IMPLEMENTATION STRATEGIES FOR ARRAY LOCALISATION ALGORITHMS BASED ON THE SRP-PHAT

Anders M. Johansson

Blekinge Tekniska Högskola, Ronneby, Sweden (e-mail: ajh@bth.se).

ABSTRACT

The steered response power - phase transform (SRP-PHAT) algorithm has become an ubiquitous tool in signal source localisation using sensor arrays. This paper investigates four different implementation strategies for the SRP-PHAT algorithm. In this context, implementation using shaped response interpolation versus plain wideband position estimation is investigated, as are two different data averaging strategies. Each of the four combinations yields a different computational complexity and localisation accuracy. Simulations are used to compare the localisation accuracy and the computational requirements are calculated based on average number of floating point operations per position estimate.

1. INTRODUCTION

Signal source positioning of wideband signal sources using sensor arrays has become a common problem encountered in a wide variety of applications, including radar [1], radio [2], sonar [3] and acoustics [4]. A common algorithm used in these applications is the steered response power - phase transform (SRP-PHAT) algorithm. The algorithm is an extension of the generalised cross correlation (GCC) algorithm pioneered by Knapp and Carter in 1976 [5]. It was later extended by Rabinkin in 1996 [6] and reached its current form in the works of Johansson in 2002 [7] and 2005 [4].

The SRP-PHAT algorithm produces a source position estimate based on the time delay difference of arrival of a wavefront across two or more sensor elements. The algorithm operates by extracting the phase difference between the microphone signals from an estimate of their cross-spectral density (CSD) function. There exist two possible approaches for reducing the uncertainty of the position estimate. The first is to average the CSD over time (referred to as CSD averaging CA), and the second is to use instantaneous CSD estimates and average the resulting position estimates (referred to as post estimation averaging PA). These averaging approaches result in two different implementation strategies for the algorithm.

The SRP-PHAT algorithm can operate both on wideband (WB) and on narrowband (NB) sensor signals. However, it is possible to apply shaped response interpolation (SRI) [8] to project the CSD of wideband signals onto a single frequency band, which makes it possible to use the narrowband version of the SRP-PHAT to estimate the location of the a wideband signal source. This results in two additional implementation approaches for the algorithm.

Given the two averaging strategies and the two implementation approaches for wideband sensor data, a total of four different implementations are possible. Each of these implementations yields a unique computational complexity and localisation accuracy.

The paper is organised as follows: Section 2 contains a description of the signal model. Section 3 presents a brief review of the fundamental SRP-PHAT and SRI algorithms and description of the different implementations. The computational requirements are investigated in Section 4, followed by an evaluation of the performance in Section 5. The paper is concluded in Section 6

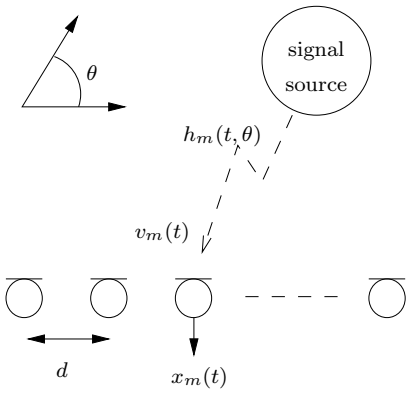

Fig. 1. Signal model.

2. SIGNAL MODEL

To limit the extent of the theory and evaluation sections, the discussion is limited to far-field localisation using linear arrays with uniformly spaced omni directional sensor elements, indexed $m \in \{1, 2, \ldots, M\}$. Consider a two dimensional Cartesian coordinate system with origin placed at the centre of the array, and a sensor spacing of d meters. Furthermore, an isotropic background noise field uncorrelated to any source signals is assumed, where the noise signal at each sensor element is denoted $v_m(t)$. The source signal is a spectrally white wideband source denoted $s(t)$, which originates from a location at an angle θ in the far-field of the array. The source signal is convolved by a set of linear time invariant transfer functions denoted $h_m(t, \theta)$, $m \in \{1, 2, \ldots, M\}$. The resulting sensor signals are denoted $x_m(t)$, and are defined as

$$x_m(t) = s(t) * h_m(t, \theta) + v_m(t) \quad m \in \{1, 2, \ldots, M\}, \quad (1)$$

where $*$ denotes convolution. The scenario is depicted in Fig. 1.

