

DOI: 10.19650/j.cnki.cjsi.J2312175

# 多特征优化下室内声源鲁棒跟踪算法

刘望生, 刘艳梅

(浙江理工大学信息科学与工程学院 杭州 310018)

**摘要:**针对室内强混响低信噪比下单特征声源跟踪精度低和稳健性差的问题,提出了一种多特征优化鲁棒跟踪算法。该算法建立了基于时延估计多假设模型的多特征优化机制,克服了混响噪声下单特征定位性能差的缺陷。为了提高多特征优化机制对说话人随机运动的稳健性,提出了交互式多模型(IMM)粒子滤波改进算法,通过对模型噪声方差和模型概率实时调整,增强了多特征优化机制的鲁棒性。仿真分析和实际测试结果表明,在多特征优化机制下,IMM改进算法比已有模型位置平均均方根误差(RMSE)降低约12%;在IMM改进算法下,多特征优化机制相对其他算法位置平均RMSE降低89.6%左右。该算法明显消除了混响噪声的不利影响,提高了声源定位与跟踪的精度和鲁棒性。

**关键词:**室内混响;多特征;声源跟踪;交互式多模型;粒子滤波

中图分类号: TH712 文献标识码: A 国家标准学科分类代码: 510.4040

## Robust tracking algorithm for indoor sound source based on Multi-feature optimization

Liu Wangsheng, Liu Yanmei

(School of Information Science and Engineering, Zhejiang Sci-Tech University, Hangzhou 310018, China)

**Abstract:** To address the issue of low accuracy and poor robustness in single-feature sound source tracking under strong indoor reverberation and low signal-to-noise ratio (SNR), a robust tracking algorithm using multi-feature optimization mechanism is presented in this paper. This algorithm establishes a multi-feature optimization mechanism based on a time-delay estimation multi-hypothesis model, overcoming the poor localization performance of single-feature tracking in reverberant noise environments. Moreover, To enhance the robustness of the multi-feature optimization mechanism against random movements of the speaker, we introduce an improved Interacting Multiple Model (IMM) particle filter algorithm. By real-time adjustment of model noise variance and model probability, the robustness of the multi-feature optimization mechanism is improved. Simulation analysis and actual test results indicate that the average root mean square error (RMSE) of the position is reduced by approximately 12% using the proposed algorithm, compared with the existing literature, under the multi-feature optimization mechanism. Based on the improved IMM algorithm, the average RMSE of the position is reduced by nearly 89.6% through the proposed algorithm, compared with the other algorithms. The proposed algorithm significantly eliminates the adverse effects of reverberation and noise, and improves the accuracy and robustness of sound source localization and tracking.

**Keywords:** room reverberation; multi-feature; sound source tracking; IMM; particle filter

## 0 引言

基于麦克风阵列的声源定位与跟踪技术在虚拟现实、无人驾驶、安防监控、车载通信以及人工智能等领域有着广泛的应用<sup>[1-3]</sup>。然而受多种不利因素的影响<sup>[4-5]</sup>,

室内声源跟踪精度还存在较大的误差,其难点源于室内声学环境的复杂性和声源运动的随机性。在众多声源位置信息提取技术中,空时相关法具有运算量低以及精度高等优点受到学者们的关注。

空时相关法中基于广义互相关的时延估计方法受室内环境影响较大<sup>[6-7]</sup>,而实际应用场景大都是在低信噪比

(signal-to-noise ratio, SNR) 和强混响下进行的,为了解决混响噪声下声源定位问题,文献[8]提出一种联合加权的广义二次相关时延估计算法,通过联合相位变换和平滑相干变换得到新的广义加权函数,在低信噪比下提高了时延估计精度。文献[9]提出了一种基于稀疏过零点信息的抗幅值失真时延估计方法,通过对两路信号幅值归一化并错位相减绘制误差曲线得到时延值,避免了幅值信息失真带来的影响。文献[10]提出一种基于互相关序列和反向传播神经网络的声源定位算法,在复杂环境下,能够明显提高声源定位性能,但跟踪精度仍有限。文献[11]提出了一种基于扩频声波的室内定位方法,通过对声波信号进行扩频调制和定位解算,取得了一定的抗干扰能力和较高的定位精度。该方法受同步精度影响大,且需四路扩频调制信号,不适合室内说话人的定位。这些改进算法提高了时延估计的鲁棒性,但当室内声学环境变差时会产生虚拟声源。为了消除虚拟声源的影响,文献[12]提出了一种多目标检测前跟踪算法,采用集成定位函数 SBF 和基于像素的检测前跟踪框架跟踪多个同时活动的语音源,取得了更稳定的性能。文献[13]提出一种可控响应功率相位变换(steered-response power-PHase transform, SRP-PHAT)改进法,通过在其周围对称体积延迟间隔上累加广义互相关相位变换(generalized cross correlation-PHase transform, GCC-PHAT)获得了较高的精度。文献[14]提出分布式 SRP 声源定位法,通过采用无线声传感器网络节点快速分布线性平均并迭代运算,在通信负载和定位性能之间取得了较好的平衡。上述方法采用波束输出能量定位,混响噪声下稳健性强,但精度有限。

为了突破单特征定位算法的局限性,文献[15]将到达时间差(time difference of arrival, TDOA)和 SRP 相结合,提出了分布式快速定位法,取得了较好的跟踪效果。文献[16]提出了基于奈奎斯特-香农采样的可控响应功率(steered-response power, SRP)法,利用对 GCC 的 TDOA 间隔进行严格抽样,以有限的插值成本近似 SRP 图,在定位性能相当的情况下降低了计算复杂度。文献[17]提出了压缩感知基追踪去噪法,该方法能有效处理水下多径效应引起的混合相干信号问题,提高了检测精度,但该方法对噪声敏感,存在检测性能不稳定的问题。文献[18]将麦克风阵列接收数据的样本协方差特征向量馈送到深度神经网络训练,在测距和测向方面取得了较高的精度。相比单特征而言,以上所提的多特征/压缩感知以及深度学习方法在声源跟踪性能上有很大提高,但这些方法不适合跟踪不同运动模式的声源。

由于多特征对说话人随机运动的鲁棒性较弱,文献[19]提出了郎之万(Langevin, L)模型,文献[20]提出了随机行走(random walk, RW)模型,文献[21]建立了

自适应当前统计模型,这些单模型改进算法改善了跟踪系统的鲁棒性,但很难全面描述说话人的随机运动。为了提高模型的适应性,文献[22]提出了一种基于分布式的交互式多模型(interacting multiple model, IMM)方法,其中模型输入交互通过相邻节点状态进行,取得了较好的跟踪效果。文献[23]提出了一种改进的交互式多模型多伯努利滤波器,该方法可以有效地解决多源机动问题并跟踪波达方向。由于 IMM 是为适应目标机动而建立的,其关键在于模型集的选择以及模型信息的交互利用,上述文献尽管做了改进,但造成了模型信息的部分损失。

针对上述问题,本文从麦克风接收信号的空时相关出发,提出了多特征优化鲁棒跟踪算法。将时延估计多假设模型和波束能量相结合,利用迭代粒子滤波提取了多特征优化(multi-feature optimization, MFO)观测量。根据说话人运动的特点,建立了 IMM 改进算法。文中利用数值模拟分析了所提算法的可行性,最后利用室内实测数据验证了所提算法的有效性。

