Investigating Co-Prime Microphone Arrays for Speech Direction of Arrival Estimation

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 investigates the application of the steered response power - phase transform (SRP-PHAT) method to co-prime microphone array (CPMA) recordings to estimate the direction of arrival (DOA) of speech sources. While existing CPMA approaches for acoustics applications are limited, especially under reverberant conditions, the proposed algorithm utilises SRP-PHAT to estimate the DOA of speech sources and then employs a histogram-based stochastic algorithm using steered response power (SRP) adjustment and kernel density evaluation (KDE) to improve the DOA estimation accuracy. Experiments are conducted for up to three simultaneous speech sources in the far field considering both anechoic and reverberant scenarios. Results suggest that the proposed approach achieves more accurate DOA estimates than a uniform linear array (ULA) with the same number of microphones under both anechoic and low reverberant conditions, and it significantly decreases the number of microphones of another equivalent ULA while maintaining similar performances. Moreover, the operating frequency of the microphone array is largely increased without changing the number of microphones, making it possible to accurately record higher-frequency components of source signals.**

I. Introduction

Speech direction of arrival (DOA) is a conventional topic in the field of acoustic signal processing and plays a crucial role in a wide variety of real-world applications, such as teleconferencing systems [1], smartphones [2], [3] and robotic systems [4], [5]. In 2010, co-prime arrays were proposed as a sparse sensing method [6], which can achieve sharper beams using relatively few elements, exceeding the normal limit imposed by the spatial Nyquist sampling theorem. For the application of a co-prime microphone array (CPMA) in broadband DOA estimation, experimental results show that co-prime beamforming is viable and, under some circumstances, preferable [7], and then the method is extended by using a model-based Bayesian framework trying to determine the number of sound sources [8]. Wideband DOA estimation using a CPMA is also investigated by developing an algorithm based on group sparsity to lower the computational complexity [9].

However, the existing methods present limited research on acoustic environments, particularly when considering speech signals and reverberations that can distort the received signals and can degrade the performance of the broadband DOA estimation approaches using CPMAs.

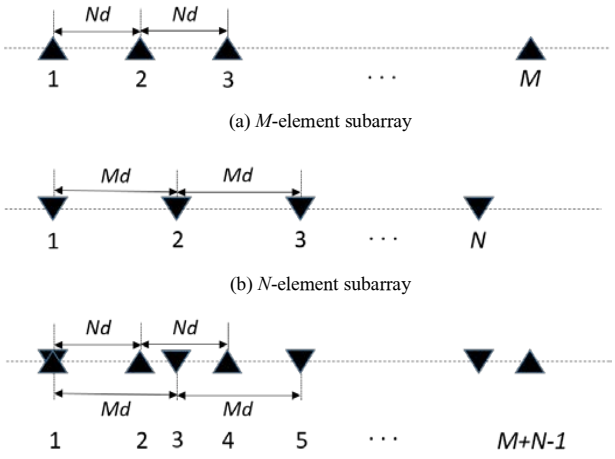
Steered response power - phase transform (SRP-PHAT) was proposed and could be seen as an extension of the steered response power (SRP) approaches [10]. It has been shown to be more robust under conditions with high noise and reverberation than other DOA estimation and source localisation algorithms, which are mostly based on time difference of arrival (TDOA) or spectral estimation [10], [11]. There are also a number of proposed improvements of the SRP-PHAT algorithm, such as the stochastic region contraction approach for multiple source localisation [12] and the scalable spatial sampling method to promote the robustness when locating the sound sources [13].

This paper shows that SRP-PHAT can be applied to the CPMA to achieve accurate speech DOA estimation under the effects of reverberation. The proposed method in this paper estimates the DOA of speech signals in both anechoic and reverberant scenarios, which provides an advantage over state-of-the-art research which do not take reverberation into account [8], [9]. By comparing the performances between the CPMA and the uniform linear array (ULA), the proposed method using a 16-element CPMA achieves better results than a ULA with the same number of microphones. It also obtains equivalent accuracy for speech DOA estimation than another ULA with a much larger number of microphone capsules, which can save a great deal of cost in real-world applications. Moreover, the spatial Nyquist frequency can be increased significantly without increasing the number of microphones so that higher-frequency components of the sources can be recorded more accurately than the conventional ULA. This indicates a potential benefit to source separation and speech enhancement algorithms based on clustering time-frequency DOA estimates [14], [15].

