A Maximum Likelihood Cross Phase Spectra Estimator for Time Delay Estimation

Xi-Lin Li

Abstract—**Accurate cross phase spectra estimation is essential for time difference of arrival (TDOA) estimators that exclusively use phase information, e.g., generalized cross correlation with phase transform (GCC-PHAT). However, many GCC-PHAT implementations do not consider the phase bias caused by spatially correlated wide band noises in practice. This brings a considerable performance gap between GCC-PHAT and more complicated subspace methods like generalized eigenvalue decomposition multiple signal classification (GEVD-MUSIC). This paper proposes a maximum likelihood (ML) cross phase spectra estimator that can use the noise spatial complex coherence function, which can be predicted in theory or measured in advance, to compensate such phase bias. Experimental results with real world speech signal TDOA task demonstrate that the new ML cross phase spectra estimator can bring significant performance gain to GCC-PHAT with virtually no extra cost.**

Index Terms—**Time difference of arrival (TDOA), generalized cross correlation (GCC), phase transform (PHAT), coherence, phase estimation.**

I. Introduction

Time difference of arrival (TDOA) estimation is a fundamental problem in signal processing. Generalized cross correlation (GCC) and its special case, GCC with phase transform (GCC-PHAT) [1], [2], are still widely used for this task due to their simplicity and reasonably good performance. This is especially true for acoustic source direction of arrival (DOA) estimation and localization [4], where GCC-PHAT and its derived methods like steered response power PHAT (SRP-PHAT) [3] are predominant in real time systems. The cross phase spectra of observations used in GCC-PHAT serve as unbiased estimations of the ones of delayed source signals only for spatially uncorrelated noises. However, real world wide band noises tend to be spatiotemporally correlated, and the spatial correlation introduces undesirable biases to the phases used in GCC-PHAT for TDOA estimation. Although it is common to compensate the correlation of noises in subspace methods [5] by whitening the observations using estimated noise spatial correlation matrix [6], little open work is found on how to improve the performance of GCC methods in the presence of spatially correlated noises. This paper develops a maximum likelihood cross phase spectra estimator assuming that the prior knowledge of noise spatial coherence, which can be predicted in theory or measured in advance, is available. We show that it could provide a simple remedy to compensate the phase bias suffered by GCC-PHAT with virtually no extra cost.

II. Background

We consider a TDOA problem with signal model as

$$x_n(t) = s(t-\tau_n) + v_n(t), \quad n = 1, 2 \tag{1}$$

where t is the time index, $x_n(t)$ the received signal or observation, $s(t)$ the source signal, $v_n(t)$ the noises uncorrelated with $s(t)$, and the target is to estimate the TDOA defined by $\tau = \tau_2 - \tau_1$. It is convenient to consider this model in the frequency domain by rewriting (1) as

$$X_n(\omega, t) = e^{-j\omega\tau_n} S(\omega, t) + V_n(\omega, t), \quad n = 1, 2 \tag{2}$$

where $j = \sqrt{-1}$, and ω is the angular frequency. For second-order stationary observations, the power spectral density (PSD) matrix of $X_n(\omega, t)$ are given by

$$\begin{aligned} \boldsymbol{P}_x(\omega) =& E[\boldsymbol{X}(\omega,t)\boldsymbol{X}^H(\omega,t)] \\ =& P_s(\omega) \begin{bmatrix} 1 & e^{j\phi(\omega)} \\ e^{-j\phi(\omega)} & 1 \end{bmatrix} + \boldsymbol{P}_v(\omega) \end{aligned} \tag{3}$$

