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　基于麦克风树设计的声源定位研究

摘　要

随着多媒体技术的进一步发展，听声辩位技术已经被广泛应用于各领域。基于麦克风树的声源定位技术是阵列信号处理中的关键技术之一，是声音信号处理领域一个新的研究热点。针对声源方位进行定位、追踪成为声音信号处理领域中关键的组成部分，本文运用麦克风树技术对室内声音定位展开研究。

对于问题一，通过对声音信号进行预处理，并在分析了传统四元十字麦克风阵列定位算法的基础上，为了提高室内声场中声源定位的精确性与实时性，提出了一种基于 PHAT 的三维七元麦克风阵列的声源定位方法，该方法建立了七元麦克风阵列模型与室内声场模型，利用该麦克风阵列接收室内声源声音信号，对接收的声音信号进行去噪、去混响处理，再利用相位变换加权广义互相关方法，得各麦克风之间的时间延迟，运用三维空间定位方法与坐标旋转数字式计算机方法，确定声源位置。

对于问题二，针对室内轮廓构图原理进行系统理论研究，在理论分析的基础上提出了两种声音信号的处理算法，欧式距离法和 TDOA 法，并对接收到的声音信号进行优化处理，通过提出一种基于遗传算法的麦克风树优化设计算法，对不同阵列形式麦克风树进行优劣对比耦合分析，最终选取大厅轮廓、声源位置估算结果较为精准、计算过程较为简洁的十字阵列麦克风树进行现场试验研究并详细给出了每支麦克风的相对位置信息，最后的现场试验结果数据证实了 TDOA 算法的优越性，验证了麦克风树阵列设计的合理性，整体平均误差为 0.05m，满足实用要求。

关键词：相位变换加权广义互相关，时间延迟，信号预处理，TDOA 法，对比分析

Abstract

With the further development of multimedia technology, listening to speech position technology has been widely used in various fields. Microphone tree-based sound source localization technology is one of the key technologies in array signal processing, and is a new research hotspot in the field of sound signal processing. Positioning and tracking the sound source orientation has become a key component in the field of sound signal processing. This paper uses microphone tree technology to conduct research on indoor sound localization.

For problem one, based on the preprocessing of the sound signal and the analysis of the traditional four-element cross microphone array localization algorithm, in order to improve the accuracy and real-time nature of the sound source localization in the indoor sound field, a PHAT-based A sound source localization method for a three-dimensional seven-element microphone array. This method establishes a seven-element microphone array model and an indoor sound field model. The microphone array is used to receive indoor sound source sound signals, and the received sound signals are denoised and deriver berated Then, the phase transformation weighted generalized cross-correlation method is used to obtain the time delay between the microphones. The three-dimensional space positioning method and the coordinate rotation digital computer method are used to determine the position of the sound source.

For problem two, a systematic theoretical study was conducted on the principle of indoor contour composition. Based on the theoretical analysis, two sound signal processing algorithms were proposed, the Euclidean distance method and the TDOA method, and the received sound signal was optimized. A microphone tree optimization design algorithm based on genetic algorithm, the coupling analysis of the advantages and disadvantages of different array microphone trees is selected, and finally the cross-array microphone tree with more accurate estimation of the hall contour and sound source position estimation and simpler calculation process is selected for field test The relative position information of each microphone was studied and given in detail. The final field test data confirmed the superiority of the TDOA algorithm and verified the rationality of the microphone tree array design. The overall average error was 0.05m, which met practical requirements.

Keywords: Phase transform weighted generalized cross-correlation, Time delay, Signal preprocessing, TDOA method, Comparative analysis

目录

一、问题重述

把若干(⩾1)支同样型号的麦克风固定安装在一个刚性的枝形架子上(架子下面带万向轮，在平地上可以被水平推动或旋转，但不会歪斜)，这样的设备称为一个麦克风树。不同的麦克风由于位置不同，录制到的声音往往也有细微的不同，所以通过对多支麦克风接收到的声音进行对比分析，可以得到更多的有关声源的信息。我们假设每个麦克风都是全向的，也就是单麦克风无法分辨声源的方向。现在有一个地面、墙壁和天花板都是光滑大理石的大厅，大厅内空旷而安静。在大厅里只有一个走动的人，发出清晰的脚步声。我们准备在大厅里安放一个麦克风树，希望通过检测声音来进行一些测量，包括尽量准确地实时确定这个人的位置，也包括测量这个大厅的某些几何参数。请你建立合理的数学模型，设计一个成本尽量低、而且可以达到使用要求的麦克风树。要求给出每支麦克风的相对位置以及相对于地面的高度，至于枝形架子的具体力学结构则不需要考虑。

根据已知和所求信息的不同，这个问题可以有不同的复杂程度。我们需要在以下两种情形中对麦克风树进行合理的设计，并给出对接收到的声音信号进行分析的算法。

1、已知大厅的平面形状是矩形，地面和天花板都是水平的。假设我们已知大厅轮廓的准确尺寸(长×宽×高)。但在实际施工中，由于操作和设备所限，在麦克风树放置在地面上的时候，无法精确测量放置点的坐标以及水平的旋转角度。我们希望确定此人的位置。

2、大厅轮廓的尺寸未知，其余条件同上。我们希望确定此人的位置，并尽 量准确地确定大厅的轮廓尺寸(长×宽×高)。

二、问题分析

2.1 问题一的分析

问题一主要解决的是分辨声源位置，通过声波传输、反射和多径传输以及麦克风接受声波进行处理，并通过分析延时估计法，利用不同的麦克风接收到声源信号的时延差以及麦克风之间的几何位置关系来定位声源。

对于声源定位问题，我们可以首先建立一个四元十字麦克风阵列定位算法，然后建立一种基于 PHAT 的三维七元麦克风阵列的声源定位方法，对结果分别进行预测，并将结果进行比较。

2.2 问题二的分析

对于本问题中确定人的位置与大厅轮廓这个问题，人的位置（声源）信息确认相对较为容易，可在接收声音信号确定大厅轮廓过程中进行相应解决，因此本问题的主要难点在于如何较为准确的确定大厅轮廓尺寸。

针对于本问题，我们可以建立麦克风树的优化设计算法对麦克风树阵列进行合理设计，并建立两种不同的算法来确定声源位置信息和大厅轮廓尺寸，最后通过合理的现场试验来进行对比耦合确定计算结果较为优秀的算法加以采纳，验证麦克风树阵列设计的可行性。

三、模型假设

1. 假设麦克风阵列为线性阵列；
2. 假设麦克风间距为定值；
3. 假设在大厅内的障碍物必须保证障碍物尺寸小于声波波长；
4. 假设在大厅内的任意位置都可以接收到完整、清晰易于判别的脉冲信号；
5. 假设不存在其他因素对人的位置信息、大厅轮廓估算结果造成影响。

四、定义与符号说明

符号	定义
$S_w(n)$	加窗后的信号
$s(n)$	原始信号
$w(n)$	窗函数
N	窗口函数的长度
n	该帧信号在原始信号中的起始位置
$\mathrm{sgn}[\ ]$	符号函数
$R_{ss}(\tau)$	声源的自相关函数
$R_{v_iv_j}(\tau)$	噪声的互相关函数
$G_{ij}(\omega)$	麦克风m_i和m_j采集到信号的互功率谱。
α	过减因子
β	增益补偿因子。
FFT	表示傅里叶变换
IFFT	表示傅里叶反变换
ln	表示自然对数
$h(t)$	房间时域冲激响应
a_i	每一个信道的增益
τ_i	每一个信道的时延
s	声源坐标
s_i	镜像声源坐标
p_i	声源一阶镜像声源点连线与墙线交点坐标
$s(t)$	系统输入信号
$y(t)$	系统输出信号
$h(t)$	系统的冲激响应

V	房间体积
S_k	室内反射面的吸声系数
α_k	室内反射面的面积
c	声信号在空气中的传播速度
r_o	第$1\sim n$个接收端的位置
s_i	待求的第$1\sim m$个实声源或者镜像声源
t_{io}	声源发出的信号传播到第$1\sim n$个接收端的传播时间

五、模型的建立与求解

5.1 问题一的解答

5.1.1 信号预处理

由于麦克风收到的信号直接来自于室内环境，因此其信号中除了声源发出的有用信号，还会混杂了很多环境噪声，所以在进行声源定位之前，首先要对信号进行预处理。声音信号的预处理一般分为三个步骤，预滤波、加窗分帧和端点检测，如图 5.1 所示。

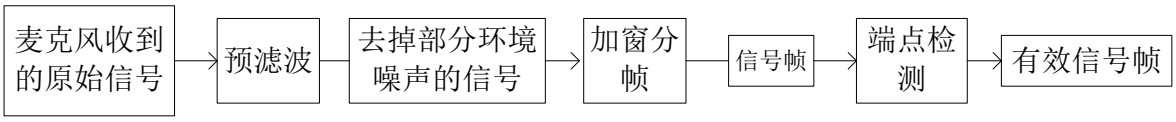


图 5.1 信号预处理流程图

Fig.5.1 Flow chart of signal preprocessing

预滤波主要用于滤除接收信号中的环境噪声。因为声音信号属于非平稳信号，只具有短时平稳性，所以需要通过加窗分帧来获得短时平稳的信号段。声音信号不是连续持续的信号，很多时候会不包含有用信息，因此需要对声音信号进行端点检测，只对有用信号进行声源定位。

5.1.1.1 预滤波

预滤波主要用于对麦克风收到的声音信号中的环境噪声进行滤波。一般情况下，麦克风接收到的来自于声源可用于进行声源定位的有用信号往往处于一个有限的频段内，所以预滤波一般采用的是带通滤波，去除有效频带之外的高频噪声及低频噪声[1]。人耳能感受到的振动频率是 $20Hz \sim 20kHz$，而人类语言的频率主要分布在 $300Hz \sim 3000Hz$，因此，在对人的声音信号进行处理时，可设带通滤波器的上截止频率 $f_h = 3000Hz$，下截止频率 $f_l = 300Hz$。在其他要求较高的场景下，可根据有效频带的范围适当改变上下截止频率，但一般不超过人耳能感受的振动频率。

预滤波，只能够去除有效频带之外的噪声信号，但是对于有效频带内的噪声信号则不做处理，因此预滤波只能够实现对信号的初步处理。但是通过预滤波，可以有效排除部分噪声，从而有效降低噪声信号对声源定位算法的影响，减小定位误差，因此对提高定位精度是十分必要的[1]。

5.1.1.2 加窗分帧

在基于麦克风阵列的声源定位系统中，麦克风接收到的信号为非平稳信号，只具有短时平稳性，所以必须要将信号变成短时的信号帧才能进行处理，这就需要对信号进行分帧处理[2]。通常情况下，可认为声音信号在 $10ms \sim 30ms$ 内是平稳的。所以每帧信号的时长一般取 $10ms \sim 30ms$。

对声音信号的分帧处理可分为连续分段和交叠分段两种方式。

连续分段是连续且不重叠地对声音信号进行截取，按照声音信号的平稳特性，可得到每秒的帧数为 33~100 帧。可根据实际需求进行选择。

在实际应用中，为让相邻的两帧之间过渡平滑，通常会采用交叠分段的方式。交叠分段得到的声音段中，相邻的两段声音会有一段固定长度的重叠，该重叠段称为帧移，帧移的长度一般不超过帧长的 1/2。

分帧是采用固定长度的窗口对原始信号进行加权处理来实现的，窗口的长度决定帧长，得到的加窗后的一帧声音信号可表示为

$$S_w(n) = s(n) \cdot w(n) \tag{5.1}$$

式中，$S_w(n)$ 为加窗后的信号，$s(n)$ 为原始信号，$w(n)$ 称为窗函数。

窗函数也叫做截断函数，采用不同的窗函数，对于截取的信号的频谱泄漏的大小有着不同的影响，且频率分辨率也不同，因此选择不同的窗函数，会对声源定位的结果产

生一定的影响。

在数字声音信号处理中，常采用矩形窗和汉明窗对信号进行处理，设窗长为N，则矩形窗可表示为

$$w(n)=\begin{cases}1 & (0\le n\le N-1)\\ 0 & else\end{cases} \tag{5.2}$$

而汉明窗的表达式为

$$w(n)=\begin{cases}0.54-0.46\cos\left[2\pi/(N-1)\right] & (0\le n\le N-1)\\ 0 & else\end{cases} \tag{5.3}$$

通过对两种窗函数的频率特性进行分析，可得到表 5.1

表 5.1 矩形窗和汉明窗对比表

Tab.5.1 Comparison between rectangular window and Hamming window

窗类型	矩形窗（Rectangle）	汉明窗（Hamming）
中心峰全宽	2.0	4.0
最高旁瓣高度	-13	-41
主瓣峰值	$4\pi/N$	$8\pi/N$
最小阻带衰减	-21	-53

由表 5.1 可以看出，汉明窗的主瓣宽度为矩形窗的 2 倍，这说明汉明窗的带宽大约增加了一倍。而矩形窗虽然具有较好的平滑性，但是却损失部分的高频信号的波形，从而丢失部分细节。

利用计算机对声音信号进行加窗分帧处理，可得到结果如下，其中图 5.2 是原始信号图，共持续 0.1 秒。

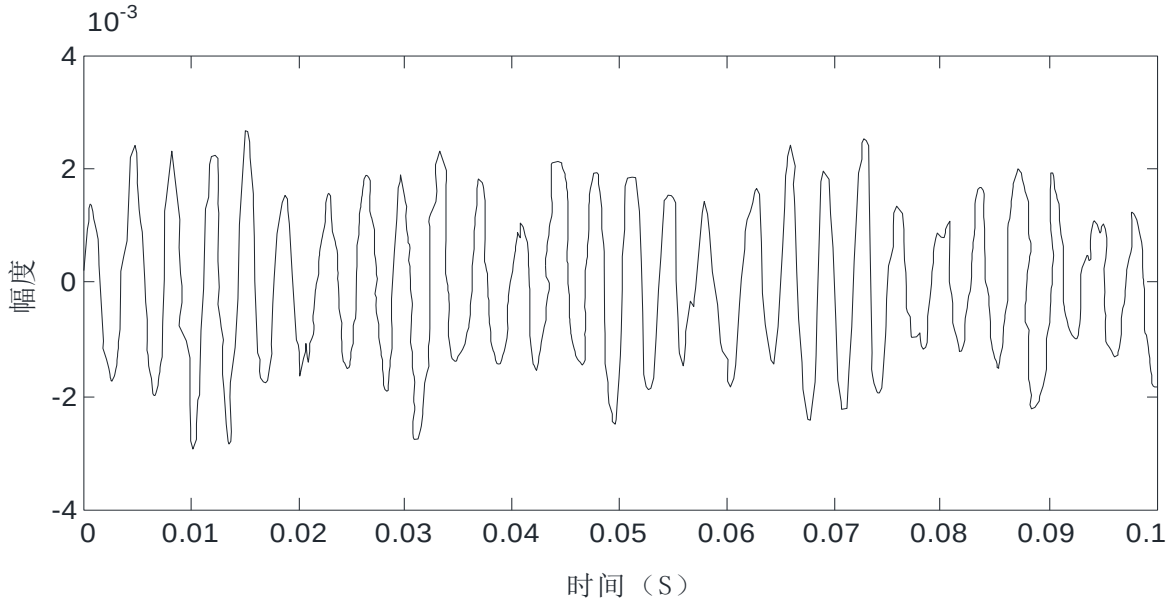


图 5.2 未加窗分帧的原始信号

Fig.5.2 The original signal without windowing and framing

图 5.3 和图 5.4 分别对原始信号进行加窗分帧处理后的波形，其中，帧长为30ms。

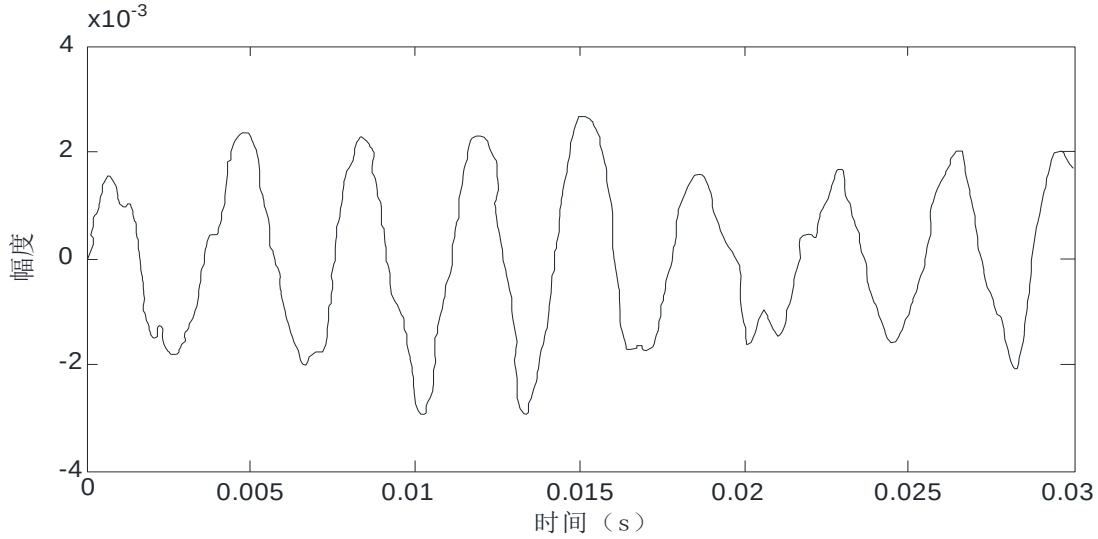


图 5.3 矩形窗进行加窗分帧后的波形

Fig.5.3 The waveform of the rectangular window after windowing and framing

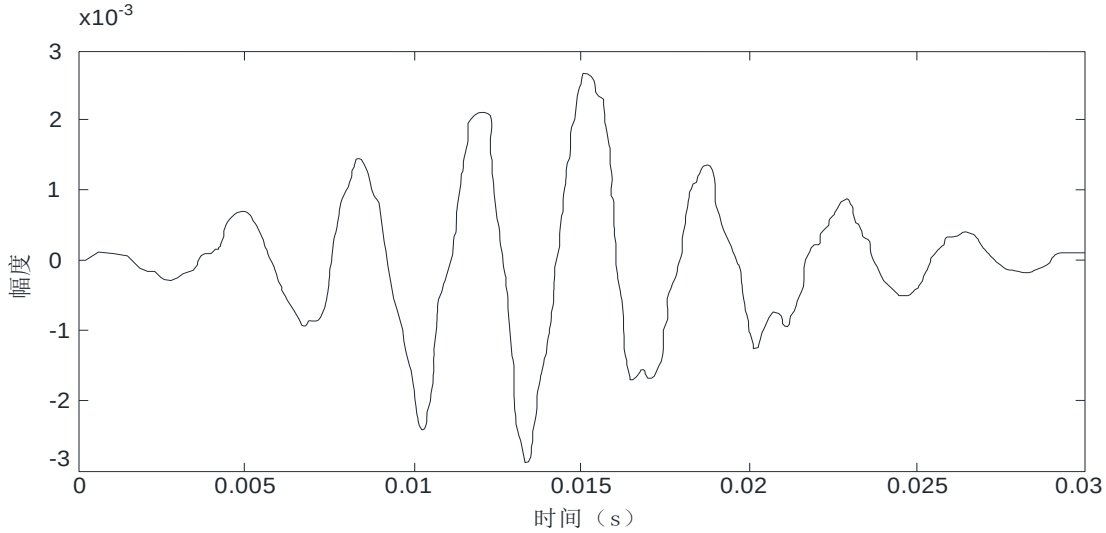


图 5.4 汉明窗进行加窗分帧后的波形

Fig.5.4 Waveform of Hamming window after windowing and framing

5.1.1.3 端点检测

通常情况下，声音并不是连续不断的，中间会有很多停顿，因此麦克风收到的声音信息中并不是任何时刻都含有有用信息。这就意味着，经过加窗分帧后的声音段并不是每一段都含有有效定位信息，不含有效定位信息的声音段称为噪声帧。如果对噪声帧信息进行声源定位的运算，则不仅会加大系统的计算量，影响系统实时性，而且在一些前一帧的定位结果对后一帧的结果有影响的声源定位算法中，对噪声帧的定位运算结果会影响有效帧的定位结果，从而引入多余的误差，影响定位精度。因此，必须能够正确的过滤噪声帧，而我们可以利用声音信号本身的特征，通过对声音信号的端点检测来避免噪声帧的干扰[3,4]。