The source signals are band-limited and sampled with a sample period T, $x_m(\ell) = x_m(t \cdot T)$.

Assuming a direct line of sight between the source and the sensors, the transfer functions $h_m(t, \theta)$ can be modelled

as $h_m(t,\theta) = A_m\delta(t-\Delta_m) + g_m(t)$, where $\delta(t)$ is a Dirac, A_m is an amplitude function, Δ_m is bulk delay in seconds and $g_m(t)$ is the transfer function minus the direct path. When combined with the fact that the the source is in the far-field, the wavefront will arrive at the sensors as a distorted plane wave, where the time delay difference of arrival in samples between the sensor elements m and n is

$$\begin{aligned}\tau_{m,n}(\theta) &= (\Delta_m - \Delta_n)/T \\ &\simeq \frac{(m-n)\cos\theta}{c\cdot T},\end{aligned} \tag{2}$$

where c is the propagation speed of the wave. Please note that $\tau_{m,n}(\theta)$ contains the angle of arrival for the wave front, which is our parameter of interest.

3. THE SRP-PHAT ALGORITHM

Given the above signal model, the CSD between sensor elements m and n for the frequency $\omega_k, k \in \{1,2,\ldots,K\}$, is defined as

$$\begin{aligned}\Phi_{\mathrm{X}m,n}(\omega_k) &= \mathcal{F}\left\{\mathbb{E}\left[x_m(\ell)x_n(\ell+\rho)\right]\right\} \\ &\simeq P_{m,n}e^{j\omega_k\tau_{m,n}(\theta)} + \Phi_{\mathrm{G}m,n}(\omega_k) + \Phi_{\mathrm{V}m,n}(\omega_k),\end{aligned} \tag{3}$$

where $\mathcal{F}\{\cdot\}$ and $\mathbb{E}[\cdot]$ denotes the Fourier transform and the expected value, respectively. The power of the direct wave is denoted $P_{m,n}$, and $\Phi_{\mathrm{V}m,n}(\omega_k)$ and $\Phi_{\mathrm{G}m,n}(\omega_k)$ are CSDs of the background noise and the transfer function minus the direct path, respectively.

The CSD between two sensor elements can be estimated from the sensor signals, by splitting up the signals into K element segments

$$\widehat{\Phi}_{\mathrm{X}m,n}(\omega_k) = \sum_{\ell=1}^{L} X_\ell(\omega_k,m)X_\ell^*(\omega_k,n), \tag{4}$$

where $X_\ell(\omega_k,m)$ is the kth FFT bin of sensor m and signal segment ℓ, and $()^*$ denotes complex conjugate.

The CSD matrix $\boldsymbol{\Phi}_{\mathrm{X}}(\omega_k)$ is defined by gathering the CSD for each microphone pair into a matrix as $[\boldsymbol{\Phi}_{\mathrm{X}}(\omega_k)]_{m,n} = \Phi_{\mathrm{X}m,n}(\omega_k)$.

3.1. The PHAT transform

The PHAT transform is originally an ad-hoc method to improve phase function estimates. It is used here based on the assumption that the power of the sensor signals does not carry useful information about the source location and is therefore discarded. The PHAT transform yields the relative phase function $\psi_{m,n}(\omega_k)$ and is defined as

$$\psi_{m,n}(\omega_k) = \frac{\widehat{\Phi}_{\mathrm{X}m,n}}{\left|\widehat{\Phi}_{\mathrm{X}m,n}(\omega_k)\right|}. \tag{5}$$

Note that in a noise free scenario ($\Phi_{\mathrm{V}} \equiv 0$) and free space propagation ($\Phi_{\mathrm{G}} \equiv 0$), the relative phase function simplifies to $\psi_{m,n}(\omega_k) = e^{j\omega_k\tau_{m,n}(\theta)}$.

3.2. The SRP algorithm

The shaped response function for a single frequency band is defined as

$$P(\omega_k,\nu) = \sum_{m=1}^{M}\sum_{n=1}^{M}\psi_{m,n}(\omega_k)e^{j\omega_k\tau_{m,n}(\nu)}. \tag{6}$$

Thus, in a noise free scenario and free space propagation, the main peak of the function is found at $\nu \equiv \theta$. This peak will become increasingly distorted as the noise increases and as the tail of the transfer function grows.

