A New Calibration Method for Microphone Array with Gain, Phase, and Position Errors
Hua Xiao, Huai-Zong Shao, and Qi-Cong Peng

Abstract—Microphone array can be used in sound source localization and separation. But gain, phase, and position errors can seriously influence the performance of localization algorithms such as multiple signal classification (MUSIC) algorithm. In this paper, a new calibration method for microphone array with gain, phase, and position errors is proposed. Unlike traditional calibration methods for antenna array, the proposed method can be used in the broadband and near-field signal model such as microphone array with arbitrary sensor geometries in one plane. Computer simulations are presented and simulation results show the new method having good performance.

Index Terms—Calibration, microphone array, multiple signal classification (MUSIC).

1. Introduction

Multiple signal classification (MUSIC) algorithm[1] based on eigendecomposition has high resolution. If the algorithm is used in microphone array, it can realize sound source localization[2][6]. Computer simulations have verified high performance. However, gain, phase, and position errors exist in actual microphone array. MUSIC algorithm requires precise knowledge of the signals received by the sensor array from a standard source located at any direction[7], the performance of this algorithm used in actual microphone array degrades seriously. It is necessary to calibrate microphone array before implementing sound source localization algorithm.

A few calibration methods have been proposed. Benjamin Friedlander and Anthony J. Weiss[7] developed a self-calibration method based on eigendecomposition. It does not need exact source positions, but requires a lot of computation and has a long convergent time. Zhang and Zhu[8] presented an improved self-calibration method. Compared to the former, it has less computation. But the performance is inferior to active-calibration method. Jia, Bao, and Wu[9] presented an active-calibration method which can estimate gain, phase, and position errors with high accuracy. But it can only be used in narrowband and far-field signal model. By now, the presented calibration methods aim at antenna array whose signal model is narrowband and far-field. However, the signal model of microphone array is broadband and near-field[10]. In this paper, an active-calibration method for microphone array with gain, phase, and position errors is proposed. Microphone array receives only one speech signal in different positions that is known exactly. The frequency with maximum amplitude is chosen to calibrate and locate. Gain, phase errors, and element positions can be estimated. The gain and phase errors in other frequencies can be worked out by repeating the same procedure. After calibration, the performance of MUSIC algorithm used in microphone array is highly improved.

2. Problem Formulation

Consider a microphone array with \( N \) elements and \( M \) uncorrelated and directional sound sources in near-field scale. Each element receives white noise with the same power. Consider the first element as origin of planar rectangular coordinates and geometry center of microphone array as reference point. The beeline between the reference point and origin is \( x \)-axis. According to this coordinates system, denote the coordinates of each element as \((x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)\), the coordinates of sound sources as \((x'_1, y'_1), (x'_2, y'_2), \ldots, (x'_M, y'_M)\) and reference point as \((x_c, y_c)\), respectively. Some factors are considered in this paper: speech signal is nonstationary and only has short time stationarity; short time Fourier transform (STFT) is used for received signal, and the data in frequency domain is generated; the amplitudes of signals received by each element at the same time are different, they are inverse proportion to propagation distance. Steering vector of the \( i \)th \((i = 1, 2, \ldots, M)\) sound source is described by

\[
a_i(\theta, r, f) = \left[ \frac{r_1}{r_1} e^{-j2\pi f \Delta t_1}, \frac{r_2}{r_2} e^{-j2\pi f \Delta t_2}, \ldots, \frac{r_N}{r_N} e^{-j2\pi f \Delta t_N} \right]^T
\]
where $\theta_i$ is the $i$th sound source azimuth angle, $r$ is distance between the $i$th sound source and the reference point, $f$ is frequency,

$$r_j = \sqrt{(x'_j - x_j)^2 + (y'_j - y_j)^2}, \quad (j = 1, 2, \cdots, N) \tag{2}$$

is the distance between the $i$th sound source and each element, and

$$\Delta t_j = (r_j - r)/c \tag{3}$$

is time difference, where $c$ is sound velocity.

