

DOI: 10.13382/j.jemi.B2407433

多麦克风阵列数据融合声源定位方法研究*

刘扬 张传营 赵景玉 卜凡亮

(中国人民公安大学信息安全学院 北京 102600)

摘要:针对TDOA估计精度受采样频率限制的问题以及单一麦克风阵列声源定位误差较大的问题,提出了两种改进方法并进行了仿真验证。首先,为了解决采样频率低导致的TDOA估计误差较大的问题,提出了一种基于三次样条插值的改进二次互相关算法进行TDOA估计的方法,在第1次互相关后进行三次样条插值提高采样频率,之后进行第2次互相关求得TDOA。在10倍插值的条件下仿真表明该方法将TDOA误差从7.6%降低到0.6%,有效的提升了估计精度。其次,为了解决单个麦克风阵列声源定位误差较大的问题提出了一种多麦克风阵列数据融合的声源定位方法,以四元十字麦克风阵列为单元,首先在每个单元阵列上进行单独的声源定位,之后将多个单元阵列的定位结果进行融合得到最终的声源位置,选取空间中的10个点利用双麦克风阵列以及四麦克风阵列对比单麦克风阵列进行远场定位仿真,结果表明,单麦克风阵列在声源距离较远时的定位误差达1 m以上,而利用所提方法设计的双麦克风阵列的定位距离误差在0.3 m以下,四麦克风阵列的定位距离误差在0.2 m以下,大大提高了声源定位的精度。

关键词: 到达时间差;麦克风阵列;声源定位;三次样条插值;互相关

中图分类号: TN912.3 **文献标识码:** A **国家标准学科分类代码:** 510.40

Research on sound source localization method using data fusion of multiple microphone arrays

Liu Yang Zhang Chuanying Zhao Jingyu Bu Fanliang

(College of Information and Network Security, People's Public Security University of China, Beijing 102600, China)

Abstract: Addressing the issues of TDOA estimation accuracy being limited by sampling frequency and the large localization errors of a single microphone array, this paper proposes two improved methods and conducts simulation verification. Firstly, to tackle the problem of significant TDOA estimation errors caused by low sampling frequency, a modified cross-correlation algorithm based on cubic spline interpolation is proposed. This method involves performing cubic spline interpolation after the first cross-correlation to increase the sampling frequency, followed by a second cross-correlation to obtain the TDOA. Simulations under 10x interpolation conditions demonstrate that this method reduces the TDOA error from 7.6% to 0.6%, effectively minimizing the estimation error. Secondly, to address the large localization errors associated with a single microphone array, a multi-microphone array data fusion method for sound source localization is proposed. This approach involves fusing the localization results of multiple microphone arrays to obtain the final sound source position. Using a quaternary cross array as an example, simulations were conducted at 10 points in space, comparing far-field localization accuracy between dual-microphone arrays, quad-microphone arrays, and single microphone arrays. The results show that the localization error for a single microphone array exceeds 1 meter when the sound source is distant, whereas the proposed dual-microphone array design achieves localization errors below 0.3 meters, and the quad-microphone array achieves errors below 0.2 meters, significantly enhancing the accuracy of sound source localization.

Keywords: time difference of arrival; microphone array; sound source localization; cubic spline interpolation; cross-correlation

0 引言

随着科技的发展,基于麦克风阵列的声源定位算法的研究得到重视,在远程视频会议、汽车鸣笛监测、机械部件异常声音识别等领域问题的探索越来越深入。麦克风阵列是实现波束形成^[1-2]、语音增强^[3-4]、声源定位^[5]的重要载体,它通过接收和处理声源信号实现具体的功能。麦克风阵列通常具有一定的方向性,即在特定方向上对声音更为敏感。这种方向性可以通过麦克风的布置方式和信号处理技术来调节,从而实现对特定方向声源的定位或捕捉。

声源定位方法主要有基于到达时间差(time difference of arrival, TDOA)的声源定位方法^[6-8]、基于高分辨率谱估计的声源定位方法^[9]和基于波束形成的声源定位方法^[10]。其中基于 TDOA 的方法计算成本低,是一种基于声音信号在多个麦克风之间到达时间差异的声源定位技术。计算 TDOA 的方法有很多,最常见的是互相关算法^[11](cross correlation algorithm, CC)、广义互相关算法^[12-13](generalized cross correlation algorithm, GCC)、二次互相关算法^[14](second-order cross-correlation algorithm, SCC)。这些方法落实到计算机仿真平台上都有一个共同的局限性,即估计精度与采样频率直接相关。且对信噪比较低的环境下误差明显,而在通过 TDOA 进行位置解算时,微小的误差可能就会造成巨大的定位误差。为了解决这个问题,提出了对二次互相关算法进行改进,首先让原始语音信号进行互相关计算,然后对互相关的结果进行插值,提高采样精度,最后进行二次互相关计算得到时延差,这种方法明显提升算法在较低信噪比水平下时延估计的正确率的同时也保证了估计精度,避免微小误差对定位结果的巨大影响。

对于位置解算方法,目前主要有迭代算法和非迭代算法。其中迭代算法包括 Taylor 算法、最小二乘法^[15]等,非迭代算法包括几何解法^[16-17]、Chan 算法^[18]等。迭代算法较为复杂,而且对初始值的选择有比较大的依赖性,当迭代的初始值选择不合适时,最后迭代的结果与真实值会有较大偏差。非迭代算法几何算法可以通过麦克风阵列之间位置关系得到最终的声源位置,计算过程简单有效并且同时适用于一些平面阵列和三维阵列。但是这种方法可能会因为噪声等环境条件导致定位的巨大偏差,为解决此问题,可以利用多个麦克风阵列同时对声源进行定位^[19],首先每个子麦克风阵列分别进行声源位置解算,然后通过求解出的多个声源位置进行数据融合得到融合后的声源位置,有效地解决了使用单一阵列进行定位的局限性,提高了声源定位的精度。

综上所述,该研究提出了一种基于三次样条插值的

二次互相关 TDOA 估计方法和基于多麦克阵列数据融合的声源定位算法。改进的二次互相关算法有效提升了时延估计方法的准确性,实现了在较低信噪比环境下进行高精度的时延估计。基于多麦克阵列数据融合的声源定位算法有效提升了单一阵列定位的准确度。通过推导双麦克风阵列和四麦克风阵列的融合定位方法并进行 MATLAB 软件仿真,证明了所提出的声源定位方法简单有效,实际应用价值高。

1 麦克风阵列接收模型

在基于 TDOA 的声源定位算法中,利用声音信号在不同麦克风之间到达的时间差来确定声源的位置。利用两两麦克风之间的声源信号到达时间差,结合特定算法就可以得到声源具体位置。在存在噪声的环境下,以四元十字麦克风阵列为例,麦克风 i 接收到的信号为:

$$x_i(t) = s(t - \tau_{ii}) + n_i(t) \quad (1)$$

其中, $s(t)$ 为麦克风接收的声源信号, τ_{ii} 为其余麦克风相对于参考麦克风 1 的时延差, n_i 表示麦克风接收到的噪声信号。

以四元十字阵列中的两个麦克风 M_1 和 M_2 为例,如图 1 所示, t_1 和 t_2 分别表示声源到麦克风 M_1 和 M_2 的时间, c 表示声速,一般取 340 m/s。则这两个麦克风的接收模型:

$$x_1(t) = s(t) + n_1(t) \quad (2)$$

$$x_2(t) = s(t - \tau_{12}) + n_2(t) \quad (3)$$

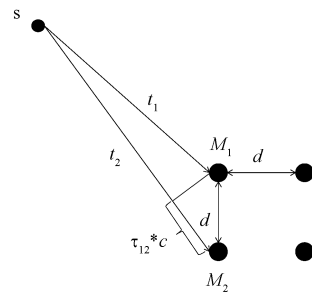


图 1 麦克风阵列接收模型

Fig. 1 Microphone array signal reception model

2 基于三次样条插值的二次互相关时延估计算法

时延估计常用的方法是互相关技术,其中互相关算法以及二次互相关算法是最主要的方法。但是实际操作中对于采样频率具有较高的要求,采样频率会严重影响时延估计的精度,影响互相关的性能。利用插值算法,可以提升信号的采样频率。使得 TDOA 估计更加精准,避

