

基于时间差估计声源定位技术的研究

摘要：近年来，声源定位技术在军事、工业、航空航天、智能制造等领域都有广泛的应用需求。本文以室内近场声源定位为研究背景，并结合国内外前沿研究，对时间差估计定位技术进行了深入研究。主要做了以下工作：

① 对基于麦克风阵列的时间差估计定位原理进行阐述，提出使用基本互相关法和广义互相关法以计算时间差。

② 对根据时间差的声源定位方法模型进行了阐述，提出了常见的几何模型并取其中一个作为设计主线。

③ 对整个实验过程进行了仿真模拟，并使用预处理信号、LMS 滤波、加窗分帧等方式对模拟数据进行处理，说明实验的可行性。

④ 依据模拟实验过程，在室内进行实验，构建了基于 ELVISIII 和 labview 的定位装置。

⑤ 基于算法和实验对各部分进行误差分析，提出改进实验和减小误差的方法，并据此得出实验定位准确性的经验规律。

⑥ 最后，笔者通过以上工作提出进一步工作的重点。

关键词：时间差估计，广义互相关，labview

Research On Sound Source Localization Technology Based On Time Delay Estimation

Abstract: In recent years, sound source localization technology has been widely used in military, industrial, aerospace, intelligent manufacturing and other fields. In this paper, based on the research background of indoor near-field sound source localization, combined with domestic and foreign frontier research, an in-depth study of time difference estimation and localization technology is carried out. Mainly do the following work:

① The principle of time difference estimation and positioning based on the microphone array is expounded, and the basic cross-correlation method and the generalized cross-correlation method are proposed to calculate the time difference.

② The sound source localization method model based on time difference is expounded, and common geometric models are proposed and one of them is taken as the main design line.

③ The whole experimental process is simulated, and the simulated data is processed by means of preprocessing signal, LMS filtering, windowing and framing, etc., to illustrate the feasibility of the experiment.

④ According to the simulation experiment process, the experiment was carried out indoors, and a positioning device based on ELVIS III and labview was constructed.

⑤ Analyze the error of each part based on the algorithm and experiment, propose methods to improve the experiment and reduce the error, and obtain the empirical law of the accuracy of the experimental positioning.

⑥ Finally, the author puts forward the focus of further work through the above work.

Keywords: delay estimation, generalized cross-correlation, labview

目 录

一、 前言	7
二、 麦克风阵列声音定位技术原理	7
2.1 基于麦克风阵列的声音定位技术	7
2.2 TDOA 定位技术的确定	8
三、 基于时间差估计的声音方位定位方法	8
3.1 五传感器算法	8
3.1.1 模型图像描述	9
3.1.2 模型公式	9
3.2 从时延估计开始的误差传递	11
3.3 装置大小对模型精度的影响	11
四、 时间差估计方法	12
4.1 信号预处理	12
4.1.1 信号端点检测	13
4.1.2 加窗分帧	14
4.2 基于麦克风阵列的声音传播模型	16
4.3 互相关分析的时间差估计法	17
4.3.1 基本互相关算法	17
4.3.2 广义互相关(GCC)算法	18
4.3.3 基于互相关函数的仿真模拟	20
五、 仿真模拟	24
5.1 声源坐标	25
5.2 采样频率	26
5.3 传感器距离	28
5.4 声音频率	30
六、 实验装置介绍	31
6.1 硬件部分	31
6.1.1 声音传感器阵列	31
6.1.2 NI ELVIS	33
6.2 软件部分	35
6.3 仪器成本	37
七、 实验数据处理与结果	37
7.1 实验流程	37
7.1.1 收集信号	37
7.1.2 信号处理	38

7.2 数据处理.....39

7.3 误差分析.....39

 7.3.1 麦克风采样精度.....39

 7.3.2 采样频率.....40

 7.3.3 声速误差.....40

 7.3.4 计算误差.....40

八、 工作总结与展望.....40

 8.1 工作总结.....40

 8.2 研究展望.....40

九、 参考文献.....42

图形索引

图 1 五元阵装置模型图.....	9
图 2 XOS 平面.....	10
图 3 根据时间差估计声源位置流程图.....	12
图 4 信号端点检测示意.....	14
图 5 加窗分帧示意.....	15
图 6 声音传播模型.....	16
图 7 GCC 算法流程	19
图 8 尖叫和掌声时域频域图.....	20
图 9 三种加权函数的互相关函数(SNR=5, 尖叫).....	21
图 10 三种加权函数的互相关函数(SNR=-5, 尖叫).....	21
图 11 三种加权函数的互相关函数(SNR=-15, 尖叫).....	22
图 12 两种加权函数的互相关函数(SNR=5, 掌声).....	23
图 13 两种加权函数的互相关(SNR=-15, 掌声).....	24
图 14 偏差随距离增加变化图.....	26
图 15 相对偏差随采样频率变化图.....	28
图 16 不同传感器距离下的相对偏差随采样频率变化图.....	30
图 17 实验总体物理模型图.....	31
图 18 五通道的麦克风传感器阵列.....	32
图 19 麦克风传感器连接细节.....	32
图 20 麦克风电子积木.....	33
图 21 ElvisIII 面板简介.....	33
图 22 ElvisIII AI 模拟输入底层逻辑.....	34
图 23 ElvisIII 模拟输入 8 通道的标定曲线.....	34
图 24 ElvisIII 支持麦克风电路 1.0 版.....	35
图 25 LabVIEW 程序框图	35
图 26 SCOT 广义互相关子 vi 程序框图	36
图 27 重构算法子 vi 程序框图	36
图 28 收集信号流程图.....	38
图 29 信号处理模块.....	38

图表索引

表 1 加权函数表..... 19

表 2 声源坐标仿真数据..... 25

表 3 相对偏差随采样频率变化表..... 27

表 4 不同传感器距离下相对偏差随频率变化..... 28

表 5 实验器材成本..... 37

表 6 实验结果..... 39

表 7 实验数据计算偏差表..... 39

一、前言

数十年来，智能系统领域大多集中在视频图像处理和分析上，而随着时代的发展，其局限性逐渐凸显出来：在一些有障碍物遮挡的领域，视频监控无法准确得到其位置。而声音作为一种波，则可以轻易穿过障碍物，因此通过声音监控定位越来越为人们所重视。声音定位结合数字信号处理技术完美地利用了声波信号作为视觉信号的补充，在声电学、信号与信息处理、通信系统以及软件工程设计等诸多领域有着重要作用，具体应用场景为：武器的精确制导，视频实时跟拍，人机语音交互，医学助听等领域。

声音定位是通过麦克风收集声音，通过提取信号中所需要的信息，基于一定的算法将声源的位置确定的技术。单个麦克风定位的实现比较困难，仅对环境噪声较小、声源位置较近且固定时有较好的效果，然而现实中很难达到这样理想化的要求，在这种情况下，麦克风阵列技术逐渐发展起来。顾名思义，麦克风阵列技术是一系列麦克风组成一定形状的拓扑结构，麦克风阵列可以通过不同麦克风之间信号的差异和组合以确定声源的位置，其准确率较高。20 世纪 80 年代，Flanagan 将麦克风阵列引入到大型会议的语音增强应用中，之后 Silverman 和 Brandstein 又将其应用到语音识别和声源定位中，近年来麦克风阵列技术已经成为一个研究热点。

根据麦克风阵列技术对声源定位的算法原理，可以有多种方法，本实验中我们采用基于时间差估计的定位技术。

我们的工作共分为 5 个章节，第二章我们介绍麦克风阵列定位技术原理，第三章是我们声音定位的模型，第四章我们介绍了声音传播基本模型，同时介绍了基于互相关函数计算时延的方法原理及仿真实验，并对各种情形进行了简单的比较，第五章是我们对整个实验过程的仿真，在第六章我们介绍了我们实验装置，并阐述了我们实验的过程，第七章我们对做过的实验数据进行处理，分析误差的传递，并验证了我们实验的可行性，最后一章我们对所做的工作进行了总结，并对以后的工作进行了展望。

二、麦克风阵列声音定位技术原理

本章节首先列举了目前主流的声音定位方法及各自优缺点，最后确定使用基于到达时间差差(Time Difference of Arrival, TDOA)估计的定位方法。

2.1 基于麦克风阵列的声音定位技术

在生活中，我们通过人耳可以简单辨别声音的方位，即所谓的“听声辩位”，原理就是通过辨别声音到达双耳的差异进行辩位，事实上，这就是麦克风点阵技术的基本原理：将麦克风阵列做成一定的形状，以不同麦克风收到声源信息的差异性，通过一定的算法确定声源的位置。依据麦克风对声源信息收集以及算法原理的不同，可以大致分为 3 类，分别为：基于

最大输出功率的可控波束形成技术(Beamforming)、高分辨率谱估计技术、和基于时间差估计(TDOA)的定位技术。

① 基于最大输出功率的可控波束形成技术

可控波束形成技术是以波束形成原理为基础, 首先对麦克风点阵进行建模, 以声源传播到各麦克风探测器的信号为基础, 对各麦克风信号波束进行加权求和, 得到总波束。之后在空间中搜索声源的位置, 使得总波束功率达到最大输出功率, 此时搜索的声源位置点就是声源位置, 这种既可以在频域中使用, 也可以在时域中使用。

② 高分辨率谱估计技术

高分辨率谱估计技术是计算麦克风阵列的空间矩阵和声源的空间谱矩阵来确定方位角, 进而确定声源的位置。具体包括 MUSIC 算法等。高分辨率谱估计技术适用于多个声源的情形, 要求环境噪声尽量平稳, 同时要求声源时不变, 因此在生活中很难实现。

高分辨率谱估计技术计算量大, 因此在定位系统中并不常见。

③ 基于时间差估计(TDOA)的定位技术

基于 TDOA 定位技术是根据声音到达各麦克风的时间差来进行定位。具体包含两个步骤, 首先通过麦克风阵列收到信号进行处理, 得到麦克风阵列之间发时间差矩阵, 其次, 根据时间差矩阵利用几何模型定出声源的位置。

基于 TDOA 定位技术是近年来最为常用的方法, 其原理简单定位精确度高, 而且定位的实时性好, 广泛用于生产生活中。但其也有缺点, 在多个声源定位问题上略显乏力, 而且有两个步骤存在误差传递导致误差放大的问题。

2.2 TDOA 定位技术的确定

基于最大输出功率的可控波束形成技术需要在全空间内搜索计算, 计算量大过程复杂, 而且有时可能会出现多个峰值的结果, 定位不准确; 高分辨谱估计技术对环境要求较高, 同时也要求声源时不变, 在生活中很难实现; 而基于时间差估计(TDOA)的定位技术原理简单, 而且定位准确度高, 且有很好的实时性。

因此在本实验中, 我们着重研究基于时间差估计(TDOA)的定位技术。

三、基于时间差估计的声音方位定位方法

在本章我们主要讨论在已知时间差的情况下得出声源位置的算法。我们列举了几种种常用的几种声音定位方法, 同时提出我们自己的定位方法。其次, 通过仿真实验验证理论的可行性, 并据此对误差进行分析。

3.1 五传感器算法

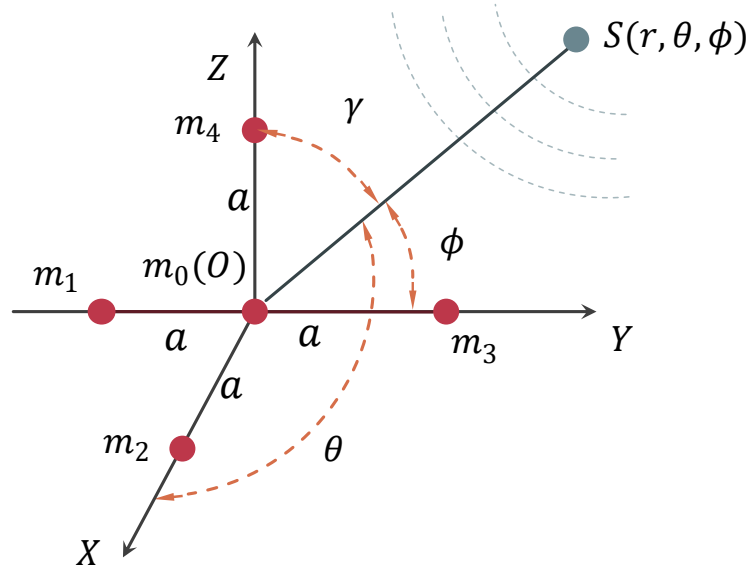


图 1 五元阵装置模型图

3.1.1 模型图像描述

声程差 $d = \tau * v$ ，其中 τ 为时延， v 为声速。声源 S 到坐标原点的距离 $r = |\mathbf{r}| = \sqrt{x^2 + y^2 + z^2}$ ，其中 (x, y, z) 为声源 S 在图 1 所示右手笛卡尔系的坐标， \mathbf{r} 表示原点指向声源 S 的向量。传感器 m_0 在原点处，传感器 m_1, m_2, m_3, m_4 到 m_0 的距离相等，记该距离为 a 。用 \mathbf{m}_1 表示原点指向传感器 m_1 的向量。

θ, ϕ, γ 分别为向量 \mathbf{r} 的三个方向余弦所对应的角，并有 $\theta, \phi, \gamma \in (0, \pi)$ 。记声程差 d_{01} 为声源 S 到 m_0 的距离与声源到 m_1 的距离之差，且有 $d_{01} = -d_{10}$ 。

3.1.2 模型公式

首先我们推导 r 的计算公式：

m_1 和 m_3 的坐标分别是 $(0, -a, 0)$ 和 $(0, a, 0)$ ，则

$$\sqrt{x^2 + (y+a)^2 + z^2} - \sqrt{x^2 + y^2 + z^2} = d_{10} \quad (3.1)$$

$$\sqrt{x^2 + (y-a)^2 + z^2} - \sqrt{x^2 + y^2 + z^2} = d_{30} \quad (3.2)$$

把 $r = \sqrt{x^2 + y^2 + z^2}$ 移到方程右边，平方得

$$2ya + a^2 = d_{10}^2 + 2rd_{10} \quad (3.3)$$

$$-2ya + a^2 = d_{30}^2 + 2rd_{30} \quad (3.4)$$

(3.3)与(3.4)相加得出

$$r = \frac{a^2 - \frac{d_{10}^2 + d_{30}^2}{2}}{d_{10} + d_{30}} \quad (3.5)$$

其次，我们取出 XOS 平面，见图 2 所示

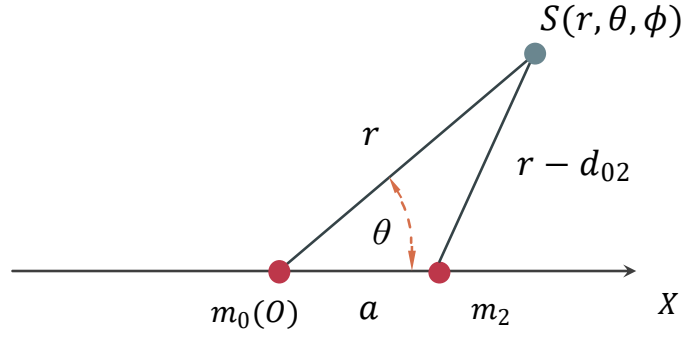


图 2 XOS 平面

根据余弦定理，有

$$\cos\theta = \frac{r^2 + a^2 - (r - d_{02})^2}{2ar} = \frac{d_{02}}{a} + \frac{a}{2r} - \frac{d_{02}^2}{2ar} \quad (3.6)$$

