

第十三届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会
电话：0471-4969085

邮编：010021

网址：www.tzmcm.cn
Email: service@tzmcm.cn

2020 年第十三届“认证杯”数学中国 数学建模网络挑战赛第一阶段论文

题 目 基于时间延迟定位法的麦克风阵列的混响模型

关 键 词 时间延迟定位法 麦克风阵列的混响模型

GCC 时延估计法 球形插值法 几何镜像法

摘 要：

目前使用麦克风阵列定位声源研究较多的有高分辨率谱定位法、可控波束定位法、时间延迟定位法三种方法，考虑我们的声源信号是人为产生的音源信号，具有实时性，不连续性，短暂平稳性，因此我们采用时间延迟定位法作为本问题的解决方法。

考虑到室内环境有混响和噪音的干扰，我们利用 MATLAB 来模拟出麦克风接受的真实信号并利用广义互相关的方法求时延，来削弱混响信号带来的影响。

针对于问题一，由于房间尺寸已知，求解脚步声源位置便是二维平面求解，在二维平面高度是忽略不计的，所以麦克风阵列的高度越低越好。由理论可得二维平面至少需要三个麦克风，由于麦克风阵列的摆放位置和旋转角度不能精确得知，我们使用线性等间距结构和放在房间角落的方法来方便确定位置和旋转角度，再通过不断步进增加间距，根据所求结果的精度以及控制成本的前提下来确定合适的间距，最终得到合适的间距为 20cm，问题一最后的麦克风阵列三元线性等间距阵列，间距为 20cm。

针对问题二，需要求解房间尺寸，所以我们将二维平面转化为三维平面，由于三维平面的位置求解更为复杂，所以我们引入球形插值法来求解声源位置，麦克风个数也由最少的三个增加到了四个，通过实验对比我们发现正四面体结构的麦克风阵列精确度最高，再通过逐步增加间距的方法求解出间距为 25cm 时最合适。由于互相关函数的 R 有多个峰值，所以我们可以根据不同的峰值来估算出，除直达路径以外的其他路径信号的时延，在假设墙壁只反射一次的情况下求出声源和镜像声源的位置，并利用几何原理求出房间的尺寸。问题二最后的麦克风阵列为正四面体结构，每个麦克风位于四面体的每个顶点，间距为 25cm。

参赛队号： 1037

所选题目： A 题

参赛密码 _____
(由组委会填写)

第十三届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会
电话：0471-4969085

邮编：010021

网址：www.tzmcm.cn
Email: service@tzmcm.cn

英文摘要

Three methods of using microphone array to locate sound source are currently studied more highresolution spectral positioning method, controllable beam positioning method, time delay positioning method, considering that our sound source signal is an artificially generated sound source signal, with realtime , Discontinuity, transient stability, so we use the time delay positioning method as the main method.

Considering the reverberation and noise interference in the indoor environment, we use MATLAB to simulate the real signal received by the microphone and use the generalized cross-correlation method to find the time delay to reduce the impact of the reverberation signal.

For problem one, since the room size is known, the solution to the position of the footstep sound source is the two dimensional plane. The height of the twodimensional plane is negligible, so the lower the height of the microphone array, the better. It can be obtained from theory that at least three microphones are needed for the two dimensional plane. Since the placement and rotation angle of the microphone array cannot be accurately known, we use the linear equalspace structure and the method of placing it in the corner of the room to easily determine the position and rotation angle. Then continue to increase the spacing step by step, and determine the appropriate spacing according to the accuracy of the required results and the cost of control, and finally get the proper spacing of 20cm. Problem 1 The final microphone array is a ternary linear equal spacing array with a spacing of 20cm .

For problem two, the room size needs to be solved, so we convert the two-dimensional plane into a three-dimensional plane. Since the position of the three-dimensional plane is more complicated, we introduce the spherical interpolation method to solve the position of the sound source. The number of microphones has also been increased from three to four. Through experimental comparison, we have found that the microphone array with a regular tetrahedral structure has the highest accuracy, and then the method of gradually increasing the spacing is used to find the most suitable when the spacing is 25cm. Since R of the cross-correlation function has multiple peaks, we can estimate the delay of the signal of the path other than the direct path according to different peaks, Assuming that the wall reflects only once, find the positions of the sound source and the mirrored sound source. Finally, according to the coordinates of the sound source and the mirror sound source, the geometrical principle is used to obtain the size of the room. Problem 2: The final microphone array is a regular tetrahedron structure, each microphone is located at each vertex of the tetrahedron, with a spacing of 25cm.

目录

一、问题重述.....	2
1.1 相关背景	2
1.2 问题提出	2
二、模型假设.....	2
三、符号说明.....	2
四、模型准备.....	3
4.1 麦克风阵列信号模型.....	3
4.1.1 理想模型	3
4.1.2 实际模型	4
4.2 声源定位模型	6
4.2.1 高分辨率谱定位法.....	6
4.2.2 可控波束定位法.....	6
4.2.3 时间延迟定位法.....	6
五、问题一模型的建立与求解.....	7
5.1 构建基于时间延迟的声源定位模型.....	7
5.1.1 TDOA 法简介.....	7
5.1.2 GCC 方法估计时延.....	7
5.1.3 几何法求解声源位置	9
5.2 构建最少成本麦克风阵列	11
六、问题二模型的建立与求解.....	12
6.1 模型拓展	12
6.2 球形插值法	14
6.3 声源与镜像位置的确定	14
6.4 房屋尺寸估量方法.....	15
七、模型分析与评价	18
7.1 敏感性分析	18
7.2 模型的优缺点	19
参考文献	19
附 录	20

一、问题重述

1.1 相关背景

传统视频监控只能记录图像，而无法计算准确的距离，而听音辨位系统的麦克风树则可以准确地计算出声源的位置信息。由于每个麦克风所处的位置不同，因此每个麦克风接收到的声音信息也会有细微的差别，利用这些差别，我们可以计算出声源的方向和距离。

听音辨位系统具有很高的应用潜力，在安全领域上，它能够解决传统摄像头存在监控死角的问题，通过声音辨别来辅助监控系统。在地质勘探中，它能够快速地测量洞穴的深度。因此，研究听声辨位的应用以及它的特性，具有很高的价值与意义。

1.2 问题提出

(1) 在一个平面形状是矩形，地面和天花板都是水平的大厅（假设我们已知大厅的轮廓的准确尺寸，包括：长、宽、高）中建立一个麦克风树结构模型，此模型能使用较少的麦克风，且较为准确地计算发出声音的人的位置。

(2) 大厅的尺寸未知，我们需要设计出一个模型，改模型在能够确定发出声音的人的位置的同时，能最大限度地测量出大厅的轮廓尺寸(长 \times 宽 \times 高)。

二、模型假设

1. 假设所有麦克风具有相同的性能参数，且工作在额定功率。
2. 假设大厅环境中的干扰噪声仅有高斯噪声。
3. 发出声音的声源只有一个人。
4. 声波在墙面上只反射一次。

三、符号说明

符号	含义
$s(t)$	声源信号
r_s	声源的坐标
α_n	传播效应衰减因子
$v_n(t)$	环境噪声信号
τ	麦克风之间的时间差
g_n	麦克风对声源信号的脉冲响应函数
M_i	第 i 个麦克风
d_i	声源到麦克风 M_i 的距离
k	单位距离衰减系数
R	互相关系数
θ	声源信号与 x 轴夹角
L	大厅的长度
S	大厅的宽度
H	大厅的高度

四、模型准备

4.1 麦克风阵列信号模型

麦克风阵列^[1]是利用各个阵元接收到的语音信号作为后续处理的输入的，麦克风阵列所接收到的信号中有直接到达的信号、经过障碍物反射的信号以及噪声信号。通常将麦克风阵列信号模型分为理想模型和实际模型两种。