免位置解算的较大误差。

2.1 互相关算法

在时域上,互相关函数表示一个信号与另一个信号的延迟版本之间的相似程度。互相关算法(cross-correlation, CC)首先计算互相关函数;将两个信号进行卷积操作,产生互相关函数,用于衡量两个信号之间的相似性。分析互相关函数并找到其中的峰值。互相关函数的峰值位置对应于信号之间的最大相似性,因此可以用来估计信号之间的时延差。

两个麦克风接收信号的互相关函数为:

$$R_{x_1x_2} = E\{x_1(t)x_2(t-\tau)\} \quad (4)$$

其中, E 代表期望值运算,将式(2)和(3)代入式(4)

可得:

$$R_{x_1x_2} = E\{(s(t) + n_1(t))(s(t - \tau - \tau_{12}) + n_2(t - \tau))\} \quad (5)$$

通常情况下,假设声源 s 与噪声 n_1 、 n_2 均两两互不相关,式(5)可直接改写为:

$$R_{x_1x_2} = R_{ss}(\tau - \tau_{12}) \quad (6)$$

由互相关函数的性质可以得到,当 $\tau = \tau_{12}$ 时函数 $R_{x_1x_2}$ 取得最大值,且正好对应所需要的时延估计值。因此可以通过峰值搜索找到 $R_{x_1x_2}$ 最大值对应的自变量值得到所求的时延差。

2.2 广义互相关加权算法

广义互相关算法是对传统互相关算法的扩展,用于处理非平稳信号和噪声环境下的信号处理问题。传统的互相关算法假设信号是平稳的,但在实际应用中,许多信号都是非平稳的,并且受到噪声的影响。广义互相关算法通过引入加权函数来对非平稳信号进行加权处理,以增强信号之间的相关性,并抑制噪声的影响。

互相关函数与互功率谱函数的关系可以表示为:

$$R_{x_1x_2}(\tau) = \int_0^\pi G_{x_1x_2}(\omega) e^{-j\omega\tau} d\omega = \int_0^\pi X_1(\omega) X_2^*(\omega) e^{-j\omega\tau} d\omega \quad (7)$$

为了避免噪声对互相关函数峰值的影响,提高时延估计的准确性,可以利用先验知识在频域上加权信号和噪声,以增强互相关函数的峰值:

$$R_{x_1x_2}(\tau) = \int_0^\pi \varphi_{12}(\omega) X_1(\omega) X_2^*(\omega) e^{-j\omega\tau} d\omega \quad (8)$$

其中, $\varphi_{12}(\omega)$ 是频域条件下的加权函数。加权函数有许多形式,其中最常用的一种加权函数为相位变换(phase transformation)加权函数,其表达式如式(9)所示。

$$\varphi_{12}(\omega) = \frac{1}{|G_{x_1x_2}(\omega)|} \quad (9)$$

此方法全称为广义互相关-相位变换加权算法(generalized cross-correlation with phase transform, GCC-PHAT)。

2.3 基于插值的改进二次互相关算法

在信噪比较低的情况下,互相关算法的结果的峰值会受到明显影响,实验证明二次互相关算法可以有效减少噪声的影响。另一方面,时延估计的精度会收到采样频率的影响。远场条件下,麦克风阵列中的阵元间距通常比声源到麦克风阵列的距离小的多, TDOA 估计是一个很小的量,所以需要较高的采样频率进行处理。通常情况下,直接增加采样频率会对计算机硬件有较高的要求,而且运算速度很慢,运用插值算法可以提高时域的采样率,且相比直接提升采样率速度更快。

在假设信号与噪声两两互不相关的条件下,针对两路信号的二次互相关模型如式(10)和(11)所示。

$$R_{x_1x_1} = E\{x_1(t)x_1(t-\tau)\} = R_{ss}(\tau) \quad (10)$$

$$R_{x_1x_2} = E\{x_1(t)x_2(t-\tau)\} = R_{ss}(\tau - \tau_{12}) \quad (11)$$

然后再对以上两式进行互相关运算:

$$R_{R_{x_1x_1}R_{x_1x_2}} = E\{R_{x_1x_1}(t)R_{x_1x_2}(t-\tau)\} \quad (12)$$

最终可以推导出:

$$R_{R_{x_1x_1}R_{x_1x_2}} = R_{R_s}(\tau - \tau_{12}) \quad (13)$$

常用的插值算法有线性插值、拉格朗日插值、三次样条插值等,线性插值对于非线性变化的函数性能一般,拉格朗日插值在数据点较多的情况下会出现明显的龙格现象,本研究以三次样条插值为例进行时延估计,提出了基于三次样条插值的二次互相关算法(secondary cross-correlation algorithm based on cubic spline, SCC-CSI),实际中可以续安泽其他插值方法去有效的提高采样率,如图2所示。

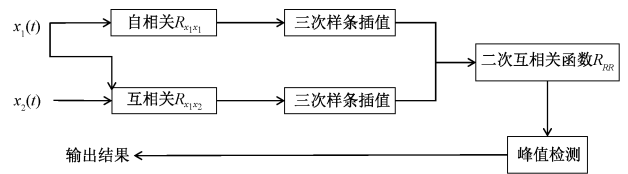


图2 基于三次样条插值的二次互相关算法

Fig. 2 Interpolation based quadratic cross-correlation algorithm

互相关之后可以通过峰值检测得到最后的 TDOA 估计值,本文在二次互相关之前的两路信号进行三次样条插值处理,提升信号的采样频率。

3 基于改进 TDOA 估计的多麦克风阵列声源定位方法

在改进 TDOA 算法以后需要进一步利用这个结果进行声源的位置解算,该研究进一步提出了一种多麦克风阵列数据融合的定位方法,致力于解决单一阵列声源定位结果不准确的问题,通过将阵列分解为多个子阵列,在全局阵上列建立全局坐标系以及在子阵列上建立子坐标

系,通过旋转平移等操作建立坐标系的变换关系。在子坐标系下进行声源定位后变换到全局坐标系下进行数据融合,得到融合后的坐标位置。在研究中用四元十字阵列作为子阵列分析多麦克风阵列数据融合方法的可行性,其余形状的麦克风阵列定位依旧适用。

3.1 四元十字麦克风阵列声源定位

四元十字麦克风阵列声源定位算法中,将 4 个麦克风全部置于坐标轴上,环境模型如图 3 所示。通过该阵列进行声源定位的方法主要利用 TDOA 估计值进行几何运算。每个麦克风距离坐标原点的长度设置为 d ,4 个麦克风在坐标轴上分别固定在 $(d,0,0)$, $(0,d,0)$, $(-d,0,0)$, $(0,-d,0)$ 。声源到麦克风 M_i 和 M_j 的时延差用 τ_{ij} ($i,j=1,2,3,4$) 表示, θ ($0^\circ \leq \theta \leq 90^\circ$), φ ($0^\circ \leq \varphi \leq 360^\circ$) 分别表示为声源的俯仰角和方位角。 C 用表示声速。