Using the steered response function the narrowband SRP-PHAT algorithm is defined as the optimisation problem

$$\hat{\theta}_{\mathrm{NB}} = \arg\max_{\nu} P(\omega_0,\nu), \tag{7}$$

where $\hat{\theta}_{\mathrm{N}}$ is an estimate of the angle of arrival of the wave front. The wideband version of the algorithm simply performs the optimisation across all frequency bands as

$$\hat{\theta}_{\mathrm{WB}} = \arg\max_{\nu} \sum_{k=1}^{K} P(\omega_k,\nu). \tag{8}$$

In both cases the optimisation is performed over one variable and can be solved using for example the golden search algorithm.

3.3. Shaped response Interpolation (SRI)

SRI works by designing a set of projection matrices $\boldsymbol{T}(\omega_0,\omega_k), k \in \{1,2,\ldots,K\}$ that are used to project the CSD matrix for every frequency band onto a single frequency ω_0 according to

$$\boldsymbol{\Phi}_{\mathrm{X}}(\omega_0) = \sum_{k=1}^{K}\boldsymbol{T}(\omega_0,\omega_k)\boldsymbol{\Phi}_{\mathrm{X}}(\omega_k)\boldsymbol{T}^T(\omega_0,\omega_k). \tag{9}$$

The narrowband version of the SRP-PHAT can subsequently be applied to $\boldsymbol{\Phi}_{\mathrm{X}}(\omega_0)$ to locate the signal source.

3.4. Averaging approaches

As stated in the introduction it is possible to perform averaging either on the CSD matrix (CA) or on the location estimate (PA). When PA is used, the number of signal segments used to estimate the CSD matrix yields $L \equiv 1$. This can be used to rewrite (7) by insertion of (2), (4) and (5), as

$$P_\ell(\omega_k,\nu) = \left|\sum_{m=1}^{M}\frac{X_\ell(\omega_k,m)e^{\frac{j\omega_k m\cos(\nu)}{cT}}}{|X_\ell(\omega_k,m)|}\right|^2. \tag{10}$$

This expression requires far fewer calculations to evaluate compared to (7).

When PA is used in conjunction with SRI, it is possible to save calculations by rewriting (9) according to

$$\boldsymbol{\Phi}_{\mathrm{X}}(\omega_0) = \sum_{k=1}^{K}\boldsymbol{B}(\omega_0,\omega_k)\boldsymbol{B}(\omega_0,\omega_k)^H. \tag{11}$$

where $()^H$ denotes Hermitian transpose and $\boldsymbol{B}_\ell(\omega_0,\omega_k) = \boldsymbol{T}(\omega_0,\omega_k)\boldsymbol{X}_\ell(\omega_k)$ in which

$$\boldsymbol{X}_\ell(\omega_k) = \begin{bmatrix}X_\ell(\omega_k,1) & X_\ell(\omega_k,2) & \cdots X_\ell(\omega_k,M)\end{bmatrix}^T. \tag{12}$$

4. COMPUTATIONAL REQUIREMENTS

4.1. Optimisation algorithm

The optimisation algorithm employed here is the Golden search algorithm with parabolic interpolation, with a termination threshold of 10^{-4} samples. A Mone Carlo simulation of the different implementations indicated that the average number of objective function evaluations is 9.5 for the WB

Prior to optimisation				
Operation	WB-CA	WB-PA	NB-CA	NB-PA
Real divisions	KM^2	KM	M^2	M^2
Real additions	KM^2	KM	M^2	M^2
Real multiplications	$2KM^2$	$2KM$	$2M^2$	$2M^2$
Complex to real multiplications	KM^2	KM	$KM^{\mathbf{3}}$	$K(M^2+M)$
Complex additions	$2KM^2$		$3KM^2$	KM^2
Complex multiplications	KM^2		KM^2	
Real square root	KM^2	KM	M^2	M^2
During each step of the optimisation				
Operation	WB-CA	WB-PA	NB-CA	NB-PA
Real additions	K	K		
Real multiplications	KM	KM	M	M
Complex additions	KM^2	KM	M^2	M^2
Complex multiplications	$2KM^2$	$2KM$	$2M^2$	$2M^2$
Complex exponential	$2M$	$2M$	$2M$	$2M$

Table 1. Number of operations for each implementation.

algorithms and 9.2 for the NB algorithms. The simulations also showed that averaging method did not impact on the number of objective function evaluations.

The Golden search algorithm itself has very low computational requirements compared to the number of calculations required to evaluate the objective functions. The computations required to execute the Golden search algorithm are therefore excluded from the final results to improve clarity.