4.1.1 理想模型

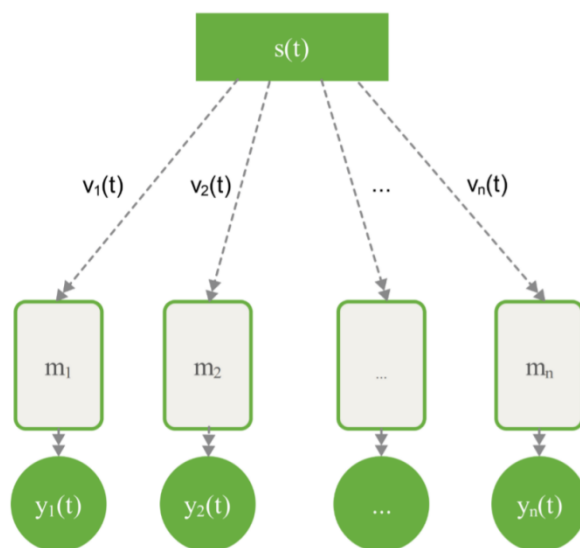


图 1 麦克风阵列理想信号模型

在理想的声音环境下，我们可以认为麦克风阵列阵元所接收的所有语音信号都是直达的。因为我们的问题只涉及单声源的定位，因此单声源多麦克风理想模型如图 1 所示，声源信号 $s(t)$ 呈辐射状传播，且语音信号的能量与距离声源的距离成反比。我们将第 1 个麦克风定为参考麦克风，则第 1、2、 n 个麦克风在 t 时刻所输出的信号分别为：

$$y_1(t) = \alpha_1 s(t - t_0) + v_1(t) = \alpha_1 s(t - t_0 - \Gamma_1(\tau)) + v_1(t) = x_1(t) + v_1(t) \quad (1-1)$$

$$y_2(t) = \alpha_2 s(t - t_0 - \tau_{21}) + v_2(t) = \alpha_2 s(t - t_0 - \Gamma_2(\tau)) + v_2(t) = x_2(t) + v_2(t) \quad (1-2)$$

$$y_n(t) = \alpha_n s(t - t_0 - \tau_{n1}) + v_n(t) = \alpha_n s(t - t_0 - \Gamma_n(\tau)) + v_n(t) = x_n(t) + v_n(t) \quad (1-3)$$

其中 $n = 1, 2, \dots, N$ ， α_n 是语音信号的传播效应衰减因子，取值范围从 0 到 1； $s(t)$ 是我们不知道位置的声源信号， t_0 是信号从声源到第一个麦克风的传播时间， $v_n(t)$ 是第 n 个麦克风的噪声信号，我们认为 $v_n(t)$ 与声源信号不相关，且两两 $v_n(t)$ 之间互不相关， $\tau_{n1} = \Gamma_n(\tau)$ 是第 n 个麦克风与第 1 个麦克风之间的时间差，其中 $\Gamma_1(\tau) = 0$ ， $\Gamma_2(\tau) = \tau$ ，当 n 大于等于 3 时， $\Gamma_n(\tau)$ 由 τ 和阵元之间的几何位置关系共同决定。

4.1.2 实际模型

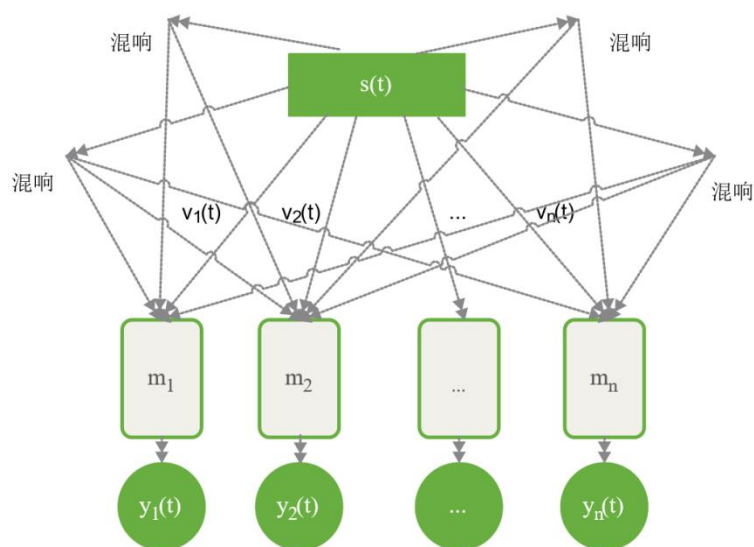


图 2 麦克风阵列混响信号模型

我们在理想环境暂不考虑把多径效应，这样使计算变得简单，但是这种模型在室内混响环境下是不适当的。因此建立麦克风阵列的混响模型，使其更容易和较多的理解多径传播对模型的影响，如图 3 所示，现在就可以将这个实际模型看成是单输入多输出系统，照例将第 1 个麦克风看成是参考元，则第 1、2、 n 个麦克风在 t 时刻所接收到的语音信号分别为：

$$y_1(t) = g_1(t) * s(t) + v_1(t) = x_1(t) + v_1(t) \quad (2-1)$$

$$y_2(t) = g_2(t) * s(t) + v_2(t) = x_2(t) + v_2(t) \quad (2-2)$$

...

$$y_n(t) = g_n(t) * s(t) + v_n(t) = x_n(t) + v_n(t) \quad 1 \leq n \leq N \quad (2-3)$$

其中, g_n 是第 n 个麦克风对声源信号的脉冲响应函数, ' $*$ ' 表示卷积, $v_n(t)$ 是第 n 个麦克风的噪声信号, $v_n(t)$ 与声源信号不相关且 $v_n(t)$ 之间两两互不相关。在单音源多元麦克风阵列信号混响模型中, 时延并不是一个可直接获得的参数。由于声源信号未知, 只有当这个单输入多输出系统可以被辨识时, 才能求出时延。

我们利用 MATLAB 的工具箱^[2]来构造生成每个麦克风在室内混响条件下所接受到的声音信号模型。

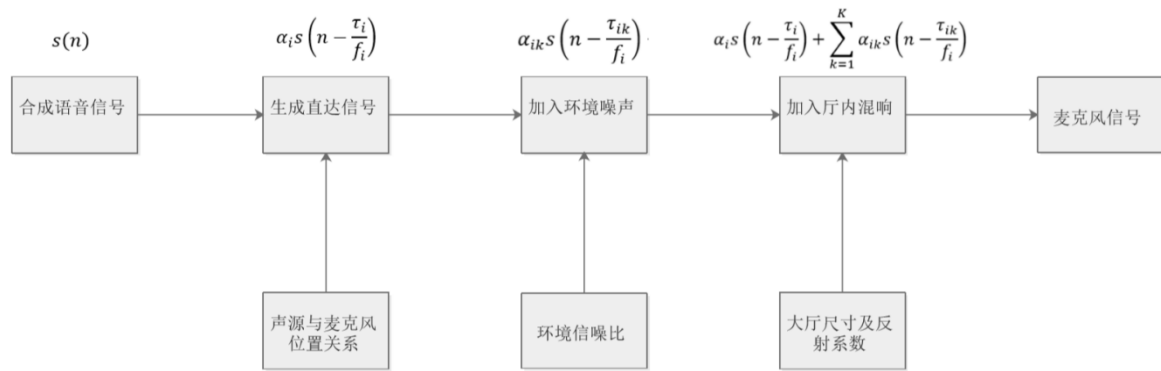


图3 麦克风信号合成流程图

由于录制的语音信号中含有外界噪声信息, 为避免外界噪声干扰, 保证测试准确性, 本模型中采用的原始语音信号由计算机语音合成。

对于合成的原始声源信号, 计算机通过对其进行处理生成各路麦克风的语音信号。对原始声源信号的处理主要分为①延时衰减、②加噪声和③加混响三部分, 经过处理后的第 i 路麦克风的语音信号表示为:

$$x_i(n) = \alpha_i s\left(n - \frac{\tau_i}{f_i}\right) + \sum_{k=1}^K \alpha_{ik} s\left(n - \frac{\tau_{ik}}{f_i}\right) + n_i(n) \quad (i = 1, 2, 3, 4) \quad (3)$$

