

中国科学技术大学

硕士学位论文工作中期报告考核表

学号	SA21219031	姓名	侯祎夫	导师	陈香
工程领域	电路与系统		研究方向	室内定位	
论文题目	复杂非视距场景下高精度音频定位研究				
论文类型	①工程设计（ ）		②研究论文（ √ ）		

论文进展情况

一、研究目标

旨在推动室内音频定位技术的应用，本论文的研究目标是为解决在复杂 NLOS 场景下，基于 TDOA 的音频定位方案鲁棒性不佳的问题，需要探索一种具有识别 LOS、NLOS 信号类型并根据结果自动优化定位流程的高精度定位方案。

二、研究内容

- (1) 对 LOS、NLOS 信号有效的特征提取方法的探索及分类器实现。
- (2) 强 NLOS 场景下优化定位算法的开发。
- (3) 复杂 NLOS 场景下定位算法方案的整体实现。

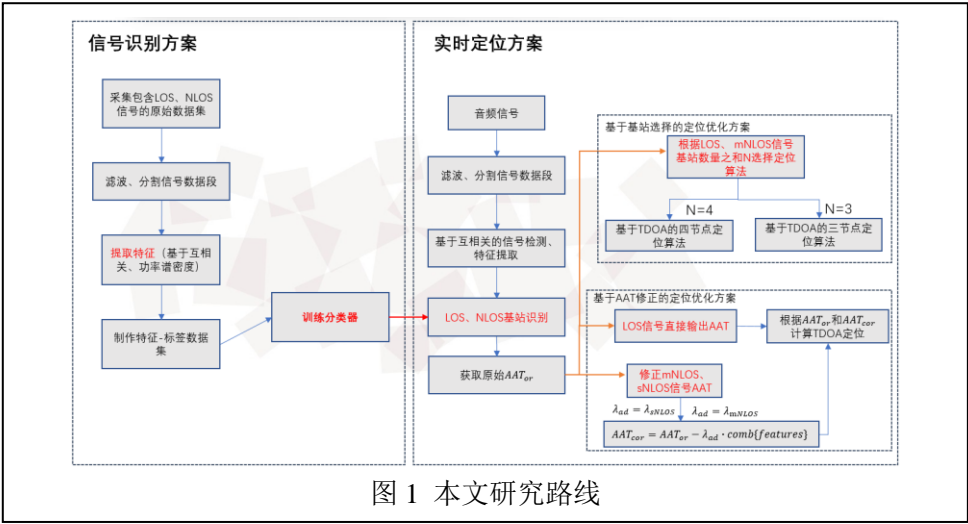


图 1 本文研究路线

本论文整体研究路线如图 1 所示，主要包括视距（line-of-sight, LOS）、非视距（non-line-of-sight, NLOS）信号识别方案和实时定位方案。信号识别方案首先需要构建包含 LOS、NLOS 信号的原始数据集，在提取有效的音频信号段后对数据集作特征提取制作“特征——标签数据集”，应用得到的“特征——标签数据集”训练分类器。在得到训练好的分类器后，可以进行实时定位方案的研究。该部分研究前半段流程是相同的，首先终端接收到指定频段的音频信号，以基于互相关的方法对其作信号检测和特征提取，在获取信号原始到达时刻（audio arrival time, AAT）并应用训练好的分类器作信号类型分类后，基于不同的定位优化思路该部分研究后半段可分为两部分。第一条路径是基于基站选择的定位优化方案，根据基站信号类型识别结果，选择 LOS、mNLOS（mild NLOS）信号定位，从而减小定位误差并保证系统刷新率。第二条路径是基于 AAT 修正的定位优化方案，根据基站信号识别结果，保留 LOS 信号的原始 AAT，对 NLOS 信号的 AAT 作不同程度的修正，利用修正之后的 AAT 进行定位。

## 1. 对 LOS、NLOS 信号有效的特征提取方法的探索及分类器实现。

该部分研究可分为三步骤实现：

(1) 第一步：采集原始信号数据集。

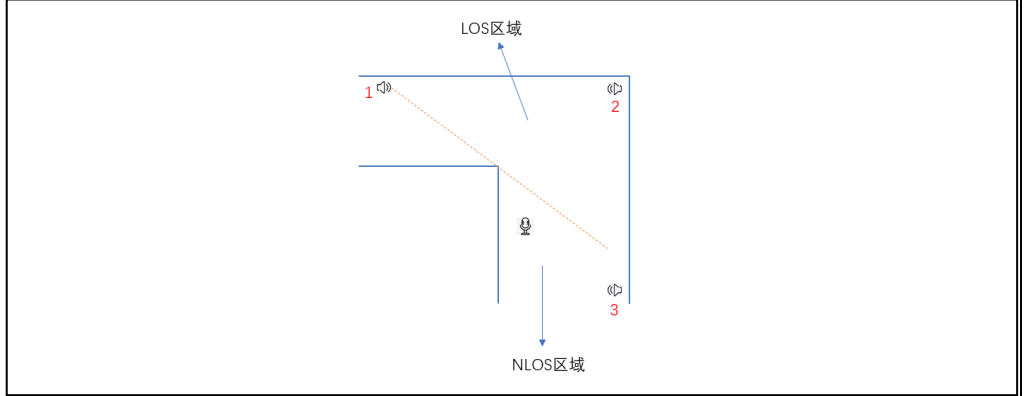


图 2 数据采集示意图

由于本研究将音频信号分为 LOS、mNLOS、sNLOS 三种类型，为了正确地给 mNLOS、sNLOS 两种信号打上标签，本研究提出一种基于 TDOA 的三节点定位精度判断数据类型的方法。如图 2 所示，为确定某数据采集点对 1 号基站来说是否为合适的 NLOS 采集点，选取两个 LOS 基站（2，3 号）与 1 号基站组成基本的三节点定位组合，通过考察其平均定位误差（mean positioning error, MPE）来判断在该店采集到的信号类型。MPE 由下式给出：