设 $x(m)$ 为经过加窗分帧处理之后得到的帧信号，则 $x(m)$ 可以表示为

$$x(m) = w(m)x(n+m) \qquad (0 \le m \le N-1) \tag{5.4}$$

其中， N 为窗口函数的长度， n 是该帧信号在原始信号中的起始位置。

利用每一帧信号的短时能量可以区分有效帧和噪声帧。 $x(m)$ 的短时能量 E_n 为

$$E_n = \sum_{m=0}^{N-1} x^2(m) \tag{5.5}$$

根据实际情况，可以设定合适的阈值，当E_n大于该阈值，则可以认为该信号帧属于有效帧。但是在实际应用过程中，利用短时能量进行端点检测有一些不足。由于在计算E_n的过程中设计平方运算，就会导致该结果对信号的电平过于敏感，可能会将包含突发的极短高电平信号的噪声帧识别为有效帧，有一定的识别误差。因此可采用短时平均幅值来代替短时平均能量，以近似表征短时能量。其数学表达式如下：

$$M_n = \sum_{m=0}^{N-1} \left| x_n(m) \right| \tag{5.6}$$

与通过短时能量进行端点检测的方法相似，通过设定合理的阈值来区分有效帧和噪声帧。

除短时能量外，信号的短时平均过零率也可以用于端点检测。短时平均过零率是指一帧信号中通过零的次数所占的比例。对于麦克风采集到的离散信号，若相邻的两个样本点之间的正负号不同，则称为“过零”，表明此时信号在时间域上通过零电平，然后对该帧信号的过零次数进行统计平均，就可以得到此帧的平均过零率[5]。

对于每一帧信号，其过零平均率可定义为：

$$Z_n = \frac{1}{2} \sum_{m=0}^{N-1} \left| \mathrm{sgn}\left[x_n(m) \right] - \mathrm{sgn}\left[x_n(m-1) \right] \right| \tag{5.7}$$

其中，sgn[ ]是符号函数，为：

$$\mathrm{sgn}\left[x \right] = \begin{cases} 1, (x \geq 0) \\ -1, (x < 0) \end{cases} \tag{5.8}$$

因为在相邻的采样点过零时，其$\left| \mathrm{sgn}\left[x_n(m) \right] - \mathrm{sgn}\left[x_n(m-1) \right] \right| = 2$，因此需要将统计后的结果乘上 1/2。在实际应用过程中，麦克风收到的声音中会包含直流偏移及噪声干扰，对过零率会造成较大的干扰，为避免其干扰，可以把过零率改为过门限率。设零电平上下的正负门限为$\pm T$，则短时平均过门限率可以表示为：

$$\begin{aligned} Z_n = \frac{1}{4} \sum_{m=0}^{N-1} \{ & \left| \mathrm{sgn}\left[x(m)T \right] - \mathrm{sgn}\left[x(m-1) - T \right] \right| \\ & + \left| \mathrm{sgn}\left[x(m) + t \right] - \mathrm{sgn}\left[x(m-1) + T \right] \right| \} \end{aligned} \tag{5.9}$$

过门限率表明电平通过正负门限的次数，其优势在于比过零率增加抗噪声能力，排除过零但没有超过$[-T, +T]$范围的情况，从而提高端点检测的效果。

在实际应用中，为提升端点检测的准确性，保证能够充分滤除噪声帧，可以用能量检测和过门限检测两种方法同时对信息帧进行检验，信号帧只有通过两种端点检测，才会被认为是有效声音帧[6-8]。结合两种方式的端点检测的系统流程图如图 5.5

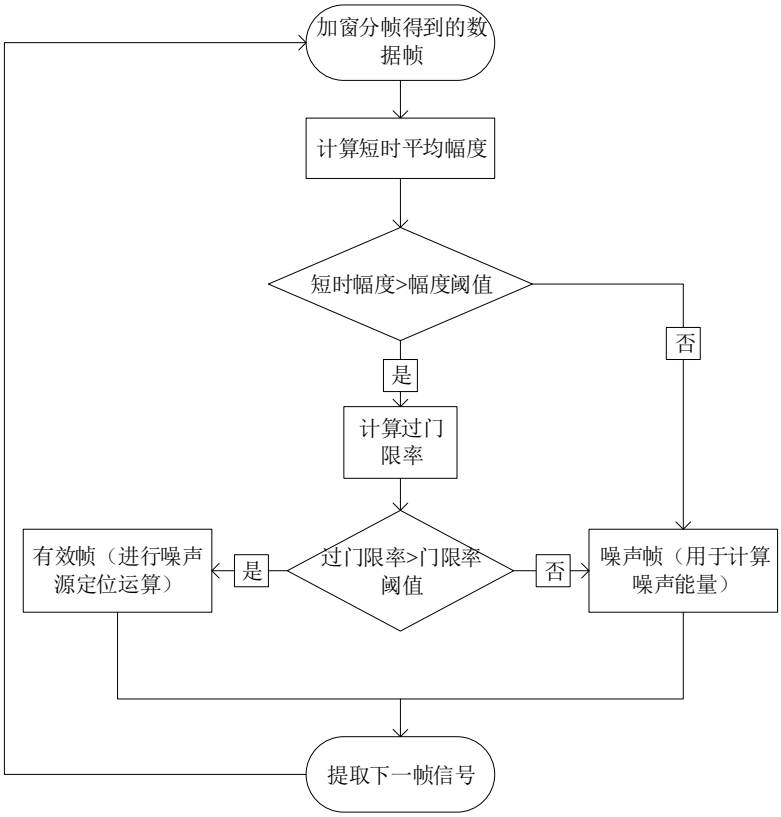
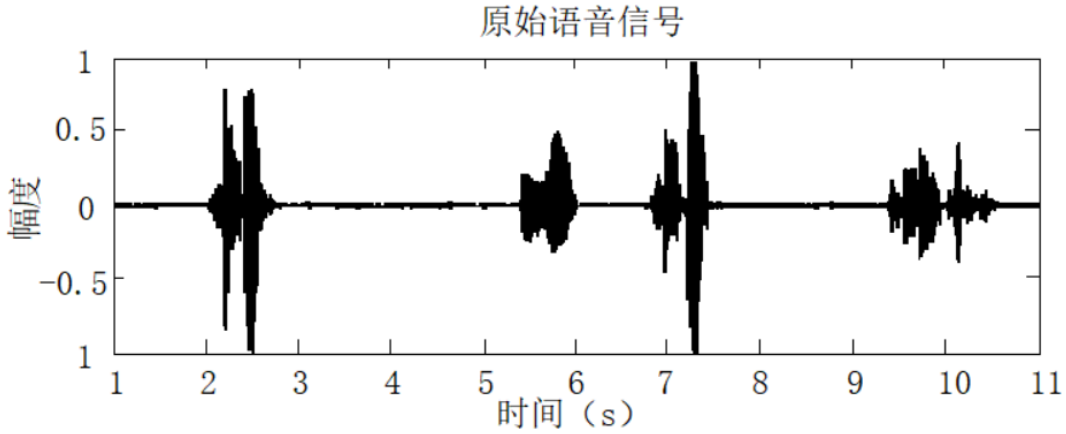


图 5.5 端点检测流程图

Fig.5.5 Endpoint detection flowchart

图 5.5 中，首先对信号进行短时幅度的检测，通过短时幅度检测，可以滤除部分不含有效声音信号的环境噪声，然后对上一步的结果进行短时过零率检测，通过这一步检测，可以滤除环境中的短时突发噪声。当两次检测都能通过时，该帧信号被认为是有效帧。

图 5.6 是对端点检测进行的计算机仿真的结果图。声音信号采用的是在实际室内环境下录制的声音信号，采样频率 $f=22.05kHz$，采样时间 $T=11s$，通过端点检测后，结果如图 5.6 所示，其中在检测结果图中，值为 1 代表该信号为有效信号。



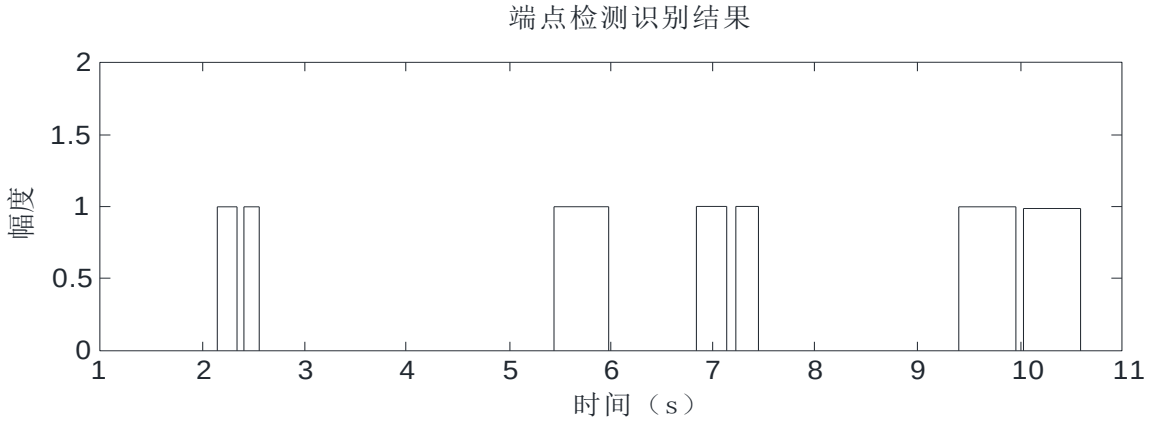


图 5.6 端点检测结果

Fig.5.6 Endpoint detection results

可以看出，通过端点检测，可以准确的识别出有效信号和非有效信号。

5.1.1.4 时间延迟

在室内环境下利用麦克风阵列在进行声源定位时，为得到声源的精确位置，需要考虑每个麦克风相对于声源位置的方向角的不同，所以将声源发出的波形用球面波表示。在不考虑反射的情况下，麦克风位置的不同，会使得声音到达每个麦克风的路径也会有所不同，当两条路径的距离不同时，不同的路径会导致声源到达麦克风的时间的不同，该时间的差值就是所述的时间延迟，即时延（TDOA）。

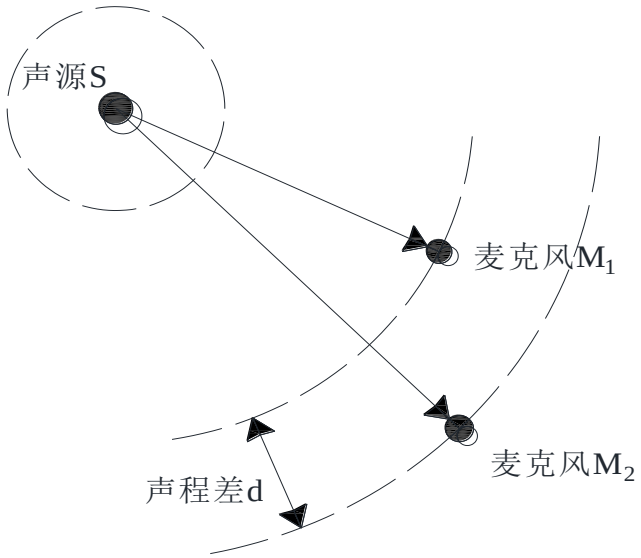


图 5.7 声波传播声程差示意图

Fig.5.7 Schematic diagram of sound path difference of sound wave propagation

声源到两个不同的麦克风之间的距离差称作声程差，由物理学知识可知，声程差和时延的关系为

$$d = v \cdot \tau \qquad (5.10)$$

在这里v指声速，通过式(4.1)，可以把时延差转换为声程差，然后通过得到的声程差，利用几何知识即可得到声源位置，因此式(5.10)是基于 TDOA 定位算法的两个步骤的之间的桥梁，而时延是算法的基础，时延的误差会对相应的产生定位结果的误差。

5.1.2 基于 PHAT 的三维七元麦克风阵列声源定位算法

5.1.2.1 传统四元十字麦克风阵列声源定位算法

广义互相关（Generalized Cross Correlation, GCC）法是一种传统时延估计方法[9,10]，根据麦克风阵列拾取声源信号间的相关性，估计时延进而求得声源位置。

在麦克风阵列中，对麦克风 m_i 、 m_j ，根据式（5.11）：

$$x_i(n)=\alpha_i s(n-\tau_i)+v_i(n) \tag{5.11}$$

可得其采集信号的互相关函数 $R_{ij}(\tau)$ 为：

$$R_{ij}(\tau)=\alpha_i\alpha_j E\left[s(n-\tau_i)s(n-\tau_i-\tau)\right]+\alpha_i\left[s(n-\tau_i)v_j(n-\tau)\right] \tag{5.12}$$
$$+\alpha_j E\left[s(n-\tau_i-\tau)v_i(n-\tau)\right]$$

根据声源信号与噪声信号的不相关性，式（5.12）巧化简为：

$$R_{ij}(\tau)=\alpha_i\alpha_j R_{ss}\left(\tau-\tau_{ij}\right)+R_{v_iv_j}(\tau) \tag{5.13}$$

式中， $R_{ss}(\tau)$ 为声源的自相关函数， $R_{v_iv_j}(\tau)$ 为噪声的互相关函数。

由互相关函数与互功率谱关系得：

$$R_{ij}(\tau)=\int_0^{\pi} G_{ij}(\omega)e^{-j\omega\tau}d\omega \tag{5.14}$$

式中， $G_{ij}(\omega)$ 为麦克风 m_i 和 m_j 采集到信号的互功率谱。

根据 $R_{ij}(\tau)$ 的峰值估测出时延值 τ_{ij} 。

然后利用四元麦克风阵列，如图 5.8 所示，求得声源的角度位置 φ 、 θ ，即

$$\varphi=\arctan\frac{2R_1d_{14}+d_{14}^{\ 2}}{2R_1d_{12}+d_{12}^{\ 2}} \tag{5.15}$$

$$\theta=\arccos\sqrt{\frac{\left(\frac{2R_1d_{14}+d_{14}^{\ 2}}{2L}\right)^2+\left(\frac{2R_1d_{12}+d_{12}^{\ 2}}{2L}\right)^2}{R_1^{\ 2}}} \tag{5.16}$$

式中， $d_{12}=\tau_{12}c$ 、 $d_{13}=\tau_{13}c$ 、 $d_{14}=\tau_{14}c$ ， τ_{12} 、 τ_{13} 、 τ_{14} 为时延值。

$$R_1=\frac{d_{13}^{\ 2}-d_{12}^{\ 2}-d_{14}^{\ 2}}{2(d_{12}+d_{14}-d_{13})} \tag{5.17}$$

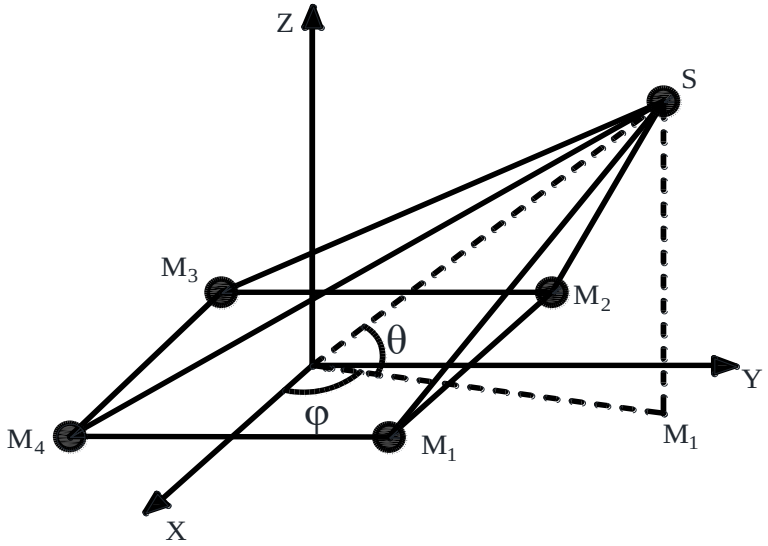


图 5.8 四元麦克风阵列

Fig.5.8 The four microphone array

5.1.2.2 基于相位变换加权的广义互相关七元麦克风阵列声源定位算法

1、相位变换加权的广义互相关算法

相位变换加权的广义互相关算法是在广义互相关法中引入相位变化加权函数。设

$x_a(t)$ 为麦克风阵列中第 a 个麦克风所接受到的声音信号，依据时-频域二元性，通过傅里叶变换，其频域信号为 $X_a(f)$。

在时域中的自相关函数表示为:

$$R_{x_a x_a}(\tau)=\int_{-\infty}^{+\infty} x_a(t)x_a(t+\tau)dt \tag{5.18}$$

在频域中的自相关函数表示为：

$$R_{X_a X_a}(\tau)=\int_{-\infty}^{+\infty} X_a(f)X_a^{*}(f)e^{j2\pi f\tau}df \tag{5.19}$$

若 x_b 为麦克风阵列中第 b 个麦克风所接受到的声音信号，则 $x_a(t)$ 与 $x_b(t)$ 的互相关函数表示为：

$$R_{X_a X_b}(\tau)=\int_{-\infty}^{+\infty} X_a(f)X_b^{*}(f)e^{j2\pi f\tau}df \tag{5.20}$$

如果 $x_b(t)$ 是 $x_a(t)$ 的延迟信号，式（5.19）中通过搜索峰值得到对应的时延，表示为：

$$\tau_{ab}=\arg\max_{\tau} R_{X_a X_b}(\tau) \tag{5.21}$$

式中，$\arg\max$ 表示使 $R_{X_a X_b}(\tau)$ 取得最大值时 τ_{ab} 的取值。

考虑到在有限的时间间隔里需要得到准确、高精度的时延。相位变换（Phase Transform, PHAT）加权的广义互相关方法（GCC-PHAT）的优点在于，在宽带信号中，利用加权函数，可以较为精确地估计出时延，并对噪声和混响都有较强的抑制力，加权函数表示为：

$$\psi_{PHAT}(f)=\frac{1}{\left|G_{X_a X_b}(f)\right|} \tag{5.22}$$

$G_{X_a X_b}(f)$ 为相关函数的功率谱，表示为：

$$G_{X_a X_b}(f)=\int_{-\infty}^{+\infty} R_{x_a x_a}(\tau)e^{j2\pi f\tau}df \tag{5.23}$$

$$R_{x_a x_a}(\tau)=E\left[x_a(t)x_b(t+\tau)\right] \tag{5.24}$$

式中，E 表示期望值

对于 $x_a(t)$ 、 $x_b(t)$ 声音信号，由 PHAT-GCC 得到的相关函数，表示为：

$$R^{(g)}_{x_a x_a}(\tau)=\int_{-\infty}^{+\infty} \psi_{PHAT}(f)X_a(f)X_b^{*}(f)e^{j2\pi f\tau}df \tag{5.25}$$

时延值，表示为：

$$\hat{\tau}_{ab}=\arg\max R^{(g)}_{X_a X_b}(\tau) \tag{5.26}$$

式中，$\arg\max$ 表示使 $R^{(g)}_{X_a X_b}(\tau)$ 取得最大值时 $\tau=\hat{\tau}_{ab}$，即 $\hat{\tau}_{ab}$ 为第 a 个麦克风与第 b 个麦克风所接收到信号的时延估计值。

理论上，PHAT 是考虑对信号与白化处理而导出的，它的加权适用于宽带信号，并且其在大信噪比的条件下效果更为显著，在实际应用中，可能会降低相关处理的信噪比。但是我们必须注意的是，PHAT 权函数一个明显的缺点是，由于加权 $\psi_{PHAT}(f)$ 是 $G_{X_a X_b}(f)$ 的倒数，故其在能量小的地方会出现大的误差，从而影响时延精度。

2、基于 PHAT 的三维七元麦克风阵列声源定位算法

建立七元麦克风阵列模型，如图 5.9 所示。

坐标原点 $O(0,0,0)$，七个麦克风的位置坐标分别为 $M_0(0,0,0)$ 、 $M_1(L,0,0)$，$M_2(0,L,0)$，$M_3(-L,0,0)$，$M_4(0,-L,0)$，$M_5(0,0,L)$，$M_6(0,0,L)$，L 表示第 i 个麦克风位置 M_i 与坐标原点处第 0 个麦克风位置 M_0 间的距离；声源位置的直角坐标为 $S(x,y,z)$，

方位角为φ（$0°\le\varphi\le360°$），俯仰角为θ（$0°\le\varphi\le90°$）。声源到坐标原点（就是第 0 个麦克风的位置M_0）的距离为r，在远场条件下阵列接收的信号为平面波。d_i（$1\le i\le6$）为声源到第 0 个麦克风的位置M_0与声源到第i个麦克风的位置M_i的距离差值。

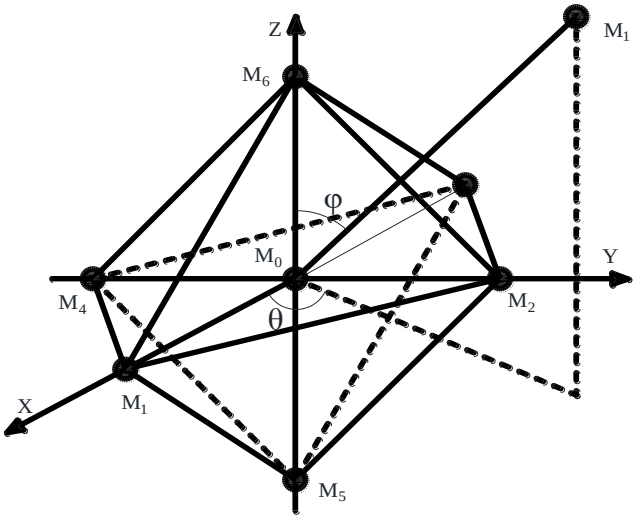


图 5.9 七元麦克风阵列

Fig.5.9 The seven microphone array

空气中声速公式，表示为：

$$c=20.05\sqrt{273.15+T} \tag{5.27}$$

式中T代表室温，单位为$°C$。

由麦克风阵列与声源的几何位置关系得：

$$\begin{cases} OS: x^2+y^2+z^2=r^2 \\ SM_1: (x-L)^2+y^2+z^2=(r+d_1)^2 \\ SM_2: x^2+(y-L)^2+z^2=(r+d_2)^2 \\ SM_3 := (x+L)^2+y^2+z^2=(r+d_3)^2 \\ SM_4 := x^2+(y+L)^2+z^2=(r+d_4)^2 \\ SM_5 := x^2+y^2+(z+L)^2=(r+d_5)^2 \\ SM_6 := x^2+y^2+(z-L)^2=(r+d_6)^2 \end{cases} \tag{5.28}$$