## 1 接收信号多特征描述

### 1.1 GCC 时延估计

在实际室内声学环境中,麦克风接收信号是多个信号源的延迟与噪声的叠加。设  $M$  个麦克风中第  $i$  个麦克风  $k$  时刻的接收信号为:

$$x_i(k) = h_i(\mathbf{l}, k) * s(k) + n_i(k) \quad (1)$$

式中:  $h_i(\mathbf{l}, k)$  为房间冲激响应,  $i = 1, 2, \dots, M$ ,  $\mathbf{l}$  为声源空间位置向量;  $*$  表示卷积;  $s(k)$  为声源信号;  $n_i(k)$  表示第  $i$  个麦克风的接收噪声。设  $n_i(k)$  互不相关,  $s(k)$  与  $n_i(k)$  不相关,则上式的频域表示为:

$$X_i(\omega) = H_i(\mathbf{l}, \omega) S(\omega) + N_i(\omega) \quad (2)$$

其中,  $X_i(\omega)$ 、 $H_i(\mathbf{l}, \omega)$ 、 $S(\omega)$  和  $N_i(\omega)$  分别为  $x_i(k)$ 、 $h_i(\mathbf{l}, k)$ 、 $s(k)$ 、 $n_i(k)$  的离散傅立叶变换,  $\omega$  为角频率。设两路麦克风  $i, j$  接收信号的互相关函数为:

$$G(\tau_d) = \int_{\Omega} \phi_d(\omega) X_i(\omega) X_j^*(\omega) e^{j\omega\tau_d} d\omega \quad (3)$$

式中:  $j = 1, 2, \dots, M$ , 当  $i \neq j$  时,  $G(\tau_d)$  称为广义互相关函数,  $\tau_d$  为时间延迟,  $d$  是两路麦克风  $i, j$  组成的麦克风对的索引;  $\phi_d(\omega)$  为互相关函数的权值,  $X_j^*(\omega)$  表示  $X_j(\omega)$  取复共轭,  $\Omega$  是频域。当  $\phi_d(\omega) = 1/|X_i(\omega)X_j^*(\omega)|$  时, 广义互相关函数被称作 GCC-PHAT<sup>[8]</sup>。使式(3)取最大值时的自变量就是时延估计(time delay estimation, TDE)值。研究表明,在弱混响高信噪比下,由 GCC-PHAT 估计的时延有较高的精度,但真实环境中的混响噪声会导致时延估计不准确。

## 1.2 延迟求和波束法

延迟求和波束有两种方法,一种是将信号延迟求和再计算波束能量,称为 SBF 法,另一种是可控时延相关函数求取波束能量。其中 SBF 的输出为:

$$P_{\text{SBF}}(\mathbf{I}) = \int_{\Omega} \left| \sum_{i=1}^M X_i(\omega) \exp(j\omega \|\mathbf{I} - \mathbf{I}_i\| c^{-1}) \right|^2 d\omega \quad (4)$$

式中:  $\mathbf{I}_i = [x_i \ y_i]^T$  为第  $i$  个麦克风的位置,  $T$  表示转置;  $c$  为声速。设  $\tau_d = \tau_i - \tau_j$ ,  $\tau_i$  与  $\tau_j$  为阵列可控时延,则可控制时延波束为:

$$P_{\text{CDS}}(\mathbf{I}) = P_{\text{CDS}}(\tau_d) = \sum_{d=1}^L G(\tau_d) \quad (5)$$

其中,当  $i \neq j$ ,  $L = \lfloor M/2 \rfloor$ ,  $\phi_d(\omega)$  取 1 时,表示相邻麦克风构成麦克风对的输出能量  $P_{\text{PAM}}(\mathbf{I})$ ,  $P_{\text{PAM}}(\mathbf{I})$  主要用来优化多特征向量;当  $i \neq j$ ,  $L = M(M-1)/2$ ,  $\phi_d(\omega) = 1/|X_i(\omega)X_j^*(\omega)|$  时,表示 SRP 定义的输出能量  $P_{\text{SRP}}(\mathbf{I})$ ,称为 SRP-PHAT。在混响环境中,SBF 定位性能差,但适合低信噪比环境;SRP-PHAT 有较强的抗混响性能,但 SRP-PHAT 受噪声影响大。为了提高 SRP-PHAT 定位精度,文献[24]建立了区域 SRP-PHAT 法,即:

$$P_{\text{SRP}}^c(\mathbf{I}) = 2\pi \sum_{i=1}^M \sum_{j=i+1}^M G(\tau_{i,j}) \quad (6)$$

式中:  $\mathbf{I}$  表示以  $\mathbf{I}$  为圆心的圆形网格上的点。已有文献表明<sup>[25]</sup>,混响噪声下,延迟求和波束法鲁棒性较强,但跟踪精度不够。

## 2 多特征粒子滤波优化

### 2.1 声源跟踪粒子滤波

受室内声学环境和测量方法的影响,时延观测向量表现出非线性和非高斯统计特性,因此可采用粒子滤波估计说话人的运动状态。声源跟踪粒子滤波是将说话人的跟踪问题转化为贝叶斯后验概率密度最大化建模问题,设说话人的状态空间模型为:

$$\mathbf{X}_k = f_k(\mathbf{X}_{k-1}) + \mathbf{v}_{k-1} \quad (7)$$

$$\mathbf{z}_k = g_k(\mathbf{X}_k) + \mathbf{n}_k \quad (8)$$

式中:  $f_k(\cdot)$  为状态转移函数;  $g_k(\cdot)$  为非线性观测函数;  $\mathbf{X}_k$  为状态向量;  $\mathbf{z}_k$  是观测序列;  $\mathbf{v}_{k-1}$ 、 $\mathbf{n}_k$  分别为零均值过程噪声和量测噪声,两者相互独立,其方差分别为  $\mathbf{Q}$  和  $\mathbf{R}$ 。给定观测  $\mathbf{z}_k$ ,则  $\mathbf{X}_k$  在最小均方误差准则下由下式条件均值给出<sup>[26]</sup>:

$$\hat{\mathbf{X}}_k = E[\mathbf{X}_k | \mathbf{z}_{1:k}] = \int \mathbf{X}_k P(\mathbf{X}_k | \mathbf{z}_{1:k}) d\mathbf{X}_k \quad (9)$$

其中,  $P(\mathbf{X}_k | \mathbf{z}_{1:k})$  为后验密度函数 (posterior density function, PDF)。粒子滤波是利用状态空间随机采样的粒子表示需要的 PDF,当粒子数目非常大时,这组具有权

值的粒子就接近状态变量实际 PDF。基于这些粒子和权值可实现目标状态的最优估计。由于直接从 PDF 分布中采样比较困难,当状态转移满足一阶马尔可夫过程时,重要采样可描述为:

$$\mathbf{X}_k^{(p)} \sim P_r(\mathbf{X}_k^{(p)} | \mathbf{X}_{k-1}^{(p)}) \quad (10)$$

其中,  $p$  为粒子序号,且  $0 \leq p \leq P$ ,  $P$  是粒子数目,  $P_r(\mathbf{X}_k^{(p)} | \mathbf{X}_{k-1}^{(p)})$  表示粒子先验分布。此时,粒子权值更新为:

$$w_k^{(p)} \propto w_{k-1}^{(p)} P_r(\mathbf{z}_k | \mathbf{X}_{k-1}^{(p)}) = w_{k-1}^{(p)} P_r(\mathbf{z}_k | \mathbf{X}_k^{(p)}) \quad (11)$$