$$MPE = \frac{1}{N_c} \sum_{i=1}^{N_c} \|p(i) - \hat{p}(i)\|. \quad (1)$$

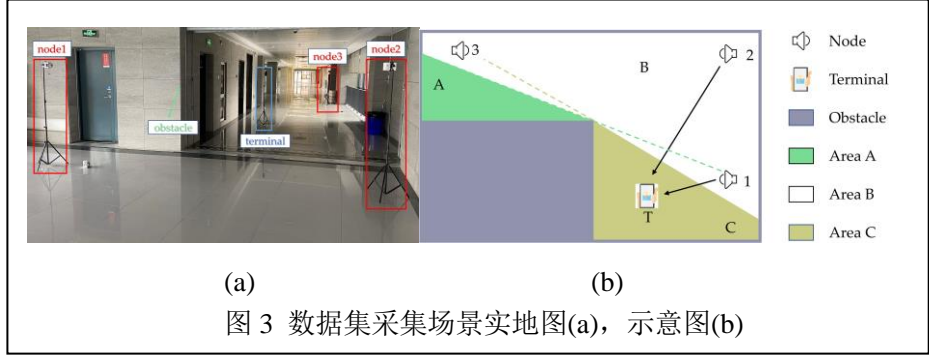
(2) 第二步：音频信号特征提取研究。在采集到合适的数据之后，对 LOS、NLOS 信号的时频域特性开展特征提取研究，提取适用于 LOS、mNLOS、sNLOS 三分类的特征，获得“特征——标签”训练矩阵。

(3) 第三步：分类器设计。选择支持向量机（support vector machine, SVM）作为分类器，设置合适的训练参数。应用第二步获得的训练矩阵对分类器进行训练，需要探索最佳的特征组合并对比其他论文中的特征提取方法。

该部分已取得进展：

(1) 原始信号数据集的构建

如图 3 所示，分别为实验数据采集场景的实景图(a)和示意图(b)，实验采集场景被划分为三部分 A, B, C 区域，对在 A 区域的采集点来说，基站 2, 3 是 LOS 基站，基站 1 是 NLOS 基站；对在 B 区域的采集点来说，三个基站全部为 LOS 基站；对在 C 区域的采集点而言，基站 1, 2 是 LOS 基站，而基站 3 是 NLOS 基站。



在 A, B, C 三个采集区域均匀设置约十个数据采集点，在每个采集点采集约一分钟的数据，由于信号发射周期为 1s，则每个采集点包含约 60 个周期的数据。实验的 MPE 阈值设置为 0.8m，当采集的 NLOS 信号结合三节点定位误差 MPE 小于 0.8m 时，判断采集到的 NLOS 信号为 mNLOS 类型，反之则为 sNLOS 类型。

采集的数据集包含约 1200 个 sNLOS 信号样本，1300 个 mNLOS 样本，8000 个 LOS 样本，为了使训练数据集各类型样本数相同，每类信号随机选取 1000 个样本，则最终的数据集包含三个类型的样本各 1000 个。

## (2) 音频信号特征提取方法研究

基于互相关函数 (cross-correlation function) 方法和区间搜索方法共提取了六种特征  $\{\hat{\tau}_{mf}, \hat{\alpha}_{mf}, R_{an}, \tau_{gm}, \tau_{gsd}, num_{gm}\}$ ，其中  $\hat{\tau}_{mf}, \hat{\alpha}_{mf}, R_{an}$  三种特征直接基于 CCF 提取得到，定义分别为：

$$\hat{\tau}_{mf} = \hat{\tau}_m - \hat{\tau}_f \quad (2)$$

$$\hat{\alpha}_{mf} = \frac{\hat{\alpha}_f}{\hat{\alpha}_m} = \frac{|R_x(\hat{\tau}_f)|}{|R_x(\hat{\tau}_m)|} \quad (3)$$

$$R_{an} = \frac{|R_x(\hat{\tau}_f)|}{mean(|R_x(1:\hat{\tau}_f)|)} \quad (4)$$

其中， $\hat{\tau}_m$  和  $\hat{\tau}_f$  分别是 CCF 峰值和信号 AAT 对应 CCF 取值， $|R_x(\tau)|$  表示 CCF 函数取绝对值。剩余三种特征  $\tau_{gm}, \tau_{gsd}, num_{gm}$  基于本研究提出的区间搜索方法得到。如图 4 所示，将 CCF 从 AAT 开始截取 45ms 的数据段长度得到  $|R'_x(\tau)|$ ，按峰值比例将其划分为 L 个搜索区间，对每一个搜索区间寻找其平均时延  $\bar{\tau}_i$ ，从而得到  $\{\bar{\tau}_1, \bar{\tau}_2, \dots, \bar{\tau}_L\}$ ， $\bar{\tau}_i$  由式 (5) 得到：