The remainder of this paper is organised as follows. In Section II, the mathematical model for co-prime microphone array recordings is introduced. Section III describes the proposed method of fitting SRP-PHAT to CPMAs on speech DOA estimation. Section IV demonstrates a stochastic algorithm based on histograms utilising SRP adjustment and kernel density modelling, which aims at boosting the DOA estimation accuracy. In Section V, experiments of the proposed approach in multiple scenarios are performed, and then the results and evaluations are given. The paper is concluded in Section VI.

II. Mathematical Model for the Co-prime Microphone Array Recording

A CPMA is composed of two uniform linear microphone subarrays, which can be illustrated as in Fig. 1. The numbers



(a) *M*-element subarray

(b) *N*-element subarray

(c) *M*·*N*-element co-prime microphone array

Fig. 1 An illustration of the geometry of co-prime microphone arrays.

of elements of the subarrays, *M* and *N*, are a pair of co-prime numbers, which mathematically mean that the only positive integer that divides both is 1. Assuming that there are *Q* uncorrelated far-field speech sources (which can be narrowband or wideband in theory) impinging on the CPMA from different DOAs θ_i ($i = 1, 2, \dots, Q$), and the number of the elements is *n*. So the received signal vector $\mathbf{y}(t)$ can be expressed as

$$\mathbf{y}(t) = \boldsymbol{\alpha}\mathbf{s}(t) * \mathbf{r}(t) + \mathbf{v}(t) \tag{1}$$

where $\mathbf{y}(t) = [y_1(t), \dots, y_n(t)]^T$, and $\boldsymbol{\alpha} = [\boldsymbol{\alpha}(\theta_1), \dots, \boldsymbol{\alpha}(\theta_Q)]$ is the attenuation factors due to propagation effects. Also, $\mathbf{s}(t) = [s_1(t), \dots, s_Q(t)]^T$, $\mathbf{r}(t) = [r_1(t), \dots, r_n(t)]^T$ and $\mathbf{v}(t) = [v_1(t), \dots, v_n(t)]^T$ represent the source signals, the reverberation from the environment and the additive noise signals, respectively. Normally, the sources $\mathbf{s}(t)$ are seen as uncorrelated, and the noise $\mathbf{v}(t)$ is also assumed to be uncorrelated zero mean noise with similar power levels at each element.

According to the spatial Nyquist sampling theorem, the spacing of any two elements of the microphone array should satisfy $d \leq \lambda / 2$, where d denotes the spacing and λ is the wavelength of the received signal, otherwise the beampattern of the microphone array will suffer from spatial aliasing. As a result, the design frequency of a ULA with *D* microphones can be defined as

$$f_{op\_ULA} = \frac{c}{2d} = \frac{cD}{2L_U} \tag{2}$$

where L_U is the aperture of the ULA. For CPMAs, it has been derived that the design frequency or the operating frequency of a CPMA can be calculated as [16]

$$f_{op\_CPMA} = \frac{cMN}{2L_C} \tag{3}$$

where L_C is the CPMA's aperture, which takes a virtual source into account, being larger than its physical length [16].

