Joint Time Delay and Frequency Estimation of Multiple Sinusoids

G. Liao*, H. C. So† and P. C. Ching*

* Department of Electronic Engineering, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong

†Department of Electronic Engineering, City University of Hong Kong, Tat Chee Avenue, Hong Kong

e-mail: gsliao@ee.cuhk.edu.hk, hcso@ee.cityu.edu.hk, pcching@ee.cuhk.edu.hk

ABSTRACT

In this paper, we devise a new subspace method for estimating the differential time delay of a signal received at two separated sensors as well as the frequencies of the source signal, assuming that it consists of multiple sinusoids. The time delay and frequency estimates are related to the eigenvalues and eigenvectors of a matrix obtained from the covariances of the received signals. The effectiveness of the proposed algorithm is demonstrated via computer simulations using sinusoidal signals as well as real speech data.

I. Introduction

Time delay estimation between signals received at spatially separated sensors [1] and sinusoidal frequency estimation [2] are two research topics which have been widely studied for many years. Recently, the problem of joint time delay and frequency estimation of sinusoidal signals has also attracted considerable attention. Application examples for this problem include speech enhancement and pitch estimation using a microphone array [3], synchronization in CDMA systems [4], analysis of thalamocortical seizure pathways [5] and FSK demodulation using multiple segments [6].

Let the discrete-time sinusoidal signals received at two sensors be

$$r_1(n) = s(n) + q_1(n), \qquad \text{and}$$
$$r_2(n) = s(n-D) + q_2(n), \quad n = 0, 1, \cdots, N-1 \tag{1}$$

where

$$s(n) = \sum_{m=1}^{P} \alpha_m \exp(j\omega_m n) \tag{2}$$

The source signal $s(n)$ is modeled by a sum of P complex sinusoids where the amplitudes $\{\alpha_m\}$ are unknown complex-valued constants and the normalized radian frequencies $\{\omega_m\}$ are distinct. Without loss of generality, we assign $\omega_1 < \omega_2 < \cdots < \omega_P$. It is assumed that P is known *a priori* or an accurate estimate of P has been obtained [7]. The additive noises $q_1(n)$ and $q_2(n)$ are uncorrelated zero-mean complex white Gaussian processes with variances $\sigma_{q_1}^2$ and $\sigma_{q_2}^2$, respectively. The parameter D represents the difference in arrival times at the two receivers and N is the number of samples collected at each channel. Our goal is to estimate both the time difference of the received signals and the frequencies of their constituent components.

When $s(n)$ is a single sinusoid, a discrete-time Fourier transform based approach [8] can be used to achieve the optimum time delay and frequency estimation. For $P > 1$, Sherman *et al* [5] had presented an ESPRIT algorithm [9] to estimate D while a generalized Yule-Walker solution was suggested in [6] to determine $\{\omega_m\}$ separately. In this paper, a novel subspace method for joint estimation of the time delay and frequencies is developed. Basically, the estimation procedure involves finding the eigenvalues and eigenvectors from the product of the cross-correlation matrix of the two received signals and the pseudoinverse of the correlation matrix of $s(n)$. Simulation results show that the mean square errors of the time delay and frequency estimates are close to their corresponding Cramér-Rao lower bounds (CRLBs). Applicability of the proposed method for a voiced speech arriving at spatially separated microphones is also demonstrated.

II. The Proposed Algorithm

Using the received signals, we form the following set of vectors

$$\mathbf{X_1}(k) = [r_1(k), r_1(k+1), \cdots, r_1(k+M-1)]^T$$
$$\mathbf{X_2}(k) = [r_2(k), r_2(k+1), \cdots, r_2(k+M-1)]^T \tag{3}$$

where $k = 0, 1, \cdots, K-1$ with $K = N - M + 1$ and T denotes the transpose operation. The parameter M is the length of each vector and K has a value lies between $P+1$ and $N-P+1$ so that the span of any K of $\mathbf{X_1}(k)$ or $\mathbf{X_2}(k)$ has no rank less than P. Substituting (1) and (2) into (3) gives

$$\mathbf{X_1}(k) = \mathbf{A}(\omega)\mathbf{S}(k) + \mathbf{Q}_1(k)$$
$$\mathbf{X_2}(k) = \mathbf{A}(\omega)\mathbf{\Delta}(\omega)\mathbf{S}(k) + \mathbf{Q}_2(k) \tag{4}$$