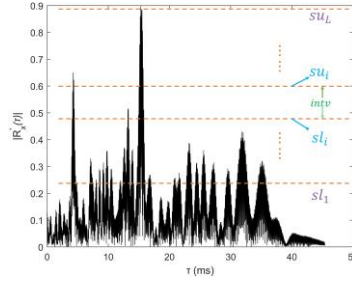


图 4 区间搜索方法示意图

$$\bar{\tau}_i = \text{mean} \left( \arg(sl_i \leq |R'_x(\tau_j)| \leq su_i) \right), \quad j = 1, 2, \dots, \text{length}(R'_x). \quad (5)$$

其中 $sl_i$ 为第 $i$ 个搜索区间的下限， $su_i$ 为第 $i$ 个搜索区间的上限， $sl_1 = \alpha_{thd} \cdot su_L$ ， $\alpha_{thd}$ 是系数阈值，本研究设置为 0.2， $su_L$ 就是 CCF 的最大值。每个区间的宽度为 $intv$ ，可由式（6）计算得到。

$$intv = (1 - \alpha_{thd}) \cdot \max(|R'_x(\tau)|) / L. \quad (6)$$

与 $\{\bar{\tau}_1, \bar{\tau}_2, \dots, \bar{\tau}_L\}$ 集合对应还可以得到集合 $\{num_1, num_2, \dots, num_L\}$ ，其中 $num_i$ 代表在对第 $i$ 个区间进行搜索时，满足条件的 $\tau_j$ 个数。

三种特征 $\tau_{gm}$ 、 $\tau_{gsd}$ 、 $num_{gm}$ 分别基于集合 $\{\bar{\tau}_1, \bar{\tau}_2, \dots, \bar{\tau}_L\}$ 和 $\{num_1, num_2, \dots, num_L\}$ 提取相关统计特性提出，具体由下式给出：

$$\tau_{gm} = \frac{\sum_{i=1}^L r_i^2 \bar{\tau}_i}{\sum_{j=1}^L r_j^2}, \quad \tau_{gsd} = \sqrt{\frac{\sum_{i=1}^L \frac{r_i^2}{\sum_{j=1}^L r_j^2} \bar{\tau}_i - \tau_{gm}}{L}} \quad (7)$$

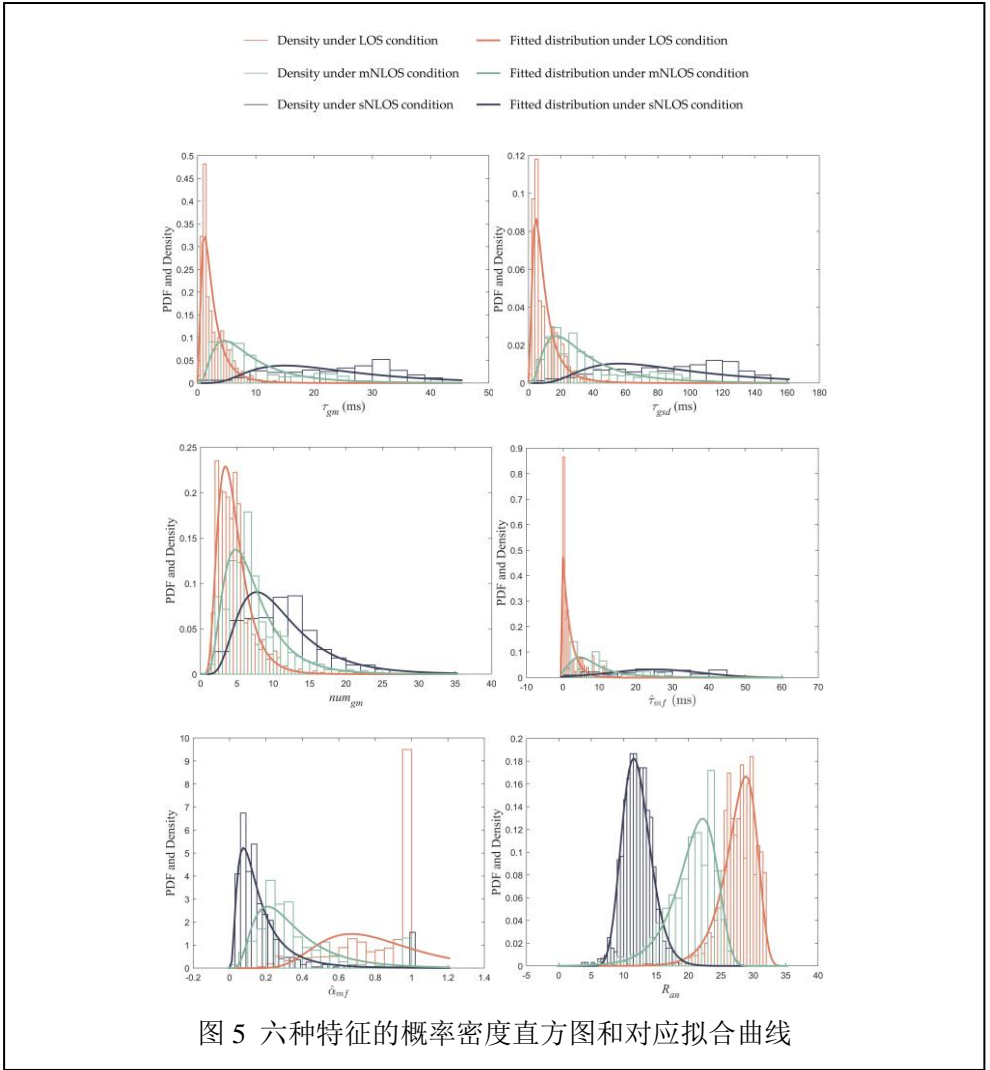
$$num_{gm} = \frac{\sum_{i=1}^L r_i^2 num_i}{\sum_{j=1}^L r_j^2}. \quad (8)$$