where $\boldsymbol{X}(\omega,t) = [X_1(\omega,t); X_2(\omega,t)]$ is a column vector, superscript H denotes Hermitian transpose, $P_s(\omega) = E[|S(\omega,t)|^2]$, $\phi(\omega) = \omega\tau$, and $\boldsymbol{P}_v(\omega)$ the noise PSD matrix. GCC-PHAT extracts the TDOA information exclusively from the phase of cross PSD, $E[X_1(\omega,t)X_2^*(\omega,t)]$, where superscript $*$ denotes conjugate. From (3), it is clear that this phase is an unbiased estimate of $\phi(\omega)$ if and only if $V_1(\omega,t)$ and $V_2(\omega,t)$ are uncorrelated, which in practice is seldom true for wide band noises. It is possible to prepare an estimate of $E[V_1(\omega,t)V_2^*(\omega,t)]$, and use the phase of noise compensated cross PSD, i.e., $E[X_1(\omega,t)X_2^*(\omega,t)] - E[V_1(\omega,t)V_2^*(\omega,t)]$, in GCC-PHAT. However, estimating the noise cross PSD brings inconvenience, or can be difficult for certain types of nonstationary noises, e.g., reverberations in acoustic TDOA problems. On the other hand, the complex coherence function of noise fields, which is defined as

$$\gamma_v(\omega) = \frac{E[V_1(\omega,t)V_2^*(\omega,t)]}{\sqrt{E[|V_1(\omega,t)|^2]E[|V_2(\omega,t)|^2]}} \tag{4}$$

is more readily available than the absolute level noise cross PSD, where $|\cdot|$ denotes absolute value. It is clear that $|\gamma_v(\omega)| \le 1$ by definition. In practice, $|\gamma_v(\omega)|$ is strictly less than 1. One example is the complex coherence function for the field of acoustic reverberations or diffuse noises given by [7], [8]

$$\gamma_v(\omega) = \mathrm{sinc}\,(\omega d/c) \tag{5}$$

where sinc is the sinc function, d the distance between sensors, i.e., microphones here, and c the speed of sound. This paper shows how to exploit the noise complex coherence function to compensate the phase bias in cross PSD $E[X_1(\omega,t)X_2^*(\omega,t)]$.

The Author is with GMEMS Technologies, Inc., 366 Fairview Way, Milpitas, CA 95035 (e-mail: lixilinx@gmail.com).

III. Maximum Likelihood (ML) Cross Phase Spectra Estimation

As the cross phase of each bin is estimated in the same way, we will drop out the parameter ω in this section to simplify our writings when doing so causes no misunderstanding.

A. ML Phase Estimation

We are interested in only the cross phase spectra. Hence, it is acceptable to normalize the empirical variances of observed $X_1(\omega, t)$ and $X_2(\omega, t)$ to the same level, say 1 without loss of generality. Then, an estimation of the PSD matrix defined in (3) will have the following normalized form

$$\hat{\mathbf{\Gamma}}_x = \begin{bmatrix} 1 & \hat{\gamma}_x \\ \hat{\gamma}_x^* & 1 \end{bmatrix} \tag{6}$$

where $\gamma_x(\omega) = \frac{E[X_1(\omega,t)X_2^*(\omega,t)]}{\sqrt{E[|X_1(\omega,t)|^2]E[|X_2(\omega,t)|^2]}}$, and $\hat{\gamma}_x(\omega)$ is an estimate of $\gamma_x(\omega)$ obtained by replacing the expectations with sample averages. We assume that $V_1(\omega, t)$ and $V_2(\omega, t)$ have the same power such that $\gamma_v(\omega)$ and $E[|V_1(\omega, t)|^2]$ are sufficient to describe the noise PSD matrix. Then, the true PSD matrix should have form

$$\mathbf{\Gamma}_x = P_s \begin{bmatrix} 1 & e^{j\phi} \\ e^{-j\phi} & 1 \end{bmatrix} + P_v \begin{bmatrix} 1 & \gamma_v \\ \gamma_v^* & 1 \end{bmatrix} = P_s \mathbf{\Phi} + P_v \mathbf{\Gamma}_v \tag{7}$$