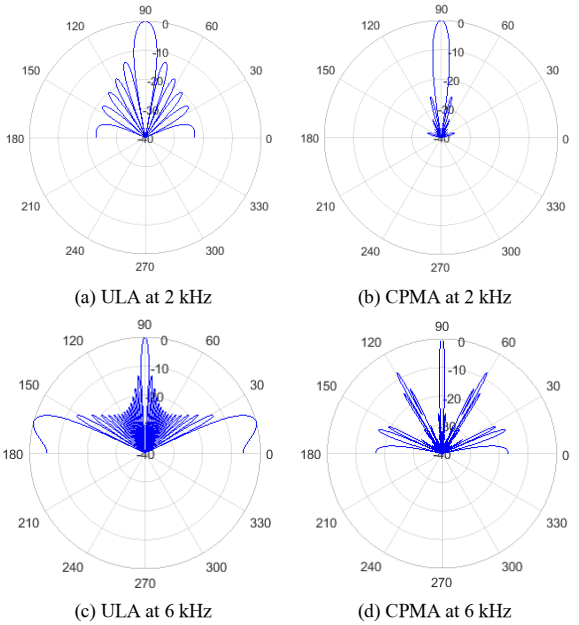
Comparing (3) with (2), it can be found that *D* is replaced by *M*·*N*, so an *M*·*N*-element CPMA can be seen equivalent to a *D*-element ULA, thus allowing the CPMA to receive higher frequency components of speech signals with fewer elements than the ULA and exceeding the separation limit determined by the spatial Nyquist sampling theorem.

To describe the reason for CPMAs to achieve much higher freedom than ULAs with the same number of microphones, the fundamental theory of beamforming is also investigated. The beampatterns of a conventional ULA (B_U) and a co-prime microphone array (B_C) can be expressed as (4) and (5), respectively.

$$B_U = \sum_{k=0}^{D-1} r(k) e^{j\alpha k} \tag{4}$$

$$B_C = B_M \times B_N^* \tag{5}$$

where $r(k)$ $(k = 0, 1, \dots, D-1)$ is the received signal by each microphone and $\alpha = 2\pi d sin\theta / \lambda$ with θ being the azimuth and $j = \sqrt{-1}$. B_M and B_N are the beampatterns of the two subarrays of the CPMA, and $*$ denotes the conjugate operation. It can be seen that the beampattern is a dependent variable of the wavelength of the signal, thus being dependent with the variation of the frequency of the signal. The shapes of the beampatterns of a conventional ULA and a CPMA, at 2 kHz and 6 kHz separately, can be illustrated as in Fig. 2. The 2 kHz is a representative frequency that is lower than the operation frequency of both microphone arrays. In (a) and (b) of Fig. 2, it can be seen that the side lobes of the CPMA are cancelled to



(a) ULA at 2 kHz

(b) CPMA at 2 kHz

(c) ULA at 6 kHz

(d) CPMA at 6 kHz

Fig. 2 Comparison of beampatterns for ULA and CPMA for 2 kHz and 6 kHz source frequencies

a large extent by staggering the two linear subarrays, and the main lobe is also narrower than the ULA, which means that the directionality is better at the desired steering angle. Moreover, the 6 kHz is also a selected frequency, which is above the operating frequency of the ULA and below that of the CPMA. In (c) and (d) of Fig. 2, it is shown that the beampattern of the ULA shows two grating lobes, which means the signals received from these two directions will be unreasonably amplified, while the beampattern for the CPMA has big side lobes, but they are not as big as the main lobe.

III. SRP-PHAT Based DOA Estimation Applied to Co-prime Microphone Arrays

For linear microphone arrays, it has been shown that the SRP can be obtained by summing the generalised cross-correlations (GCC) for all combinations of microphone pairs [10]. The GCC for one microphone pair can be calculated as

$$\tau_{y_1y_2}(p) = F^{-1}\left[\emptyset_{y_1y_2}(f)\right] \tag{6}$$

where y_1 and y_2 represent the outputs of the two microphones. $F^{-1}[\cdot]$ denotes the inverse discrete-time Fourier transform (IDTFT) and f stands for the variable in the frequency domain. Equation (6) reaches its maximum when $\tau = p$.

$$\emptyset_{y_1y_2} = w(f)\varphi_{y_1y_2}(f) \tag{7}$$

is the generalised cross-spectrum, with $w(f)$ being a frequency-domain weighting function and $\varphi_{y_1y_2}(f)$ being the cross-spectrum which can be expressed as

$$\varphi_{y_1y_2}(f) = E[Y_1(f)Y_2^*(f)] \tag{8}$$

where $E[\cdot]$ calculates the mathematical expectation and "*" denotes the conjugate operation, with the microphone outputs $Y_n(f), (n = 1, 2)$ being the summation of all possible values at time instants k. The equation of $Y_n(f)$ is shown as follows.