其中 $r_i$ 为第 $i$ 个区间的中值，即 $r_i = (sl_i + su_i) / 2$ 。

应用上述特征提取方法对第 1 步获得的包含 3000 个样本的数据集作特征提取，得到“特征——标签”训练数据集。数据集内的六种特征分布如图 5 所示，每幅图表示不同的特征概率密度直方图和对应的拟合曲线。可以看到，每种特征都有较为明显的区分度，这为后续分类器提供良好的训练数据集支撑。

### （3）分类器设计及训练结果

本研究训练使用 SVM 分类器完成 LOS、mNLOS、sNLOS 三分类任务。由于 SVM 本身是二分类器，因此本研究训练三个 SVM 分类器 $SVM_l$ 、 $SVM_s$ 和 $SVM_m$ ， $SVM_l$ 将 LOS 信号与其他两种信号分开， $SVM_s$ 将 sNLOS 信号与其他两种信号分开， $SVM_m$ 将 mNLOS 信号与其他两种信号分开。对于一个新的待分类样本，经过三个分类器的评估后最终得到三个分类评分 $score_l$ 、 $score_s$ 和 $score_m$ ，选取最高评分对应的信号类型作为该待分类样本的分类结果。分类的准确率准则由式（9）确定。



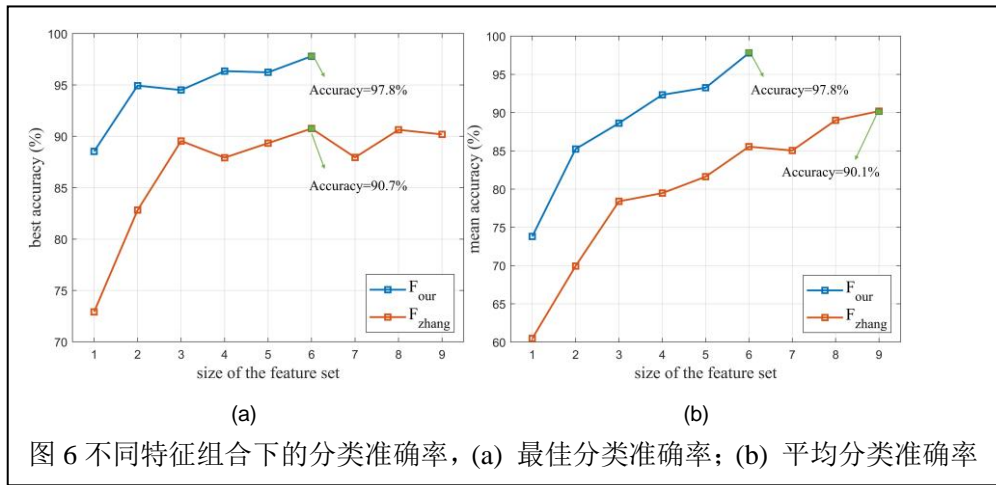
$$accuracy = \text{mean} \left( \frac{a_l}{n_l} + \frac{a_m}{n_m} + \frac{a_s}{n_s} \right) \cdot 100\%.$$

(9)

其中,  $a_l$ ,  $a_m$  和  $a_s$  表示 LOS、mNLOS、sNLOS 信号的分类正确样本数,  $n_l$ ,  $n_m$  和  $n_s$  表示 LOS、mNLOS、sNLOS 总的信号样本数。

为研究最佳的特征组合, 以不同的特征集合大小训练测试 SVM 分类器, 如当特征集合大小为  $M$  时, 共有  $C_6^M$  个特征组合,  $M = 1, 2, \dots, 6$ 。为了对比显示本研究提出的特征提取方法的优越性, 另一篇论文中的特征提取方法被用作比对, 其特征集合记为  $F_{\text{Zhang}}$ , 相应的, 本文提出的特征集合记为  $F_{\text{Our}}$ 。其中, 对比特征集合  $F_{\text{Zhang}} = \{\tau_{\text{med}}, \tau_{\text{rms}}, k, s, K_R, g_m, g_{\text{rms}}, k_f, s_f\}$ , 共有 9 个特征, 则对特征集合大小  $M$ , 共有  $C_9^M$  个特征组合,  $M = 1, 2, \dots, 9$ 。图 6 显示了在不同的特征组合大小下,  $F_{\text{Our}}$  和  $F_{\text{Zhang}}$  各自的最佳分类准确率和平均分类准确率。图 6(a) 展示了在最佳分类准

准确率的评价下，两种特征提取方法的表现。当  $M=6$  时， $F_{our}$  取得最佳的表现，达到 97.8% 的分类准确率，而  $F_{zhang}$  也在  $M=6$  下取得最佳的分类准确率，达到 90.7%，两