where

$$\mathbf{A}(\omega) = [\mathbf{a_1}, \mathbf{a_2}, \cdots, \mathbf{a_P}]$$

$$\begin{aligned}
\mathbf{a_m} &= \left[1, e^{j\omega_m}, \cdots, e^{j\omega_m(M-1)}\right]^T, \; m = 1, 2, \cdots, P \\
\mathbf{S}(k) &= \left[\alpha_1 e^{j\omega_1 k}, \alpha_2 e^{j\omega_2 k}, \cdots, \alpha_P e^{j\omega_P k}\right]^T \\
\mathbf{Q}_1(k) &= [q_1(k), q_1(k+1), \cdots, q_1(k+M-1)]^T \\
\mathbf{Q}_2(k) &= [q_2(k), q_2(k+1), \cdots, q_2(k+M-1)]^T \\
\mathbf{\Delta}(\omega) &= \mathrm{diag}\{e^{-jD\omega_1}, e^{-jD\omega_2}, \cdots, e^{-jD\omega_P}\}
\end{aligned} \tag{5}$$

We can see that the time delay and frequency information is embedded in the Vandermonde matrix $\mathbf{A}(\omega)$ as well as the diagonal matrix $\mathbf{\Delta}(\omega)$, which have dimensions of $M \times P$ and $P \times P$, respectively.

Using (4) and (5), the auto-correlation matrix of $\mathbf{X_1}(k)$ can be shown to be

$$\begin{aligned}
\mathbf{R}_{11} &\stackrel{\mathrm{def}}{=} E\left[\mathbf{X}_1(k)\mathbf{X}_1^H(k)\right] \\
&= \mathbf{A}(\omega)\mathbf{R}_s\mathbf{A}^H(\omega) + \sigma_{q_1}^2\mathbf{I}
\end{aligned} \tag{6}$$

where

$$\mathbf{R}_s = E[\mathbf{S}(k)\mathbf{S}^H(k)] \tag{7}$$

has full rank because the P sinusoidal frequencies are distinct. The notations E and H denote the expectation and Hermitian transposition, respectively, and $\mathbf{I}$ is the $M \times M$ identity matrix. Similarly, the cross-correlation matrix of $\mathbf{X_1}(k)$ and $\mathbf{X_2}(k)$ can be written as

$$\begin{aligned}
\mathbf{R}_{21} &\stackrel{\mathrm{def}}{=} E\left[\mathbf{X}_2(k)\mathbf{X}_1^H(k)\right] \\
&= \mathbf{A}(\omega)\mathbf{\Delta}(\omega)\mathbf{R}_s\mathbf{A}^H(\omega)
\end{aligned} \tag{8}$$

In practice, $\mathbf{R}_{11}$, $\mathbf{R}_{21}$ and $\sigma_{q_1}^2$ are unknown and they must be determined from the N measurements of $r_1(n)$ and $r_2(n)$. Consistent estimates of $\mathbf{R}_{11}$ and $\mathbf{R}_{21}$ are given by

$$\begin{aligned}
\hat{\mathbf{R}}_{11} &= \frac{1}{K}\sum_{k=0}^{K-1}\mathbf{X}_1(k)\mathbf{X}_1^H(k) \\
\hat{\mathbf{R}}_{21} &= \frac{1}{K}\sum_{k=0}^{K-1}\mathbf{X}_2(k)\mathbf{X}_1^H(k)
\end{aligned} \tag{9}$$

while the noise power estimate $\hat{\sigma}_{q_1}^2$ is found by averaging the $(M-P)$ smallest eigenvalues of $\hat{\mathbf{R}}_{11}$. A good estimate of the noise-free covariance matrix of $\mathbf{X_1}(k)$, that is, $\mathbf{A}(\omega)\mathbf{R}_s\mathbf{A}^H(\omega)$, denoted by $\hat{\mathbf{C}}_{11}$, is of the form

$$\hat{\mathbf{C}}_{11} = \sum_{i=1}^{P}(\lambda_i - \hat{\sigma}_{q_1}^2)\mathbf{V}_i\mathbf{V}_i^H \tag{10}$$