due to (3), where P_s and P_v are the powers of source signal and noises, respectively, and $\mathbf{\Phi}$ and $\mathbf{\Gamma}_v$ the source signal and noise coherence matrices, respectively. We assume that the normalized observations are circular and Gaussian. Note that this Gaussianess assumption cannot exactly hold since the two observations are normalized to the same variance. Nevertheless, it is accurate enough for reasonably large sample sizes and works well on simulated data, as shown in Section III.B. Then, the sample size normalized negative natural logarithm of the likelihood function for the power normalized observations is given by [9]

$$J(\hat{\mathbf{\Gamma}}_x | P_s, P_v, \phi) = \log\det\mathbf{\Gamma}_x + \mathrm{tr}(\mathbf{\Gamma}_x^{-1}\hat{\mathbf{\Gamma}}_x) + 2\log\pi \tag{8}$$

where $\det$ and tr denote the determinant and trace of a square matrix, respectively. ML estimations for (P_s, P_v, ϕ) can be obtained by minimizing the cost in (8). Fortunately, their closed-form solutions are available by solving the system of equations $\partial J/\partial P_s = 0$, $\partial J/\partial P_v = 0$ and $\partial J/\partial \phi = 0$, where

$$\begin{aligned}
\frac{\partial J}{\partial P_s} &= \mathrm{tr}(\mathbf{\Gamma}_x^{-1}\mathbf{\Phi}) - \mathrm{tr}(\mathbf{\Gamma}_x^{-1}\hat{\mathbf{\Gamma}}_x\mathbf{\Gamma}_x^{-1}\mathbf{\Phi}) \\
\frac{\partial J}{\partial P_v} &= \mathrm{tr}(\mathbf{\Gamma}_x^{-1}\mathbf{\Gamma}_v) - \mathrm{tr}(\mathbf{\Gamma}_x^{-1}\hat{\mathbf{\Gamma}}_x\mathbf{\Gamma}_x^{-1}\mathbf{\Gamma}_v) \\
\frac{1}{P_s}\frac{\partial J}{\partial \phi} &= \mathrm{tr}(\mathbf{\Gamma}_x^{-1}\frac{d\mathbf{\Phi}}{d\phi}) - \mathrm{tr}(\mathbf{\Gamma}_x^{-1}\hat{\mathbf{\Gamma}}_x\mathbf{\Gamma}_x^{-1}\frac{d\mathbf{\Phi}}{d\phi})
\end{aligned} \tag{9}$$

It is a tedious process to show that these closed-form ML solutions for (P_s, P_v, ϕ) are

$$\begin{aligned}
P_{v,\mathrm{ml}} =& \frac{1-\mathrm{Re}(\gamma_v\hat{\gamma}_x^*)}{1-|\gamma_v|^2} - \sqrt{\left(\frac{1-\mathrm{Re}(\gamma_v\hat{\gamma}_x^*)}{1-|\gamma_v|^2}\right)^2 - \frac{1-|\hat{\gamma}_x|^2}{1-|\gamma_v|^2}} \\
P_{s,\mathrm{ml}} =& 1 - P_{v,\mathrm{ml}} \\
e^{j\phi_{\mathrm{ml}}} =& \gamma_{s,\mathrm{ml}} = \frac{\hat{\gamma}_x - P_{v,\mathrm{ml}}\gamma_v}{1-P_{v,\mathrm{ml}}}
\end{aligned} \tag{10}$$

where Re takes the real part of a complex variable. Yet, it is relatively easy to verify that (10) indeed give the ML solutions to our problem. Here, we only demonstrate the process verifying the correctness of solutions (10), and omit the less insightful procedure for arriving at them. With (10), it is easy to show that $\mathbf{\Gamma}_{x,\mathrm{ml}} = \hat{\mathbf{\Gamma}}_x$, where $\mathbf{\Gamma}_{x,\mathrm{ml}}$ is obtained by replacing the (P_s, P_v, ϕ) in (7) with their ML estimates. Hence, as required, all those derivatives in (9) become zero. Still, we further need to show that $0 \le P_{v,\mathrm{ml}} \le 1$ such that $P_{v,\mathrm{ml}}$ and $P_{s,\mathrm{ml}}$ are two valid power estimates, and $|\gamma_{s,\mathrm{ml}}| = 1$ such that equation $e^{j\phi_{\mathrm{ml}}} = \gamma_{s,\mathrm{ml}}$ is consistent. The key observation for proving these two properties is that $P_{v,\mathrm{ml}}$ is an eigenvalue of $\mathbf{\Gamma}_v^{-1}\hat{\mathbf{\Gamma}}_x$, which can be verified by straightforward algebra calculations.