$\alpha_i s\left(n - \frac{\tau_i}{f_i}\right)$ 表示声源到麦克风 M_i 的直达信号。其中, τ_i 为声波从声源到麦克风 M_i 所需的时间, 用 d_i 表示声源到麦克风 M_i 的距离, v 表示声速, 则 $\tau_i = \frac{d_i}{v}$ 。 α_i 表示信号从声源直达麦克风 M_i 的幅值衰减系数, 为简化模型, 设声波的幅值和其传播的距离成反比, 即 $\alpha_i = 1 - kd_i$, k 为单位距离衰减系数, 在本模型中, 取 $k = 0.9m^{-1}$ 。

$n_i(n)$ 为麦克风 M_i 收到的噪声信号, 本设计中为纯高斯噪声, 其幅值取决于环境信噪比。

$\sum_{k=1}^K \alpha_{ik} s\left(n - \frac{\tau_{ik}}{f_i}\right)$ 为麦克风 M_i 收到的混响信号。 α_{ik} 为混响信号衰减系数, 其值取决于该路反射路径的长度及墙面的反射系数。用 d_{ik} 表示声音从声源到达麦克风 M_i 的第

k 条路径的距离, γ 表示墙面的反射系数,则 α_{ik} 的计算公式为:

$$\alpha_{ik} = kd_{ik}\gamma^{\frac{n}{2}} \quad (4)$$

其中, n 为反射次数, τ_{ik} 为声音从声源到达麦克风 M ,的第 k 条路径所花费的时间, 其计算公式为 $\tau_{ik} = \frac{d_{ik}}{v}$ 。

室内混响情况的模拟采用 IMAGE 模型, 在此模型中, 我们仅考虑单次反射后的回声, 因此求出声源相对于墙面对称的镜像声源的位置, 然后计算出镜像声源对每个麦克风产生的混响信号。至此, 我们的信号仿真构造完成。

4.2 声源定位模型

目前相关专业学者研究的麦克风阵列的声源定位方法大致分为三种: 高分辨率谱定位法^[3]、可控波束定位法、时间延迟定位法^[4]。

4.2.1 高分辨率谱定位法

此方法根据不同的麦克风所接受到的语音信号, 利用这些信号之间的相关矩阵来计算出信号的方位角, 进而确定出声源的准确位置。此方法在传统的阵列信号处理中应用较为广泛, 要求信号在计算时段中处于平稳状态, 但是人发出的语音信号或者是脚步声信号是非平稳的信号, 短平稳时间内无法采集到足够数据, 难以满足高分辨谱对计算精度的要求。因此, 此方法不适合于我们的问题。

4.2.2 可控波束定位法

该方法对各个麦克风所接收到的信号进行滤波处理并加权求和, 从而通过搜索可能的声源方位来引导波束, 修正权值使阵列的输出功率最大, 使麦克风转至有最大输出功率的波束的方向, 那么声源的位置就是功率最大的那一点。可控波束形成定位的主要方法分为延迟累加波束算法与自适应波束算法。前者计算量比较小, 信号失真比较少, 但易受噪声的影响, 需要较多的阵元才会有较好效果; 后者由于加了自适应滤波, 计算量增大, 并且输出信号有一定的失真, 但需要的阵元数目不多, 在无混响环境下效果较好。但是可控波束方法要求声源信号与环境噪声的频谱特性是已知的, 而在我们的现实问题中, 这些参数往往是不可能已知的, 因此采用可控波束定位法容易得出局部极小值, 而不能计算出准确的位置。

4.2.3 时间延迟定位法

此方法先计算出不同麦克风接收到语音信号之间的时间差, 然后利用时延估计声源距离麦克风的距离差, 最后通过几何方法确定声源的位置。任意两个麦克风接收到声源发出的语音信号都有一个时间差, 而声速是固定不变的, 因此我们可以在两个麦

麦克风之间建立两个曲线，而曲线的交点就是声源的位置。由于两曲线大部分情况至少有两个交点，因此计算出所有麦克风的曲线的交集就是最为准确的声源位置。

考虑我们的声源信号是人为产生的音源信号，具有实时性，不连续性，短暂平稳性，因此我们采用时间延迟定位法作为主要方法。

五、问题一模型的建立与求解

5.1 构建基于时间延迟的声源定位模型

5.1.1 TDOA 法简介

基于时间延迟的声源定位方法，由于其具有概念简单、计算量小、定位精度高、实时性好等优点，近几年发展较好，且在实际中得到了广泛的应用。如图 3.1 所示，该方法分成两步：估计时延和定位^{[4][5]}。



图 4 TDOA 法的操作流程

基于两步定位的 TDOA 方法由于其概念简单，运算量小，研究的历史相对较长，因此在实际中得到了广泛的应用。

目前实现的绝大部分相关系统，都是使用这种方法。两步定位方法的方法：

- 第一步是对时延的估计。
- 第二步是由得到的时延，结合麦克风阵列的物理结构，得到声源的位置。

5.1.2 GCC 方法估计时延

我们假设声源的信号为 $s(n)$ ，声源坐标 $r_s = (x, y, z)$ ，麦克风阵列共有 M 个麦克风，且每个麦克风的坐标为 $r_i = (x, y, z), i = 1, 2, \dots, M$ 。且在采样时刻每个麦克风收到的信号为 $x_i(n)$ ^[6]。

对于麦克风 p ，我们忽略传播中幅度的衰减，所以它收到的信号为：

$$x_p(n) = s(n - \tau_p) + n_p \quad (5)$$

其中， τ_p 是从声源到麦克风 p 过程中的传输时延， n_p 为噪声，则：

$$\tau_p = \frac{|r_s - r_p| * fs}{c} \quad (6)$$

同理，对于麦克风 q 而言，则有：

$$x_q(n) = s(n - \tau_p) + n_q \quad (7)$$

其中， τ_q 是声源到麦克风 q 的传输时延， n_q 为噪声：

$$\tau_q = \frac{|r_s - r_q| * fs}{c} \quad (8)$$

接下来，我们计算两个麦克风接收信号 $x_p(n)$ 和信号 $x_q(n)$ 的互相关函数：

$$R_{pq}(\tau) = E[x_p(n)x_q(n - \tau)] \quad (9)$$

$$\begin{aligned} R_{pq}(\tau) = & E[s(n - \tau_p)s(n - \tau_q - \tau)] + E[s(n - \tau_p)n_q(t)] \\ & + E[s(n - \tau_q)n_p(t)] + E[n_p(t)n_q(t)] \end{aligned} \quad (10)$$

由于上面的噪声是不相关的假设，于是有：

$$R_{pq}(\tau) = E[s(n - \tau_p)s(n - \tau_q - \tau)] \quad (11)$$

我们根据相关函数的性质了解到，当 $\tau = \tau_p - \tau_q$ 时，即等于节点间信号到达时间差时，相关函数此时达到最大值。在实际计算中常以积分式表示：

$$\hat{R}_{pq}(\tau) = \frac{1}{T - \tau} \int_{\tau}^T x_p(t)x_q(t - \tau)dt \quad (12)$$

来估计相关函数值。但对于离散时间序列，上述积分无法直接完成。为了更精确的求解相关函数的最大值。所以利用相关定理^[7]得到：

$$R_{pq} = \frac{1}{2\pi} \int_{-\infty}^{+\infty} X_p(\omega)X_q^*(\omega)e^{j\omega\tau}d\omega \quad (13)$$

但在实际应用中， $x_p(n)$ 和 $x_q(n)$ 的观测都是有限长的，所以我们只能得到有限长，在离散域上可以表示为：

$$R_{x_p(n)x_q(n)}(\tau) = \sum_{k=0}^{L-1} X_p(k)X_q^*(k)e^{j2\pi k\tau/L} \quad (14)$$