where $\lambda_1 > \lambda_2 > \cdots > \lambda_M$ and $\{\mathbf{V}_i\}$ are a set of orthonormal vectors which are obtained from the eigenvalue decomposition of estimate of $(\hat{\mathbf{R}}_{11} - \hat{\sigma}_{q_1}^2\mathbf{I}) = \sum_{i=1}^{M}\lambda_i\mathbf{V}_i\mathbf{V}_i^H$. It can be easily shown that the pseudoinverse of $\hat{\mathbf{C}}_{11}$ is

$$\hat{\mathbf{C}}_{11}^{\dagger} = \sum_{l=1}^{P}\lambda_l^{-1}\mathbf{V}_l\mathbf{V}_l^H \tag{11}$$

Since $\mathrm{span}\{\mathbf{a_1}, \mathbf{a_2}, \cdots, \mathbf{a_P}\} = \mathrm{span}\{\mathbf{V}_1, \mathbf{V}_2, \cdots, \mathbf{V}_P\}$ in the absence of noise and from (8), (10)-(11), we can express the term $\hat{\mathbf{R}}_{21}\hat{\mathbf{C}}_{11}^{\dagger}\mathbf{A}(\omega)$ as follows,

$$\begin{aligned}
&\hat{\mathbf{R}}_{21}\hat{\mathbf{C}}_{11}^{\dagger}\mathbf{A}(\omega) \\
&\approx \mathbf{A}(\omega)\mathbf{\Delta}(\omega)\left(\mathbf{R}_s\mathbf{A}^H(\omega)\right)\sum_{l=1}^{P}\lambda_l^{-1}\mathbf{V}_l\mathbf{V}_l^H\mathbf{A}(\omega) \\
&= \mathbf{A}(\omega)\mathbf{\Delta}(\omega)\Bigg(\left(\mathbf{A}^H(\omega)\mathbf{A}(\omega)\right)^{-1}\mathbf{A}^H(\omega) \\
&\quad \cdot\sum_{i=1}^{P}\lambda_i\mathbf{V}_i\mathbf{V}_i^H\Bigg)\cdot\sum_{l=1}^{P}\lambda_l^{-1}\mathbf{V}_l\mathbf{V}_l^H\mathbf{A}(\omega) \\
&= \mathbf{A}(\omega)\mathbf{\Delta}(\omega)\left(\mathbf{A}^H(\omega)\mathbf{A}(\omega)\right)^{-1} \\
&\quad \cdot\left(\mathbf{A}^H(\omega)\sum_{i=1}^{P}\mathbf{V}_i\mathbf{V}_i^H\mathbf{A}(\omega)\right) \\
&= \mathbf{A}(\omega)\mathbf{\Delta}(\omega)\left(\mathbf{A}^H(\omega)\mathbf{A}(\omega)\right)^{-1}\left(\mathbf{A}^H(\omega)\mathbf{A}(\omega)\right) \\
&= \mathbf{A}(\omega)\mathbf{\Delta}(\omega)
\end{aligned} \tag{12}$$

As $N \to \infty$ and when noise is absent, $\hat{\mathbf{R}}_{21} \to \mathbf{R}_{21}$ and $\hat{\mathbf{C}}_{11} \to \mathbf{A}(\omega)\mathbf{R}_s\mathbf{A}^H(\omega)$, $\mathbf{a_m}$ will be the same as the eigenvector of $\hat{\mathbf{R}}_{21}\hat{\mathbf{C}}_{11}^{\dagger}$ associated with the eigenvalue $e^{-j\omega_m D}$, for $m = 1, 2, \cdots, P$. Nevertheless, under other situations, reasonable estimates of $\mathbf{a_m}$ and $\{e^{-j\omega_m D}\}$ can still be obtained from $\hat{\mathbf{R}}_{21}\hat{\mathbf{C}}_{11}^{\dagger}$. In our study, we use the following steps to find D and $\{\omega_m\}$:

1. Compute the P largest eigenvalues of $\hat{\mathbf{R}}_{21}\hat{\mathbf{C}}_{11}^{\dagger}$, $\hat{\lambda}_1, \hat{\lambda}_2, \cdots, \hat{\lambda}_P$. Hence find the corresponding P eigenvectors, denoted by $\{\hat{\mathbf{a}}_\mathbf{m}\}$, $m = 1, 2, \cdots, P$.