即

$$\theta = \arccos\left[\frac{r^2 + a^2 - (r - d_{02})^2}{2ar}\right] \quad (3.7)$$

且有近似

$$\theta = \arccos\left(\frac{d_{02}}{a}\right) (r \rightarrow \infty) \quad (3.8)$$

同理得

$$\cos\phi = \frac{r^2 + a^2 - (r - d_{03})^2}{2ar} = \frac{d_{03}}{a} + \frac{a}{2r} - \frac{d_{03}^2}{2ar} \quad (3.9)$$

$$\phi = \arccos\left[\frac{r^2 + a^2 - (r - d_{03})^2}{2ar}\right] \quad (3.10)$$

$$\cos\gamma = \frac{r^2 + a^2 - (r - d_{04})^2}{2ar} = \frac{d_{04}}{a} + \frac{a}{2r} - \frac{d_{04}^2}{2ar} \quad (3.11)$$

$$\gamma = \arccos\left[\frac{r^2 + a^2 - (r - d_{04})^2}{2ar}\right] \quad (3.12)$$

把(3.5)解出的径向距离 r 待回(3.7)(3.10)(3.12)即可得到声源的角向位置 (θ, ϕ, γ) ，由坐标变换

$$x = r \cos\theta, y = r \cos\phi, z = r \cos\gamma \quad (3.13)$$

可得到声源的右手笛卡尔系坐标 (x, y, z) 。

至此，五传感器模型的算法实现了全空间上声源位置的无近似解析解。

3.2 从时延估计开始的误差传递

通过仿真实验，我们得出了采样率越高误差越小的结论。实验误差主要来自于两方面： a 这个直接测量量和 τ 这个间接测量量。仿真实验中只能考察互相关算法带来的误差， a 没有误差，故时延误差 $\Delta\tau$ 与采样率有直接关系，写为

$$\Delta\tau_{0i} = g(f_s) \quad (3.14)$$

其中 f_s 是采样率，且采样率越高 $\Delta\tau_{0i}$ 越小，时延曲线近似指数衰减（未拟合）。声程差的误差为：

$$\Delta d_{0i} = v * \Delta\tau_{0i} \quad (3.15)$$

由于 a 的测量可以达到一个相当高的精度，其引起的实验结果误差相较于 Δd_{0i} 引起的误差可以忽略不计，故可以假设传感器 m_1, m_2, m_3, m_4 到 m_0 的距离 a 为一定值，然后考察 $\Delta\tau_{0i}$ 对定位误差的影响，以获得参数调整的指导和解析误差计算的准备。

由(3.5)得 $r = r(d_{01}, d_{03})$ ，绝对误差传递公式为：

$$\Delta r = \left| \frac{\partial r}{\partial d_{01}} \Delta d_{01} \right| + \left| \frac{\partial r}{\partial d_{03}} \Delta d_{03} \right| \quad (3.16)$$

由(3.5)和(3.16)，得

$$\Delta r = \left| -\frac{d_{10} + r}{d_{10} + d_{30}} \Delta d_{10} \right| + \left| -\frac{d_{30} + r}{d_{10} + d_{30}} \Delta d_{30} \right| \quad (3.17)$$

由(3.7)及绝对误差传递公式

$$\Delta\theta = \left| \frac{\partial\theta}{\partial r} \Delta r \right| + \left| \frac{\partial\theta}{\partial d_{02}} \Delta d_{02} \right| \quad (3.18)$$

得

$$\Delta\theta = \left| \frac{a^2 - d_{02}^2}{2ar^2\sqrt{1 - \cos^2\theta}} \Delta r \right| + \left| \frac{d_{02} - r}{ar\sqrt{1 - \cos^2\theta}} \Delta d_{02} \right| \quad (3.19)$$

由(3.13)及绝对误差传递公式

$$\Delta x = \left| \frac{\partial x}{\partial r} \Delta r \right| + \left| \frac{\partial x}{\partial \theta} \Delta\theta \right| \quad (3.20)$$

得

$$\Delta x = |\Delta r * \cos\theta| + |\Delta\theta * r * \sin\theta| \quad (3.21)$$

3.3 装置大小对模型精度的影响

要考察装置大小（周围传感器到原点的间距）对实验结果误差的影响，主要依靠(3.17)(3.19)(3.21)式。

首先考察 a 的变化对 d_{0i} 的影响。由(3.9)和(3.13)式得

$$\frac{y}{r} = \frac{d_{03}}{a} + \frac{a}{2r} - \frac{d_{03}^2}{2ar} \quad (3.22)$$

进而

$$\frac{(d_{30} + r)^2}{r^2 - y^2} - \frac{(a - y)^2}{r^2 - y^2} = 1 \quad (3.23)$$

由该双曲线图像和性质, a 减小时, d_{0i} 也减小, 只是减小幅度略微小于 a 。考察(3.17)式,

a 减小时, d_{0i} 减小, $\left| \frac{d_{10} + r}{d_{10} + d_{30}} \right|$ 和 $\left| \frac{d_{30} + r}{d_{10} + d_{30}} \right|$ 增大, 则在固定的 $\Delta\tau_{0i}$ 下 Δr 增大。同样的, 对(3.19)式, $\left| \frac{d_{02} - r}{ar\sqrt{1 - \cos^2 \theta}} \right|$ 增大且数量级高于 $\left| \frac{a^2 - d_{02}^2}{2ar^2\sqrt{1 - \cos^2 \theta}} \right|$, 故在固定的 $\Delta\tau_{0i}$ 下 $\Delta\theta$ 也增大, 由(3.21)式 Δx 也就增大。

综上, 在 $\Delta\tau_{0i}$ 与 a 相互独立的情况下, a 的减小会导致实验结果误差的增大。故我们想要减小装置大小, 就要提高采样率以控制误差一定。

四、时间差估计方法

在上一章我们介绍了基于时间差的声源方位定位方法, 基于 TDOA 定位方法包括两个步骤, 即计算时间差和根据时间差确定声源位置, 流程见图 3 所示:

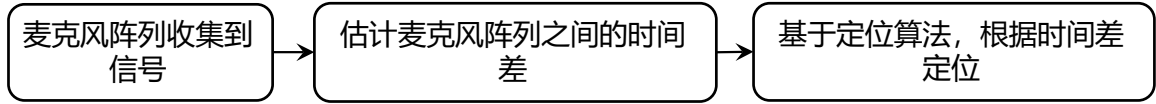


图 3 根据时间差估计声源位置流程图

本章我们讨论计算信号时间差估计的算法, 具体包括互相关法和最小均方自适应(Least Mean Square, LMS)法, 互相关法我们又讨论了基本互相关法和广义互相关(Generalized Cross Correlation, GCC)法, 在进行互相关运算之前要进行信号预处理, 包括带通滤波、降噪、信号端点检测和加窗分帧等步骤, 每一步骤我们均进行仿真实验, 并计算与理论的偏差, 确定较好的实验条件。

4.1 信号预处理

我们了解到真实世界的信号并不理想, 总是掺杂着各种各样的干扰, 包括白噪声和混响, 混响是因为在较小的房间中, 声音同过与各种障碍物的反弹形成的, 噪声和混响对我们进行实验估计时有很大的干扰, 因此在进行时间差估计前需要对信号进行一定的预处理, 较为有

效的手段包括带通滤波、降噪、信号端点检测和加窗分帧的手段，这些的目的是为了将我们不必要或不想要的噪声消除，以此达到改进实验精度的效果。^[1]

4.1.1 信号端点检测

由于目前的实验都是在传感器开始与结束采集之间播放一段短暂的声音用以定位，传感器采集到的数据总有一些没有记录任何有用的声源信息，如图所示，红色虚线之外的数据都是传感器未收到声源或是收到混响信息，如果能将其排除，便能有效降低噪音带来的误差，在信噪比较低的情况下效果更加明显。我们采用文献的语音端点监测方法，简述如下^[2]：

语音信号的监测通常有短时能量以及短时平均过零率两种判据。短时指将一段语音信号进行加窗分帧处理，每一扇窗代表一个“短时”，即

$$X_n(m) = w(m)x(n+m) \quad (4.1)$$

其中 n 、 m 为整数且 $0 \leq m \leq N-1$ ， N 为窗的长度， $n=1N, 2N \dots$ 表示帧数，通常选用汉明窗，即

$$w(n) = 0.54 - 0.46 \times \cos\left(\frac{2\pi n}{N-1}\right) \quad (4.2)$$

则第 n 帧的短时能量表示为

$$E_n = \sum_{m=0}^{N-1} X_n^2(m) \quad (4.3)$$

由于是信号的平方和，当该帧内含有语音信号，短时能量过大容易造成溢出，所以一般用短时平均能量 M_n 代替，即

$$M_n = \sum_{m=0}^{N-1} |X_n(m)| \quad (4.4)$$

第 n 帧的短时平均过零率表示为

$$Z_n = \frac{1}{2} \sum_{m=0}^{N-1} \left\{ \left| \text{sgn}[x(m)-T] - \text{sgn}[x(m-1)-T] \right| + \left| \text{sgn}[x(m)+T] - \text{sgn}[x(m-1)+T] \right| \right\} \quad (4.5)$$

由这两个判据便可以监测每帧是否有语音信号，我们采取了双门限-两级判决：

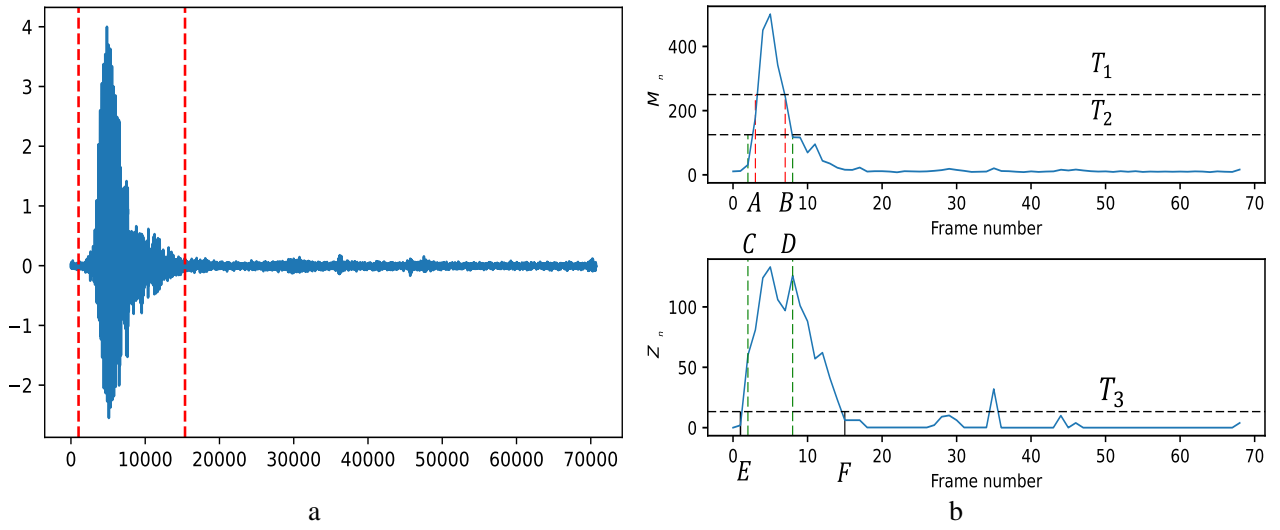


图 4 信号端点检测示意

如图 4 所示，在短时平均能量曲线上设置较高门限 T_1 和较低门限 T_2 ，A、B 点分别为曲线与 T_1 的交点，从 A 点往起点、从 B 点往终点，分别搜索曲线与 T_2 的交点 C、D 点，然后在短时平均过门限率曲线上从 C 点往起点、从 D 点往终点，分别搜索曲线与 T_3 的交点 E、F 点，这两点之间则是有语音的信号。

4.1.2 加窗分帧

为减小采样率导致的时延计算的误差，文献^[3]介绍了将信号等距分段，分别计算时延再取平均的方法。理论上真实的连续信号 $x(t)$ 被传感器离散采集成 $x(k\Delta t)$ 实际上做了这样的变换

$$x(k\Delta t) = \sum_{k=-\infty}^{\infty} x(t) \delta(t - k\Delta t) \quad (4.6)$$

将信号截断则是乘一个窗函数 $W(t)$

$$x_{\omega}(k\Delta t) = \sum_{k=0}^{\infty} x(t) \delta(t - k\Delta t) W(t) \quad (4.7)$$

如果分段仅仅是把信号截成多个小段，则窗函数是矩形窗

$$W(t) = \begin{cases} \frac{1}{T} & 0 \leq |t| \leq T \\ 0 & |t| > T \end{cases} \quad (4.8)$$