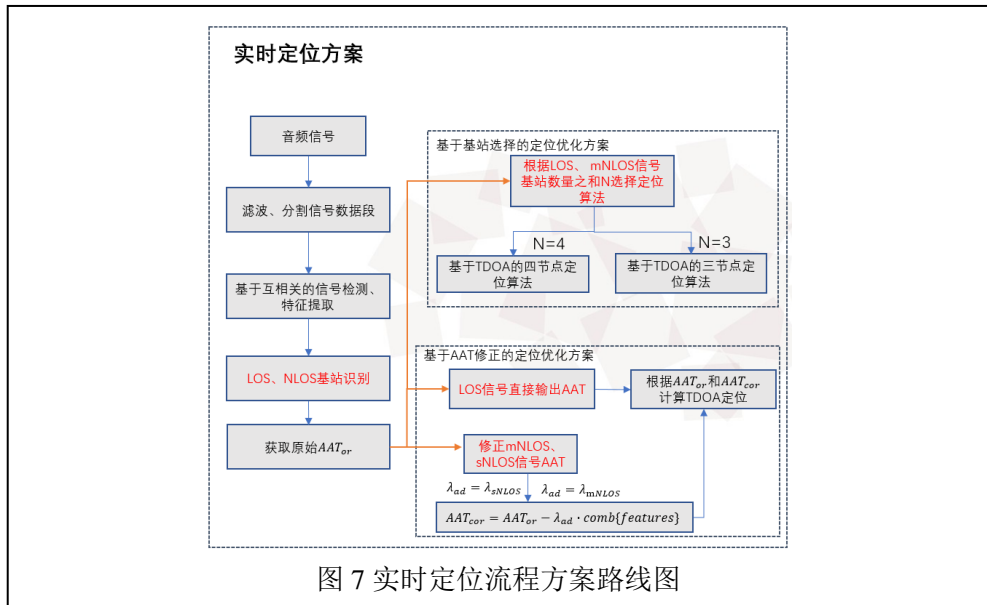


者相差 7 个百分点。而在平均分类准确率的评估下， $F_{our}$  依然领先于  $F_{zhang}$  近 8 个百分点。可以看出，本文提出的特征提取方法在 LOS、mNLOS、sNLOS 三分类任务中取得了更为优异的表现。

## 2. 强 NLOS 场景下优化定位算法的开发

### 该部分研究已取得进展：

当对各基站发射的音频信号进行 NLOS/LOS 识别并检测获得原始 AAT 以后，本课题采用图 7 所示实时定位流程，实现高精度音频定位。具体的，根据不同的定位优化思路，后续的定位优化方案可分为两条路径。第一条路径是基于基站选择的定位优化方案，根据基站信号类型识别结果，选择 LOS、mNLOS（mild NLOS）信号

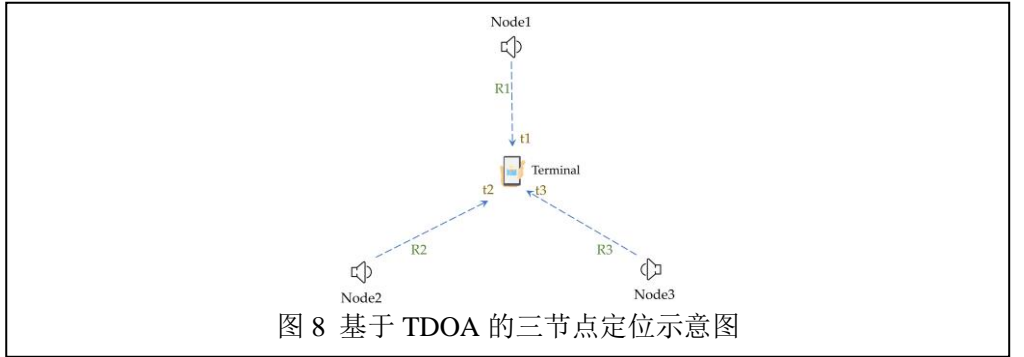


定位，从而减小定位误差并保证系统刷新率。第二条路径是基于 AAT 修正的定位优化方案，根据基站信号识别结果，保留 LOS 信号的原始 AAT，对 NLOS 信号的 AAT 作不同程度的修正，利用修正之后的 AAT 进行定位。

### (1) 基于基站选择的定位优化方案

根据信号识别结果，直接抛弃 sNLOS 基站信号，利用 LOS、mNLOS 信号作定位估计。具体的，根据 LOS、mNLOS 信号数量之和  $N$  选择定位算法，当  $N=4$  时，选择基于 TDOA 的四节点定位算法；当  $N=3$  时，选择基于 TDOA 的三节点定位算法，当  $N<3$  时，不输出定位结果。因为在定位优化过程中会抛弃 sNLOS 信号，所以在 LOS、mNLOS 信号数量  $N$  不满足三节点定位算法的时候，系统拒绝给出定位结果，造成系统的刷新率下降。但是由于系统只利用 LOS、mNLOS 信号，因此在复杂 NLOS 场景下的定位误差会大幅度减小。