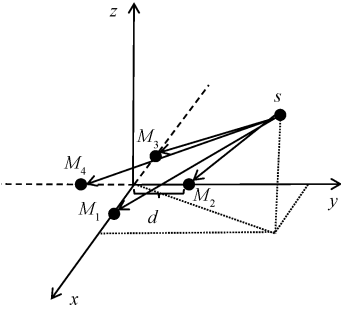


图 3 四元十字麦克风阵列建模

Fig. 3 Modeling of quaternion cross-shaped microphone array

根据图 3 建立的四元十字麦克风阵列模型,得到以下方程组:

$$\begin{cases} x^2 + y^2 + z^2 = r^2 \\ (x-d)^2 + y^2 + z^2 = r_1^2 \\ x^2 + (y-d)^2 + z^2 = r_2^2 \\ (x+d)^2 + y^2 + z^2 = r_3^2 \\ x^2 + (y+d)^2 + z^2 = r_4^2 \end{cases} \quad (14)$$

其中:

$$\begin{cases} r_3 - r_1 = \tau_{13}C \\ r_2 - r_1 = \tau_{12}C \\ r_4 - r_1 = \tau_{14}C \end{cases} \quad (15)$$

$$\begin{cases} x = r \sin \theta \cos \varphi \\ y = r \sin \theta \sin \varphi \\ z = r \cos \theta \end{cases} \quad (16)$$

其中, r_1, r_2, r_3, r_4 分别表示声源到 4 个麦克风的距离,整理方程可得:

$$r = \frac{C(\tau_{31}^2 - \tau_{41}^2 - \tau_{21}^2)}{2(\tau_{21} + \tau_{41} - \tau_{31})} \quad (17)$$

$$\tan \varphi = \frac{y}{x} = \frac{(r_4 - r_2)(r_4 + r_2)}{(r_3 - r_1)(r_3 + r_1)} \quad (18)$$

考虑远场状态下的模型,因此 $\frac{r_4 + r_2}{r_3 + r_1} \approx 1$, 则上式可近似为:

$$\tan \varphi \approx \frac{\tau_{14} - \tau_{12}}{\tau_{13}} \quad (19)$$

$$\sin \theta \approx \frac{C\sqrt{\tau_{13}^2 + (\tau_{14} - \tau_{12})^2}}{2d} \quad (20)$$

得到声源到坐标轴原点的距离,方位角和俯仰角之后,根据式(16)可以解算出声源的具体位置。

3.2 数据融合方法

数据融合的关键就是找到距离两条及以上直线最近的点。首先以单麦克风阵列为研究对象建立子坐标系,以全部麦克风为研究对象建立全局坐标系;然后在子阵列上进行定位得到初始结果,并且利用坐标变换建立在全局坐标系到子坐标系的位置变换关系,之后将所有计算的初始结果变换到全局坐标系下;最后进行数据融合得到最终的结果。

坐标变换的具体操作为平移和旋转。将一个坐标系下的坐标平移到另一个坐标系下的变换公式为:

$${}^o m_s = {}^c m_s + {}^o m_o \quad (21)$$

其中, ${}^o m_s$ 为声源在原坐标轴下的坐标, ${}^c m_s$ 为声源在变换后的坐标轴的坐标, ${}^o m_o$ 为原坐标轴原点在变换后坐标轴下的坐标。

坐标轴以原点为轴逆时针旋转 β 角度则旋转后的新坐标可以表示为:

$$\begin{pmatrix} s \\ t \end{pmatrix} = \begin{pmatrix} \cos \beta & -\sin \beta \\ \sin \beta & \cos \beta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \quad (22)$$

在子坐标系进行声源定位后,得到的声源具体位置坐标结果根据坐标变换变换到全局坐标系下,在全局坐标系下,每个子坐标系的原点位置和在此子坐标系下定位的声源位置之间形成一条连线。线条的个数为子阵列的个数。若忽略各种误差影响,这些直线会在声源位置相交,但考虑噪声和计算误差的影响,这几条线的交点实际上会有所偏差,因此要对定位结果进行数据融合。通过寻找离这些直线最近的点作为最终的定位结果。

在空间中两个点即可确定一条直线,表示为 $p_s = (p_{sx}, p_{sy}, p_{sz})$ 和 $p_e = (p_{ex}, p_{ey}, p_{ez})$ 。数学中,可以用仅有一个变量的方程表示一条直线:

$$p = p_s + s \cdot u \quad (23)$$

其中, s 是变量, $u = p_e - p_s = (u_x, u_y, u_z)$, 当 $s = 0$ 时, $p = p_s$, 当 $s = 1$ 时, $p = p_e$ 。对于两条直线:

$$\begin{cases} p = p_{s_1} + a_1 \cdot u_1 \\ p = p_{s_2} + a_2 \cdot u_2 \end{cases} \quad (24)$$

假设 (x_1, y_1, z_1) 和 (x_2, y_2, z_2) 分别是上面两条直线上的点, 则:

$$\begin{cases} 1 \cdot x_1 + 0 \cdot y_1 + 0 \cdot z_1 - u_{1x} \cdot a_1 - 0 \cdot a_2 = p_{s_1x} \\ 0 \cdot x_1 + 1 \cdot y_1 + 0 \cdot z_1 - u_{1y} \cdot a_1 - 0 \cdot a_2 = p_{s_1y} \\ 0 \cdot x_1 + 0 \cdot y_1 + 1 \cdot z_1 - u_{1z} \cdot a_1 - 0 \cdot a_2 = p_{s_1z} \\ 1 \cdot x_2 + 0 \cdot y_2 + 0 \cdot z_2 - 0 \cdot a_1 - u_{2x} \cdot a_2 = p_{s_2x} \\ 0 \cdot x_2 + 1 \cdot y_2 + 0 \cdot z_2 - 0 \cdot a_1 - u_{2y} \cdot a_2 = p_{s_2y} \\ 0 \cdot x_2 + 0 \cdot y_2 + 1 \cdot z_2 - 0 \cdot a_1 - u_{2z} \cdot a_2 = p_{s_2z} \end{cases} \quad (25)$$

假设 (x_1, y_1, z_1) 和 (x_2, y_2, z_2) 为同一个点, 或者时距离两条直线最近的一个点 (x, y, z) , 据此创建向量 \mathbf{n} :

$$\mathbf{n} = \begin{bmatrix} x \\ y \\ z \\ a_1 \\ a_2 \end{bmatrix} \quad (26)$$

接着定义 \mathbf{M}, \mathbf{d} :

$$\mathbf{M} = \begin{bmatrix} 1 & 0 & 0 & -u_{1x} & 0 \\ 0 & 1 & 0 & -u_{1y} & 0 \\ 0 & 0 & 1 & -u_{1z} & 0 \\ 1 & 0 & 0 & 0 & -u_{2x} \\ 0 & 1 & 0 & 0 & -u_{2y} \\ 0 & 0 & 1 & 0 & -u_{2z} \end{bmatrix}, \mathbf{d} = \begin{bmatrix} p_{s_1x} \\ p_{s_1y} \\ p_{s_1z} \\ p_{s_2x} \\ p_{s_2y} \\ p_{s_2z} \end{bmatrix} \quad (27)$$

通过矩阵运算, 可以求得:

$$\mathbf{n} = (\mathbf{M}^T \mathbf{M})^{-1} \mathbf{M}^T \mathbf{d} \quad (28)$$

式中: \mathbf{n} 的前三维是要求解的离两条直接最近的点, 后两位可以推导出两条直线上相邻最近的两个点, 多于两条直线时, 计算方法与此一致, 扩展 \mathbf{M}, \mathbf{d} 的大小即可。