Verification of property $0 \le P_{v,\mathrm{ml}} \le 1$: Note that $\mathbf{\Gamma}_v^{-1}\hat{\mathbf{\Gamma}}_x$ and $\mathbf{\Gamma}_v^{-0.5}\hat{\mathbf{\Gamma}}_x\mathbf{\Gamma}_v^{-0.5}$ share the same eigenvalues since $\det(\mathbf{\Gamma}_v^{-1}\hat{\mathbf{\Gamma}}_x - \lambda\boldsymbol{I}) = \det(\mathbf{\Gamma}_v^{-0.5}\hat{\mathbf{\Gamma}}_x\mathbf{\Gamma}_v^{-0.5} - \lambda\boldsymbol{I})$ for any λ. Thus, $P_{v,\mathrm{ml}}$ is real and nonnegative as $\mathbf{\Gamma}_v^{-0.5}\hat{\mathbf{\Gamma}}_x\mathbf{\Gamma}_v^{-0.5}$ is positive semidefinite. The condition ensuring $P_{v,\mathrm{ml}} \le 1$ can be shown to be

$$|\hat{\gamma}_x|^2 + |\gamma_v|^2 \ge 2\mathrm{Re}(\gamma_v\hat{\gamma}_x^*) \tag{11}$$

which is always true, and the equal sign holds only when $\hat{\gamma}_x = \gamma_v$.

Verification of property $|\gamma_{s,\mathrm{ml}}| = 1$: Recalling that $P_{v,\mathrm{ml}}$ is an eigenvalue of $\mathbf{\Gamma}_v^{-1}\hat{\mathbf{\Gamma}}_x$. Thus, $\det(\hat{\mathbf{\Gamma}}_x - P_{v,\mathrm{ml}}\mathbf{\Gamma}_v) = \det(\mathbf{\Gamma}_v)\det(\mathbf{\Gamma}_v^{-1}\hat{\mathbf{\Gamma}}_x - P_{v,\mathrm{ml}}\boldsymbol{I}) = 0$. On the other hand, we have $\det(\hat{\mathbf{\Gamma}}_x - P_{v,\mathrm{ml}}\mathbf{\Gamma}_v) = (1-P_{v,\mathrm{ml}})^2 - |\hat{\gamma}_x - P_{v,\mathrm{ml}}\gamma_v|^2$. Hence, we must have $1 - P_{v,\mathrm{ml}} = |\hat{\gamma}_x - P_{v,\mathrm{ml}}\gamma_v|$, which implies $|\gamma_{s,\mathrm{ml}}| = 1$.

It is also insightful to consider a few special solutions for $\phi_{\mathrm{ml}} = \angle\gamma_{s,\mathrm{ml}}$ as below, where $\angle$ takes the angle of a complex variable.

Case $\gamma_v = 0$: From (10), we have $P_{v,\mathrm{ml}} = 1 - \sqrt{1-(1-|\hat{\gamma}_x|^2)} = 1 - |\hat{\gamma}_x|$, and thus $\gamma_{s,\mathrm{ml}} = \hat{\gamma}_x/|\hat{\gamma}_x|$. Hence, ϕ_{ml} reduces to the naive cross phase spectra estimator used in the standard GCC-PHAT method.

