and additive noise to the uth microphone, respectively. For SCPMAs, the output signal is obtained as a function of the outputs from all three sub-arrays, which needs a processor to calculate the sub-array outputs.

B. Processors and Operating Frequencies

There are a variety of processors that are utilized to process the sub-array outputs, and the common workflow is illustrated in Fig. 3. Taking the SCPMA as an example, the received signal of each sub-array x_i (i = 1, 2, 3) is weighted by w_i (i = 1, 2, 3), resulting in the inputs of processor *F*, which are z_i (i = 1, 2, 3). The output of each sub-array is obtained through a beamforming operation with weights w_a applied to the set of recorded microphone signals, x_a, for each sub-array.

$$z_a = w_a^H x_a \quad (2)$$

where *H* denotes the conjugate-transpose operation, and a ranges from 1 to *A*, with *A* being the number of sub-arrays. The final output from combining all sub-arrays through some processing function

$$z = F(z_1, z_2, \ldots, z_A), \quad (3)$$

where for CPMAs, *A* = 2 and for SCPMAs, *A* = 3.

Two most popular processors for traditional CPMAs are the product processor and min processor, which calculate the product and minimum of all weighted sub-array outputs, respectively. Performance evaluations in this paper are mainly around these two processors.

With regard to the spatial Nyquist sampling theorem, the ULAs suffer from spatial aliasing problem when the distance between neighboring microphones δ is larger than half of the wavelength λ, i.e. $\delta > \lambda / 2$, resulting in ambiguity in distinguishing the desired source and source from other

directions. The operating frequency describes the maximum frequency that enables the array to avoid spatial aliasing, and it satisfies $\delta = \lambda / 2$, in which $\lambda = c / f_{op}$. Thus, the operating frequency of a typical ULA with N_0 microphones is defined as

$$f_{op\_ULA} = \frac{c}{2\delta} = \frac{cN_0}{2L_{ULA}}, \quad (4)$$

where L_{ULA} is the aperture of ULA. A conventional CPMA's operating frequency equals that of its equivalent full ULA, which is [11]

$$f_{op\_CPMA} = \frac{cM_cN_c}{2L_{CPMA}}, \quad (5)$$

where M_c and N_c are microphone numbers of the two sub-arrays, and L_{CPMA} is the aperture of CPMA. In this same way, this paper derives the operating frequency of the SCPMA as

$$f_{op\_SCPMA} = \frac{cMNPQ}{2L_{SCPMA}}, \quad (6)$$

where L_{SCPMA} is the aperture of SCPMA.

C. Performance Measures

There are two key metrics to evaluate the performance of microphone array characteristics, which are the beam pattern and array gain. The beam pattern illustrates the gain of a beam former to sources arriving from different directions. This paper considers the horizontal beam pattern, which means the desired source is assumed to arrive from 90 degrees. Therefore, the beam pattern of a ULA with N_0 microphones can be formulated as follows.

$$\boldsymbol{B}[\boldsymbol{w}(\omega), \theta] = \boldsymbol{w}^H(\omega)\boldsymbol{d}(\omega, \theta) \quad (7)$$

where $\omega = 2\pi f$ is the radian frequency, and $\boldsymbol{w}(\omega)$ of length N_0 represent complex beamforming weights, and this paper assumes equal weights to achieve a unity gain at the desired source DOA. $\boldsymbol{d}(\omega, \theta)$ of length N_0 is the steering vector that is expressed as [1]

$$\boldsymbol{d}(\omega, \theta) = [d_1(\omega, \theta) \quad d_2(\omega, \theta) \quad \cdots \quad d_{N_0}(\omega, \theta)]^T, \quad (8)$$

where $d_n(\omega, \theta) = e^{j\omega(n-1)\delta c^{-1}\cos\theta} (n = 1, 2, \ldots, N_0)$, and the range of θ is 0 to 180 degree.