另外在实际环境中会还有回响的影响，相应的传播模型为：

$$x_p(n) = s_p(n) + h_p(n) * s(n) + n_p(n) = s(n - \tau_p) + h_p(n) * s(n) + n_p(n) \quad (15)$$

$$x_q(n) = s_q(n) + h_q(n) * s(n) + n_q(n) = s(n - \tau_q) + h_q(n) * s(n) + n_q(n) \quad (16)$$

已有的文献中表明，在低回响和低噪声的情况，依靠单纯计算互相关函数的方法基本可以准确的计算两个信号的时延差，但在高回响和高噪声的环境中，这种方法的时延估计会受到很大的影响。为了克服高噪声和高回响的影响，我们引入信号处理相关知识。由信号处理的相关理论^[8]我们知道信号的相关函数在频域上表现在对频谱的乘积。目前的很多方法集中在对信号频谱的处理上面。

$$R_{x_p(n)x_q(n)}(\tau) = \sum_{k=0}^{L-1} X_p(k)X_q^*(k)e^{j2\pi k\tau/L} \quad (17)$$

其中 $\omega(k)$ 就是对频谱做出的处理，我们称之为加权函数，其中 $X_p(k)$ 为 $x_p(n)$ 的傅立叶变换， $X_q(k)$ 为 $x_q(n)$ 傅立叶变换， L 为每帧的数据长度。

在我们的实现中，我们采用了 GCC-PHAT 算法，因此上面的式子转化为：

$$R_{x_p(n)x_q(n)}(\tau) = \sum_{k=0}^{L-1} \frac{X_p(k)X_q^*(k)}{|X_p(k)||X_q^*(k)|} e^{j2\pi k\tau/L} \quad (18)$$

5.1.3 几何法求解声源位置

在问题一的条件中，房间的尺寸我们是已知的(假设已知，可以为任意数)，无需我们自己求解，而人发出的脚步声以及其他声音信号相较于开阔的大厅都可以视为在底面上发出，因此求解声源位置可以化简为在二维平面上进行，因此我们可以使用几何法来对时延进行分析，从而得到声源位置。

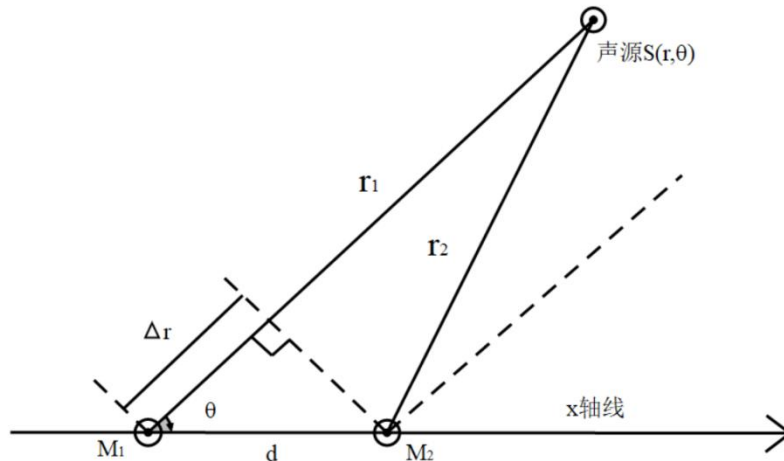


图5 麦克风阵列远场模型

我们设 M_1 、 M_2 、 M_3 分别代表麦克风阵列中的三个等间距的阵元，阵元的间距为 d ， s 为声源信号。假设阵元 M_2 作为时延估计的参考点，三个麦克风阵元的坐标分别为 $M_1(-d,0)$ 、 $M_2(0,0)$ 、 $M_3(d,0)$ ，声源信号的坐标为 $s(r,0)$ ， c 为声音在空气介质中传播的速度。 r_i 分别表示声源 s 到麦克风阵元的距离。其中 θ 为 r 与麦克风阵列 x 轴的夹角。 M_2 与 M_1 接收信号的时延差， M_2 与 M_3 接收信号的时延差为 τ_{12} 、 τ_{23} ，且满足几何关系如下：

$$\begin{cases} \tau_{12} = \frac{1}{c}(r_2 - r_3) \\ \tau_{23} = \frac{1}{c}(r_2 - r_3) \end{cases} \quad (19)$$

$$\theta = \cos^{-1} \frac{r}{d} = \cos^{-1} \left(\frac{\tau_{12} \times c}{d} \right) \quad (20)$$

$$\cos \theta = \frac{d^2 + r_1^2 - r_2^2}{2dr_1} \quad (21)$$

$$\cos \theta = \frac{d^2 - \tau_{12}^2 c^2 + 2r_1 \tau_{12} c}{2dr_1} \quad (22)$$

整理上式我们得到：

$$\theta = \cos^{-1} \frac{d^2 - \tau_{12}^2 c^2 + 2r_1 \tau_{12} c}{2dr_1} \quad (23)$$

另外，当目标声源距离麦克风较近时，我们认为这是一个为近场信号^[9]，信号模型如图6所示，在声源 s 、 M_1 、 M_2 组成的三角 A_{12} 和声源 s 、 M_2 、 M_3 组成的三角形(未画出，同图6)中，根据余弦定理可得：

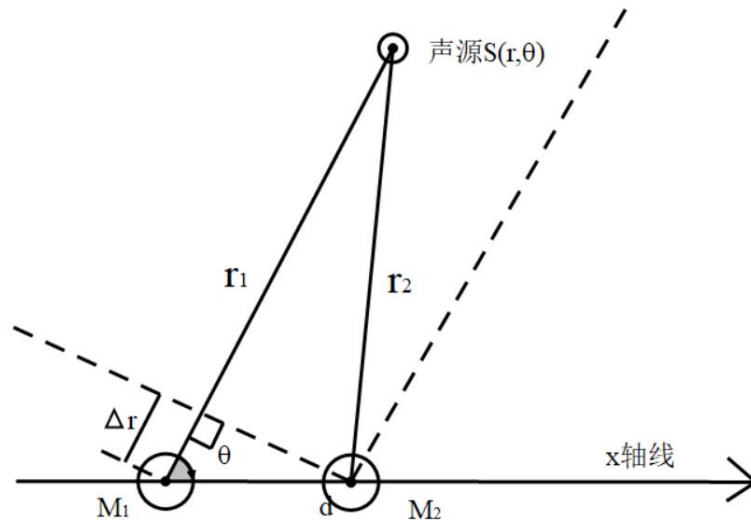


图6 麦克风阵列近场模型

$$\begin{cases} \cos(\pi - \theta) = \frac{r^2 + d^2 - r_1^2}{2rd} \\ \cos \theta = \frac{r^2 + d^2 - r_3^2}{2rd} \end{cases} \begin{cases} r_1 = \sqrt{r^2 + d^2 + 2rd \cos \theta} \\ r_2 = r \\ r_3 = \sqrt{r^2 + d^2 - 2rd \cos \theta} \end{cases} \begin{cases} ct_{21} = \sqrt{r^2 + d^2 + 2rd \cos \theta} - r \\ ct_{23} = \sqrt{r^2 + d^2 - 2rd \cos \theta} - r \end{cases} \quad (24)$$

经推导可得出声源信号距离和方位角如下：

$$\begin{cases} r = \frac{c^2(t_{21}^2 + t_{23}^2) - 2d^2}{2c(t_{21} + t_{23})} \\ \theta = \cos^{-1} \left[\frac{cd^2(t_{21} - t_{23}) + c^3(t_{21}^2 t_{23} - t_{21} t_{23}^2)}{dc^2(t_{21}^2 + t_{23}^2) - 2d^3} \right] \end{cases} \quad (25)$$