Case $\hat{\gamma}_x = 0$: From (10), we have $P_{v,\mathrm{ml}} = 1/(1+|\gamma_v|)$, and thus $\gamma_{s,\mathrm{ml}} = -\gamma_v/|\gamma_v|$. Hence, the complex coherence functions of source signal and noises are in opposite phase in order to generate spatially uncorrelated observations.

Case $|\hat{\gamma}_x| = 1$: From (10), we have $P_{v,\mathrm{ml}} = 0$, and thus $\gamma_{s,\mathrm{ml}} = \hat{\gamma}_x$. This is the case when either no noise is present, or $\hat{\gamma}_x$ is estimated just on one sample of $\boldsymbol{X}(\omega, t)$.

Case $\gamma_v = \hat{\gamma}_x$: Unfortunately, the formula for $\gamma_{s,\mathrm{ml}}$ by (10) is not applicable since $P_{v,\mathrm{ml}} = 1$. This is not astonishing as no signal can be detected. Thus, ϕ_{ml} is undefined.

We propose the following proposition to summarize the main conclusion of this subsection.

Proposition 1: Assume that $E[|V_1(\omega, t)|^2] = E[|V_2(\omega, t)|^2]$, and $S(\omega, t)$ and $V_n(\omega, t)$ are uncorrelated in the signal model (2). Then, given $\hat{\gamma}_x(\omega)$ and $\gamma_v(\omega)$, the $\phi_{\mathrm{ml}}(\omega)$ in (10) provides an ML cross phase spectra estimator for the two delayed source signals.

B. Performance of the ML Phase Estimator

Note that the J in (8) is an negative logarithm likelihood normalized by sample size. Thus, Fisher information of the

ML phase estimator in (10) can be derived as $TE[(\partial J/\partial \phi)^2]$ or $TE[\partial^2 J/\partial \phi^2]$, where T is the number of independent samples used for the estimation of γ_x. It is more convenient to work with the second order derivative here. Starting from (9), we have

$$\begin{aligned}\frac{\partial^2 J}{\partial \phi^2} =& P_s \mathrm{tr}(\mathbf{\Gamma}_x^{-1}\frac{d^2\mathbf{\Phi}}{d\phi^2}) - P_s \mathrm{tr}(\mathbf{\Gamma}_x^{-1}\hat{\mathbf{\Gamma}}_x\mathbf{\Gamma}_x^{-1}\frac{d^2\mathbf{\Phi}}{d\phi^2}) \\ &- P_s^2\mathrm{tr}[(\mathbf{\Gamma}_x^{-1}\frac{d\mathbf{\Phi}}{d\phi})^2] + 2P_s^2\mathrm{tr}[\mathbf{\Gamma}_x^{-1}\hat{\mathbf{\Gamma}}_x(\mathbf{\Gamma}_x^{-1}\frac{d\mathbf{\Phi}}{d\phi})^2]\end{aligned} \tag{12}$$

Since $\mathbf{\Gamma}_{x,\mathrm{ml}} = \hat{\mathbf{\Gamma}}_x$, (12) can be simplified to

$$\begin{aligned}\frac{\partial^2 J}{\partial \phi^2} =& P_s^2\mathrm{tr}[(\mathbf{\Gamma}_x^{-1}\frac{d\mathbf{\Phi}}{d\phi})^2] \\ =& P_{s,\mathrm{ml}}^2\mathrm{tr}[(\hat{\mathbf{\Gamma}}_x^{-1}\begin{bmatrix} 0 & je^{j\phi_{\mathrm{ml}}} \\ -je^{-j\phi_{\mathrm{ml}}} & 0\end{bmatrix})^2]\end{aligned} \tag{13}$$

for the ML solutions. Thus, eventually $\partial^2 J/\partial\phi^2$ boils down into a function of $\hat{\gamma}_x$ and γ_v. Still, the exact evaluation of $TE[\partial^2 J/\partial \phi^2]$ seems computationally intractable. Here, we simply approximate $E[f(\hat{\gamma}_x)]$ with $f(\gamma_x)$ for a continuous function $f(\cdot)$ since $\hat{\gamma}_x \to \gamma_x$ for $T \to \infty$. Then, straightforward algebra calculations lead to

$$\mathcal{I}(\phi) = TE[\frac{\partial^2 J}{\partial \phi^2}] \approx \frac{2TP_s^2[1-\mathrm{Re}(\gamma_x^2 e^{-2j\phi})]}{(1-|\gamma_x|^2)^2} \tag{14}$$

where $\gamma_x = P_s e^{j\phi} + P_v\gamma_v$.