由式（5.28）得：

$$r=\frac{6L^2-c^2\sum_{i=1}^{6}(\hat{\tau}_i)^2}{2c\sum_{i=1}^{6}\hat{\tau}} \tag{5.29}$$

$$\begin{cases} \tan\varphi=\dfrac{(\hat{\tau}_2-\hat{\tau}_4)\left[2r+c(\hat{\tau}_2+\hat{\tau}_4)\right]}{(\hat{\tau}_1-\hat{\tau}_3)\left[2r+c(\hat{\tau}_1+\hat{\tau}_3)\right]} \\ \tan\theta=\dfrac{\sqrt{\left\{(\hat{\tau}_1-\hat{\tau}_3)\left[2r+c(\hat{\tau}_1+\hat{\tau}_3)\right]\right\}^2+\left\{(\hat{\tau}_2-\hat{\tau}_4)\left[2r+c(\hat{\tau}_2+\hat{\tau}_4)\right]\right\}^2}}{(\hat{\tau}_6-\hat{\tau}_5)\left[2r+c(\hat{\tau}_6+\hat{\tau}_5)\right]} \end{cases} \tag{5.30}$$

在远场$r\ge c\hat{\tau}_i$（$1\le i\le6$），式（5.30）可化简为：

$$\begin{cases} \tan\varphi = \dfrac{(\hat{\tau}_2 - \hat{\tau}_4)}{(\hat{\tau}_1 - \hat{\tau}_3)} \\ \tan\theta = \dfrac{\sqrt{(\hat{\tau}_1 - \hat{\tau}_3)^2 + (\hat{\tau}_2 - \hat{\tau}_4)^2}}{(\hat{\tau}_6 - \hat{\tau}_5)} \end{cases} \tag{5.31}$$

式中，$\hat{\tau}_i$（$1 \le i \le 6$）表示为第 0 个麦克风与第i个麦克风所接收到信号的时延估计值。

在此我们需要计算反正切函数以确定角度φ、θ，由于反正切函数为超越函数，此处利用坐标旋转数字式计算机（CORDIC）算法来计算，一方面由于其计算简洁，仅由简单移位和加减运算组成，另外其具有较高的迭代精度。

由式（5.29）得的r与式（5.31）得φ、θ，因此声源的球面坐标为(r,φ,θ)。

再根据球坐标公式为：

$$\begin{cases} x = r\sin\varphi\cos\theta \\ y = r\sin\varphi\sin\theta \\ z = r\cos\varphi \end{cases} \tag{5.32}$$

可确定声源位置坐标(x,y,z)。

3、声音信号处理

在混响、噪声等干扰信号存在的室内环境下，为了提高定位精度，我们需要对用七元麦克风阵列采集到的声音信号进行处理，去除声音信号中各种干扰，使定位精度提高。此处我们针对声音信号特征，依次考虑噪声、室内混响等对于采集到的声音信号的干扰，同时兼顾不能对声音信号有所破坏等前提条件下我们提出利用多窗谱估计谱减法（去噪声）、倒谱法（去混响），分成以下几个步骤来分析、处理拾取的声音信号。

（1）多窗谱估计谱减法

1）对采集的声音信号先分帧、加窗，然后再进行 FFT 变换，分别求其第i帧的幅度谱$|X_i(k)|$和相位谱$\delta_i(k)$在相邻帧之间做平滑处理，以i帧为中心前后各取M帧共有$2M+1$帧，计算平均幅度谱$|\bar{X}_i(k)|$：

$$|\bar{X}_i(k)| = \frac{1}{2M+1}\sum_{j=-M}^{M}|X_{i+j}(k)| \tag{5.33}$$

2）把分帧后的信号$x_i(m)$进行多窗谱估计，得到多窗谱功率谱密度$P(k,i)$，并计算其平滑功率谱密度$P_y(k,i)$为：

$$P(k,i) = PMTM\left[x_i(m)\right] \tag{5.34}$$

$$P_y(k,i) = \frac{1}{2M+1}\sum_{j=-M}^{M}P(k,i+j) \tag{5.35}$$

3）根据前导无话段（噪声）占有的 NIS 帧，计算出噪声的平均功率谱密度值$P_n(k)$为

$$P_n(k) = \frac{1}{NIS}\sum_{i=1}^{NIS}P_y(k,i) \tag{5.36}$$

4）计算增益因子$g(k,i)$为：

$$g(k,i)=\begin{cases}\dfrac{P_y(k,i)-\alpha P_n(k)}{P_y(k,i)} & P_y(k,i)-\alpha P_n(k)\geq 0\\ \dfrac{\beta P_n(k)}{P_y(k,i)} & P_y(k,i)-\alpha P_n(k)\leq 0\end{cases} \tag{5.37}$$

式中，α 为过减因子，β 为增益补偿因子。

5）利用增益因子 $g(k,i)$ 和平均幅度谱 $\left|\bar{X}_i(k)\right|$ 可求得谱减后的幅度谱 $\left|\hat{X}_i(k)\right|$ 为：

$$\left|\hat{X}_i(k)\right|=g(k,i)\times\left|\bar{X}_i(k)\right| \tag{5.38}$$

再结合 1）中所求得的相位谱 $\delta_i(k)$ 进行 IFFT，就得到减噪后的声音信号元 $\hat{x}_i(m)$ 为:

$$\hat{x}_i(m)=IDFT\left[\left|\hat{X}_i(k)\right|\right]\exp\left[j\delta_i(k)\right] \tag{5.39}$$

（2）倒谱法

对去噪时域信号 $x_w(k)$ 由倒谱法进行消除混响处理，得到纯声音时域信号 $\hat{x}(k)$，其处理过程为：

1）对去噪时域信号进行分帧处理，计算每帧的复倒谱信号 $\hat{x}_c(k)$ 为：

$$\hat{x}_c(k)=IFFT\left\{\ln\left[FFT\left[\hat{x}_i(m)\right]\right]\right\} \tag{5.40}$$

式中，FFT 表示傅里叶变换；IFFT 表示傅里叶反变换；ln 表示自然对数。

2）对复倒谱信号 $\hat{x}_c(k)$，利用复倒谱域低通滤波器（由通带、过渡带、阻带组成进行滤波，其中通带的最高截止点 $P=4$，过渡带的带宽 h 为 64;，得到滤波信号。

3）对滤波信号，在时域条件下将每帧信号叠接相加，得到纯声音时域信号。

最后，利用相位变换加权广义互相关方法（GCC-PHAT）计算第 i 个麦克风接收的经声音信号处理所得纯声音时域信号 $\hat{x}_i(k)$ 与坐标原点处第 0 个麦克风接收的经声音信号处理所得纯声音时域信号 $\hat{x}_0(k)$ 间的时延值 τ_{i0}，简记为 τ_i（$1\leq i\leq 6$），则互相关函数表示为：

$$R_{\hat{x}_0\hat{x}_i}^{(g)}(\tau_i)=\int_{-\infty}^{+\infty}\psi_{PHAT}(f)\hat{X}_0(f)\hat{X}_i^{*}(f)e^{j2\pi f\tau_1}df \tag{5.41}$$

时延值表示为

$$\hat{\tau}_i=\arg\max R_{\hat{x}_0\hat{x}_i}^{(g)}(\tau_i) \tag{5.42}$$

其中 $\arg\max$ 表示使 $R_{\hat{x}_0\hat{x}_i}^{(g)}(\tau_i)$ 取得最大值时 $\tau_i=\hat{\tau}_i$，即 $\hat{\tau}_i$ 为第 0 个麦克风与第 i 个麦克风所接收到信号的时延估计值。

利用求得的时延估计值 $\hat{\tau}_i$，结合三位空间定位算法中式（5.29）、式（5.31）、式（5.32），即可确定声源位置的直角坐标。

5.1.2.3 仿真实验对比

1、四元与七元麦克风阵列定位性能比较

仿真环境为设为 6m×5m×3.5m 的房间模型。

为使仿真实验更加接近真实环境，所建立的房间模型中，声源为一个走动的人发出清晰的脚步声的声音信号，持续时间设为 20s，信号采样频率为 16kHz，64kHz，声源位置 S 为（325,215,155）。根据室温为 14.5℃，取声速 c 为 340.053m/s。

阵元间距为 0.1m 时，七元麦克风的位置依次为：$M_0(0.05,0,0.05)$，$M_1(0.1,0.05,0.05)$，$M_2(0.05,0.1,0.05)$，$M_3(0,0.05,0.05)$，$M_4(0.05,0,0.05)$，$M_5(0.05,0.05,0)$，$M_6(0.05,0.05,0.1)$；四元麦克风的位置依次为：$M_1(0.1,0.05,0.05)$，

$M_2(0.05,0.1,0.05)$， $M_3(0,0.05,0.05)$， $M_4(0.05,0,0.05)$。

阵元间距为 0.2m 时，七元麦克风的位置依次为：$M_0(0.1,0.1,0.1)$，$M_1(0.2,0.1,0.1)$，$M_2(0.1,0.2,0.1)$，$M_3(0,0.1,0.1)$，$M_4(0.1,0,0.1)$，$M_5(0.1,0.1,0)$，$M_6(0.1,0.1,0.2)$；四元麦克风的位置依次为：$M_1(0.2,0.1,0.1)$，$M_2(0.1,0.2,0.1)$，$M_3(0,0.1,0.1)$，$M_4(0.1,0,0.1)$。

图 5.10 为七元麦克风阵列

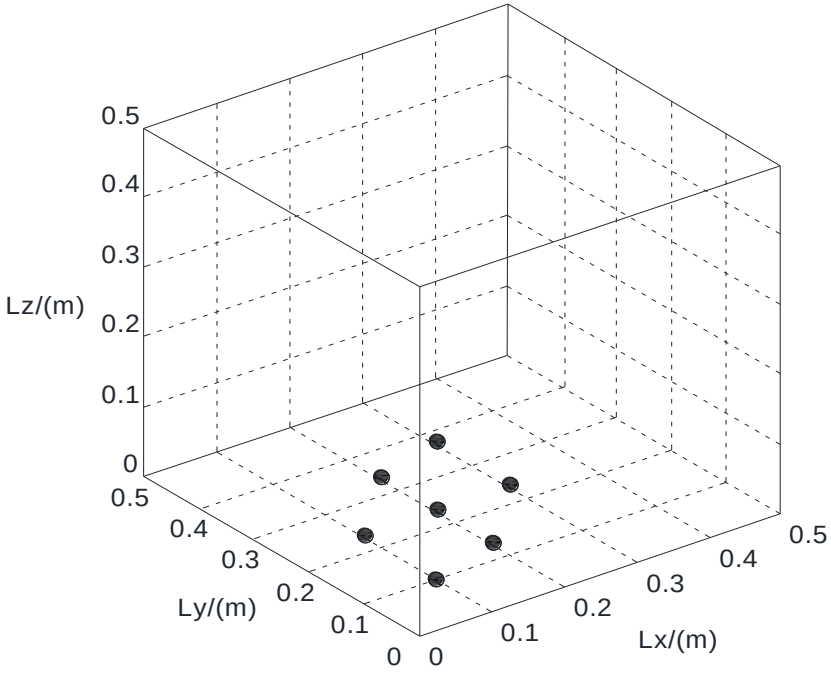


图 5.10 七元麦克风阵列

Fig.5.10 The seven microphone array

图 5.11 为M_0麦克风采集到的声音信号，图 5.12 为经过谱减法、倒谱法处理后的声音信号。

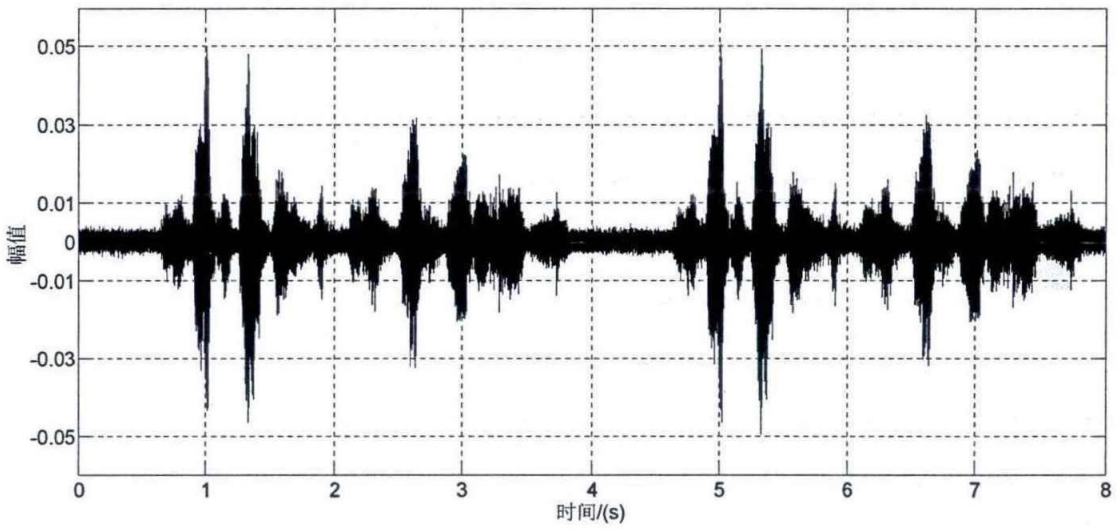


图 5.11 M_0麦克风采集到的声音信号

Fig.5.11 The speech signal acquired M_0 microphone

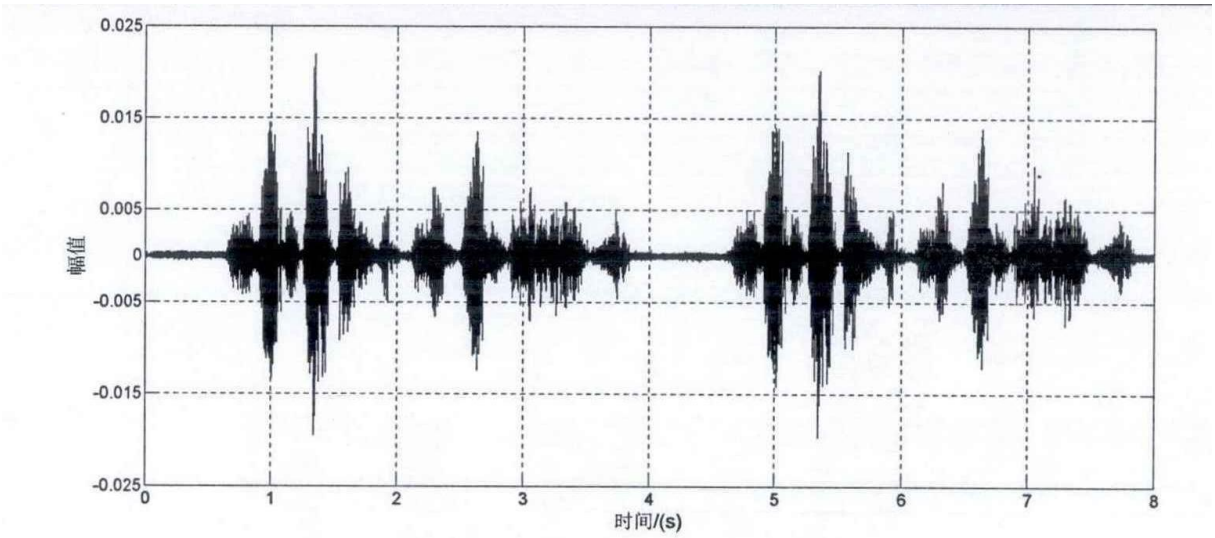


图 5.12 处理后的声音信号

Fig.5.12 The processed speech signal

由图 5.11、图 5.12 对比可以看出，在进行定位前要先进行声音信号处理步骤，对采集到的声音信号先去除噪声、混响的干扰。对于 $M_1 \sim M_6$ 麦克风采集到的信号也需要进行此类处理，然后在用经过处理后的声音信号，依据算法求出时延，进而求出声源位置。

表 5.2 给出了阵元间距为 0.1m、信号采样率为 16kHz 时的声源定位结果比较；表 5.3 给出了阵元间距为 0.2m、信号采样率为 16kHz 时的声源定位结果比较；表 5.4 给出了阵元间距为 0.1m、信号采样率为 64kH 时的声源定位结果比较。

表 5.2 阵元间距为 0.1m,信号采样率为 16kHz 下的定位

Tab.5.2 The location results for 2L=0.1m and fs=16kHz

	X/m	Y/m	Z/m	绝对误差/m		
原始声源位置	3.25	2.15	1.55	ΔX	ΔY	ΔZ
七麦克风声源位置	3.3604	1.9875	1.7030	0.1104	0.1625	0.1530
四麦克风声源位置	2.2606	2.2605	1.8655	0.3855	0.1105	0.3155

表 5.2 表明，在阵元间距为 0.1m、信号采样率为 16kHz 条件下，用七元麦克风阵列所得声源位置 X、Y、Z 的绝对误差分别比四元麦克风阵列减少 0.2751m、0.0520m、0.1625m。

表 5.3 阵元间距为 0.2m，信号采样率为 16kHz 下的定位

Tab.5.3 The location results for 2L=0.2m and fs=16kHz

	X/m	Y/m	Z/m	绝对误差/m		
原始声源位置	3.25	2.15	1.55	ΔX	ΔY	ΔZ
七麦克风声源位置	3.1551	2.1224	1.6174	0.0949	0.0267	0.0674
四麦克风声源位置	3.1365	2.2825	1.8148	0.1135	0.1325	0.2918

表 5.3 表明，在阵元间距为 0.2m、信号采样率为 16kHz 条件下，用七元麦克风阵列

所得声源位置 X、Y、Z 的绝对误差分别比四元麦克风阵列减少 0.0186m、0.1058m、0.2244m。

表 5.4 阵元间距为 0.1m，信号采样率为 64kHz 下的定位
Tab.5.4 The location results for 2L=0.1m and fs=64kHz

	X/m	Y/m	Z/m	绝对误差/m		
原始声源位置	3.25	2.15	1.55	ΔX	ΔY	ΔZ
七麦克风声源位置	3.2466	2.1179	1.5396	0.0034	0.0321	0.0104
四麦克风声源位置	3.4538	1.9854	1.2352	0.2038	0.1646	0.3148

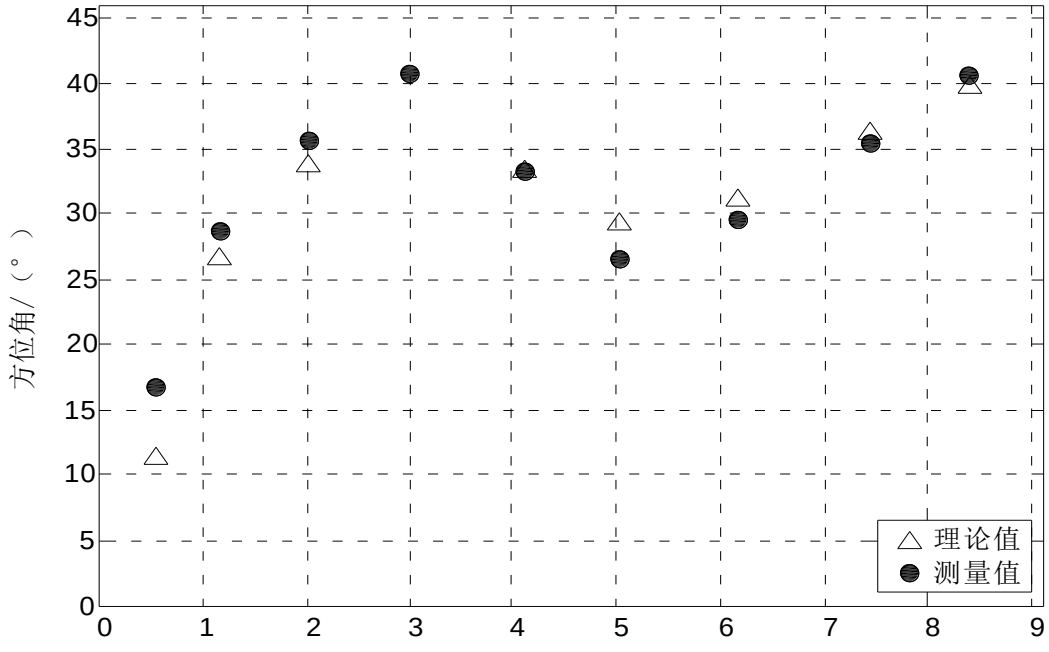
表 5.4 表明，在阵元间距为 0.1m、信号采样率为 64kHz 条件下，用七元麦克风阵列所得声源位置 X、Y、Z 的绝对误差分别比四元麦克风阵列减少 0.2004m、0.1325m、0.3044m。