3.3 双麦克风阵列声源融合定位

得到声源到原点的距离, 方位角和俯仰角之后, 根据上述公式可以解算出声源的具体位置。但用单一四元阵列的定位结果容易受到噪声、算法准确度的影响而产生较大偏差。结合 3.2 节的数据融合方法, 用两个四元十字麦克风阵列来进行声源定位。首先建立模型, 如图 4 所示。首先在全局坐标系下建立阵列基本模型, 6 个麦克风 $M_1 \sim M_6$ 的坐标分别为 $(d, 0, 0)$ 、 $(d, d, 0)$ 、 $(0, d, 0)$ 、 $(-d, d, 0)$ 、 $(-d, 0, 0)$ 和 $(0, 0, 0)$ 。其中 d 为阵元间距。其中 M_3 和 M_6 是共有麦克风, 可根据实际情况设置麦克风位置。

下面分别在 M_1, M_2, M_3, M_6 以及 M_3, M_4, M_5, M_6 4 个麦克风上建立两个子坐标系, 使其分别满足 3.2 节四元麦克风阵列的坐标分布。然后进行数据融合的过程如图

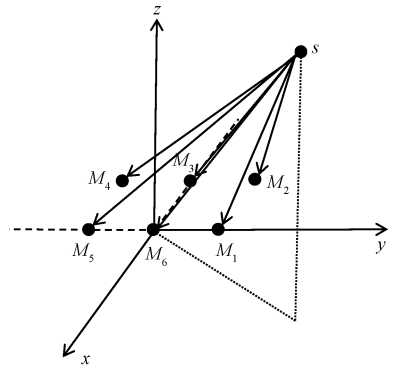


图 4 双麦克风阵列接收模型

Fig. 4 Dual microphone array reception model

5 所示。图中的 S_1 和 S_2 是两个子坐标系下分别进行声源定位并坐标转换到全局坐标系下的坐标位置数字 1 和 2 表示子阵列的编号。根据 3.2 节的理论, l_1 和 l_2 分别为全局坐标系下两个估计的声源位置和子坐标系坐标原点的连线。根据 3.3 节的数据融合方法找到离两条直线最近的点融合的位置。

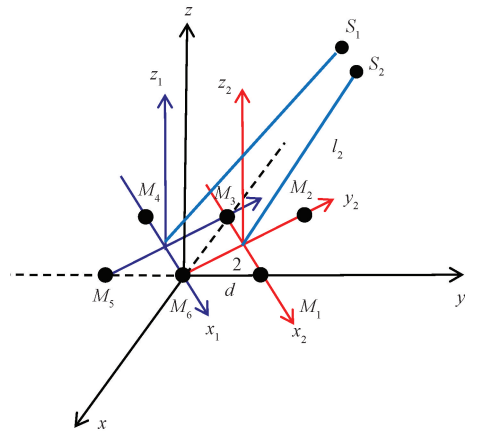


图 5 双麦克风阵列定位结果数据融合

Fig. 5 Data fusion of dual microphone array localization results

3.4 四麦克风阵列声源融合定位

建立在全局坐标系下的阵列接收模型, 全局坐标系下 9 个麦克风 $M_1 \sim M_9$ 的坐标分别为 $(0, 0, 0)$ 、 $(2d, 0, 0)$ 、 $(d, d, 0)$ 、 $(0, 2d, 0)$ 、 $(-d, d, 0)$ 、 $(-2d, 0, 0)$ 、 $(-d, -d, 0)$ 、 $(0, -2d, 0)$ 、 $(d, -d, 0)$ 。通过借鉴整合双阵列融合方法的分析过程, $M_1 M_7 M_8 M_9$ 、 $M_1 M_3 M_4 M_5$ 、 $M_1 M_5 M_6 M_7$ 、 $M_1 M_2 M_3 M_6$ 分别为 4 个子阵列, 满足 3.1 节的四元十字阵列定位方法的解算条件。平面建模如图 6 所示, 其中数字表示分解的子阵列坐标系。通过 3.3 节提到的数据融合方法得到融合位置。

3.5 总体架构

综合以上两个章节提出的基于三次样条插值的二次

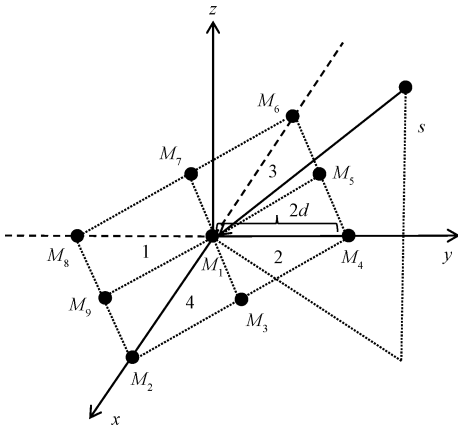


图 6 四麦克风阵列接收模型

Fig. 6 Quadruple microphone array reception mode

互相关算法改进方法以及多麦克风阵列数据融合方法, 得到该研究的总体流程。以四元十字麦克风阵列为基本子阵列单元, 首先利用基于插值的二次互相关算法对每个四元十字麦克风阵列声源定位所需要的 3 个时延差进行求解, 然后对得到的多组时延差估计值分别利用, 单独进行基本的四元十字麦克风阵列声源定位得到位置结果, 最后利用提出的多麦克风声源定位融合方法得到最终的声源位置结果。研究的总体架构如图 7 所示。

4 实验仿真分析

4.1 改进二次互相关时延估计方法性能仿真

取一段语音信号为例进行分析, 其中设置信噪比

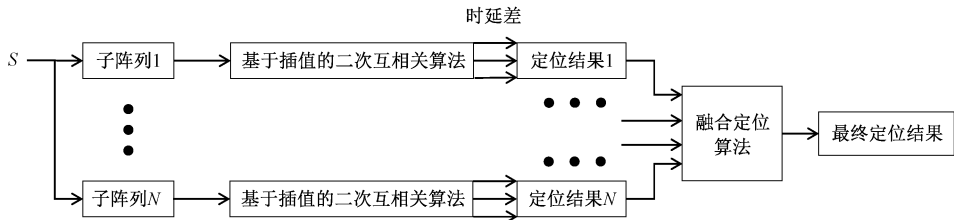


图 7 多麦克风阵列数据融合定位方法总体架构

Fig. 7 General framework for multi-microphone array data fusion localization method

SNR 为 10, 采样率为 220 50 Hz。设置两个麦克风 M_1 和 M_2 在坐标轴上, 分别固定在 $(d, 0, 0)$, $(-d, 0, 0)$ 。其中 d 取 0.1 m 声源 S 假设在 $(1.2, 0.6, 3.3)$, 本次仿真声源到两个麦克风 M_1 和 M_2 的真实时延差为 $\tau_{12} = 1.963 5 \times 10^{-4}$ s, 模型如图 8 所示。

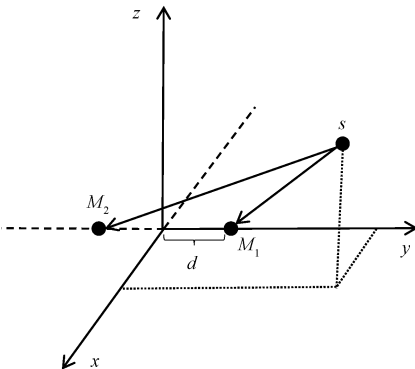


图 8 时延估计环境模型

Fig. 8 Time delay estimation environment model

值倍数为 2、5、10、15 分别进行实验, 表 1 展示在 4 种插值倍数下, M_1 和 M_2 TDOA 估计的计算值以及运算速度, 并计算时延估计与真实值的误差。