其中 T 为窗的宽度。

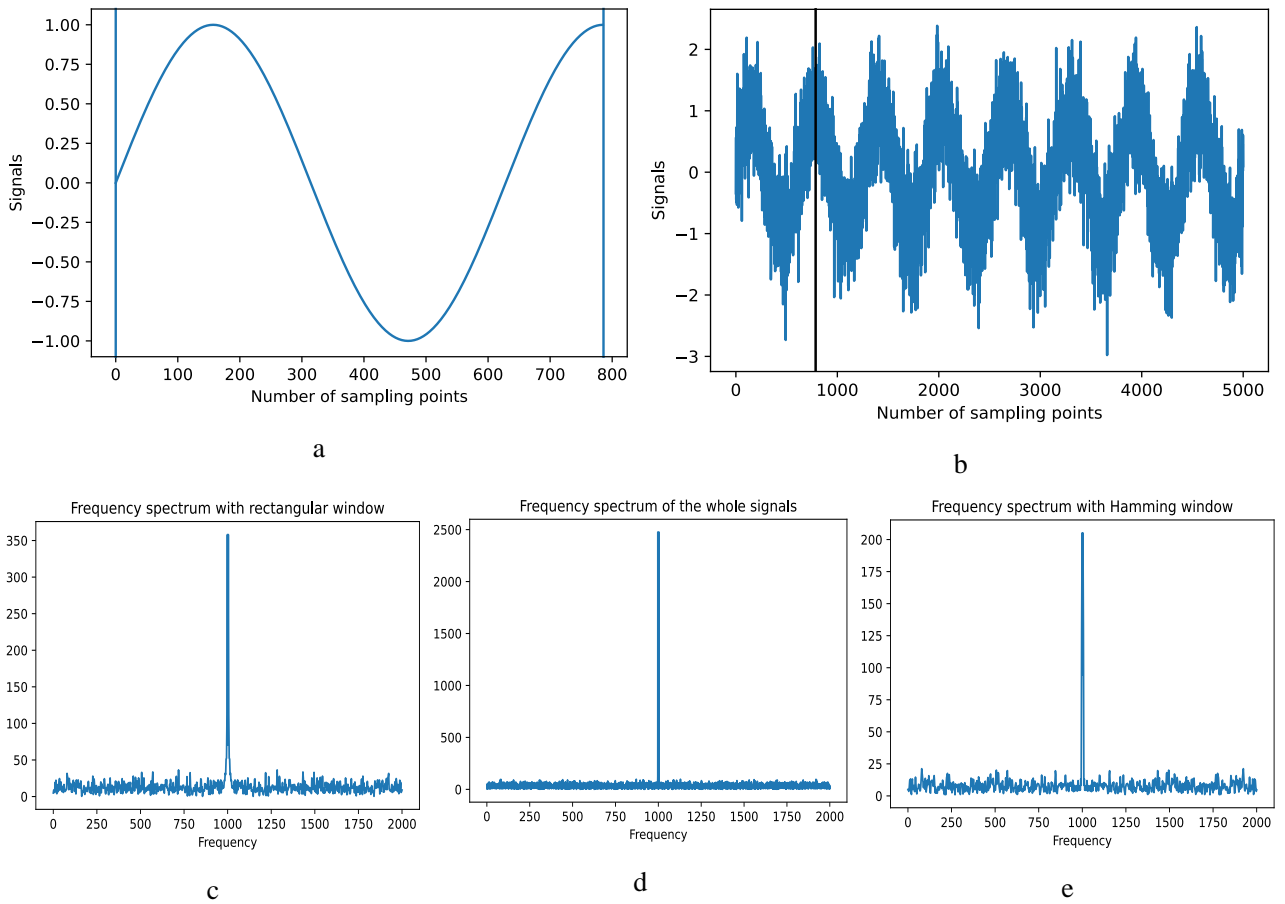


图 5 加窗分帧示意

这样的处理存在一定问题，如图 5.a，当截断宽度不是信号周期的整数倍，在该窗时域内有波形间断，再将该窗内的信号用于时延计算，则会有明显的边缘效应，以线性互相关函数为例说明，窗右侧“伪造”的补零信号与窗内信号没有连续性，将引起互相关函数计算的失真。从频谱上来看，加矩形窗后的信号的频谱 $X_{\omega}(f)$ 为原信号频谱 $X(f)$ 与窗频谱 $W(f)$ 的卷积，即

$$X_{\omega}(f) = X(f) * W(f) = \int_{-\infty}^{\infty} X(p)W(f-p)dp \quad (4.9)$$

造成的影响是，频谱 $X_{\omega}(f)$ 不再是原来的脉冲，而是出现了附件的频率分量，这种现象称为“泄漏”。举个例子，如图 5-b 加上高斯分布噪音的正弦信号，对整段作离散傅立叶变换，频谱是尖锐的脉冲，如图 5-d，用矩形窗截取图 5-b 黑色线左侧的信号，频谱的脉冲变宽，且附加上了次级频率分量，如图 5-c 换成汉明窗后，窗边缘的信号都趋于 0，边缘效应将减弱，这时发现频谱的脉冲变窄了，如图 5-e。

4.2 基于麦克风阵列的声音传播模型^[5]

首先阐述基于麦克风阵列的基本声音传播模型，以 M_0 和 M_1 指代麦克风点阵的基本单元， S 为声源位置，示意图见图 6 所示：

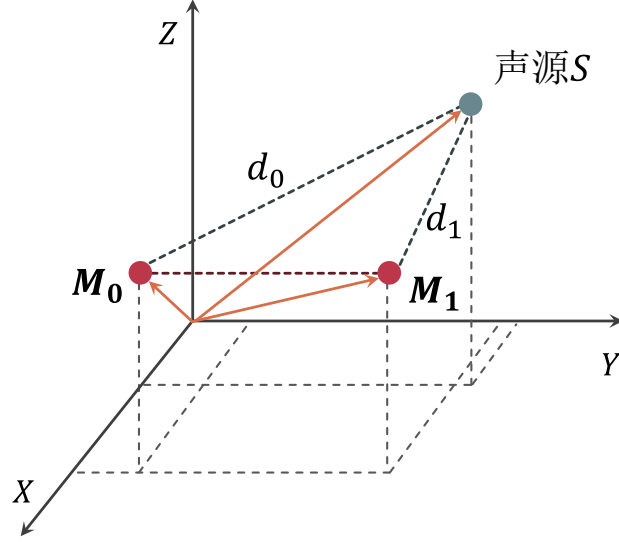


图 6 声音传播模型

如图，假设声源信号为 $s(t)$ ，设 M_0 、 M_1 、 S 为传感器、声源的位置，则有

$$d_0 = \|M_0 - S\| \quad (4.10)$$

$$d_1 = \|M_1 - S\| \quad (4.11)$$

设传感器收到的信号分别为 $x_0(t)$ 、 $x_1(t)$ 则有

$$x_0(t) = \alpha_0 s(t - \tau_0) + n_0(t) \quad (4.12)$$

$$x_1(t) = \alpha_1 s(t - \tau_1) + n_1(t) \quad (4.13)$$

其中 $\tau_0 = \frac{d_0}{v}$ ， $\tau_1 = \frac{d_1}{v}$ ， v 为声速， α_0 、 α_1 分别为声音衰减系数， n_0 、 n_1 分别为噪音信号，根据文献，噪音一般是高斯分布的。

根据上述模型，能够发现 d_0 、 d_1 的大小是有约束的：

令 $a = \|M_0 - M_1\|$ ，有三角形不等式

$$|d_0 - d_1| < a \quad (4.14)$$

所以后文计算时间差时，应当加上这一数学上的判据。

接下来计算声音衰减系数 α_0 、 α_1 ：

在此之前先定义分贝的概念，显然声音以波动形式传播，以单频信号为例，用三角函数表示有

$$s(t) = A_0 \cos(\mathbf{k} \cdot \mathbf{x} - \omega t + \phi_0) \quad (4.15)$$

其中 A_0 为声源的振幅, \mathbf{k} 为波矢量, 方向与声波传播方向相同, \mathbf{x} 为场点, ω 为声波的圆频率, ϕ_0 为声音的初相位。则描述声音大小的分贝为:

$$1dB = 10 \log_{10} \left(\frac{A}{A_0} \right)^2 = 20 \log_{10} \frac{A}{A_0} \quad (4.16)$$

根据国标给出的分贝衰减系数^[6]设常温为 25℃, 声音频率为 1000Hz, 相对湿度为 50%,

可得 $\alpha_{dB} = -5.68dB/km$ 。所以, $\alpha_{dB_i} = d_i \alpha_{dB}$ ($i=1, 2, 3 \dots$), 所以 $a_i = 10^{\frac{\alpha_{dB_i}}{20}} = 10^{\frac{d_i \alpha_{dB}}{20}}$ ($i=1, 2 \dots$)。

4.3 互相关分析的时间差估计法

互相关法是计算时间差的常用方法, 由最初的基本互相关, 到更为常用的广义互相关 (GCC), 都是利用了两个信号的相关性。本节我们介绍两种互相关算法并进行仿真以验证其合理性。

4.3.1 基本互相关算法

如图 1 所示, 有同一个声源 S 传播来的声音被两个麦克风 m_0 和 m_1 接受, 接收到的信号具有很强的相关性, 二者由于到声源的距离存在差值 $d_{01} = d_0 - d_1$, 先讨论两个传感器的情况。最基本的时间差计算方法是采用互相关函数来计算。

如果信号是连续的, 互相关函数表达如下:

$$R_{01}(\tau) = \int_{-\infty}^{+\infty} x_0(t) x_1(t + \tau) dt \quad (4.17)$$

但实际上传感器只能在有限时间, 而且离散采样, 所以采集到的信号为

$$x_0(n) = \alpha_0 s(n - \tau_0) + n_0(n) \quad (4.18)$$

$$x_1(n) = \alpha_1 s(n - \tau_1) + n_1(n) \quad (4.19)$$

则互相关函数应为

$$R_{01}(m) = E[x_0(n) x_1(n + m)] = \sum_{n=0}^N x_0(n) x_1(n + m) \quad (4.20)$$

其中 n 为 x_i 采集到的第 n 个数据, N 为采样点数, N 、 n 、 m 均为整数。

将式(4.18)和式(4.19)代入式(4.20), 得:

$$\begin{aligned} R_{01}(m) = & \alpha_0 \alpha_1 E[s(n - \tau_0) s(n + m - \tau_1)] + \alpha_0 E[s(n - \tau_0) w_1(n + m)] \\ & + \alpha_1 E[s(n - \tau_1 + m) w_0(n)] + E[w_0(n) w_1(n + m)] \end{aligned} \quad (4.21)$$

由于噪音是高斯分布, 所以噪音与信号不相关^[7], 所以后面三项为 0, 所以

$$R_{01}(m) = \alpha_0 \alpha_1 E[s(n - \tau_0) s(n + m - \tau_1)] \quad (4.22)$$

根据自相关函数的性质，当且仅当 $n - \tau_0 = n + m - \tau_1$ 即 $m = \tau_1 - \tau_0$ ，互相关函数取最大值，最大值就是时间差 $\tau_1 - \tau_0$ ，所以，到两个麦克风之间的时间延迟可表示为

$$\tau = \arg \max R_{01}(m) / Fs \quad (4.23)$$

式中， $\tau \in [-\tau_{\max}, \tau_{\max}]$ ， τ_{\max} 为可能的最大时间差值。

利用基本互相关计算时间差特点就是简单，计算量小，只需要计算两个信号的互相关函数，然后寻找峰值对应的横坐标就是时间差，峰值越尖锐说明互相关程度越好，时间差的计算结果也越准确。然而，这种方法的要求是噪声和噪声之间两两不相关，即它们之间的互相关函数为 0，这样的要求在实际应用中很难实现。在实际中，由于噪声和混响的存在，麦克风收到信号的互相关程度变低，导致峰值不明显，最终影响时间差的精度。为了降低噪声的影响，产生了广义互相关计算算法。

4.3.2 广义互相关(GCC)算法^[8]

基本的互相关函数可以根据时间差信号的相关性，峰值对应的时间差就是时间差。在现实中，由于噪声和混响的原因，基本互相关函数的相关性被弱化，这就给时间差估计带来很大的困难。在考虑噪声和混响的情况下，通常使用的是广义互相关函数代替基本的互相关进行计算。

互相关和卷积的关系为 $y_0(x)$ 和 $y_1(x)$ 的卷积为 $y_0(x)$ 和 $y_1(-x)$ 的互相关函数，而互相关运算和互功率谱运算互为傅里叶变换，据此，我们可以先计算两个函数的互功率谱，在互功率谱上进行一定的加权运算，使得我们想要的频率的峰值变大，削弱噪声和混响和峰值，之后在反变换回到时域中得到二者的互相关函数。

设两个信号分别为 $y_0(t)$ 和 $y_1(t)$ ，设二者的时间差为 τ ，有 $y_1(t) = y_0(t - \tau)$ ，如果 $y_0(t)$ 的傅里叶变换为 $Y_0(j\omega)$ ，则 $y_2(t)$ 的傅里叶变换为 $Y_1(j\omega) = Y_0(j\omega)e^{-j\omega\tau}$ ，用 $G_{01}(j\omega)$ 表示二者的互功率谱，有：

$$G_{01}(\omega) = Y_0(j\omega) \cdot Y_1(j\omega)^* \quad (4.24)$$

互相关函数可以表示为：

$$R_{01}(t) = \frac{1}{2\pi} \int \psi_{01}(\omega) G_{01}(\omega) e^{j\omega t} d\omega \quad (4.25)$$

其中 $\psi_{01}(\omega)$ 为两个信号在频域上的加权函数， $Y_1(j\omega)^*$ 为 $Y_1(j\omega)$ 的复共轭，这样一来时间差

$$\tau = \arg \max [R_{01}(t)] / Fs \quad (4.26)$$

GCC 算法流程图见图 7 所示：

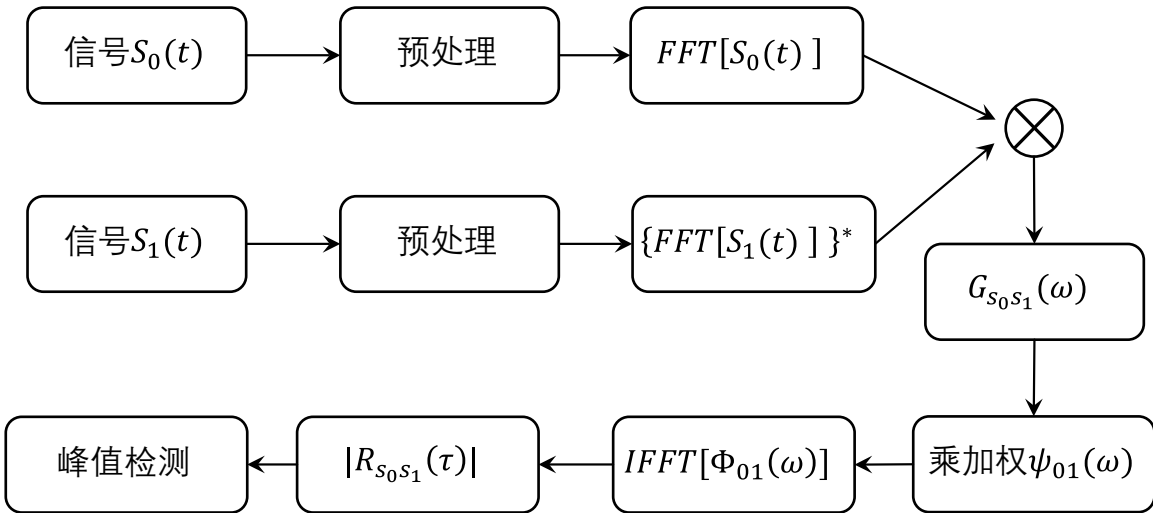


图 7 GCC 算法流程

图 7 中， $\{FFT[s_1(t)]\}^*$ 为信号 $s_1(t)$ 傅里叶变换后的共轭函数，整个流程为将信号 $s_0(t)$ 和信号 $s_1(t)$ 预处理之后，分别取傅里叶变换和傅里叶变换的共轭函数，之后将二者相乘得到信号的互功率谱函数 $G_{s_0s_1}(\omega)$ ，将互功率谱乘加权函数 $\psi_{01}(\omega)$ 得到广义互功率谱 $\Phi_{01}(\omega)$ ，对其做傅里叶反变换得到两个信号的互相关函数，最后进行峰值检测即可得到时间差。