式中:  $P_r(\mathbf{z}_k | \mathbf{X}_k^{(p)})$  表示似然函数。

为了准确提取说话人的位置信息,采用匀速模型 (constant velocity, CV) 对多特征观测量进行优化。设说话人运动的状态方程为:

$$\mathbf{X}_k = \mathbf{A}\mathbf{X}_{k-1} + \mathbf{u}_k \quad (12)$$

式中:  $\mathbf{X}_k = [x_k \ \dot{x}_k \ y_k \ \dot{y}_k]^T$  用来描述说话人的运动状态;

$\mathbf{A} = \begin{bmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{bmatrix}$  表示状态转移矩阵,  $\mathbf{B} = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix}$ ,  $T$  是采样间隔;

$\mathbf{u}_k$  表示过程噪声,其方差为  $\mathbf{\Gamma} = q_0 \mathbf{I}$ ,  $\mathbf{I} = \text{diag}(1, 1, 1, 1)$ ,  $q_0$  为过程噪声方差强度。与声源状态  $\mathbf{X}_k$  相关的第  $d$  组麦克风对的 TDOA 由下式给出:

$$\tau_k(\mathbf{X}_k) = \frac{(\|\mathbf{X}_{k,x} - m_1^{(d)}\| - \|\mathbf{X}_{k,x} - m_2^{(d)}\|)}{c} \quad (13)$$

其中,  $\mathbf{X}_{k,x}$  表示  $\mathbf{X}_k$  的坐标,  $m_1^{(d)}$ 、 $m_2^{(d)}$  为第  $d$  对麦克风的坐标,  $d = 1, 2, \dots, D$ ,  $D$  为麦克风对的个数。

### 2.2 多特征优化机制

针对单一特征定位性能差的问题,在粒子滤波的基础上,采用基于时延估计多假设模型,利用波束能量对多特征观测量进行优化。设  $k$  时刻麦克风对所有候选时延估计值为:

$$\tilde{\mathbf{y}}_k = [(\tilde{\mathbf{y}}_k^{(1)})^T, (\tilde{\mathbf{y}}_k^{(2)})^T, \dots, (\tilde{\mathbf{y}}_k^{(d)})^T]^T \quad (14)$$

其中,  $\tilde{\mathbf{y}}_k^{(d)}$  为第  $d$  组麦克风对的时延假设,即:

$$\tilde{\mathbf{y}}_k^{(d)} = [(\tilde{\mathbf{y}}_k^{(d,1)})^T, (\tilde{\mathbf{y}}_k^{(d,2)})^T, \dots, (\tilde{\mathbf{y}}_k^{(d,n)})^T]^T \quad (15)$$

式中:  $\tilde{\mathbf{y}}_k^{(d,n)}$  表示麦克风对  $d$  的第  $n$  个假设时延估计,  $1 \leq n \leq N$ ,  $N$  为假设时延个数。设  $H_n$  为  $D$  组麦克风对的真实时延,  $1 \leq n \leq N^D$ , 则  $H_n$  下麦克风对的时延估计为:

$$\tilde{\mathbf{y}}_k^{(\bar{n})} = [(\tilde{\mathbf{y}}_k^{(1,\bar{n}_1)})^T, \dots, (\tilde{\mathbf{y}}_k^{(2,\bar{n}_2)})^T, \dots, (\tilde{\mathbf{y}}_k^{(D,\bar{n}_D)})^T]^T \quad (16)$$

式中:  $\bar{n}_d$  表示麦克风对  $d$  的时延估计序号,且  $1 \leq \bar{n}_d \leq N$ 。将  $N^D$  组时延假设观测量分别代入式(5)并令  $L = D$ , 则每组时延对应的波束能量和为:

$$P_{\text{PAM}}(\tau_d) = \sum_{d=1}^D G(\tau_d) \quad (17)$$

选取  $N_k$  组最大波束能量对应的时延估计作为候选时延观测,记  $\hat{n}$  为其索引,  $1 \leq \hat{n} \leq N_k$ 。将  $N_k$  组时延观测分别代入式(8)进行粒子滤波,即:

$$\mathbf{X}_{k|k-1}^{(p)} = f_k(\mathbf{X}_{k-1}^{(p)}) + \mathbf{v}_{k-1} \quad (18)$$

$$w_k^{(p,\bar{n})} = w_k^{(p)} \cdot P_r(\tilde{\mathbf{y}}_k^{(\bar{n})} | \mathbf{X}_{k-1}^{(p)}, H_n) \cdot w_k^{(l)} \quad (19)$$

$$\hat{\chi}_k^{\hat{n}} = \sum_{p=1}^P w_k^{(p, \hat{n})} / \sum_n w_k^{(p, \hat{n})} \cdot \chi_{k|k-1}^{(p)} \quad (20)$$

式中:  $P_r(\bar{\mathbf{y}}_k^{(\hat{n})} | \chi_{k-1}^{(p)}, H_n)$  表示  $H_n$  下的似然函数;  $\hat{\chi}_k^{\hat{n}}$  为  $k$  时刻  $H_n$  下滤波估计值, 假设其对应的位置向量为  $\mathbf{l}_x$ , 由式(4) 可求得  $M$  个麦克风上的波束能量和为:

$$P_{\text{SBF}}(\mathbf{l}_x) = \int_{\Omega} \left| \sum_{i=1}^M X_i(\omega) \exp(j\omega \|\mathbf{l}_x - \mathbf{l}_i\| c^{-1}) \right|^2 d\omega \quad (21)$$

使上式波束能量最大所对应的时延估计为:

$$\bar{\mathbf{y}}_k = \bar{\mathbf{y}}_k^{(\hat{n}_k)} = [(\bar{\mathbf{y}}_k^{(1, \hat{n}_1)})^T, \dots, (\bar{\mathbf{y}}_k^{(D, \hat{n}_D)})^T]^T \quad (22)$$

其中,  $\hat{n}_k = \operatorname{argmax}_{\hat{n} \in \{1, 2, \dots, N_k\}} \{P_{\text{SBF}}(\mathbf{l}_x)\}$  为  $\bar{\mathbf{y}}_k$  在多假设测量  $\bar{\mathbf{y}}_k^{(\hat{n}_k)}$  中的索引,  $\bar{\mathbf{y}}_k^{(d, \hat{n}_d)}$  为麦克风对  $d$  的第  $\hat{n}_d$  个时延值。此时  $\bar{\mathbf{y}}_k$  对应的似然函数为:

$$P_r(\bar{\mathbf{y}}_k^{(\hat{n})} | \chi_{k-1}^{(p)}, H_n) = \mathcal{N}(\bar{\mathbf{y}}_k^{(\hat{n})}; \hat{\boldsymbol{\tau}}_{k|k-1}, \mathbf{R}) \quad (23)$$

式中:  $\hat{\boldsymbol{\tau}}_{k|k-1}$  为一步预测时延值, 按式(13) 和(18) 求得;  $\mathcal{N}(x; \bar{x}, \sigma^2)$  表示  $x$  服从高斯分布, 其中  $\bar{x}$  为均值,  $\sigma^2$  为方差。

为了对 SBF 和 SRP-PHAT 定位性能进行互补, 设特征向量距离函数为:

$$v_f = \|\mathbf{P}_f(\mathbf{l}) - \mathbf{P}_{f, \tau}(\mathbf{l})\| \quad (24)$$