#### a) 基于 TDOA 的三节点定位算法



如图8所示为基于TDOA的三节点定位示意图，终端接收来自各个基站的信号， $t_1, t_2, t_3$  为各个基站信号到达终端的时间， $R_1, R_2, R_3$  为基站距离各个终端的距离。以基站 Node1 为参考，终端的位置可由式（10）解出：

$$\begin{cases} (x-x_1)^2 + (y-y_1)^2 = R_1^2 \\ (x-x_2)^2 + (y-y_2)^2 = R_2^2 = (R_1 + d_{21})^2 \\ (x-x_3)^2 + (y-y_3)^2 = R_3^2 = (R_1 + d_{31})^2 \end{cases} \quad (10)$$

其中， $x_i, y_i$  为各个基站的横纵坐标， $i=1,2,3$ .  $\hat{\mathbf{p}}(x, y)$  为终端的坐标， $d_{21}=R_2-R_1$ ,  $d_{31}=R_3-R_1$ .

#### b) 基于 TDOA 的四节点定位算法

终端的最终位置  $\hat{\mathbf{p}}_f$  由式（11）给出：

$$\hat{\mathbf{p}}_f = \sum_{i=1}^{M_{tc}} \omega_i \hat{\mathbf{p}}_i, \quad (11)$$

其中， $M_{tc}$  是三节点基站的组合数，当基站数目  $N=4$  时， $M_{tc}=C_N^3=4$ .  $\omega_i$  是每种三节点组合定位下的权重，由式（12）给出：



$$\omega_i = \frac{(1/\widehat{CRLB}_i)}{\sum_{j=1}^{M_{tc}} (1/\widehat{CRLB}_j)} \quad (12)$$

其中， $\widehat{CRLB}$ 是克拉美劳下界，可以给出任何一个无偏估计的方差下界，通过式（12）来确定每种三节点定位组合下的权重，通过加权平均的方式确定最终的终端位置估计值 $\hat{\mathbf{p}}_i$ 。

## （2）基于 AAT 修正的定位优化方案

获得信号类型识别结果和原始 AAT 量测后，保留 LOS 信号的 AAT，对 sNLOS、mNLOS 信号的 AAT 作不同程度的修正。修正公式如下所示：

$$AAT_{cor} = AAT_{or} - \lambda_{ad} \cdot comb\{features\} \quad (13)$$

其中， $AAT_{or}$ 是原始 AAT， $AAT_{cor}$ 是修正之后的 AAT， $\lambda_{ad}$ 是自适应参数，在 sNLOS 和 mNLOS 信号类型下有不同的取值。该定位方案需要考察以不同的特征作为 AAT 修正参考的前提下， $\lambda_{sNLOS}$ 和 $\lambda_{mNLOS}$ 最优的取值组合情况。并在多种不同的 NLOS 场景下开展实验验证定位优化效果。因为该定位优化的思路是不抛弃任何 AAT 量测，只对特定的 NLOS 信号的 AAT 作出修正，所以系统的刷新率可以达到 100%。但由于 $\lambda_{ad}$ 的设置值具有场景局限性，因此定位优化效果的鲁棒性不强。

## 3. 复杂 NLOS 场景下定位算法方案的整体实现

该部分研究已取得进展：

### （1）基于基站选择的定位优化方案

#### a)静态定位实验结果

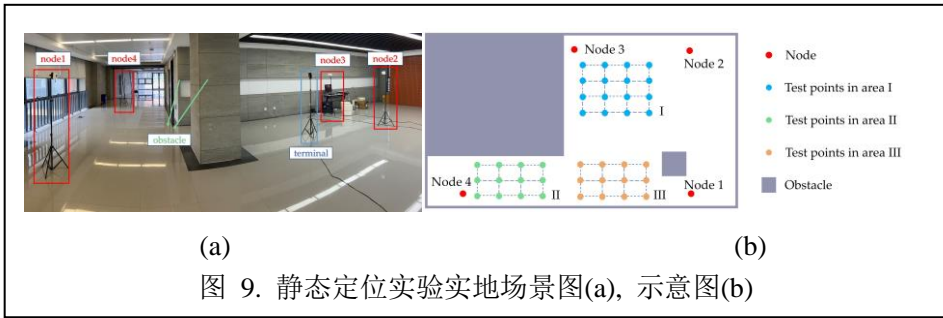


图 9. 静态定位实验实地场景图(a), 示意图(b)

如图 9 所示，为静态定位实验实地场景图和示意图。在该复杂 NLOS 场景下布设了四个音频发射基站 Node 1, 2, 3, 4。定位测试点采集被划分为三个区域 area I, II, III。对区域 I 的定位测试点来说，Node 2, 3 是 LOS 基站，Node 1, 4 是 NLOS 基站；对区域 II 的测试点来说，Node 1, 4 是 LOS 基站，Node 2, 3 是 NLOS 基站；对区域 III 的测试点来说，Node 1, 3, 4 是 LOS 基站，Node 2 有可能是 NLOS 基站。

假设本研究提出的基于基站选择的定位优化方案为 Scheme A。实验对比了其他两种方案 Scheme B, Scheme C。Scheme B 不对基站作任何识别，直接利用原始 AAT

定位；Scheme C 对基站信号类型作出识别，并抛弃 mNLOS、sNLOS 信号，只使用 LOS 信号定位，当 LOS 基站数目小于 3 时不输出定位结果。在图 9 所示的定位场景