For the SCPMA and CPMA, the beam pattern of each sub-array is conventionally obtained separately and then combined to generate the overall beam pattern based on different processors. In order to make the magnitudes of beam patterns of the product processor and min processor remain at the same scale, this paper proposes a square root operation to the product-processed beam pattern, and the min processor compares the absolute values of beam patterns, leading to the following beam pattern expressions.

$$\boldsymbol{B}_{CPMA-product} = \sqrt{|\boldsymbol{B}_1 \times \boldsymbol{B}_2^*|} = \sqrt{|\boldsymbol{B}_1| \times |\boldsymbol{B}_2|} \quad (9)$$

$$\boldsymbol{B}_{SCPMA-product} = \sqrt[3]{|\boldsymbol{B}_1'| \times |\boldsymbol{B}_2'| \times |\boldsymbol{B}_3'|}, \quad (10)$$

$$\boldsymbol{B}_{CPMA-min} = min(|\boldsymbol{B}_1|, |\boldsymbol{B}_2|) \quad (11)$$

$$\boldsymbol{B}_{CPMA-min} = min(|\boldsymbol{B}_1'|, |\boldsymbol{B}_2'|, |\boldsymbol{B}_3'|) \quad (12)$$

where * denotes the complex conjugation, and $\boldsymbol{B}_3'$ in (10) and (12) represents the beam pattern of the short sub-array with Q elements of the SCPMA.

Another crucial measure is the array gain (AG), termed as the ratio between the gain to the desired signal and the average gain to spatial noises from all undesired directions [12]. The AG can be given by [13]

$$D[\boldsymbol{w}(\omega)] = \frac{|\boldsymbol{B}[\boldsymbol{w}(\omega), \theta_S]|^2}{1/\Theta \sum_{\theta \neq \theta_S} |\boldsymbol{B}[\boldsymbol{w}(\omega), \theta]|^2}, \quad (13)$$

where θ_s is the steering angle, and Θ is the number of discrete angles used in calculating the beam pattern $\boldsymbol{B}$. The condition $\theta \neq \theta_s$ is specified in (13) to make the calculation clear.

III. DOA Estimation Using SRP-PHAT with Adaption to the Processors

A. Processed SRP-PHAT and DOA Estimation

Ref. [8] has shown that conventional SRP-PHAT can be utilized to accurately estimate the speech DOA, while this paper proposes an approach to adapt SRP-PHAT to different processors.

For SCPMAs and CPMAs, the adapted SRP-PHAT firstly calculates the SRP value $P(\tau)$ at each bearing by summing the PHAT-weighted generalized cross-correlations (GCC) of all combinations of microphone pairs of each sub-array [14].

$$P(\tau) = \sum_{i_1=1}^{N_0} \sum_{i_2=i_1+1}^{N_0} \int_{-\infty}^{+\infty} \frac{\vartheta_{y_1y_2}(f)}{|\vartheta_{y_1y_2}(f)|} e^{j2\pi f\tau} df \quad (14)$$

where $\vartheta_{y_1y_2}(f)$ is the cross-spectrum expressed as follows.

$$\vartheta_{y_1y_2}(f) = E[Y_1(f)Y_2^*(f)] \quad (15)$$

where $E[\cdot]$ obtains the mathematical expectation, and $Y_i(f)$ $(i = 1, 2)$ are the outputs of selected microphone pairs in the frequency domain. Consequently, the SRPs of all sub-arrays are processed using the product processor and min processor. To maintain all results at the same scale to make comparisons, a square root operation is added to the sub-arrays' SRPs before using the product processor.