The STFT of input vector is described by

$$Z(f) = \sum_{i=1}^{M} a_i(\theta, r, f)s_i(f) + N(f)$$

$$= AS(f) + N(f) \tag{4}$$

where

$$A = [a_1(\theta, r, f), a_2(\theta, r, f), \cdots, a_M(\theta, r, f)]$$

is the steering matrix,

$$Z(f) = [z_1(f), z_2(f), \cdots, z_N(f)]^T$$

is the STFT vector of received signals,

$$S(f) = [s_1(f), s_2(f), \cdots, s_M(f)]^T$$

is the STFT vector of source signals, and

$$N(f) = [n_1(f), n_2(f), \cdots, n_N(f)]^T$$

is the STFT vector of noise. While microphone array errors exist, the $i$th steering vector is converted to

$$a'_i(\theta, r, f) = \Gamma(f)$$

$$= \begin{bmatrix}
    r e^{-j2\pi f \Delta t_1} \\
    r e^{-j2\pi f \Delta t_2} \\
    \vdots \\
    r e^{-j2\pi f \Delta t_{N}}
\end{bmatrix}$$

where

$$\Gamma(f) = \text{diag}[1, g_1(f) e^{j\phi_1(f)}, \cdots, g_N(f) e^{j\phi_N(f)}] \tag{5}$$

In (6), $g_j(f)$ and $\phi_j(f)$ are gain and phase error of the $j$th element at frequency $f$ compared to the first element. The STFT of input vector under errors is changed into

$$Z'(f) = \sum_{i=1}^{M} a'_i(\theta, r, f)s_i(f) + N(f)$$

$$= [a'_1(\theta, r, f), a'_2(\theta, r, f), \cdots, a'_M(\theta, r, f)]S(f) + N(f)$$

$$= A'S(f) + N(f) \tag{7}$$

The problem addressed here is as follows: Given

$Z'(f)$ and sound source positions, $\Gamma(f)$ and coordinates of each element are estimated.

\section{3. Calibrating Method}

This method is based on the following assumptions: 1) each element is omni-directional, 2) gain and phase errors of all elements are constant and unrelated to direction, and 3) the positions of sound source are known exactly.

The covariance matrix under errors is described by

$$R_{xx}(f) = A'E[S(f)S^H(f)]A^H + E[N(f)N^H(f)]$$

$$= A'R_{xx}(f)A^H + R_{NN}(f) \tag{8}$$

where $R_{xx}(f)$ is source signal covariance matrix, $R_{NN}(f)$ is noise covariance matrix.

Denote eigenvalues of $R_{xx}(f)$ as $\lambda_1, \lambda_2, \cdots, \lambda_N$ (listed from large to small) and related eigenvectors as $u_1, u_2, \cdots, u_N$. Consider only one sound source in near-field scale. Denote the steering vector of this sound source as $a'$. According to theory of MUSIC algorithm, there is only one eigenvector in signal subspace. The steering vector can be linearly expressed by the signal eigenvector, i.e.

$$a' = lu_i \tag{9}$$

where $l$ is obtained by dividing the first element of $a'$ by the first element of $u_i$. Denote $lu_i$ as $u'_i$, where

$$u'_i = [q_1, q_2, \cdots, q_N]^T$$

We can obtain

$$u'_i = \begin{bmatrix}
    r e^{-j2\pi f \Delta t_1} \\
    r e^{-j2\pi f \Delta t_2} \\
    \vdots \\
    r e^{-j2\pi f \Delta t_{N}}
\end{bmatrix}$$

The phase parts in (10) follow that

$$\left[ \begin{array}{c}
    \arg(q_1) \\
    \vdots \\
    \arg(q_N)
\end{array} \right] = \left[ \begin{array}{c}
    -2\pi f \Delta t_1 + 2\pi l_1 \\
    \vdots \\
    -2\pi f \Delta t_{N} + 2\pi l_{N}
\end{array} \right] \tag{11}$$

where $\arg(\cdot) = \arctan(\text{Im}(\cdot)/\text{Re}(\cdot))$. Because the position of sound source is exactly known, we can determine $2\pi l_j (j = 1, 2, \cdots, N)$ in (11). Eliminating $2\pi l_j$ in (11), we obtain

$$P = \begin{bmatrix}
    p_1 \\
    p_2 \\
    \vdots \\
    p_N
\end{bmatrix}$$

We lay one sound source in $K(K \geq 3)$ positions at different time. Denote the coordinates of the $K$ positions
as \((x_i', y_i'), (x_2', y_2'), \ldots, (x_K', y_K')\). Thus, we can obtain \(K\) vectors in the form of (12), i.e. \(p_m\) \((m = 1, 2, \ldots, K)\).