加权函数可以有不同的类型，不同的加权函数适用于不同的条件，理想情况下是希望得到一个尖锐的峰，峰约尖锐说明相关性越好，也就是噪声和混响的干扰程度小。表 1 列出了常用的加权函数及其作用：

表 1 加权函数表

加权	表达式	特点
CC（基本互相关）	1	基本互相关，对环境噪声和混响较为敏感
PHAT（互功率谱相位）	$\frac{1}{ G_{01}(\omega) }$	相当于白化滤波，对大噪声有较好的结果
ROTH（处理器）	$\frac{1}{G_{11}(\omega)}$	等价于维纳滤波，能有效抑制噪声大的频带，但会使得相关函数的谱峰展宽
SCOT（平滑相干变换）	$\frac{1}{\sqrt{G_{00}(\omega)G_{11}(\omega)}}$	同时考虑两个通道的影响
ML（最大似然法）	$\frac{ \gamma(\omega) ^2}{ G_{01}(\omega) \cdot (1 - \gamma(\omega) ^2) }$	是统计意义上的一种最优滤波器，高信噪比加大权值，收环境噪声影响小

4.3.3 基于互相关函数的仿真模拟

针对上面的加权函数，选取基本互相关(CC)、PHAT、和 SCOT 平滑相干变换进行仿真模拟，为了保持与真实世界的联系性，我们分别选取了男性尖叫声和掌声等奇异信号，人的尖叫声带有一定呈现一定的周期性，即频率较为离散，而掌声则周期性较弱，频率更偏向连续，采样频率为 50000Hz，两种信号时域图和频域图见图 8 所示：

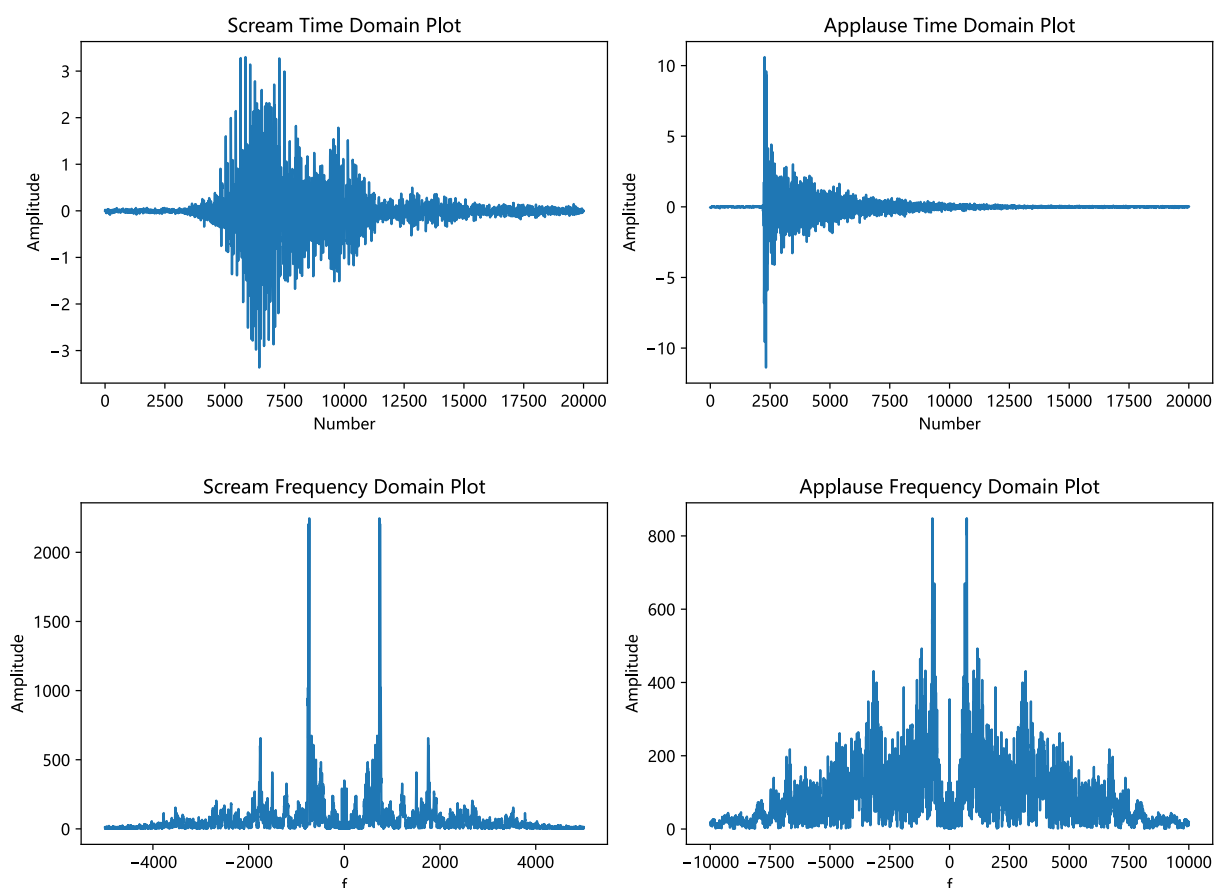


图 8 尖叫和掌声时域频域图

可以观察到人的尖叫声频域图中频谱跟偏向于离散，而掌声频域图中频谱更偏向于连续，也就是说，尖叫声的周期性更好，掌声的冲击性更好，这一点可以在时域图中看出。

模拟过程是我们手动将上述信号移位 100 个样品点，生成一个新的信号，在对原信号和新的信号进行互相关运算，加权函数分别选基本互相关(CC)、PHAT、和 SCOT 平滑相干变换通过控制不同的信噪比达到控制波形和相关程度的目的。

下面我们分别对不同信噪比下的不同加权函数的功能进行对比：

信噪比较大，取 $\text{SNR}=5$ 时，三种加权函数的互相关函数见图 9 所示：

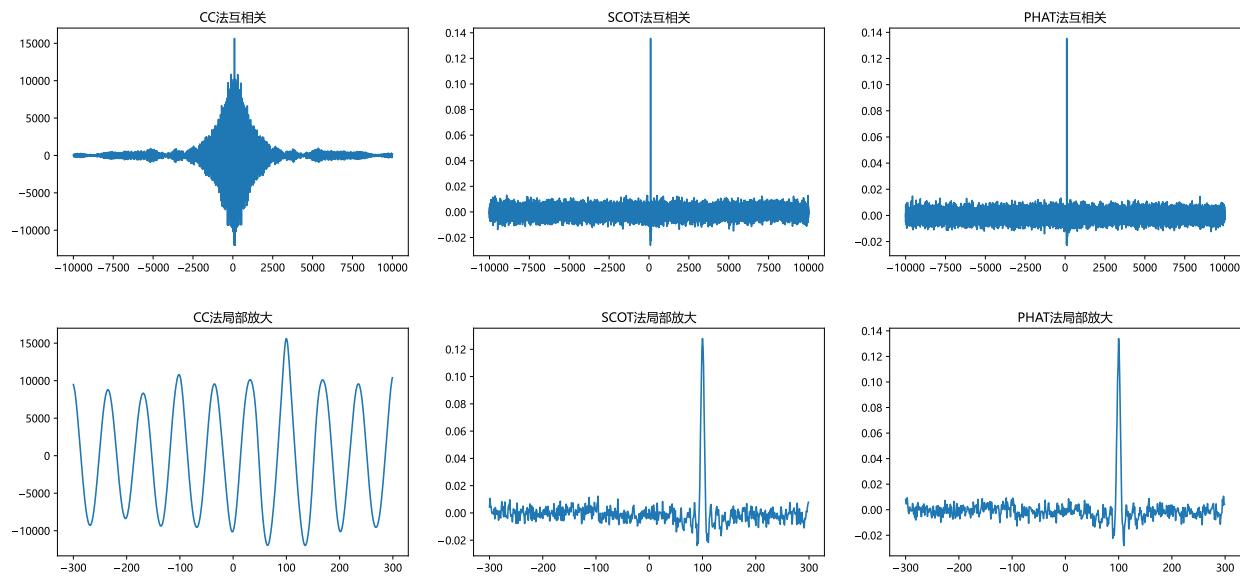


图 9 三种加权函数的互相关函数(SNR=5，尖叫)

信噪比一般时，取 $SNR=-5$ ，三种加权函数的互相关函数见图 10 所示：

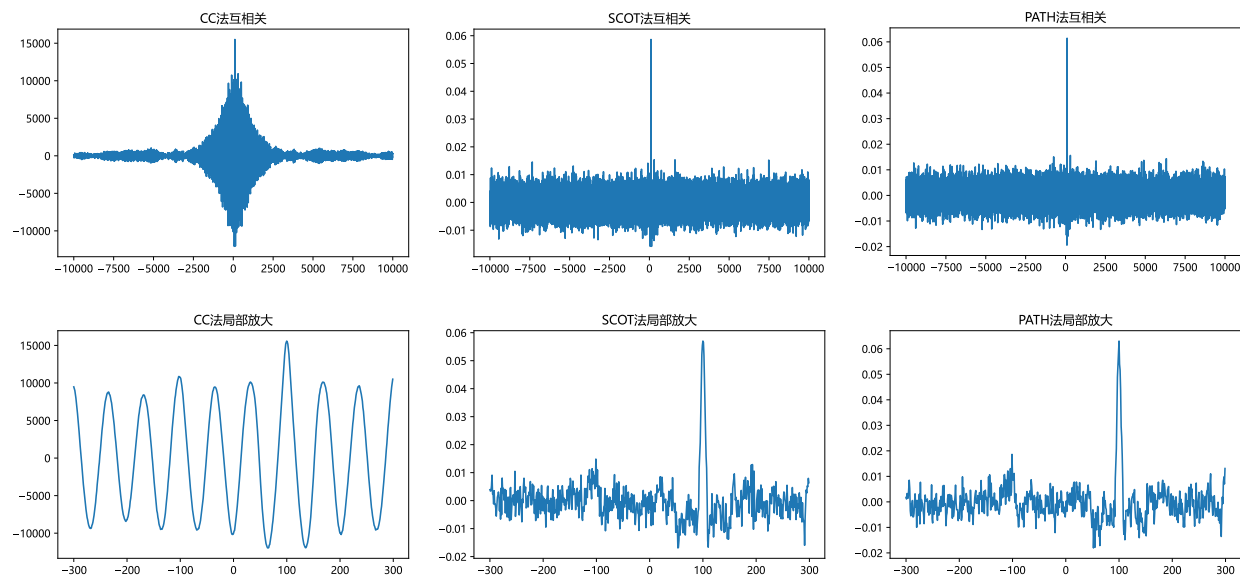


图 10 三种加权函数的互相关函数(SNR=-5，尖叫)

信噪比一般时，取 $SNR=-15$ ，三种加权函数的互相关函数见图 11 所示：

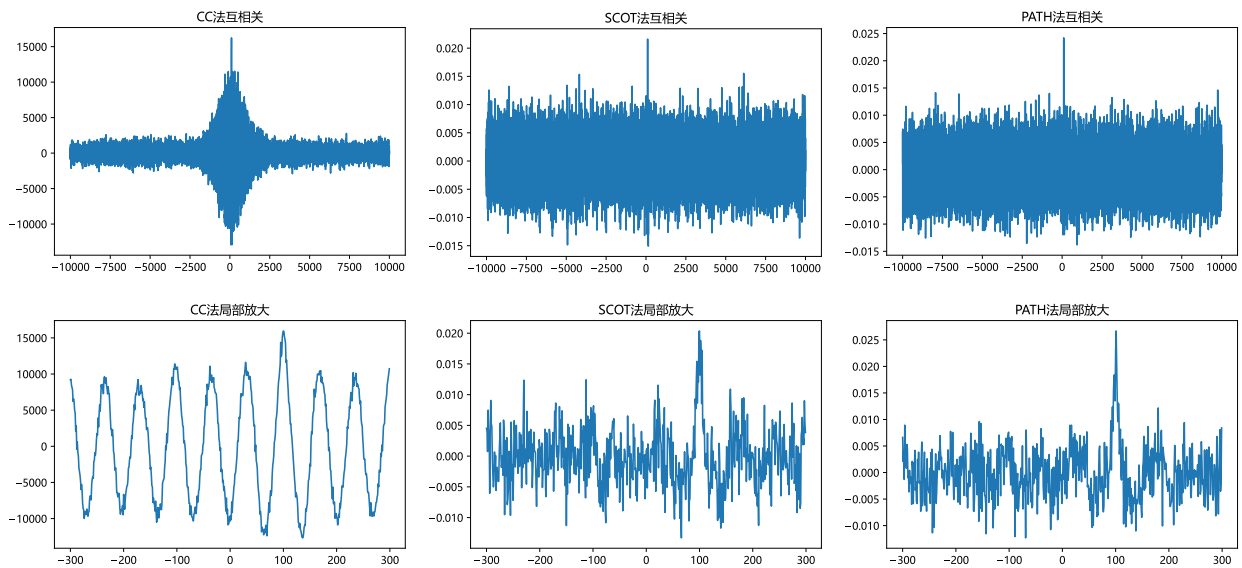


图 11 三种加权函数的互相关函数(SNR=-15, 尖叫)

通过对信号的加权和运算，我们得到了不同信噪比不同加权函数下的互相关函数，通过以上仿真，我们可以得到以下结论：

1) SCOT 法和 PATH 对互相关函数的影响近似，均是将互相关函数变为冲击函数 $a \cdot \delta(t - \tau_{12})$ 的形式，当 $t = \tau_{12}$ 时可以取到冲击函数的峰值，由于人的声音具有一定的周期性，因此未加权的基本互相关可以得到多个峰值，在最大峰取时间差。

2) 随着信噪比减小，即噪声的增大，三者的毛刺现象变多，峰值减弱，SNR=-15 时噪声已经很大，此时三个函数均出现 1~2 个样品点的误差，但 SCOT 法和 PHAT 法仍然出现峰值，基本的互相关函数也不再平滑，出现的毛刺使得时间差出现了 1~2 个样品点的误差。

3) 尽管 CC 基本互相关整个函数图像出现较好的“金字塔”形，但放大看，我们可以观察到其主峰和其他峰之间的高度差并不明显，而加权的广义互相关则只会出现一个峰，这就是广义互相关与基本互相关在类似周期信号的区别。

掌声的情形与人叫声不尽相同，掌声更偏向于冲激信号，周期性减弱而冲击性增强，特点是抵抗噪声的能力较强，这里我们分别取信噪比大和小的情形，由于 SCOT 和 PHAT 加权函数对互相关函数的波形影响类似，因此我们仅作 CC 基本互相关和 SCOT 法互相关的情形。