$$Y_n(f) = \sum_k y_n(k)e^{-j2\pi fk}, n = 1, 2 \tag{9}$$

Substituting IDTFT into (6), the GCC function can be derived as

$$\tau_{y_1y_2}(p) = \int_{-\infty}^{+\infty} w(f)\varphi_{y_1y_2}(f)\, e^{j2\pi fp} df \tag{10}$$

To determine the weighting function, phase transform (PHAT) has been proved to be a very effective one for TDOA estimation in reverberant scenarios [17], which is expressed as

$$w(f) = \frac{1}{|\varphi_{y_1y_2}(f)|} \tag{11}$$

In this case, only the information conveyed in the phase is taken into account, and then considering all possible microphone pairs in the summation operation, the equation of SRP-PHAT algorithm can be shown as

$$P(\tau) = \sum_{m=1}^{S}\sum_{l=m+1}^{S}\int_{-\infty}^{+\infty}\frac{\varphi_{y_1y_2}(f)}{|\varphi_{y_1y_2}(f)|}e^{j2\pi f\tau}df \tag{12}$$

where P is the SRP of the microphone array, and τ is the time delay of the sound propagation from the m^{th} microphone to the l^{th} microphone, taking the leftmost microphone as the first one ($m, l = 1, 2, \ldots, S$). For each scanning DOA θ ($0^\circ \le \theta \le 180^\circ$), the time delay in terms of samples can be formulated as

$$\tau = \frac{|d_{ml}|\cdot F_s\cdot\cos\theta}{c} \tag{13}$$

where $|d_{lm}|$ is the magnitude of the distance from the m^{th} microphone to the l^{th} microphone, and F_s and c are the sampling frequency and the speed of sound, respectively.

Having considered all the above deductions, the DOA estimation algorithm based on SRP-PHAT can be mathematically modelled as

$$\tau_{opt} = \underset{\tau}{\operatorname{argmax}}\left(\sum_{m=1}^{S}\sum_{l=m+1}^{S}\int_{-\infty}^{+\infty}\frac{\varphi_{y_1y_2}(f)}{|\varphi_{y_1y_2}(f)|}e^{j2\pi f\tau}\,df\right) \tag{14}$$

where the τ_{opt} represents the time lag that leads to the largest value of the SRP, and then the estimated DOA is

$$\theta_{est} = arccos\left(\frac{c\cdot\tau_{opt}}{|d_{ml}|\cdot F_s}\right) \tag{15}$$

If the CPMA is given, the distance between any two of the microphones is known, so the DOA estimates can be achieved.

IV. Histogram-based Stochastic DOA Estimation Algorithm Using SRP Adjustment and Kernel Density Modelling

After achieving all the DOA estimates, a histogram can be formed from the number of estimates for each angle in the steering range. However, in reverberant and multisource scenarios the microphones will receive many signals from other directions in addition to the direct source path, which can result in a spreading in the histogram and poses a negative influence on estimating true DOAs. To solve this problem, this section considers a power-adjusted histogram that is then modelled using kernel density estimation (KDE) to locate peaks corresponding to the estimated DOA.

A. SRP Adjusted DOA Histogram

An energy-weighted histogram has been proposed [14], which considers the energy of the time-frequency instants when analysing the histogram (similar to other weighting approaches [18], [19]). The DOA estimates with low energy will have insignificant contributions to the energy weighted DOA histogram. A similar approach is used here whereby the SRP value is used to adjust the histogram, referred to as the SRP-adjusted histogram (SAH) and described as

$$sahist(\theta_m) = \begin{cases} hist(\theta_m) - 1, & P(\theta_m) < T \\ hist(\theta_m), & P(\theta_m) \ge T \end{cases} \tag{16}$$

where θ_m ($0^\circ \le \theta_m \le 180^\circ$) represents each possible DOA under concerned, $hist(\theta_m)$ is the original DOA histogram and $sahist(\theta_m)$ is the value of the histogram bin at θ_m after considering the SRP for DOA θ_m, $P(\theta_m)$ that is determined from (12). T is a pre-defined threshold, which is the key point of the method. If the threshold is too small, the SAH method will not make a large difference, while if it is too big, most of the values in $hist$ will be cancelled, which is not as expected. According to informal experiments, one third of the subtraction of the maximum and the minimum of all the SRPs is found to be a good threshold to allow the peaks in the histogram to be more distinguishable, thus leading to reliable DOA estimates.