因此，信号采样率为 16kHz 不变的条件下，阵元间距增加时，或阵元间距为 0.lm 不变的条件下，信号采样率增加时，无论是七元麦克风阵列还是四元麦克风阵列，声源定位的绝对误差都得到减小，定位精度提高。

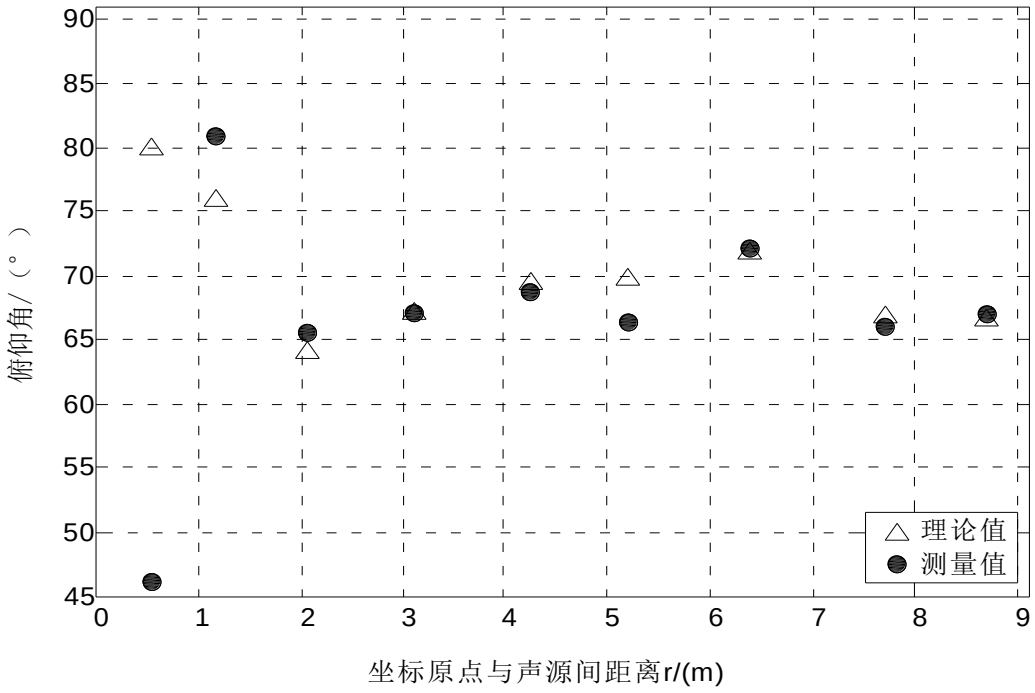
2、声源处于不同位置时七元麦克风在阵列定位结果

为了验证七元阵列对于声源处于不同位置时定位结果的精确度，改变坐标原点与声源间距离 r，运用阵列进行定位计算。

图 5.13 给出麦克风间距为 0.1m，采样率为 16kHz 条件下，图 5.13（a）为麦克风间距与方位角关系图，图 5.13（b）为麦克风的间距与俯仰角的关系。



（a）方位角

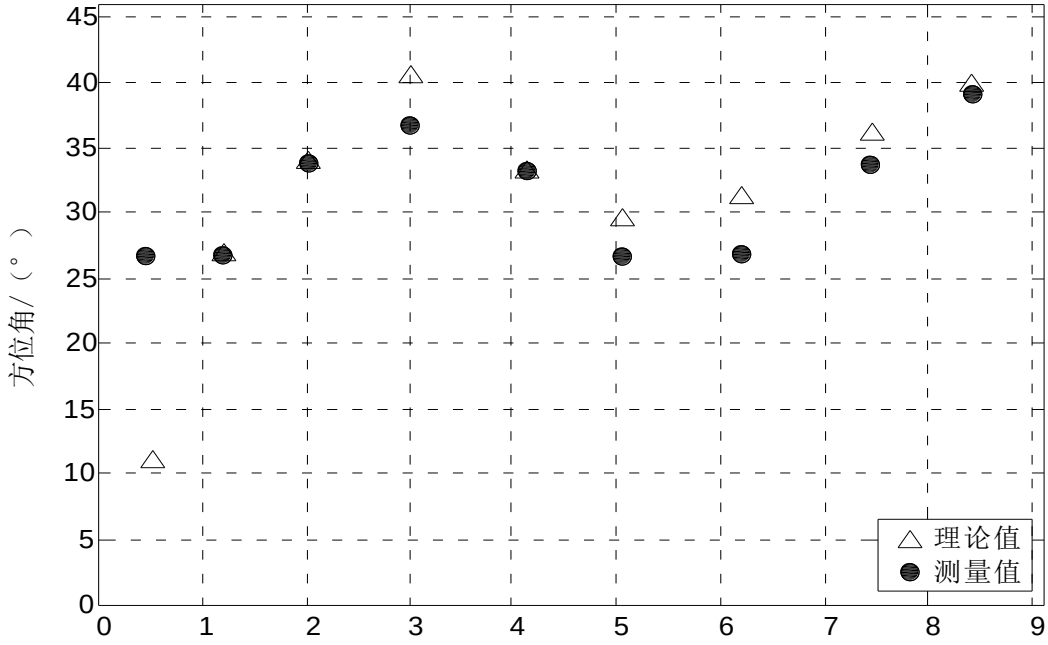


（b）俯仰角

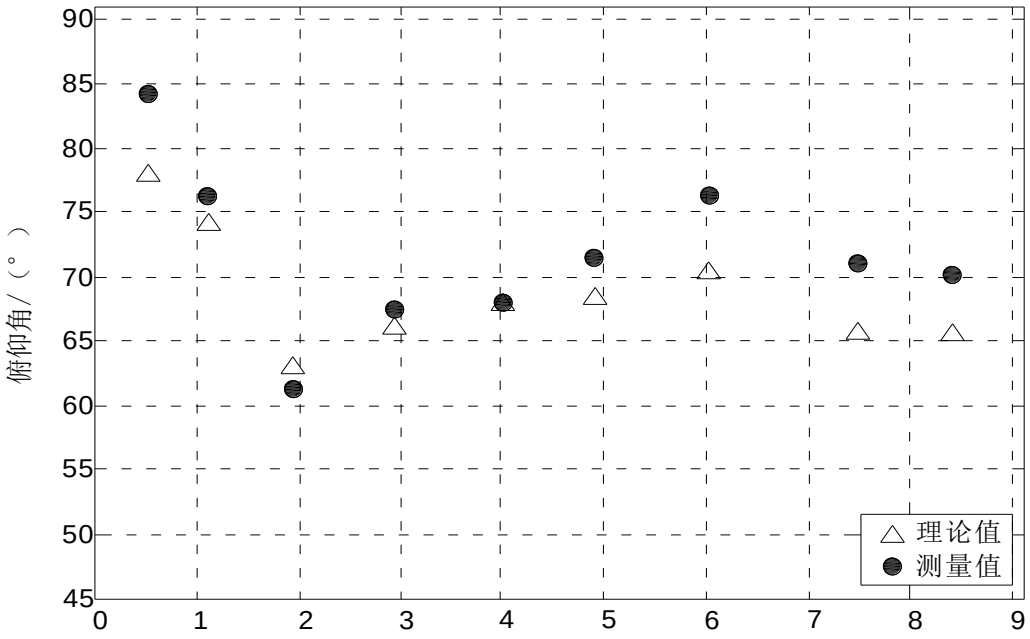
图 5.13 麦克风间距为 0.1m，采样率为 16kHz

Fig.5.13 The azimuth angle and the pitch angle for 2L=0.1m and fs=16kHz

图 5.14 给出麦克风间距为 0.2m，采样率为 16kH 条件下，图 5.14（a）为麦克风间距与方位角关系图，图 5.14（b）为麦克风的间距与俯仰角的关系。理论值为空心三角，仿真测量值为实心圆点。



（a）方位角



（b）俯仰角

图 5.14 麦克风间距为 0.2m，采样率为 16kHz

Fig.5.14 The azimuth angle and the pitch angle for 2L=0.2m and fs=16kHz

由图 5.13、图 5.14 可得，随着坐标原点与声源间距离增加，方位角、俯仰角的测量值与真实值间的误差减小。但由于声源随着距离的增大会产生衰减所以声源的位置不能处于无限远处。

改变坐标原点与声源间距离 r，运用七元麦克风阵列进行定位计算。表 5.5 给出阵元间距为 0.1m，信号采样率为 16kH 条件下的定位结果，表 5.6 给出阵元间距为 0.2m，信号采样率为 16kHz 下的定位结果。

表 5.5 阵元间距为 0.1m，信号采样率为 16kHz 下的定位

Tab.5.5 The location results for 2L=0.1m and f=16kHz

真实值/（m）				测量值/（m）			绝对误差值/（m）			平均绝对误差值/（m）	
X	Y	Z	r	X	Y	Z	ΔX	ΔY	ΔZ		
0.55	0.25	0.15	0.5	0.4883	0.2442	0.0437	0.0617	0.0058	0.1063		
1.05	0.55	0.35	1.1	1.0027	0.5014	0.2563	0.0473	0.0486	0.0937	$\Delta\bar{X}$	0.1357
1.55	1.05	0.95	2.0	1.5937	1.0625	0.6254	0.0437	0.0125	0.3246		
2.15	1.85	1.25	3.0	2.1797	1.7567	1.1197	0.0297	0.0933	0.1303		
3.25	2.15	1.55	4.1	3.3604	1.9875	1.7030	0.1104	0.1625	0.1530	$\Delta\bar{Y}$	0.1481
4.15	2.35	1.85	5.0	4.0791	2.4090	1.7199	0.0708	0.0590	0.1301		
5.05	3.05	2.05	6.1	5.1328	2.8836	1.9673	0.0828	0.1664	0.0827		
5.55	4.05	3.05	7.4	5.6224	3.8917	2.9143	0.0724	0.1583	0.1357	$\Delta\bar{Z}$	0.2083
5.95	4.95	3.45	8.3	6.6525	4.3202	2.7313	0.7025	0.6289	0.7187		

表 5.6 阵元间距为 0.2m，信号采样率为 16kHz 下的定位

Tab.5.6 The location results for 2L=0.2m and fs=64kHz

真实值/（m）				测量值/（m）			绝对误差值/（m）			平均绝对误差值/（m）	
X	Y	Z	r	X	Y	Z	ΔX	ΔY	ΔZ		
0.2	0.3	0.2	0.5	0.3789	0.1137	0.3789	0.1789	0.1863	0.1789		
1.1	0.6	0.4	1.1	0.9931	0.5435	0.2019	0.1069	0.0565	0.1981	$\Delta\bar{X}$	0.1112
1.6	1.1	1.0	2.0	1.4867	1.0619	0.8496	0.1133	0.0381	0.1504		
2.2	1.9	1.3	3.0	2.1000	1.8000	1.2000	0.1000	0.1000	0.1000		
3.3	2.2	1.6	4.1	3.1911	2.0927	1.5287	0.1089	0.1073	0.0713	$\Delta\bar{Y}$	0.1108
4.2	2.4	1.9	5.0	4.0959	2.2982	1.8118	0.1041	0.1018	0.0882		
5.1	3.1	2.1	6.1	5.0834	2.8837	1.9603	0.0166	0.2163	0.1397		
5.6	4.1	3.1	7.4	5.4845	3.9175	3.1340	0.1155	0.1825	0.0340	$\Delta\bar{Z}$	0.1246
6.0	5.0	3.5	8.3	5.8434	5.0086	3.3391	0.1566	0.0086	0.1609		

表 5.5 表明：在阵元间距为 0.1m，信号采样率为 16kHz 条件下，距离从 0.5m 到 8m，七元麦克风阵列的声源定位精度基本在 0.3m 之内。

表 5.6 表明：在阵元间距为 0.2m，信号采样率为 16kHz 条件下，距离从 0.5m 到 8.5m，七元麦克风阵列的声源定位精度基本在 0.15m 之内。

因此，在信号采样率不变的条件下，随着坐标原点与声源间距离增加，阵元间距为 0.2m 的阵列比阵元间距为 0.1m 的阵列，声源定位的平均绝对误差得到减小，定位精度提高。

5.2 问题二的解答

5.2.1 室内房间轮廓构图的基本原理

5.2.1.1 室内声场简介

研究声信号构图首先需要熟悉声音信号在室内传播的特性。声波传播的空间被称为声场，声场分为自由声场、扩散声场(混响声场)和半自由声场[11][12]。

自由声场指是在声信号传播的过程中不经过反射面反射，在声场中的任意一点只存在直达声没有反射信号的声场。

扩散声场又称混响声场，指的是封闭空间内，被激发足够多的简正方式（简正方式指的是房间内驻波的一种振动方式），空间内任意一点到达的声波包含各种入射方向的声场。在这种声场中，室内声压级几乎相等，声能密度近乎相等。

半自由声场指的是在宽阔的广场上，或者在特殊房间内有一个面是反射面，其余面都是吸声面的声场。

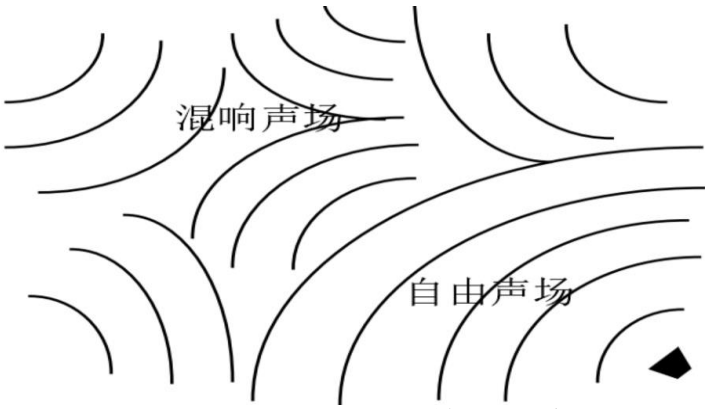


图 5.15 自由声场与扩散声场示意图

Fig.5.15 Schematic diagram of free sound field and diffused sound field

常见的室内环境大部分属于扩散声场，只有发出声音的一段时间内属于自由声场，如上图 5.15 所示。在室内，声信号在封闭空间中的传播远比自由空间复杂。这时，声信号将受到封闭空间各个界面，如天花板、地面、墙壁等反射、吸收与透射[12]。此时室内声场属于扩散声场，扩散声场作为封闭室内声场存在以下四个特点：

(1)声波在各个界面发生反射，吸收与透射；

(2)声波在室内声场拥有与自由声场不同的音质；

(3)由于房间的共振可能引起某些频率的声音被加强或减弱；

(4)声音在空间的分布发生了变化。

5.2.1.2 室内声场的 Image 模型

基于几何声学模型的方法主要有两种，一种是射线跟踪法，另一种是 Image 法。射线跟踪法由挪威的 A.Krokstand 提出，他利用携带能量的射线来描述声源能量的辐射，由于每条射线的能量在传播过程中和遇到反射面都会能量衰减，需要在接收端记录声线的每条路径以得到房间的冲激响应。声线追踪法适合处理复杂声场，对于具有规则几何图形的房间来说，Image 镜像源法更适合，这种方法理论简单，易于理解而且利于实现，被认为是研究声学的基础理论，在声场设计，室内声学，工学方面都得到了大范围利用。

Image 模型是由 Allen J 和 Berkeley D 在 1979 年经典文献[13]提出的一种描述小房间混响的声学模型。这种模型假设声源发出的信号一部分沿直线传播直接被麦克风所接收，另一部分经过墙面或者障碍物反射被麦克风接收，反射符合镜面反射规律，入射角等于反射角。该方法利用反射阶数（声波经过墙壁、地板、天花板反射的次数）和房间维数表示房间混响，适用于具有规则几何结构的房间。声音在房间内传播模型类似于一个多径信道，描述房间信道的冲激响应模型可以用以下公式表示：

$$h(t)=\sum_i a_i\delta(t-\tau_i) \qquad (5.43)$$

在式（5.43）中，$h(t)$ 代表房间时域冲激响应，a_i 代表每一个信道的增益，τ_i 代表每一个信道的时延。Image 模型如图 5.16 所示：

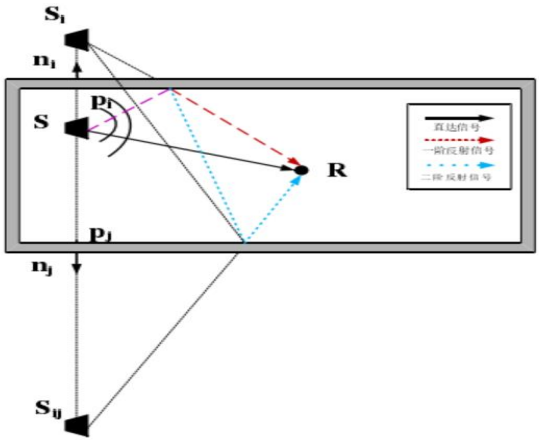


图 5.16 Image 模型下的一阶镜像源和二阶镜像源

Fig.5.16 First-order and second-order mirror sources in the Image model

根据 Image 模型：一阶镜像声源坐标和声源坐标关于每一面反射面（墙、天花板、地板）轴对称，而镜像源和麦克风的连线与反射面的交点为反射点。声源坐标定义为 s，镜像声源坐标定义为 s_i，声源一阶镜像声源点连线与墙线交点坐标定义为 p_i，黑斜体代表向量，n_i 定义为与实声源一阶镜像声源连线共线的指向镜像声源的单位向量，$p_i - s$ 指声源指向 p_i 点的向量，$\langle p_i - s, n_i \rangle$ 指两个向量的内积也就是声源点 s 到 p_i 的距离，一阶镜像声源 s_i 可以表示为[14]：

$$s_i = s + 2\langle p_i - s, n_i \rangle n_i \tag{5.44}$$

一阶镜像声源点二阶镜像声源点连线与墙线交点坐标定义为 p_j，n_j 定义为与一阶镜像声源二阶镜像声源连线共线的指向二阶镜像声源的单位向量，$p_j - s_i$ 指一阶镜像声源指向 p_j 的向量，$\langle p_j - s_i, n_j \rangle$ 指两个向量的内积也就是一阶镜像声源点 s_i 到 p_j 的距离，二阶镜像声源和一阶镜像声源关于另一面墙轴对称用 s_{ij} 表示，s_{ij} 可以表示为[14]：

$$s_{ij} = s_i + 2\langle p_j - s_i, n_j \rangle n_j \tag{5.45}$$

当二阶以上等多阶反射发生时，根据镜像源的原理，声源点将在房间各个平面产生镜像源，镜像源还会在产生新的镜像源，由此可以推测镜像源数目将会无限增大。假设当一个房间存在 N 个反射面，当考虑 M 个阶数时，镜像数目可以达到 $N(N-1)^{M-1}$ 个。随着镜像源数目的增加，总镜像源数目按照一下规律指数增加[15]。

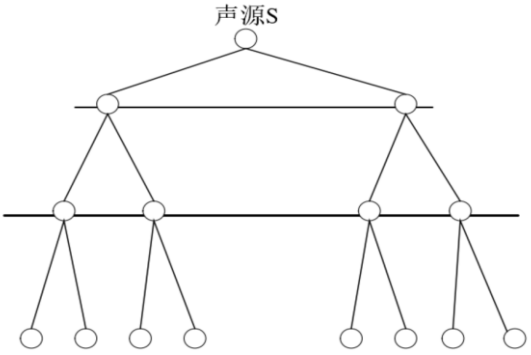


图 5.17 镜像源数目分析图

Fig.5.17 Analysis diagram of the number of mirror sources

虽然镜像源数目很多，但不是每个镜像源都是有效的。按照 $N(N-1)^{M-1}$ 的推测方法只考虑到了反射面和阶数的关系，并没有考虑到有些反射面由于角度问题造成实际有一

些镜像源是不存在的，如下图所示：

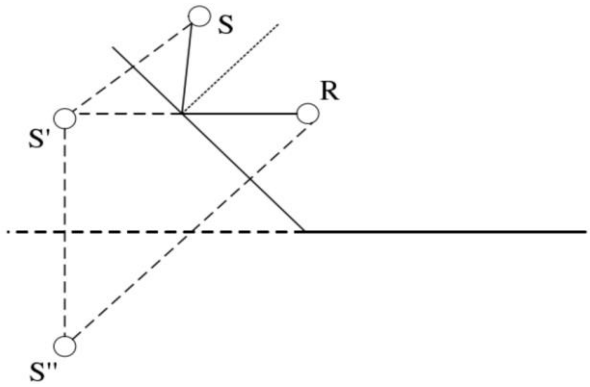


图 5.18 镜像源可见性判断依据

Fig.5.18 Visibility criteria of mirror source

上图声源点 s 关于左侧倾斜壁面的镜像声源 s' 是存在的，而一阶镜像源 s' 关于平行平面的二阶镜像声源 s'' 是不存在的。

5.2.1.3 室内脉冲响应简介

利用 Image 法模拟的室内声场可以用室内脉冲响应表示，因此测量室内脉冲响应是模拟声场的关键。声场脉冲指的是声场中某一位置接收到声源发出的一系列信号序列，即为声场系统在单位冲激函数激励下引起的零状态响应被称之为该系统的“脉冲响应”。它与系统的传递函数互为傅里叶变换关系。在同一声场环境下，声源收到的脉冲响应是唯一的，它代表在这一点声波的传播情况，声源距离此测量点的距离，此测量点收到声波的声压值，所有信息都被包含在脉冲响应中。当声源是单位脉冲声源时，接收到的信号就是该测试点的房间脉冲响应。但是，由于单位脉冲声源实际很难产生一般情况下求解系统脉冲响应函数的方法来求解[16]。