4.2. Objective functions

The following steps can be taken to reduce the number of calculations required to execute the different implementations:

1. In the WB algorithms, it is possible to rewrite the complex exponential according to
$$e^{j\omega_k \tau_{m,n}(\nu)} = e^{j\omega_{k-1}\tau_{m,n}(\nu)} \cdot e^{j\omega_1 \tau_{m,n}(\nu)},$$
as the frequency points ω_k are uniformly spaced. This can be used to speed up the WB algorithms significantly as the evaluation of the complex exponential requires a large number of floating point operations (FLOPs).
2. The PHAT and the SRI transformations can be performed in advance of the optimisation algorithm.
3. The computational cost of the complex exponential used in all implementations (except WB PA), can be reduced by using (2)
$$e^{j\omega_k \tau_{m,n}(\nu)} = e^{\frac{j\omega_k m \cos(\nu)}{cT}} \cdot e^{-\frac{j\omega_k n \cos(\nu)}{cT}}.$$
This further reduces the number of complex exponentials that must be calculated in the various algorithms from M^2 to M.
4. The computational complexity in calculating the average in (4) can be reduced using a sliding window implementation, according to
$$\widehat{\Phi}_{\mathrm{x}m,n,\ell}(\omega_k) = \widehat{\Phi}_{\mathrm{x}m,n,\ell-1}(\omega_k) + X_\ell(\omega_k, m)X^*_\ell(\omega_k, n) - X_{\ell-(L+1)}(\omega_k, m)X^*_{\ell-(L+1)}(\omega_k, n).$$

The number of operations required to evaluate the objective function for each of the implementations are presented in Table 1. The table is split in two, where the first part tabulates the calculations that can be made prior to the optimisation algorithm and the second part tabulates the calculations required for each evaluation of the objective function. The table shows that all implementation strategies exhibits linear growth of the computational complexity with respect to the number of samples per segment K. The table also shows that the growth of the computational complexity with regards to the number of sensors M, is linear for the WB-PA, quadratic for the WB-CA and the NB-PA, and cubic for the NB-CA implementations. Furthermore, the table shows that the computational complexity is in general quite large for both the CA implementations.

4.3. Computational load

The number of FLOPs per operation is displayed in Table 2. The values displayed in the table are taken from typical implementations of the different operations. The total number of FLOPs per position estimate (FLOPE) can be calculated by combining Tables 1 and 2. The result from the calculation for different values of M and K is displayed in Figs. 2 and 3, respectively. The figures show that the most computationally efficient implementation is the WB-PA if the number of sensors is large and the NB-PA if the number of sensors is small and the number of samples per segment is large.

Real divisions	16
Real additions	1
Real multiplications	1
Complex to real multiplications	2
Complex additions	2
Complex multiplications	6
Real square root	56
Complex exponential	270

Table 2. FLOPs per operation.

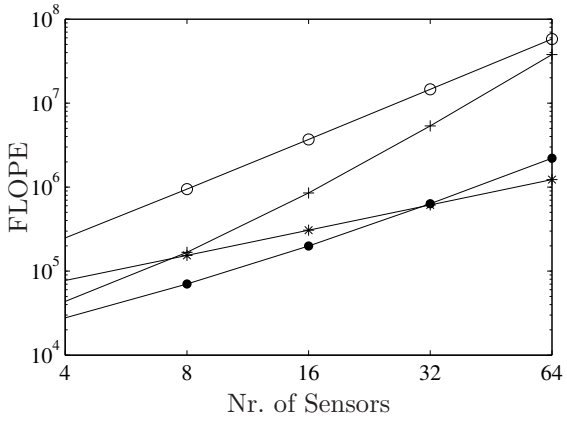

Fig. 2. Average number of floating point operations per position estimate versus the number of sensors for $K = 256$ samples per segment. Key: WB-CA (∘), WB-PA (∗), NB-CA (+), NB-PA (•)