B. Kernel Density Estimation (KDE) Modelling of the SAH

The SAH approach enables the histogram to focus on the most representative contributors in terms of energy, but due to the influence of reverberation and multiple sources, there can be a few discrete angles possessing higher energy than the true sources. In addition, if there is too much spreading in the histogram, the energy of all DOAs can be similar, which shows difficulty in finding the peaks of the histogram. Therefore, to improve the accuracy and reliability of DOA estimation, a stochastic algorithm based on kernel density estimation (KDE) [20] is applied to obtain the probability density function (PDF) of the SAH, and then the DOAs can be achieved by searching for the local maximum of the PDF.

By utilizing KDE, the DOA histograms are smoothened using a suitable kernel function, and the discrete histograms is transformed into continuous ones. In this way, the probabilities of all DOA estimates are taken into account so that the DOA estimation method is more reasonable, weakening the results of the discrete high-energy scanning points and accumulating the effects of close histogram bins to find a local peak. The density function can be achieved as [14]

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\frac{x-x_i}{h} \tag{17}$$

where K is a selected kernel function, $h\ (h > 0)$ is the bandwidth which is a smoothing parameter, and $x_i (i = 1,2,\dots,n)$ is an independent and evenly distributed sample, the distribution of which is f. In this paper, a Gaussian kernel is used, but in fact, the impact of the kernel function is much less than the selected bandwidth. The bandwidth is not a fixed value but should be chosen such that the correct peaks in the PDF can be distinguished. The bandwidth of the Gaussian kernel is calculated as [21]

$$h_{opt}^{GAUSS} = 3\left(\frac{1}{70n\sqrt{\pi}}\right)^{1/5}\hat{\sigma} \tag{18}$$

where $\hat{\sigma}$ is the standard deviation of the observed samples. After choosing the appropriate bandwidth, the DOA estimation can be obtained by searching for peaks in the SAH by analysing the first derivative and the second derivative of the PDF as

$$\theta_{est\_KDE} = \theta_m, if\ f'_{est}(\theta_m) = 0, and\ f''_{est}(\theta_m) < 0 \tag{19}$$

TABLE I
THE PROPOSED DOA ESTIMATION METHOD

1)	Start with the recorded speech signals by a co-prime microphone array, which is $\mathbf{y}(t)$ from (1).
2)	Convert the signals to the short-term frequency domain.
3)	Apply SRP-PHAT to the co-prime microphone array to achieve DOA estimates as shown in (14) and (15), and their corresponding SRP values are also obtained using (12), thus gaining time-frequency DOA estimation θ_{est} and the power $P(\tau)$.
4)	Derive the SRP-adjusted histogram according to (16).
5)	Apply the KDE method to the histogram, resulting in a continuous distribution and then searching for the final DOA estimation results θ_{est_final} from (19).

Having considered all the aforementioned theories, the proposed DOA estimation method is summarised in Table I.

V. EXPERIMENTS AND RESULTS

In this section, different scenarios in terms of levels of reverberation and numbers of sources are simulated and the proposed speech DOA estimation approach as shown in Table I is applied to investigate the influences of reverberation and configurations of microphone arrays on the performance of speech DOA estimation.