求解脉冲响应的意义在于：当声源点和接收位置被固定，接受位置接收到的脉冲响应是唯一的，室内声场的瞬时变化以及衰减过程均能够通过脉冲响应准确表示。利用室内脉冲响应经过数字信号处理可以获得早期反射信号、混响时间、清晰度、声场强度等统计声学参量，这些参量不仅对于房间结构的重构而且对于如何更合理地设计或布置房间以改善室内听音质量都有很重要的意义[17]。

根据室内声场内的声能和声压等参数可以叠加且满足齐次的特性，构成室内声场的声源、空间、接收端可以共同构成一个线性系统。假设此系统输入为 $s(t)$，输出为 $y(t)$，$h(t)$ 就代表声场的冲激响应。

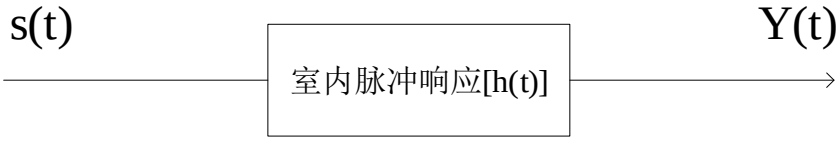


图 5.19 线性系统示意图

Fig.5.19 Schematic diagram of linear system

$s(t)$ 代表系统输入信号，$y(t)$ 代表系统输出信号，$h(t)$ 代表系统的冲激响应。根据信号系统的理论，输入信号 $s(t)$，输出信号 $y(t)$ 满足以下公式：

$$y(t)=\int_{-\infty}^{+\infty}s(t-\tau)h(\tau)d\tau \qquad (5.46)$$

根据频域传递函数与冲激响应函数互为傅里叶变换关系可知，在频域上：

$$Y(\omega)=S(\omega)H(\omega) \qquad (5.47)$$

其中 $Y(\omega)$、$S(\omega)$、$H(\omega)$ 分别是 $y(t)$、$s(t)$、$h(t)$ 经过傅里叶变换在频域的形式。

为了求解时域房间冲激响应，变换上（5.46）、（5.47）两式，可以得到：

$$Y(\omega) = F\{y(t)\} \quad (5.48)$$

$$S(\omega) = F\{s(t)\} \quad (5.49)$$

$$h(t) = F^{-1}\left\{\frac{Y(\omega)}{S(\omega)}\right\} \quad (5.50)$$

F 代表傅里叶变换，F^{-1} 代表傅里叶逆变换，利用接收到的信号按照式（5.48）经过傅里叶变换转换为频域，发射信号按照式（5.49）经过傅里叶变换转换为频域，利用接收信号的频域形式与发射信号的频域形式做商，商的结果按照式（5.50）经过傅里叶逆变换即可以得到室内冲激响应的时域形式。

在室内声学中，室内冲激响应又称为室内或房间脉冲响应。从信号和系统的角度分析，室内冲激响应合理地解释了声场中声传播与接收信号的关系，室内声场冲激响应的准确测量可以实现声场的重现和对室内音响效果的分析[18]。

5.2.1.4 基于声学概念对于室内脉冲响应的分类

按照统计声学的概念，室内冲激响应图包含直达声、早期反射声、混响声[19]。

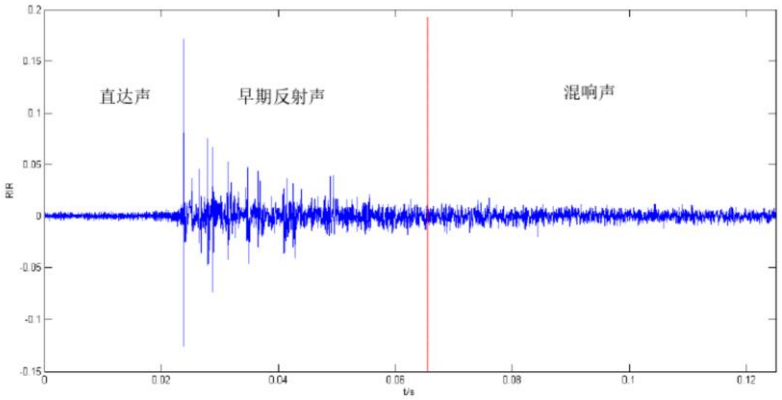


图 5.20 实测室内脉冲响应图

Fig.5.20 Actual indoor impulse response diagram

直达声指的是：从发射端发射不经过任何反射面反射直接到达接收端的信号。由于直达声不经过任何反射，从声线角度分析直达声可以类比为一条射线。它经过的路程最短，因此传播时间也最短，在冲激响应图里指的是第一个接收到的信号。另外，由于直达声没有经过反射面吸收，能量衰减最小，一般来说也是早期收到的最强信号。从通信角度来看，直达声相当于电磁波传播过程中的视距信号。

早期反射指的是：发射端发出的信号经过一次反射到达接收端的信号，如果信号是语音信号早期反射指的是接收到直达声后 50ms 内接收到的信号；如果信号是音乐信号则代表接收到直达声后 80ms 内接收到的信号。一般情况下，室内音响设计者认为这一部分信号对于加强直达声响度提高清晰度的作用，从室内声学和室内声信号处理的观点来看，这一部分信号清晰度和明晰度都非常好，最重要的一点是早期反射不影响对于声源方向与位置的判断，也就是说，这一部分信号用来实现室内声源定位和室内空间结构重建是非常理想的信号。

混响指的是：声音在室内传播的过程中，接收端收到的经过多次反射得到的反射声，这些反射声造成了混响现象，当所有反射声互相叠加就得到了混响信号。引发混响的主要原因在于声波在封闭空间内发生的反射相比与发射信号产生了相位的偏移，其波形在未改变的情况下，声波产生了叠加。混响产生的经过如下图所示：

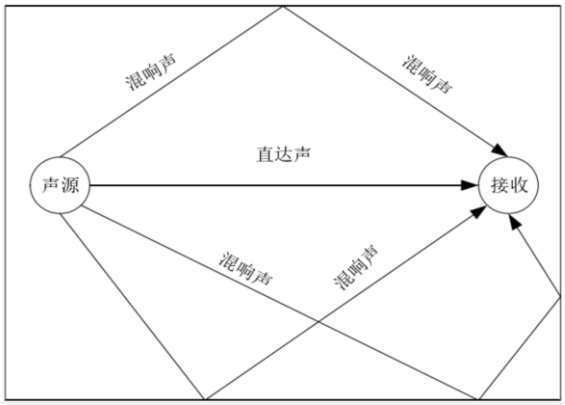


图 5.21 室内混响产生的过程

Fig.5.21 Indoor reverberation generation process

从人类听觉出发，人类感受到混响在时间上的延迟就是混响时间。如今混响时间的主要概念是由美国的塞宾提出的，混响时间作为只与房间结构大小特性相关的量在房间的任意位置测量结果都是相同的。混响时间作为描述室内声学最重要的特性，可以准确反映能量随时间衰减的特性，这对于室内声音品质的评价以及在研究声信号在室内的传播情况有着极其重要的价值以及意义。塞宾不仅给出混响时间的定义，即声能衰减 60dB 所用的时间，而且给出了计算混响时间的公式：

$$T_{60} = 0.163 \frac{V}{\sum S_k \alpha_k} \tag{5.51}$$

其中，V 代表房间体积，S_k 代表室内反射面的吸声系数，α_k 代表室内反射面的面积。对于知道房间体积面积的情况，可以利用这种方法计算估计房间的混响时间，如果在不知道房间情况的条件下需要利用冲激响应求解或者利用混响计进行计算。

5.2.1.5 室内房间轮廓构图的基本思路和难点

（1）室内房间轮廓构图的基本思路

实际上，实声源和一阶镜像声源连线的中垂面就代表了房间的墙面、天花板、地板所在的位置，得到了实声源的坐标和所有一阶镜像声源的坐标相当于得到了房间的几何构型，求解实声源和一阶镜像声源的位置就是实现室内构图的途径。

（2）室内房间轮廓构图的难点

可以注意到，按照 Image 模型如图所示，在一个凸多面体的房间内，实声源和一阶镜像声源连线的中垂面就代表了房间的墙面、天花板、地板等反射面所在的位置。得到了实声源的坐标和所有一阶镜像声源的坐标相当于得到了房间的几何构型，求解实声源和一阶镜像声源的位置就是实现室内构图的途径。

为了求解实声源和一阶镜像声源坐标，根据声信号传播公式：

$$t_{io} = \frac{\|s_i - r_o\|}{c}, i = 1 \sim m, o = 1 \sim n \tag{5.52}$$

$$\|s_i - r_o\| = \sqrt{(x_{si} - x_{ro})^2 + (y_{si} - y_{ro})^2 + (z_{si} - z_{ro})^2} \tag{5.53}$$

在式（5.52）中，c 代表声信号在空气中的传播速度，r_o 代表第 $1 \sim n$ 个接收端的位置，s_i 代表待求的第 $1 \sim m$ 个实声源或者镜像声源，t_{io} 代表此声源发出的信号传播到第 $1 \sim n$ 个接收端的传播时间。式中声速 c 已知，t_{io} 可以通过各个接收端测得的房间冲激来得到，$\|s_i - r_o\|$ 代表发射端到接收端的欧式距离，具体求解参照公式（5.53）。接收端 r_o 的位置属于已知信息，取决于麦克风阵列的布阵方式。但要求解声源 s_i 还需要解决以下问题。

难点一是声音反射信号到达次序不一的造成的分类困难问题。TOA（Time of Arrival）代表信号从发射端到达接收端的时间值。对于直达信号来说，信号到达的先后顺序代表 TOA 的大小反映出声源到麦克风位置的远近，由此可以很容易判断每一个信号来自哪一个发射端。例如下图右侧的两个冲激响应图，上图传音器 1 接收到的直达信号明显晚于下图传音器 2 接收到的直达信号，因此可以判断传音器 2 到声源的距离小于传音器 1 到声源的距离。

但是对于一阶反射信号来说，由于每一个接收端收到的信号的先后顺序取决于接收端与反射面的相对位置，也就是说，由于接收端离各个反射面距离不同造成了关于各个反射面的镜像声源位置也不同。从接收端来看，接收端收到的信号可以来自任何一个反射面。如下图 5.22[20]所示：从右侧上方麦克风 1 的冲激响应图来看，先收到的是来自于蓝色 A 墙面的一阶反射信号后收到来自红色 B 墙面的一阶反射信号；而从右侧下方麦克风 2 的冲激响应来看，先收到的是来自于红色 B 墙面的一阶反射信号后接受到来自蓝色 A 墙面的反射信号。这种反射信号到达顺序颠倒的问题直接决定了利用多传感测量房间冲激响应时不能认为反射信号是同一次序到达的，这样就无法判断信号的来源，这也给定位造成了极大的困难。

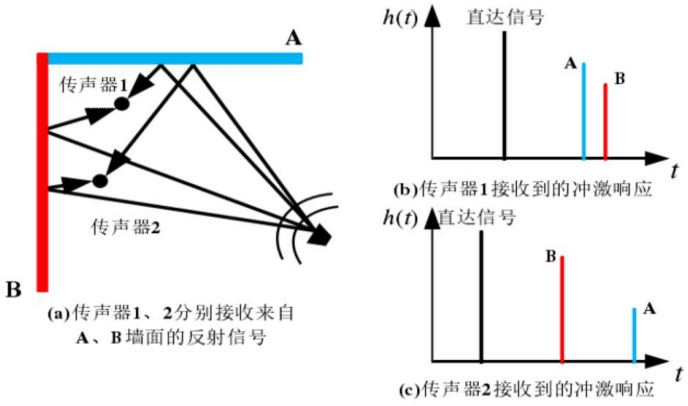


图 5.22 反射信号到达次序不一致

Fig.5.22 Inconsistent arrival order of reflected signals

也就是说先收到的反射信号不一定来自同一反射面，反射信号受到声源麦克风反射面位置关系的共同影响，有时会出现反射次序颠倒的问题，同一个声源发出的信号麦克风 1 先收到蓝色面反射的，而麦克风 2 先收到红色面反射的。因此不能认为反射信号都是以同一次序到达的。无法判断反射信号的来源，定位声源坐标也就无从谈起。

难点二是高阶混响干扰与一阶反射信号的辨别问题。早期反射声也就是只经过一次反射的信号由于有着非常好的清晰度和明晰度，其信号对于方位识别度高的缘故可以视为理想信号。反射信号到达顺序和发射端、接收端、反射面三者位置相关，只根据冲激响应图无法判断一阶反射信号和二阶反射信号的界限，因为有些二阶反射信号甚至会先于一阶反射信号到达。

为了选择合适的信号来重建房间结构，这里必须设计一种能够识别区分一阶、二阶信号的方法，如下图所示：横轴代表时间轴，纵轴代表时间域房间冲激响应 RIR（room impulse response），其中黑色标记为实测直达信号，红色标记为实测一阶反射信号，黄色标记为二阶极其以上信号。只有在冲激响应中区分了直达信号、早期反射声、混响声三种不同时期的声音才能去除高阶反射信号实现室内空间结构的重建。

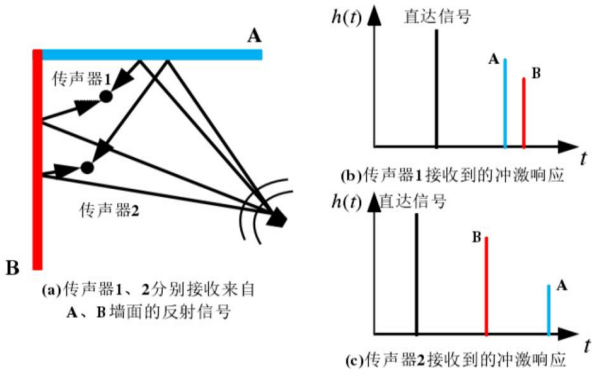


图 5.23 高阶反射信号干扰

Fig.5.23 Interference of high order reflected signal

5.2.2 关于声音反射信号分类的对比研究

5.2.2.1 基于欧氏距离法对于反射信号的分类

由于反射信号次序不一致无法判断反射信号来源的问题，最早研究利用声信号实现房间重构的 Dokmanić I 等人在经典文献[20]中提出了利用欧式距离阵 EDM（Euclidean Distance Matrices）特性为反射信号归类的方法。

欧式距离阵指的是：在欧几里得空间中有 N 个点，这 N 个点两两之间求欧式距离，以求得的欧式距离为元素组成一个 $N\times N$ 的对称矩阵，矩阵中包含 $N\times(N-1)/2$ 个距离信息[21]。用公式表示为：欧式空间 X 中的 N 个点每个点的坐标是 d 维的，则

$$X=\left[x_1,x_2,x_3,x_4\cdots x_N\right] \tag{5.54}$$

X 中任意两点之间的欧式距离的平方可以表示为：

$$d_{ij}=\left\|x_i-x_j\right\|^2 \tag{5.55}$$

展开上式得到

$$d_{ij}=(x_i-x_j)^T(x_i-x_j)=x_i^Tx_i-2x_i^Tx_j+x_j^Tx_j \tag{5.56}$$

所有欧式距离平方可以构成欧式距离阵 $D=\left[d_{ij}\right]$，$diag$ 代表对角阵，定义为：

$$EDM(X)=1diag(X^TX)^T-2X^TX+diag(X^TX)1^T \tag{5.57}$$

EDM 被认为包含完整的节点信息，而且其矩阵形式易于处理，已经在心理测量学、晶体学、机器学习等方向得到了广泛应用，利用 EDM 可以重建这 N 个点的位置。

这种方法首先建立基于接收端传音器阵列的欧式距离阵，将麦克风之间的欧式距离作为矩阵的一部分元素，另一部分元素来源于冲激响应图的反射信号到达时间，到达时间转换为声源到麦克风之间的欧式距离，这部分距离以作为扩展与原矩阵组成增广矩阵，如下图所示：

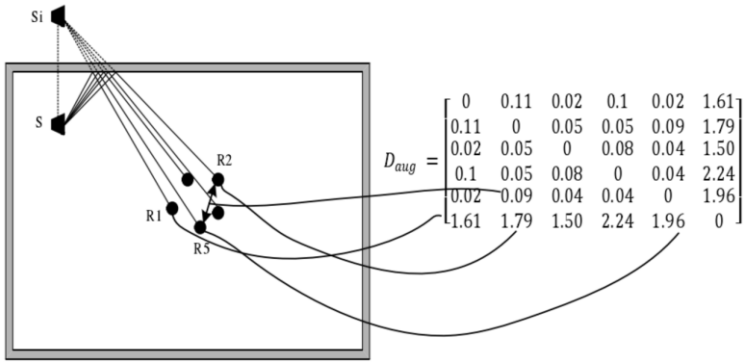


图 5.24 基于欧式距离阵对反射信号归类的方法

Fig.5.24 Classification of reflected signals based on Euclidean range array

得到矩阵后利用多维尺度分析法 MDS（Multidimensional Scaling）来反推出符合约束条件的声源位置以及麦克风位置[22]。

假设 X 为待求的声源位置与麦克风位置的坐标。从 EDM 中得到正确的 X 需要使用 MDS 算法。假设 x_1 坐标为原点，

$$d_{i1} = \|x_i - x_1\|^2 = \|x_i - 0\|^2 = \|x_i\|^2 \tag{5.58}$$

于是

$$1diag(X^T X) = 1d_1^T \tag{5.59}$$

其中 $d_1 = De_1$，1代表所有元素都为1的向量，解算 $X^T X$ 可得：

$$X^T X = -\frac{1}{2}(D - 1d_1^T - d_1 1^T) \tag{5.60}$$

由于 $X^T X$ 为方阵，利用特征值分解可以得到：

$$X^T X = U\Lambda U^T \tag{5.61}$$

其中 $\Lambda = diag(\lambda_1, \cdots, \lambda_n)$，$\Lambda$ 为以 $X^T X$ 特征值为元素的对角阵，U 是正交化特征向量。求解 X，可以得到：

$$X = \left[diag(\sqrt{\lambda_1}, \cdots, \sqrt{\lambda_n}), 0_{d\times(n-d)} \right] U^T \tag{5.62}$$

从室内冲激响应中得到的 TOA 由于发射端、反射面和接收端位置的影响经过各个反射面反射的信号到达次序各不相同，这里需要一种方法来判断这一组传播距离是否是来自与同一个镜像声源的也就是经过同一反射面的以确保增广矩阵得到的结果是正确的。Dokmanić I 等人借用了心理学领域的 Takane、Young 等人在 1977 年提出了 s-stress 概念，提出了一种验证增广矩阵是否正确的公式如下式所示[23]：

$$\sum_{(i,j)} (EDM(X)_{ij} - di_j)^2, X \in R^{d\times n} \tag{5.63}$$

其中，EDM 代表 X 构成的欧式距离阵，$\tilde{d}_{ij}$ 代表 MDS 算法求出的 X 这组坐标点每两个点之间的欧式距离。

也就是说，随机从每个接收端得到的冲激响应中选出一个 TOA，转换为传播距离，与原欧氏距离阵组成增广矩阵。对新组成的矩阵利用 MDS 算法求出符合距离条件的点集，求点集中每两个点的距离与原增广矩阵中的对应元素做差求解两个矩阵之间的欧式距离。重复上述过程，遍历所有的待选 TOA，求出所有重构后点集与原点集两两之间距离的差距，求出差距最小的一组。差距最小就代表重建成功，那么这一组 TOA 就可以视为是经过同一个反射面反射的反射信号。假设存在 N 个反射面，循环上述过程 N 次则可以将待分类的 TOA 分成 N 类，这就是基于欧氏距离阵特性方法为反射信号分类的主要思想。

Dokmanić I 等人提出的基于欧氏距离阵特性的分类方法是利用声音反射信号重建房间构型的一种经典算法，这种方法将麦克风阵列和声源的几何特性利用矩阵来表示，利用 MDS 算法实现待测镜像声源以及麦克风位置的重建，利用 s-stress 的思想检验重建结果，实现对于反射信号的分类。但是，随着后期关于利用室内冲激响应构建房间结构的研究更加深入，这种方法的不足之处也逐渐显现了出来。最重要的一点在于，分类的成功与否完全取决于声源到麦克风距离的准确性。基于信号 TOA 的分类算法最大的弊端在于当发射端和接收端之间存在延迟，这些延迟包括信号处理收发的时间、信号响应时间和介质中传输的时间，介质传输的时间受温度以及室内环境影响很大。因此，这些延迟不是精确已知的，是随着时间而变化的变量。由于在 20 摄氏度的常温条件下声速高达 $343m/s$，即使存在 $10^{-3}s$ 级别的延迟都会引入 3、$40cm$ 的误差，利用存在误差的