Noting that $\gamma_x e^{-j\phi} = P_s + P_v|\gamma_v|e^{-j(\phi - \angle\gamma_v)}$, it is possible to rewrite $\mathcal{I}(\phi)$ as a function $|\gamma_v|$, $\phi - \angle\gamma_v$ and the signal-to-noise ratio (SNR), P_s/P_v. Fig. 1 shows a few typical results of the predicted mean squared errors (MSEs) given by $1/\mathcal{I}(\phi)$ and test ones. The approximation in (14) is fairly accurate except when the SNR is low and $|\gamma_v|$ is close to 1. Obviously, both the predicted and test MSEs decrease with the increase of SNR, and are symmetric with respect to $\phi - \angle\gamma_v$. The MSEs are relatively flat with respect to $\phi - \angle\gamma_v$ for weakly spatially correlated noises, regardless of the SNR level. For the case of high SNR, the spatial correlation of noises helps to improve the performance when $|\phi - \angle\gamma_v|$ is small, but, becomes destructive for large $|\phi - \angle\gamma_v|$. However, the interaction between source signal and highly spatially correlated noises is a little complicated in the case of low SNR. Both predicted and test MSEs are minimized only when ϕ somewhat deviates from $\angle\gamma_v$.

Lastly, Fig. 2 shows the typical comparison results between the ML estimator and $\angle\hat{\gamma}_x$. Clearly, this naive phase estimator is biased when $|\gamma_v| \neq 0$. Increasing sample size is far less effective than increasing SNR for its performance improvement.

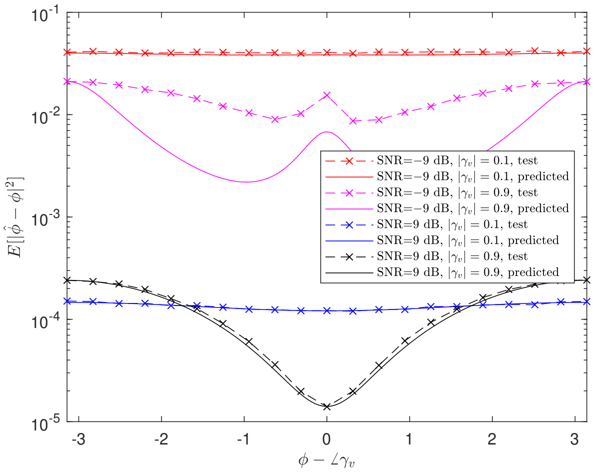

Fig. 1. Predicted and test MSEs versus ϕ with $T = 1000$ and varying SNR levels. Each test point is an MSE averaged $10,000$ runs.

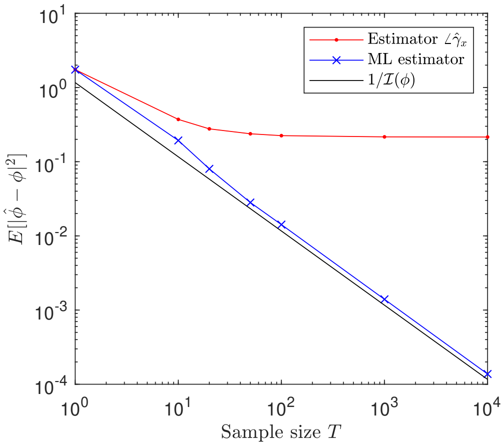

Fig. 2. Test and predicted accuracy of $\hat{\phi}$ versus sample size T. Settings of the signal model are SNR $= 0$ dB, $\phi = 0.5\pi$, and $\gamma_v = 0.5$. Each test MSE is averaged over $10,000$ runs.