A. Experiment Conduction

Speech utterances from the IEEE-standard corpus [22] and the NOIZEUS speech corpus (clear sources) [23] are utilised to simulate scenarios with 1 to 3 speakers talking simultaneously. As shown in Table II, a co-prime linear microphone array and two contrastive ULAs with the same physical length are placed at exactly the same position separately in each experiment. All sources are designed to be of the same distance from the centre of the microphone array and are located in 3 fixed positions (S_1, S_2 and S_3) in the far field as illustrated in Fig. 3. Sources are rotated over 3 positions resulting in a total of 9 different trials for single source cases and 3 trials for scenarios of multiple sources, over which the average error of DOA estimation is calculated in terms of root mean square error (RMSE), which is defined as

$$\text{RMSE} = \sqrt{\frac{1}{H}\sum_{j=1}^{H}(\theta_{est\_final,j} - \theta_t)^2} \tag{20}$$

where H is the number of estimates and θ_t is the true DOA (unknown before estimating), respectively.

All the experiments are completed in Matlab R2016b. It is assumed that the aforementioned settings are placed in a room, and the source signals can be degraded by reverberation, which is simulated using the IMAGE algorithm [24]. All of the configurations and parameters of the experiments are demonstrated in Table III. The physical length of the CPMA and the ULAs are equal to $0.89m$. Time-domain signals are converted to the short-term frequency domain using the Fast Fourier Transform, which is applied to 25 ms Hamming-windowed frames. DOA estimates of multiple frames are then

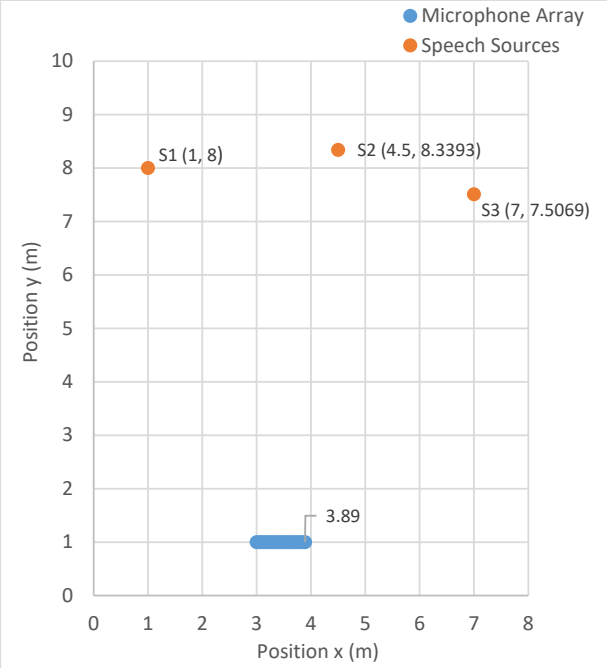


Fig. 3 Recording Configurations in two dimensions (maintaining z = 2m)

achieved by applying the SRP-PHAT algorithm to the CPMA and the ULAs. A histogram of the estimated DOAs is then analysed to find the DOA that achieves the highest local probability, which is used as the estimated source DOA [14].

For the selection of the bandwidth of the KDE method, there is a built-in algorithm to calculate the optimal value for Gaussian kernel in Matlab as shown in (18), which is utilised directly in the experiments except for the three-source scenarios. When there are three simultaneous sources, it is found that the optimal bandwidth determined by Matlab does not result in three distinct peaks, and instead a value of 4 is found to be a suitable bandwidth in this case (informal experiments find other values ranging from 3 to 5 produce similar results).

B. Performance Analysis

Applying the configurations of Table II and Table III to the proposed method, the results are shown in Table IV.

Firstly, results show that the DOA estimation accuracy of the 16-element co-prime microphone array is generally higher than that of the 16-element ULA under both anechoic and lower reverberant conditions (RT_{60} = 200ms). For the higher reverberant condition (RT_{60} = 400ms), the 16-element CPMA performs slightly worse in the single and two source cases but much better in the 3-source case. It should be noted that for the 2-source case with 400 ms reverberation, all three microphone arrays perform quite poorly, which is deemed to be due to the DOA histograms often resulting in a large peak that is in between the peaks corresponding to direct source directions and is likely resulted from a large reflected source.