$$P_{CPMA-product} = \sqrt{P_1} \times \sqrt{P_2} \quad (16)$$

$$P_{SCPMA-product} = \sqrt{P_1'} \times \sqrt{P_2'} \times \sqrt{P_3'} \quad (17)$$

$$P_{CPMA-min} = min(P_1, P_2) \quad (18)$$

$$P_{SCPMA-min} = min(P_1', P_2', P_3') \quad (19)$$

where P_3' in (17) and (19) represents the output power of the Q-element sub-array of SCPMA. Therefore, the preliminary DOA estimates are achieved as

$$\theta_{est} = arccos\left(\frac{c\tau_{opt}}{|\delta_{i_1i_2}| \cdot F_s}\right), \quad (20)$$

where F_s is the sampling frequency, and τ_{opt} is the optimal time lag leading to the largest SRP, which is calculated by

$$\tau_{opt} = \underset{\tau}{argmax}(P_0), \quad (21)$$

where P_0 is used to represent the overall power of array output.

B. SRP-Adjusted DOA Histogram

An SRP-adjusted histogram (SAH) approach has been discussed in [13] to deal with the spreading in the DOA histogram caused by noisy, reverberant or multi-source conditions. By calculating the energy of time-frequency instants (similar to other weighting methods [15], [16]), the low-SRP DOA estimates are removed by using

$$hist_{sah}(\theta_j) = \begin{cases} hist(\theta_j) - 1, & P(\theta_j) < T \\ hist(\theta_j), & P(\theta_j) \geq T \end{cases}, \quad (22)$$

where θ_j ($0^\circ \leq \theta_j \leq 180^\circ$) denotes each possible DOA, $hist$ is the initial histogram and $hist_{sah}$ is the resulting SRP-adjusted histogram. Additionally, T is a pre-defined energy threshold, which is set to the minimum energy value plus one third of the difference between the maximum energy and minimum energy.

Moreover, the DOA estimation accuracy is further increased by modelling the kernel density estimation (KDE) to search for the local maximum of the probability density function (PDF), which can be formulated as [17]

$$\widehat{F'}(x) = \frac{1}{nh}\sum_{k=1}^{n} Z\left(\frac{x - x_k}{bw}\right), \quad (23)$$

where Z is a pre-defined kernel function, bw ($bw > 0$) is known as the bandwidth, x_k ($k = 1, 2, \ldots, n$) are evenly distributed samples, and F' is the distribution of x. The final DOA estimation is achieved by locating peaks of the PDF.

IV. Results and Discussion

A. Experimental Settings

TABLE I
EXPERIMENTAL MICROPHONE ARRAY PARAMETERS

Type of array	Number of microphones	Aperture (m)	f_{op} (Hz)
SCPMA	10	0.8	7717.5
CPMA	10	0.8	6431.3
ULA10	10	0.8	2143.8
ULA36	36	0.8	7717.5

TABLE II
SIMULATION CONFIGURATION

Sampling frequency (F_s)	25 kHz
Frequency bin number for FFT	200
Frame duration	25 ms
Frame overlap	50%
Number of frames	180
Azimuthal range	0° - 180°
Azimuthal resolution	0.1°
Room dimensions	$8 \times 10 \times 5$ m^3
Reverberation time (RT60)	{0, 200, 400} ms
Noise levels (SNRs)	{10, 20, 30, 40, ∞} dB
Ground truth DOAs (S_1, S_2, S_3)	{107.0°, 75.7°, 56.3°}
Source-array distance	6.5 m
Speed of sound (c)	343 m/s

The experiments evaluate performance in terms of the beam pattern, array gain and speech DOA estimation where six types of microphone arrays are considered, which are shown in Table I, and the SCPMA and CPMA are both processed by the product processor and min processor. The four coefficients of SCPMA are set as $P = 2$, $Q = 3$, $M = 2$, $N = 3$, and the configuration of CPMA is $M_c = 5$, $N_c = 6$, so that the number of microphones of both is 10. Two ULAs are also selected for comparison, including a 10-element ULA and a 36-element ULA that is equivalent to the SCPMA in terms of having same resolution and operating frequency. All microphone arrays considered have the same aperture of 0.8 meters, and their operating frequencies are listed in Table I.