IV. Enhanced GCC-PHAT with ML Cross Phase Spectra

One immediate application of Proposition 1 is to enhance the performance of GCC-PHAT. The inputs are $\hat{\gamma}_x(\omega)$ and $\gamma_v(\omega)$, where $\hat{\gamma}_x(\omega)$ is the estimated complex coherence function between the two observations $x_1(t)$ and $x_2(t)$ in the frequency domain, and $\gamma_v(\omega)$ the noise complex coherence function, e.g., (5) for acoustic reverberations and diffuse noises. For each frequency bin, we calculate the ML cross phase spectra as in (10). Then, we convert $\gamma_{s,\mathrm{ml}}(\omega)$ to a time domain function of τ as in the standard GCC-PHAT method. Lastly, the τ associated with the maximum value of this time domain function serves as the TDOA estimate.

V. Experimental Results

We have tested enhanced GCC-PHAT on a real world speech signal TDOA problem. Matlab/Octave[1] code and audio data for reproducing the results reported here can be obtained from www.github.com/lixilinx/EnhancedPHAT. An array of two microphones separated by 11 cm was used to record a single far field speech source in a conference room with considerable reverberations and back ground noises. The true TDOA is about 5 sampling periods at sampling rate 16 KHz. The total length of speech signal is about half a minute. A

[1] www.mathworks.com and www.gnu.org/software/octave, respectively

short time Fourier transform (STFT) with frame size 32 ms and hop size 10 ms is used to obtain $X_n(\omega, t)$ from $x_n(t)$.

Fig. 3 shows the theoretical complex coherence function predicted by (5), and the estimated one on 5.6 s noises. Real part of the estimated noise complex coherence function matches that of theoretical prediction reasonably well. But, its imaginary part significantly deviates from zero at around 1.6 KHz, possibly due to directional noises caused by air conditioner. Nevertheless, we will use the theoretical noise complex coherence function in enhanced GCC-PHAT.

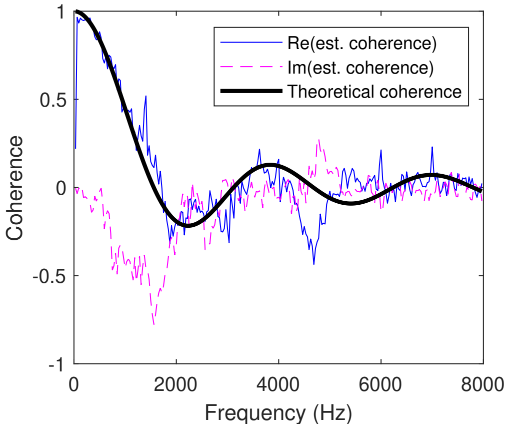

Fig. 3. Theoretical and estimated noise complex coherence functions. Only real part of the theoretical prediction is plotted since its imaginary part is zero.

Except for the standard GCC-PHAT base line, we also have included the wide band generalized eigenvalue decomposition multiple signal classification (GEVD-MUSIC) method from [6] for comparison. The noise PSD matrices required to whiten the microphone signals in GEVD-MUSIC is estimated on the same 5.6 s noises. Note that GEVD-MUSIC is a significantly more complicated method than GCC-PHAT and its enhanced version. Furthermore, it is allowed to access the noise PSD matrices, while GCC-PHAT and its enhanced version do not rely on such extra information.

Fig. 4 shows the percentage of failures averaged over $20,000$ runs versus the length of speech signals used for TDOA estimation. We say that a TDOA estimation fails if $|\hat{\tau}-\tau| \ge 1$, i.e., $\hat{\tau} \ne \tau$ as both are integers. From Fig. 4, we see that the enhanced GCC-PHAT significantly outperforms GCC-PHAT, and is comparable with GEVD-MUSIC in performance.