距离进行重建显然会严重影响对反射信号归类的判断而且还会影响求解出的声源位置的精度。因此，需要考虑利用到达时间做差消除这种时间延时带来的影响。

5.2.2.2 基于 TDOA 的最小二乘定位

现在常用的利用到达时间差进行定位的方法叫做到达时间差法 TDOA（Time Difference of Arrival），又称为双曲线定位法。

声学中的 TDOA 法用来定位声源位置，首先搭建麦克风阵列的平台，保证所有麦克风均处于同一状态，即能保证同时打开和同时关闭以及不会出现相互干扰或共振现象。分别利用各个麦克风接收信号求解冲激响应并计算信号到达时间，估计到达时间之间的时间延迟并计算出待测声源位置的坐标。TDOA 法原理简单，操作方便在声源定位系统中得到了广泛的利用。

在三维空间中，至少需要 4 个麦克风来定位声源。声源发出的声信号到达主麦克风和其他麦克风之间的时间差形成了以这两个麦克风为焦点的双曲面，四个麦克风可以形成三对双曲面，其中每两个单边双曲面相交于一条线，这两条线的交点就是声源的位置。

设声源坐标为 (x,y,z)，已知四个麦克风坐标分别为：(x_1,y_1,z_1)，(x_2,y_2,z_2)，(x_3,y_3,z_3)，(x_4,y_4,z_4)，其中 (x_1,y_1,z_1) 为主麦克风。计算声源到达各个麦克风的时间差可以得到：

$$\Delta_{r_i} = r_i - r_1 = c * \Delta_{t_{i1}} (i = 2,3,4) \tag{5.64}$$

其中，上式中 $\Delta_{t_{i1}}$ 代表麦克风 2、3、4 的到达时间与主麦克风 1 到达时间之间的时间差。r_i 代表从声源到达 2、3、4 麦克风的传播距离，r_1 代表从声源到主麦克风的传播距离，用公式表示为：

$$\Delta_{t_{i1}} = t_i - t_1 (i = 2,3,4) \tag{5.64}$$

$$r_1^2 = (x - x_1)^2 + (y - y_1)^2 + (z - z_1)^2 \tag{5.65}$$

$$r_i^2 = (x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2 \tag{5.66}$$

将（5.64）中的 r_i 移到第一个等号的左边，两边取平方，再带入式（5.65）、（5.66），可得

$$(x_1 - x_i)x + (y_1 - y_i)y + (z_1 - z_i)z = k_i + r_1^* \Delta r_i \tag{5.67}$$

其中，

$$k_i = 1/2(\Delta r_i^2 + x_1^{\,2} + y_1^{\,2} + z_1^{\,2} - x_i^{\,2} - y_i^{\,2} - z_i^{\,2}) \tag{5.68}$$

将式（5.67）改写为矩阵形式：

$$AX = B \tag{5.69}$$

$$A = \begin{bmatrix} X_1 - X_2 & Y_1 - Y_2 & Z_1 - Z_2 \\ X_1 - X_3 & Y_1 - Y_3 & Z_1 - Z_3 \\ X_1 - X_4 & Y_1 - Y_4 & Z_1 - Z_4 \end{bmatrix}; X = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}; B = \begin{bmatrix} k_1 + r_1 \Delta r_2 \\ k_2 + r_1 \Delta r_3 \\ k_3 + r_1 \Delta r_4 \end{bmatrix} \tag{5.70}$$

当 2、3、4 传感器与主麦克风不在同一条直线上，方阵 A 可逆，可以得到：

$$X = A^{-1}B \tag{5.71}$$

$$A^{-1} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \tag{5.72}$$

那么 X,Y,Z 参数解分别为：

$$\begin{aligned} X &= m_1 r_1 + n_1 \\ Y &= m_2 r_1 + n_2 \\ Z &= m_3 r_1 + n_3 \end{aligned} \tag{5.73}$$

其中，

$$m_i = \sum_{j=1}^{3} a_{ij}\Delta r_w ; n_i = \sum_{j=1}^{3} a_{ij}\Delta k_w \tag{5.74}$$

将式（5.73）带入式（5.75）、（5.76）中可得：

$$a r_1^2 + 2 b r_1 + c = 0 \tag{5.75}$$

$$a = m_1^2 + m_2^2 + m_3^2 - 1 \tag{5.76}$$

$$b = m_1 (n_1 - x_1) + m_2 (n_2 - y_1) + m_3 (n_3 - z_1) \tag{5.77}$$

$$c = (n_1 - x_1)^2 + (n_2 - y_1)^2 + (n_3 - z_1)^2 \tag{5.78}$$

根据求根公式可得：

$$X = \frac{-b \pm \sqrt{b^2 - ac}}{a} \tag{5.79}$$

将求得的 r_1 带入式（5.73）中即可得到声源位置坐标的估计值，实现对声源的定位。

前面提到过，基于同步的定位算法之所以越来越少的被利用在于其存在巨大的缺陷。缺陷在于真正的收发同步实际很难达到，一些线路带来的延迟以及收发的延迟很难精确统计。虽然在非同步状态下，得到的到达时间直接转换为传播距离的话会引入很大误差，但是这并不影响我们参考同步状态下为声音反射信号归类的思想。参考同步状态下归类信号的方式，可以注意到当两个传播距离做差可以表示为一组双曲线，那么取其中 1 个传播距离分别与其他 n 个传播距离做差就可以形成 n 条双曲线。如果 n 条双曲线可以确定唯一的一个交点，这个点就是可能的一个声源坐标，这组距离分类成功。否则，如果不存在交点，这组距离匹配失败，原理如下图 5.25 所示：

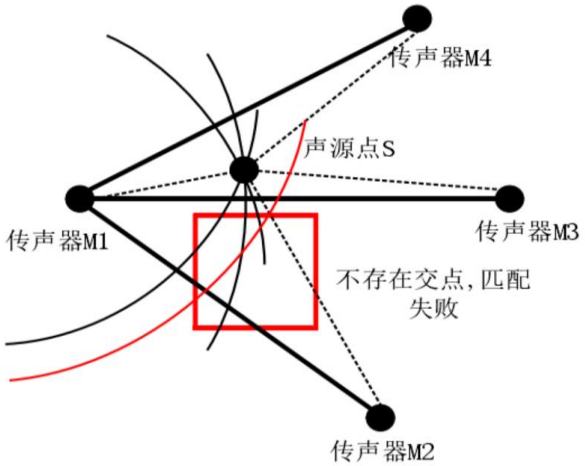


图 5.25 利用双曲线法进行反射声信号归类的过程

Fig.5.25 Classification of reflected acoustic signals by hyperbolic method

也就是说，利用最小二乘误差来检验待测的一组到达时间转换成的传播距离构成的双曲线是否存在交点。假设待测声源坐标为 (x, y, z)，存在 $n-1$ 个麦克风，坐标分别为 (x_1, y_1, z_1)，(x_2, y_2, z_2)，…，(x_n, y_n, z_n)，其中 (x_1, y_1, z_1) 为主麦克风坐标，利用 TDOA 双曲线法求解坐标 (x, y, z) 的过程如公式（5.64）到（5.79）所示。

引入最小二乘法的思想来判断求得的 (x, y, z) 的准确性，即有目标函数如下所示：

$$\varepsilon(x, y, z) = \sum_{i=2}^{n} \omega_i \left[c(t_i - t_1) - d_{i1} \right]^2 \tag{5.80}$$

$$d_{i1}=\sqrt{(x_i-x)^2+(y_i-y)^2+\left(z_i-\hat{z}\right)^2}-\sqrt{(x_1-x)^2+(y_1-y)^2+\left(z_1-\hat{z}\right)^2} \quad (5.81)$$

上式中的$\varepsilon(x,y,z)$代表定位的最小二乘误差，用来衡量求得的声源坐标(x,y,z)的准确性，d_{i1}代表求得的声源坐标到主麦克风 1 之间的距离与声源坐标到其他$n-1$个麦克风之间的距离差。基于以上论证，无需收发同步，通过到达时间作差利用双曲线定位法来判断也可以完成声音反射信号的分类，算法可以分为以下步骤：

1.利用接收到的反射信号分别计算各个麦克风接收端的房间冲激响应。

2.在求得的各冲激响应中随机选取信号的到达时间，并经过上述公式转换为信号的传播距离。

3.选取一个麦克风为主麦克风，其他为辅麦克风，分别计算信号到达主麦克风传播距离与信号到达辅麦克风传播距离的差。

4.将得到的传播距离分别代入公式中计算是否存在交点。

5.若存在交点将交点带入公式（5.80）、（5.81）中，计算结果$\varepsilon(x,y,z)$，若不存在交点d_i趋向于无穷大，ε随之也趋向于无穷大，代表函数不收敛，这组 TOA 不属于一类，令$\varepsilon(x,y,z)=0$。

6.遍历所有传播距离的组合，有时在误差的影响下，匹配时会出现多个ε值的现象，也就是说同一个到达时间 TOA 被分配到了两组。理论上，当声源位置距离所有反射面距离都不相同时，一个 TOA 只能和唯一的一组 TOA 形成交点。当出现这种情况出现需要比较ε大小，选择误差最小的那一组到达时间 TOA 作为一组经过同一反射面的反射信号。因为当存在使$\varepsilon(x,y,z)$最小时，认为$t=[t_1,t_2\ldots t_n]^T$就是最符合定位结果的一组反射到达声，因此可以被归为一类。与此同时，$\left[\hat{x}\ \hat{y}\ \hat{z}\right]^T$可以被认为是一个镜像声源。

7.每当归类好一组反射信号便去除已经归类好的反射信号，重复以上过程就可以不断为反射信号归类，直至为所有反射信号归类成功。具体流程如下图所示：

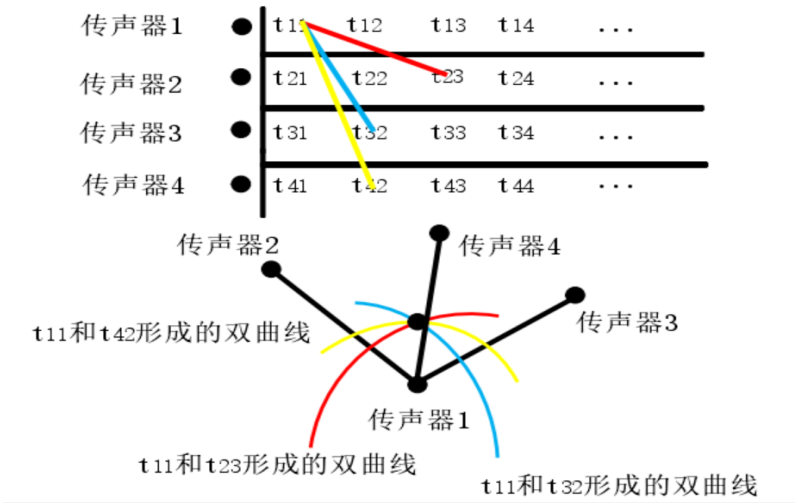


图 5.26 反射信号的匹配分类过程

Fig.5.26 Matching and classification process of reflected signals

假设存在m个麦克风组成阵元的麦克风阵列，分别可以计算得到m个冲激响应序列，记录每个冲激响应的到达时间组成m个 TOA 序列，假设每个序列包含n个信号到达时间，建立数据库如上图所示。直达信号第一个到达不需要分组，去除首个直达信号的 TOA，t代表接收到的反射信号 TOA，第一个小标代表麦克风序号，第二个小标代表接收到信号的次序，例如t_{23}代表第 2 个麦克风收到的第 3 个反射信号。为了更好遍历

所有到达信号，以麦克风 1 为主麦克风，依次取$t_{11}\sim t_{1n}$分别与麦克风 2 的$t_{21}\sim t_{2n}$，麦克风 3 的$t_{31}\sim t_{3n}$一直到麦克风n的$t_{m1}\sim t_{mn}$组成一组$1^{*}m$维的数组进行最小二乘误差匹配，计算最小二乘误差$\varepsilon(x,y,z)$。如果得到的坐标值最小二乘误差最小那么就将其归为一类并输出声源位置作为这些到达信号的镜像声源。需要说明的是，考虑到优化算法的复杂度的问题，不需要对每一个 TOA 值都进行匹配，根据三角形原理，两边之差小于第三边，来源于同一声源的信号之间的波程差必然小于阵列中任意两个阵元之间距离的最大值，如下图所示：

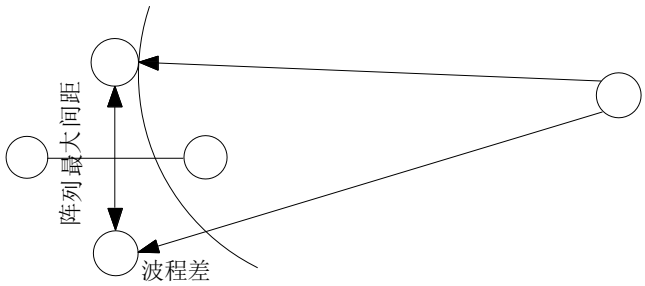


图 5.27 波程差小于麦克风阵列的最大间距示意图

Fig.5.27 Schematic diagram of the maximum spacing of the microphone array with wave range difference smaller than that of the microphone array

因此，在匹配的过程中设置搜索门限，减小搜索范围为：

$$t_{1n}-\tau<t_{in}<t_{1n}+\tau\left(\tau=\frac{d_{\max}}{c},i=1,2,3\cdots,m\right) \tag{5.82}$$

其中，上式中，$d_{\max}$代表阵元中的最大距离，i代表除主麦克风外的$m-1$个麦克风。由于设置搜索门限可以大大提高匹配效率，因此在设计麦克风阵列时，需要考虑麦克风阵列阵元之间的距离以达到最好的效果。

考虑到要匹配的组合的数目，假设存在 4 个麦克风，每个传声检测 20 个反射信号，并且这 4 个麦克风之间的最大距离为$1m$，对于每一个峰值必须在剩下 3 个麦克风里寻找，设置门限大小为$2\times\frac{1m}{343m/s}$，大约是$6ms$的时间。一般来说，$6ms$的时间窗内含有 5 个反射信号，因此必须计算$20\times5^{4}=12500$个ε值，这对一台普通电脑来说几秒内就可以完成。实际上，当成功分类一组反射信号就可以在后续的分类中将已经分类好的排除在外，因此根据对计算复杂度的估计，这是完全可以接受的。

综上所述，一个四阵元的反射信号分类算法流程可以用下图表示：

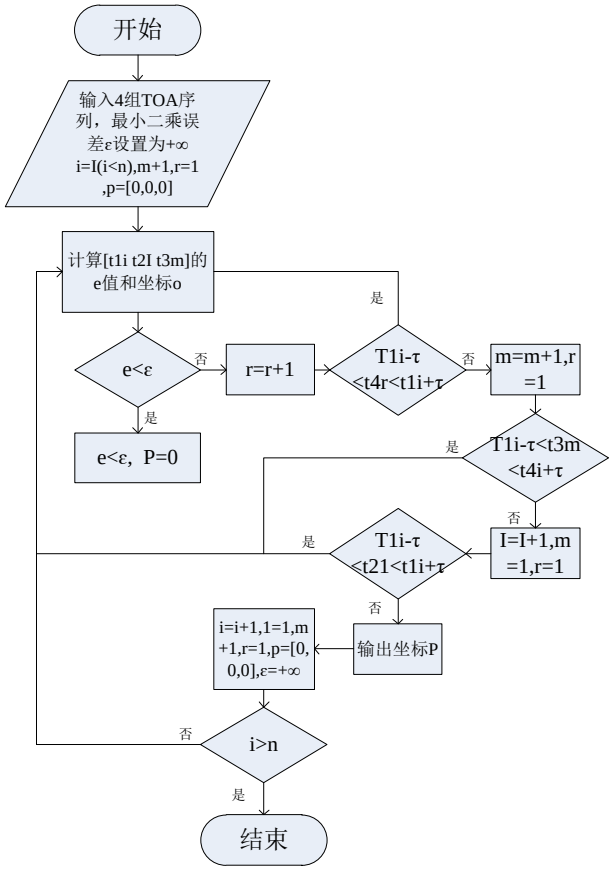


图 5.28 分类算法流程图

Fig.5.28 Flow chart of classification algorithm

5.2.3 接收信号的处理与优化

早期反射声也就是只经过一次反射的信号由于有着非常好的清晰度和明晰度，其信号可以视为重构室内图形的理想信号。但是，由于反射信号到达顺序和发射端、接收端、反射面三者位置相关，有些二阶反射信号甚至会先于一阶反射信号到达，造成只根据冲激响应图无法判断一阶反射信号和二阶反射信号的界限的问题。如果错选了二阶及其以上高阶信号，不仅可能会引入高阶镜像声源造成一阶高阶镜像声源的混淆造成建图失败，还会造成一阶到达信号和高阶到达信号进行匹配造成时间的浪费。因此，在最后一部分建图时，对于直接到达声、早期反射声（一阶反射信号）、混响声（二阶极其以上高阶反射信号）的辨别就显得尤为重要。

5.2.3.1 脉冲响应选峰的优化问题

通常概况下，在确定信号到达时间时，普遍采用选择信号冲激响应的最高峰当作信号到达时间。实际上，由于麦克风阵列无法做到完全的同步，再加上人为测量误差的影响实际上选择最高峰作为信号到达时间是不准确的，在远场定位情况下这种误差会被放大从而影响对实际声源和镜像声源的判断。假设在 44.1kHz 的采样频率下，一个冲激响应由 10 个采样点组成，如果信号到达时间分别位于第 1 个采样点和第 10 个采样点，如果依然采用最高峰到达时间的话就会引入 7 厘米的误差。由于直达信号易于分离，本文采用利用直达信号对反射 TOA 进行修正。

如下图 5.29 所示，取完整直达信号峰值内的几个采样点（如下图红线内）分别进行最小二乘匹配，找到实际代表信号到达时间的位置，计算出最佳采样点位置与峰值的差距 τ，在后面进行反射信号匹配时分别以 tnm± τ 来代替反射信号到达时间 tnm，可以起到修正不同步的作用，实验证明利用这种方法进行修正可以消除 4~5 厘米误差。

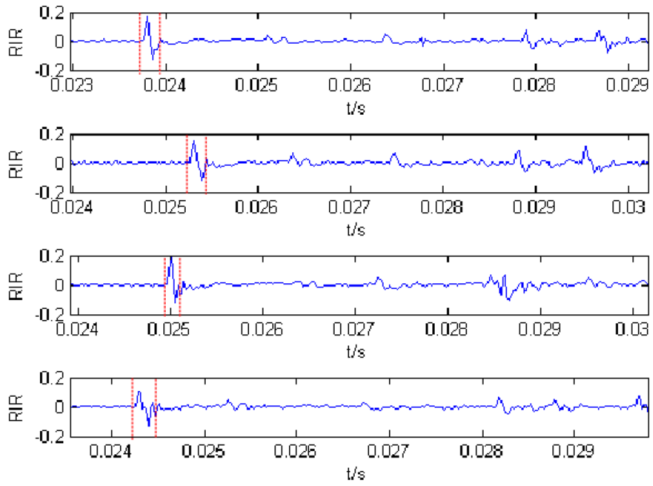


图 5.29 利用直达信号修正反射波 TOA

Fig.5.29 Correction of reflected wave TOA by direct signal

5.2.3.2 脉冲响应分类问题的研究

根据前面提到过的镜像源生成原理，在 Image 模型中，假设声源位置为 X，接收位置为 X'，室内脉冲可以表示为：

$$h(t,X,X')=\sum_{p=0}^{1}\sum_{r=-\infty}^{\infty}\beta_{x,1}^{|n-q|}\beta_{x,2}^{|n|}\beta_{y,1}^{|l-j|}\beta_{y,2}^{|l|}\beta_{z,1}^{|m-k|}\beta_{z,2}^{|m|}\frac{\delta\left[t-\left(\left|R_p+R_r\right|/c\right)\right]}{4\pi\left|R_p+R_r\right|} \tag{5.83}$$

其中，

$$R_p=\left(x-x'+2qx',y-y'+2jy',z-z'+2kz'\right) \tag{5.84}$$

$$p=\left(q,j,k\right) \tag{5.85}$$

$$R_r=2\left(nL_x,lL_y,mL_z\right)n,l,m=\ldots-1,0,1\ldots \tag{5.86}$$

$$R_p+R_r=\left(x-x'+2qx'+2nL_x,y-y'+2jy'+2lL_y,z-z'+2kz'+2mL_z\right) \tag{5.87}$$

R_p 指的是相对于坐标原点声源到镜像声源位置，p 表示向量 R_p 元素的个数，q，j，k 分别取 0 或 1。R_r 是高阶情况下反射的虚拟尺寸，r 代表反射阶数，n，l，m 的取值与 r 相关，$\beta_{x,1},\beta_{x,2},\beta_{y,1},\beta_{y,2},\beta_{z,1},\beta_{z,2}$ 分别代表房间各个面的反射系数。从脉冲响应表达式可以看出，一阶、二阶以至于多阶反射信号的不同之处在于到达时间不同和反射次数不同而导致的脉冲大小不同，分类也从这两点予以考虑。

（1）直达信号与一阶反射信号的分类研究

按照上述室内脉冲的表达形式，直达信号表达式为：

$$h(t,X,X')=\frac{\delta\left[t-\left|(x-x',y-y',z-z')\right|/c\right]}{4\pi\left|(x-x',y-y',z-z')\right|} \tag{5.88}$$