2. Let $\hat{\mathbf{a}}_\mathbf{m} = [\hat{a}_m(1), \hat{a}_m(2), \cdots, \hat{a}_m(M)] \approx \hat{a}_m(1) \cdot \left[1, e^{j\hat{\omega}_m}, \cdots, e^{j\hat{\omega}_m(M-1)}\right]^T$ where $\hat{\omega}_m$ denotes the estimate of ω_m. The frequency estimate is obtained by minimizing a least squares cost function $\sum_{l=1}^{M-1}(l\omega_m - \angle(\hat{a}_m(l+1)/\hat{a}_m(1)))^2$ where $\angle(v)$ represents the phase angle of v, and it is given by

$$\hat{\omega}_m = \frac{6\sum_{l=1}^{M-1} l\angle\left(\frac{\hat{a}_m(l+1)}{\hat{a}_m(1)}\right)}{M(M-1)(2M-1)} \tag{13}$$

3. Based on the P eigenvalues and frequency estimates, the estimated time delay $\hat{D}$ is evaluated as $\hat{D} = (-1/P)\sum_{m=1}^{P} \angle(\hat{\lambda}_m)/\hat{\omega}_m$.

Notice that if $s(n)$ is a periodic signal, then the P complex sinusoids in (2) are harmonically related and the frequencies must satisfy

$$\omega_m = m\omega_1, \qquad m = 1, 2, \cdots, P \qquad (14)$$

It is expected that the accuracies of $\{\omega_m\}$ and D should be increased by utilizing the harmonic signal information in the estimation process. In this case, instead of using (13), we can minimize another least squares cost function of the form $\sum_{m=1}^{P}\sum_{l=1}^{M-1}(lm\omega_1 - \angle(\hat{a}_m(l+1)/\hat{a}_m(1)))^2$ to estimate the fundamental frequency, ω_1, as follows,

$$\hat{\omega}_1 = \frac{36\displaystyle\sum_{m=1}^{P}\sum_{l=1}^{M-1} lm\angle\left(\frac{\hat{a}_m(l+1)}{\hat{a}_m(1)}\right)}{M(M-1)(2M-1)P(P+1)(2P+1)} \qquad (15)$$

III. Simulation Results & Conclusions

Computer simulations were conducted to evaluate the time delay and frequency estimation performance of the proposed method in the presence of white Gaussian noise. All results provided were averages of 200 independent runs. In the first test, the source signal $s(n)$ was a sinusoidal signal of the form $s(n) = \alpha_1 e^{j\omega_1 n} + \alpha_2 e^{j\omega_2 n}$ with $\alpha_1 = 1$, $\alpha_2 = e^{-j\pi/4}$, $\omega_1 = 0.2\pi$ rad/s and $\omega_2 = 0.4\pi$ rad/s, which was a periodic signal. The sampling interval was 1 s and the time delay D was selected to be 1.7 s. We assigned $\sigma_{q_1}^2 = \sigma_{q_2}^2$ and different signal-to-noise ratios (SNRs) were obtained by proper scaling of the noise sequences. The number of samples used was $N = 200$ while the vector length M had a value of 4. Figures 1 to 3 plot the mean square errors (MSEs) of the frequency and time delay estimates versus SNR, together with the corresponding CRLBs. The frequency variances for ω_1 and ω_2 obtained from the conventional ESPRIT method [9] were also included for comparison. In Figures 1 and 2, it can be seen that the two variants of the proposed algorithm, that is, with and without exploiting the periodic property of $s(n)$, were superior to the ESPRIT method in frequency estimation. However, it is noted that the performance of the proposed method will be greatly improved if (15) is employed instead of (13). At high SNRs, the variances of the frequency estimates based on (15) were roughly 6 dB above the CRLB for frequency. For example, the MSE of $\hat{\omega}_1$ was -63.5 dB at SNR = 20 dB and this implies that $\hat{\omega}_1 \in (\omega_1 - 2\times10^{-3}, \omega_1 + 2\times10^{-3})$ with a probability of 99.75%, assuming that the frequency error was Gaussian distributed. While in Figure 3, we observe that for SNR $\geq$ 0 dB, the MSE of $\hat{D}$ was larger than the corresponding CRLB by approximately 2 dB but there was no obvious improvement when the harmonic signal information was utilized.