We also tried to test enhanced GCC-PHAT with the estimated noise complex coherence function. Interestingly, no meaningful performance gain is observed compared with the use of theoretical prediction by (5). One possible reason could be that the reverberations favorite the theoretical prediction, and at the same time, make the coherence estimation exclusively using background noises less accurate when speech is active.

VI. Conclusion

We have derived a closed-form maximum likelihood (ML) cross phase spectra estimator in the setting of known noise spatial complex coherence function. It can be used to enhance the performance of generalized cross correlation with phase transform (GCC-PHAT) for time difference of arrival (TDOA) estimation. Experimental results on real world speech TDOA task suggest that the enhanced GCC-PHAT could perform as well as the significantly more complicated generalized eigenvalue decomposition multiple signal classification (GEVD-MUSIC) method. Our enhanced GCC-PHAT provides an attractive alternative for the standard GCC-PHAT since the performance boost is offered as a virtually free lunch.

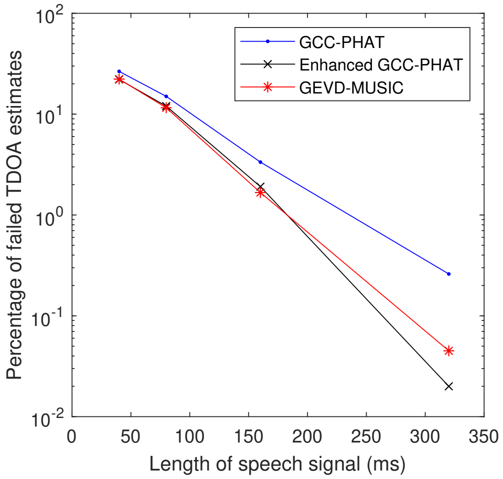

Fig. 4. Percentage of TDOA estimation failures averaged over $20,000$ runs versus the length of speech signal used for observation coherence estimation.

References

[1] C. Knapp and G. Carter, “The generalized correlation method for estimation of time delay”, *IEEE Transactions on Acoustics, Speech and Signal Processing.*, vol. 24, no. 4, pp. 320–327, Aug. 1976.

[2] G. C. Carter, “Coherence and time delay estimation,” *Proceedings of the IEEE*, vol. 75, no. 2, pp. 236–255, Feb. 1987.

[3] J. DiBiase, H. Silverman, and M. Brandstein, “Robust localization in reverberant rooms,” in *Microphone Arrays* (Digital Signal Processing), M. Brandstein and D. Ward, Eds. Berlin, Germany: Springer-Verlag, 2001.

[4] A. Schmidt, H. W. Löllmann, and W. Kellermann, “Acoustic self-awareness of autonomous systems in a world of sounds,” *Proceedings of the IEEE*, vol. 108, no. 7, pp. 1127–1149, Jul. 2020.

[5] R. Schmidt, “Multiple emitter location and signal parameter estimation,” *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, Mar. 1986.

[6] K. Nakamura, K. Nakadai, and H. G. Okuno, “A real-time super-resolution robot audition system that improves the robustness of simultaneous speech recognition,” *Advanced Robotics*, vol. 24, no. 5-6, pp. 933–945, Apr. 2013.

[7] R. K. Cook, R. V. Waterhouse, R. D. Berendt, S. Edelman, and M. C. Thompson Jr, “Measurement of correlation coefficients in reverberant sound fields,” *The Journal of the Acoustical Society of America*, vol. 27, no. 6, pp. 1072–1077, Nov. 1955.

[8] F. Jacobsen and T. Roisin, “The coherence of reverberant sound fields,” *The Journal of the Acoustical Society of America*, vol. 108, no. 1, pp. 204–210, Jul. 2000.

[9] X. L. Li, T. Adali, and M. Anderson, “Noncircular principal component analysis and its application to model selection,” *IEEE Transactions on Signal Processing*, vol. 59, no. 10, pp. 4516–4528, Oct. 2011.