一阶信号的表达式同上式，n，l，m 取 1 或者-1。由此可以看出，一阶信号的到达时间为 $\left|x-x'+2qx'+2nL_x,y-y'+2jy'+2lL_y,z-z'+2kz'+2mL_z\right|/c$ 是无论如何都大于直达波的。从脉冲的大小来看，由于反射系数不大于 1，在没有噪声恒定没有突发噪声的情况下，一阶反射脉冲的大小也是小于直达波脉冲大小的。但是，以上条件都是在视距传播 LOS（line-of-sight propagation）的条件下满足的。但是在室内这种复杂空间中，可能会存在阻挡声波传播的障碍物，造成非视距现象的出现。因此在布置实验场景时，需要注意保证麦克风阵列中的每一个麦克风与声源之间没有障碍物遮挡。如果存在障碍物遮挡，那么就有可能接收不到直达信号。

（2）一阶反射信号与高阶反射信号的分类研究

对于一阶反射信号和二阶反射信号来说，它们都属于早期反射声的范畴。早期反射声指的是直达声到达后 50ms 之内接收到的声波，其中包含一阶反射波和部分二阶反射波。前面提到过一阶反射波的到达时间为：

$$t_1 = \frac{\left|x - x' + 2qx' + 2nL_x, y - y' + 2jy' + 2lL_y, z - z' + 2kz' + 2mL_z\right|}{c} \tag{5.89}$$

其中，n、l、m 至少一个为 1 或-1。二阶反射波的到达时间为：

$$t_2 = \frac{\left|x - x' + 2qx' + 2nL_x, y - y' + 2jy' + 2lL_y, z - z' + 2kz' + 2mL_z\right|}{c} \tag{5.90}$$

其中，n、l、m 至少一个为 2 或-2。为了分类一阶、二阶反射信号，考虑最后一个到达的一阶反射信号与第一个到达的二阶反射信号的边界。最后一个一阶反射波到达的情况下，q、j、k 取 1 或者-1，而对于第一个二阶反射波，q、j、k 取 0，对比两个到达时间，假设一阶反射波里的qx'，jy'，kz'为正值，那么对应的t_1式里的的n、l、m应该为正值且为 1。t_2 里面的n、l、m 只有一个为 2 或-2，其它为 0。比较两者的大小，发现两者大小取决与L_x、L_y、L_z的大小，也就是说在不知道房间尺寸的情况下无法证明一阶反射信号均早于二阶反射信号到达，那么就无法从时间域找到一条分界线区分一阶反射信号和二阶反射信号。

再来分析脉冲信号的能量，一阶脉冲信号的幅度表达式为：

$$a_1 = \frac{\beta_{x,1}^{|n-q|}\beta_{x,2}^{|n|}\beta_{y,1}^{|l-j|}\beta_{y,2}^{|l|}\beta_{z,1}^{|m-k|}\beta_{z,2}^{|m|}}{4\pi\left|R_p + R_r\right|} \tag{5.91}$$

其中，n、l、m 至少一个为 1 或-1。二阶脉冲信号的幅度表达式为：

$$a_2 = \frac{\beta_{x,1}^{|n-q|}\beta_{x,2}^{|n|}\beta_{y,1}^{|l-j|}\beta_{y,2}^{|l|}\beta_{z,1}^{|m-k|}\beta_{z,2}^{|m|}}{4\pi\left|R_p + R_r\right|} \tag{5.92}$$

其中，n、l、m 至少一个为 2 或-2。经过分析一阶反射信号的取值范围为：

$$\frac{\beta_{x,1}^{2}\beta_{x,2}^{1}\beta_{y,1}^{2}\beta_{y,2}^{1}\beta_{z,1}^{2}\beta_{z,2}^{1}}{4\pi\left|R_p + R_r\right|} \le a_1 \le \frac{\beta}{4\pi\left|R_p + R_r\right|} \tag{5.93}$$

其中，上式中β取$\beta_{x,1}$、$\beta_{x,2}$、$\beta_{y,1}$、$\beta_{y,2}$、$\beta_{z,1}$、$\beta_{z,2}$中的一个。而二阶信号的取值范围为：

$$\frac{\beta_{x,1}^{3}\beta_{x,2}^{2}\beta_{y,1}^{3}\beta_{y,2}^{2}\beta_{z,1}^{3}\beta_{z,2}^{2}}{4\pi\left|R_p + R_r\right|} \le a_2 \le \frac{\beta_{x,y,z}\beta_{x,y,z}^{2}}{4\pi\left|R_p + R_r\right|} \tag{5.94}$$

由此可见，二阶反射信号的取值与一阶反射信号的取值有重合的地方。因此同样没有明确的界限来区分一阶反射信号和高阶反射信号。但是，可以借鉴的是，可以根据墙面、地板、天花板的反射系数来设立门限阈值来缩小选峰范围，提高效率。

已知直达信号的脉冲响应幅值为：

$$a_0 = \frac{1}{4\pi\left|R_p + R_r\right|} \tag{5.95}$$

由于直达波的路径$R_p + R_r$小于反射波路径，因此上门限可以设为

$$a_1 < \beta a_0 \tag{5.96}$$

下门限考虑实际声源和阵列的布置方式，一般来说麦克风阵列放置在实验场景的中间位置，声源布置在靠近角落的位置，一阶反射波的传播距离近乎为直达波的 2 倍，因此下门限可以设为：

$$\frac{\beta_{x,1}^{2}\beta_{x,2}^{1}\beta_{y,1}^{2}\beta_{y,2}^{1}\beta_{z,1}^{1}\beta_{z,2}^{1}a_0}{2}<a_1 \quad (5.97)$$

相当于在室内脉冲响应中挑选一阶反射波时，在时域设置直达波到达后 50ms 的时域门限（如果早期反射后期已经出现由于混响严重造成的选峰困难也可以考虑缩小门限）满足早期反射声定义，在幅度域设置如上门限，这样做可以显著提高选峰的效率。

（3）二阶反射信号的去除方法

上述的方法提出了门限的方法来缩小取值范围和隔离排除一部分高阶信号，但是门限还存在一些漏检现象。在这里，我们设计一种利用一阶、二阶镜像源关系的方法来去除二阶反射信号。首先，按照匹配的步骤从先到达的 TOA 匹配到后到达的 TOA，先求出镜像声源坐标可以认为是低阶镜像声源，后求出的为高阶。考虑到二阶及其以上的声源满足几何关系，在这里规定，如果新求出的声源与已有的声源满足如下规律:

$$s_2 = s_1 + 2\left\langle p_2 - s_1, n_2 \right\rangle n_2 \quad (5.98)$$

$$p_2 = \frac{(s_1 + s_2)}{2} \quad (5.99)$$

$$n_2 = \frac{(s_2 - s_1)}{\|s_2 - s_1\|} \quad (5.100)$$

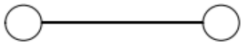
那么就认为新求出的声源坐标是一个二阶及其以上阶的声源。其中，s_1，s_2 分别代表已有的声源和新求出的声源，公式（5.99）中的 p_2 代表新旧声源连线的中点，公式（5.100）n_2 代表已有声源指向新声源的单位向量。去除高阶干扰后再结合待测房间结构去除不合理的镜像声源就可以得到真正的一阶墙面镜像声源，从而通过几何关系得到房间的几何构型以及尺寸。

5.2.4 麦克风树的设计与优化

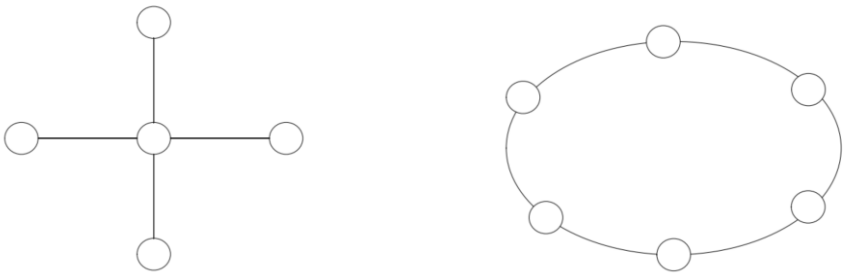
5.2.4.1 麦克风树的基本设计研究

单个麦克风数据量小而且单一，无法对目标进行定位分析，必须采用麦克风阵列来进行研究。麦克风阵列的布置对于定位的精度具有非常大的影响，对于室内地图重建来说对于麦克风阵列具有如下要求：1.能够实现全方位定位。2.结构简单易于布置。就目前来说，麦克风阵列的主要形状有：一维线阵、L 形阵、十字形阵列、圆形阵列、三维立体阵列这几种阵列。其中，一维线阵是最简单的阵列形式，可以在任意场合使用，但是一维线阵的定位范围有限，无法进行全方位定位，当声源存在于线阵两端或者靠近中心的地方误差都会快速增大。在二维平面阵列中，L 型阵列是比较简单的一种，如图所示，声源在其开口方向内的定位效果较好。十字形阵列是 L 形阵列的升级，十字形阵列分别在各个坐标轴上等距离放置麦克风，相当于来自任何位置的声源都处于阵列的开口方向内。圆形阵列如图所示，分别在圆或者椭圆轨迹上等距离放置麦克风，一般在语音增强方面被广泛使用。三维立体阵列需要的麦克风数量较多，一般在 10 个以上。其全方位定位效果非常好，没有盲点，但是由于计算比较复杂较少使用。

线性阵列



L 形阵列



十字形阵列　　圆型阵列

图 5.30 常见麦克风阵型图

Fig.5.30 Common microphone formation diagram

除了麦克风阵元的布置会影响定位结果外，麦克风的数量也会对定位结果产生影响。一般来说，麦克风数量越多，定位精度越高。但是麦克风数量过多也会带来相应的问题：1.增加麦克风数量无疑必须增加麦克风的体积，在空间有限的情况下必须考虑麦克风阵列体积的大小。2.麦克风数量过多会使定位算法复杂,继而导致较高的运算复杂度。综上考虑，选择数量合适的麦克风进行声源定位是非常重要的。

综上考虑，在进行室内重构时，由于镜像声源分别与实际声源关于各个墙面、天花板、地面对称，为了保证定位的准确性，选择十字型阵列进行实验。与传统平面十字型阵列不同的是为了定位天花板和地板所产生的镜像源，稍微调整麦克风高度做出高度差，数量的话为方便进行计算验证，暂时选择三维定位最少的 4 个，因此设计麦克风树的十字阵列中心点相对位置为（0，0，0），一号麦克风 M1 位置为（0，0.3m，1.31m），二号麦克风 M2 位置为（0.3m，0，1.31m），三号麦克风 M3 位置为（-0.3m，0，1.31m），四号麦克风 M4 位置为（0，-0.3m，1.31m），麦克风树高度以十字阵列中心点进行限定，高度设置为 2m 即可。

5.2.4.2 麦克风树的优化设计研究

1、遗传算法的基本原理

遗传算法是一种仿生算法，是受到自然界繁衍生息的自然规律的启发，它模拟生物体的遗传与进化过程，首先把待解决问题的参数编成码（即基因），若干基因组合成一个码串，码串的形式和长度与采用的编码方式有关。每个码串可看成是一个独立的个体，许多这样的个体就组成了一个种群，种群的大小取决于个体数的多少。对种群中的个体进行类似于自然复制、配对互交叉和个别基因突变等运算，经过多代繁衍直到最后的优良个体（即问题的最优结果）出现。其中，衡量个体优劣的标准是其适应度函数值的大小，适应度函数值越大，表示个体越优良（即解的质量越好）。适应度函数是由目标函数经一定变换得来的，目标函数的合理确定将决定着适应度函数的好坏。遗传算法的基本流程图如图 5.31 所示。

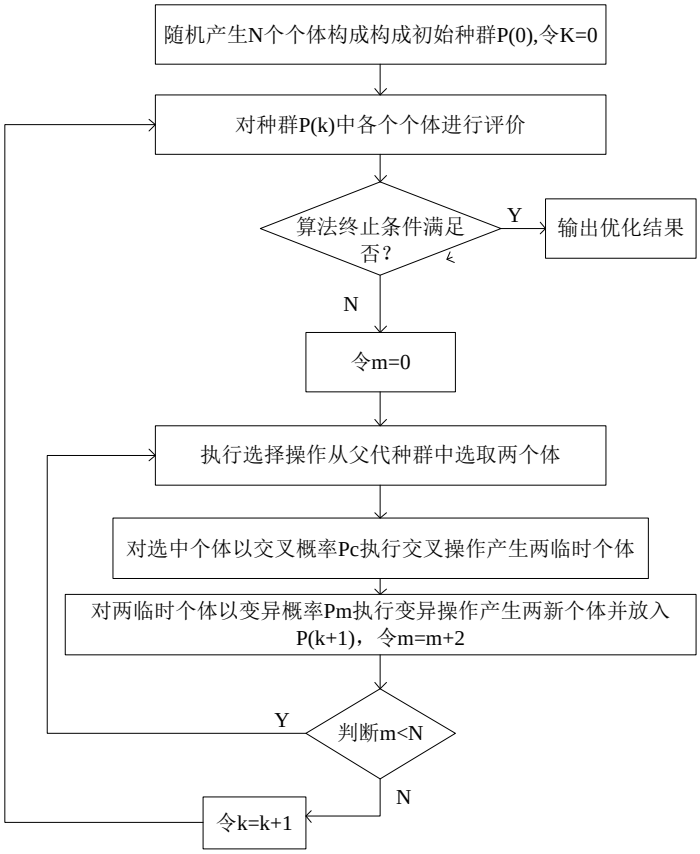


图 5.31 遗传算法框图

Fig.5.31 A diagram of genetic algorithm

2、遗传算法设计

在封闭环境下麦克风接收到的信号会受到混响和噪声的影响，模型如图 5.32 所示，第 i 个麦克风接收到的信号为 $x_i(n)$ 。

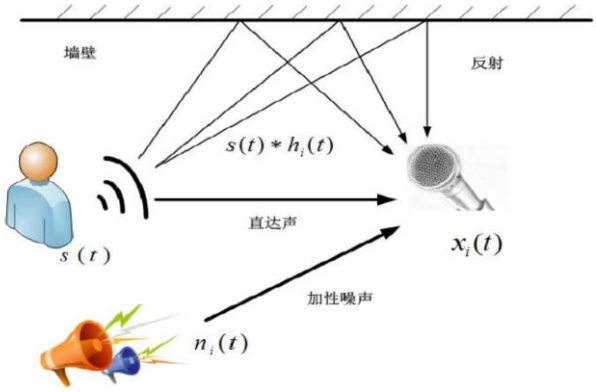


图 5.32 声学模型

Fig.5.32 Acoustic Model

$$x_i(n)=h_i(n)*s(n)+v_i(n) \qquad (5.101)$$

麦克风阵列接收到的语音 $x(n)$ 为

$$x(n)=\left[x_1(n)、x_2(n)、x_M(n)\right] \qquad (5.102)$$

其中 $i=1,\cdots M$ ， $h_i(n)$ 是声源 $s(n)$ 到第 i 个麦克风的冲激响应， $v_i(n)$ 为加性噪声，M 为麦克风个数，*表示卷积。

（1）编码规则

选取十字麦克风阵列的简化版-均匀 L 型麦克风树阵列作为实际问题参数集进行分

析，对不同的麦克风阵列结构进行二进制编码，选择其中 L 形阵元作为参考阵元，以一定的规律进行编码，1 代表此位置有阵元存在，0 代表此位置无阵元存在。例如一维均匀线阵，1101110110111111 编码代表 16 元一维均匀线阵中的第三、七和十位置上无阵元存在，其余位置有阵元存在。例如二维均匀矩形阵，1001101110101011 编码代表 4*4 元二维均匀矩形阵，1001 代表矩形阵第一行的第二和三列位置无阵元存在，其余位置有阵元存在；1101 代表矩形阵第二行的第三列位置无阵元存在，其余位置有阵元存在；1010 代表矩形阵第三行的第二和四列位置无阵元存在，其余位置有阵元存在；1011 代表矩形阵第四行的第二列位置无阵元存在，其余位置有阵元存在。

（2）停止条件设置

遗传算法停止条件多采用遗传进化终止代数，即先设置固定的遗传代数值，但是算法运行时，有时会在设置的固定遗传代数值之前优化得出最好结果，有时却在设置的固定遗传代数值之后优化得出最好结果，即这样设置停止条件不能保证在满足停止准则时所得结果为最优。

于是，将声源角度估计误差值代替传统优化设计中采用的遗传进化终止代数作为停止条件，即这里采用的算法停止条件为，优化前后麦克风阵列对应估计出的声源的二维来波角度差值。

（3）目标函数与适应度函数构造

本文构造的目标函数是由二维 MUSIC 空间谱函数值间的欧式距离和阵元个数共同组成。设有 D 个信号入射到 M 个（M>D）阵元的阵列上，则接收的信号为

$$\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_M \end{bmatrix} = \begin{bmatrix} a(\theta_1,\varphi_1) & a(\theta_2,\varphi_2) & \cdots & a(\theta_D,\varphi_D) \end{bmatrix} \begin{bmatrix} S_1 \\ S_2 \\ \vdots \\ S_D \end{bmatrix} \begin{bmatrix} N1 \\ N_2 \\ \vdots \\ N_M \end{bmatrix} \tag{5.103}$$

式（5.103）可简写为

$$X^* = A^* S + N \tag{5.104}$$

上式中 A 为麦克风阵列的方向向量，N 为阵列接收的随机噪声向量，S 为声源信号向量。

对式（5.104）中 X 的协方差矩阵进行特征值分解，得到对应于 D 个大特征值的特征向量构成的信号子空间 Us 以及 M-D 个特征向量构成的噪声子空间 Un。构造二维 MUSIC 空间谱函数为

$$P_{MUSIC}(\theta,\varphi) = \frac{1}{a^H(\theta,\varphi) U_n U_n^H a(\theta,\varphi)} \tag{5.105}$$

构造的目标函数为

$$objvalue = \alpha * \sqrt{\sum (P_{MUSIC1})(\theta,\varphi) - (P_{MUSIC})(\theta,\varphi)^2} + \beta * L \tag{5.106}$$

假设 PMUSIC1 和 PMUSIC 分别代表优化前后阵列对应求得的二维 MUSIC 空间谱函数，L 为优化后阵列中阵元的总个数，α、β 为权重，需要满足 α + β =1。对目标函数进行适当变换后得到的适应度函数为

$$fitvalue = \begin{cases} C_{\max} - objvalue, & objvalue < C_{\max} \\ 0, & objvalue > C_{\max} \end{cases} \tag{5.107}$$

上式中 Cmax 为一个适当的相对比较大的数，是目标函数 objvalue 的最大值估计。

（4）仿真实验结果与分析

实验在有混响和噪声的环境下进行，选取的房间规格为 7m×5m×4m，房间混响时间为 263ms，实验选取实际问题参数集中阵元个数为 16，声源为男生朗读“第一课认识

新同学”，信噪比 SNR=5dB，声源坐标为（4，3，1.43），入射方向为（44.682°，45.000°），停止条件是当角度差值小于等于 1°，交叉概率为 Pc=0.9，变异概率为 Pm=0.005。ULA 麦克风阵列参考阵元坐标为（3，2，0），阵元间距均为 0.04 米，语音的采样率为 16kHz；使用 Image 模型构造房间冲激响应，长度为 2048 点。对此方法进行仿真时，选取数据矩形窗，窗长为 128，通过实验选取最优步长 u 为 0.001。实验选取的平面 16 元均匀 L 型阵如图 5.33 所示。

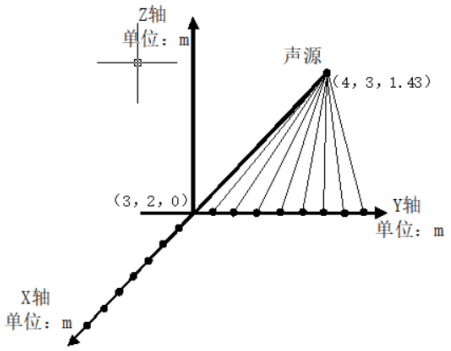
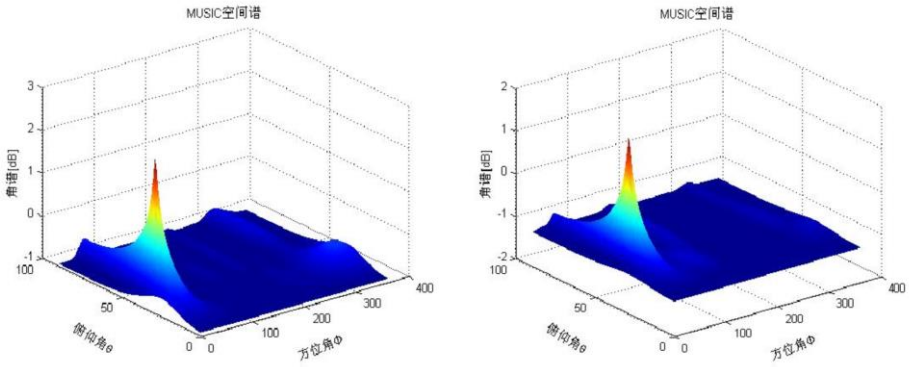