Table II explains configuration details of the DOA estimation algorithm, simulated environment and source signals. Speech recordings are calculated using the image method [18], and the original speech sources are a sub-set of three utterances selected from the IEEE corpus (wideband), with the sampling frequency being 25 kilohertz [19]. All sources are located in three fixed positions in the far field, and their distances to the center of microphone arrays are identical. Signals are recorded under different levels of reverberation and additive (white) noise and then are transformed to the short-term frequency domain by utilizing fast Fourier transform (FFT), which makes use of 50% overlapped Hamming windowed frames of 25ms duration. Compared with [8], the effects of the noise are tested. After that, the adapted SRP-PHAT are applied to estimate the DOA of speeches, with the error found by calculating the root mean square error (RMSE). The expression is $RMSE = \sqrt{\frac{1}{C}\sum_{k=1}^{C}(\theta_k - \theta_{true})^2}$, where C is the number of estimates, θ_k ($k = 1, 2, \ldots, C$) is the DOA estimation result, and θ_{true} is the ground truth DOA.

B. Performance Evaluation of Arrays and Processors

Fig. 4 illustrates beam patterns of the 10-element SCPMA and CPMA processed by the product processor and min processor as well as the ULAs using 10 microphones and 36 microphones, separately. Fig. 4 (a) to (d) show plots at an example frequency 5 kilohertz, which is above the Nyquist frequency of 10-element ULA. Fig. 4 (a) and (b) demonstrate the advantages of the min processor over the product processor in terms of the PSL level and overall side lobe areas. For the 10-microphone CPMA, the PSL level is the same no matter which processor is used, while the min-processed SCPMA beam pattern has lower PSL. In addition, the side lobe areas are smaller in both figures. Fig. 4 (c) shows that compared with CPMA, the min-processed SCPMA has a beam pattern of smaller PSL level, which indicates an advantage of cancelling side lobes. In Fig. 4 (d), there are two large grating lobes caused by spatial aliasing in the 10-element ULA beam pattern, leading to ambiguity in distinguishing the desired source and signals from those two directions. The 36-microphone ULA's beam pattern is also plotted, which has small side lobes due to the usage of a large number of microphones.

Beam patterns at frequencies which range over the recorded signals are investigated in Fig. 4 (e) to (h). It can be found that