In the second experiment, the application of the proposed algorithm for joint time delay and pitch estimation of voiced speech received at two microphones was investigated. The signal $s(n)$ was now a portion of an Cantonese word and its waveform is depicted in Figure 4. The voiced speech was sampled at 16 kHz. This speech frame was shifted by one sampling interval to obtain another frame to simulate the speech data from a second microphone. We used six complex sinusoids ($P = 6$) or three real sinusoids to model the speech [10] while the parameters N and M were chosen to be 320 and 10, respectively. We assumed the frequencies of the speech were harmonically related and thus (15) was used in the frequency estimation procedure. The MSEs of the time delay estimates of the proposed algorithm as well as the direct cross-correlation (DCC) and generalized cross correlation (GCC) methods [1] are shown in Figure 5. We see that the subspace method outperformed both the DCC and GCC for a wide range of SNRs. On the other hand, the estimated normalized pitch frequency was 0.11 rad/s which agreed with the fundamental frequency of the voiced speech.

To conclude, a subspace algorithm for joint time delay and frequency estimation of sinusoidal signals received at two separated sensors has been proposed. The time delay and frequency estimates are derived using the eigenvalues and eigenvectors of a matrix obtained from the covariances of the received signals. It is shown that for periodic signals, the time delay and frequency estimation performances are inferior to the CRLBs by only a few dB. The proposed method has also been successfully applied in joint pitch and delay estimation using two microphones.

References

[1] G.C.Carter, *Coherence and Time Delay Estimation: An Applied Tutorial for Research, Development, Test and Evaluation Engineers*, IEEE press, 1993

[2] P.Stoica and R.Moses *Introduction to Spectral Analysis*, Prentice-Hall, 1997

[3] X.Qian and R.Kumaresan, "Joint estimation of time delay and pitch of voiced speech signals," *Conf. Rec. of the 29th Asilomar Conf. Signals, Systems & Computers*, vol.1, pp.735-739, 1995

[4] S.R.Dooley and A.K.Nandi, "Adaptive time delay and frequency esitmation for digital signal synchronization in CDMA systems," *Conf. Rec. of the 32th Asilomar Conf. Signals, Systems & Computers* , vol.2, pp.1838-1842, 1998

[5] D.L.Sherman, Y.C.Tsai, L.A.Rossell, M.A.Mirski and N.V.Thakor, "Narrowband delay estimation for thalamocortical epileptic seizure pathways," *Proc.*

IEEE Int. Conf. Acoust. Speech, Signal Processing, vol.5, pp.2939-2942, May 1995, Detroit, Michigan, USA

[6] J.A.Sills and Q.R.Black, "Frequency estimation from short pulses of sinusoidal signals," *Proc. IEEE MILCOM '96*, vol.3, pp.979-983, 1996

[7] H.T.Wu, J.F.Yang and F.K.Chen, "Source number estimators using transformed Gerschgorin radii," *IEEE Trans. Signal Processing*, vol.43, pp.1325-1333, 1995

[8] H.C.So, "Delay estimation using sinusoidal signals," *Proc. IEEE Int. Conf. Acoust. Speech, Signal Processing*, vol.5, pp.3168-3171, June 2000, Istanbul, Turkey

[9] R.Roy, T.Kailath, "ESPRIT - Estimation of signal parameter via rotational invariance techinques," *IEEE Trans. Acoust. Speech, Signal Processing*, vol.37, no.7, pp.984-995, July 1989

[10] R.J.Mcaulay and T.F.Quateri, "Speech analysis/synthesis based on a sinuosoidal representation," *IEEE Trans. Acoust. Speech, Signal Processing*, vol.34, no.4, pp.744-754, April 1986