表 1 不同倍数插值下计算的时延差

Table 1 Time delay difference calculated under different interpolation factors

插值倍数	计算值/s	误差/s	运算速度/s
2	$2.040 8 \times 10^{-4}$	$7.730 0 \times 10^{-6}$	0.033 7
5	$1.995 5 \times 10^{-4}$	$3.200 0 \times 10^{-6}$	0.075 4
10	$1.950 1 \times 10^{-4}$	$1.340 0 \times 10^{-6}$	0.114 2
15	$1.965 2 \times 10^{-4}$	$1.700 0 \times 10^{-7}$	0.232 5

根据表 1 可以看出, 随着插值倍数的增大, 计算得到的时延误差也会变小, 但是运算速度也会相应的增加, 并结合实际情况下后期位置解算精度的需要, 所以本研究中三次样条插值倍数设为 10 为例进行实验。分别经过 CC、GCC-PHAT、SCC, 以及本文提出的 SCC-CSI 算法计算后的互相关函数如图 9 所示, 观察其函数形状和峰值。四种方法在不同信噪比下的 TDOA 估计的均方根误差如图 10 所示, 其中每个信噪比下进行 100 次计算。

图 9 中, 4 种方法下, 图像峰值都较为明显, CC 算法和 GCC-PHAT 算法在一定程度上受噪声影响较大, 表明, 在信噪比较低且采样率一定的条件下, SCC 和 SCC-

在此应用场景下, 为避免 TDOA 误差带来的巨大位置误差, 选取合适的三次样条插值倍数减小误差尤为重要, 理论上, 越大的插值倍数误差越小, 但是过大的插值倍数会严重影响计算的速度, 不具有实用性。为选取合适的插值倍数, 在信噪比为 30 dB 的理想情况下选取插

CSI 估计的准确度会比互相关和广义互相关加权算法高。另外,SCC-CSI 算法相比于 SCC 算法,旁瓣会更加明显,因为插值算法引入了额外的频谱成分,但是其峰值依

旧明显,然而更重要的,利用插值提高采样率以获得精度更高的结果是后续位置解算的基础。

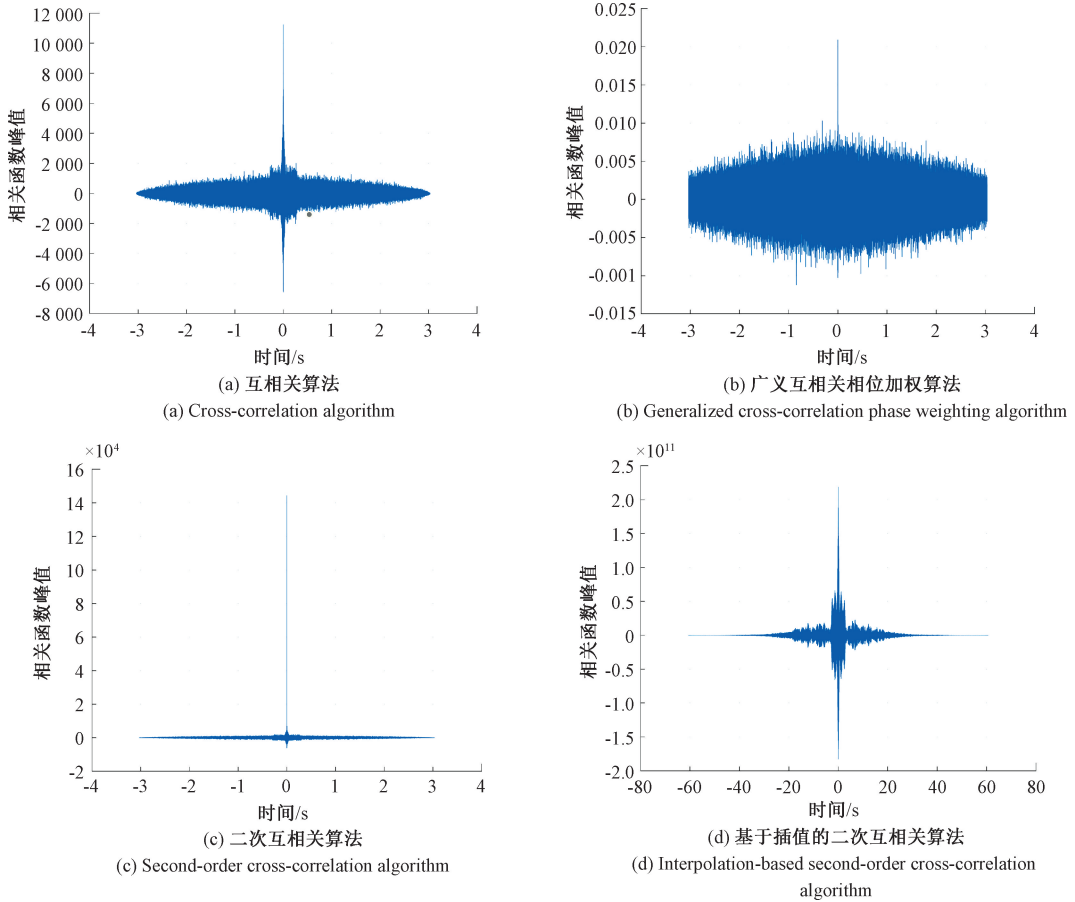


图9 4种方法的互相关函数

Fig. 9 Cross correlation functions of four methods

图10中可以看出,在同一采样率下,信噪比较低时4种方法的效果均不佳,但SCC-CSI依旧明显优于其他方法。随着信噪比增益的增加,TDOA估计的均方根误差显著下降,到信噪比15之后趋于稳定。本方法较其他方法最大的不同在于通过插值提高了采样频率,使得TDOA估计的精度显著提高,使得均方根误差明显低于其他传统方法,这一结果也验证了本方法并不受旁瓣的影响,正确率高于其他方法,便于之后精确的声源位置解算。在信噪比合适时,其他方法的时延差在 $\tau_{12} = 1.814 \times 10^{-4} \text{ s}$,SCC-CSI的结果为 $\tau_{12} = 1.950 \times 10^{-4} \text{ s}$ 。算法的误差从7.6%降低到了0.6%,显著的提高精度,可以满足之后利用时延差进行位置解算的要求。

4.2 多麦克风阵列数据融合定位方法性能仿真

依托于4.1节的方法,本节对四元十字阵列、双麦克风阵列以及四麦克风阵列的定位方法进行在MATLAB

软件上进行仿真分析,考虑远场条件下的声源定位,信噪

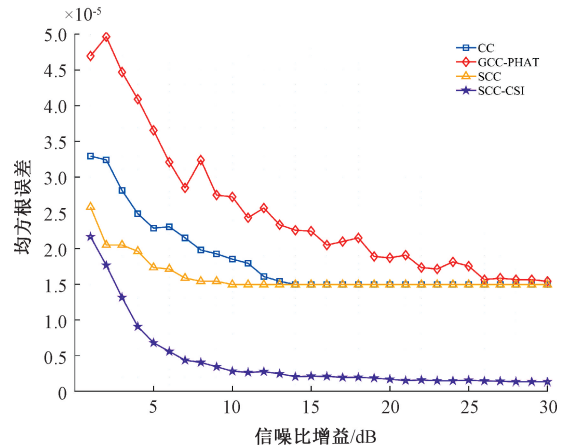


图10 不同信噪比下的时延估计均方根误差

Fig. 10 Root mean square error of time delay estimation under different signal-to-noise ratios