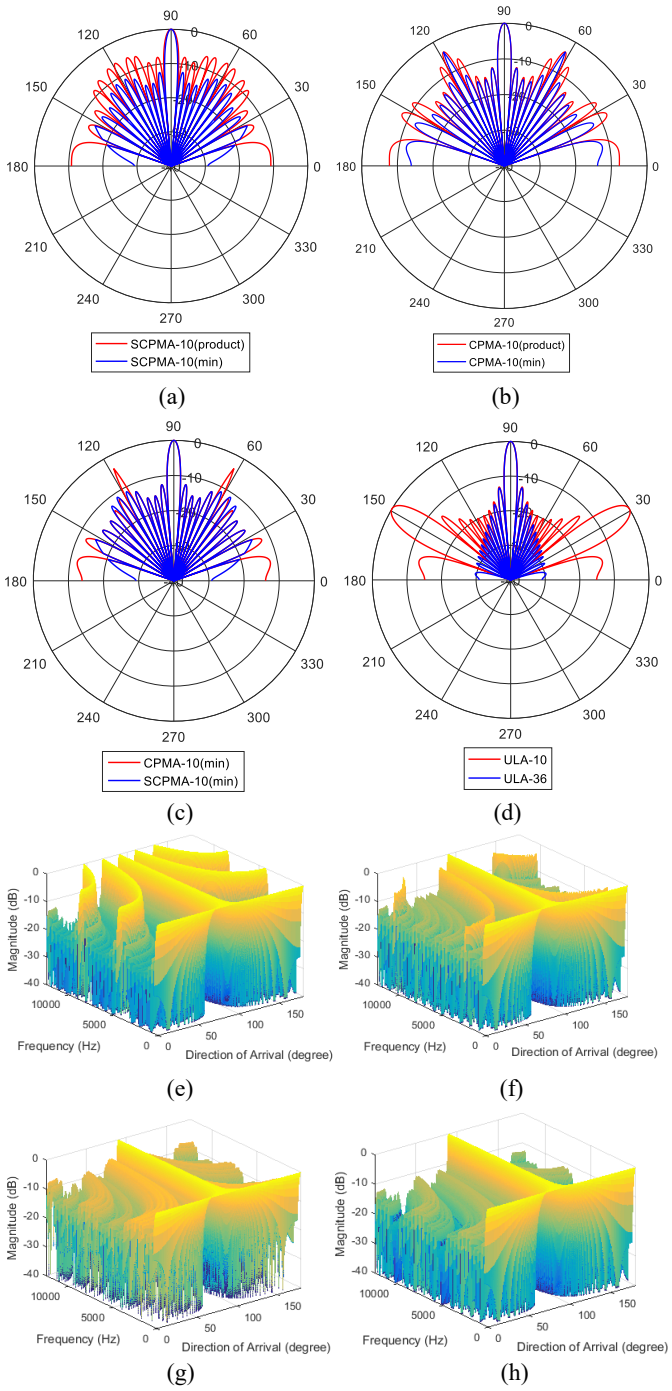

Fig. 4 Beam patterns of the 10-microphone SCPMA and CPMA with the product processor and min processor as well as the contrastive ULAs: (a) – (d) are plotted for 5 kHz sources while for broadband sources: (e) 10-microphone ULA, (f) min-processed CPMA, (g) product-processed SCPMA and (h) min-processed SCPMA. Conditions of simulation: $\theta_s = 90^\circ$.

the total side lobe magnitude in the min-processed SCPMA's beam pattern is smaller than that of the product-processed SCPMA and min-processed CPMA. The 10-element ULA's beam pattern shows large grating lobes, which do not exist in the other three plots.

Fig. 5 plots AGs of the SCPMA and CPMA with the product processor and min processor, separately. Results of the ULAs using 10 and 36 microphones are also illustrated. It can be seen

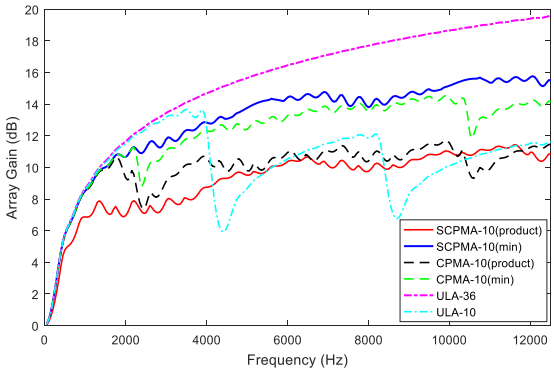

Fig. 5 Comparison of AGs of the 10-microphone SCPMA and CPMA with the product processor and min processor as well as contrastive ULAs. Conditions of simulation: $\theta_s = 90^\circ$.

