Multiple Frequency and Source Angle Estimation by Gaussian Mixture Model with Modified Microphone Array Data Model

Bandhit Suksiri\(^1\) and Masahiro Fukumoto\(^2\)

School of Information, Kochi University of Technology
Kami Campus, Tosayamada, Kami City, Kochi 782-8502, Japan
E-mail: \(^1\)208005s@gs.kochi-tech.ac.jp, \(^2\)fukumoto.masahiro@kochi-tech.ac.jp

Abstract
This paper presents an alternative and efficient one-dimensional direction-of-arrival estimation method for wide-band sources. The proposed method is compatible with most classical subspace-based methods, such as, conventional and root multiple signal classification. Although it only employs a Gaussian mixture model with a maximum likelihood estimation algorithm, it is sufficient for exhibiting wide-band sources angle estimation. Modification of the microphone array data model is proposed and investigated in order to avoid confusion caused by unwanted side lobes in uniform linear arrays radiation. The performance is evaluated in terms of the root-mean-squared error over a range of the signal-to-noise ratio. In conclusion, the proposed method enables the synthesis of signal sources and provides a potential alternative to intelligent source localization systems.

1. Introduction

The fundamental competence of source localization has been extensively implemented in navigation systems for the exploration of sources, which is known as direction-of-arrival (DOA) estimation. The main factors reducing the efficiency of source localization in navigation systems are background noise, multiple source locations, and variation in the sound patterns and source frequencies [1]. A number of methods have been introduced in order to improve the source localization efficiency, including, time-difference-of-arrival (TDOA)-based DOA estimation [2], delay and sum beamforming [3] and subspace method [4]-[6].

Subspace methods are increasingly utilized for the DOA estimation of multiple sources in array signal processing. The conventional subspace method known as multiple signal classification (MUSIC) decomposes the observation space into a signal subspace and a noise subspace by using eigenvalue decomposition [4]. Such conventional methods calculate the directions of sources using the noise subspace, leading to superior multiple-source localization. Nonetheless, only narrow-band sources can be localized.

In order to estimate the DOA for wide-band sources, the coherent signal subspace (CSS) method was introduced [7]. This method constructively focuses the observation subspace at distinct frequency bins to a single subspace associated with a reference frequency, producing a smoothed correlation matrix. Conventional narrow-band DOA estimators such as MUSIC may then be used for DOA estimation. A common shortcoming of CSS methods is the requirement of an initial DOA estimation. It was previously found that the estimation performance is sensitive to the initial conditions [8]. Poor initial conditions can lead to biased estimates.

This paper therefore aims to propose a new DOA estimation scheme for wide-band sources without any requirement of DOA initial estimation. Unlike the CSS method, in which the signal and noise subspaces in each frequency bin must be arranged into a single viable correlation matrix, the proposed method focuses the entire observation subspace in each bin with a distinct frequency to a multiple correlation matrix. After narrow-band DOA estimation is employed in each correlation matrix, the set of estimated DOAs can be classified into groups by a Gaussian mixture model with a maximum likelihood estimation algorithm. In order to avoid confusion caused by unwanted side lobes in uniform linear arrays (ULA) radiation, a modification of the microphone array data model is proposed and investigated. Furthermore, the proposed method enables the synthesis of signal sources.

2. Microphone Array Data Model

The classical phase difference function can be simply presented in Euler’s formula as \(\exp(i2\pi ft)\), where \(f\) is the source frequency and \(t\) is time difference of arrival due to the distance between the source and microphones. In addition, the ULA phase difference function can be expressed as follows:

\[
w_m(\theta, f) = \exp \left( i2\pi f (m - 1) \frac{d \sin \theta}{c} \right)
\]

where \(d\) is the spacing of the microphone elements, \(m\) is the index of the microphone elements, \(\theta\) is the angle and \(c\) is the...
speed of sound. According to [1], in DOA estimation techniques, to avoid confusion caused by many grating lobes and the lower power beam-width of the lobe in ULA radiation, \( d \) should be set to \( \frac{\lambda}{2} \), where \( \lambda = \frac{\pi}{f} \) is the wavelength and \( f \) is the frequency corresponding to the wavelength. Assuming that the power beam-width of the grating lobes in ULA radiation also depends on the source frequency and the number of microphone elements, the phase difference function is redefined as follows:

\[
w_m(\theta, \omega) = \exp\left(\frac{i2\pi\omega(m-1)d\sin\theta}{\omega\lambda}\right) \tag{2}
\]

where \( d \) is the spacing of microphone elements, \( m \) is the index of microphone elements, \( \theta \) is the angle and \( \lambda \) is the wavelength. Note that \( \omega = 2\pi f \) and \( \omega_\lambda = 2\pi f\lambda \).

Figure 1 shows a radiation patterns of eight- and 16-element microphone arrays normalized by the modified phase difference function in Eq.(2). As \( \omega \) decreases from \( \frac{\lambda}{2} \) to \( \frac{\lambda}{4} \), the power beam-width decreases. As the number of elements in the microphone array \( M \) increases from eight to 16, the power beam-width increases. This study indicates that the power beam-width of the grating lobe pattern depends on \( \omega \) and \( M \). These results suggest that source frequency range should be selected to be \( \omega \leq \omega_\lambda \). When a lower source frequency is chosen, the results suggest that the number of microphone elements should be increased to avoid the confusion.

### 3. DOA Estimation Scheme

Consider a one-dimensional ULA model in which the microphones are uniformly spaced [1] with longitudinal wavefront sources, and the output of each microphone is uniformly sampled and decomposed into snapshots or frames by a short-time Fourier transform (STFT), with each frame containing a set of frequencies. The received signal vector from \( P \) incident sources at \( \tau_k \) snapshot can be represented as:

\[
x(\tau_k, \omega_j) = W(\theta, \omega_j) s(\tau_k, \omega_j) + n(\tau_k, \omega_j) \tag{3}
\]

where \( \tau_k, k = 1, 2, \ldots, K \) are snapshots, \( \omega_j, j = 1, 2, \ldots, J \) are discrete frequencies, \( x \in \mathbb{C}^M \) is the sum of received signal vectors for all sources with different phases or angles, \( s \in \mathbb{C}^P \) is a signal source vector, \( n \in \mathbb{C}^P \) is a noise vector, \( \theta \in \mathbb{R}^P \) is an angle vector and \( W \in \mathbb{C}^{M \times P} \) is a phase difference matrix. Each attribute in \( W \) depends on the data model, which is described in Eq.(2).

The classical subspace-based method called MUSIC [4] and the estimation of signal parameters via a rotational invariance technique (ESPRIT) [5] are the two most popular DOA estimation methods. In this paper, the MUSIC method is chosen to estimate the angle sources using a modified phase difference function. The MUSIC spectrum is expressed as follows:

\[
P_{\text{MUSIC}}(\theta, \omega_j) = \frac{w(\theta, \omega_j)^H w(\theta, \omega_j)}{w(\theta, \omega_j)^H E(\omega_j) E(\omega_j)^H w(\theta, \omega_j)} \tag{4}
\]

where \( P_{\text{MUSIC}}(\theta, \omega_j) \) is the reciprocal of the ordinary Euclidean distance from \( w(\theta, \omega_j) \) to \( E(\omega_j) \). \( w(\theta, \omega_j) \) is the angle steering vector, which can also be described by (2). \( E \in \mathbb{C}^{M \times (M-P)} \) is the set of eigenvectors of noise known as the noise subspace, which is calculated from the covariance matrix of \( x(\tau, \omega_j) \). One previous study suggested that the number of microphones \( M \) is larger than number of incident sources \( P \) [4]. Prior information of \( P \) can be determined from the eigenvalues [4]. The DOAs of signal sources are estimated by locating the peaks of the MUSIC spectrum for each \( \omega_j \). Because the MUSIC method has a large computational requirement, the Root MUSIC method has been proposed in order to refine the DOA estimation and to alleviate the algorithm complexity [6]. Root MUSIC with the modified phase difference function can be expressed as follows:

\[
\theta = \sin^{-1}\left(-\text{angle}(z) \frac{\lambda\omega_\lambda}{2\pi\omega_j d}\right) \tag{5}
\]

where \( \text{angle}(\cdot) \) is the phase angle operator in radians. \( z \) is the roots of the polynomial function \( f(z, \omega_j) \). The polynomial function \( f(z, \omega_j) \) can be represented as

\[
f(z, \omega_j) = \sum_{m=-M+1}^{M-1} z^m C_m(\omega_j) \quad \text{where} \quad C_m(\omega_j) = \sum_{a+b=m} V_{a,b}(\omega_j) \quad \text{and} \quad V(\omega_j) = E(\omega_j) E(\omega_j)^H \] among a roots in \( f(z, \omega_j) \), the closer a root is to the unit circle, the
more likely it corresponds to DOA. The remaining roots that are farther from the unit circle are noises. Note that the unit circle can be considered as the unit complex numbers. In general, for a comparable DOA estimation performance, Root MUSIC is more computationally efficient than MUSIC.

While the Root MUSIC method generates a large set of DOAs at each frequency and angle, unsupervised machine learning has been suggested as a potential solution to solve classification problem [9]. In this paper, a Gaussian mixture model (GMM) with a maximum likelihood estimation algorithm is utilized to cluster local maxima of angle observations into \( P \) clusters or sources, where each observation belongs to a cluster with the parametric probability density function (PDF) represented as a weighted sum of Gaussian component densities [9]. The PDF can be written as a linear superposition of Gaussian in the form

\[
p(\theta) = \sum_{p=1}^{P} \pi_p \mathcal{N}(\mu_p, \Sigma_p)
\]

where \( \theta \) is the set of DOAs obtained by the Root MUSIC method Eq.(5) without a frequency space, \( \pi_p, p = 1, 2, ..., P \) are the mixture weights, \( \mathcal{N}(\mu_p, \Sigma_p) \) is the multivariate normal distribution or multivariate Gaussian distribution, \( \Sigma_p, p = 1, 2, ..., P \) are DOA covariance matrices and \( \mu_p, p = 1, 2, ..., P \) are the estimated DOA. The estimated number of sources \( P \) is defined by the maximum of the likelihood function of the observation in GMM algorithm, which is obtained by changing the \( P \) free parameters in Eq.(6).

4. Signal Synthesis

Once the DOAs of the \( P \) incident signals have been estimated, \( s(\tau_k, \omega_j) \) can be estimated for synthesis of the signal sources by using Eq.(3) as follows:

\[
s(\tau_k, \omega_j) = (W(\theta_{DOA}, \omega_j))^{-1} (x(\tau_k, \omega_j) - n(\tau_k, \omega_j))
\]

where \( \theta_{DOA} \) is the DOA of sources, which is obtained from \( \mu_p \) in Eq.(6). The sources in the time domain can be estimated by the inverse STFT of \( s(\tau_k, \omega_j) \).

5. Numerical Simulation

The proposed system was tested by computer simulation considering a 16-sensor ULA. The statistical performance was evaluated by performing 100 Monte Carlo runs for each scenario. Three far-field uncorrelated sources with the same power were placed at 0\(^\circ\), 10\(^\circ\) and 20\(^\circ\). Source frequencies were 1, 1.75 and 2.5 kHz, respectively. The output of each microphone was decomposed into 192000 snapshots. The sampling frequency was thus 192 kHz, the spacing of microphone elements was 5 cm and speed of sound was assumed to be 340 m/s.

Figure 2: Estimation results of the proposed method by (a) Root MUSIC and GMM fusion and (b) MUSIC.