比设置为 $SNR = 20$ 。为方便研究,统一设置矩阵的阵元间距皆为 0.1 m ,则在四元十字阵列上的 $d = d_1 = (0.1/\sqrt{2})\text{ m}$, 双麦克风阵列中 $d = d_2 = 0.1\text{ m}$,四麦克风阵列 $d = d_3 = (0.1/\sqrt{2})\text{ m}$ 。声速为 340 m/s 。首先,在坐标系下选取 10 个不同位置进行声源位置估计,定义误差为相对于真实位置的距离误差,并进行误差分析。需要注意的是,在四元阵列的定位 $\tau_{21} + \tau_{41} - \tau_{31}$ 中由于时延本身就是一个很小的值,所以可能因为采样频率相同导致出现等于 0 的情况,造成定位结果趋于无穷大,利用插值的方法可以在不同通道重采样不同的插值倍数,消除奇异值,实现较为精确的声源定位,因此选择 3 个时延 τ_{21} 、 τ_{31} 、 τ_{41} 分

别选择插值倍数为 5、10、15 进行实验分析。四元阵列的 10 次定位结果及其定位误差分别如表 2 所示,另外取两个点 $(1, 0.3, 2.3)$, $(-0.3, 1.6, 3.8)$ 在空间范围进行仿真,结果如图 11 所示。双麦克风阵列分为两个四元矩形阵列,双麦克风阵列的定位结果如表 3 所示,对比分析了两个子阵列以及融合定位的结果。两个子阵列定位结果以及融合定位结果的空间仿真如图 12 所示。四麦克风阵列 4 个子阵列分别定位的结果以及融合后的结果如表 4 所示。四麦克风阵列定位的误差如表 5 所示。最后 4 个子阵列定位结果以及融合定位结果的空间仿真如图 13 所示。

表 2 四元十字阵列定位结果

Table 2 Localization results of quaternion cross-shaped array

真实坐标	估计坐标	坐标绝对误差	总距离误差
(1, 0.3, 2.3)	(0.852 0, 0.255 6, 1.962 8)	(-0.148 0, -0.044 4, -0.337 2)	0.370 9
(0.5, 0.8, 1.3)	(0.510 3, 0.817 5, 1.327 3)	(0.010 3, 0.017 5, 0.027 3)	0.034 0
(1.2, 0.3, 3.5)	(1.206 3, 0.300 4, 3.524 3)	(0.006 3, 0.000 4, 0.024 3)	0.025 1
(-0.3, 1.6, 3.8)	(-0.238 1, 1.228 1, 2.923 3)	(0.068 2, -0.371 9, -0.876 7)	0.954 8
(-1, 0.3, 1.6)	(-0.863 7, 0.260 1, 1.383 1)	(0.136 3, -0.039 9, -0.216 9)	0.259 3
(3.2, 1.4, 1.5)	(3.314 1, 1.449 1, 1.552 7)	(0.114 1, 0.049 1, 0.052 7)	0.134 9
(1.7, -0.8, 5.0)	(1.045 0, -0.510 4, 3.185 9)	(-0.655 0, 0.289 6, -0.1814 1)	1.950 4
(3.2, 2.4, 5.4)	(3.692 8, 2.763 3, 6.230 2)	(0.492 8, 0.363 3, 0.830 2)	1.031 6
(0.5, 1.4, 5)	(0.131 4, 0.365 6, 1.309 2)	(-0.368 6, -1.034 4, -3.690 8)	3.850 7
(2.0, 4.0, 4.5)	(3.626 7, 7.268 4, 8.159 4)	(1.626 7, 3.268 4, 3.659 4)	5.169 1

表 3 双麦克风阵列定位结果

Table 3 Localization results of dual microphone array

真实坐标	子阵列 1 估计坐标	子阵列 2 估计坐标	融合方法估计坐标	子阵列 1 误差	子阵列 2 误差	融合方法误差
(1,0.3,2.3)	(1.470 0,0.462 6,3.308 1)	(0.716 2,0.230 3,1.599 6)	(0.903 7,0.297 6,2.079 1)	1.124 1	0.758 9	0.241 0
(0.5,0.8,1.3)	(0.447 6,0.721 1,1.182 9)	(0.449 7,0.715 5,1.152 4)	(0.486 5,0.774 1,1.268 2)	0.150 6	0.177 4	0.043 1
(1.2,0.3,3.5)	(1.104 1,0.245 0,2.979 1)	(0.718 9,0.195 0,2.304 6)	(1.146 9,0.278 6,3.343 8)	0.150 6	0.166 3	0.134 6
(-0.3,1.6,3.8)	(-0.260 1,1.301 7,3.120 8)	(-0.237 8,1.344 6,3.164 5)	(-0.293 1,1.548 2,3.707 7)	0.742 9	0.687 7	0.106 0
(-1,0.3,1.6)	(-0.944 0,0.277 9,1.505 5)	(-0.951 8,0.289 7,1.532 6)	(-1.012 7,0.298 4,1.623 9)	0.112 1	0.083 6	0.027 1
(3.2,1.4,1.5)	(2.916 6,1.275 2,2.1374 4)	(3.406 1,1.486 7,1.593 6)	(3.122 5,1.362 2,1.463 6)	0.334 2	0.242 4	0.093 6
(1.7,-0.8,5)	(1.463 4,-0.705 4,4.332 1)	(1.649 3,-0.777 1,4.839 0)	(1.701 4,-0.815 5,5.001 6)	0.714 8	0.170 4	0.015 6
(3.2,2.4,5.4)	(4.264 6,3.202 4,7.154 7)	(3.018 1,2.269 7,5.091 1)	(3.084 9,2.334 9,5.207 2)	2.203 6	0.381 4	0.233 8
(0.5,1.4,5)	(0.950 9,2.559 7,8.951 4)	(0.339 8,0.898 6,3.124 2)	(0.423 0,1.218 1,4.225 8)	4.142 6	1.948 2	0.799 0
(2.0,4.0,4.5)	(1.765 7,3.543 4,3.989 2)	(2.669 1,5.342 8,6.043 6)	(2.039 2,4.062 8,4.581 1)	0.724 1	2.152 6	0.109 8

表 4 四麦克风阵列定位结果

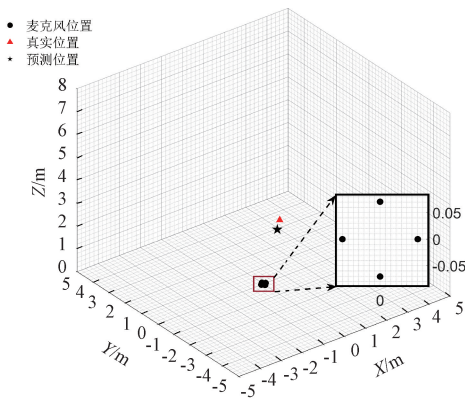
Table 4 Localization results of quadruple microphone array