我们通过求解上述公式能够得出声源与麦克风的相对位置，但由于麦克风阵列的准确位置和旋转角度无法精确得知，因此我们可以将麦克风放在墙角以固定坐标，并采用线性等间距结构来判断旋转角度。在依次将麦克风阵列中的不同麦克风作为参考麦克风，将不同麦克风求出来的声源位置求均值来减小随机误差。

至此，利用麦克风阵列求解声源位置模型已经构造完毕。如何在满足精度的同时使成本最少，这是我们下一步所要解决的问题。

5.2 构建最少成本麦克风阵列

麦克风阵列的成本来源于两大部分，第一部分为麦克风本身的价格，因此为了节约成本，在满足精确度的前提下，我们应该使用更少的麦克风；第二部分来自于支架的长度，因此，在满足精确度的前提下，麦克风之间的间距应该越小越好，查阅相关论文我们可以得出，麦克风数目的增加和麦克风之间距离的增加都可以更精确的测得时延，以精准后面对声源位置的计算，而在问题一中，人的脚步声相较于宽阔的大厅来说只存在地面，所以我们仅仅只需要二维平面即可对声源进行定位，考虑到麦克风的成本远远大于支架的成本，所以我们使用最少的麦克风来构成麦克风阵列，再通过提高麦克风之间的距离来提高精确度。

对二维平面进行分析可得，对二维平面声源定位最少需要三个麦克风，因此我们便可以得到我们麦克风的个数，接下来，在三个麦克风的基础上，逐渐增加麦克风之间的距离，再计算测量位置与实际位置的偏差，我们通过模拟得到以下数据：

表 1 麦克风间距与相对误差对照表

麦克风间距	相对误差
5.0	22.00%
10.0	7.00%
15.0	6.00%
20.0	4.00%

25.0	3.90%
30.0	3.80%
35.0	3.60%
40.0	3.40%
45.0	3.20%
50.0	3.10%
55.0	3.00%
60.0	2.90%
65.0	2.85%
70.0	2.82%
75.0	2.70%

对数据进行绘图可得：

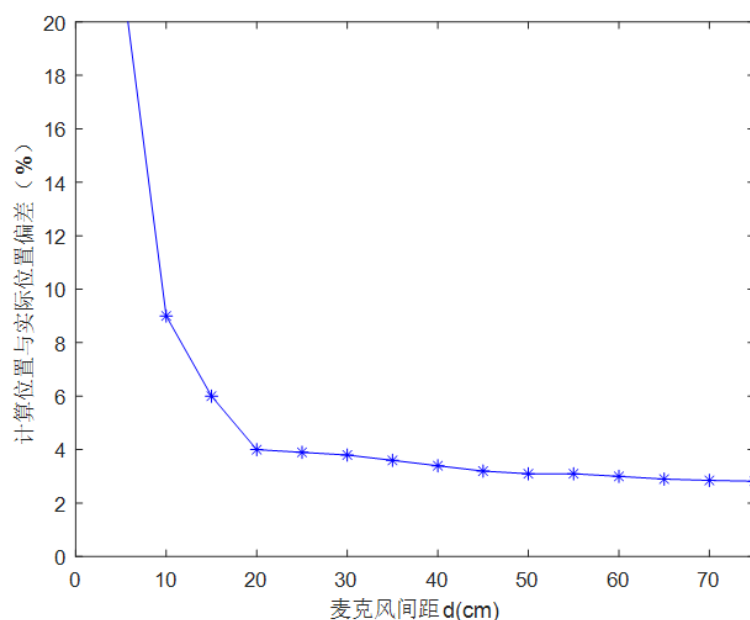


图 7 麦克风间距与位置偏差关系图

通过图 7，我们可以清楚的看到在间距为 0~20cm 时，增加麦克风之间的间距，准确度会大大提高，在间距大于 20cm 时，逐步增加间距，虽然精度依旧有提高，但是提高的范围十分有限，所以我们的麦克风阵列应为 3 个线性排列的单元，之间间距为 20cm，此时为最小成本阵列。

六、问题二模型的建立与求解

6.1 模型拓展

相比问题一，问题二中我们缺少了房屋的尺寸，所以怎么在确定人的位置的同时

得到房屋的尺寸是我们需要解决的问题。根据问题一中我们用到的广义互相关时延算法我们知道，当 τ 等于声音信号到达两个麦克风的时间差时，互相关函数 R 到达最大值，此时的声音信号为直达声，而问题二中我们需要得出房间的大致尺寸，所以我们要考虑反射声到达麦克风的时延，并根据麦克风、声源和房间墙壁的相对位置来判断房间的尺寸。反射声的时延我们同样根据不同麦克风的互相关函数 R 计算，互相关函数 R 具有多个峰值，其中最高的峰值可得出直达声时延，根据其他几个相对较低的峰值，我们可以得知反射声的时延。

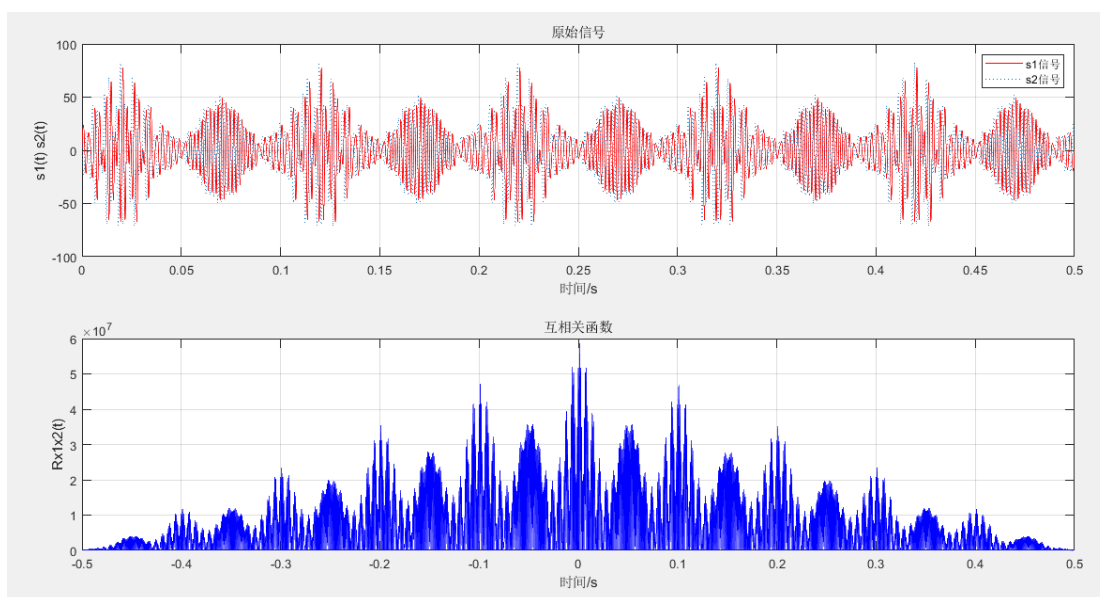


图 8 互相关系数 R 峰值图

之后我们计算时延，并根据时延计算声源位置，算上直达声得到的声源位置我们总共需得出六个位置，根据这六个位置得出房间尺寸的方法用下图做简要解释。

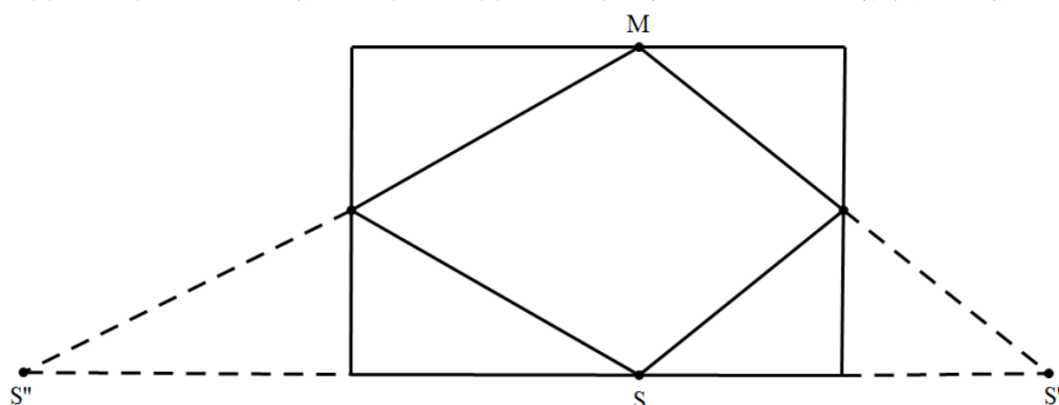


图 9 声源以及确定声源的镜像位置示意图(利用相似三角形等知识)