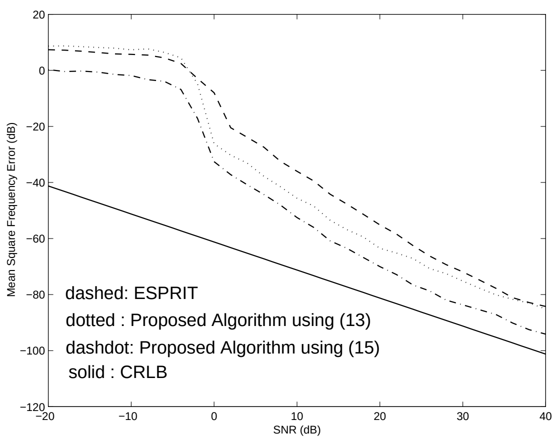

Figure 1: MSEs of $\hat{\omega}_1$ versus SNR for periodic signal

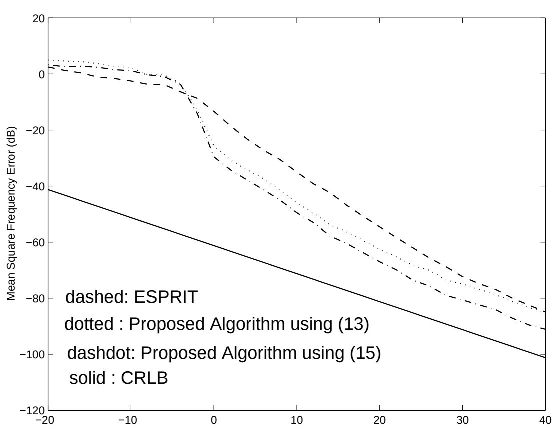

Figure 2: MSEs of $\hat{\omega}_2$ versus SNR for periodic signal

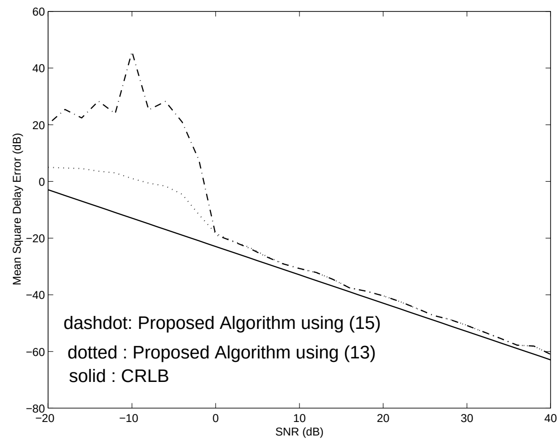

Figure 3: MSEs of $\hat{D}$ versus SNR for periodic signal

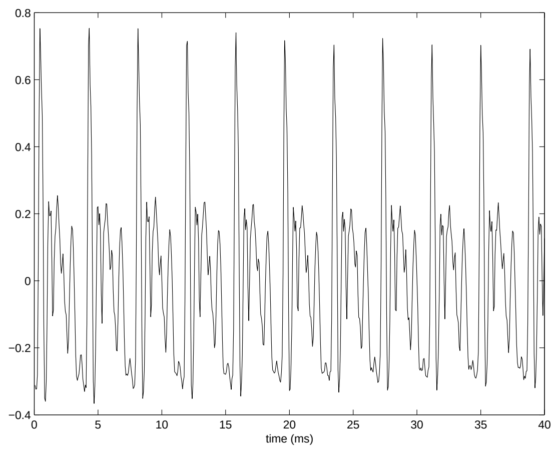

Figure 4: Waveform of voiced Cantonese speech

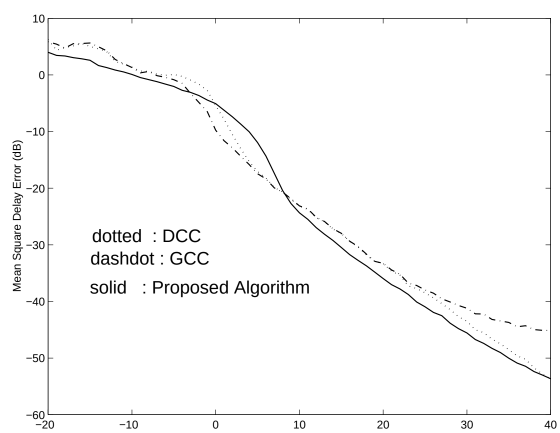

Figure 5: MSEs of $\hat{D}$ versus SNR for voiced speech