式中:  $\mathbf{P}_{f, \tau}(\mathbf{l}) = \mathbf{P}_f(\mathbf{l}) \cdot \mathbf{P}_r(\bar{\mathbf{y}}_k | \hat{\chi}_k)$ ,  $\mathbf{P}_{f, \tau}(\mathbf{l})$  为  $\mathbf{P}_f(\mathbf{l})$  与时延特征的融合;  $\mathbf{P}_f(\mathbf{l})$  表示  $\mathbf{P}_{\text{SBF}}(\mathbf{l})$  或  $\mathbf{P}_{\text{SRP}}^c(\mathbf{l})$  特征。由于时延估计已经优化一次, 通过引入距离函数, 可得出优化后的波束输出能量为:

$$w_k^{(l)} = \mu \mathbf{P}_{\text{SBF}}(\mathbf{l}) + (1 - \lambda) \mathbf{P}_{\text{SRP}}^c(\mathbf{l}) \quad (25)$$

$$\lambda = \frac{v_{\text{SRP}}^2}{v_{\text{SBF}}^2 + v_{\text{SRP}}^2} \quad (26)$$

其中,  $\lambda$  为波束能量加权因子。将优化出的波束能量和  $N_k$  组候选测量分别融合并再次进行粒子滤波, 则对应波束能量最大的一组时延为最优时延。

### 2.3 优化机制实现

在多特征优化机制下, 多特征优化观测量提取步骤如下:

1) 初始化粒子集:  $k = 0$ , 采样  $\chi_0^{(p)} \sim P(\chi_0)$ , 即由  $P(\chi_0)$  采样得到  $\chi_0^{(p)}$ ,  $p = 1, 2, \dots, P$ 。

2) 多假设下滤波更新:  $k$  时刻, 选取  $N_k$  组使式(17) 最大的时延估计作为多假设测量集, 按式(18) ~ 式(20) 滤波, 求出使式(21) 最大时所对应的时延估计  $\bar{\mathbf{y}}_k$ 。此时,  $w_{k, \text{opt}}^{(l)} = \lambda_{\text{opt}} \mathbf{P}_{\text{SBF}}(\mathbf{l}) + (1 - \lambda_{\text{opt}}) \mathbf{P}_{\text{SRP}}^c(\mathbf{l})$ ,  $\lambda_{\text{opt}}$  根据式(23) 和(26) 优化。

3) 将  $N_k$  组测量对应的似然函数和  $w_{k, \text{opt}}^{(l)}$  分别融合进行滤波, 有:  $w_k^{(p)} = w_k^{(p)} \cdot P_r(\bar{\mathbf{y}}_k | \chi_{k-1}^{(p)}) \cdot w_{k, \text{opt}}^{(l)}$ ,  $w_k^{(p)} = w_k^{(p)} / \sum_{p \in P} w_k^{(p)}$ 。为防止粒子退化<sup>[26]</sup>, 设重采样后的粒子集为  $\bar{\chi}_k^{(p)}$ , 归一化权值为  $w_k^p$ 。

4) 多测量下状态更新:  $\hat{\chi}_k = \sum_{p=1}^P w_k^p \cdot \bar{\chi}_k^{(p)}$ 。将  $N_k$  组

$\hat{\chi}_k$  分别代入式(21) 求出  $N_k$  组波束能量, 对应波束能量最大的时延值即为最佳时延观测向量。

5) 递归:  $k = k + 1$  时, 重复步骤 2) ~ 5)。

## 3 IMM 粒子滤波改进算法

说话人运动可分解为说话人行走运动和说话人的头部运动, 对这两类运动分别采用 CV 模型和 RW 模型描述, 并在 IMM 算法框架下进行滤波。由于说话人的两类运动均具有随机性, 且 IMM 滤波性能取决于各子模型的性能, 因此, 本文对 CV 模型和 RW 模型进行了改进。设 IMM 算法中模型集的状态方程为:

$$\mathbf{x}_k^i = \mathbf{F}_{k-1}^i \mathbf{x}_{k-1}^i + \mathbf{v}_{k-1}^i \quad (27)$$

其中,  $\mathbf{F}_{k-1}^i$  为模型  $i$  ( $i = 1, 2$ )  $k - 1$  时刻状态转移矩阵,  $\mathbf{x}_k^i$  为  $k$  时刻模型  $i$  的状态向量,  $\mathbf{v}_{k-1}^i$  为过程噪声。模型集中 CV 模型的表达式如式(12), RW 模型参数设置见文献[20]。说话人机动或随机运动可视为说话人状态方程过程噪声的增加, 此时模型  $i$  的过程噪声方差可表示为:

$$\mathbf{P}_{k|k}^i = \mathbf{R}_i + \operatorname{diag}(q_{ix} I_2, q_{iy} I_2) \quad (28)$$

式中:  $q_{ix} = C |x_{k|k}^i - x_{k|k-1}^i| / T$ ,  $C$  是常数,  $q_{iy}$  表达式类似  $q_{ix}$ ,  $I_2 = \operatorname{diag}(1, 1)$ ;  $\mathbf{R}_i$  为模型  $i$  的初始过程协方差。过程噪声方差的自适应改变了粒子重要采样的分布概率, 使粒子更加接近说话人的真实运动状态。

IMM 算法的突出优点是使用模型信息并进行交互<sup>[27-28]</sup>, 各子模型似然函数通过新息及其协方差估计。由于 IMM 粒子滤波不具备一步预测功能, 不能按传统方法计算模型概率, 本文定义模型概率为:

$$L_k^i = L_{1,k}^i \cdot L_{2,k}^i \quad (29)$$

其中,  $L_{1,k}^i = \mathcal{N}(\bar{\mathbf{Z}}_k^i; \mathbf{0}, \mathbf{S}_{k|k}^i)$  是第  $i$  个模型的新息似然函数,  $\bar{\mathbf{Z}}_k^i$  和  $\mathbf{S}_{k|k}^i$  分别为  $k$  时刻模型  $i$  的新息及协方差;  $L_{2,k}^i = \mathcal{N}(\hat{\chi}_k^i; \hat{\chi}_{k-1}^i, \mathbf{P}_{k|k}^i)$  表示第  $i$  个模型的状态似然函数。其中,  $\mathbf{S}_{k|k}^i$  定义为:

$$\mathbf{S}_{k|k}^i = \sum_{n=0}^N [w_n^i (\bar{\mathbf{y}}_{n,k|k-1}^i - \hat{\mathbf{y}}_k^{i-}) (\bar{\mathbf{y}}_{n,k|k-1}^i - \hat{\mathbf{y}}_k^{i-})^T] \quad (30)$$

式中:  $w_n^i$  是粒子权值,  $\bar{\mathbf{y}}_{n,k|k-1}^i$  表示每个粒子的量测预估值;  $\hat{\mathbf{y}}_k^{i-}$  为  $N$  个粒子加权后的量测估计值。由式(29) 可知, 模型概率除了与新息似然函数有关外, 还与模型状态似然函数有关, 这两种似然函数的乘积反映了模型描述说话人运动状态性能的好坏。设  $k$  时刻 IMM 算法中模型集为  $M_k$ , 模型概率转移矩阵为  $\mathbf{\Pi}$ ,  $\pi_{ij} \in \mathbf{\Pi}$ , 模型  $i$  初始概率为  $\mu_{k-1}^i$ , 不同时刻模型集中各子模型过程方差不同。此时, IMM 粒子滤波改进算法为:

1) 模型条件重初始化: 对第  $j$  个模型, 其输入估计和

协方差分别为:

$$\hat{\chi}_{k-1|k-1}^j = \sum_{m_i \in M_{k-1}} \hat{\chi}_{k-1|k-1}^i \mu_{k-1}^{(i,j)}, \quad \forall m_i \in M_{k-1} \quad (31)$$

$$\hat{P}_{k-1|k-1} = \sum_{m_i \in M_{k-1}} [\mathbf{P}_{k-1|k-1}^i + \hat{\chi}_{k-1|k-1}^i (\mathbf{X}_{k-1|k-1}^i)^T] \mu_{k-1}^{(i,j)} \quad (32)$$

$$\mu_{k-1}^{(i,j)} = \pi_{ij} \mu_{k-1}^i / \mu_{k-1}^j \quad (33)$$

$$\mu_{k-1}^j = \sum_{m_i \in M_{k-1}} \pi_{ij} \mu_{k-1}^i \quad (34)$$

$$\tilde{\chi}_{k-1|k-1}^i = \hat{\chi}_{k-1|k-1}^i - \hat{\chi}_{k-1|k-1}^j \quad (35)$$

式中:  $\mu_{k-1}^j$ 、 $\mu_{k-1}^{(i,j)}$  分别表示模型预测概率和混合权重,  $m_i$  为  $k-1$  时刻的匹配模型。

2) 自适应滤波:

$$\Delta \mathbf{X} = \hat{\chi}_{k-1|k-1}^j - \hat{\chi}_{k-1|k-1}^i \quad (36)$$

$$\hat{\chi}_{k-1|k-1}^{(p,j)} = \hat{\chi}_{k-1|k-1}^{(p,i)} + \text{cvecrep}(\Delta \mathbf{X}, N) \quad (37)$$

其中,  $\hat{\chi}_{k-1|k-1}^{(p,j)}$  为  $k-1$  时刻模型  $j$  的粒子集,

$\text{cvecrep}(\Delta \mathbf{X}, N)$  表示向量  $\Delta \mathbf{X}$  重复  $N$  列。将  $\hat{\chi}_{k-1|k-1}^{(p,j)}$  和  $\hat{P}_{k-1|k-1}^j$  分别代入下一时刻与模型  $m_j$  相匹配的滤波器进行滤波, 模型  $m_j$  的过程方差按式(28) 计算, 可得各个滤波器各自状态估计及协方差分别为  $\hat{\chi}_{k|k}^j$  和  $\mathbf{P}_{k|k}^j$ 。

3) 模型概率实时更新: 对第  $i$  个模型, 概率描述为:

$$\mu_k^i = \frac{\mu_{k-1|k-1}^i L_k^i}{\sum_{m_j \in M_k} \mu_{k-1|k-1}^j L_k^j}, \quad \forall m_j \in M_k \quad (38)$$

其中,  $L_k^i$  按式(29) 计算。

4) 多模型输出融合: 系统状态次优估计值为:

$$\hat{\chi}_{k|k} = E(\mathbf{X}_k | \mathbf{y}_k) = \sum_{m_i \in M_k} \hat{\chi}_{k|k}^i \mu_k^i, \quad \forall m_i \in M_k \quad (39)$$

$$\mathbf{P}_{k|k} = \sum_{m_i \in M_k} [\mathbf{P}_{k|k}^i + (\hat{\chi}_{k|k}^i - \hat{\chi}_{k|k}) (\hat{\chi}_{k|k}^i - \hat{\chi}_{k|k})^T] \mu_k^i \quad (40)$$

## 4 数值仿真与实验验证

采用数值仿真验证不同声学环境和不同跟踪路径下所提算法的可行性, 再利用实测数据验证其跟踪效果。用于仿真的房间尺寸设置为  $6 \text{ m} \times 5 \text{ m} \times 2.7 \text{ m}$ , 室内平均反射系数设为 0.92。麦克风阵列摆放如图 1 所示, 图中黑点代表麦克风, 麦克风距地面和墙壁分别为 1.5 m、0.12 m。麦克风接收信号通过式(1) 获取, 式中房间冲激响应函数由能量衰减 image 模型产生<sup>[29]</sup>, 为了模拟说话人行走时头部转角变化, 叠加标准差  $\sigma = 0.02$  的高斯白噪声。仿真过程中, 采样频率设为 8 kHz, 窗函数

取汉明窗, 每帧长为 32 ms, 相邻两帧重叠 50%, 快速傅里叶变换长度设为 512。IMM 中各子模型过程噪声标准差取 0.07, 量测噪声标准差设为 0.031 6, 粒子数设为 300, 跟踪过程只考虑二维情况。采用位置平均均方根误差 (average root mean square error, ARMSE) 评估算法性能, 即<sup>[30]</sup>:

$$\text{ARMSE} = \frac{1}{I} \sum_{i=1}^I \sqrt{\frac{1}{K} \sum_{k=1}^K \|\hat{s}_{k,i} - s_k\|^2} \quad (41)$$

式中:  $I$  为 Monte Carlo 实验次数;  $K$  表示说话人状态个数;  $\hat{s}_{k,i}$  是第  $i$  次实验  $k$  时刻声源位置估计值;  $s_k$  是  $k$  时刻声源真实值;  $\text{ARMSE}$  表示说话人位置估计值偏离真实值的程度,  $\text{ARMSE}$  值越小, 跟踪结果越准确。当  $i=1$  时,  $\text{ARMSE}$  转化为均方根误差 (root mean square error, RMSE)。

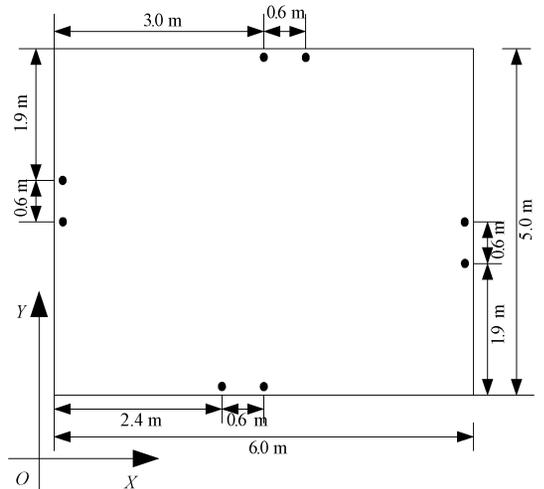


图 1 室内麦克风布置图

Fig. 1 Indoor microphone layout

### 4.1 抗混响抗噪声性能分析

采用一段自录时长约 4.75 s 的语音作为声源, 设声源从 (1.5 m, 1.5 m) 开始以速度 0.8 m/s 和 0.5 m/s 分别沿 X、Y 轴运动, 沿声源轨迹将接收信号分成 147 帧。为便于对比, 构建 5 种不同观测特征跟踪给定轨迹。第 1 种采用 SBF 与 TDE 构建观测测量, 即 SBFTDE 算法, 第 2 种采用 SRP 与 TDE 融合跟踪<sup>[15]</sup>, 即 SRPTDE 算法, 第 3 种采用 SBF 和 SRP 相结合定位<sup>[25]</sup>, 即 SBFSRP 算法, 第 4 种采用多特征直接融合跟踪, 即 MF 算法, 第 5 种采用本文所提算法, 即 MFO 算法。MFO 时间复杂度为  $O(N^D + N_k P n_x^2)$ , 空间复杂度为  $O(N^D + N_k (P + n_x) n_x)$ , 其中  $n_x$  为状态维数。实验过程中均采用 CV 模型, 在不同混响时间和不同 SNR 下分别做 30 次 Monte Carlo 试验, 5 种观测特征的位置平均 RMSE 如图 2~3。