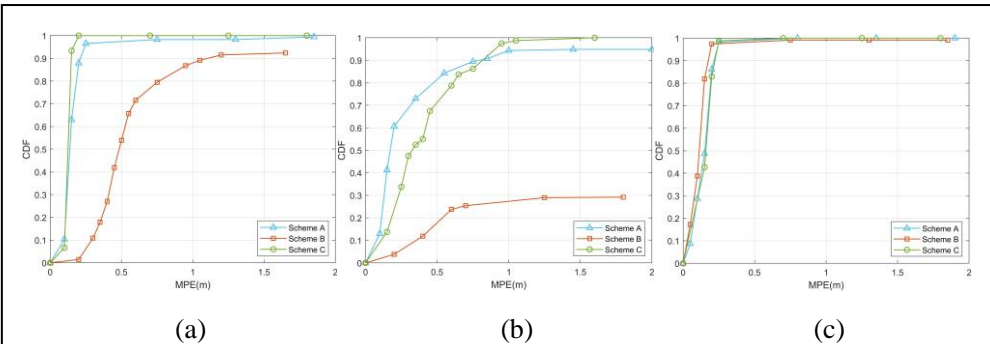


图 10. 三种实验方案的定位误差 CDF 图。(a) 区域 I; (b) 区域 II; (c) 区域 III。

下的静态定位实验结果如图 10 所示。

图 10 给出了静态定位实验中 MPE 的累积分布函数(CDF)。方案 A 和方案 C 获得了较好的定位精度。方案 A 中，90%的 MPE 在 Area I 小于 0.2 米，80%的 MPE 在 Area II 小于 0.5 米，90%的 MPE 在 Area III 小于 0.3 米。方案 C，100%的 MPE 在 Area I 小于 0.2 米，70%的 MPE 在 Area II 小于 0.5 米，90%的 MPE 在 Area III 小于 0.3 米。50%的 MPE 在 Area I 大于 0.5m，80%的 MPE 在 Area II 大于 0.5m，90%的 MPE 在 Area III 小于 0.3m。

表 1 显示了使用三种定位方案在三个定位区域获得的 AMPE 和 SRR。AMPE 是一个定位区域内所有测试点的 MPE 的平均值。在 Area I 和 Area II，由于一些测试点从相应的节点接收到两个 NLOS 信号，方案 B 直接采用多节点定位算法，获得了最高的定位误差(米级)，系统刷新率为 100%。相比之下，方案 A 和方案 C 的定位误差较小(分米级)，且 SRR 也较低。与方案 C 在 LOS 节点少于 3 个时拒绝输出定位结果相比，采用 mLOS 节点的优化方案 A 的定位误差略高，两者之间的差值在 13 cm 以内。然而，由于方案 C 对 LOS 节点的数量要求严格，其 SRR 相对较低，仅为 13.6%，远低于方案 A 的 61%，这意味着方案 C 每 5 秒甚至 6 秒进行一次位置估计。显然，C 方案不能胜任复杂 NLOS 环境下的实时定位任务。

	Area I		Area II		Area III	
	AMPE	SRR	AMPE	SRR	AMPE	SRR
<b>Scheme A</b>	0.261	61	0.563	96.6	0.138	98.1
<b>Scheme B</b>	0.967	100	4.823	100	0.232	100
<b>Scheme C</b>	0.137	13.6	0.413	18.8	0.167	90.7

表 1. 静态定位实验的 AMPE(m)和 SRR(%)

## b) 动态定位实验结果

如图 11 所示，有三条动态定位路线 Track I, Track II, Track III。这三个轨迹分别是在静态定位实验中，Area I、Area II 和 Area III 的外缘。在动态定位实验中，实验者沿着矩形轨道顺时针方向行走，末端拿着三脚架，行走速度控制在 1m/s 左右，每个轨道走两次。对于 Track II 和 Track III 收集了大约 60 个定位周期的数据，对于 Track I 由于其较长的长度，收集了大约 90 个定位周期的数据。

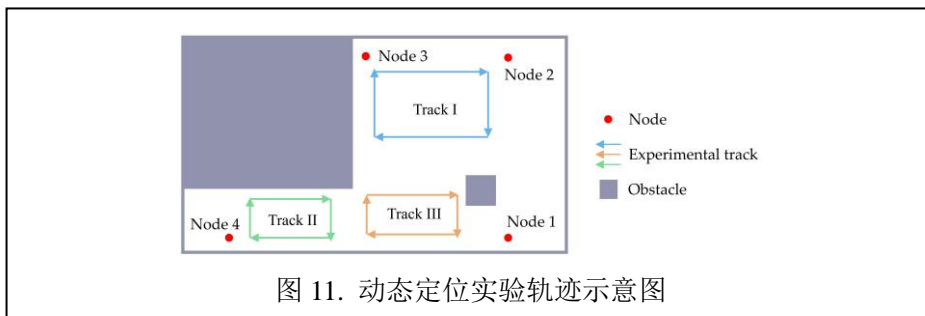


图 11. 动态定位实验轨迹示意图

如图 12 所示，方案 A 和方案 C 获得了较好的定位精度。方案 A 的 Track I 和 Track II 的 80% MPE 小于 0.4m，Track III 的 90% MPE 小于 0.5m。方案 C Track I 的 90% MPE 小于 0.5m，Track II 的 90% MPE 小于 0.4m，Track III 的 90% MPE 小于 0.5m。同时，方案 B 的定位精度最差：Track I 的 50% MPE 大于 0.5m，Track II 的 60% MPE 小于 0.4m，Track III 的 90% MPE 小于 0.5m。