信噪比 SNR=5，两种加权函数的互相关函数见图 12 所示：

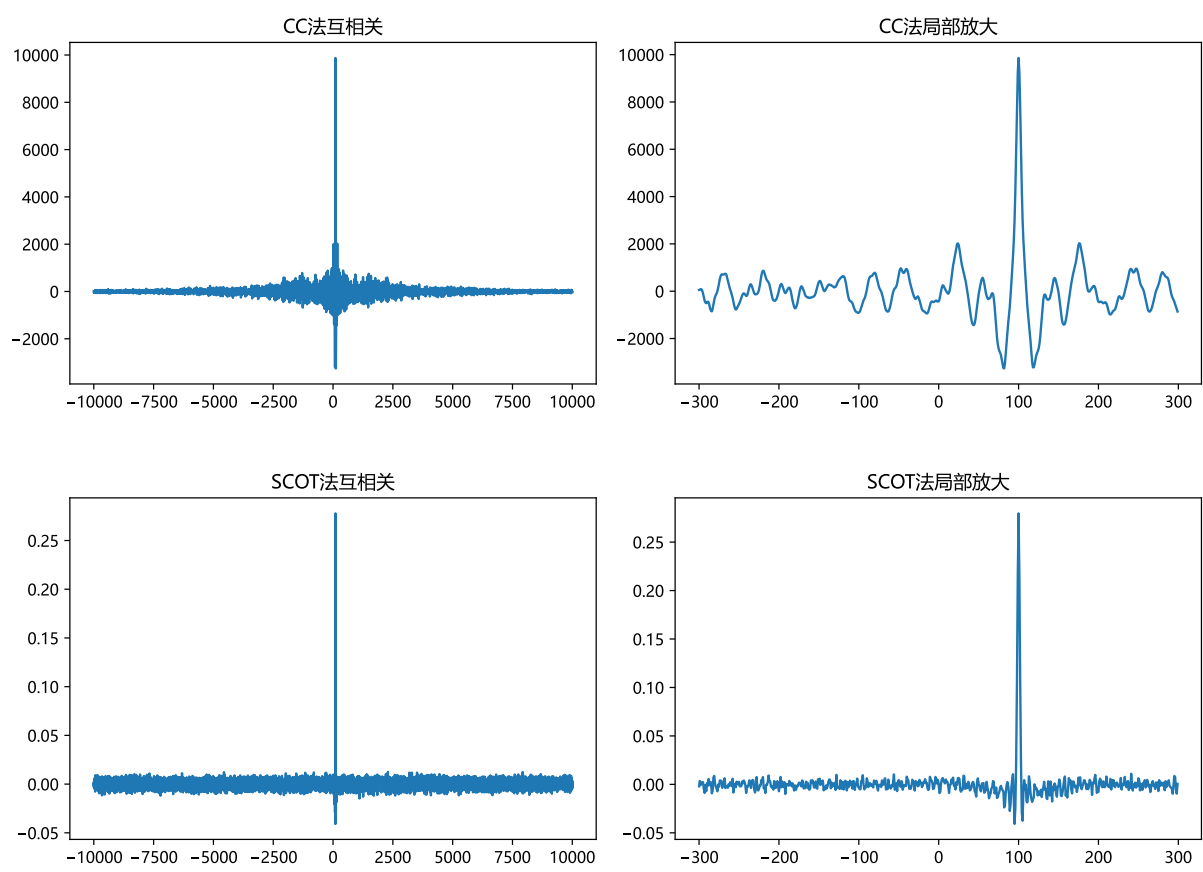


图 12 两种加权函数的互相关函数(SNR=5，掌声)

信噪比 SNR=-15，两种加权函数的互相关函数见图 13 所示：

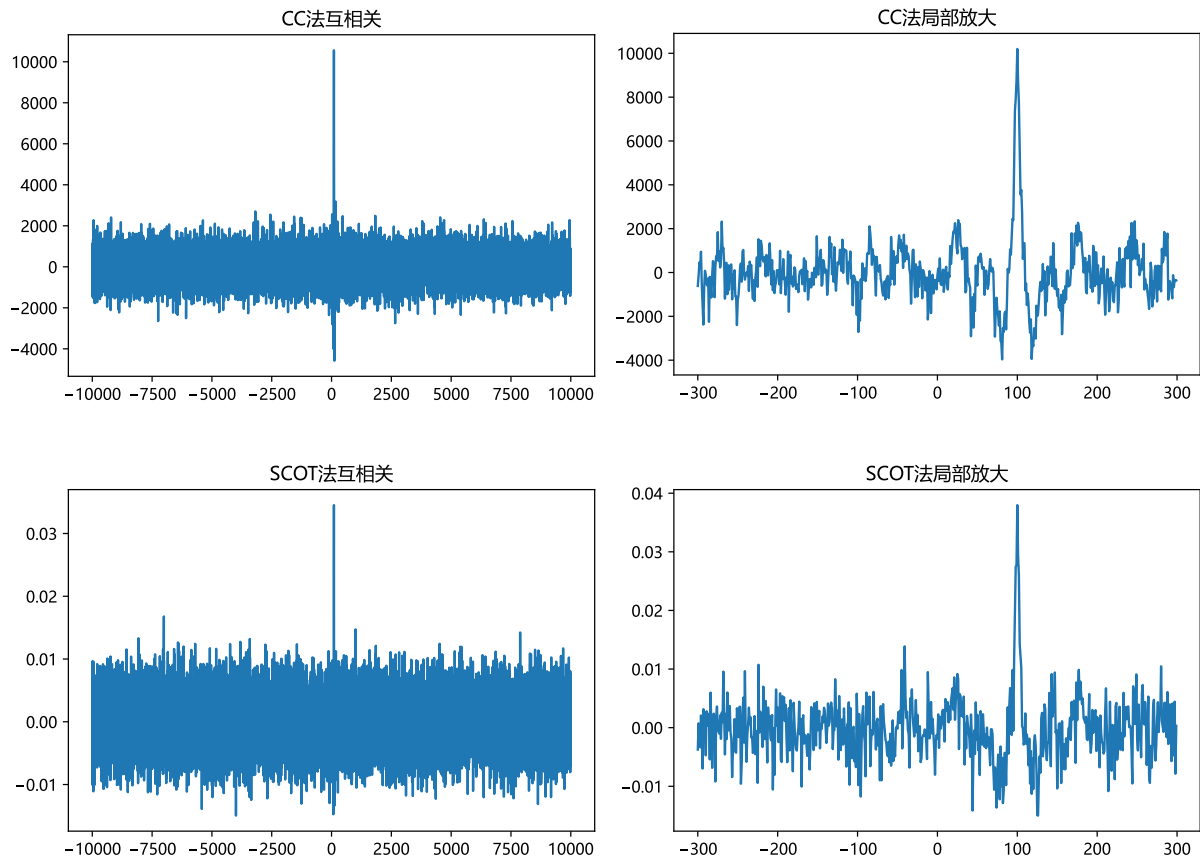


图 13 两种加权函数的互相关(SNR=-15, 掌声)

可以看到，对于冲击性较为明显的信号，CC 加权和 SCOT 加重的结果类似，互相关函数的形状也类似，也就是说，对于掌声等冲击性明显的信号，基本互相关已经可以得到较好的结果，加权法反而会对波形产生一定的变化，在后续实验时，我们应该将周期性明显的信号和冲击性明显的信号进行不同的处理。

五、仿真模拟

针对以上的实验步骤和方法，我们根据 4.2 的麦克风点阵声音传播模型进行整个实验流程的仿真模拟，同时根据仿真的结果得到误差的传递规律，同时探索得到较好的实验条件以及模型适用的条件。

使用的信号录制的成年男子的叫喊声，声音定位模型为五元定位，声音衰减系数取 $\alpha_{dB} = -5.68dB/km$ 。

误差来源主要为到原点距离、信噪比、声音频率以及采样频率，我们一次对这几个因素进行仿真实验，对一个仿真时需控制其他量不变。

5.1 声源坐标

信号取自我们录制的成年男子的叫喊声，传感器距离取 $a = 40cm$ ，采样频率为 $50000Hz$ ，不加入噪声，取不同声源坐标下计算仿真结果，结果如下：

表 2 声源坐标仿真数据

理论值(cm)				计算结果(cm)				绝对误差(cm)			
X	Y	Z	r	\hat{X}	\hat{Y}	\hat{Z}	\hat{r}	ΔX	ΔY	ΔZ	$ \mathbf{r} - \hat{\mathbf{r}} $
45	15	20	51.48	45.36	15.30	20.00	52.26	0.36	0.3	0	0.4685
75	45	50	100.75	76.24	45.34	51.64	102.31	1.24	0.34	1.64	2.0819
95	80	86	151.07	96.17	79.99	87.00	152.01	1.17	0.01	1	1.5426
110	110	130	202.73	98.21	98.21	117.87	181.92	11.79	11.79	12.13	20.6233
125	165	145	252.74	113.35	145.10	131.11	223.29	11.65	19.9	13.89	26.9199
155	220	125	303.46	163.73	248.11	131.29	326.76	8.73	28.11	6.29	21.9361
230	195	175	348.64	260.11	221.18	194.98	395.84	30.11	26.18	19.98	44.6227
290	200	190	400.25	251.01	172.36	166.65	345.83	38.99	27.64	23.35	53.1873
315	180	265	449.28	337.86	193.44	190.15	481.97	22.86	13.44	74.85	36.5478
350	160	290	481.87	366.11	169.46	306.94	511.2	16.11	9.46	16.94	25.2183
380	240	300	540.37	260.78	164.22	206.85	368.39	119.22	75.78	93.15	169.2146

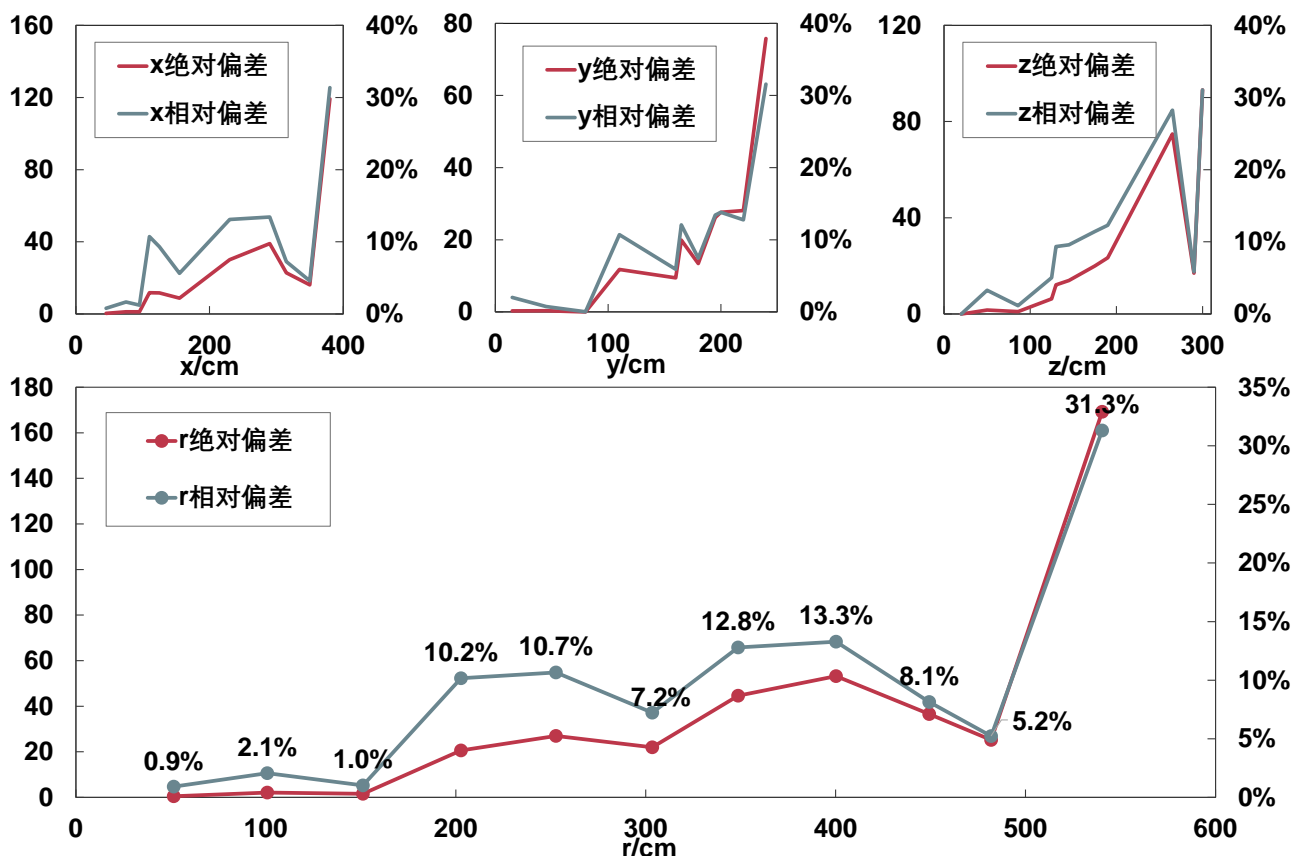


图 14 偏差随距离增加变化图

计算不同位置坐标下的相对偏差和绝对偏差，分别绘制 x 、 y 、 z 和 r 的偏差随距离增加的图像，见图 14 所示。

在图中我们可以明显的观察到，三个坐标增加时，绝对偏差和相对偏差均增加，虽然又是会出现下降的情况，但总体趋势是增加的； r 的偏差代表的是理论声源位置与计算得到声源位置之间的距离，并不是二者与原点距离之差， r 的偏差可以包含三个坐标的偏差，可以看到：

① 随距离的增加，相对偏差和绝对偏差均增加，在 500cm 以内，相对偏差基本在 15% 以内，在 150cm 以内，相对偏差在 3% 以下，可以认为此时定位基本准确。

② 随距离增加，偏差时增时减，这是不可避免的，这是由于由连续转为离散的不可预测性造成的，但总体趋势是增加，因此，定位算法有一定的距离限制。但是我们可以控制其他指标来增加定位算法的适用范围。

5.2 采样频率

采样频率决定的是由连续到离散的过程，理想情况下是采样率无限大此时系统误差为零，但是我们无法做到采样率无限大的情况，因此我们研究采样率多大可以达到我们的定位精度。

位置坐标选用的是上面位置中误差较大的两个点(290,200,190)以及(380,240,300)，用这两个点增加采样频率，探究采样频率对实验定位误差的影响。选用采样率从 50000Hz 依次递增，结果见下表所示：