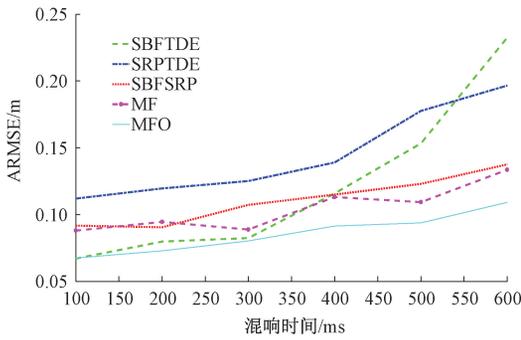


图 2 位置平均 RMSE 随混响时间的变化  
Fig. 2 Average RMSE versus reverberation time

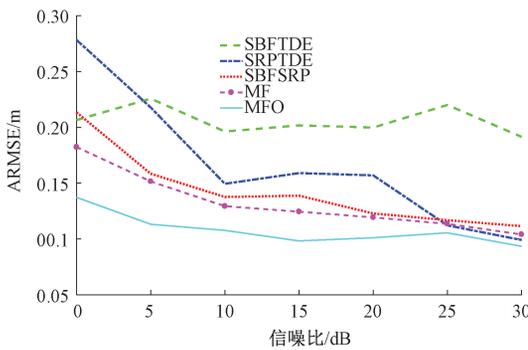


图 3 位置平均 RMSE 随信噪比的变化  
Fig. 3 Average RMSE versus SNR

图 2 是  $SNR = 10$  dB 时位置平均 RMSE 随混响时间变化的曲线。从图中可看出,在强混响下,SBFTDE 和 SRPTDE 跟踪性能明显下降,SBFSRP 跟踪效果好于 SBFTDE 和 SRPTDE,反映了 TDE 鲁棒性能差。MF 位置平均 RMSE 低于 SBFTDE、SRPTDE 以及 SBFSRP,这说明多特征能较好地抑制混响的影响。MFO 因采用优化机制,在 100~600 ms 混响时间范围内,其跟踪精度最高,表明了 MFO 抗混响性能强。

图 3 是在  $T_{60} = 600$  ms 时计算的位置平均 RMSE。图中 SBFTDE 和 SRPTDE 跟踪性能受 TDE 不能有效抑制背景噪声的影响变差。随着噪声减弱,5 种算法受混响影响变大,融合 SRP 特征的 MF、SBFSRP 和 SRPTDE 跟踪性能在高信噪比时较为接近,而 SBFTDE 跟踪性能最差。采用多特征优化机制的 MFO 位置平均 RMSE 最小,取得了比其他算法更高的抗噪声性能。

#### 4.2 不同运动模式跟踪分析

跟踪路径 1:在  $T_{60} = 300$  ms、 $SNR = 20$  dB 下,将一段自录时长约 7.28 s 的语音作为声源,假定声源从点 (1.0 m, 1.0 m) 开始作速度为 1 m/s 运动,停止于点 (5.0 m, 1.0 m),在 (1.0 m, 2.6 m)、(5.0 m, 2.6 m) 作转向机动。采用 4 种模型对说话人进行 30 次 Monte Carlo

跟踪。第 1 种采用 L 模型<sup>[19]</sup>,第 2 种采用 RW 模型<sup>[20]</sup>,第 3 种采用 CV 模型,第 4 种采用改进的 IMM,即鲁棒跟踪 (robust tracking, RT)模型,观测量均采用 MFO,4 种算法分别称作 MFOL、MFORW、MFOCV 和 MFORT,4 种算法位置平均 RMSE 如图 4 所示。跟踪路径 2:声源作速率为 0.9 m/s 的逆时针半圆弧运动,始点为 (1.0 m, 2.5 m),终点为 (5.0 m, 2.5 m),其余设置同跟踪路径 1,4 种算法位置平均 RMSE 如图 5 所示。如表 1 所示为不同路径下 4 种算法对 Monte Carlo 次数和仿真时间点数的求取 ARMSE。

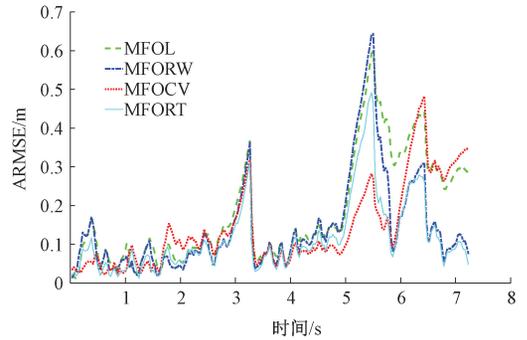


图 4 转向机动位置估计 RMSE  
Fig. 4 RMSE of location estimation for steering maneuver

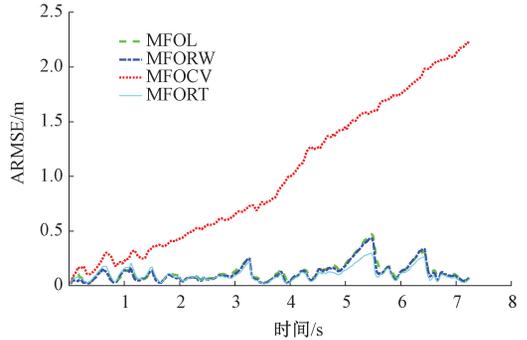


图 5 圆弧机动位置估计 RMSE  
Fig. 5 RMSE of location estimation for circular arc

表 1 4 种跟踪模型位置平均均方根误差  
Table 1 Position average root mean square error of the four tracking models

模型	转向 位置 ARMSE/m	圆弧 位置 ARMSE/m
MFOL	0.182 4	0.128 4
MFORW	0.141 0	0.116 6
MFOCV	0.140 8	0.997 7
MFORT	0.115 4	0.102 5

从图 4 和表 1 可知,当说话人转向时,MFORT 跟踪误差最低。由于 L、CV 模型均不能刻画说话人转向,

导致 MFOL、MFOCV 跟踪性能差。MFORW 尽管收敛速度快,但不适合描述匀速运动。由表 1 可知,MFORT 位置平均 RMSE 相对于 MFOCV 减少了 18% 左右,说明 IMM 改进算法能正确描述转向机动。从图 5 和表 1 可看出,圆弧机动时 MFOCV 失跟,说明 CV 模型不能匹配变加速运动,其他 3 种算法基本能跟踪圆弧运动,但 MFORT 跟踪精度高。表 1 结果显示,圆弧机动时 MFORT 相对于 MFORW 位置平均 RMSE 减少了约 12%。两种不同机动仿真结果表明,MFORT 的跟踪精度高于其他 3 种算法,改进的 IMM 提高了 MFO 对不同运动模式的鲁棒性,但当 IMM 中两个子模型都不适用说话人机动时,IMM 模型跟踪失效。