真实坐标	子阵列 1 估计坐标	子阵列 2 估计坐标	子阵列 3 估计坐标	子阵列 4 估计坐标	融合定位坐标
(1,0.3,2.3)	(0.966 8,0.386 1,2.230 4)	(0.561 3,0.113 7,1.294 0)	(0.677 8,0.189 0,1.419 6)	(1.613 2,0.531 9,4.124 3)	(1.012 4,0.305 1,2.333 0)
(0.5,0.8,1.3)	(0.669 1, 1.205 9, 1.739 9)	(0.256 8, 0.357 9, 0.668 1)	(0.469 3, 0.624 0, 1.019 0)	(0.464 1, 0.930 6, 1.513 2)	(0.493 3, 0.788 7, 1.284 5)
(1.2,0.3,3.5)	(1.166 8, 0.384 3, 3.399 8)	(0.845 1, 0.141 7, 2.457 5)	(0.745 9, 0.170 4, 2.014 6)	(1.575 7, 0.430 9, 5.015 6)	(1.228 6, 0.305 6, 3.583 4)
(-0.3,1.6,3.8)	(-0.299 4, 1.699 4, 3.787 1)	(-0.181 9, 0.896 5, 2.284 8)	(-0.110 4, 0.877 2, 2.091 7)	(-0.326 3, 1.306 8, 3.103 2)	(-0.291 5, 1.548 7, 3.683 9)
(-1,0.3,1.6)	(-0.922 3, 0.369 9, 1.480 7)	(-0.860 5, 0.173 6, 1.376 9)	(-0.890 6, 0.293 3, 1.582 6)	(-0.925 4, 0.254 9, 1.345 1)	(-1.004 7, 0.301 9, 1.608 5)
(3.2,1.4,1.5)	(3.824 1, 1.782 8, 1.792 8)	(2.559 2, 1.042 4, 1.197 8)	(4.254 0, 1.791 2, 1.938 7)	(2.540 3, 1.153 3, 1.224 5)	(3.278 5, 1.432 3, 1.535 8)
(1.7,-0.8,5)	(1.436 3, -0.653 2, 4.065 7)	(1.771 8, -0.942 1, 5.217 6)	(1.145 0, -0.510 4, 3.185 9)	(0.993 7, -0.503 7, 3.098 7)	(1.750 1, -0.830 5, 5.149 0)
(3.2,2.4,5.4)	(1.563 8, 1.226 3, 2.629 6)	(2.004 8, 1.445 2, 3.385 1)	(4.311 3, 3.114 6, 7.052 9)	(7.817 6, 6.035 3, 13.600 4)	(3.078 4, 2.307 2, 5.188 9)
(0.5,1.4,5)	(0.535 6, 1.610 4, 5.359 1)	(0.270 7, 0.701 2, 2.681 2)	(0.513 3, 1.194 8, 4.281 1)	(0.240 1, 0.841 3, 3.026 5)	(0.500 1, 1.398 7, 4.999 7)
(2.0,4.0,4.5)	(1.489 5, 3.040 8, 3.342 8)	(0.914 4, 1.775 8, 2.057 3)	(7.271 8, 13.912 6, 15.596 6)	(6.006 5, 12.682 0, 14.260 3)	(1.946 5, 3.892 2, 4.380 4)

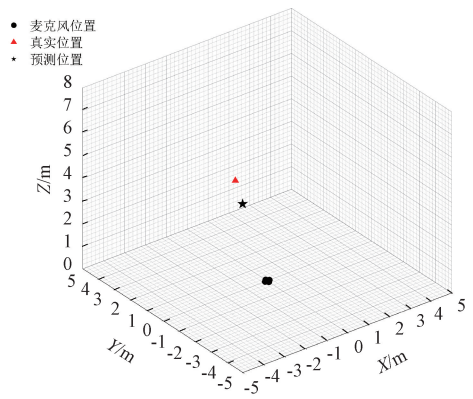
表5 四麦克风阵列定位误差

Table 5 Localization error of quadruple microphone array

子阵列1误差	子阵列2误差	子阵列3误差	子阵列4误差	融合定位误差
0.078 3	1.100 9	0.982 7	1.972 5	0.035 6
0.561 9	0.758 6	0.356 4	0.258 1	0.020 3
0.106 7	1.102 8	1.590 7	1.594 7	0.088 3
0.012 9	1.635 2	1.857 1	0.7596	0.127 2
0.145 5	0.264 4	0.020 9	0.312 3	0.009 9
0.745 1	0.753 9	1.120 6	0.670 8	0.092 1
0.981 8	0.233 0	1.950 4	2.017 5	0.160 1
3.460 4	2.493 7	2.065 3	10.135 0	0.260 7
0.377 4	7.974 9	0.752 6	2.057 3	0.001 3
1.649 8	3.414 4	15.752 5	13.693 2	0.169 6



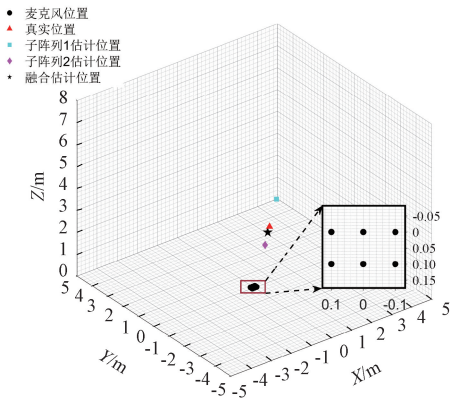
(a) 四元十字阵列声源定位仿真图1
(a) Simulation of sound source localization using quaternary cross-shaped array (Simulation 1)



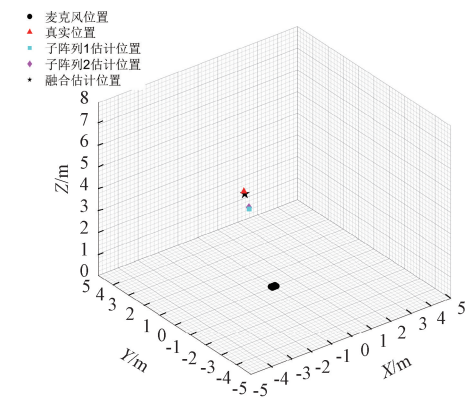
(b) 四元十字阵列定位仿真图2
(b) Simulation of sound source localization using quaternary cross-shaped array (Simulation 2)

图11 四元十字阵列声源定位仿真

Fig. 11 Simulation of sound source localization using quaternary cross-shaped array



(a) 双麦克风阵列声源定位仿真图1
(a) Simulation of sound source localization using dual microphone array (Simulation 1)



(b) 双麦克风阵列声源定位仿真图2
(b) Simulation of sound source localization using dual microphone array (Simulation 2)

图12 双麦克风阵列声源定位仿真

Fig. 12 Simulation of sound source localization using dual microphone array

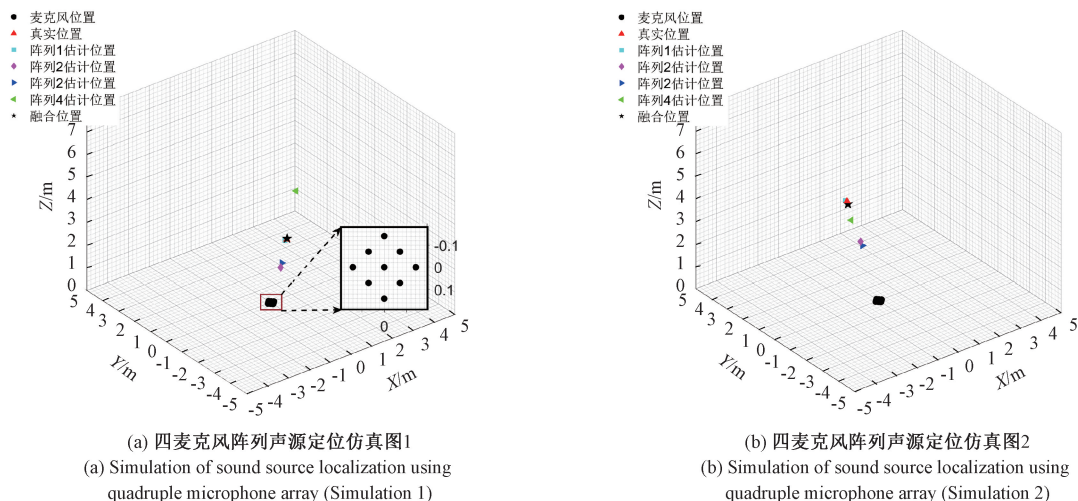


图 13 四麦克风阵列声源定位仿真

Fig. 13 Simulation of sound source localization using quadruple microphone array

从图 11 和表 2 可以看出,在设置的条件下,利用四元阵列在声源较近的条件下大体上可以定位出声源位置,但是距离较远下定位结果偏差巨大,无法得到真正的位置。在实际应用中有可能导致较大的偏差。所以仅仅用一个四元十字阵列进行定位会有很大的不稳定性。