表 3 相对偏差随采样频率变化表

频率 ($\times 10^4 \text{ Hz}$)	坐标(290,200,190), $r = 400.25$			坐标(380,240,300), $r = 540.37$		
	计算坐标	$ r - \hat{r} $	相对偏差	计算坐标	$ r - \hat{r} $	相对偏差
5	(251,172,167)	53.19	13.29%	(261,164,207)	169.21	31.31%
7.5	(289,199,189)	1.95	0.49%	(486,306,385)	150.81	27.91%
10	(276,192,182)	17.52	4.38%	(433,273,341)	74.78	13.84%
12.5	(286,196,188)	6.02	1.50%	(361,227,286)	27.01	5.00%
15	(267,185,177)	30.23	7.55%	(355,224,280)	36.14	6.69%
20	(293,203,193)	4.99	1.25%	(402,252,316)	29.42	5.44%
25	(285,196,186)	8.03	2.01%	(381,241,301)	1.91	0.35%
30	(278,192,182)	16.36	4.09%	(370,234,293)	13.38	2.48%
35	(302,208,197)	16.3	4.07%	(394,250,312)	20.8	3.85%
40	(295,203,193)	6.08	1.52%	(371,234,293)	13.14	2.43%
50	(298,205,195)	10.64	2.66%	(382,241,301)	2.38	0.44%
60	(294,200,190)	5.69	1.42%	(380,240,300)	0.5	0.09%
70	(287,198,188)	4.18	1.04%	(386,244,305)	9.39	1.74%
80	(290,200,190)	0.19	0.05%	(377,238,298)	3.64	0.67%
90	(290,200,290)	1.94	0.48%	(383,242,302)	4.08	0.76%
100	(287,198,188)	3.55	0.89%	(382,241,302)	2.69	0.50%

绘图，见图 15 所示

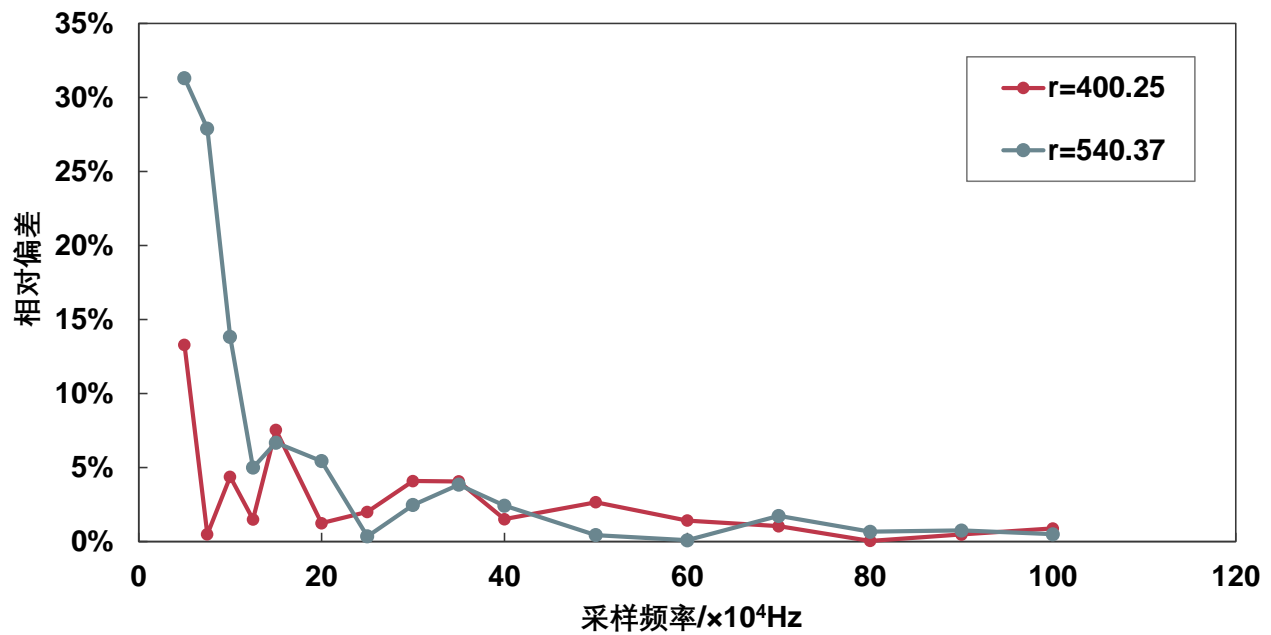


图 15 相对偏差随采样频率变化图

与我们预料的相同，随采样频率的增加，相对误差有明显的减小，在采样频率大于 $5 \times 10^5 \text{ Hz}$ 时，二者的相对偏差均在 5% 以下，可以认为达到了实验精度。因此，通过增加采样频率，可以有效扩大模型的适用范围。

5.3 传感器距离

传感器之间距离决定的是传播到不同传感器的时间差，时间差大则实验精度大，但是距离越大的话最终得到的装置尺度会更大，装置希望更小，这样就需要我们取到一个平衡。

选取采样频率从 $1 \times 10^5 \text{ Hz}$ 到 $2 \times 10^6 \text{ Hz}$ ，传感器距离选择 40cm、30cm 和 20cm，选择坐标为 (380,240,300) 的声源位置点，计算相对偏差和绝对偏差，结果见下表所示。

表 4 不同传感器距离下相对偏差随频率变化

频率(×10 ⁴ Hz)	a = 40cm		a = 30cm		a = 20cm	
	$ r - \hat{r} $	相对偏差	$ r - \hat{r} $	相对偏差	$ r - \hat{r} $	相对偏差
10	74.78	13.84%	150.08	27.77%	377.36	69.83%
15	36.14	6.69%	20.67	3.83%	77.25	14.30%
20	29.42	5.44%	53.59	9.92%	381.73	70.64%
25	1.91	0.35%	34.44	6.37%	36.42	6.74%
30	13.38	2.48%	21.91	4.05%	11.06	2.05%

35	20.8	3.85%	19.35	3.58%	104.62	19.36%
40	13.14	2.43%	21.92	4.06%	79.19	14.65%
50	2.38	0.44%	5.91	1.09%	27.5	5.09%
60	0.5	0.09%	1.07	0.20%	12.99	2.40%
70	9.39	1.74%	18.74	3.47%	47.92	8.87%
80	3.64	0.67%	21.01	3.89%	12.96	2.40%
90	4.08	0.76%	6.82	1.26%	13.08	2.42%
100	2.69	0.50%	6.32	1.17%	2.59	0.48%
110	1.27	0.24%	16.42	3.04%	23.4	4.33%
120	6.84	1.27%	10.34	1.91%	37.04	6.85%
130	1.55	0.29%	0.82	0.15%	18.64	3.45%
140	2.12	0.39%	2.06	0.38%	2.15	0.40%
150	2.41	0.45%	5.95	1.10%	13.39	2.48%
160	6.23	1.15%	13.09	2.42%	27.38	5.07%
170	0.3	0.06%	5.67	1.05%	17.82	3.30%
180	4.6	0.85%	1.07	0.20%	4.8	0.89%
190	0.94	0.17%	0.04	0.01%	7.16	1.33%
200	2.4	0.44%	5.94	1.10%	2.65	0.49%

将上面的图表绘图，见图 16 所示：

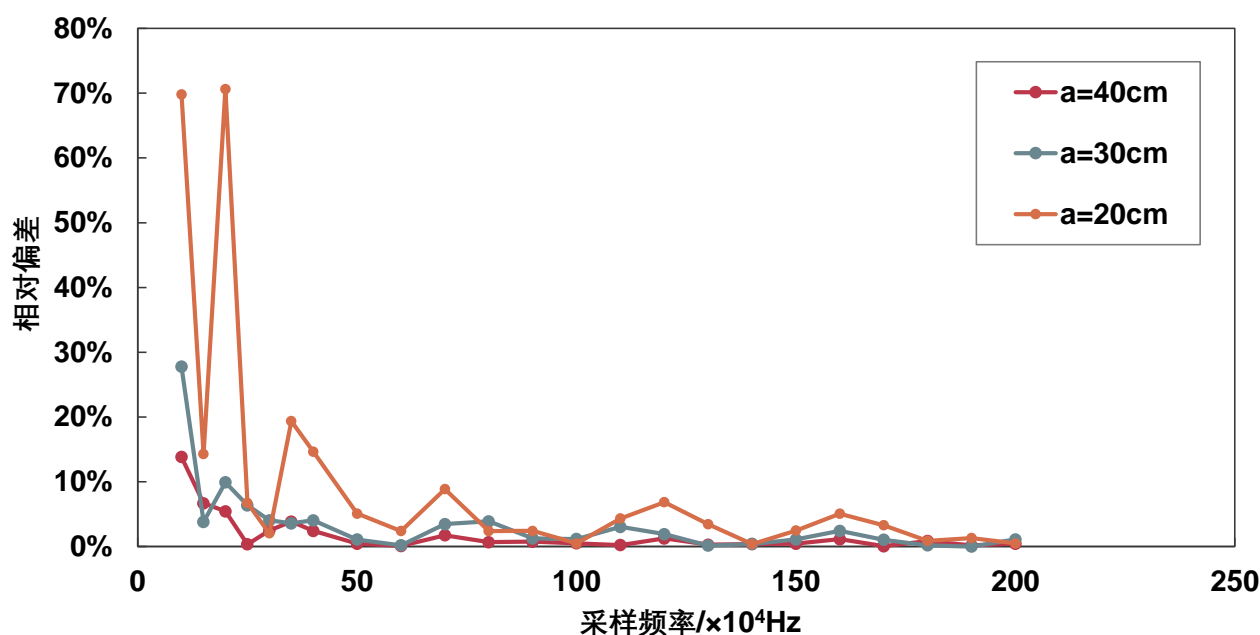


图 16 不同传感器距离下的相对偏差随采样频率变化图

三条曲线横向对比看，传感器距离小的相对偏差明显比传感器距离大的相对偏差大，验证了我们的猜想；纵向对比看，对于 $a = 20\text{cm}$ ，随采样频率，相对偏差呈现起伏上升和下降，但上升的幅度随采样频率的升高逐渐减小，当采样频率大于 $1.8 \times 10^6 \text{Hz}$ 之后，传感器距离 $a = 20\text{cm}$ 的相对偏差可以降至 5% 以下，采样率大于 $1 \times 10^6 \text{Hz}$ 之后，传感器距离 $a = 30\text{cm}$ 的相对偏差可以降到 5% 以下，若要再降低传感器距离，需要将采样频率增加到更高，就形成了技术上的困难。

进行实际实验时，需根据采样频率和定位声源范围的要求选择合适的传感器尺度。

5.4 声音频率

采样频率和声音频率存在一定关系，采样频率与声音频率之比越大，在一个声音周期内采集到的点的个数越大，这样越可以保留声音完整信号。同时我们还可以证明，声音频率和采样频率同时扩大相同的倍数，采集到的声音信号的波形不变，这样的话最后定位的位置不变，也就是说，改变声音频率或采样频率，实际改变的是采样频率与声音频率之比。

在声音不是单频信号时，用声音频谱图的最大频率代替作为声音频率。

仿真中原始信号的频率是 $5 \times 10^4 \text{Hz}$ ，在频谱图中，声音的最大频率为 5000Hz 左右，根据 5.2 中得到的数据，采样率与声音频率之比大于 100 时， $a = 40\text{cm}$ 传感器距离下可以实现精确定位， $a = 30\text{cm}$ 和 $a = 20\text{cm}$ 分别需要比为 200 和 360。

正式实验时，可以根据装置采样率和声音频率大致范围合理选择装置线度。

六、实验装置介绍

本章节我们对实验用到的装置和实验过程进行模拟，本实验通过 LabVIEW 调取实时模块 myRIO 与远程的设备 ElvisIII 设备连接，实现了实时声音定位，这套实验平台完全由我们团队自主研发和设计。实验物理模型图见图 17 所示：

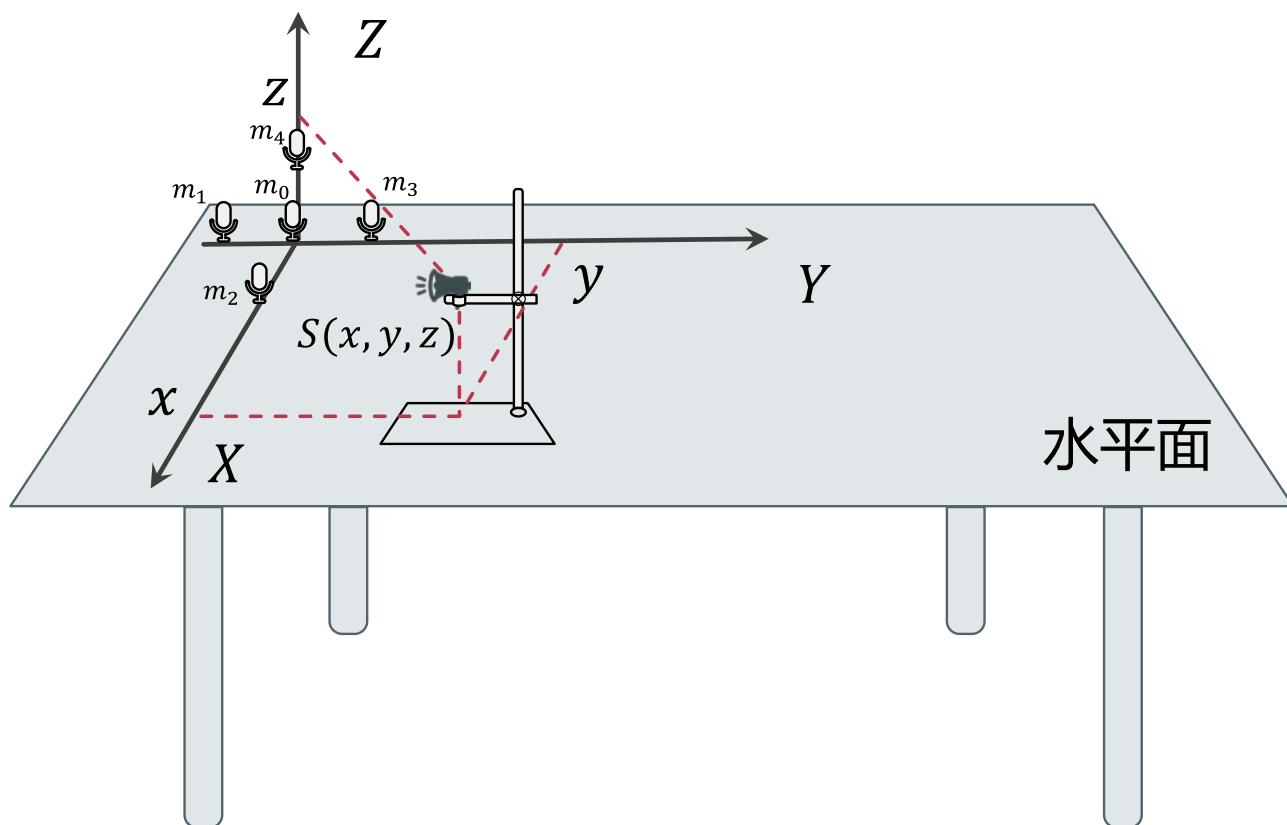


图 17 实验总体物理模型图

6.1 硬件部分

五通道的麦克风传感器阵列并由 ElvisIII 自带的面包板提供工作电路并采集模拟输入 (A/AI)，基本可以实现三维全空间的无死角定位。软件部分：多通道数据采集，预处理，互相关计算时延，重构算法，基本实现不间断采样和实时计算，返回计算所得声源位置三维空间中坐标。