4.3 实际实验验证

第 4.1 和 4.2 节仿真结果验证了所提算法理论分析的可行性,为了更充分地评估算法实际跟踪的有效性,在真实房间内进行了实际测试实验。房间尺寸为 6.74 m×5.69 m×4.05 m,传声器阵列摆放参如图 1 所示,4 组传声器对的中心坐标分别设置在 (3.0 m, 0 m)、(5.4 m, 2.7 m)、(2.4 m, 4.8 m)、(0 m, 2.1 m)。通过 MPA436A 传声器读取语音信号,利用 DEWE-51-PCI-16 设备采集传声器阵列读取信号。实际测试过程中,说话人按预设路径运动:1) 由点 (2.1 m, 4.18 m) 开始沿 Y 轴以 -0.7 m/s 速度运动,在 (2.1 m, 1.4 m) 作 90°转向,终于点 (4.2 m, 1.4 m);2) 作速率为 1 m/s 的逆时针半圆弧运动,始点为 (3.6 m, 4.2 m),止点为 (3.2 m, 0.65 m)。室内混响时间实测约 819 ms,背景噪声级实测约 43 dBA。从第 4.1 节中选取 MFO、MF 作为观测特征,采用 4.2 节的 4 种不同跟踪模型对实际实验进行跟踪。第 1 种采用 MFO 和 L 模型,即 MFOL 算法,第 2 种采用 MFO 和 RW 模型,即 MFORW 算法,第 3 种采用 MFO 和 CV 模型,即 MFOCV 算法,第 4 种采用 MF 和 RT 模型,即 MFRT 算法,第 5 种采用所提算法 MFORT。5 种算法的位置 RMSE 如图 6~9,位置平均 RMSE 如表 2 所示。

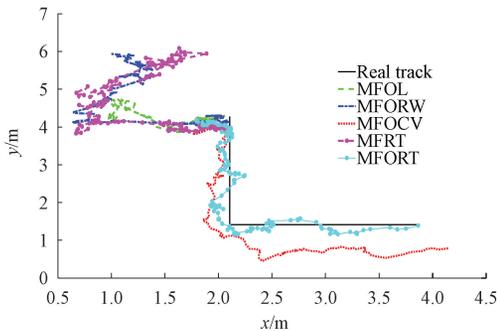


图 6 转向轨迹

Fig. 6 Steering track

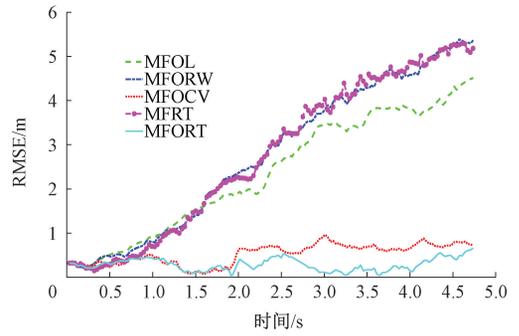


图 7 转向轨迹位置估计 RMSE

Fig. 7 RMSE of location estimation for steering track

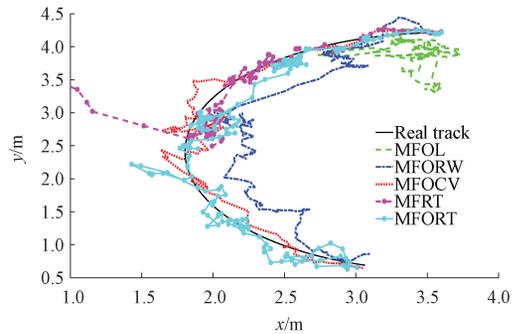


图 8 圆弧轨迹

Fig. 8 Circular arc track

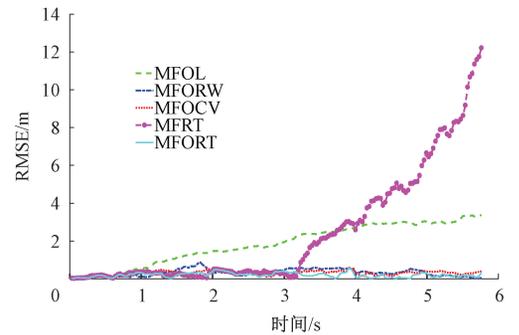


图 9 圆弧轨迹位置估计 RMSE

Fig. 9 RMSE of location estimation for circular arc track

表 2 实测五种算法位置平均均方根误差

Table 2 Position average root mean square error of the five algorithms for the real trajectory

算法	转向	圆弧
	位置 ARMSE/m	位置 ARMSE/m
MFOL	2.371 8	2.033 2
MFORW	2.558 5	0.467 1
MFOCV	0.525 2	0.317 9
MFRT	2.768 6	3.856 0
MFORT	0.288 4	0.250 8

由图6、7和表2可知,转向机动时,MFORT位置RMSE最小。L模型、RW模型均不能与声源运动模式匹配,跟踪发散。CV模型收敛速度慢,跟踪性能有限。在IMM改进算法下,MFORT比MFRT位置平均RMSE减少了约89.6%,这说明多特征直接融合不能准确提取说话人位置信息。

图8和9以及表2表明,圆弧运动时,MFRT跟踪发散。L模型与说话人运动模式失配,MFOL呈发散趋势。RW模型和CV模型在与MFO相结合条件下,基本能描述说话人圆弧运动,但跟踪精度不高。MFORT克服了其他算法的缺陷,能有效跟踪圆弧运动,其位置RMSE最小。

## 5 结 论

本文建立了基于时延估计多假设模型的多特征优化机制和鲁棒跟踪模型,解决了复杂声学环境下单特征跟踪性能差的问题。采用多假设时延估计模型和多特征优化机制从混响噪声中准确地提取了多特征观测量。通过对说话人运动进行分解,建立了说话人随机运动的鲁棒跟踪模型,提高了多特征优化机制检测的稳健性。将多特征观测量和鲁棒跟踪模型相结合,能有效跟踪复杂场景下随机运动的声源。仿真分析和实际测试结果表明,所提算法跟踪精度高,数值稳定性强,能有效处理说话人的不同运动模式,适用于低信噪比强混响下随机运动的单个声源。对于不同声源高度所带来误差时的所提算法还有待进一步研究。

## 参考文献

[1] ZAKEN O B, KUMAR A, TOURBABIN V, et al. Neural-network-based direction-of-arrival estimation for reverberant speech-the importance of energetic, temporal, and spatial information[J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2024, 32: 1298-1309.

[2] GARCÍA-BARRIOS G, LATORRE IGLESIAS E, GUTIÉRREZ-ARRIOLA J M, et al. Exploiting spatial diversity for increasing the robustness of sound source localization systems against reverberation[J]. *Applied Acoustics*, 2023, 202: 1-11.

[3] WANG L, CAVALLARO A. Deep-learning-assisted sound source localization from a flying drone[J]. *IEEE Sensors Journal*, 2022, 22(21): 20828-20838.

[4] ZHU X C, ZHANG H, FENG H T, et al. Ifan: An

icosahedral feature attention network for sound source localization[J]. *IEEE Transactions on Instrumentation and Measurement*, 2024, 73: 1-13.