从图 12 和表 3 可以看出,双麦克风阵列声源定位算法由于融合了两个四元阵列的结果,定位精度明显提升,距离误差大部分都在 0.1 m 以下,且在较远距离的定位依旧准确,仅有点(0.5, 1.4, 5)定位误差较大,但对比单阵列精度依旧有巨大的提升。

从图 13 以及表 4、表 5 可以看出,四麦克风阵列声源定位中,使用单一阵列的定位误差巨大,经过融合定位之后精度得到巨大提升,另外对比双麦克风阵列,四麦克风阵列的定位结果得到进一步提升。

综上所述,多麦克风阵列数据融合的方法相比于单一阵列来说有效提升了定位精度,且通过双麦克风阵列和四麦克风阵列的定位仿真验证了该方法可以适用于两个及以上阵列的融合定位,并且得到了良好的结果。

5 结 论

本文提出了一种基于插值算法的改进二次互相关算法以及一种多麦克风阵列数据融合的声源定位方法,改进的二次互相关算法通过三次样条插值提高了采样频率,使得时延估计精度更高,并且可以通过使用不同插值倍数避免同一采样率下时延估计出现奇异值的问题。得到时延差之后进行声源位置的计算,提出的多麦克风阵列数据融合方法避免了单阵列定位误差较大的问题,通过在两个及以上麦克风阵列上分别进行声源定位后进行数据融合得到最终声源位置,并在实验中用两个以及 4

个四元麦克风阵列进行远场声源定位的仿真,实验结果表明多麦克风阵列数据融合方法相比于单个四元阵列可以更加精准的估计出声源位置。

本研究不仅展示了在声源定位领域的先进性和必要性,也为未来进一步提高定位速度、优化算法以适应实时应用、以及在复杂声学环境中的应用提供了新的思路 and 方向。下一步,可以考虑在不改变定位精度的条件下进一步提高定位速度,满足多场景的需求。

参考文献

- [1] 郭业才,陈小燕,王超. LCMV 频分的改进维纳滤波后置波束形成算法[J]. 电子测量与仪器学报, 2017, 31(10): 1646-1652.
GUO Y C, CHEN X Y, WANG CH. Improved wiener post-filter beamforming algorithm based on LCMV divided-frequency[J]. Journal of Electronic Measurement and Instrumentation, 2017, 31(10): 1646-1652.
- [2] JIN J, BENESTY J, CHEN J, et al. Differential beamforming from a geometric perspective[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2023, 31: 3042-3054.
- [3] 杨诗童,杨飞. 改进 TF-GSC 和改进后置滤波语音增强算法[J]. 电子测量技术, 2023, 46(17): 118-124.
YANG SH T, YANG F. Improved TF-GSC and improved post filter speech enhancement algorithm[J]. Electronic Measurement Technology, 2023, 46(17): 118-124.
- [4] GANNOT S, VINCENT E, MARKOVICH-GOLAN S, et al. A consolidated perspective on multimicrophone speech enhancement and source separation[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2017, 25(4): 692-730.

- [5] YIN J, WAN Q, YANG S, et al. A simple and accurate TDOA-AOA localization method using two stations [J]. IEEE Signal Processing Letters, 2015, 23 (1): 144-148.
- [6] DANG X, MA W, HABETS E A P, et al. TDOA-based robust sound source localization with sparse regularization in wireless acoustic sensor networks [J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2022, 30: 1108-1123.
- [7] 张梅,张双双,袁宏永,等.燃气管道泄漏的次声源定位算法研究[J].电子测量与仪器学报,2020,34(3):187-194.
ZHANG M, ZHANG SH SH, YUAN H Y, et al. Research on infrasound source location algorithm of gas pipeline leakage [J]. Journal of Electronic Measurement and Instrumentation, 2020, 34(3): 187-194.
- [8] WANG X, HUANG G, BENESTY J, et al. Time difference of arrival estimation based on a kronecker product decomposition [J]. IEEE Signal Processing Letters, 2021, 28: 51-55.
- [9] 李蜀丰,徐永绍,刘秉政,等.基于改进 MUSIC 的声源定位方法 [J]. 电子测量与仪器学报, 2021, 35(8): 212-219.
LI SH F, XU Y SH, LIU B ZH, et al. Sound source localization method based on improved MUSIC [J]. Journal of Electronic Measurement and Instrumentation, 2021, 35(8): 212-219.
- [10] LIU M, HU J, ZENG Q, et al. Sound source localization based on multi-channel cross-correlation weighted beamforming [J]. Micromachines, 2022, 13(7): 1010.
- [11] 刘路路. 基于麦克风阵列的声源定位技术研究 [D]. 西安:西安电子科技大学, 2020.
LIU L L. Research on sound source localization technology based on microphone array [D]. Xi'an: Xidian University, 2020.
- [12] JI H, CUI X, GAO Y, et al. 3-D ultrasonic localization of transformer patrol robot based on EMD and PHAT- β algorithms [J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 1-10.
- [13] 李保伟, 张兴敢. 基于广义互相关改进的麦克风阵列声源定位方法 [J]. 南京大学学报 (自然科学版), 2020, 56(6): 917-922.
LI B W, ZHANG X G. Improved microphone array sound source localization method based on generalized cross correlation [J]. Journal of Nanjing University (Natural Sciences), 2020, 56(6): 917-922.
- [14] LIU M, ZENG Q, JIAN Z, et al. A sound source localization method based on improved second correlation time delay estimation [J]. Measurement Science and Technology, 2023, 34(4): 045102.
- [15] ZHANG L, ZHANG T, SHIN H S, et al. Efficient underwater acoustical localization method based on time difference and bearing measurements [J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 1-16.
- [16] KRALJEVIC L, RUSSO M, STELLA M, et al. Free-field TDOA-AOA sound source localization using three soundfield microphones [J]. IEEE Access, 2020: 87749-87761.
- [17] QIU Y, LI B, HUANG J, et al. An analytical method for 3-D sound source localization based on a five-element microphone array [J]. IEEE Transactions on Instrumentation and Measurement, 2022, 71: 1-14.
- [18] DESAI D, MEHENDALE N. A review on sound source localization systems [J]. Archives of Computational Methods in Engineering, 2022, 29(7): 4631-4642.
- [19] LIU H, CHEN Y, HUANG Y, et al. Study on the localization method of multi-aperture acoustic array based on TDOA [J]. IEEE Sensors Journal, 2021, 21(12): 13805-13814.

作者简介



刘扬, 2022 年于山东科技大学获得学士学位, 现为中国人民公安大学硕士研究生, 主要研究方向为阵列信号处理与安全防范技术。

E-mail: 1419311746@qq.com

Liu Yang received his B. Sc. degree from Shandong University of Science and Technology in 2022. Now he is a M. Sc. candidate at People's Public Security University of China. His main research interests include array signal processing and security and prevention technology.



张传营, 2022 年于河北科技大学获得学士学位, 现为中国人民公安大学硕士研究生, 主要研究方向为阵列及语音信号处理。

E-mail: 962409974@qq.com

Zhang Chuanying received his B. Sc. degree from Hebei University of Science and Technology in 2022. Now he is a M. Sc. candidate at People's Public Security University of China. His main research interests include array and speech signal processing.



赵景玉, 2021 年于东北大学获得学士学位, 现为中国人民公安大学硕士研究生, 主要研究方向为阵列信号处理与自动化控制。

E-mail: 1754633076@qq.com

Zhao Jingyu received his B. Sc. degree