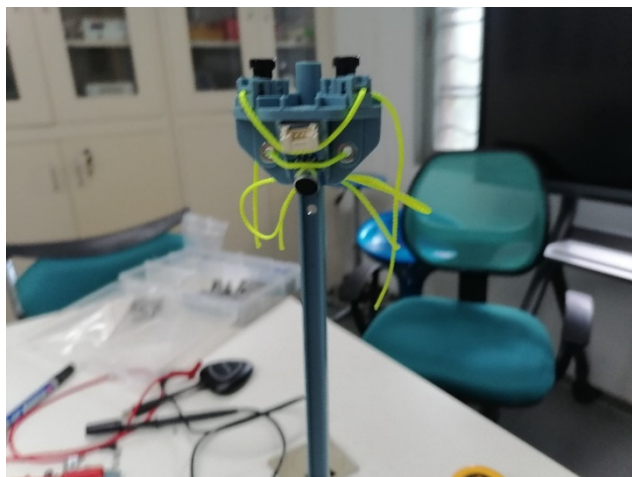
6.1.1 声音传感器阵列

本实验中使用五通道的麦克风传感器阵列接收声源端的信号，见图 18 所示。



图 18 五通道的麦克风传感器阵列

连接细节如下两图所示：（为了避免刚连接造成固体共振传声，我们的阵列用细绳的方式进行软连接以有效减震，同时我们传感器的四壁使用了相同的工件以保证中心传感器到达其他四个传感器的距离相同）。



a



b

图 19 麦克风传感器连接细节

实验中我们用到的麦克风型号及相关电气参数如下



规格

PCB 大小	33.0mm X 14.0mm X 1.6mm
工作电压	5V DC
操作电压	5V DC
兼容接口	2.54 间距的三针接口和 Grove 的四针接口 ^[1]

注 1：S 表示模拟输出端口，V 和 G 表示电源和地。

电气特性

参数	最小	典型	最大	单位
工作电压	4.5	5	5.5	VDC
模拟输出电压 (VCC=5V)	0	-	5	V
工作电流 (VCC=5V)	-	260	-	uA
频率范围	100	-	10000	Hz
灵敏度	-	-50	-	dB

图 20 麦克风电子积木

该声音传感器带有灵敏的麦克风，可显示 45 到 100 分贝范围内声波的波形，频率响应范围为 20-16000Hz，能过较好地适用于人耳能够听到的声音频率；信噪比大于 60dB；该款麦克风没有截断现象，可以记录偏离其工作电压±5V 内的全部信号；该款麦克风响应灵敏，可以记录 0.00001s 内拍手声音的完整信号；此外该款麦克风价格便宜。

6.1.2 NI ELVIS

NI ELVIS 是一个基于项目的学习解决方案，适用于工程基础、系统设计，它结合了仪器仪表、嵌入式设计、web 连接。为学生提供了一个全面的教学解决方案，包括模拟电路，机电一体化，电力电子，仪器仪表，数字通信，数字电子，控制等动手实验平台。

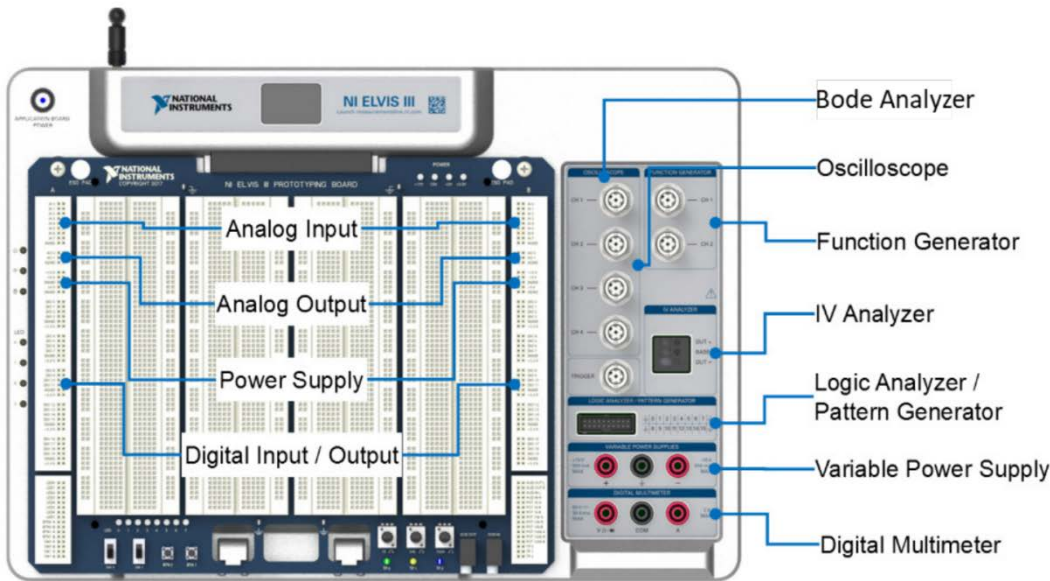


图 21 ElvisIII面板简介

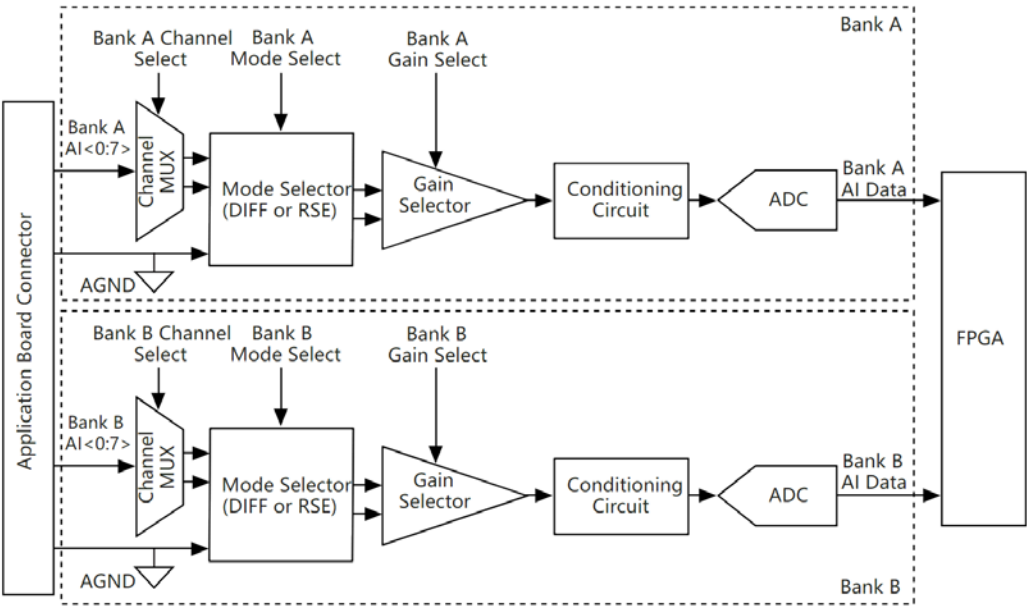


图 22 ElvisIII AI 模拟输入底层逻辑

我们主要使用 ElvisIIIAI 模拟输入通道，如上图所示 ElvisIIIAI 采用轮询的方式对不同通道进行采集，我们巧妙设计了标定实验，利用互相关算法，发现在一次轮询中第一通道 AI0 和第八通道 AI7 在采样率的情况下，之间因调整通道所带来的误差时间不超过一个采样点。

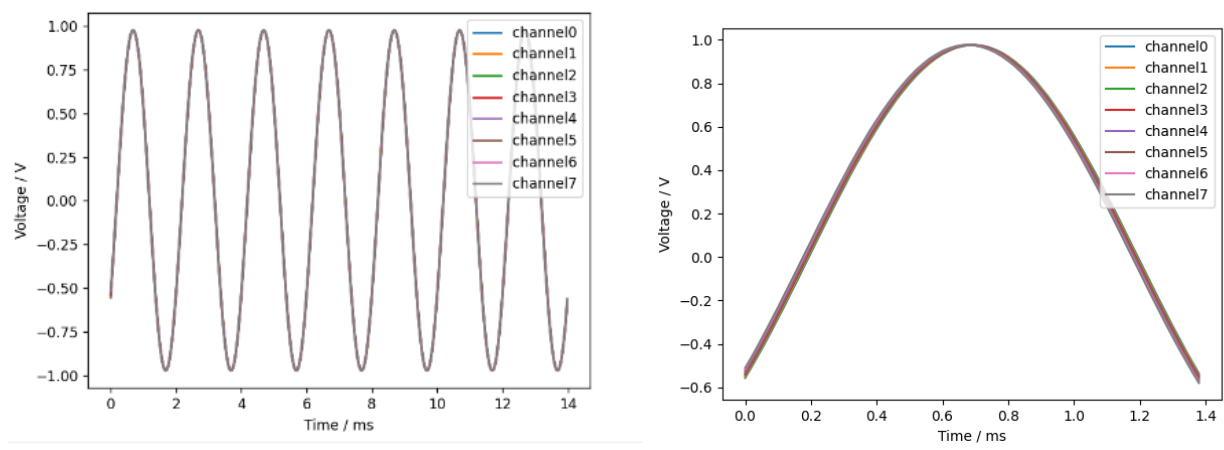


图 23 ElvisIII模拟输入 8 通道的标定曲线

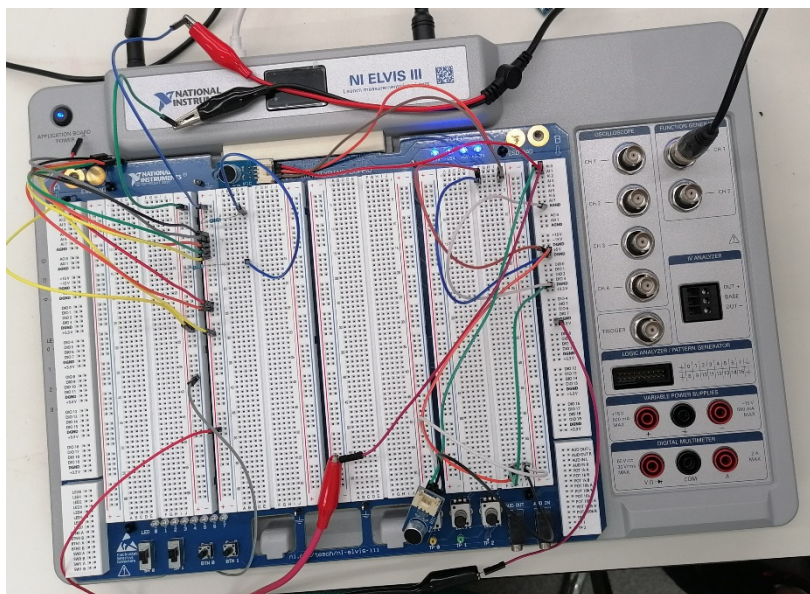


图 24 ElvisIII支持麦克风电路 1.0 版

6.2 软件部分

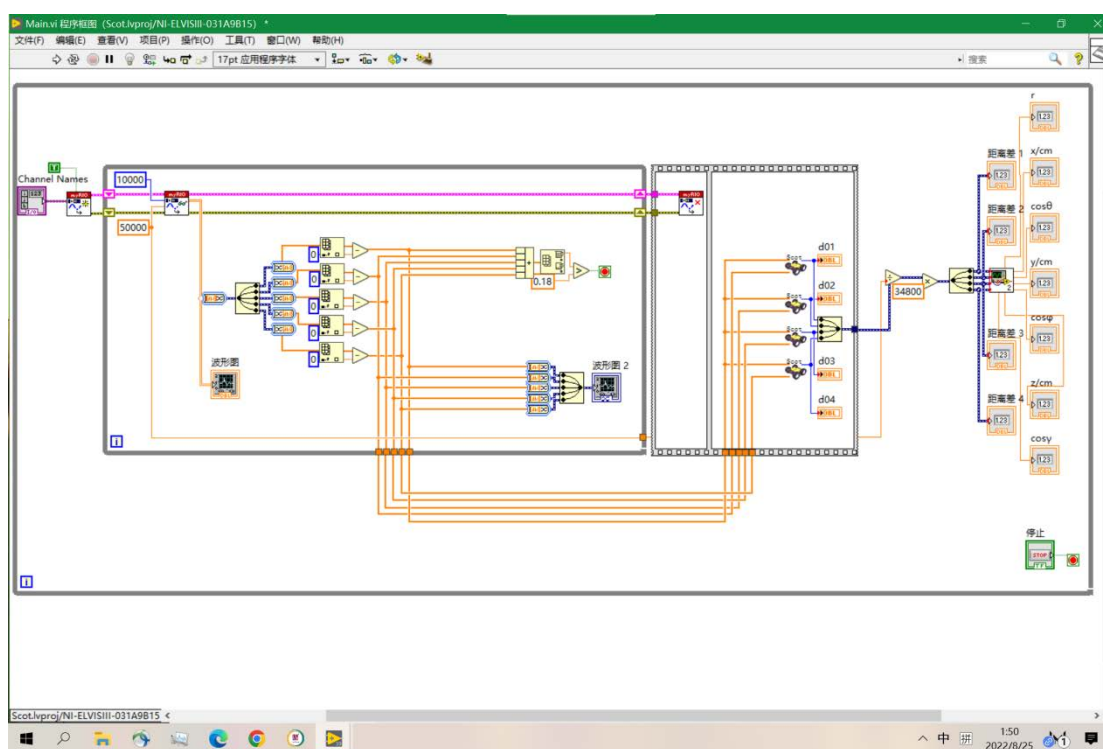


图 25 LabVIEW 程序框图

软件部分由使用 labVIEW 编程，包括了多通道数据采集，预处理，互相关计算时延，重构算法。

模拟输入通道使用单端输入的方式，如图 24，所有的 AI 通道共地；采样方式使用 N 采样，兼顾了实时性和数据流可计算性；数据的预处理将不同麦克风的静态工作电压扣除，有利于减少之后互相关计算的误差。对信号进行计算的触发信号为所有通道当前所采样之和达到电压阈值（如上图 0.18V）即跳出采样，进行之后的计算，计算时延我们采用了广义互相关中的 SCOT 平滑处理，我们将其封装为子具体的程序以及接口见附录重构算法我们将其封装为子 v i 具体的程序以及接口，最后可以得到计算所得的声源在三维空间的坐标。