其中 S 为声源位置， M 为麦克风位置， S' 和 S'' 是我们根据其中两个数据得出的反射声的虚拟位置，根据这两个虚拟位置与的坐标与实际位置的绝对值的距离的绝对值我们可以得到长宽高中的一个，他们的关系是两个绝对值的和，等于尺寸值的二倍。由此，根据这六个位置，两两结成一队，我们可以得出房间的大概尺寸^[10]。

整体的计算思路和问题一相同，共分两步进行，第一步利用 GCC 方法估计时延，计算步骤同问题一，第二步利用时延和相对位置求实际位置时，由于我们需要考虑计

算出房间的高度所以我们做出了调整，将二维平面转化为三维空间算法计算实际位置。综合考虑后我们采用球形插值法^[11]求解。

6.2 球形插值法

该方法不同于极大似然法需要距离差概率分布的先验知识，且利用最陡下降算法估计位置的计算复杂度高，球形插值法根据最小二乘估计方法求解声源与麦克风的位置关系，计算过程简单，计算量小，实时性好，适用于实际的系统。球形插值法中声源和麦克风之间的位置关系如下图：

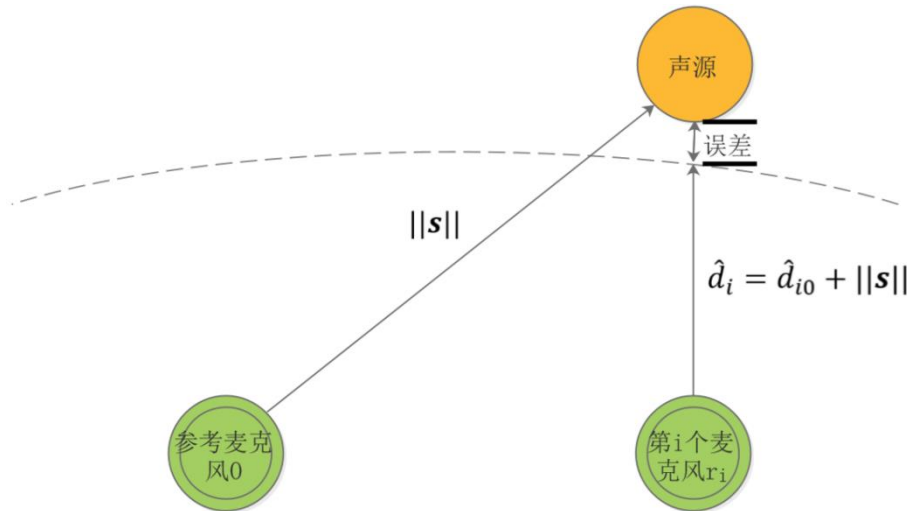


图 10 球形插值法示意图

球形插值法的误差准则是声源到麦克风阵列的距离的误差函数，以这些麦克风为球心的球面的交点就是要获得的声源位置，因此，声源位置的最优解就是到这些球面的最短距离^[13]。

6.3 声源与镜像位置的确定

从上文就已知声源到参考元的距离为 s ，到其他麦克风的距离为 d_i ，则有：

$$\hat{d}_{i0} = d_i - ||s||$$

根据毕氏定理可知，信号的球面波距离可被定义为：

$$d_i^2 = ||\mathbf{s} - \mathbf{r}_i||^2 = S^2 - 2\mathbf{s}^T \mathbf{r}_i + R_i^2 \quad (26)$$

则误差函数可被定义为：

$$e_{SI}(\mathbf{s}) = \frac{1}{2}(\hat{d}_i^2 - d_i^2) = \mathbf{s}^T \mathbf{r}_i + \hat{d}_{i0}^2 S - \frac{1}{2}(R_i^2 - \hat{d}_{i0}^2) \quad i = 1, 2, \dots, N \quad (27)$$

因为麦克风阵列共有 N 对时延，因此可以得到一个 N 维的方程组

$$e_{SI}(\mathbf{s}) = \mathbf{A}\boldsymbol{\varsigma} - \mathbf{e} \quad (28)$$

其中,

$$A = [S\hat{M}], S = \begin{bmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ \vdots & \vdots & \vdots \\ x_n & y_n & z_n \end{bmatrix}, \zeta = \begin{bmatrix} x \\ y \\ z \\ S \end{bmatrix}, e = \begin{bmatrix} R_1^2 - \hat{d}_{10}^2 \\ R_2^2 - \hat{d}_{20}^2 \\ \vdots \\ R_N^2 - \hat{d}_{N0}^2 \end{bmatrix} \quad (29)$$

则对应的误差准则是: $J_{SI}(\mathbf{s}) = \mathbf{e}_{SI}^T \mathbf{e}_{SI} = [\mathbf{A}\zeta - \mathbf{e}]^T [\mathbf{A}\zeta - \mathbf{e}]$

然后我们利用球面误差准则, 分两个步骤求解声源的方位, 首先, 求出声源在最小二乘上的矢量坐标 \mathbf{s} 为:

$$\mathbf{s} = \mathbf{Q}^*(\mathbf{e} - \mathbf{S}\mathbf{d}) \quad (30)$$

其中, $\mathbf{Q}^* = (\mathbf{Q}^T \mathbf{Q})^{-1} \mathbf{Q}^T$ 是矩阵 \mathbf{Q} 的伪逆, 将上式带入到方程 $\mathbf{S}^2 = \mathbf{s}^T \mathbf{s}$ 中得:

$$\mathbf{S}^2 = [\mathbf{Q}^*(\mathbf{e} - \mathbf{S}\mathbf{d})]^T [\mathbf{Q}^*(\mathbf{e} - \mathbf{S}\mathbf{d})] \quad (31)$$

将上式化简, 得 $a\mathbf{S}^2 + b\mathbf{S} + c = 0$

其中, $a = 1 - \|\mathbf{Q}^* \mathbf{d}\|^2$, $b = 2\mathbf{e}^T \mathbf{Q}^{*T} \mathbf{Q}^* \mathbf{d}$, $c = -\|\mathbf{Q}^* \mathbf{e}\|^2$ 。

再通过韦达定理, 我们可以求出上述方程的根为 $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$ 。

所以, 声源的位置为:

$$\mathbf{x} = \frac{1}{2} \mathbf{Q}^* (2\mathbf{e} - \frac{\mathbf{d}\sqrt{b^2 - 4ac} - b}{2a}) \quad (32)$$

利用以上公式, 我们可以得到声源和镜像声源的坐标。

6.4 房屋尺寸估量方法

依照上述的方法以及我们获得的六个时延值, 根据求出来的六个坐标, 我们可以判断出 X 轴相距最远的两个点为声源在房间长方向的两个镜像声源; Y 轴相距最远的两个点为声源在房间宽方向的两个镜像声源; Z 轴最大的那个点为声源在房间高方向的一个镜像声源, 我们可以得到一个实际声源位置和五个镜像声源位置, 这五个位置 and 实际声源分别关于房屋除地面外的屋面墙对称, 根据这三组关于对向两面墙的对称的声源位置, 可以得到房屋的大致尺寸。因此, 我们可以得出走路的人所在的位置以及房屋的大致尺寸。我们利用其中一个实际声源位置和任意任意两个相对面所测量出的声源镜像位置, 可得出这两个面之间的距离为:

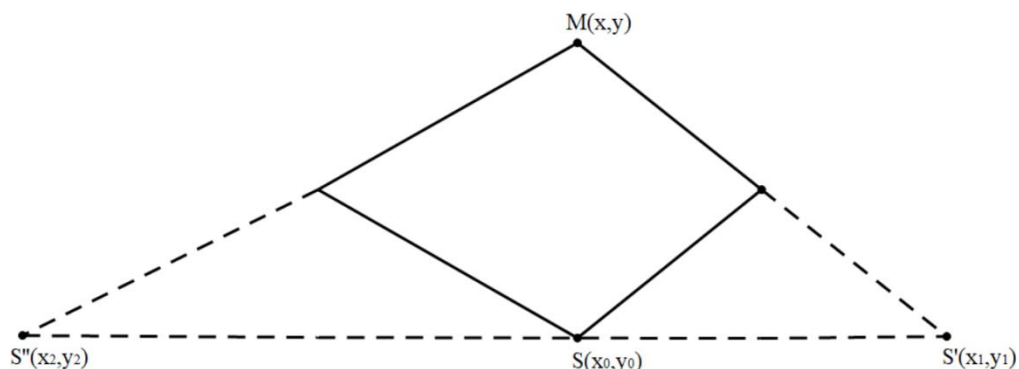


图 11 麦克风 M 与声源 S 及其镜像的位置关系(大厅尺寸未知情况下)

$$L = \sqrt{(x_1^2 - x_2^2)^2 + (y_1^2 - y_2^2)^2} \quad (33)$$

同理，可用相同的方法选取其他对立的镜像声源求出房间的宽度 S 以及高度 H。

6.5 麦克风阵列结构设计

同样在考虑最小成本的前提下，我们希望能使用最少的麦克风来构造出我们的麦克风阵列，而当空间二维变成三维时，麦克风的最少个数也会从三个变为四个，结构也不再是简单的线性拓扑结构这些都是我们要关心的问题。我们参考相关论文^[11]得到不同结构的四元麦克风阵列的准确度。

表 2 四种平面型麦克风阵列的定位精度误差表

时延估计最大相对误差 (%)	一字型麦克风阵列定位误差 (%)	正方形麦克风阵列定位误差 (%)	T 型麦克风阵列定位误差 (%)	四面体麦克风阵列定位误差 (%)
0%	1.6%	2%	0.989%	0.364%
1%	2.4%	2%	1.185%	0.558%
2%	4.2%	2%	1.842%	0.951%
3%	6.0%	3%	2.602%	1.392%
4%	8.0%	4%	3.394%	1.813%
5%	9.7%	5%	4.097%	2.273%
6%	11.5%	6%	4.975%	2.705%
7%	13.5%	7%	5.737%	3.187%
8%	15.1%	8%	6.435%	3.626%
9%	17.3%	8%	7.257%	4.064%
10%	19.2%	10%	8.151%	4.507%
11%	21.2%	10%	8.914%	4.928%
12%	22.4%	12%	10.048%	5.397%
13%	24.7%	13%	10.681%	5.906%
14%	26.5%	13%	11.317%	6.304%

15%	27.6%	15%	12.279%	6.793%
16%	29.6%	15%	13.025%	7.220%
17%	32.1%	16%	13.692%	7.572%
18%	33.7%	17%	14.694%	8.088%
19%	35.2%	19%	15.759%	8.578%
20%	37.1%	20%	16.446%	9.014%

将以上数据以折线图表示出：

四种平面型麦克风阵列的定位精度误差

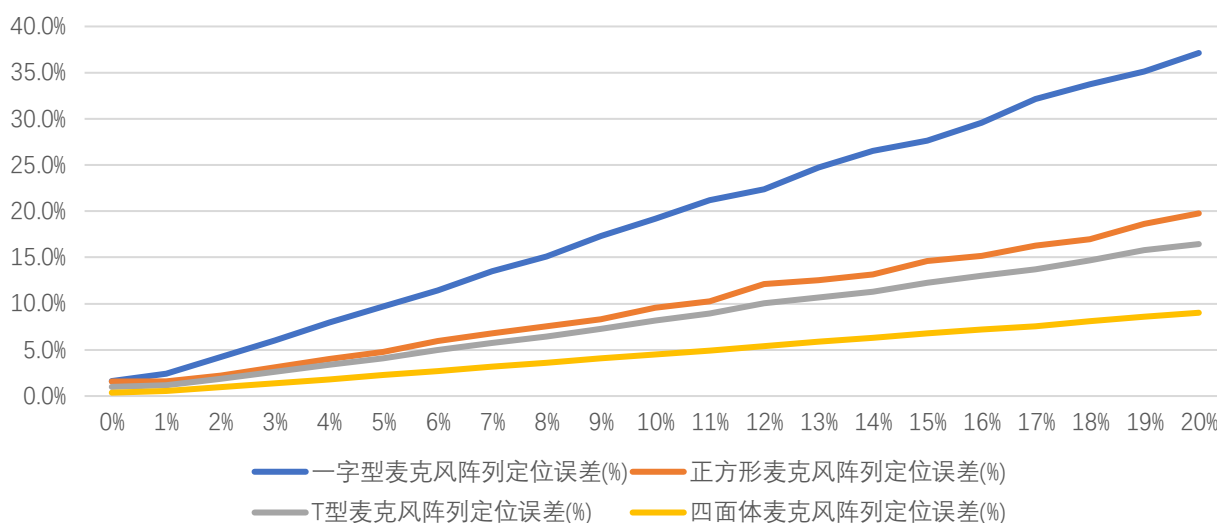


图 12 四种平面型麦克风阵列的定位精度误差对比

从图 12 中我们可以看出，在四种麦克风阵列中，在相同的条件下，一字型麦克风阵列的定位误差最大，正四面体结构的麦克风阵列的定位误差最小，正方形结构和 T 字形结构的结果居中，且 T 字形的定位误差要小于正方形结构。在时延估计误差较小时，除一字型结构外，其他结构的定位误差均较小，但是随着时延估计的不断增大，定位误差也逐渐增大，当达到最大延估计误差 20% 时，正四面体的平均误差小于 10%，而一字型麦克风阵列的平均误差大于 35%，正方形和 T 字形麦克风阵列的误差在 15% 到 20% 之间。我们可以得出，四个麦克风构成的多种拓扑结构的麦克风阵列中，正四面体结构的麦克风阵列其定位误差最小，因此本设计中采用四麦克风构成的正四面体结构麦克风阵列，其中四个麦克风分别位于正四面体的四个顶点。

我们仍使用问题一中判断间距采用的办法，通过步进不断增加间距距离，观察精确度的变化，来确定合适的间距，不过这次的精确度不仅包括声源位置，还要包括房间尺寸。通过仿真，我们得到以下结果：