图 5.33 均匀 L 型阵空间坐标

Fig.5.33 ULA spatial coordinates

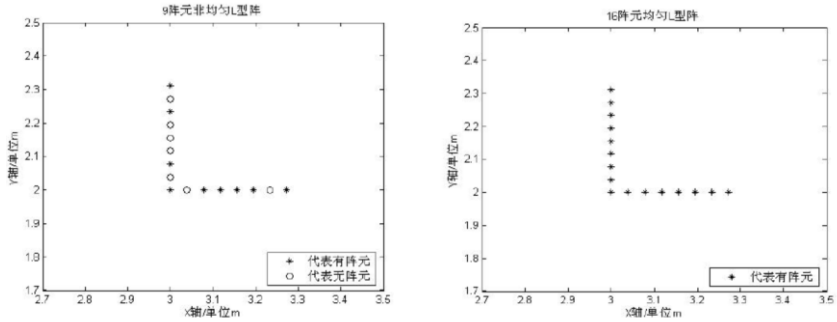
采用上述方法优化上述 L 形阵列的仿真结果如图 5.34 所示，优化前后 L 型阵对应阵元在 XOY 平面上的摆放位置分别如图 5.35 所示。



（a）优化后阵元数 M=9　　（b）优化前阵元数 M=16

图 5.34 优化前后 L 型阵对应求得估计结果

Fig.5.34 L matrix corresponding to the estimated results before and after optimization



（a）优化后　　（b）优化前

图 5.35 优化前后 L 型阵对应阵元摆放位置

Fig.5.35 Before and after the optimization of L array corresponding element location

由图 5.34 可知，同等条件下优化后的 L 型阵与优化前的 L 型阵的 MUSIC 空间谱图相比，优化后的 L 型阵的峰值主瓣宽度与优化前的 L 型阵相差不大，峰值尖锐程度优于

优化前的 L 型阵，但是优化后的 L 型阵的旁瓣影响较明显。总体来说，优化后的 L 型阵可以得到较为准确的位置估计信息，基本可以达到优化前的 L 型阵的位置估计性能。

由图 5.35 可知，在阵列孔径保持一致的前提下，优化后的 L 型阵比优化前的 L 型阵可节省 7 个阵元。仿真结果也证明了优化后的 L 型阵对单声源的 2-DDOA 估计性能并没有因阵元数目的减少而受到影响。

由此，针对于不同使用类型、用途，均可使用此基于遗传算法的麦克风树优化算法进行优化计算，使得麦克风树设计结果更为合理。

5.2.5 麦克风树设计的实例验证

5.2.5.1 麦克风树设计的验证试验

Image 模型用来描述小房间混响模型。小房间模型意味着必须保证直达波和一阶反射信号在房间的任意位置都可以全部接收到。考虑到室内环境的复杂性以及常用扬声器发声所能传播的距离，对于实验场景来说必须满足三个条件：

（1）在实验场景内的障碍物必须保证障碍物尺寸小于声波波长。

（2）在实验场景内的任意一点都可以接收到完整、清晰易于判别的脉冲信号。

（3）实验场景符合规则多边形的结构。

根据以上三个限制条件，在教室、教师办公室、走廊等场所均进行了测试，教室、教师办公室由于房间比较大而且桌椅等障碍物很多往往收到的信号会出现不完整以及不清晰等很明显的混响现象。相比教室以及教室办公室，走廊的好处在于除了地板墙面以及天花板之外，内部一般没有别的障碍物，而且走廊是非常规则的四边形结构便于实现重建。本实验选择在山东科技大学一号教学楼走廊进行，实验环境如下图所示：

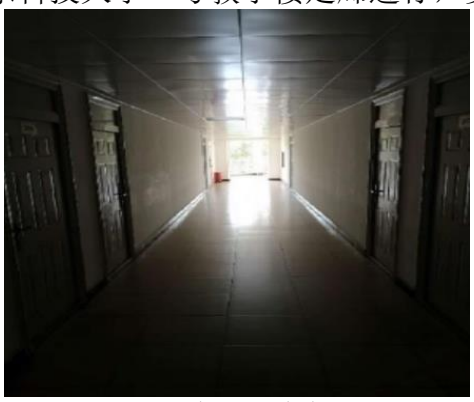


图 5.36 实验环境实景图

Fig. 5.36 real scene diagram of the experimental environment

走廊宽为 2.3m 高度为 2.37m，前后没有墙壁遮挡，左右两面墙和天花板以及地板足够平整，走廊的轮廓符合规则四边形结构。走廊地板材料为白色釉面瓷砖，天花板为铝板材质，左右两侧墙面为涂刷白色墙漆的水泥墙体。查询各材质反射系数，白色釉面瓷砖反射系数为 0.80，铝板反射系数为 0.70，白色墙漆反射系数为 0.84，实验场景如图所示。

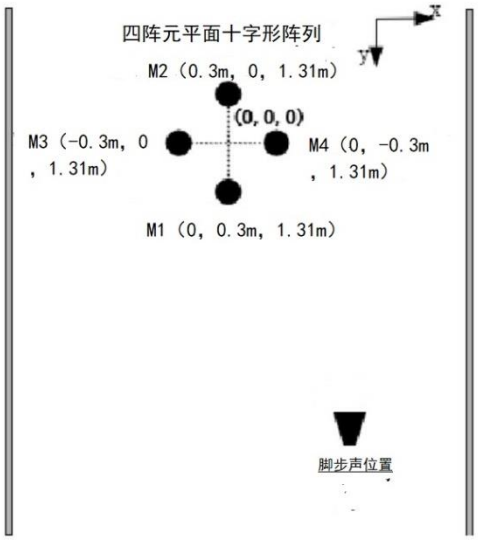


图 5.37 实验场景示意图

Fig.5.37 Schematic diagram of the experimental scene

5.2.5.2 实验结果及对比分析

利用频域接收信号 Y(ω)和频域发射信号 S(ω)经过计算得到的四个通道冲激响应经过傅里叶逆变换得到的时域冲激响应如下所示：

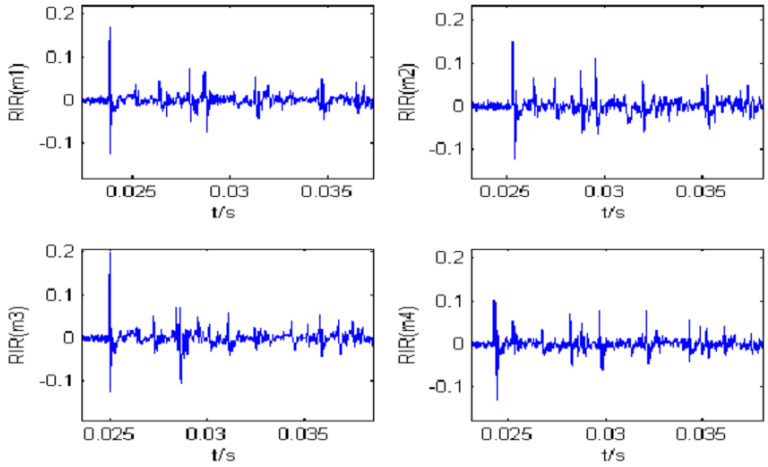


图 5.38 麦克风阵列四阵元接收到的脉冲响应图

Fig.5.38 Pulse response received by microphone array quad

根据麦克风阵列接收到的冲激响应图可以看出，率先到达的峰对应直达信号，随后反射信号依次到达。以麦克风正中心靠近地面的位置为坐标原点，以走廊通行方向为 y 轴，以垂直两墙壁方向为 x 轴，以地板到天花板的垂直方向为 z 轴，利用四个直达信号的 TOA 得到声源坐标为（0.76 m, 2.01 m, 0.70m），接下来利用幅度阈值和时间阈值来进行选峰，时间阈值设置为 50ms 之内，幅度阈值按照第四章的分析结果进行设置分别选择了 10 个峰值到达时间组成 10*4 的一组数据，经过匹配算法分别求得声源位置坐标为：（1.46m,2.02m,0.72m），（-3.05m,1.96m,0.82 m），（0.83 m,1.97m,3.45 m），（0.77 m,2.03 m, -0.94 m），（0.40 m,1.95 m,1.13m），（0.94 m,1.39m,-1.46 m），（1.09 m,-0.42 m,2.03 m）(由于后 3 个峰值到达时间不满足间隔小于阵列最大间距所对应的时间值，因此没有产生声源坐标)。绘制以上坐标在 x、z 方向的投影如下图所示：

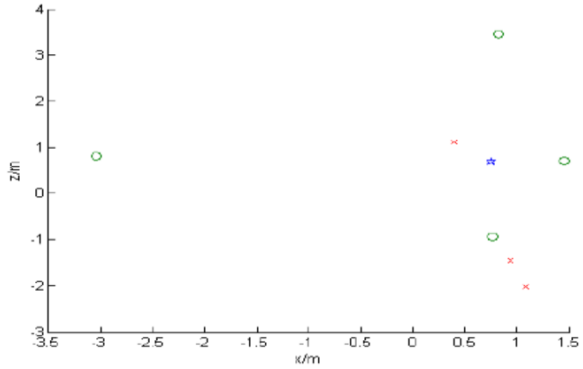


图 5.39 利用 TDOA 最小二乘误差法得到的镜像源坐标

Figure 5.39 the mirror source coordinates obtained by using the TDOA least squares error method

图 5.39 五角星代表实际声源，x 坐标经过式（5.98）（5.99）（5.100）检验不符合二阶声源定义，（0.40m, 1.95m, 1.13m）离实际声源太近显然不是符合镜像声源常理，（0.94m, 1.39m, -1.46m），（1.09m, -0.42m, -2.03m）这两个坐标 y 轴偏差太大，也可以排除。因此它们代表由噪声产生的错误声源，圆圈代表一阶镜像声源。利用实际声源和镜像声源，作它们连线的垂直平分线可以得到下图：

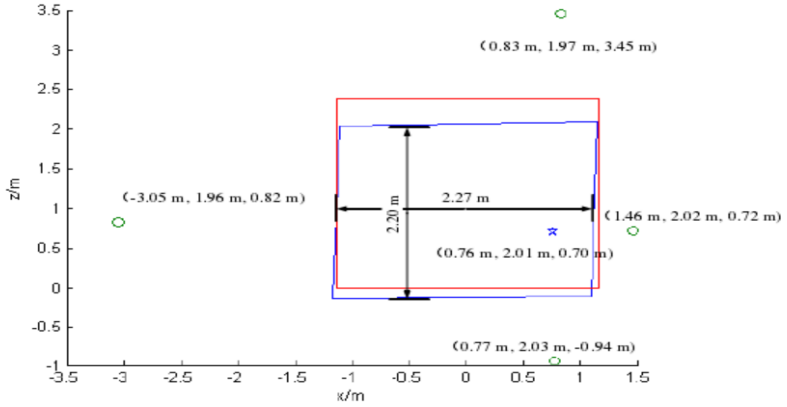


图 5.40 利用 TDOA 最小二乘误差法得到的走廊结构与实际走廊结构的对比

Fig.5.40 Comparison between the corridor structure obtained by the least squares error method of TDOA and the actual corridor structure

表 5.7 TDOA 最小二乘法得到结果与实际情况的对比

Tab.5.7 Comparison between the results obtained by TDOA least square method and the actual situation

4 个反射面	镜像源坐标/m	理论镜像源坐标/m	声源到各个墙面距离与实际距离的误差/m
左墙面	(-3.05 1.96 0.82)	(-3.07 2 0.82)	0.01
右墙面	(1.46 2.02 0.72)	(1.53 2 0.82)	0.02
天花板	(0.83 1.97 3.45)	(0.79 2 3.43)	0.05
地板	(0.77 2.03 -0.94)	(0.79 2 -0.82)	0.12
平均			0.05

图 5.40 对应代表利用 TDOA 最小二乘法求出的走廊左右墙面、天花板、地面的位置以及与真实值的对比，其中蓝色代表利用 TDOA 最小二乘法求出的走廊轮廓，红色代表理论值。从表 5.7 可以读出声源到各个墙面距离的误差分别为 0.01m、0.02m、0.05 m、0.12m，x 轴方向平均误差为 0.03m，z 轴方向平均误差为 0.17 m，整体平均误差为 0.05m。在一定误差范围内实现了对走廊的几何重构。

作为对比，利用 EDM 法得到走廊结构、用 TDOA 最小二乘法得到的走廊结构以及与实际情况的对比如下图 5.41 所示：

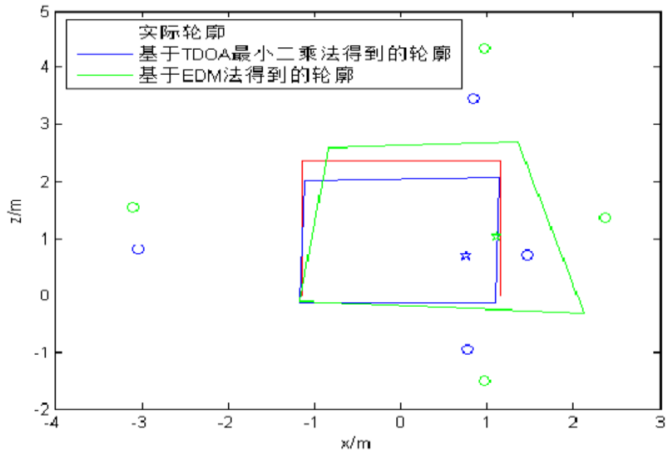


图 5.41 利用欧氏距离法得到的走廊结构与实际走廊结构的对比

Fig.5.41 Comparison between the corridor structure obtained by Euclidean distance method and the actual corridor structure

表 5.8 EDM 法得到结果与实际情况的对比

Tab.5.8 Comparison between the results obtained by EDM method and the actual situation

4 个反射面	镜像源坐标/m	理论镜像源坐标/m	声源到各个墙面距离与实际距离的误差/m
左墙面	(-3.11 2.12 1.56)	(-3.07 2 0.82)	0.20
右墙面	(2.37 1.93 1.36)	(1.53 2 0.82)	0.28
天花板	(0.96 1.81 4.33)	(0.79 2 3.43)	0.10
地板	(0.96 2.00 -1.5)	(0.79 2 -0.82)	0.45
平均			0.26

表 5.8 可以得到声源到各个墙面距离的误差分别为 0.20m、0.28m、0.10m、0.45m，x 轴方向平均误差为 0.48m，z 轴方向平均误差为 0.55m，整体平均误差为 0.26m。可以明显看出，本文提出的基于 TDOA 最小二乘法的重构方案是优于 EDM 法的，通过此试验，不但验证了所使用的基于 TDOA 的最小二乘法算法的优越性，同时也证实了麦克风树设计的合理性。

六、结果分析

1、问题一的结果分析

通过研究传统时延估计四元十字麦克风阵列的研究基础上，建立一种基于相位变换加权的广义互相关七元麦克风阵列声源定位算法，该算法在考虑到背景噪声、房间混响的情况下，首先建立了七元麦克风阵列模型，利用该麦克风阵列接收室内声源声音信号，对接收的声音信号进行谱减去噪、倒谱去混响处理，然后利用相位变换加权广义互相关方法 GCC-PHAT，得各麦克风之间的时间延迟，最后运用三维空间定位方法与坐标旋转数字式计算机方法，确定声源位置。

2、问题二的结果分析

本文针对于室内轮廓构图原理进行了系统理论研究，在理论分析的基础上提出了两种声音信号的处理算法，欧式距离法和 TDOA 法，并对接收到的声音信号进行优化处理，通过提出一种基于遗传算法的麦克风树优化设计算法，对不同阵列形式麦克风树进行优劣对比耦合分析，最终选取大厅轮廓、声源位置估算结果较为精准、计算过程较为简洁的十字阵列麦克风树进行现场试验研究并详细给出了每支麦克风的相对位置信息，最后的现场试验结果数据证实了 TDOA 算法的优越性，验证了麦克风树阵列设计的合理性，整体平均误差为 0.05m，满足了实用要求。

七、模型评价与推广

7.1 模型的优点

1、问题一模型优点

一种基于相位变换加权的广义互相关七元麦克风阵列声源定位算法在混响、噪声等干扰信号存在的仿真室内环境下，能准确确定声源位置，提高了定位精度，得到满意的解，使结果更加合理，极大的减小了误差，能够使模型得到最大限度的简化。

2、问题二模型优点

基于 TDOA 的最小二乘分类法可对于信号到达时间进行作差处理，消除了时延带来的影响，保证了较高的声音信号分类准确度。

基于遗传算法的麦克风树阵列设计模型可在保证声音信号处理效果的基础上，减少麦克风的数量，兼顾了估算精度、实用经济、计算简洁三个方面。

7.2 模型的缺点

1、问题一模型缺点

该模型对外界环境要求较高，其声源定位的精确性与声音信号的去噪、去混响紧密相关。

2、问题二模型缺点

TDOA 法定位存在一些精度恶化的情况，尤其是在基线过短或者定位物体方向与基线垂直的情况下。

7.3 模型的改进

1、问题一模型改进

基于 PHAT 的三维七元麦克风阵列声源定位算法，对外界环境要求较高声源定位的精确性与语音信号去噪、去混响紧密相关，可对此进行进一步研究，降低这种相关性，提高定位的适用性。

2、问题二模型改进

可利用 AOA 等定位方法与 TDOA 法相结合的方式获得更为精确的结果，解决 TDOA 法可能存在的定位存在一些精度恶化的情况。

由于时间限制与计算简洁方面考虑，仅对问题二采用的十字阵列麦克风树的简洁版-L 型阵列进行了验证分析，因此，可对完整的十字阵列设计的麦克风树进行下一步研究，使得设计完成的麦克风树阵列表现更为优秀。

八、参考文献

[1] He Q, Zhang Y. On prefiltering and endpoint detection of speech signal [J]. International Conference on Signal Processing Proceedings, ICSP. 1998,1:749-752.

[2] 安栋，杨杰.数字音频基础[M].上海:上海音乐学院出版社,2011.

[3] 苏健民，黄英来，于慧伶.基于语音信号端点检测技术的研究应用[J].林业机械与木工设备.2006(06):49-50.

[4] 沈宏余，李英.语音端点检测方法的研究[J].科学技术与工程.2008(15):4396-4397.

[5] Wall R W. Simple Methods for Detecting Zero Crossing[C]. Roanoke, VA, United states: IEEE Computer Society. 2003.

[6] 吴艳花.语音短时幅度和短时过零率分析与应用[J].电脑知识与技术.2009(33):9314-9315.

[7] 吕卫强，黄荔.基于短时能量加过零率的实时语音端点检测方法[J].兵工自动化.2009(09):69-70.

[8] Liu K, Xiang J, Hou T, et al. Joint method of zero-crossing detection and energy detection[J]. Xuebao/Acta Armanentarii. 2008,29(9):1044-1048.

[9] 赵圣崔牧凡尤磊。基于小型麦克风阵列的声源定位技术[J].华中科技大学学,2013,41(I)：188-191.

[10] Ali Pourmohammad, Seyed Mohammad Ahadi. Real Time High Accuracy 3-D PHAT-Based Sound Source Localization Using a Simple 4-Microphone Arrangement[J]. IEEE Systems Journal, 2012, 6(3):455-468.

[11] 张辉,王盼,肖军浩等.一种基于三维建图和虚拟现实的人机交互系统[J].控制与决策,2018,33(11):58-65.

[12] 杨莎莎.基于低频声场特征的数值模拟与分析[D].南京理工大学,2015.

[13] Allen J, Berkley D. Image method for efficiently simulating small room acoustics[J]. J. Acoust. Soc. Am., 1979, 65(4):943-950.

[14] Dokmanic I, Parhizkar R , Walther A , et al. Acoustic echoes reveal room shape[J]. Proceedings of the National Academy of Sciences, 2013, 110(30):12186-12191.

[15] Yan J, Kleijn W B. Fast simulation method for room impulse responses based on the mirror image source assumption[C]// IEEE International Workshop on Acoustic Signal Enhancement. 2016.

[16] Neely S T, Allen J B. Invertibility of a room impulse response[J]. Journal of the Acoustical Society of America, 1979, 66(1):165-169.

[17] Mourjopoulos J. On the variation and invertibility of room impulse response functions[J]. Journal of Sound & Vibration, 1985, 102(2):217-228.

[18] Kuster M, Vries D D ,Hulsebos E M, etal. Acoustic imaging in enclosed spaces: Analysis of room geometry modifications on the impulse response[J]. Journal of the Acoustical Society of America, 2004, 116(116):2126-2137.

[19]Seetharaman P, Pardo B. Reverbalize:A Crowdsourced Reverberation Controller[J]. 2014.

[20] Dokmanic I , Parhizkar R, Walther A, etal. Acoustic echoes reveal room shape[J]. Proceedings of the National Academy of Sciences, 2013, 110(30):12186-12191.

[21] Dokmanic I, Parhizkar R, Ranieri J, etal. Euclidean Distance Matrices: Essential theory, algorithms, and applications[J]. IEEE Signal Processing Magazine, 2015, 32(6):12-30.

[22] Takane Y. Applications of multidimensional scaling in psychometrics[J]. Psychometrics, 2007,26(06):359-400.

[23] Takane Y, Young F W, De Leeuw J. Nonmetric individual differences multidimensional scaling:An alternating least squares method with optimal scaling features.[J]. Psychometrika,1977,42(1):7-67.