- [5] DESAI D, MEHENDALE N. A review on sound source localization systems[J]. *Archives of Computational Methods in Engineering*, 2022, 29(7): 4631-4642.
- [6] JI X Y, CHENG L, ZHAO H F. Multipath time-delay estimation with impulsive noise via Bayesian compressive sensing[J]. *IEEE Signal Processing Letters*, 2023, 30: 937-941.
- [7] 卢焯华, 薛齐凡, 刘志恩, 等. 基于SincNet增强的时延估计声源定位算法研究[J]. *武汉理工大学学报*, 2023, 45(10): 127-134.
- LU CH H, XUE Q F, LIU ZH EN, et al. Research on time delay estimation based sound source localization algorithm enhanced by SincNet[J]. *Journal of Wuhan University of technology*, 2023, 45(10): 127-134.
- [8] 余萍, 杨乘, 王紫薇, 等. 基于联合加权的广义二次相关时延估计算法[J]. *计算机仿真*, 2023, 40(3): 400-404.
- YU P, YANG CH, WANG Z W, et al. Generalized secondary correlation delay estimation algorithm based on joint weighting[J]. *Computer Simulation*, 2023, 40(3): 400-404.
- [9] 唐俊, 汪照, 白宇田. 基于稀疏过零点信息的抗幅值失真时延估计方法[J]. *天津大学学报(自然科学与工程技术版)*, 2022, 55(2): 211-220.
- TANG J, WANG ZH, BAI Y T. Anti-amplitude distortion time delay estimation method based on sparse cross-zero information[J]. *Journal of Tianjin University (Science and Technology)*, 2022, 55(2): 211-220.
- [10] 曾庆宁, 苏盼, 王红丽, 等. 基于互相关序列和BP网络的声源定位算法[J]. *科学技术与工程*, 2023, 23(16): 6945-6951.
- ZENG Q N, SU P, WANG H L, et al. Sound source localization algorithm based on cross-correlation sequence and BP network[J]. *Science Technology and Engineering*, 2023, 23(16): 6945-6951.
- [11] 王帅帅, 贾学东, 陈国军, 等. 一种基于扩频声波的室内定位方法及实现[J]. *测绘科学技术学报*, 2020, 37(2): 221-225.
- WANG SH SH, JIA X D, CHEN G J, et al. An indoor positioning method and implementation based on spread spectrum acoustic wave[J]. *Journal of Geomatics*

- Science and Technology, 2020, 37(2): 221-225.
- [12] FALLON M F, GODSILL S. Acoustic source localization and tracking using track before detect [J]. IEEE Transactions on Audio, Speech, and Language Processing, 2010, 18(6): 1228-1242.
- [13] BOORA R, DHULL S K. Iterative modified SRP-PHAT with adaptive search space for acoustic source localization[J]. IETE Technical Review, 2022, 39(1): 28-36.
- [14] CAKMAK B, DIETZEN T, ALI R, et al. A distributed steered response power approach to source localization in wireless acoustic sensor networks [C]. International Workshop on Acoustic Signal Enhancement (IWAENC). Germany, Bamberg: IEEE, 2022:1-5.
- [15] DING X F, WANG C L, LIU H P. FRL: Fast and reconfigurable accelerator for distributed sound source localization[J]. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 2022, 41(11): 3922-3933.
- [16] DIETZEN T, DE SENA E, VAN WATERSCHOOT T. Low-complexity steered response power mapping based on Nyquist-Shannon sampling [C]. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA). America, New Paltz: IEEE, 2021: 206-210.
- [17] SHEN X, LU T T, ZHANG H, et al. A high resolution DOA estimation based on random array in underwater multipath environment[C]. 2023 IEEE 3rd International Conference on Power, Electronics and Computer Applications (ICPECA). China, Shenyang: IEEE, 2023: 220-225.
- [18] YAN J J, ZHAO W L, WU Y I, et al. Indoor sound source localization under reverberation by extracting the features of sample covariance[J]. Applied Acoustics, 2023, 210: 1-9.
- [19] ONG J, VO B T, NORDHOLM S. Blind separation for multiple moving sources with labeled random finite sets[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2021, 29: 2137-2151.
- [20] LEVY A, GANNOT S, HABETS E A P. Multiple-hypothesis extended particle filter for acoustic source localization in reverberant environments [J]. IEEE Transactions on Audio, Speech, and Language Processing, 2011, 19(6): 1540-1555.
- [21] PAN X, BAO Y, ZHU Y T, et al. Deconvolved conventional beamforming and adaptive cubature Kalman filter based distant speech perception system[J]. IEEE Access, 2020, 8: 187948-187958.
- [22] WANG R F, CHEN Z, YIN F L. Speaker tracking based on distributed particle filter and interacting multiple model in distributed microphone networks[J]. Applied Acoustics, 2021, 174: 1-12.
- [23] DONG X D, ZHANG X F, ZHAO J, et al. Multi-maneuvering sources DOA tracking with improved interactive multi-model multi-Bernoulli filter for acoustic vector sensor (AVS) array[J]. IEEE Transactions on Vehicular Technology, 2021, 70(8): 7825-7838.
- [24] WU K, KHONG A W H. Acoustic source tracking in reverberant environment using regional steered response power measurement [C]. 2013 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA 2013). Kaohsiung, 2013:1-6.
- [25] 蔡卫平, 吴镇扬. 一种基于粒子滤波的鲁棒声源跟踪算法[J]. 电子测量与仪器学报, 2010, 24(5): 407-413.
- CAI W P, WU ZH Y. Robust acoustic source tracking algorithm based on particle filtering [J]. Journal of Electronic Measurement and Instrument, 2010, 24(5): 407-413.
- [26] ARULAMPALAM M S, MASKELL S, GORDON N, et al. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking [J]. IEEE Transactions Signal Processing, 2002, 50(2): 174-188.
- [27] 刘望生, 潘海鹏, 王明环. 混响噪声下声源定位与跟踪的多特征自适应 IMM 粒子滤波算法[J]. 控制理论与应用, 2023, 40(3): 477-484.
- LIU W SH, PAN H P, WANG M H. An adaptive imm particle filter algorithm based on multi-feature for sound source tracking in reverberant and noisy environments[J]. Control Theory & Applications, 2023, 40(3): 477-484.
- [28] YOUNG W, HUANG Y L, MYUNG H. Robust localization using IMM filter based on skew Gaussian-Gamma mixture distribution in mixed LOS/NLOS condition[J]. IEEE Transactions on Instrumentation and Measurement, 2020, 69(7): 5166-5182.
- [29] LEHMANN E A, JOHANSSON A M. Prediction of energy decay in room impulse responses simulated with an image-source model[J]. Journal of the Acoustical Society

of America, 2008, 124(1): 269-277.

- [30] ZHOU N, LIU Q Y, YANG Y C, et al. An indoor positioning algorithm based on particle filter and neighbor-guided particle optimization for wireless sensor networks[J]. IEEE Transactions on Instrumentation and Measurement, 2024, 73: 1-16.

## 作者简介



**刘望生** (通信作者), 2003 年于长春工业大学获得硕士学位, 2008 年于中国舰船研究院获得博士学位, 现为浙江理工大学副教授, 主要研究方向为声源定位与跟踪、信号检测与处理。

E-mail: lwsh22@hotmail.com

**Liu Wangsheng** (Corresponding author) received his M. Sc. degree in 2003 from Changchun University of Technology,

received his Ph. D. degree in 2008 from China Ship Research Institute. Now he is an associate professor in Zhejiang Sci-Tech University. His main research interests include sound source location and tracking, signal detection and processing.



**刘艳梅**, 2001 年于沈阳航空工业学院获得学士学位, 2004 年于浙江大学获得硕士学位, 现为浙江理工大学讲师, 主要研究方向为信号处理、智能感知。

E-mail: lmail80@163.com

**Liu Yanmei** received her bachelor's degree from Shenyang Institute of Aeronautics and Astronautics in 2001 and her master's degree from Zhejiang University in 2004. She is currently a lecturer in Zhejiang Sci-Tech University. Her main research interests include signal processing and intelligent perception.