SCOT 广义互相关子 vi 程序框图见图 26 所示

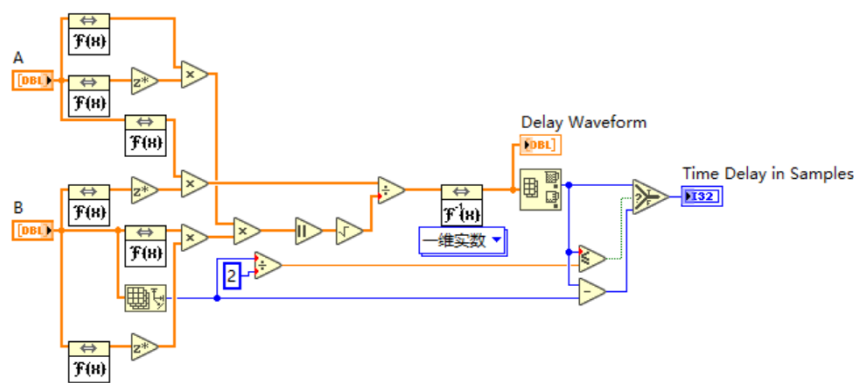


图 26 SCOT 广义互相关子 vi 程序框图

输入接口为两个 double 数组，输出接口是计算所得时延点数以及广义互功率谱图像，相关原理见实验原理部分。

根据四个时延对声源位置的重构算法子 vi 程序框图

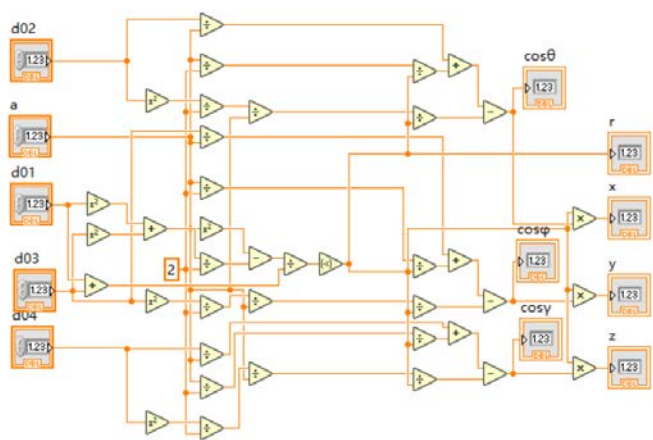


图 27 重构算法子 vi 程序框图

6.3 仪器成本

以上仪器成本价格见下表所示：

表 5 实验器材成本

仪器	单价/元	数量
NI	34800	1
麦克风	7	5
其他耗材	500	

总价为 35335 元。

七、实验数据处理与结果

本章节介绍我们根据以上算法和装置，进行实验，并对实验结果进行处理。首先介绍时延流程，包括收集信号和信号处理两个部分。之后我们阐述我们实验的参数和条件，并对得到的数据进行计算并于真实值比较，得出误差和相对偏差。最后我们对误差进行分析，分析误差的来源和误差的传递，并提出减小误差和改进实验的方法。

7.1 实验流程

按照前面所讲，实验流程包括收集信号和信号处理过程，本节按照这两个方面进行阐述。

7.1.1 收集信号

收集信号流程见图 28 所示。首先四个传感器收集到信号并提取，并在 250Hz~4500Hz 带通滤波，之后进行信号的检测，只有当短时能量和短时过零率大于阈值时才确定为有信号，其余均为未检测到信号，未检测到信号可以作为背景噪声供以后消噪使用。

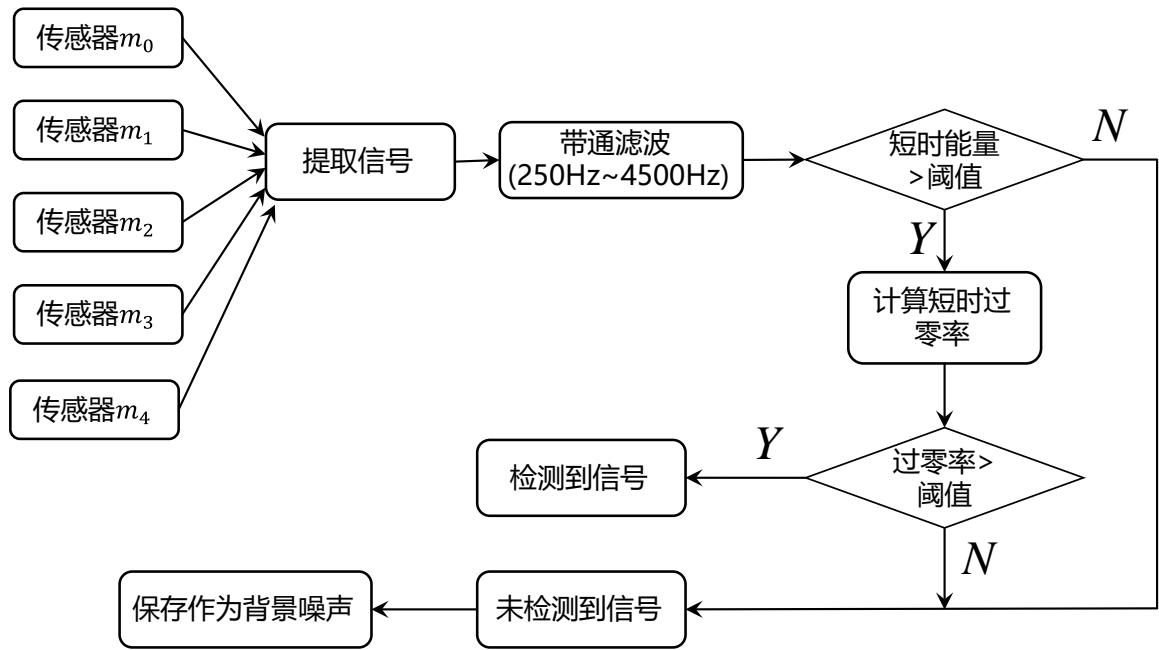


图 28 收集信号流程图

7.1.2 信号处理

根据前面所阐述的原理，信号的处理仅需要将前面的原理步骤组合到一起。如图 29 所示，确认检测到信号之后，通过半重叠 512 点汉明窗分帧、LMS 自适应降噪等步骤之后通过前面的互相关函数计算时延，得到四个时延值 τ_{01} 、 τ_{02} 、 τ_{03} 、 τ_{04} ，将三个时延值代入互相关算法，即可得到声源位置。

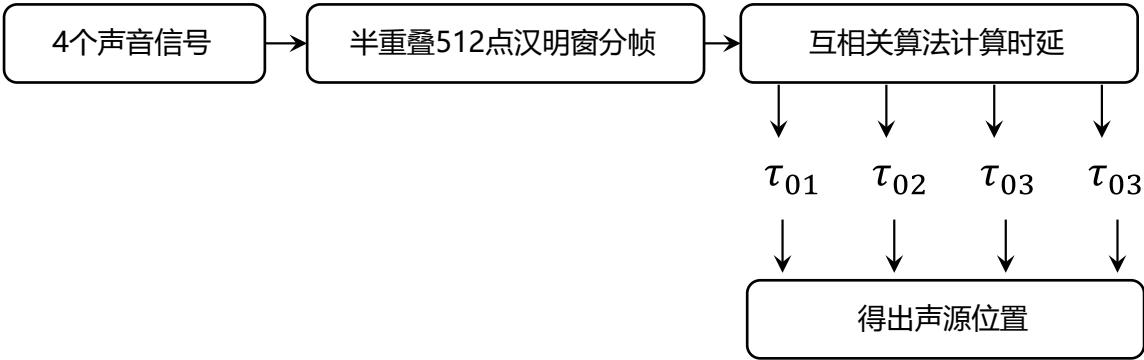


图 29 信号处理模块

7.2 数据处理

由于现有声音判别方法过于简单，只能对振幅达到设置的阈值的近场静止声源才能进行检测，我们实验数据不能完整覆盖全空间，且集中在近场。我们在 的情况下，围绕以麦克风建立的坐标系原点，做了三组近场声源的实验，得到的数据较为准确。

我们令采样率为 50kHz，声源采用人的掌声，由于每次拍掌发出的声音必然不同，对此声源能够定位将说明我们的定位系统对声源具有一定的鲁棒性，使用激光测距仪近似测量出声源的位置，得到的实验结果如下。

表 6 实验结果

理论坐标/cm	计算坐标/cm
(-18.9,-14.6,36.2)	(-26.8986,-23.5944,43.4256)
(51.3,-18.2,29.8)	(32.4997,-14.7417,36.2446)
(22.5,23.2,24.6)	(26.9645,16.5,25.4703)

计算得到的偏差如下：

表 7 实验数据计算偏差表

理论坐标/cm	计算坐标/cm	x 相对偏差	y 相对偏差	z 相对偏差	r 相对误差
(-18.9,-14.6,36.2)	(-26.9,-23.6,43.4)	42.3%	61.6%	20.0%	29.7%
(51.3,-18.2,29.8)	(32.5,-14.7,36.2)	36.6%	19.0%	21.6%	18.0%
(22.5,23.2,24.6)	(26.9,16.5,25.5)	19.8%	28.9%	3.54%	0.0479%

实验结果分析：

从实验结果来看我们的实验误差还有很大，分析主要原因如下：

①为了在视频中体现实时效果我们使用拍手这种方式作为声源，但是拍手的目标声源所占范围过大，造成了实验上的很大误差

②我们装置中的麦克风接收口是直径大约为 1cm 的圆盘，而我们理论上的重构算法将这些麦克风阵列全部视为点，这里需要我们理论进一步修正

7.3 误差分析^[9]

7.3.1 麦克风采样精度

本实验中使用麦克风并非完全一样的麦克风，我们不同的麦克风对声音的敏感度不同，这就导致两个信号的相关程度降低。麦克风是将声音信号转化到电信号的装置，但在本实验

中存在无法较零的问题，在信号为零时，麦克风收集到噪声信号对实验有干扰。同时在距离较远时，声压较低，收集到的信号相关程度进一步减弱，噪声在信号中占的比例太低，信噪比低。

7.3.2 采样频率

采样频率对实验误差也有很大影响，理论的连续系统下，时间差可以取到连续的值，而在实际情况下，我们只能取到离散的时间差，而在这离散的时间差中，很可能没有我们真实的时间差，尽管在时间差上相对误差很小，但是根据误差的传递，最后定位到声源位置时误差可能很大。例如，加入采样频率为 20000Hz，在声速 340m/s 的情况下，相邻数据点的时间差换算到距离差为 1.7cm，也就是计算得到的距离差值为 1.7cm 的整数倍，但是我们的 r 的尺度为 1m，这样根据误差传递关系，声源位置会有很大的偏差，因此，提高采样率能有效减小定位误差。

7.3.3 声速误差

声速受多方面的影响，如压强、温度、空气流动等因素，在我们实验中，我们默认声速为常温下的 340m/s，尽管后来添加了温度的影响，但也不能完全得到产地的准确声速，这就给实验带来误差。

7.3.4 计算误差

我们使用的 PC 机储存能力和计算能力有限，无法将数据完成的呈现，而且根据误差的传递，前面的由于很小的误差都可能对后面带来很大的误差，出现余弦值求解时更是如此，累计误差增大，最终导致实际偏理理论。

八、工作总结与展望

8.1 工作总结

通过对声源定位原理分析和实验验证，我们对麦克风点阵技术进行了深入研究，并设计完成了一套根据时延的定位方案，主要取得工作如下：

- ① 完成了对声源定位模型算法的两种麦克风阵列的推导和研究，并根据参数值得到了了解的情况，准确度较好。
- ② 研究和学习了基于 GCC 算法进行声源定位的流程，学习了计算时延的方法，并进行仿真模拟以验证算法的准确性。
- ③ 基于 NI Elvis III 和 LabVIEW 编程，利用低成本麦克风模块搭建了实时声音定位系统
- ④ 详细分析了各因素对本系统定位精度的影响。

8.2 研究展望

在取得阶段性成就的同时，我们也看到了我们实验的不足：①在实验过程中相邻传感器的距离为 33cm，导致用于定位的实验装置的尺寸较大，不满足日常生活的应用场景。②声

音位置的确定精度还有待提高，这说明我们的设备和算法都有很大改进的空间。③只能对振幅达到设置的阈值的近场静止声源才能进行检测，对远场或运动声源的定位结果并不理想。④目前我们是通过对传感器信号设置阈值来分辨声音信号，阈值的设定受麦克风的自身性质和环境因素的影响较大。

据此，笔者提出下一步工作的重难点：

① 现有的定位程序只会给出声源确定的直角坐标，下一步我们将继续优化声音定位程序的前面板设计，使声源位置的标记更加具体形象。

② 我们将继续研究小型化，轻量化装置的方法。

③ 因此下一步我们将继续研究远场声源和运动声源的定位方法

④ 探究近场静止声源和运动声源的定位方法

⑤ 接下来我们将根据环境噪音短时能量和短时平均能量的计算方法优化声音判定的程序。

总之，如果要想实现更好的定位需求，必须要在算法研究和信号设备采集两头并进，以达到精确的结果。

九、参考文献

- [1] 王森,宫玉琳,王春阳,李明秋,李健建.MATLAB 加窗滤波在广义互相关时延估计中的应用[J].长春理工大学学报(自然科学版),2016,39(06):48-52.
- [2] 白顺先,马瑞士.语音端点检测中判决机制的研究[J].大连民族学院学报,2010,12(03):251-253+260.DOI: 10.13744/j.cnki.cn21-1431/g4.2010.03.023.
- [3] 罗倩.麦克风阵列声源定位的实验设计与实现[J].实验技术与管理,2020,37(06):185-188.DOI:10.16791/j.cnki.sjg.2020.06.040.
- [4] 毛青春,徐分亮.窗函数及其应用[J].中国水运(学术版),2007(02):230-232.
- [5] X. Alameda-Pineda and R. Horaud, "A Geometric Approach to Sound Source Localization from Time-Delay Estimates," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 22, no. 6, pp. 1082-1095, June 2014, doi: 10.1109/TASLP.2014.2317989.
- [6] GB/T 17247.1-2000
- [7] 刘英翠. 基于传声器阵列的被动声音定位技术研究[D].哈尔滨工业大学,2012.
- [8] 仝卜匀. 基于时延估计技术的声源定位方法研究[D].华北电力大学,2019.
- [9] 张明瀚. 基于时延估计的声源定位系统研究[D].重庆大学,2009.