Subtracting \(p_{m+1}\) from \(p_m\) \((m = 2, 3, \ldots, K)\), we obtain

\[
p_{m+1} - p_m = \begin{bmatrix}
2\pi f & \Delta t_1^{(m)} - 2\pi f & \Delta t_1^{(m-1)} \\
2\pi f & \Delta t_2^{(m)} - 2\pi f & \Delta t_2^{(m-1)} \\
\vdots & \vdots & \vdots \\
2\pi f & \Delta t_N^{(m)} - 2\pi f & \Delta t_N^{(m-1)} 
\end{bmatrix}
\] (13)

where

\[
\Delta t_j^{(m)} = \frac{1}{c} \sqrt{(x_m' - x_j)^2 + (y_m' - y_j)^2}
\]

For the \(j\)th element of array, it follows that

\[
\begin{bmatrix}
2\pi f & \Delta t_1^{(j)} - 2\pi f & \Delta t_1^{(j-1)} \\
2\pi f & \Delta t_2^{(j)} - 2\pi f & \Delta t_2^{(j-1)} \\
\vdots & \vdots & \vdots \\
2\pi f & \Delta t_N^{(j)} - 2\pi f & \Delta t_N^{(j-1)} 
\end{bmatrix} \begin{bmatrix}
p_1(j) - p_2(j) \\
p_2(j) - p_3(j) \\
\vdots \\
p_{m-1}(j) - p_m(j)
\end{bmatrix}
\] (15)

In (15), \(p_m\), \((x_m', y_m')\) \((m = 1, 2, \ldots, K)\), and \((x_j', y_j')\) are known. While \(K=3\), equation (15) has only one solution that is the coordinates of the \(j\)th element. While \(K \geq 3\), we select every different combination of two equations from (15) to get solution, then add all possible solutions and average them as final result.

According to (12) and above conditions, we obtain

\[
\begin{bmatrix}
\varphi_1(f) \\
\varphi_2(f) \\
\vdots \\
\varphi_N(f)
\end{bmatrix} = p_m + \begin{bmatrix}
2\pi f & \Delta t_1^{(m)} \\
2\pi f & \Delta t_2^{(m)} \\
\vdots \\
2\pi f & \Delta t_N^{(m)}
\end{bmatrix}
\] (16)

Substituting the elements’ coordinates and \(p_m\) into (16), we can solve phase errors \(\varphi_j(f)\) at frequency \(f\).

The absolute value of each side in (10) follows that

\[
\begin{bmatrix}
g_1(f) \\
g_2(f) \\
\vdots \\
g_N(f)
\end{bmatrix} = \begin{bmatrix}
|g_1|r_1/r \\
|g_2|r_2/r \\
\vdots \\
|g_N|r_N/r
\end{bmatrix}
\] (17)

Substituting the elements’ coordinates and \(u_j'\) into (17), we can solve gains at frequency \(f\).

### 4. Simulation Result

Consider a uniform linear microphone array with 8 elements. Ideal distance between each element is 0.08 m. Number of snapshot is 752 and length of frame is 256. One sound source with SNR of 27 dB is located in 73.74°, 106.26°, and 163.7° at different time.

Table 1 shows the actual value of gain and phase errors at 1000 Hz and elements’ coordinates. Estimate value of gain and phase errors and elements’ coordinates are also given in the table.