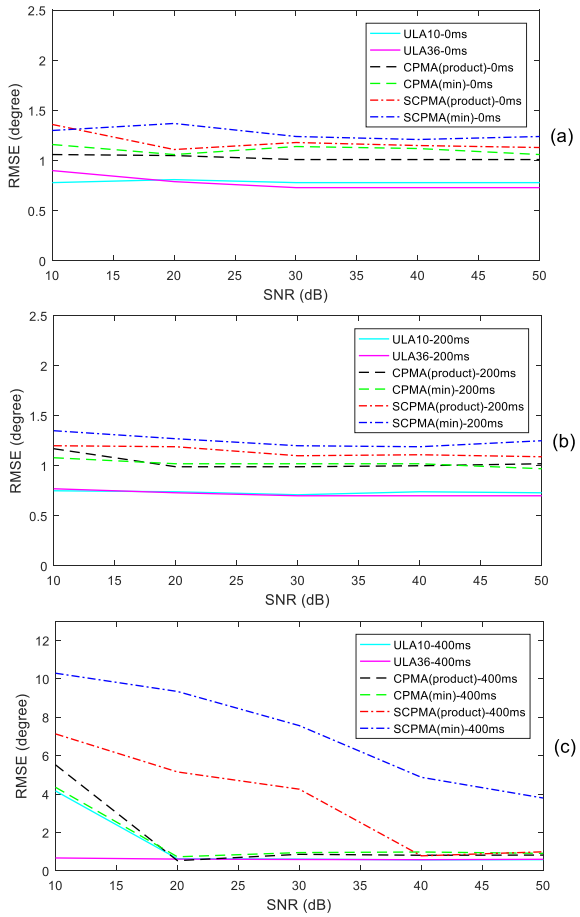

Fig. 6 Evaluating DOA estimation results using adapted SRP-PHAT for the 10-microphone SCPMA, CPMA and the contrastive ULAs under multiple levels of noise and reverberation: (a) anechoic, (b) 200 ms RT60 and (c) 400 ms RT60.

that the AGs of the SCPMA and CPMA with min processor are generally larger than that of the product processor, and the min-processed SCPMA has a greater AG than other microphone arrays with same numbers of microphones. For example, at 6 kilohertz, the elevation of the min-processed SCPMA's gain is about 1 dB compared with the min-processed CPMA's gain. This elevation is around 4 dB if comparing with the gain of the product-processed microphone arrays and 10-element ULA.

The equivalent full ULA has the highest AG, as it utilizes a large number of microphones to achieve a full resolution. The 10-element ULA's AG deteriorates at frequencies larger than 4 kilohertz, which is around twice its Nyquist frequency.

Fig. 6 compares speech DOA estimation results from the SCPMA and CPMA with the product processor and min processor as well as two contrastive ULAs. Fig. 6 (a), (b) and (c) illustrate the results under conditions of anechoic environment, 200-millisecond RT60 and 400-millisecond RT60, separately. The SCPMA with both processors achieve accurate results and have close DOA estimation accuracy with all other types of microphone arrays in Fig. 6 (a) and (b), while under higher room reverberation in Fig. 6 (c), the SCPMA is less accurate in estimating the DOA. The reason may lie in the big side lobes in beam patterns and the degradation in array gains of the SCPMA with both processors.

V. CONCLUSIONS

This paper evaluates performance of the SCPMA and CPMA with the product processor and min processor as well as two contrastive ULAs in terms of the beam pattern, array gain and speech DOA estimation accuracy. A new approach based on SRP-PHAT is adapted for these processors. Simulation results indicate that the min processor leads to better beam patterns and array gains than the product processor, and the SCPMA further cancels the side lobe in the CPMA's beam pattern by having smaller PSL level and total side lobe area. Compared with the ULA using same number of microphones, the SCPMA significantly increases the operating frequency and has a beam pattern without grating lobes, leading to advantages in accurately recording high frequency components of speech signals. This is potentially beneficial for applications such as source separation and speech enhancement based on time-frequency DOA estimation. In addition, the min-processed SCPMA possesses the largest array gain among the discussed microphone arrays using the same number of microphones. Speech DOA estimates show that the SCPMA processed by both processors achieve accurate results, having equivalent accuracy with the CPMA with both processors, while the SCPMA shows less accurate DOA estimates under higher reverberation. The reason may lie in that the side lobes in the SCPMA's beam pattern causing amplification of the source reflections. Overall, the geometry of SCPMA shows positive potentials and is worth investigating.

Future work will focus on further cancelling the side lobes in the SCPMA's beam pattern to obtain more accurate DOA estimates. Additionally, the design of frequency-invariant beam patterns will also be researched.