TABLE II
CHARACTERISTICS OF THE MICROPHONE ARRAYS

Type of array	Number of elements	Length (m)	f_{op} (Hz)
CPMA	16	0.89	**12348.00**
ULA	16	0.89	2890.45
ULA	72	0.89	**13681.46**

TABLE III
EXPERIMENTAL SETTINGS

Sampling frequency (f_s)	8 kHz
N	200
Frame length	25 ms
Frame overlap	50%
Number of frames concerned	180
Azimuth concerned	0° – 180°
Azimuth scanning resolution	0.01°
Reverberation time (RT_{60})	{0, 200, 400} ms
Room dimensions (x, y, z)	(8 m, 10 m, 5 m)
True DOAs (S_1, S_2, S_3)	{109.3°, 81.9°, 62.9°}
Source-array distance	7.4 m
Speed of sound (c)	343 m/s

Secondly, another comparison is made between the 16-element CPMA and the 72-element ULA. Within the 9 experiments, the 16-element CPMA performs better than the 72-element ULA in 5 of the scenarios, with inferior accuracy in the other 4 cases. These can be regarded as equivalent in terms of their performances. An important result is that the 16-element CPMA can estimate all the three sources much more accurately than the 72-element ULA when the RT_{60} is 400ms.

C. Further Analysis of the advantage using KDE

Under high reverberation (RT_{60} = 400ms), one example of the DOA estimation of three sources utilising the proposed algorithm is shown in Fig. 4. It can be seen visually that if using the discrete histogram to estimate DOAs, the first three biggest contributors are all close to 90 degrees, which are not correct. When taking the probability distribution into account,

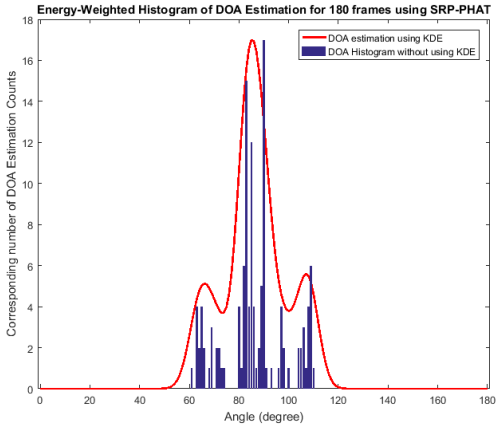


Fig.4 Comparison between DOA estimation with and without using KDE

TABLE IV
DOA ESTIMATION RESULTS MEASURED BY RMSE (UNIT: DEGREE)

Type of array	RT_{60} = 0 ms			RT_{60} = 200 ms			RT_{60} = 400 ms		
	Single Source	Two Sources (S1, S2)	Three Sources	Single Source	Two Sources (S1, S2)	Three Sources	Single Source	Two Sources (S1, S2)	Three Sources
16-element CPMA	1.02	**0.14**	**0.89**	**0.81**	0.25	**0.84**	2.87	12.05	**2.95**
16-element ULA	1.15	0.21	1.08	1.16	0.25	0.97	2.51	**11.79**	9.30
72-element ULA	**0.99**	0.19	1.01	0.99	**0.13**	0.92	**2.36**	11.82	9.21

the angles contributing lower energy but having concentrated contribution are paid much attention to, thus achieving more accurate DOA estimation results. So by applying KDE, all of the bins in the histogram are considered, rather than only finding the biggest ones. This strategy of calculation is more reasonable when estimating the DOA.

VI. CONCLUSIONS

This paper proposes a DOA estimation method using CPMA recordings of speech sources in the far field. The conventional SRP-PHAT algorithm is applied to the CPMA recordings and raw DOA estimates are obtained. The accuracy of these results is then enhanced by adjusting the DOA histogram based on SRP outputs and using KDE to obtain smooth histogram peaks, which are the final DOA estimates. The experimental evaluation is based on conditions including single source, two sources and three sources in both anechoic and reverberant environments. Results indicate that the proposed approach using a 16-element CPMA leads to accurate speech DOA estimation when the environment is anechoic or the reverberation is low, and its performance is better than that of a ULA with the same number of microphones. Moreover, the proposed algorithm achieves similar accuracy to that obtained by a ULA with a much larger number of 72 microphones. Compared with a ULA with the same number of microphones, the CPMA has the advantage of significantly increasing the operating frequency of the microphone array, enabling high frequency components of signals to be more accurately recorded. This offers a potential advantage for source separation and speech enhancement algorithms based on time-frequency DOA estimation. As the beampattern of a CPMA can have big side lobes in certain directions, future work will involve developing algorithms to deal with these side lobes. In addition, the performance of the proposed method under high reverberation and different levels of noise will be further improved in the future. The issues of source separation and speech enhancement using the proposed approach will also be investigated.