**Table 1**

<table>
<thead>
<tr>
<th>Array element</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>x-axis actual value (m)</td>
<td>0.085</td>
<td>0.1612</td>
<td>0.2354</td>
<td>0.3223</td>
<td>0.3989</td>
<td>0.4814</td>
<td>0.5569</td>
<td></td>
</tr>
<tr>
<td>x-axis estimate value</td>
<td>0.086</td>
<td>0.1609</td>
<td>0.2352</td>
<td>0.3222</td>
<td>0.3992</td>
<td>0.4819</td>
<td>0.5579</td>
<td></td>
</tr>
<tr>
<td>y-axis actual value (m)</td>
<td>0.0024</td>
<td>0.0036</td>
<td>-0.0015</td>
<td>0.0041</td>
<td>0.0012</td>
<td>-0.0035</td>
<td>0.0011</td>
<td></td>
</tr>
<tr>
<td>y-axis estimate value</td>
<td>0.0025</td>
<td>0.0027</td>
<td>-0.0020</td>
<td>0.0038</td>
<td>0.0012</td>
<td>-0.0034</td>
<td>0.0015</td>
<td></td>
</tr>
<tr>
<td>Gain</td>
<td>1.097</td>
<td>1.12</td>
<td>1.109</td>
<td>0.953</td>
<td>1.212</td>
<td>0.945</td>
<td>1.167</td>
<td></td>
</tr>
<tr>
<td>Gain estimate</td>
<td>1.06</td>
<td>1.12</td>
<td>1.106</td>
<td>0.951</td>
<td>1.212</td>
<td>0.948</td>
<td>1.170</td>
<td></td>
</tr>
<tr>
<td>Phase error (°)</td>
<td>6.5</td>
<td>3.9</td>
<td>4.8</td>
<td>-5.2</td>
<td>7.754</td>
<td>1.778</td>
<td>-6.12</td>
<td></td>
</tr>
<tr>
<td>Phase estimate</td>
<td>5.7</td>
<td>4.2</td>
<td>4.9</td>
<td>-5.5</td>
<td>7.340</td>
<td>1.842</td>
<td>-5.78</td>
<td></td>
</tr>
</tbody>
</table>

It can be concluded from Table 1 that this calibration method estimates gain accurately and phase errors with a bit deviation. Simulations show that the estimate accuracy advances with increase of SNR. Simulations of other frequencies also show this method to be effective.

**Table 2**

<table>
<thead>
<tr>
<th>Result of localization before and after calibration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance (m)</td>
</tr>
<tr>
<td>Actual position</td>
</tr>
<tr>
<td>Before calibration</td>
</tr>
<tr>
<td>After calibration</td>
</tr>
</tbody>
</table>

Table 2 shows the actual position of sound source and localization results before and after calibration. It is clear that gain, phase, and position errors have an enormous effect on MUSIC algorithm. After calibration, localization precision is improved. Fig. 1 shows the localization result with microphone array errors of Table 1 before calibration. The peak of spectrum is low and the position is inaccurate, MUSIC algorithm is disturbed. Fig. 2 shows the localization result after calibration. It can be observed that the spectrum peak is sharp and its value is much larger than that in Fig. 1. The performance of MUSIC is improved highly.
5. Conclusions

In this paper, a new calibration method for microphone array with gain, phase, and position errors is proposed. It is different from traditional calibration methods for antenna array. It can apply to array with broadband and near-field signal model and can calibrate gain, phase, and position errors. Computer simulations confirm effectiveness of this method. After calibration, the performance of MUSIC algorithm used in microphone array is highly improved.

References


Hua Xiao was born in Sichuan Province, China, in 1983. He graduated from University of Electronic Science and Technology of China (UESTC) in 2005. He is currently pursuing the M.E. degree with the School of Communication and Information Engineering, UESTC. His research interests include array and speech signal processing. In 2006 TI China DSP Design Challenge, he received the 3rd Class Prize in algorithm group.

Huai-Zong Shao received his M.E. degree from Sichuan University in 1998 and Ph.D. degree from UESTC in 2003. He has published over 30 academic papers. He is currently an associated professor with the School of Communication and Information Engineering, UESTC. His research interests include wireless communication technology and communication signal processing.

Qi-Cong Peng graduated from the Department of Radio Electronics, Tsinghua University and M.E. degree from UESTC. He received the 3rd Prize of National Science and Technology Progress in 1997 and the 2nd Prize of National Teaching Achievement in 2001. He has published over 60 academic papers. He is currently a professor with UESTC. His research interests include signal processing in communication and information system, real-time and high speed signal processing, and DSP technology.