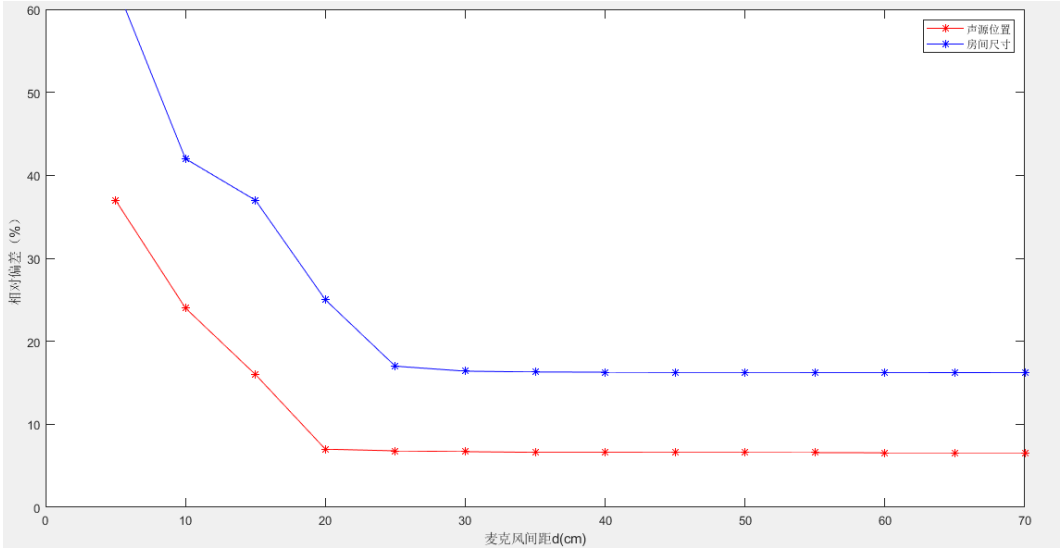


图 13 麦克风间距与相对偏差的关系图

分析以上我们可以看出，在麦克风间距为 20cm 时，对声源位置估计的准确度已经趋于最高，但是对房间尺寸的估计仍有较大的偏差，而在麦克风间距增加到 25cm 时，对声源位置和房间尺寸的计算都达到了较高的准确度。

因此我们麦克风阵列的结构为正四面体，麦克风位于正四面体的每个顶点，麦克风之间的间距(正四面体的边长)为 25cm。

七、模型分析与评价

7.1 敏感性分析

最后，我们对以上建立的模型进行敏感性分析，我们分别逐步增加麦克风个数和麦克风间距，以及麦克风阵列对地面的高度。

表 3 模型一敏感性分析

变量	增加量	声源准确度改变量
麦克风个数	1/个	3.72%
麦克风间距	5/cm	1.73%
麦克风阵列高度	5/cm	1.37%

表 4 模型二敏感性分析

变量	增加量	声源准确度改变量	房间尺寸准确度
麦克风个数	1/个	2.89%	4.71%
麦克风间距	5/cm	3.62%	2.64%
麦克风阵列高度	5/cm	-1.96%	3.77%

在模型一中我们发现虽然在逐步增加的趋势下，我们测量的精度会有所提高但是，提高的精度相对较低，考虑成本的情况下，我们无需增加麦克风个数和间距。

在模型二中，我们观察到精确度改变结果与模型一大致相同，但是高度的增加反而会引起精确度下降，这是因为如果增加麦克风阵列的高度，那么地面也会有一定的反射信号干扰，所以我们应将麦克风阵列的高度尽量低。

7.2 模型的优缺点

本文中构建的模型具有以下优点：

- 我们使用 TODA 算法，相对于其他两种声源确定算法，此方法计算量较少，精度较高，能够较好地计算出声源的位置。
- 我们采用了广义互相关函数，这种函数算法在一定程度上能够减弱室内噪音和混响带来的计算偏差，使我们的结果更加准确。

本文中构建的模型具有以下缺点：

- 麦克风接受的信号与实际接收信号有一定偏差，在模型中只考虑了墙壁反射一次的效果，实际运用中墙面的反射情况要复杂的多。
- 由于麦克风不能放置在较高的空中位置，因此测量大厅高度的误差相对较大。

参考文献

- [1]宗志亚. 基于麦克风阵列的室内声源定位方法研究[D]. 2012, 4 :13
- [2]罗建军. 精讲多练 Matlab[M]. 西安:西安交通大学出版社,2002.
- [3]王永良, 陈辉, 彭应宁等. 空间谱估计理论与算法[M]. 清华大学出版社. 2004 年.
- [4] Carter G. Variance bounds for passively location an acoustic source with a symmetric line array[J]. Journal of Acoustical Society of America,2004,62(2):992-996.
- [5]韩素丹, 吴东娅, 蒋彦雯, 贺维维. 基于麦克风阵列和互相关时延估计的声源定位. 国防科学技术大学电子科学与工程学院[D].
- [6] 崔玮玮, 曹志刚, 魏建强. 声源定位中的时延估计技术[J].数据采集与处理. 2007,22(1):90-99.
- [7]马雯, 黄建国. 广义相关时延估计在被动定位系统中的应用研究[J]. 探测与控制学报. 2000, 22 (3) :51~54
- [8] 张小飞, 汪飞, 徐大专. 阵列信号处理的理论与应用[M]. 北京:国防工业出版社. 2010.
- [9]居太亮, 彭启琮, 邵怀宗. 基于麦克风阵列的近场声源定位子阵算法研究[J]. 电子测量与仪器学报 2006, 10: 第 20 卷;第 5 期

[10]陈颖, 吴长奇. 基于麦克风阵列的声源定位算法研究及 DSP 实现[J]. 燕山大学学报. 2013,1: 第 37 卷第 1 期.

[11]刘云飞. 基于麦克风阵列的声源定位[D]. 杭州电子科技大学. 2018 年 3 月.

[12]张会新, 闫安斌, 王昕, 姜维. 分布式麦克风阵列的声源定位. 火力与指挥控制. 2016 年 2 月:第 41 卷第 2 期.

附录

```
%远场模型,多麦克风
clc
clear all
close all

sita = 60;
load('num.mat');
%产生脉冲响应
c = 340; % 声速
fs = 8e3; % 采样率
room_size = [5 4 6]; % 房间尺寸[x y z] (m)
d0 = 0.2; % 麦克风间距
mic_num = 2; %麦克风个数
C = nchoosek(1:mic_num,2); %所有可能的组合
xm = 0:d0:(mic_num-1)*d0;
delte_m = 0.5*room_size(1) - 0.5*(mic_num-1)*d0;
xm = xm' + delte_m; %每个麦克风的 x 坐标
ym = 1.5.*ones(mic_num,1); %麦克风的 y 坐标
zm = 2.*ones(mic_num,1);
mic_location = [xm,ym,zm]; %麦克风位置 [x y z] (m)
clear xm ym zm delte_m

r = 2; %半径
reverberation_time = 0.5; % 混响时间(s)
n = 2048; % 冲激响应长度
audio_length = fs*1;
s = wgn(audio_length,1,0); %高斯白噪声
% [s,fs0] = audioread('BAC009S0002W0122.wav');
% s = s(1:fs0);
% s = resample(s,fs,fs0);
% plot(s);

ys = r*cos(sita*(pi/180)) + mean(mic_location(:,2));
xs = r*sin(sita*(pi/180)) + mean(mic_location(:,1));
zs = mean(mic_location(:,3));
```

```

source_location = [xs ys zs]; % 声源位置[x y z] (m)
h = rir_generator(c, fs, mic_location, source_location, room_size, reverberation_time, n); %每
一行对应一个冲激响应

%产生每个麦克风接收到的信号
for j = 1:mic_num
    x(:,j) = filter(h(j,:),1,s);
end

for j = 1: length(C(:,1))
    x1 = x(:,C(j,1)); %第一个麦克风
    x2 = x(:,C(j,2)); %第二个麦克风
    d = norm(mic_location(C(j,1,:))-mic_location(C(j,2,:))); %两个麦克风的距离
    P = (fft(x1).*conj(fft(x2))); %x1、x2 的互功率谱
    A = 1./abs(P);
    R_est1 = fftshift(iff(A.*P));
    plot(R_est1);
    [~,tau] = max(R_est1); %寻找相关最大的点
    tau = tau - 0.5*fs - 1;
    t = tau*(1/fs);
    sita_re = asin((t*c)/d)*(180/pi);
    C(j,3) = real(sita_re);
end
% end

% disp(['来波方向:',num2str(sita),'度']);
% disp(['估计方向:',num2str(sita_re),'度']);

% hold on
% plot(SITA(:,1),SITA(:,1));
% plot(SITA(:,1),SITA(:,2));
% hold off

```