REFERENCES

[1] M. Togami, S. Suganuma, Y. Kawaguchi, T. Hashimoto, and Y. Obuchi, "Transient Noise Reduction Controlled by DOA Estimation for Video Conferencing System," *IEEE 13th International Symposium on Consumer Electronics*, pp. 26-29, 2009.

[2] D. Ayllón, H. A. Sánchez-Hevia, R. Gil-Pita, M. U. Manso, and M. R. Zurera, "Indoor Blind Localization of Smartphones by Means of Sensor Data Fusion," *IEEE Transactions on Instrumentation and Measurement*, vol. 65, issue 4, pp. 783-794, 2016.

[3] G. Li, X. Bao, and Z. Wang, "The Design and Implementation of A Smartphone-based Acoustic Array System for DOA Estimation," *36th Chinese Control Conference*, pp. 5416-5423, 2017.

[4] R. Levorato and E. Pagello, "DOA Acoustic Source Localization in Mobile Robot Sensor Networks," *IEEE International Conference on Autonomous Robot Systems and Competitions*, pp. 71-76, 2015.

[5] M. Togami, A. Amano, T. Sumiyoshi, and Y. Obuchi, "DOA Estimation Method Based on Sparseness of Speech Sources for Human Symbiotic Robots," *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 3693-3696, 2009.

[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] D. Bush and N. Xiang, "A Model-based Bayesian Framework for Sound Source Enumeration and Direction of Arrival Estimation Using A Coprime Microphone Array," *22nd International Congress on Acoustics*, 2016.

[9] Q. Shen, W. Liu, W. Cui, S. Wu, Y. D. Zhang, and M. G. Amin, "Low-complexity Direction-of-arrival Estimation Based on Wideband Co-prime Arrays," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, issue 9, pp. 1445-1456, 2015.

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

[11] S. T. Birchfield, "A Unifying Framework for Acoustic Localization," *12th European Signal Processing Conference*, pp. 1127-1130, 2004.

[12] H. Do and H. F. Silverman, "Stochastic Particle Filtering: A Fast SRP-PHAT Single Source Localization Algorithm," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 213-216, 2009.

[13] M. Cobos, A. Marti, and J. J. Lopez, "A Modified SRP-PHAT Functional for Robust Real-time Sound Source Localization with Scalable Spatial Sampling," *IEEE Signal Processing Letters*, vol. 18, issue 1, pp. 71-74, 2011.

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

[15] Y. X. Zou, W. Shi, B. Li, C. H. Ritz, M. Shujau, and J. Xi, "Multisource DOA Estimation Based on Time-frequency Sparsity and Joint Inter-sensor Data Ratio with Single Acoustic Vector Sensor," *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 4011-4015, 2013.

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

[17] C. H. Knapp and G. C. Carter, "The Generalized Correlation Method for Estimation of Time Delay," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, issue 4, pp. 320-327, 1976.

[18] 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.

[19] 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.

[20] B.W. Silverman, "Density Estimation for Statistics and Data Analysis," *Monographs on Statistics and Applied Probability*, London: Chapman and Hall, 1986.

[21] G. R. Terrell, "The Maximal Smoothing Principle in Density Estimation," *Journal of the American Statistical Association*, vol. 85, no. 410, pp. 470–477, Jun. 1990.

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

[23] Y. Hu and P. Loizou, "Subjective Evaluation and Comparison of Speech Enhancement Algorithms," *Speech Communication*, vol.49, pp. 588-601, 2007.

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