REFERENCES

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

[2] Brandstein, M., and D. Ward, *Microphone Arrays: Signal Processing Techniques and Applications*, Springer-Verlag: Berlin, 2001.

[3] S. Pasha, C. Ritz and Y. Zou, "Detecting Multiple, Simultaneous Talkers through Localising Speech Recorded by Ad-hoc Microphone Arrays," *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC 2016)*, December 2016.

[4] J. Batke and H. Hake, "Design Aspects for an Improved B-format Microphone", *European Signal Processing Conference (EUSIPCO 2009)*, pp. 2554-2558, August 2009.

[5] Y. Haneda, K. Furuya, S. Koyama and K. Niwa, "Close-talking Spherical Microphone Array Using Sound Pressure Interpolation Based on Spherical Harmonic Expansion", *IEEE International Conference on Acoustics, Speech and Signal Processing (ICCASP 2014)*, pp. 604-608, May 2014.

[6] P. P. Vaidyanathan and P. Pal, "Sparse Sensing with Coprime Arrays," *the Forty-Fourth Asilomar Conference on Signals, Systems and Computers*, pp. 1405-1409, 2010.

[7] D. Bush and N. Xiang, "Broadband Implementation of Coprime Linear Microphone Arrays for Direction of Arrival Estimation," *Journal of the Acoustical Society of America*, vol. 138, issue 1, pp. 447-456, July 2015.

[8] J. Zhao and C. Ritz, "Investigating Co-Prime Microphone Arrays for Speech Direction of Arrival Estimation," *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC 2018)*, pp. 1658-1664, November 2018.

[9] K. Adhikari, "Beamforming with Semi-Coprime Arrays," *The Journal of the Acoustical Society of America*, vol. 145, issue 5, pp. 2841-2850, May 2019.

[10] Y. Liu and J. R. Buck, "Gaussian Source Detection and Spatial Spectral Estimation Using a Coprime Sensor Array With the Min Processor," *IEEE Transactions on Signal Processing*, vol. 66, no. 1, January 2018.

[11] N. Xiang and D. Bush, "Experimental Validation of a Coprime Linear Microphone Array for High-resolution Direction-of-arrival Measurements," *The Journal of the Acoustical Society of America*, vol. 137, issue 4, April 2015.

[12] H. Cox, R. Zeskind and M. Owen, "Robust Adaptive Beamforming," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 35, issue 10, pp. 1365-1376, 1987.

[13] J. Zhao and C. Ritz, "Co-Prime Circular Microphone Arrays and Their Application to Direction of Arrival Estimation of Speech Sources," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICCASP 2019)*, pp. 800-804, May 2019.

[14] J. H. DiBiase, "A High-accuracy, Low-latency Technique for Talker Localization in Reverberant Environments Using Microphone Arrays," *Brown University*, 2000.

[15] M. I. Mandel, R. J. Weiss and D. P. W. Ellis, "Model-based Expectation-maximization Source Separation and Localization," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, issue 2, pp. 382-394, 2010.

[16] S. Araki, H. Sawada, R. Mukai and S. Makino, "Underdetermined Blind Sparse Source Separation for Arbitrarily Arranged Multiple Sensors," *Signal Processing*, vol. 87, pp. 1833-1847, 2007.

[17] X. Zheng, C. Ritz and J. Xi, "Encoding and Communicating Navigable Speech Soundfields," *Multimedia Tools and Applications*, vol. 75, pp. 5183-5204, 2016.

[18] J. Allen and D. Berkley, "Image Method for Efficiently Simulating Small-room Acoustics," *The Journal of the Acoustical Society of America*, vol. 65, issue 4, pp. 943-950, April 1979.

[19] IEEE subcommittee on subjective measurements, "IEEE Recommended Practices for Speech Quality Measurements," IEEE Transactions on Audio and Electroacoustics, vol. 17, pp. 227-46, 1969.