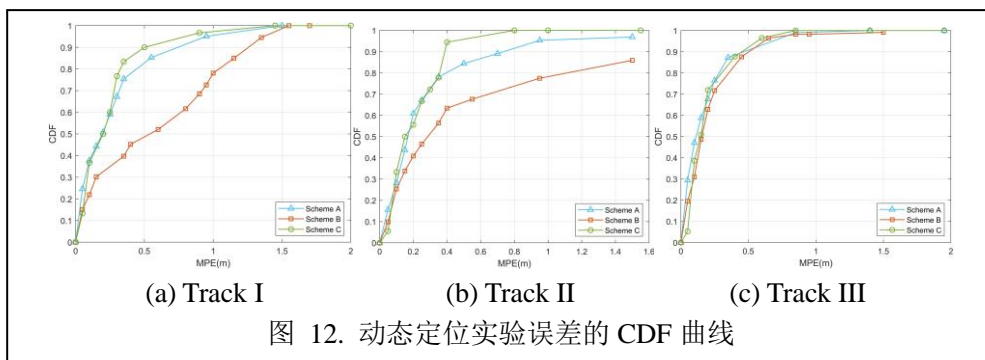


图 12. 动态定位实验误差的 CDF 曲线

	Track I		Track II		Track III	
	AMPE	SRR	AMPE	SRR	AMPE	SRR
<b>Scheme A</b>	0.285	83.6	0.392	90.1	0.171	90.3
<b>Scheme B</b>	0.588	100	1.221	100	0.311	100
<b>Scheme C</b>	0.234	31.1	0.221	25.3	0.178	88.4

表 2. 静态定位实验的 AMPE(m)和 SRR(%)

与静态定位实验的结果类似，对于 Track I 和 Track II，方案 B 获得了最高的定位误差(米级)和系统刷新率(100%)，因为一些测试点接收到两个 NLOS 信号。相比之下，方案 A 和方案 C 的定位误差较小，且刷新率也较低。与方案 C 相比，方案 A 的定位误差略高(分米级)，Track I 的定位误差差值在 5 cm 以内，Track II 的定位误差差值在 17 cm 以内。但另一方面，由于方案 C 对 LOS 节点数的严格要求，其 SRR 相对较低，仅为 25.3%，远低于方案 A 的 90.1%。这表明方案 C 在复杂的 NLOS 环境下每 4 秒进行一次位置估计，不适合实时定位系统。对于 Track III 上的大多数测试点，所有四个节点都是 LOS。三种方案的定位误差均小于 32 cm，SRR 大于 88%。总体来说，A 方案的定位性能最好，定位误差为 17.1 cm，信噪比高达 90.3%。

(2) 基于 AAT 修正的定位优化方案初步探究

如图 13 所示，在有墙壁遮挡的场景下布置两个基站，在两基站的连接线上采集 4 个定位测试点，在接收到音频信号后，对信号作分类识别，保留 LOS 信号的 AAT，对 NLOS 信号 AAT 按式（5）作修正，最后应用原始 AAT 和修正之后的 AAT 作定位。AAT 修正依据 $\tau_{gm}$ 这一特征，修正值为 $\tau_{gm}/\lambda_{ad}$ ， $\lambda_{sNLOS}$ 和 $\lambda_{mNLOS}$ 均取 10。

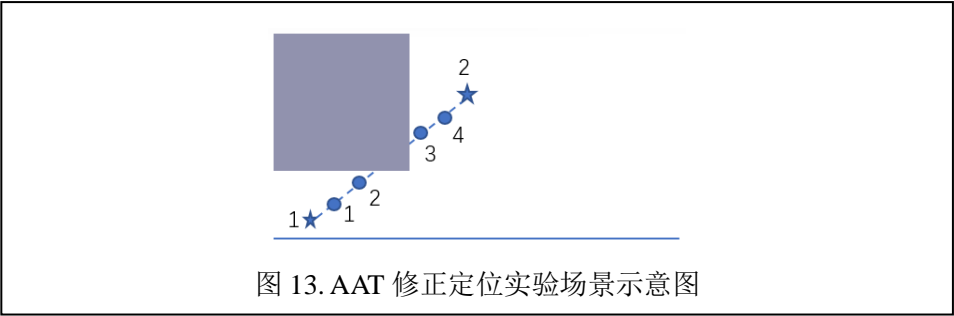


图 13. AAT 修正定位实验场景示意图

	直接定位(m)	AAT 修正定位(m)
1	0.30	0.21
2	0.34	0.20
3	0.21	0.13
4	0.30	0.27

表 3. AAT 修正方案定位优化结果

由表 3 看出，经过对 NLOS 信号的 AAT 修正之后，定位误差有较明显的下降，说明基于 AAT 修正的定位优化思路是可行的。

有待改进之处

- 1.需进一步补充基于 AAT 修正的定位优化方案中关于不同特征下对应的最佳 $\lambda_{sNLOS}$ 和 $\lambda_{mNLOS}$ 参数组合实验结果。
- 2.需进一步补充不同 NLOS 场景下的定位优化实验结果