Three other methods were tested for comparison: the generalized cross-correlation phase transform (GCC-PHAT) [2], the conventional CSS method [7] and a modification of the total least-squares CSS method (CSS-MTLS) [8]. For the proposed method, 218 frequency bins were utilized in the range from \( \frac{\omega_\Lambda}{2\pi} \) to \( \omega_\Lambda \), where \( \omega_\Lambda \) was 3.4 kHz. The conventional CSS method and CSS-MTLS used 435 frequency bins in the range from \( f_0 - \frac{BW}{4} \) to \( f_0 + \frac{BW}{4} \), where the central frequency \( f_0 \) was 3.4 kHz, bandwidth \( BW \) was 5.1 kHz and the preliminary approximate DOA was 10\(^\circ\). Both the CSS method and CSS-MTLS were recursively processed five times to update the focusing DOA and the preliminary approximate DOA. For GCC-PHAT, only two microphones were selected with a maximum spacing of microphone elements of 75 cm.

Figures 2(a) and 2(b) show examples of the proposed method output for a 16-sensor ULA model with three human voice sources at angles of 0\(^\circ\), 10\(^\circ\) and 20\(^\circ\). The example indicates that there are three sources because one can see three peaks at the correct DOAs. Note that the magnitude of the parametric PDF is not necessarily proportional to the power of the signal sources.

Figure 3 compares the performance of the four methods in terms of the root-mean-square error (RMSE) of the sources at 0\(^\circ\), 10\(^\circ\) and 20\(^\circ\), over the signal-to-noise ratio (SNR) range from -10 to 40 dB. The results indicate that the performance of CSS-MTLS is not significantly dominated by the SNR, whereas the other three methods are significantly dominated by the SNR. As the SNR decreases to less than -3 dB,
CSS-MTLS exhibits better performance than the proposed method and the other two methods. The proposed method could not resolve the issue caused by higher variance in the GMM. When the SNR is as high as -3 dB, the proposed method begins to exhibit a smaller RMSE than CSS-MTLS and the other two methods. The results indicate that the proposed method exhibits high performance in the whole range of SNRs in terms of RMSE while CSS-MTLS exhibits high performance at high SNRs. The RMSE performances for the conventional CSS method and GCC-PHAT are similar.

6. Conclusions

This paper presented an alternative DOA estimation method for wide-band sources by using a GMM with a maximum likelihood estimation algorithm for multiple frequencies and source angle estimation. The proposed method is compatible with classical subspace-based methods, such as, conventional and Root MUSIC. The investigation of a modified microphone array data model indicates that confusion caused by a lower source frequency in ULA radiation can be avoided by increasing the number of microphone elements. The performance evaluation results suggest that the proposed method may be a particularly effective method of DOA estimation. Furthermore, the proposed method enables the synthesis of signal sources and provides an alternative to intelligent source localization systems.

References

[1] S. A. Zekavat and R.M. Buehrer: Handbook of Position Location: Theory, Practice, and Advances, Wiley-IEEE Press, 2011.
[2] T. Nishiura, M. Nakamura, A. Lee, H. Saruwatari and K. Shikano: Talker tracking display on autonomous mobile robot with a moving microphone array, ICAD2002: 8th International Conference on Auditory Display, July 2002.
[3] Y. Sasaki, M. Kabasawa and S. Thompson: Spherical microphone array for spatial sound localization for a mobile robot, IEEE/RSJ International Conference on Intelligent Robots and Systems, 2012.
[4] R. Schmidt: Multiple emitter location and signal parameter estimation, IEEE Transactions on Antennas and Propagation, Vol. 34, No. 3, pp. 276-280, 1986.
[5] R. Roy and T. Kailath: Estimation of signal parameters via rotational invariance techniques, IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 37, No. 7, pp. 984-995, 1989.
[6] B. D. Rao and K. V. S. Hari: Performance analysis of Root-MUSIC, IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 37, No. 12, pp. 1939-1949, 1989.
[7] H. Wang and M. Kaveh: Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wide-band sources, IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 33, No. 4, pp. 823-831, 1985.
[8] S. Valaee, B. Champagne and P. Kabal: Localization of wideband signals using least-squares and total least-squares approaches, IEEE Transactions on Signal Processing, Vol. 47, No. 5, pp. 1213-1222, 1999.
[9] C. M. Bishop: Pattern Recognition and Machine Learning, Springer, 2006.