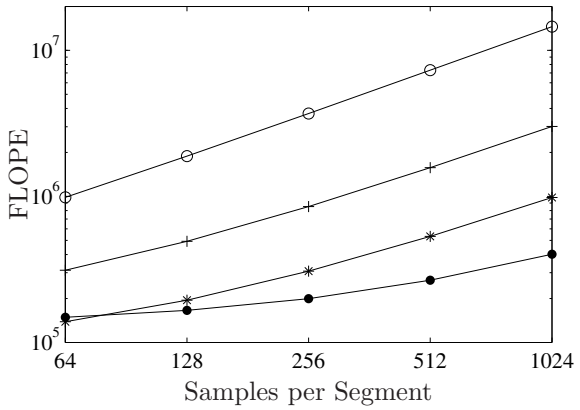

Fig. 3. Average number of floating point operations per position estimate versus the number of samples per segment for $M = 16$ sensors. Key: WB-CA (∘), WB-PA (∗), NB-CA (+), NB-PA (•)

5. PERFORMANCE EVALUATION

The four implementations are evaluated with respect to root mean square error (RMSE) between the estimated and actual source direction for different signal to noise ratios (SNR). The results are presented in Fig. 4, and show that the NB implementations perform slightly worse than the WB implementations of the algorithm. The figure also shows that the difference between averaging the CSD matrix compared to performing post averaging of the location estimates is close to negligible.

6. CONCLUSIONS

Four different implementations of the commonly used SRP-PHAT wideband sensor array localisation algorithm are presented. The theoretical description of the algorithm lists a number of methods for reducing the number of calculations required to execute the implementations. An analysis of the resulting computational complexity shows that if the number of sensors is large, the lowest computational load is given by the wideband post estimation averaging implementation of the algorithm. The implementation are also evaluated with respect to localisation accuracy. This evaluation shows that the difference between the CSD and post estimation averaging approaches is negligible and that the wideband implementations yields the lowest estimation error.

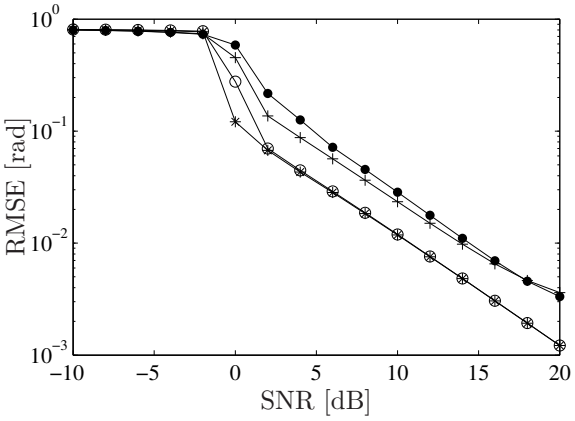

Fig. 4. Estimation accuracy RMSE vs SNR. Key: WB-CA (∘), WB-PA (∗), NB-CA (+), NB-PA (•)

7. REFERENCES

[1] B. J. Skinner, J. P. Donohoe, and F. M. Ingels. Matched FSK/PSK radar. In *Radar Conference*, pages 251–255. IEEE, March 1994.

[2] D. P. Young, C. M. Keller, D. W. Bliss, and K. W. Forsythe. Ultra-wideband (UWB) transmitter location using time difference of arrival (TDOA) techniques. *SSC*, 2:1225–1229, 9-12 Nov. 2003.

[3] S. A. L. Glegg, M. P. Olivieri, R. K. Coulson, and S. M. Smith. A Passive Sonar System Based on an Autonomous Underwater Vehicle. *IEEE Journal of Oceanic Engineering*, 26(4), October 2001.

[4] Anders Johansson and Sven Nordholm. Robust acoustic direction of arrival estimation using Root-SRP-PHAT, a realtime implementation. In *ICASSP*, March 2005.

[5] C. H. Knapp and G. C. Carter. The Generalized Correlation Method for Estimation of Time Delay. *IEEE Transactions on Acoustics Speech and Signal Processing*, 24(4):320–327, August 1976.

[6] D. Rabinkin, R. Renomeron, J. French, and J. Flanagan. Estimation of Wavefront Arrival Delay Using the Cross-Power Spectrum Phase Technique. In *132nd Meeting of the Acoustical Society of America*, volume 100, page 2697, Honolulu, USA, December 1996.

[7] Anders Johansson, Nedelko Grbić, and Sven Nordholm. Speaker localisation using the far-field SRP-PHAT in conference telephony. In *ICICS*, Kaohsiung, Taiwan ROC, 2002.

[8] G. J. Cook, B. K. Lau, and Y. H. Leung. An Alternative Approach to Interpolated Array Processing for Uniform Circular Arrays. In *APCCS*, volume 1, pages 411–414, Febryary 2002.